

**UNIVERSIDADE DE LISBOA
INSTITUTO DE EDUCAÇÃO**



**PLANEAMENTO ESTATÍSTICO
E ANÁLISE DE DADOS
COM ALUNOS DE 8.º ANO**

Cristina Maria Martins Roque

Dissertação

Mestrado em Educação

Área de especialização em Didática da Matemática

2013

**UNIVERSIDADE DE LISBOA
INSTITUTO DE EDUCAÇÃO**



**PLANEAMENTO ESTATÍSTICO
E ANÁLISE DE DADOS
COM ALUNOS DE 8.º ANO**

Cristina Maria Martins Roque

**Dissertação orientada pelo Professor Doutor João Pedro
Mendes da Ponte**

Mestrado em Educação

Área de especialização em Didática da Matemática

2013

Trabalho realizado no âmbito do Projeto Desenvolver a literacia estatística: Aprendizagem do aluno e formação do professor, financiada pela FCT- Fundação para a Ciência e Tecnologia (contrato PTDC/CPE-CED/117933/2010).

Resumo

Esta investigação tem como principal objetivo conhecer as aprendizagens realizadas por alunos de 8.º ano, no quadro de uma experiência de ensino que visa desenvolver a capacidade de planeamento estatístico e de análise de dados, dando especial atenção à análise, planeamento e realização de estudos estatísticos envolvendo questões significativas para os alunos. Pretendo também identificar e compreender as dificuldades demonstradas pelos alunos, com o intuito de melhorar o processo ensino-aprendizagem. Este estudo constituiu uma investigação sobre a minha prática profissional, segue uma metodologia de investigação qualitativa e interpretativa, utilizando estudos de caso. A recolha de dados foi realizada numa turma do 8.º ano, sendo estudada a própria turma e, de forma mais aprofundada, dois alunos, Ana e Rui, objetos de estudo de caso. A recolha de dados resultou de observação participante, recolha documental e, registos áudio e/ou vídeo das aulas e das entrevistas.

Os resultados mostram que os alunos foram capazes de distinguir entre população e amostra, ponderar elementos que podem afetar a representatividade de uma amostra em relação à população e identificar e minimizar possíveis fontes de enviesamento na recolha de dados. Além disso, os alunos evoluíram de modo significativo ao nível da formulação de questões de pesquisa e da adequação do plano de recolha de dados com vista a obter resposta para determinado problema estatístico. Consta-se que os alunos melhoraram a capacidade de análise de dados ao apresentarem uma progressão positiva no que respeita à construção, análise e interpretação de diferentes representações dos dados, tendo em consideração a sua adequação ao tipo de dados e ao mobilizar de modo articulado, as medidas estatísticas centrais e/ou de dispersão com as características gerais da distribuição representada e evidenciaram a utilização das medidas estatísticas na interpretação dos dados. Finalmente, no que se refere à generalização das conclusões à população, foi evidente o conflito que surge entre a representatividade e a variabilidade do processo de amostragem, pelo que no momento de estabelecer inferências as respostas dos alunos ficaram marcadas por: (i) inferências baseadas nas próprias opiniões e ignorando os dados dos estudos realizados, (ii) inferências baseadas nos dados, quando intuem que o processo de amostragem terá capturado a variabilidade populacional.

Palavras-chave: *planeamento estatístico, formulação de questões, amostragem, análise de dados.*

Abstract

This investigation has as a main goal the acknowledgement of the learning achieved by students of the 8th grade, in the framework of a teaching experience which seeks to develop the capacity of statistical planning and data analysis focusing especially on analysis, planning and conducting of statistical studies involving significant questions for the students. It is aimed towards identifying and understanding the difficulties presented by the students, with the intention of bettering the teaching-learning process. The study was based on an investigation of my own professional practice, following a methodology of qualitative and interpretative investigation, using case studies. The data collection was carried out in a class of the 8th grade. The whole class was studied but more closely two students, Ana and Rui, were object of the case study. The data collection came as a result of participant observation, documentation, and audio and/or video recordings of the classes and of the interviews.

The results obtained show that the students were capable of making a distinction between the population and the sample, as well as to consider elements which could affect the representativeness of the sample in what regards the population and identify and minimize possible sources of bias in the data collection. In addition, students improved significantly in what regards formulation of research questions and adequacy of the data collection plan in order to seek answers to a given statistical problem. It is noted that the students strengthened their capacity of data analysing while presenting positive improvement in what concerns structuring, analysis and interpretation of different data representations while taking into account the adequacy of the type of the data and while mobilising in a coordinated way, the main statistical measures and/or of dispersion with the general characteristics of the represented distribution and demonstrated the use of statistical measures in the data interpretation. Finally, in what regarded generalizing the findings to the population, an evident conflict emerged between the representativeness and the variability of the sampling process, for which in the moment to make inferences the answers of the students were marked by the following: (i) inferences were based on personal opinions while ignoring data from the study made, (ii) inferences were based on data, when intuitively felt that the sampling process matches the population variability.

Key-words: *statistical planning, forming questions, sampling, data analysis.*

Agradecimentos

Ao Professor João Pedro da Ponte, que me acompanhou nos momentos mais determinantes da vida profissional, estágio e concretização deste trabalho. Muito obrigada por ter sido sempre incansável e disponível, pela confiança e incentivo, pelas sugestões e críticas pertinentes, assim como pelos seus ensinamentos.

Aos meus colegas de escola que facilitaram e apoiaram a realização deste trabalho, em particular aos colegas da Direção por me terem possibilitado as condições necessárias à sua concretização.

A todos os alunos que participaram neste trabalho, em particular à Ana e ao Rui, pela sua disponibilidade, entusiasmo e colaboração, sem os quais, não teria sido possível concretizar este trabalho.

Aos amigos, pelo apoio e amizade.

A todos aqueles, que diretamente ou indiretamente, estiveram ao meu lado e contribuíram para a concretização deste trabalho.

E por fim, à minha família, Paulo e Íris, pelos momentos privados da minha presença, pelo carinho, incentivo e paciência que foram fundamentais na consecução de mais uma etapa importante da minha vida.

Em especial, aos meus maravilhosos pais e fantástica irmã, que me encorajaram e incentivaram a completar este trabalho!

Índice

CAPÍTULO 1	1
INTRODUÇÃO	1
1.1. <i>Motivações pessoais</i>	1
1.2. <i>Orientações curriculares para o ensino da Estatística</i>	5
1.3. <i>Objetivos e questões de investigação</i>	10
1.4. <i>Organização do estudo</i>	12
CAPÍTULO 2	13
QUADRO CONCEPTUAL DO ESTUDO.....	13
2.1. <i>Ensino da Estatísticas e seus objetivos</i>	13
2.2. <i>Ciclo de investigação estatístico</i>	14
2.2.1. <i>Formulação de questões estatísticas</i>	16
2.2.2. <i>Planeamento estatístico</i>	22
2.2.3. <i>Análise exploratória de dados</i>	27
2.2.4. <i>Desenvolver e avaliar inferências e previsões baseadas nos dados</i>	33
2.3. <i>Dificuldades na aprendizagem da Estatística</i>	38
2.3.1. <i>Dificuldades ao nível dos conceitos</i>	39
2.3.2. <i>Dificuldades ao nível de tabelas e gráficos estatísticos</i>	44
2.4. <i>Avaliação em Estatística</i>	48
CAPÍTULO 3	54
UNIDADE DE ENSINO	54
3.1. <i>Enquadramento curricular e objetivos</i>	54
3.2. <i>Tarefas</i>	56
3.3. <i>Dinâmica da sala de aula</i>	59
3.4. <i>Avaliação dos alunos</i>	62
CAPÍTULO 4	63
METODOLOGIA DE INVESTIGAÇÃO	63
4.1. <i>Opções metodológicas gerais e design</i>	63
4.2. <i>Participantes</i>	66
4.3. <i>Recolha de dados</i>	68
4.3.1. <i>Pedidos de autorização</i>	68
4.3.2. <i>Processo de recolha de dados</i>	68
4.3.3. <i>Instrumentos de recolha de dados</i>	69
4.4. <i>Análise de dados</i>	72
CAPÍTULO 5	74
DESEMPENHO DA TURMA	74
5.1. <i>Estrutura das aulas e materiais de apoio aos alunos</i>	74
5.2. <i>A realização das tarefas</i>	76
5.2.1. <i>Tarefa 1 – População e amostra</i>	77
5.2.2. <i>Tarefa 2 – Limpar os dados</i>	91
5.2.3. <i>Tarefa 3 – As vendas estão a correr bem?</i>	96
5.2.4. <i>Tarefa 4 – Um estudo na escola</i>	105
5.2.5. <i>Tarefa 5 – A frequência das vogais na língua portuguesa</i>	134
5.2.6. <i>Tarefa 6 – Previsões</i>	142
5.3. <i>Desempenho dos alunos nos testes</i>	153

5.3.1. Planeamento estatístico	153
5.3.2. Análise de dados.....	155
5.3.3. Síntese	161
5.4. <i>Balanço global das aprendizagens realizadas</i>	166
CAPÍTULO 6.....	173
O CASO DE ANA	173
6.1. <i>Apresentação</i>	173
6.2. <i>Capacidades de planeamento estatístico e análise de dados no início da unidade de ensino</i> .	174
6.2.1. Planeamento estatístico	174
6.2.2. Análise de dados.....	176
6.3. <i>Capacidades de planeamento estatístico e análise de dados durante a unidade de ensino</i>	179
6.3.1. Planeamento estatístico	180
6.3.2. Análise de dados.....	183
6.4. <i>Capacidades de planeamento estatístico e análise de dados depois da unidade de ensino</i>	187
6.4.1. Planeamento estatístico	187
6.4.2. Análise de dados.....	190
6.5. <i>Síntese global</i>	196
CAPÍTULO 7.....	198
O CASO DE RUI.....	198
7.1. <i>Apresentação</i>	198
7.2. <i>Capacidades de planeamento estatístico e análise de dados no início da unidade de ensino</i> .	199
7.2.1. Planeamento estatístico	199
7.2.2. Análise de dados.....	202
7.3. <i>Capacidades de planeamento estatístico e análise de dados durante a unidade de ensino</i>	205
7.3.1. Planeamento estatístico	205
7.3.2. Análise de dados.....	208
7.4. <i>Capacidades de planeamento estatístico e análise de dados depois da unidade de ensino</i>	212
7.4.1. Planeamento estatístico	212
7.4.2. Análise de dados.....	216
7.5. <i>Síntese global</i>	223
CAPÍTULO 8.....	225
CONCLUSÃO.....	225
8.1. <i>Síntese do estudo</i>	225
8.2. <i>Conclusões do estudo</i>	227
8.2.1. Planeamento estatístico	227
8.2.2. Análise de dados.....	230
8.3. <i>Reflexão final</i>	234
REFERÊNCIAS	238
ANEXOS	245

Índice de Anexos

Anexo 1 – Pedidos de autorização para realização do estudo.....	246
Anexo 2 – Planificação da unidade de ensino.....	248
Anexo 3 – Teste inicial e respetiva matriz de objetivos.....	249
Anexo 4 – Protocolo das entrevistas realizadas.....	254
Anexo 5 – Tarefa da 1ª entrevista e respetiva matriz de objetivos.....	255
Anexo 6 – Tarefa da 2ª entrevista e respetiva matriz de objetivos.....	261
Anexo 7 – Tarefa 1: População e amostra.....	267
Anexo 8 – Folheto da oferta formativa da escola – esclarecimento para o item 6 da tarefa 1	270
Anexo 9 – Acetatos utilizados na sistematização da tarefa 1.....	272
Anexo 10 – Tarefa 2: Limpar dados.....	274
Anexo 11 – Tarefa 3: As vendas estão a correr bem?.....	275
Anexo 12 – Tarefa 4: Um estudo na escola.....	276
Anexo 13 – Tarefa 5: Frequência das vogais na língua portuguesa.....	277
Anexo 14 – Tarefa 6: Previsões.....	281
Anexo 15 – Teste final e respetiva matriz de objetivos.....	282
Anexo 16 – Análise dos testes aplicados.....	288

Índice de Quadros

Quadro 1. Análise das questões estatísticas formuladas (Arnold, 2008).	19
Quadro 2. Questões moderadoras (Arnold, 2008).	20
Quadro 3. Categorias do desenvolvimento do conceito de amostra (Watson & Moritz, 2000). 24	
Quadro 4. Questões para apoiar a leitura de representações gráficas (Rudenstein & Thompson, 2012).	31
Quadro 5. Especificações de como determinadas tarefas mobilizam as componentes do RII (Zieffler, Garfield, delMas & Reading, 2008).....	38
Quadro 6. Exemplificação do processo de avaliação (Garfield, 1994).....	50
Quadro 7. Categorias de raciocínio correto e de equívocos no SRA.....	53
Quadro 8. Sequência de tarefas da unidade de ensino.	60
Quadro 9. Respostas apresentadas pelos alunos ao item 1 da tarefa 1.....	81
Quadro 10. Respostas apresentadas pelos alunos ao item 2 da tarefa 1.....	83
Quadro 11. Respostas apresentadas pelos alunos ao item 3 da tarefa 1.....	83
Quadro 12. Respostas apresentadas pelos alunos ao item 4 da tarefa 1.....	85
Quadro 13. Respostas apresentadas pelos alunos ao item 5 da tarefa 1.....	85
Quadro 14. Respostas apresentadas pelos alunos ao item 6 a) da tarefa 1.....	86
Quadro 15. Respostas apresentadas pelos alunos ao item 6b) da tarefa 1.....	88
Quadro 16. Inquérito construído pelo grupo turma.	89
Quadro 17. Resumo das respostas ao ponto dois.....	100
Quadro 18. Respostas apresentadas pelos alunos ao item 1 da tarefa 4.....	108
Quadro 19. Respostas apresentadas pelos alunos aos itens 2 e 3 da tarefa 4.	110
Quadro 20. Respostas apresentadas pelos alunos ao item 4 da tarefa 4.....	111
Quadro 21. Respostas apresentadas pelos alunos ao item 5 da tarefa 4.....	111

Quadro 22. Respostas apresentadas pelos alunos ao item 6 da tarefa 4.....	116
Quadro 23. Respostas apresentadas pelos alunos ao item 7 (género) da tarefa 4.	116
Quadro 24. Respostas apresentadas pelos alunos ao item 7 (altura) da tarefa 4.....	116
Quadro 25. Respostas apresentadas pelos alunos ao item 7 (envergadura) da tarefa 4.....	116
Quadro 26. Respostas apresentadas pelos alunos ao item 7 (nº calçado) da tarefa 4.....	117
Quadro 27. Respostas apresentadas pelos alunos ao item 8 da tarefa 4.....	118
Quadro 28. Respostas apresentadas pelos alunos ao item 9 da tarefa 4.....	120
Quadro 29. Nível de cumprimento de alguns aspetos relativos à construção do histograma - item 10 da tarefa 4.....	123
Quadro 30. Nível de cumprimento de alguns aspetos de análise na redação do relatório - item 11 da tarefa 4.	127
Quadro 31. Respostas apresentadas pelos alunos ao item 2 da tarefa 5.....	136
Quadro 32. Respostas apresentadas pelos alunos ao item 3 da tarefa 5.....	138
Quadro 33. Respostas apresentadas pelos alunos ao item 4 da tarefa 5.....	138
Quadro 34. Respostas apresentadas pelos alunos ao item 5 da tarefa 5.....	140
Quadro 35. Respostas apresentadas pelos alunos ao item 6 da tarefa 5.....	141
Quadro 36. Respostas apresentadas pelos alunos ao item 1 da tarefa 6.....	144
Quadro 37. Respostas apresentadas pelos alunos ao item 2 da tarefa 6.....	145
Quadro 38. Respostas apresentadas pelos alunos ao item 3 da tarefa 6.....	146
Quadro 39. Respostas apresentadas pelos alunos ao item 4 da tarefa 6.....	147

Índice de Figuras

Figura 1. Item adaptado de Estrada, Batanero e Fortuny (2004).	4
Figura 2. Ciclo investigativo segundo Wild e Pfannkuch (1999).	15
Figura 3. Tarefa proposta por Bill, Henderson e Penman (2010).....	40
Figura 4. Tarefa a realizar em díade proposta por Carvalho e César (2000).....	43
Figura 5. Questão proposta por Lee e Meletiou (2003).....	46
Figura 6. Exemplificação de um “ <i>Hat plot</i> ” / Diagrama chapéu.....	48
Figura 7. Tarefa proposta por Holmes (2002).	51
Figura 8. Processos de recolha de dados ao longo da experiência de ensino.	71
Figura 9. Tarefa “População e amostra” – Parte A e B.....	77
Figura 10. Tarefa “População e amostra” – Parte C.....	80
Figura 11. Resposta apresentada por um grupo ao item 1 da Parte C.	80
Figura 12. Resposta apresentada por um grupo ao item 1 da Parte C.	81
Figura 13. Resposta apresentada por um grupo ao item 2 da Parte C.	82
Figura 14. Resposta apresentada por um grupo ao item 2 da Parte C.	82
Figura 15. Resposta apresentada por um grupo ao item 3 da Parte C.	83
Figura 16. Resposta apresentada por um grupo ao item 4 da Parte C.	84
Figura 17. Resposta apresentada por um grupo ao item 4 da Parte C.	84
Figura 18. Resposta apresentada por um grupo ao item 5 da Parte C.	85
Figura 19. Item 6 da Parte C.....	86
Figura 20. Resposta de um grupo à alínea a) do item 6 da Parte C.	86
Figura 21. Resposta de um grupo à alínea a) do item 6 da Parte C.	86

Figura 22. Resposta de um grupo à alínea b) do item 6 da Parte C.	87
Figura 23. Resposta de um grupo à alínea b) do item 6 da Parte C.	88
Figura 24. Excerto da base de dados da tarefa 2.	91
Figura 25. Excerto de duas resoluções de dois pares de alunos.	92
Figura 26. Excerto de duas resoluções de dois pares de alunos.	93
Figura 27. Excerto da resolução de um par de alunos.	93
Figura 28. Excerto da tarefa.	97
Figura 29. Resposta de um grupo ao primeiro ponto.	97
Figura 30. Resposta de um grupo ao primeiro ponto.	98
Figura 31. Resposta de um grupo ao primeiro ponto.	98
Figura 32. Excerto dos apontamentos teóricos disponibilizados aos alunos.	100
Figura 33. Resposta de um grupo ao segundo ponto.	100
Figura 34. Excerto da tarefa 4.	105
Figura 35. Excerto da tabela de recolha de dados da turma - tarefa 4.	107
Figura 36. Resposta de um grupo ao item 1 da tarefa 4.	108
Figura 37. Resposta de um grupo ao item 1 da tarefa 4.	108
Figura 38. Resposta de um grupo ao item 1 da tarefa 4.	108
Figura 39. Respostas de dois grupos ao item 2 da tarefa 4.	109
Figura 40. Resposta de um grupo aos itens 2 e 3 da tarefa 4.	109
Figura 41. Resposta de um grupo ao item 4 da tarefa 4.	110
Figura 42. Resposta de um grupo ao item 5 da tarefa 4.	111
Figura 43. Excerto da base de dados da população em estudo e de uma das amostras utilizadas pelos grupos na tarefa 4.	113
Figura 44. Resposta de um grupo ao item 6 da tarefa 4.	114
Figura 45. Resposta de um grupo ao item 6 da tarefa 4.	115
Figura 46. Organizaram dos dados para a resolução do item 7 da tarefa 4.	115
Figura 47. Resposta de um grupo ao item 8 da tarefa 4.	117
Figura 48. Resposta de um grupo ao item 8 da tarefa 4.	118
Figura 49. Resposta de um grupo ao item 8 da tarefa 4.	118
Figura 50. Resposta de um grupo ao item 9 da tarefa 4.	119
Figura 51. Resposta de um grupo ao item 9 da tarefa 4.	119
Figura 52. Resposta de um grupo ao item 9 da tarefa 4.	120
Figura 53. Resposta de um grupo ao item 9 da tarefa 4.	120
Figura 54. Resposta de um grupo ao item 10 da tarefa 4.	121
Figura 55. Resposta de um grupo ao item 10 da tarefa 4.	121
Figura 56. Resposta de um grupo ao item 10 da tarefa 4.	122
Figura 57. Resposta de um grupo ao item 10 da tarefa 4.	123
Figura 58. Resposta de um grupo ao item 10 da tarefa 4.	123
Figura 59. Resposta de um grupo ao item 11 da tarefa 4.	124
Figura 60. Resposta de um grupo ao item 11 da tarefa 4.	124
Figura 61. Resposta de um grupo ao item 11 da tarefa 4.	125
Figura 62. Resposta de um grupo ao item 11 da tarefa 4.	127
Figura 63. Quadro resumo comparativo das medidas nas amostras com os parâmetros populacionais da tarefa 4.	129
Figura 64. Relatório – Guias orientadoras quanto ao estilo e formato.	134

Figura 65. Excerto da tarefa 5.	134
Figura 66. Preenchimento da tabela do item 1 da tarefa 5.	135
Figura 67. Resposta de um grupo ao item 4 da tarefa 5.	138
Figura 68. Resposta de um grupo ao item 5 da tarefa 5.	140
Figura 69. Resposta de um grupo ao item 5 da tarefa 5.	140
Figura 70. Excerto da tarefa 6.	142
Figura 71. Resposta de um grupo ao item 1 da tarefa 6.	144
Figura 72. Resposta de um grupo ao item 1 da tarefa 6.	144
Figura 73. Resposta de um grupo ao item 2 da tarefa 6.	145
Figura 74. Resposta de um grupo ao item 2 da tarefa 6.	145
Figura 75. Resposta de um grupo ao item 3 da tarefa 6.	145
Figura 76. Resposta de um grupo ao item 3 da tarefa 6.	146
Figura 77. Resposta de um grupo ao item 4 da tarefa 6.	147
Figura 78. Resposta de um grupo ao item 4 da tarefa 6.	147
Figura 79. Preenchimento da tabela do item 5 da tarefa 6.	148
Figura 80. Previsões apresentadas no item 6 da tarefa 6.	149
Figura 81. Previsões apresentadas no item 6 da tarefa 6.	149
Figura 82. Previsões versus dados reais correspondentes aos últimos 4 meses - item 7 da tarefa 6.	151
Figura 83. Comparação dos desempenhos no âmbito do planeamento estatístico.	165
Figura 84. Comparação dos desempenhos no âmbito da construção e interpretação de representações e cálculo de medidas estatísticas de tendência central.	165
Figura 85. Comparação dos desempenhos no âmbito da extração de conclusões.	166
Figura 86. Resposta do grupo de Ana à questão 6b da tarefa 1.	174
Figura 87. Resposta da Ana à questão 1.1 do teste inicial.	176
Figura 88. Resposta da Ana à questão 4.4 do teste inicial.	178
Figura 89. Resposta ao item 5.2 do teste inicial.	178
Figura 90. Resposta ao item 2a) da 1.ª entrevista.	181
Figura 91. Resposta ao item 2b) da 1.ª entrevista.	182
Figura 92. Resposta ao item 4 a) da 1.ª entrevista.	182
Figura 93. Resposta ao item 1 a) da 1.ª entrevista.	184
Figura 94. Questões de inquérito propostas por Ana no item 4 da 2.ª entrevista.	188
Figura 95. Resposta ao item 2 a) da 2.ª entrevista.	190
Figura 96. Resposta ao item 2b) da 2.ª entrevista.	192
Figura 97. Resposta ao item 3 a2) da 2.ª entrevista.	192
Figura 98. Resposta ao item 3b) da 2.ª entrevista.	193
Figura 99. Resposta ao item 1b da 2.ª entrevista.	194
Figura 100. Resposta ao item 3d) da 2.ª entrevista.	195
Figura 101. Resposta ao item 4.3 do teste inicial.	202
Figura 102. - Resposta ao item 3 do teste inicial.	203
Figura 103. Resposta ao item 3 a) da primeira entrevista.	211
Figura 104. Excerto do esboço do plano proposto no item 4 da 2.ª entrevista.	212
Figura 105. Resposta ao item 3.2 do teste final.	213
Figura 106. Resposta ao item 3.1 do teste final.	215
Figura 107. Resposta ao item 1.1 do teste final.	217

Figura 108. Resposta ao item 2b) da 2. ^a entrevista.....	217
Figura 109. Resposta ao item 3b) da 2. ^a entrevista.....	219
Figura 110. Resposta ao item 1a) da 2. ^a entrevista.....	220
Figura 111. Resposta ao item 6.4 do teste final.....	221
Figura 112. Resposta ao item 3d) da 2. ^a entrevista.....	221

“It has been said that the mark of a truly educated person is to be deeply moved by statistics.”

Bill Moyers (2006)
(citação atribuída a George Bernard Shaw)

Capítulo 1

Introdução

Neste capítulo apresento as ideias principais que motivam a realização deste estudo e as questões a que procuro responder. Faço referência às orientações curriculares existentes e ao cenário de mudança que se vive atualmente no ensino da Estatística, em especial no 3.º ciclo do ensino básico, e descrevo os objetivos do estudo e forma como está organizado.

1.1. Motivações pessoais

Os grandes avanços tecnológicos do século XX contribuíram de modo decisivo para a alteração da sociedade, gerando-se, entre muitos outros aspetos, uma enorme produção de informação. Como refere Branco (2000), a Estatística tem vindo, pouco a pouco, a ser reconhecida no ensino “como uma necessidade evidente de proporcionar aos alunos um sistema coerente de ideias estatísticas e de capacidades de usar essas ideias numa sociedade cada vez mais baseada em dados e informação numérica” (p. 11). Para Turkman e Ponte (2000), a aprendizagem de noções fundamentais de Estatística e do seu papel nos processos de pesquisa e análise de dados, faz parte dos objetivos da educação para a cidadania.

As crescentes necessidades de informação, por parte dos estados, das organizações sociais e profissionais e do cidadão comum, motivaram uma expansão da Estatística, bem como as suas aplicações na sociedade (NCTM, 1991). Segundo Shaughnessy (1992) e Batanero (2000a), isto não foi inicialmente acompanhado pela investigação

realizada por educadores matemáticos ou estatísticos. No entanto, alguns anos mais tarde, Garfield e Ben-Zvi (2007) já referem a realização de numerosos estudos de investigação, em diferentes áreas disciplinares, nomeadamente Estatística, Psicologia e Matemática, bem como congressos científicos e publicações dedicadas à pesquisa em educação estatística entre 1995 e 2007, que a encaram como uma disciplina nova e emergente, quando comparada com outras áreas de estudo e investigação. Em Portugal, também se tem sentido um crescente interesse na educação estatística, visível em encontros de educação matemática.

A minha formação de base em Estatística obtida no curso de Licenciatura em Ensino da Matemática foi complementada pela parte curricular da pós-graduação em Investigação Operacional e Engenharia de Sistemas com uma forte componente de Estatística. Na verdade, quer enquanto aluna, quer como professora, nunca tive uma atitude negativa e/ou de insegurança em relação à Estatística. No entanto, considero que, como professora, fui realizando um trabalho “superficial” no ensino deste tema, relegando o seu estudo para o final do ano letivo e associando-o, fundamentalmente, a aspetos processuais e ao desenvolvimento de um ou outro pequeno projeto de investigação. Ao longo de meu percurso profissional, assumi com alguma ligeireza que a Estatística era um tema simples, de fácil aprendizagem e para o qual os alunos se sentem motivados, o que decorreu de uma abordagem do tema com cariz fundamentalmente técnico (Fernandes, Sousa & Ribeiro, 2004). Como refere Hawkins (1997), não estava consciente da sofisticação do conhecimento estatístico, do quanto é difícil a sua compreensão para alguns alunos e do pouco que tinha investido no domínio didático deste tema.

Repensei o ensino da Estatística quando, em 2001/02, iniciei a minha preparação para a leção da disciplina de Matemática Aplicada às Ciências Sociais (MACS), cujo programa tem uma presença forte de Estatística e Probabilidades. De forma intencional passei a integrar na discussão das situações a natureza dos dados estatísticos e a valorizar de modo determinante o papel do contexto, selecionando para isso, questões de natureza social. Procurando ir ao encontro das características dos alunos a quem leciono a disciplina, tomo como eixo central das minhas práticas letivas o desenvolvimento de conceitos estatísticos recorrendo a problemas da vida real, mais numa perspectiva de formação cultural do que de formação estritamente técnica. Procuro enfatizar o facto do raciocínio estatístico estar intimamente relacionado com a incerteza e a sua importância determinante em áreas do saber que constituem as suas possíveis futuras opções académicas – Sociologia, Psicologia, Direito, Política. Na minha prática, mudei

gradualmente a ênfase de uma abordagem computacional para uma abordagem centrada no conhecimento e na compreensão dos conceitos estatísticos, valorizando a interpretação da informação no contexto da situação em estudo e a capacidade de desenvolver e avaliar conclusões e previsões.

A vida em sociedade depende cada vez mais, da capacidade do Homem fazer previsões e tomar decisões, com base em informação organizada segundo métodos estatísticos e probabilísticos, uma vez que os fenómenos aleatórios estão presentes em toda a atividade humana. Por isso, a escola deve proporcionar uma visão equilibrada e não determinística da realidade, proporcionando aos alunos momentos que promovam o desenvolvimento do pensamento estatístico e probabilístico (Abrantes, Serrazina & Oliveira, 1999). Sendo o objetivo principal da Estatística a perceção da variabilidade dos dados, é fundamental que os alunos vivam experiências que lhes permitam trabalhar com dados recolhidos em função de questões significativas.

O planeamento estatístico reveste-se de grande importância uma vez que constitui uma oportunidade de dar a entender o que é a Estatística como ciência. Tal como refere o Relatório GAISE (2005), a fase de “formulação de uma questão de natureza estatística requer a compreensão da diferença entre uma questão que antecipa uma resposta determinística e uma resposta baseada em dados que variam” (p. 11). A análise exploratória de dados, enquanto abordagem global em que se procura compreender a forma como estes se distribuem, apoiando-se nas medidas de localização e de dispersão e em diferentes formas de representação, como atitudes e não de modo meramente descritivo, é essencial para gerar novas “hipóteses”, ou seja, para se poder conjecturar sobre os dados de que se dispõe (Batanero, 2001; Batanero, Estepa & Godino, 1991). Como refere Carvalho (2001b), os alunos devem ser capazes de apreciar o papel do raciocínio estatístico que permite ler e interpretar fenómenos reais. Devem também perceber que, dada a sua natureza, este raciocínio não permite obter uma solução única e que, uma vez obtida, uma solução não pode ser considerada como totalmente correta ou totalmente errada, uma vez que está dependente da qualidade do raciocínio e da adequação dos métodos utilizados face à natureza dos dados.

A título de exemplo, perante o problema da figura 1, a generalidade dos alunos considera os nove dados, apesar da razoabilidade da variabilidade no contexto – a variabilidade da medição, sugerir a necessidade de uma “limpeza de dados”. No caso de se usar o conjunto dos nove dados, o recurso à mediana em detrimento da média está associado à compreensão de que esta não é robusta face a dados atípicos (*outliers*).

Nove estudantes pesaram um pequeno objecto com o mesmo instrumento numa aula de ciências. Os pesos registados por cada aluno (em gramas) indicam-se abaixo:
6,2 6,0 6,0 15,3 6,1 6,3 6,2 6,15 6,2
Os estudantes querem determinar tão precisamente quanto possível o peso real do objecto. De que modo devem proceder?”

Figura 1. Item adaptado de Estrada, Batanero e Fortuny (2004).

Nesta situação, um número significativo de alunos revela dificuldades em assumir que uma possível resposta à questão poderia ser: (i) rejeitar o valor 15.3, e determinar a média dos restantes dados recolhidos, determinando o ponto de equilíbrio ou (ii) considerar o valor da mediana, tendo ou não rejeitado o valor 15.3, dado que se trata de uma medida robusta que devolve, por exemplo, como informação que cerca de 50% das pesagens se encontra abaixo ou igual ao seu valor.

O contraste entre a natureza do raciocínio estatístico e a natureza dos raciocínios algébrico e geométrico também constitui um desafio com o qual me tenho debatido, pois exige um considerável domínio científico no âmbito da Estatística, bem como conhecimento didático do ensino da Estatística de modo a identificar situações de outros campos do saber ou relacionados com aspetos do quotidiano dos alunos, em que a variabilidade esteja inerente. Exige também o desenvolvimento de uma atmosfera de aprendizagem, onde os alunos (i) se sintam seguros para explorar, conjecturar, supor, (ii) se sintam motivados a trabalhar com problemas que podem não ter uma solução certa ou errada e exigir um reinvestimento de esforços na reestruturação da recolha e nos métodos de organização, (iii) se sintam confortáveis com a confusão temporária ou ao encontrar resultados inconclusivos, e (iv) não tenham medo de experimentar a aplicação de diferentes ferramentas ou métodos estatísticos (Gal & Ginsburg, 1994).

Tendo em conta as orientações do *Programa de Matemática do Ensino Básico* (ME, 2007), senti que seria pertinente pôr em prática algum do conhecimento da Estatística e da Didática da Estatística que fui desenvolvendo na leção da disciplina da Matemática Aplicada às Ciências Sociais. Tenho também interesse em ampliar ambas as vertentes do conhecimento, alterando a minha anterior postura como professora do ensino básico, em que tenho promovido uma aprendizagem centrada numa transmissão do conhecimento, para promover uma aprendizagem mais ativa da literacia estatística e do processo de investigação estatística.

Como na altura da realização do estudo estava a lecionar Matemática a duas turmas do 8.º ano numa escola que estava a implementar o programa de 2007, considerei oportuno estudar a aprendizagem de alunos deste ano de escolaridade no quadro de uma experiência de ensino que visa desenvolver a capacidade de planeamento estatístico e de análise de dados. Espero assim contribuir para um melhor conhecimento das aprendizagens ao nível do planeamento estatístico e análise de dados por parte de toda a comunidade educativa, nomeadamente para os professores que manifestam interesse e preocupações no âmbito da educação estatística e também para o meu próprio desenvolvimento profissional.

1.2. Orientações curriculares para o ensino da Estatística

Atualmente, a informação e sua descodificação é essencial, sendo impossível falar de alfabetização sem atribuir um papel primordial ao ensino da Estatística. Deste modo é internacionalmente aceite que o ensino deste tópico deve ter uma importância determinante no currículo (Australian Educational Council, 1991; National Council of Teachers of Mathematics, 1991, 2007).

O documento *Normas para o Currículo e a Avaliação em Matemática Escolar* (NCTM, 1991), ao introduzir normas relativas ao tópico Estatística e Probabilidades para todos os níveis de ensino, enfatiza a importância de uma educação matemática que promova a formação de cidadãos informados e consumidores inteligentes. Mais recentemente, o documento *Princípios e Normas para a Matemática Escolar* (NCTM, 2007) define dez normas para a Matemática escolar, sendo cinco delas no âmbito dos conteúdos matemáticos, assumindo a Análise de Dados e Probabilidades o mesmo peso que Números e Operações, Álgebra, Geometria e Medida. Houve um reajustamento da designação usada na edição anterior das *Normas* (NCTM, 1991), pois a “Estatística e Probabilidades” passou a designar-se “Análise de Dados e Probabilidades”, indicando-se que, para compreenderem as “ideias fundamentais de Estatística, os alunos deverão trabalhar diretamente com dados” (NCTM, 2007, p. 52).

Deste modo, os alunos devem adquirir competências num nível preliminar da análise estatística de dados, muitas vezes denominados por «organização e tratamento de dados» ou «análise exploratória de dados» (Loura, 2009), o que não inclui os processos de inferência formal. O documento NCTM (2007) reconhece tratar-se de um conteúdo matemático em que surge naturalmente um ambiente propício a “conexões entre

ideias e procedimentos de Números, Álgebra, Medida e Geometria” (p. 52), bem como a conexões entre a Matemática, as outras áreas disciplinares e as suas experiências quotidianas.

Os *Princípios e Normas* (NCTM, 2007) defendem que o currículo deve habilitar todos os alunos para: (i) Formular questões que possam ser abordadas por meio de dados e recolher, organizar e apresentar dados relevantes que permitam responder a essas questões; (ii) Selecionar e usar métodos estatísticos adequados à análise de dados; (iii) Desenvolver e a avaliar inferências e previsões baseadas em dados; e (iv) Compreender e aplicar conceitos básicos de Probabilidades. O trabalho direto com dados é a ideia-chave para o envolvimento progressivo dos alunos com novas ideias e procedimentos em oposição a uma atividade rotineira.

Em Portugal, os conceitos de Estatística e Probabilidades assumem visibilidade explícita no programa de Matemática do 3.º ciclo desde os anos 90 (ME, 1991a; ME 1991b). Estes conceitos surgem também de modo implícito nas indicações do programa do 1.º ciclo (ME, 1990). O documento *Matemática na Educação Básica* (Abrantes, Serrazina & Oliveira, 1999) veio a reforçar a importância do pensamento estatístico e probabilístico em todos os ciclos do ensino básico considerando que “ser competente em Estatística é fundamental para entender os julgamentos que os meios de comunicação social veiculam com base na Estatística e nas Probabilidades” (p. 76). Indo além, da promoção de uma educação estatística de cidadãos eficientes na leitura e interpretação da informação, o documento argumenta que a capacidade de ser autor de mensagens com base em Estatística e Probabilidades é fundamental a um conjunto cada vez maior de profissões e áreas de estudo. Por isso, desde cedo se deve promover a resolução de situações que envolvam a incerteza, a variabilidade e impliquem a tomada de decisão face à análise de dados recolhidos no âmbito de experiências desenhadas pelos próprios, tendo por base os interesses e situações problemáticas do quotidiano dos alunos.

Para Ponte e Fonseca (2000), as principais diferenças entre as orientações curriculares do programa de Matemática português, de 1991, o *Nacional Curriculum for Maths* inglês, de 1997 e o documento *Principles and standards for school mathematics* (NCTM, 1998), no tema da Estatística, ao nível do 3.º ciclo do ensino básico eram à data:

- Relativamente ao envolvimento dos alunos em processos de investigação, o documento americano e o documento inglês apontam clara-

mente, a sua importância, apelando diretamente à formulação de questões a investigar por parte dos alunos; o português pretende envolver os alunos em atividades de recolha, análise e interpretação de dados, mas nunca propõe explicitamente que os alunos definam as questões que pretendem estudar, não insiste na formulação de inferências nem equaciona esta atividade em termos de um processo global de investigação;

- O documento americano e o documento inglês referem-se à compreensão e utilização de medidas de tendência central e de dispersão para análise e comparação de distribuições; o português apenas aborda as medidas de tendência central;
- O documento americano é o que apresenta uma maior variedade de representações gráficas;
- Apenas o documento americano se refere ao conceito de amostra e aos diferentes modos de escolher amostras. (p. 189)

Os autores do estudo comparativo indicam, como ponto comum entre o documento português e o americano, o facto de em ambos se considerar que o tema da Estatística constitui uma boa ocasião para a realização de trabalhos interdisciplinares.

Uma visão ampla do ensino-aprendizagem da Estatística e Probabilidades é apresentada no *Currículo Nacional* (ME, 2001), que destaca que, no âmbito deste domínio, a competência matemática que todos os alunos devem desenvolver inclui os seguintes aspetos ao longo dos três ciclos:

- A predisposição para recolher e organizar dados relativos a uma situação ou a um fenómeno e para os representar de modos adequados, nomeadamente através de tabelas e gráficos e utilizando as novas tecnologias;
- A aptidão para ler e interpretar tabelas e gráficos à luz das situações a que dizem respeito e para comunicar os resultados das interpretações feitas;
- A tendência para dar resposta a problemas com base na análise de dados recolhidos e de experiências planeadas para o efeito;
- A aptidão para realizar investigações que recorram a dados de natureza quantitativa, envolvendo a recolha e análise de dados e a elaboração de conclusões;
- O sentido crítico face ao modo como a informação é apresentada. (p. 64)

Em particular, no que se refere aos aspetos específicos do 3.º ciclo, o *Currículo Nacional* (ME, 2001), procurando ir de encontro às orientações curriculares internacionais, nomeadamente *Nacional Curriculum for Maths* (1997) e *Principles and standards for school mathematics* (NCTM, 1998) aponta: (i) a compreensão das noções de moda, média aritmética e mediana, bem como a aptidão para determiná-las e para interpretar o que significam em situações concretas; (ii) a sensibilidade para decidir quais das medidas de tendência central são mais adequadas para caracterizar uma dada situação; (iii) a aptidão para comparar distribuições com base nas medidas de tendência central e numa análise da dispersão dos dados; e (iv) o sentido crítico face a afirmações baseadas em amostras não representativas.

O *Programa de Ensino Básico* (ME, 2007) procura romper com uma aprendizagem centrada numa transmissão mais ou menos passiva do conhecimento ao promover uma aprendizagem ativa da literacia estatística e do processo de investigação estatística, desde os primeiros anos de escolaridade. Incluindo nos três ciclos de ensino básico o tema “Organização e tratamento de dados”, assume-se de modo claro a sua importância no âmbito do desenvolvimento social e pessoal dos alunos, fomentando o desenvolvimento gradual dos conceitos estatísticos e das representações de modo a que os alunos compreendam e sejam capazes de produzir informação estatística e de a utilizar para resolver problemas e tomar decisões informadas. Além disso, assume-se que existem dois objetivos centrais, no ensino da Estatística, que se desenvolvem em paralelo e se reforçam mutuamente:

- Desenvolver a literacia estatística, que inclui a capacidade de ler e interpretar dados organizados na forma de tabelas e gráficos e de os usar para responder às questões mais variadas;
- Desenvolver a capacidade de planear e executar uma investigação estatística, bem como a capacidade de interpretar e avaliar criticamente os resultados de um estudo estatístico já realizado.

A aprendizagem dos conceitos e representações específicas é essencial para a realização de estudos estatísticos cada vez mais complexos e para a compreensão do que é uma investigação estatística atribuindo significado aos diversos conceitos e representações (Martins & Ponte, 2010). No programa de 2007, o papel complementar das

representações gráficas e das tabelas na análise dos dados é uma constante ao longo dos três ciclos, sendo de destacar que no âmbito do 1.º ciclo, logo no 1.º e 2.º ano, os alunos classificam dados recorrendo a diagramas de Venn e de Carroll, aprendem a construir e a interpretar tabelas de frequências absolutas, gráficos de pontos e pictogramas; no 3.º e 4.º ano as representações incluem os gráficos de barras; no 2.º ciclo estender-se-á à organização dos dados a tabelas de frequências relativas, levando os alunos a construir e interpretar gráficos circulares, permitindo a comparação de conjuntos de dimensão distinta; no 3º ciclo, à representação que anteriormente predominava – o histograma – acrescenta-se a construção e a interpretação de diagramas de caule-e-folhas e de diagramas de extremos e quartis – com forte e imediato impacto visual, em relação à apresentação numérica dos dados. Estes diagramas tornam possível a análise comparativa de dois ou mais conjuntos de dados com variabilidade distinta, em relação às medidas centrais e/ou medidas de dispersão, dada a visualização ordenada dos dados em análise, permitindo identificar de modo imediato as respetivas diferenças e semelhanças.

Espera-se que o trabalho desenvolvido ao longo dos três ciclos, habilite os alunos no final do 3.º ciclo, quanto à compreensão e utilização da linguagem básica da Estatística e das suas ideias fundamentais – formulação de questões a investigar e interpretação dos resultados à luz da questão em investigação – enfatizando deste modo todo o processo estatístico e o “sentido” dos dados. O trabalho a desenvolver no âmbito da “Organização e tratamento de dados” possibilita o estabelecimento de conexões com outros temas da disciplina. Para Martins e Ponte (2010) as conexões entre a Estatística e os diversos campos da Matemática podem ser estabelecidas com a Geometria, por exemplo, na construção de gráficos circulares; com os Números e operações, por exemplo, no cálculo de medidas estatísticas, com a Álgebra, por exemplo, no trabalho com tabelas e gráficos de linhas. Consideram também os autores que o trabalho a desenvolver no âmbito da “Organização e tratamento de dados” promove o desenvolvimento das capacidades transversais, sendo significativo o papel central da comunicação na resolução de problemas estatísticos.

No momento, a nível internacional, procura-se reformular o ensino da Estatística, privilegiando o trabalho com base em pequenas investigações de temas e questões quotidianas de interesse para os alunos. É aceite que o ensino da Estatística se deve centrar na compreensão de conceitos. Como refere Scheaffer (2000), para promover o pensamento estatístico o foco do trabalho a desenvolver com os alunos passa pela análise

exploratória de dados e, embora não se deva descurar a prática dos métodos utilizados neste tópico,

as técnicas continuam a ser úteis, e talvez sejam uma parte importante da instrução, podendo mesmo ser um ponto de partida, mas atualmente o ensino da Estatística tem de ir além do manual ou dos procedimentos (...) a educação estatística moderna tem de ter a Análise Exploratória de Dados no seu seio. (p. 158)

1.3. Objetivos e questões de investigação

Tendo por base as orientações curriculares acima indicadas, sinto interesse em desenvolver uma investigação no âmbito da Estatística. Ao ter a possibilidade de lecionar duas turmas do ensino básico numa escola que está a implementar o programa de 2007, pretendo estudar a aprendizagem de alunos de 8.º ano no quadro de uma experiência de ensino que visa desenvolver a capacidade de planeamento estatístico e de análise de dados.

O estudo que me proponho realizar tem por base a realização da unidade de ensino sugerida pela equipa de professores das turmas-piloto, no âmbito da experimentação do programa de Matemática (ME, 2007). A unidade de ensino é composta por quatro tarefas relativas ao “Planeamento estatístico”, que adaptei às características específicas das minhas turmas. As tarefas da unidade têm carácter exploratório e investigativo, procuram proporcionar aos alunos experiências diversificadas e significativas, tendo em vista o desenvolvimento das suas capacidades de planeamento, e integrando os conhecimentos desenvolvidos no âmbito do tema “Tratamento de dados”, abordado no 7.º ano.

Neste estudo, a *capacidade de planeamento*, é definida como ser capaz de:

- Formular questões e planear adequadamente a recolha de dados tendo em vista o estudo a realizar.
- Identificar e minimizar possíveis fontes de enviesamento na recolha dos dados.
- Distinguir entre população e amostra e ponderar elementos que podem afetar a representatividade de uma amostra em relação à respetiva população. (ME, 2007, p. 60)

Por *capacidade de análise de dados*, entende-se ser capaz de:

- Construir, analisar e interpretar representações dos dados (incluindo o histograma) e tirar conclusões.
- Compreender e determinar a mediana, os quartis e a amplitude inter-quartis de um conjunto de dados, e utilizar estas estatísticas na sua interpretação.
- Escolher as medidas de localização mais adequadas para resumir a informação contida nos dados.
- Comparar as distribuições de vários conjuntos de dados e tirar conclusões.
- Responder às questões do estudo e conjecturar se as conclusões válidas para a amostra serão válidas para a população. (ME, 2007, p. 60)

Assim, este estudo procura identificar eventuais contributos da unidade de ensino, baseada em tarefas de carácter exploratório e investigativo para o desenvolvimento da capacidade de planeamento estatístico e análise de dados dos alunos do 8.º ano. De modo mais específico, procuro responder às seguintes questões de investigação:

1. Que capacidades de planeamento estatístico e análise de dados têm os alunos antes da realização da experiência de ensino?
2. Que aprendizagens realizam os alunos, com a realização da unidade de ensino, ao nível do planeamento estatístico?
3. Que aprendizagens realizam os alunos, com a realização da unidade de ensino, ao nível da análise de dados?
4. Quais as principais dificuldades demonstradas pelos alunos antes, durante e no fim da experiência de ensino nos conceitos e procedimentos inerentes ao planeamento estatístico e à análise de dados?

O estudo procura responder a estas questões baseando-se, por um lado, na análise do trabalho desenvolvido pelos alunos de uma turma do 8.º ano, e por outro lado, nos estudos de caso de dois alunos dessa mesma turma. O intuito deste trabalho é contribuir para o conhecimento relativo ao desenvolvimento das capacidades de planeamento e análise de dados, identificar potencialidades da Didática da Estatística e dificuldades de concretização desta unidade de ensino. Espero, também, que a realização do estudo pos-

sa contribuir para o meu próprio desenvolvimento pessoal e profissional, melhorando o meu conhecimento científico e didático no âmbito desta matéria.

1.4. Organização do estudo

Neste primeiro capítulo exponho a minha motivação pessoal, a pertinência do estudo e o objeto de estudo. No segundo capítulo, abordo do ponto de vista teórico, os temas principais deste estudo: planeamento estatístico, análise exploratória de dados, desenvolvimento e avaliação de inferências e previsões baseadas nos dados, nomeadamente dificuldades na aprendizagem e compreensão de conceitos fundamentais em Estatística e algumas recomendações no âmbito da avaliação das aprendizagens em Estatística. No terceiro capítulo descrevo a unidade de ensino, apresento o enquadramento curricular, os objetivos de cada tarefa realizada, a sua planificação e o ambiente de sala de aula. Seguidamente, no capítulo quatro, apresento as opções metodológicas, as principais características dos participantes do estudo e as razões da sua escolha. Apresento também os procedimentos metodológicos adotados e justifico as escolhas realizadas, quanto à recolha e análise dos dados. Nos capítulos 5, 6 e 7, apresento a análise dos dados, descrevendo com pormenor o percurso global da turma (capítulo 5) e de cada um dos alunos objeto de estudo de caso – Caso da Ana (capítulo 6) e Caso do Rui (capítulo 7). Por fim, no oitavo capítulo, apresento os principais resultados do estudo e conclusões, e termino com uma reflexão pessoal sobre o que significou a sua concretização no âmbito do meu desenvolvimento pessoal e profissional, bem como sobre algumas questões que podem servir de pistas para futuras investigações.

“A Estatística estuda o comportamento dos chamados fenômenos coletivos. Projetar informações sobre um grupo ou universo, é o seu objeto material, a sua maneira própria de raciocínio, o método estatístico, é o objeto formal e procura previsões para o futuro, o que implica um ambiente de incerteza, que é a sua finalidade ou a sua causa final.”

Cabriá (1994)

Capítulo 2

Quadro conceptual do estudo

Neste capítulo abordo os objetivos do ensino da Estatística e, do ponto de vista teórico, os temas principais deste estudo, nomeadamente planeamento estatístico, análise exploratória de dados, desenvolvimento e avaliação de inferências e previsões baseadas nos dados. Especial atenção é dada às dificuldades na aprendizagem e compreensão de conceitos fundamentais em Estatística e algumas recomendações no âmbito da avaliação das aprendizagens neste tema curricular.

2.1. Ensino da Estatísticas e seus objetivos

Na maioria dos países, só recentemente a Estatística integra o currículo da Matemática escolar. Em Portugal, a sua integração como tema curricular em Matemática foi fomentada pelo movimento da Matemática Moderna, e a sua expansão pelos vários níveis de ensino foi sendo progressiva, partindo do ensino superior e expandindo-se aos restantes níveis pré universitários. Atualmente existe Estatística no programa de todos os níveis de escolaridade.

A Estatística tem sido encarada no ensino segundo uma variedade de perspetivas. Para Holmes (2000) coexistem na Europa três grandes tendências relativas ao ensino da Estatística: (i) ênfase no processo de Análise de Dados, na perspetiva em que esta ciência é utilizada na sociedade, tendo em conta que o uso de dados faz parte da vida de

todos os dias (tendência predominante em países como a Inglaterra); (ii) como capítulo da Matemática, por vezes designada por Estocástica, enfatizando aspetos conceptuais e/ou computacionais (abordagem seguida, por exemplo, na França); e (iii) como *'state' istics*, ou seja, como uma ferramenta auxiliar para o estudo de diversos assuntos e disciplinas escolares (tendência visível, por exemplo, na Suécia). Para este autor, os dois grandes objetivos da educação estatística ao nível pré universitário são: (i) levar os alunos a tomar consciência e apreciar o papel da Estatística na sociedade, percebendo os muitos e variados campos em que as ideias estatísticas são usadas; e (ii) fomentar neles a compreensão da natureza do pensamento estatístico, incluindo o poder e as limitações deste pensamento.

Para Scheaffer (2000) a Estatística é e continuará a ser uma componente principal da educação matemática nos vários níveis de escolaridade dada a sua aplicação e relevância na sociedade. Considera que, ao tomar-se como eixo principal de trabalho, a nível escolar, a análise de dados, promove-se uma visão da Estatística como um modo de pensar sobre a informação quantitativa. Indica, ainda, como fundamental que os professores encarem a Estatística como um trabalho de detetive, interativo com os dados recolhidos, envolvendo abertura, flexibilidade, multiplicidade e incerteza dos resultados.

2.2. Ciclo de investigação estatístico

O ciclo de investigação estatístico é uma estrutura que suporta a resolução de problemas estatísticos (GAISE, 2005; Graham, 2006; Konold & Higgins, 2003; Wild & Pfannkuch, 1999). Trata-se de um processo complexo e interativo entre quatro aspetos: (i) formular questões e hipóteses; (ii) recolher dados (iii) analisar dados, e (iv) comunicar e interpretar resultados. Wild e Pfannkuch (1999) subdividem a fase inicial do ciclo de investigação em duas, apresentando cinco fases interdependentes (Figura 2). Deste modo, os diversos autores apresentam o ciclo de investigação estatístico com quatro ou cinco fases sendo as semelhanças entre eles maiores do que as diferenças.

Ao apresentar o trabalho realizado no âmbito de um ciclo de investigação transmite-se a ideia de que o processo é linear, pois a sua comunicação não tem por hábito incluir os avanços, os recuos e o facto de se antecipar um determinado aspeto numa fase posterior do ciclo poder condicionar as opções realizadas numa fase anterior (Konold & Higgins, 2003), tal ocorrer porque as fases do ciclo são interdependentes (Wild & Pfannkuch, 1999).

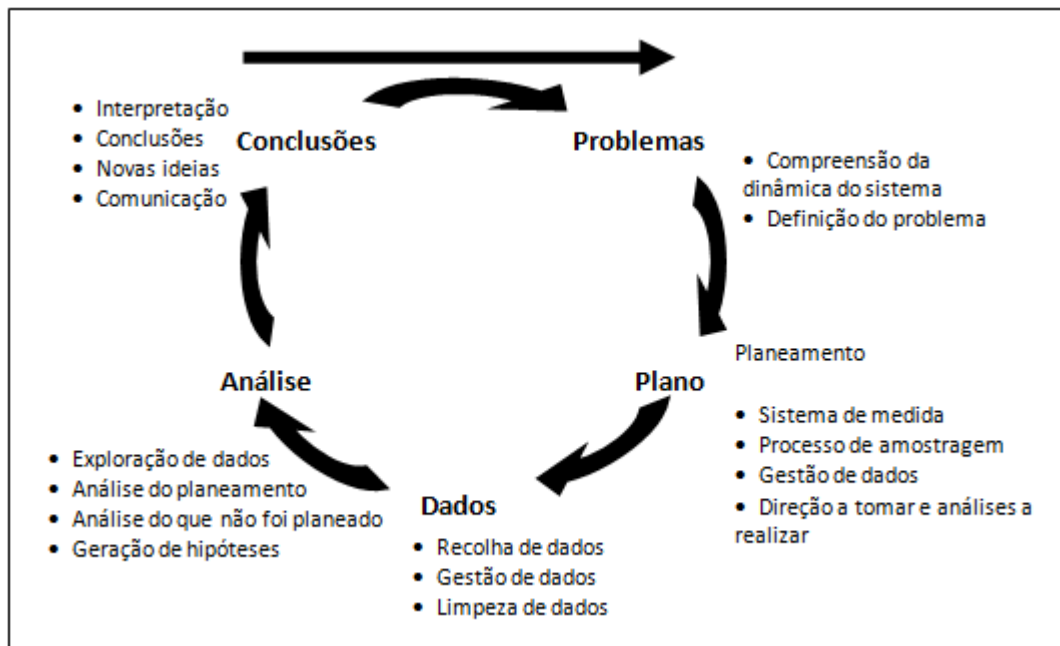


Figura 2. Ciclo investigativo segundo Wild e Pfannkuch (1999).

No desenrolar de um ciclo investigativo, Wild e Pfannkuch (1999) consideram que se mobilizam cinco componentes fundamentais do raciocínio estatístico, que apresentam como:

- *Reconhecer a necessidade de dados*: uma investigação estatística tem por base uma situação da vida real só pode ser compreendida através da análise de dados obtidos de modo correto.
- *Transnumeração*: os autores usam esta palavra para indicar uma compreensão que pode surgir ao alterar a representação dos dados, ocorre nomeadamente quando a interpretação dos dados conduz a novos significados. A título de exemplo num sistema real, na perspectiva da modelação, podem surgir três tipos de transnumeração: (i) na fase de recolha – como "captar" as qualidades ou características do mundo real, (ii) na fase de tratamento - ao passar os dados em bruto para uma representação tabular ou gráfica que permite resumir os dados, sem perda do sentido dos mesmos, (iii) na fase de análise - ao transmitir o significado, a reflexão sobre o que emergiu a partir dos dados, pode promover a compreensão sobre a situação real e outras inter-relações.
- *Perceção de variabilidade*: a recolha de dados adequados e julgamentos corretos sobre os mesmos requer uma compreensão da variabilidade que existe e que é transmitida pelos dados e pela incerteza causada pela variabilidade inexplicada. Perceber a variabilidade da situa-

ção e encontrar estratégias adequadas a cada etapa da investigação são evidências do raciocínio estatístico.

- *Raciocinar com modelos estatísticos*: qualquer medida estatística, representação gráfica ou resumo estatístico pode ser visto como um modelo, dado que é uma forma de representar a realidade. O importante é diferenciar o modelo de dados enquanto modelo relativo aos dados.
- *Integração do contexto na análise estatística* – que apontam como uma componente essencial do raciocínio estatístico: a Estatística pretende caracterizar os padrões existentes nos dados para os compreender, permite a realização de previsões, a procura de explicações e a descoberta de causas a partir do contexto. (p. 227)

2.2.1. Formulação de questões estatísticas

Dada a curiosidade e a natureza interrogadora dos jovens, um modo de os familiarizar com o contexto da investigação estatística, passa pela utilização de contextos sobre os alunos ou contextos próximos do seu “mundo”, por exemplo, estudos a nível do grupo turma e/ou recorrer a dados recolhidos em experiências laboratoriais realizada pelos alunos (NCTM, 2007). A história de fundo ou contexto permite um maior envolvimento e facilita a formulação de questões.

Segundo o Relatório GAISE (2005), a formulação de questões estatísticas implica a capacidade de reconhecer a existência de variabilidade da resposta, isto é, a compreensão da diferença entre uma questão para a qual se antecipa uma resposta determinística e uma questão para a qual se prevê a resposta com base em dados que variam. É importante reconhecer que existem diferentes fontes de variabilidade nos dados, sendo as mais importantes:

- *Variabilidade da medição* – inerentes a quem mede, ao instrumento de medição utilizado, ao momento da medição;
- *Variabilidade natural* – os elementos em análise são naturalmente distintos;
- *Variabilidade induzida* – fatores não considerados que implicam variabilidade da característica em análise;
- *Variabilidade da amostragem* – várias amostras de uma única população, não são necessariamente iguais (p.5-6).

Para Graham (2006), na fase correspondente à formulação de questões devem ter-se em consideração os seguintes cinco aspetos:

- Uma questão em vez de uma área de estudo – uma investigação com base numa questão é mais provável vir a recorrer a capacidades matemáticas e estatísticas dado que fornece uma direção clara e focaliza a investigação.
- Uma questão que seja do interesse do aluno – não só vai trazer uma maior motivação, mas também o conhecimento ao nível do senso comum sobre um determinado contexto pode ajudar a investigação a prosseguir ao longo de uma direção mais sensata.
- Uma questão que seja provável baseia-se em dados que estarão disponíveis dentro do prazo disponível – por exemplo, não investigar o crescimento de flores ou plantas durante os meses de inverno.
- Uma questão específica, de modo que seja possível responder a partir de dados – perguntas que são demasiado vagos e genéricos são mais difíceis de responder.
- Uma questão para a qual é possível pensar em aspetos da mensurabilidade – pensar com antecedência, o que se vai medir e se o que se mede ajudar a responder à pergunta. (p. 88)

Konold e Higgins (2003) analisam alguns episódios de sala de aula onde professores e alunos juntos transformam questões gerais em questões estatísticas. Defendem que os alunos podem aprender muito sobre os dados enquanto se debatem com as questões a formular, nomeadamente quando antecipam a realização do estudo, com o auxílio das questões que projetam para a recolha dos mesmos. Caso se pretenda responder à questão “As crianças são mais ativas do que os adultos?”, um dos primeiros desafios passa pela transformação da questão inicial numa questão estatística, isto é que possa ser respondida por meio de dados, como por exemplo: “Caso as distâncias totais percorridas fossem recolhidas diariamente por meio de um pedómetro, num grupo de crianças e num grupo de adultos, qual dos grupos teria maiores distâncias totais?” A formulação desta possível questão estatística teve por base questões do tipo: O que é ser ativo? De que modo se expressa ser ativo? Que variáveis posso associar a ser ativo? Como posso proceder para medir essa variável?... Os autores defendem ainda que a tomada de decisões deste tipo fazem parte do processo de transformar uma questão geral numa questão

estatística. Para Konold e Higgins (2003) um outro aspeto no âmbito da formulação de uma questão estatística passa por decidir qual a população que se pretende estudar. Os autores defendem que a aprendizagem da formulação de questões estatísticas e recolha e análise de dados para posterior resposta está associada ao equilíbrio entre: (i) a formulação de uma questão estatística específica o suficiente que permita a recolha de dados relevantes mas que por sua vez não trivialize a questão inicial e (ii) a capacidade de assumir os dados produzidos como abstrações do contexto que os gerou mas nunca esquecendo que não são apenas números mas “números num contexto”. Para estes autores, o contexto do ciclo de investigação estatístico é importante diferenciar questões que têm finalidades distintas: a questão geral/ inicial, que depois de revista do ponto de vista da estatística assume uma formulação estatística, sendo a questão que requer dados para ser respondida e a questão de inquérito/pesquisa com a qual se recolhem os dados.

Outros investigadores (como Graham, 2006; Pfannkuch & Horring, 2005) identificam diferentes tipos de questões estatísticas, dependendo da natureza dos dados: questões de descrição, de comparação ou de inter-relação. As questões de descrição, por vezes também designadas de questões de resumo, envolvem a recolha de um conjunto único de dados e têm subjacente a pretensão de resumi-lo de algum modo por palavras, medidas estatísticas ou gráficos, por exemplo: “Como se distribuem as idades dos professores da escola A?” As questões de comparação envolvem a recolha de dois ou mais conjuntos de dados do mesmo tipo e visam uma comparação gráfica ou por meio de medidas estatísticas de resumo das duas distribuições, por exemplo: “Em que escola, A ou B, se encontram professores com menos tempo de serviço?” As questões de inter-relação/relacionamento implicam a análise da relação entre duas variáveis distintas em que se procura aferir a sua associação/correlação, pelo que os dados recolhidos têm a característica de serem dados emparelhados, por exemplo: “Serão os professores com menos tempo de serviço os que têm menos idade?” Para Graham (2006) é importante que os alunos tenham em consideração os três tipos de questões estatísticas, porque tomam conhecimento da diversidade de questões face à natureza dos dados e dos objetivos envolvidos na análise e porque, conhecendo o tipo de questão de investigação, podem selecionar mais adequadamente os instrumentos analíticos para a fase de análise.

Arnold (2008) descreve um estudo que realizou envolvendo alunos com 14 e 15 anos (correspondendo ao 10.º ano), com o objetivo de identificar que diferentes categorias de questões estatísticas, nomeadamente de resumo e de comparação, os alunos propõem que sejam analisadas numa fase inicial de uma investigação estatística envolvendo

um conjunto de dados multivariado proposto. Apresenta um quadro concetual para a análise das questões estatísticas formuladas pelos alunos (Quadro 1).

Quadro 1. Análise das questões estatísticas formuladas (Arnold, 2008).

	... da categoria Resumo	... da categoria Comparação
Não são questões	<p>1. Absurdo ou não, é uma questão de resumo.</p> <p>2. Uma questão que está parcialmente relacionada com os dados, mas que não tem resposta através dos mesmos.</p> <p>3. Uma questão que aborda um dado caso individual.</p>	<p>1. Absurdo ou não, é uma questão de comparação.</p> <p>2. Uma questão que está parcialmente relacionada com os dados, mas não é respondida pelos dados.</p>
Pré-questões	<p>4. Uma questão que pretende conhecer quantos elementos existirão numa categoria particular.</p> <p>5. Uma questão que pertence conhecer quantos elementos haverá num determinado intervalo.</p> <p>6. Uma questão que solicita a moda ou os elementos mais comuns.</p>	<p>3. Uma questão que está relacionada com os dados, mas à qual não é possível responder devido a problemas relacionados com o tamanho da amostra.</p>
Questões	<p>7. Uma questão que envolve a distribuição total dos dados ou que pretende identificar o que é típico.</p> <p>8. Uma questão que envolve distribuição total dos dados ou que pretende identificar o que é típico reflete conclusões que podem ser extraídas sobre a população.</p>	<p>4a. Uma questão que é responsável pelos dados.</p> <p>4b. Uma questão que é responsável pelos dados e requer recategorização das categorias a serem comparadas.</p> <p>5a. Uma questão que é responsável pelos dados e reflete uma população para a qual se podem extrair conclusões.</p> <p>5b. Uma questão que é responsável pelos dados que requer recategorização das categorias serem comparadas e reflete a população para a qual se podem extrair conclusões.</p>

Do conjunto de questões colocadas e discutidas em aula, Arnold (2008) constatou que surgem com maior frequência questões de descrição do que de comparação. O autor considera que existe vantagem em discutir numa fase precedente um conjunto de questões moderadoras (Quadro 2) para que se possa aferir se uma determinada questão de investigação estatística é, ou não, adequada.

Quadro 2. Questões moderadoras (Arnold, 2008).

- Que questão original terá sido usada para recolher os dados (questão de pesquisa)?
- Que tipo de dados se está a usar?
- Que representação gráfica dos dados será feita?
- Que hipóteses podem ser elaboradas sobre os dados?
- A questão é interessante?
- Quem estaria interessado nas respostas a esta questão?
- Há dados suficientes disponíveis para responder à questão (questões em torno do tamanho da amostra)?
- Que informação de base tenho de disponibilizar sobre os dados (como foram recolhidos, foram recolhidos a partir de ... quando foram recolhidos, etc.)?
- É uma variável de interesse no conjunto de dados?

Numa segunda fase dos seus estudos, Arnold (2009) considera que o contexto é fundamental para que se assegure um efetivo envolvimento dos alunos nas fases de definição do problema e planeamento. Considera que os professores poderão ampliar o trabalho que habitualmente se limita às três últimas fases (dados, análise e conclusões), caso se assegurem que os alunos se envolvem nas duas fases iniciais do ciclo investigativo (problema e plano). Para tal, é fundamental que o professor valorize a história por de trás dos dados, forneça informações básicas sobre o contexto dos dados (como foram recolhidos e que questões de pesquisa foram utilizadas), disponibilize tempo para que os alunos possam interagir com a informação de fundo e a interroguem e face a uma questão de investigação colocada, por um aluno, reflita em conjunto com os alunos sobre a qualidade da questão.

O relatório GAISE (2005) alerta para a pertinência de trabalhar com os alunos a tipologia da questão/item de pesquisa e o modo como esta pode influenciar os dados obtidos, por exemplo para uma investigação sobre as preferências musicais, os inquiridos podem responder de modo distinto a um item fechado (com categorias pré determinadas), do que a um item aberto (“Qual é o tipo de música que prefere?”). As respostas ao item aberto podem incluir categorias que não se tinham incluído no item fechado, podendo tornar a organização, o tratamento dos dados e sua interpretação mais difícil. Discutir os méritos de cada tipo de item e se necessário pensar na construção de um item flexível (com categorias pré determinadas e que inclua a possibilidade de seleccionar “outra opção” [explicitamente]), para manter predominantemente a tipologia fechada.

Consideram que um outro aspeto a ter em atenção na discussão em torno da questão de pesquisa é a formulação de questões com juízo de valor e/ou que o corpo principal do item influencia a resposta do inquirido, como por exemplo, “Música clássica é o género menos popular entre os jovens. Qual é o melhor tipo de música para passar na rádio da escola?”. Os alunos têm que ter em consideração que as questões de pesquisas formuladas devem ser sempre claras e não ambíguas. Deste modo pensam que os alunos terão uma atitude crítica face às inúmeras questões de pesquisa apresentadas nos diferentes meios de comunicação social. Para Konold e Higgins (2003) é importante que os alunos antecipem o levantamento de dados com base nas questões que delinearão. Ao refletirem sobre o modo como responderiam à questão proposta, apercebem-se do grande leque de respostas, bem como da possibilidade de haver múltiplas interpretações de cada questão, e constataam o importante papel que as palavras escolhidas, para a formulação da questão, poderão ter. Os autores salientam ainda que o processo de transformar uma questão geral numa questão estatística, para além de exigir uma formulação que não conduza facilmente a múltiplas interpretações, também deve garantir a recolha da informação adequada para responder à questão inicial.

Como referem Konold e Higgins (2003) um elemento a considerar quando se colocam questões de resumo é o ponto de vista individual dos dados, versus uma visão agregada dos mesmos. Com o reconhecimento da variabilidade, muitas das vezes os alunos assumem uma visão individual dos dados, cabendo ao professor o papel de mobilizar questões interessantes que impliquem a análise dos dados de modo agregado.

Se os valores dos dados considerados variam, no entanto, por que deveriam considerar ou pensar sobre esses valores como um todo? Além disso, a resposta para muitas das perguntas que têm interesse para os alunos - por exemplo: Quem é mais alto? Quem tem mais? Quem mais gosta de mim? Exigem localização de dados individuais, especialmente dentro de um grupo (população que inclui o próprios alunos). Não devemos esperar que os alunos se concentrem nas características do grupo até que lhes seja solicitado, com uma pergunta cuja resposta exige descrever as características da distribuição. (p. 203)

Assim, constitui um desafio para os professores ajudar os alunos a formular questões do seu interesse e que estes possam por sua vez continuar de modo produtivo (Konold & Higgins, 2003).

2.2.2. Planeamento estatístico

O planeamento estatístico reveste-se de grande importância uma vez que constitui uma oportunidade de dar a entender o que é a Estatística como ciência. Neste planeamento, é fundamental a fase precedente à organização e análise de dados, designada por vezes por *produção ou aquisição de dados*. Segundo Tannenbaum e Arnold (1997), esta fase desempenha um papel fundamental na credibilidade da “história” que se pretende relatar:

Por detrás de cada estudo estatístico, há uma história, e como qualquer história tem um início, um meio, o fim, e uma moral. ... O que em Estatística significa tipicamente o início é o processo de recolha ou a recolha de dados. Os dados são a matéria-prima de que a informação estatística é feita, e a fim de obter uma boa informação estatística são necessários bons dados. (p. 426)

É necessário compreender o objeto de estudo, delinear os procedimentos de recolha de dados, criar e organizar conjuntos de dados, refletir sobre eles bem como agir em conformidade. Além disso, os alunos devem aperceber-se da existência de variabilidade nos fenómenos que são objeto de estudo por parte da Estatística.

Outra fase fundamental no planeamento é a *amostragem*. Na verdade, quando se pretende descrever as populações existe grande vantagem em reduzir a quantidade de dados a recolher, o que leva a usar amostras. Garfield e Gal (1999) apontam outras razões que podem implicar o recurso a amostras em vez de inquirir toda a população, nomeadamente quando a população é de dimensão infinita ou no caso de a análise implicar a destruição dos elementos da população, etc. O processo de amostragem é crítico no momento da recolha de dados, pelo que constitui um ponto fulcral de uma investigação estatística. O planeamento de um estudo estatístico, que envolva o recurso a uma amostra, começa com a forma de a selecionar, o que deve ser feito de forma a evitar amostras enviesadas, pelo que os alunos terão que ter em consideração questões como as seguintes: “Qual é a população? Como deverá ser selecionada a amostra de modo a ser representativa? É necessário utilizar uma amostra estratificada? Qual deverá ser a dimensão da amostra?” (NCTM, 2007, p. 384). Para Konold e Higgins (2003) é fundamental que os alunos compreendam que a escolha de uma amostra de grande dimensão não é garantia de representatividade, a inclusão de uma secção transversal da

população na amostra é o ponto principal, sendo a seleção aleatória o único processo que garante a construção de amostras representativas.

Em estudos desenvolvidos com alunos do 5.º ano, Jacobs (1999) constatou que estes optam por processo de amostragem tendenciosos, tais como participação voluntária, dada à percepção de justiça que intuem do facto de permitir aos elementos da população a possibilidade de fazer parte da amostra sem haver a necessidade de impor a uns quantos elementos selecionados por processos aleatórios a participação no conjunto amostra. A autora refere que alguns alunos rejeitam a ideia de amostragem, porque subestimam as dificuldades associada à realização de um censo e quando aceitam recorrer a um processo de amostragem optam por amostras estratificadas em detrimento de amostras aleatórias simples pois pretendem garantir que todas as tipologias de elementos populacionais sejam contempladas na amostra. Os alunos têm dificuldade em compreender que se recorre a uma amostra aleatória simples, porque na generalidade das populações em estudo é impossível identificar e estratificar os vários traços da população que pudessem ter peso no estudo de determinada característica.

Watson e Moritz (2000), num estudo dedicado ao desenvolvimento do conceito de amostra com alunos 3.º, 6.º e 9.º ano, identificaram cinco categorias de resposta e um conjunto de equívocos envolvendo a utilização de amostras (Quadro 3) e constataram que o nível de proficiência aumentava à medida que se subia no ano de escolaridade, sendo que o desempenho de quatro quintos dos alunos do 9.º ano, envolvidos no estudo, correspondia às categorias “Grandes amostras com seleção aleatória” ou “Grandes amostras sensíveis a enviesamento”.

Pfannkuch (2008) defende que à hierarquia de Watson e Moritz (2000) é possível acrescentar categorias de nível superior à categoria “grandes amostras sensíveis ao enviesamento” quando se considera a compreensão da variabilidade da amostragem. Na sua perspetiva, de modo a compreender o papel atribuído à dimensão da amostra, os alunos devem trabalhar a variabilidade entre amostras de igual dimensão e a variabilidade entre amostras de dimensões distintas. Devem ainda compreender que, ao aumentar a dimensão da amostra, existe um ponto para o qual as distribuições das amostras são semelhantes pelo que é possível extrair informação sobre a distribuição da população com base nessas amostras.

Quadro 3. Categorias do desenvolvimento do conceito de amostra (Watson & Moritz, 2000).

Categoria	Tipo de resposta
Pequenas amostras sem seleção	Aceita uma amostra de dimensão inferior a 15. Não sugere qualquer método de seleção ou sugere um método idiossincrático.
Pequenas amostras com um método primitivo de seleção	Aceita uma amostra de dimensão inferior a 15. Sugere seleção aleatória sem proceder à decisão do como se processa ou indica um instrução simples para a escolha, como por exemplo “qualquer”, “talvez de escolas diferentes”.
Pequenas amostras com pré-seleção	Aceita uma amostra de dimensão inferior a 15. Sugere seleção de pessoas por determinada característica, por exemplo “por peso”, “mistura de pessoas gordas, magras e com peso normal”.
Grandes amostras com seleção aleatória	Sugere uma dimensão de pelo menos 20 elementos ou uma % da população. Sugere um processo de seleção baseado num processo aleatório ou de distribuição estratificada.
Grandes amostras sensíveis a enviesamento	Sugere uma dimensão de pelo menos 20 elementos ou uma % da população. Sugere um processo de seleção baseado num processo aleatório ou de distribuição estratificada. Expressa preocupação com a seleção da amostra de modo a evitar enviesamento. Identifica amostras não representativas.
Equívocos envolvendo amostras	Mostra indiferença quanto à dimensão da amostra, por vezes com base em aspetos irrelevantes. Fundamenta decisões com base em amostra de pequena dimensão, mas obtida por métodos de seleção apropriados ou com atenção parcial a enviesamentos. Fundamenta decisões com base em amostras de grande dimensão obtida por métodos de seleção inapropriados.

Bakker (2004) na investigação desenvolvida com alunos de 8.º ano, cujo objetivo passava pelo desenvolvimento conceptual de algumas das grandes ideias estatísticas – variação, amostragem, dados e distribuição – constatou que o uso da ferramenta tecnológica (*Minitool*) com diferentes abordagens gráficas dos dados fomentava a noção de variabilidade do processo de amostragem. As reflexões oriundas da comparação entre as representações gráficas previstas pelos alunos de diferentes amostras, com crescente

dimensão (de uma mesma população) com a distribuição de amostras obtidas com o recurso ao *Minitool*, fomentavam uma visão agregada dos dados e contribuía para o reconhecimento que a variabilidade existente numa amostra diminuí à medida que se aumentava a dimensão da amostra. Uma outra ferramenta passível de utilizar para o reconhecimento da variabilidade entre amostras e o efeito da dimensão da amostra, são os filmes curtos (*movie snapshots*) – simulações de representações dinâmicas disponíveis no sítio <https://www.stat.auckland.ac.nz/~wild/WPRH/> – que permitem os alunos construir uma compreensão visual da variabilidade (Arnold & Pfannkuch, 2010).

Para Garfield (2002), o raciocínio correto sobre amostras inclui saber de que modo estas estão relacionadas com a população de que foram extraídas, o que se pode inferir a partir destas, reconhecer amostras representativas, conhecer aspetos que poderão gerar enviesamento na escolha de uma amostra e ser crítico face a inferências com base em amostras de pequena dimensão ou tendenciosas.

Segundo Martins e Ponte (2010) o trabalho a desenvolver com os alunos deve envolver: (i) a análise de situações em que é adequado o estudo de toda a população ou apenas de uma amostra, (ii) a análise crítica de estudos estatísticos face ao uso de amostras não representativas e (iii) a ponderação de elementos que afetam a representatividade de uma amostra. Para os autores, a análise de processos de amostragem que tendem a gerar amostras enviesadas, tais como amostra por conveniência e amostra por resposta voluntária, constitui uma situação propícia ao desenvolvimento do raciocínio sobre amostras.

Segundo o programa de Matemática, (ME, 2007) para o nível de escolaridade correspondente ao 3.º ciclo, os alunos devem ponderar elementos que podem afetar a representatividade de uma amostra em relação à respetiva população e desenvolver a sensibilidade de que uma seleção aleatória da amostra constitui o único processo que garante a representatividade. Para Martins e Ponte (2010) os alunos ao nível do 3.º ciclo da escolaridade devem ser sensíveis aos seguintes aspetos:

- A dimensão de uma amostra depende muito da variabilidade da população subjacente.
- A dimensão da amostra terá de ser tanto maior, quanto maior for a precisão exigida.
- A dimensão da amostra a recolher não é diretamente proporcional à dimensão da população a estudar. (p.189-190)

Face à crescente proliferação de sondagens, Bill, Henderson e Penman (2010) consideram pertinente que os alunos (do correspondente ao 3.º ciclo) tomem conhecimento que para populações de elevada dimensão, a fiabilidade da estimativa de amostra não depende da relação entre a dimensão da amostra e a dimensão da população (a habitual regra dos 10%), mas da dimensão absoluta da amostra, que nestes casos deve ter cerca de 1500 elementos.

Landwehr (1989, citado por Shaughnessy, Garfield & Greer, 1996) reuniu uma série de mal-entendidos sobre conceitos elementares de Estatística, associados à fase de amostragem. Segundo o autor, tende-se a acreditar que qualquer diferença no meio de dois grupos é significativa, ter confiança indevida em pequenas amostras, não atribuir significado a pequenas diferenças que ocorrem em grandes amostras e acreditar, erroneamente que não há variabilidade no “mundo real” ou então que a dimensão de uma amostra é independente da dimensão da população em geral.

Pelo seu lado, Garfield (2002) refere que na abundante investigação sobre a utilização incorreta do raciocínio estatístico se destacam os seguintes erros e mal entendidos, no âmbito da escolha de amostras:

- Acreditar, erroneamente que boas amostras devem representar uma percentagem elevada da população.
 - Associar, intuitivamente a representatividade à relação entre o tamanho da amostra e o da população.
 - Ter receio na utilização de amostra que mesmo de grande dimensão representa uma pequena percentagem da população.
- Ter confiança indevida em pequenas amostras.
 - Assumir que as amostras de uma mesma população são semelhantes entre si e com a população da qual foram extraídas, independentemente da dimensão da amostra - "lei dos pequenos números" (Kahneman, et al., 1982).
- Acreditar que uma amostra aleatória é um retrato fiel de uma população, não assumindo a aleatoriedade do processo de amostragem.
 - Assumir que numa amostra aleatória o aparecimento de cada porção/ extrato deve ser proporcional ao que surge na população. (Kahneman, et al. 1982).

- Assumir que depois de uma longa série de dados de uma porção/ extrato é mais provável recolher um dado de outro extrato. (para. 17-20)

2.2.3. Análise exploratória de dados

Aspetos gerais. Os termos “análise exploratória de dados” e “visualização de dados” foram introduzidos nos currículos universitários a partir de década de 70 por John Tukey. A sua introdução gradual nos currículos escolares pré-universitários deve-se a vários movimentos de reforma curricular, nomeadamente o *Cockcroft Report* (1982) e o *Curriculum and Evaluation Standards* (NCTM, 1989).

Segundo Batanero, Estepa e Godino (1991) até cerca de 1980, nos estudos estatísticos, a análise de dados recorria fundamentalmente ao cálculo estatístico, atribuindo pouca ou nenhuma importância à representação gráfica dos dados. O objetivo da análise era o de confirmar ou refutar a hipótese de que os dados se ajustavam a um modelo preestabelecido. Dos dados recolhidos não se extraía qualquer outra informação que deles se pudesse derivar. Na análise estava apenas em causa a procura de um modelo que expressasse a regularidade dos dados observados, isto é a estrutura simplificada de um conjunto de dados. Face à nova capacidade de cálculo e de representação gráfica proporcionada pelos instrumentos tecnológicos, surge uma nova abordagem à análise de dados, a análise exploratória de dados, para Batanero, Estepa e Godino (1991) assenta em dois princípios: a regularidade e os desvios. Os desvios (ou os resíduos) são as diferenças dos dados relativamente à estrutura, dado que normalmente o conjunto de dados não se ajusta inteiramente a uma determinada estrutura. Na análise exploratória de dados estudam-se os dados de diversas perspetivas, recorrendo a diferentes representações e procedendo a comparações, procurando extrair o máximo de informação possível, no sentido de se gerar um modelo a partir do conjunto, assumindo que são igualmente importantes a regularidade e os desvios.

Para Shaughnessy, Garfield e Greer (1996) “trabalhar na Análise Exploratória de Dados é um estado de espírito, um ambiente onde se pode explorar dados e não só um determinado conteúdo estatístico” (p. 205). Para os autores, nesta fase procuram-se os padrões, as lacunas, os agregados (*clusters*), a forma e as variações dos dados. A análise exploratória de dados, exige compreender a forma como os dados se distribuem, apoiando-se nas medidas de localização e de dispersão, bem como no apoio de diferen-

tes formas de representação. Os autores julgam ser fundamental ter em consideração que os dados são frequentemente recolhidos e apresentados por alguém que tem uma agenda particular. Deste modo, as crenças e atitudes por de trás dos dados são aspetos importantes a ter em consideração no respetivo tratamento, tal como os métodos de organização e análise utilizados. Os autores defendem, ainda, que é importante para a experiência dos alunos, no âmbito da análise exploratória dos dados, a análise dos possíveis desvios e abusos dos conceitos estatísticos.

Para Konold e Higgins (2003) a análise exploratória de dados é um processo de conversação entre as intuições de quem realiza o estudo sobre o fenómeno e o que é transmitido pelos dados relativamente as essas mesmas intuições iniciais. Os autores defendem que “o que se encontra nos dados modifica a compreensão inicial do problema, o que implica uma mudança no modo como se analisam os dados o que, por sua vez altera o seu entendimento e assim por diante” (p.194).

Assim, segundo Shaughnessy, Garfield e Greer (1996) e Konold e Higgins (2003) a análise exploratória de dados, embora assente significativamente na componente visual da representação e técnicas estatísticas, centra-se, mais do que na mera descrição, numa atitude de “dar sentido” ao conjunto de dados numa perspetiva global. Shaughnessy, Garfield e Greer (1996) defendem ainda que tomar a análise exploratória de dados como um dos eixos principais do currículo, contribui para a visão da matemática enquanto atividade humana, pois à sua natureza dedutiva se acrescenta a natureza indutiva da análise de dados.

Representações gráficas. Para esta abordagem no âmbito da análise de dados, em muito contribuiu o surgimento de duas representações – o diagrama de caule-e-folhas e o diagrama de extremos e quartis – introduzidos por Tukey, dado o forte e imediato impacto da visualização em relação à apresentação numérica dos dados. Ambas as representações são usadas para representar distribuições de variáveis quantitativas, obtendo-se em ambos os casos uma visualização ordenada dos dados em análise. Além disso, ambas as representações possibilitam a comparação de duas distribuições, observando de modo imediato diferenças e semelhanças entre elas.

O diagrama de caule-e-folhas é uma das representações mais simples, sendo uma forma sugestiva de representar os dados em que se perde apenas a informação da ordem pela qual os dados se apresentavam no conjunto recolhido. A sua forma reflete a forma da população subjacente aos dados, fornece informação quanto à simetria ou assimetria, maior ou menor concentração e eventual existência de valores atípicos (*outliers*). Para

Batanero, Estepa e Godino (1991) o diagrama de caule e folhas é similar a um histograma em que a amplitude das classes corresponde à “diferença” entre os caules, apresentando as seguintes vantagens face ao histograma: (i) facilidade de construção; (ii) facilidade na identificação das medidas estatísticas de ordem, tais como extremos, mediana, quartis, percentis, bem como na obtenção da amplitude total e interquartil e (iii) facilidade na identificação de diferenças importantes entre os valores de um mesmo caule/classe, dado que se tem acesso aos dados recolhidos. No entanto o diagrama de caule-e-folhas apresenta algumas desvantagens face ao histograma pois, na sua construção, a escolha do caule é limitada pelo sistema numérico e não é possível escolher uma escala de representação (as folhas são algarismos, ocupando assim um conjunto de caracteres específico por cada uma delas).

O diagrama de extremos e quartis fornece informação quanto à simetria e à dispersão dos dados, não só na parte central, mas também nas caudas da distribuição; evidência a posição relativa das medidas estatísticas de ordem; fornece informação sobre a existência ou não de valores atípicos, e trata-se de uma representação que facilita a comparação de dois ou mais conjunto de dados.

Quanto à representação de dados, importa ainda referir que, segundo Tuhte (1983, citado por Shaughnessy, Garfield & Greer, 1996) a sua excelência consiste em comunicar ideias complexas com clareza, precisão e eficiência. Para este autor, a representação deve contemplar os seguintes aspetos:

- Mostrar os dados;
- Induzir o leitor da representação a pensar sobre a substância e não sobre metodologia;
- Apresentar muita informação numérica num pequeno espaço;
- Tornar grandes conjuntos de dados coerentes;
- Incentivar a uma visão comparativa de diferentes “pedaços” de dados;
- Revelar os dados em vários níveis de detalhe;
- Servir a uma finalidade razoavelmente clara (tais como: descrição, exploração, tabulação ou decoração);
- Estar estreitamente integrado com a apresentação de estatísticas e descrições verbais dos dados. (p.215)

Segundo Friel, Curcio e Bright (2001) existem duas razões para a utilização de representações gráficas: análise e comunicação. As representações gráficas utilizadas para fins de análise são essencialmente ferramentas para a identificação de características importantes ou atípicas nos dados. Por outro lado, as representações gráficas usadas para comunicar são principalmente imagens com o objetivo de transmitir informações sobre os valores numéricos assumidos pelos dados e as relações entre esses valores numéricos. Para estes investigadores a estrutura de um gráfico dá informação sobre as escalas que estão a ser utilizadas e os dados que estão a ser analisados. Sendo que a maioria das representações gráficas, vistas apenas na perspetiva da construção, é composta por quatro componentes estruturais: (i) a dimensão métrica, que comporta os eixos, a escala, marcações de referência; (ii) a dimensão visual do gráfico, designada por especificadores, usada para representar os valores dos dados, por exemplo a barra de um histograma ou o símbolo de um pictograma; (iii) as etiquetas, que designam os nomes que se dá a cada um dos elementos dos especificadores; o título do gráfico, que pode ser considerado um tipo de etiqueta; e (iv) o fundo do gráfico, que pode incluir qualquer coloração, grelha e fotos sobre os quais o gráfico pode ser justaposto. No entanto, o domínio destes componentes não é condição suficiente para garantir a compreensão de uma determinada representação gráfica, pois para além dos quatro componentes enunciadas, a cada representação associa-se uma linguagem própria, pelo que, outro aspeto a ter em consideração, é o contexto.

Konold e Higgins (2003) consideram que, na construção de uma representação gráfica, os alunos não devem ficar presos a eventuais convenções estabelecidas, e sim ter em mente as questões a que pretendem responder. Deste modo, a escolha de uma escala, definição do valor mínimo e máximo os eixos, tamanho da unidade nos eixos (barras altas e estreitas, barras curtas e grossas, muito ou pouco afastadas?), não devem ser fundamentalmente baseada em convenções, porque todas as decisões anteriores irão afetar a forma como os dados se apresentam e não existe a escala ideal para representar os dados tal como são na realidade. Consideram que os alunos devem ser estimulados a construir diferentes representações com opções de escala distintas e recolher entre as várias representações o máximo de informação distinta possível. No momento de apresentar um resumo gráfico devem escolher a representação que “melhor cumprir a função de contar a história de forma mais precisa e justa” (p. 202).

Rudenstein e Thompson (2012) consideram que muitas das representações poderão ser mais transparentes para os alunos, caso os professores identifiquem e adotem

estratégias de suporte à leitura das mesmas, sugerem um conjunto de questões gerais correspondentes aos três níveis de leitura de identificadas por Curcio (1987) (Quadro 4).

Quadro 4. Questões para apoiar a leitura de representações gráficas (Rudenstein & Thompson, 2012).

Tipos de questões	Com o objetivo de...	Exemplos de questões
Orientação	Identificar as características da representação	Quais são as partes (por exemplo, título, eixos, escala, chaves, símbolos de congruência)?
“Bem ali” Leitura direta	Reconhecer e mobilizar conhecimento para uma leitura literal da representação	O que os eixos representam? Qual é a escala? O ideograma representa o quê? O que significa 3 2?
“Pensar e procurar” Obter informações	Identificar interpretações de primeiro nível	Que comparações a representação sugere? Que medidas podem ser determinadas a partir da informação disponível?
“Eu e o autor” e “No meu mundo” Relacionar a informação com o entendimento pessoal do contexto	Identificar que outras interpretações ou extensões se poderão obter ao cruzar o conhecimento sobre o contexto com a representação.	Que previsão pode estabelecer a partir desta informação? Que outros fenômenos são comparáveis com a representação? Como modificar a representação de modo a enfatizar um outro aspeto dos dados? Que outra representação seria adequada aos dados?

De acordo com Martins e Ponte (2010) a experiência matemática dos alunos ao longo do ensino básico, envolve a comparação de diversos tipos de representação para uma mesma situação, a identificação das potencialidades e adequação de cada representação face a natureza dos dados recolhidos e dos aspetos que se pretendem analisar, promovendo o desenvolvimento de uma atitude crítica relativa à utilização das representações.

O raciocínio estatístico sobre representações, para Garfield (2002) envolve: (i) compreender de que modo uma determinada representação irá representar uma amostra (conjunto de dados), (ii) compreender de que modo uma determinada representação gráfica pode ser alterada para melhor representar um conjunto de dados e (iii) ser capaz de reconhecer características gerais, como a forma, o centro e a dispersão da distribuição, isto é ter uma visão de conjunto dos dados representados. Globalmente, segundo o NCTM (2007) quando os alunos acedem a diferentes representações e às ideias por estas expressas, dominam “um conjunto de ferramentas que aumentam significativamente a sua capacidade de pensar matematicamente” (p. 75), pois distintas representa-

ções focam aspectos diferentes de uma relação ou conceitos de elevada complexidade pelo que “os alunos necessitam de uma diversidade de representações que suportem a sua compreensão” (p. 77).

Raciocínio estatístico sobre os dados. Garfield (2002) salienta que o raciocínio estatístico sobre dados, envolve o aluno ter capacidade para: (i) reconhecer/categorizar os dados como quantitativos ou qualitativos, discretos ou contínuos, e (ii) saber a razão pela qual um tipo específico de dados se adequa a um tipo específico de tabela, gráfico ou de medida estatística. No âmbito das medidas estatísticas o raciocínio estatístico envolve os seguintes aspectos: (i) compreender por que razão as medidas centrais, de dispersão, nomeadamente as de ordem, e a sua respetiva localização, fornecem informação sobre um conjunto de dados, (ii) identificar quais as medidas mais adequadas a utilizar face a diferentes condições, e de que modo representam, ou não, um conjunto de dados, (iii) saber que um resumo descritivo dos dados deve incluir uma medida central, bem como uma medida de dispersão e (iv) reconhecer que as medidas estatísticas centrais e de dispersão são úteis para comparar conjuntos de dados.

Pelo seu lado, Ben-Zvi e Friedlander (1997) descrevem algumas características do raciocínio sobre dados, observados nos processos utilizados pelos alunos na manipulação de dados e sua representação com apoio de ferramentas tecnológicas:

- Pensamento acrítico – utilização aleatória ou de forma acrítica do poder tecnológico e de métodos estatísticos,
- Utilização com compreensão das representações – uso de uma representação gráfica adequada ou de uma medida a fim de responder às suas questões de investigação e para interpretar seus resultados,
- Manuseamento com compreensão de múltiplas representações – procura de significado e interpretação de diferentes representações para alcançar resultados razoáveis, bem como no acompanhamento dos seus processos,
- Pensamento criativo – recurso a uma representação incomum ou método que expresse seus pensamentos de modo mais adequado (produção de uma representação gráfica inovadora, criação de uma medida estatística ou de um método de análise) (p. 49-52).

Para Konold e Higgins (2003), quanto mais os alunos assumirem a importância da conexão entre os dados, as ocorrências que representam e a questão que motivou a sua produção, mais se envolvem na fase de análise e dominam as ferramentas de análise

se. Os autores constataam que muitos alunos tratam dos dados como números apenas, esquecendo que estes têm um contexto e que o motivo da sua análise prende-se com o objetivo de conhecer mais sobre esse contexto. Referem ainda que, a análise descontextualizada surge com mais predominância nas situações em que os alunos não tiveram a oportunidade de trabalhar com dados recolhidos por si.

Análise exploratória de dados no ensino básico. Como indicam Batanero, Estepa e Godino (1991) e Batanero (2001), o propósito da análise exploratória de dados é extrair o máximo de informação para, tanto quando possível, gerar “hipóteses” novas, ou seja, conjecturar sobre os dados que se dispõem. Estes autores apoiam a inclusão da análise exploratória de dados na Estatística, ao nível do ensino básico pois consideram que deste modo se possibilita a criação de situações de aprendizagem envolvendo temas do interesse dos alunos, acentuando a utilização de representações gráficas variadas e não implicando uma base teórica de conhecimento matemático complexo. Para Ben-Zvi, (2004) um outro aspeto positivo da integração da análise exploratória de dados no currículo escolar prende-se com a possibilidade do desenvolvimento de situações de exploração com apoio de ferramentas tecnológicas, habilidade cada vez mais necessária aos alunos, numa fase de integração no mundo do trabalho.

2.2.4. Desenvolver e avaliar inferências e previsões baseadas nos dados

Retirar conclusões a partir dos dados e indicar a confiança que se atribui a tais conclusões é uma fase fundamental na aprendizagem da Estatística. Como refere o NCTM, o recurso a instrumentos tecnológicos permite disponibilizar mais tempo na exploração da essência estatística – analisar os dados de vários pontos de vista, estabelecer inferências, construir e avaliar argumentos:

É essencial que os alunos cheguem a perceber a diferença entre o “certo-ou-errado” que caracteriza a maior parte do conhecimento matemático, e a natureza condicionada dos resultados das análises estatísticas. Contudo, é igualmente importante que os alunos não extrapolem para além deste facto, rejeitando o pensamento estatístico porque ele permite contra-exemplos. Em vez disso, devem reconhecer que a estatística desempenha um importante papel intermediário, entre a exatidão de outros estudos matemáticos e a natureza equívoca de um mundo que depende largamente da opinião individual. (NCTM, 1991, p. 202)

Para Rubin, Bruce e Tenney (1991) as duas ideias centrais da recolha de amostras – representatividade da amostragem e variabilidade da amostragem – são fatores decisivos para compreensão da inferência estatística. A representatividade da amostra está associada à percepção de que uma amostra retirada de uma população muitas vezes, têm características semelhantes às da população da qual foi extraída. Por exemplo, a proporção de meninas em sala de aula é suscetível de se aproximar da proporção de meninas de toda a escola. A variabilidade da amostra está relacionada com a percepção de que várias amostras de uma única população, não são necessariamente iguais e, portanto, não “contradizem” a população. Assim, em algumas salas de aula de uma escola é provável que existam mais meninas do que meninos, mesmo que se tenha conhecimento de que a população escolar tenha equidade de géneros.

Rubin, Bruce e Tenney (1991) indicam que a generalidade dos alunos do ensino secundário que participou no seu estudo sente um conflito entre a variabilidade e a necessidade de representatividade da amostra. Consideram que a dependência excessiva da representatividade da amostra induz a “ideia” de que uma amostra nos diz tudo sobre uma população, por sua vez o excesso de confiança na variabilidade das amostras tem como consequência assumir que uma amostra não nos fornece informação útil sobre uma população.

Tem sido possível observar a resistência que alguns alunos opõem ao facto da informação de uma amostra poder fornecer informações úteis sobre toda a população. Metz (1999) constatou que a maioria dos alunos de 2.º, 4.º e 5.º ano que participaram no seu estudo não se dispunha a generalizar a partir de uma amostra estudada para a população, argumentando que: (i) a informação extraída só era válida no conjunto de casos observados; (ii) para caracterizar um grupo, deve-se analisar todos os membros desse grupo, e (iii) a amostragem não permite tal generalização dada a variabilidade inerente à população. Para a investigadora, o facto da generalidade dos alunos em análise, não cair no erro de confiar em pequenas amostras deve-se à extrema consciência da variabilidade, facto que invocaram predominantemente na refutação da inferência. Tal como defende Gal (2002)

O conhecimento do contexto é o principal determinante da familiaridade do leitor com as fontes de variação e erro. Se um ouvinte ou leitor não está familiarizado com o contexto em que os dados foram recolhidos, torna-se mais difícil imaginar porque é que a diferença entre os grupos pode ocorrer, que interpretações alternativas podem existir para valores

encontrados sobre a associação detetada entre determinadas variáveis, ou como um estudo poderia estar errado. (p.15)

No estudo de Jacobs (1999), face a vários estudos estatísticos, no momento de extrair conclusões, as respostas dos alunos evidenciaram as seguintes características: (i) tiveram em consideração a qualidade dos estudos realizados; (ii) agregaram todos os estudos, independentemente da qualidade; (iii) recorreram às suas próprias opiniões e ignoraram os dados dos estudos realizados, ou (iv) recusaram-se a tirar conclusões. A autora constatou ainda que embora alguns alunos tenham sido capazes de identificar possíveis fatores de enviesamento em algumas dos estudos, no momento de extrair conclusões optaram por ignorar a qualidade do mesmo.

Para o NCTM (2007) os alunos aprofundam a sua capacidade de testar a generalização das suas conclusões quando são incentivados a discutir as informações extraídas das relações observadas numa amostra e “a debater se as suas conjeturas são ou não aplicáveis a populações maiores, nas quais a amostra está contida” (p. 298). Sugerindo que, para ampliar a capacidade de testar a generalização das suas conclusões, (i) se questione os alunos sobre as conclusões extraídas, por exemplo, com base numa amostra de alunos de 7.º ano, se seriam as mesmas, caso a escolha tivesse sido uma outra amostra da mesma população; (ii) se compare, caso seja possível, diversas amostras obtidas da mesma população ou se compare as medidas estatísticas obtidas numa ou mais amostra com os parâmetros da população. Por outro lado, ao debater com os alunos as suas conjeturas, sobre se elas serão ou não aplicáveis a populações maiores, nas quais a amostra está contida, o NCTM sugere que se questione se as conjeturas que surgem com base, por exemplo, numa amostra de alunos do 7.º ano de uma escola: (i) se são ou não aplicáveis ao conjunto de todos os alunos de 7.º ano da escola; (ii) se o são para o conjunto de todos os alunos de 7.º, 8.º e 9.º ano da escola; (iii) se o serão para todos os alunos de 7.º ano dessa cidade; etc.

No âmbito desta fase do ciclo de investigação Garfield (2002) considera fundamental que a aferição do raciocínio sobre amostras inclua os seguintes aspetos: (i) saber o que se pode inferir a partir de uma amostra, (ii) ser crítico face a inferências feitas com amostras pequenas ou tendenciosas e (iii) saber que, face a amostras representativas, uma previsão será mais precisa, caso se trabalhe com uma amostra de maior dimensão. Para Martins e Ponte (2010) é importante que os alunos tenham consciência que o objetivo é inferir propriedades para a população a partir do estudo dos dados da amos-

tra, pelo que as amostras constituídas pelos dados recolhidos devem representar convenientemente as populações subjacentes. No entanto, também consideram que os alunos, no término do ensino básico, se devem aperceber das limitações inerente ao processo de recolha de dados e que "... a informação recolhida através de uma amostra, mesmo em condições ótimas de recolha de dados, está sempre associada a alguma incerteza nas conclusões formuladas sobre a população" (p. 15). No programa de 2007, assume-se que analisar possíveis causas de variabilidade e compreender as dificuldades que surgem na escolha da amostra e na inferência das conclusões para a população é um dos propósitos do planeamento estatístico (ME, 2007).

Segundo o NCTM (2007), estimular a formulação inicial de conjeturas e solicitar a sua demonstração com base nos dados, pode gerar contradição pelo facto de os resultados (ou os dados obtidos) não irem na direção espectável dos alunos, por exemplo numa questão de comparação. Este facto poderá fomentar novas hipóteses ou motivar explicações alternativas, que por sua vez podem passar pela conceção de estudos posteriores para análise. Um outro aspeto importante desta fase, passa por saber seleccionar as representações e medidas mais adequadas para comunicar as conclusões, pois, muitas das vezes, modos diferentes de representar uma mesma informação veicula mensagens muito distintas (NCTM, 1991).

Raciocínio inferencial informal (RII). Trata-se de um conceito relativamente recente na investigação realizada por educadores estatísticos. Zieffler, Garfield, delMas e Reading, (2008) compilaram algumas definições. Assim, Rubin, Hammerman e Konold (2006) definem RII como o raciocínio que envolve relações entre propriedades dos dados vistos de modo agregado (regularidades, desvios e tipo de variabilidade) com o tamanho da amostra e o controle do enviesamento. Para Pfannkuch (2006) o RII é a capacidade para interligar ideias estatísticas tais como distribuição e recolha de amostras, e tem como centro o ciclo de raciocínio empírico (Wild & Pfannkuch, 1999). Pelo seu lado, Bakker, Derry, e Konold (2006) sugerem um quadro teórico de inferência que amplia o significado de inferência estatística para permitir formas mais informais de raciocínio e incluir julgamentos com base no conhecimento contextual. Finalmente, Ben-Zvi (2006) compara RII à capacidade de argumentação, e enfatiza a necessidade deste tipo de raciocínio se basear em evidências extraídas dos dados. Procurando combinar essas perspetivas, Zieffler, Garfield, delMas e Reading (2008) apresentam como definição de raciocínio inferencial informal, o modo como os alunos utilizam os seus conhecimentos de Estatística para formular argumentos que apoiem as inferências sobre

populações desconhecidas com base em amostras observadas. Entendem o RII como um processo que inclui:

- Raciocínio sobre as possíveis características de uma população (por exemplo: forma, centro) com base numa amostra de dados;
- Raciocínio sobre possíveis diferenças entre duas populações com base na observação das diferenças entre duas amostras de dados (as diferenças são consequência de determinado fenómeno ou apenas se deve ao acaso).
- Raciocínio relativo à apreciação entre dois modelos ou declarações concorrentes de modo a escolher o(a) que é mais provável de corresponder à verdade.

Um exemplo do primeiro tipo de raciocínio foi analisado no estudo realizado por Bakker (2004) que a partir da tarefa “o crescimento de uma amostra” convida, numa segunda fase, a especular sobre as características gráficas da população com base no trabalho realizado sobre amostras de dados de dimensão cada vez maior.

O raciocínio sobre possíveis diferenças entre duas populações com base na observação das diferenças entre duas amostras de dados tem sido recorrente no trabalho de investigação de Pfannkuch (2006a, 2006b) e de Arnold e Pfannkuch (2010), recorrendo habitualmente à comparação de conjuntos de dados apresentados em diagramas de extremos e quartis, os alunos são convidados a responder a uma questão, como por exemplo “Em qual das populações se registam valores mais elevados?”, a resposta deve ser sustentada com base em pelo menos três argumentos extraídos da comparação.

Quanto ao terceiro tipo de raciocínio, Zieffler, Garfield, delMas e Reading (2008) apresentam uma tarefa de Zieffler e al. (2007) em que se recorre à apresentação de uma distribuição populacional relativa aos resultados de várias gerações de alunos a um determinado exame e se pede aos alunos para tomarem decisões sobre se um determinado resultado educativo observado numa amostra de dados (mudança na média de pontuação no teste) é consequência do acaso ou não.

No quadro 5 apresentam-se especificações de mobilização das componentes do RII em algumas tarefas típicas.

Quadro 5. Especificações de como determinadas tarefas mobilizam as componentes do RII (Zieffler, Garfield, delMas & Reading, 2008).

	Componentes do RII		
Tipo de tarefa	Fazer julgamentos ou previsões.	Utilizar ou integrar conhecimento prévio.	Requer a articulação de evidências baseadas em argumentos.
Estimar e desenhar o gráfico correspondente à distribuição da população.	Predizer características de uma população (forma, centro, dispersão) que deve constar da representação construída.	Trazer a intuição ou o que aprendeu anteriormente para prever as características da população (por exemplo, forma, uso de termos como enviesamento).	Requer uma explicação de como as características da população apresentadas no gráfico foram escolhidas.
Comparar duas amostras de dados com base em populações distintas.	Avaliar uma diferença entre as duas populações com base: em semelhanças ou diferenças nas duas amostras de dados.	Trazer a intuição ou o que aprendeu anteriormente para comparar as duas amostras de dados.	Requer uma explicação do porquê da existência ou inexistência de uma diferença nas duas populações.
Escolher entre dois modelos (duas declarações) concorrentes.	Aferir se os dados recolhidos na amostra fornecem mais suporte a um modelo que ao outro.	Trazer a intuição ou o que aprendeu anteriormente para escolher entre dois modelos concorrentes (por exemplo variabilidade da amostragem).	Requer uma explicação da razão pela qual se escolhe um modelo em detrimento de outro.

2.3. Dificuldades na aprendizagem da Estatística

Na perspetiva de Fernandes, Carvalho e Ribeiro (2007), as dificuldades dos alunos na realização de tarefas estatísticas têm duas causas: (i) uma, é inerente à Estatística, dado o facto de numa situação estatística não haver uma resposta única e decididamente certa ou errada e, agravada pelo facto de a linguagem, a notação e a terminologia utilizadas no seu âmbito serem ambíguas e confusas, e (ii) a outra, é inerente às experiências educativas vivenciadas, com maior ou menor pendor do papel do aluno na construção do conhecimento.

2.3.1. Dificuldades ao nível dos conceitos

População, amostra e variável estatística. Estudos com alunos de diferentes faixas etárias (Novaes, 2004), incluindo alunos do ensino superior (Fernandes, Carvalho e Ribeiro, 2007), constataam que estes tratam amostra como população. Para Novaes (2004) tal equívoco é consequência do modo como o assunto é tratado nos manuais escolares – enunciados de itens envolvendo dados referentes a uma amostra sem que se faça referência a estimativas para a população de onde se extraiu a amostra – e sugere que este erro poderá ser superado propondo situações práticas e de contexto real.

Para Cunha e Almeida (1996), as dificuldades de compressão do conceito população, manifestada pelo grupo de quatro alunos de 7.º ano que participou na experiência que realizaram, deve-se ao uso comum do termo população, como sendo um conjunto de pessoas, pelo que na tarefa proposta os alunos não reconheceram que a população em causa é o conjunto dos agregados familiares portugueses do ano civil 1989. Este grupo de alunos também sentiu dificuldades na identificação da variável, tendo assumido como variável a respetiva frequência absoluta: “número de famílias com uma, duas, etc., pessoas”; segundo as autoras da experiência, a deficiente compreensão poderá estar relacionada com as anteriores abordagens do conceito, uma vez que os alunos associam a variável o termo incógnita, não aceitando por isso associar variável a uma característica bem determinada. Para Martins e Ponte (2010) a confusão gerada em torno do termo variável é também consequência de algumas incorreções apresentadas em manuais escolares em que se afirma apresentar graficamente determinada distribuição, quando, na realidade, o que se expõe são dados sem proceder a qualquer redução, o que evidencia equívocos entre as noções de dado e de frequência.

Bakker (2004) constata que os alunos de 7.º e 8.º ano participantes no seu estudo, usam o termo “atípico” (*outlier*) como se fosse sinónimo de extremo inferior ou superior da distribuição e o termo “amostra” como se significasse apenas um grupo de elementos, não necessariamente um subconjunto de elementos representativos da população em estudo. Para Jacobs (1997) e Watson e Moritz (2000) as dificuldades de compreensão do conceito de amostra decorrem do significado relativamente simples do termo, em contextos exteriores à matemática escolar, por contraste com o significado mais técnico, quando se aplica como representação de uma população, com o objetivo de extrair conclusões para a mesma, ou seja, “o termo coloquial oferece uma amostra não representativa do termo estatístico” (Watson & Moritz, 2000, p. 68).

De acordo com Rubinn, Bruce e Tenney (1991) os alunos muitas vezes cometem um de dois tipos de erros quando trabalham com amostras aleatórias: (i) acreditam que uma amostra aleatória deve ser um retrato fiel da população ou (ii) têm pouca confiança na inferência com base em amostras aleatórias pois trata-se apenas de um subconjunto da população. Para que os alunos se situem a meio termo entre estes dois extremos, é necessário que reconheçam que amostras aleatórias de uma mesma população podem variar um pouco e não são necessariamente um retrato fiel da população, mas ainda assim com base numa amostra representativa se pode estabelecer inferências para a população. Para que os alunos desenvolvam esta intuição é fundamental que gerem amostras distintas de igual dimensão e analisem a variação de uma mesma medida estatística e a comparem com o parâmetro da população respetiva (GAISE, 2005).

Bill, Henderson e Penman (2010) constataam que os alunos se mostram incapazes de escolher a dimensão da amostra quando se trata de populações de grande dimensão, nomeadamente na definição da dimensão de uma amostra para a realização de uma sondagem nacional ou num estado/região, face a duas questões colocadas (Figura 3) apenas 17% dos alunos selecionou a opção (C) 1.500 eleitores na questão 1 e 4% dos alunos em ambas as questões. Sendo de salientar que 41% dos alunos considerou que a dimensão da amostra teria que ser menor no segundo caso pois a população também tinha menor dimensão, esta crença embora não se possa considerar uma conceção errada, é no entanto reveladora de uma compreensão parcial da relação entre a dimensão da amostra e a precisão da sondagem. Incrementar a dimensão da amostra nem sempre aumenta a precisão da sondagem, pois é possível que não se obtenha qualquer informação adicional significativa com essa alteração.

<p>Q1: Na população da Austrália de 20 milhões de cerca de 15 milhões (15.000.000 eleitores). Pretende-se fazer uma sondagem quanto à intensão de voto na próxima eleição nacional. Quantos eleitores considera que devem ser inquiridos? [Esta é uma pergunta sobre o tamanho da amostra; deve assumir que a amostra representa a população votante quanto ao género e a idade, etc.]</p> <p>(A) Aproximadamente 10% da população de votação (B) 15.000 (C) 1.500 (D) 150</p> <p>Q2: Sondagem, no estado de Queensland, em que existem cerca de 1,5 milhões (1.500.000 eleitores), quanto à intensão de voto na próxima eleição nacional. Quantos eleitores do estado de Queensland considera que devem ser inquiridos?</p> <p>(A) Aproximadamente 10% da população de votação (B) 15.000 (C) 1.500 (D) 150</p>
--

Figura 3. Tarefa proposta por Bill, Henderson e Penman (2010).

Medidas de tendência central. Relativamente a dados de natureza qualitativa não ordinal, Fernandes, Carvalho e Ribeiro (2007), num estudo que envolveu alunos do 7.º ano, destacam a inadequada determinação da média e da mediana, Boaventura e Fernandes (2004) observaram o mesmo erro, em alunos do 12.º ano. O equívoco mais comum entre os alunos na determinação da moda é indicar o valor da maior frequência absoluta ou da maior frequência relativa (Carvalho, 1998; Barros & Fernandes, 2001).

Garfield e Chance (2000) indicam que numa aplicação do teste Statistical Reasoning Assessment SRA, que consiste num conjunto de 20 itens de escolha múltipla, em que cada item correspondia a um problema estatístico ou probabilístico, identificaram os seguintes equívocos envolvendo a média: (i) a média é o dado numérico mais comum, (ii) a determinação da média envolve sempre a totalidade dos dados, independentemente da existência de dados atípicos (*outliers*), (iii) a média é o mesmo que mediana, e (iv) a comparação de grupos de dados assenta exclusivamente na análise da diferença das respetivas médias. Ainda no âmbito da determinação da média, outros erros foram encontrados por Carvalho (1998, 2001 a) na análise de produções de alunos do 7.º ano: i) não considerar a frequência absoluta de cada valor da variável no cálculo e ii) assumir que a média se refere à média dos valores das frequências dos dados.

Na determinação da mediana, Carvalho (1998, 2001a) salienta que os alunos comentem os seguintes erros: (i) não ordenam os dados, pois consideram que a mediana é o centro da listagem dos dados (não necessariamente ordenados); (ii) não tiveram em conta a frequência absoluta de cada valor no cálculo da mediana (quando os dados se apresentam numa tabela); (iii) calculam o dado central das frequências absolutas como se da variável se tratasse; e (iv) a mediana é o dado mais comum.

Boaventura e Fernandes (2004), no seu estudo envolvendo alunos do 12.º ano, constatou que a mediana se revelou a mais difícil das medidas de tendência central, seguindo-se a média e, finalmente, a moda. Habitualmente apresenta-se a mediana como sendo um medida estatística que divide os dados em duas metades, mas isso não é exatamente verdade, o que se pode afirmar é que a proporção de dados que são inferiores à mediana é no máximo 50%, e a proporção de dados que são superiores ou iguais à mediana é de pelo menos 50%. Para Cobo e Batanero (2000) o facto do algoritmo de cálculo da mediana não ser único, uma vez que depende do tipo de dados (de natureza ordinal ou não), da dimensão do conjunto de dados, do modo como se apresentam

(agrupados ou não) e do valor obtido também nem sempre ser único, implica que o estudo da mediana seja mais difícil.

Strauss e Bichler (1988) num estudo que envolveu alunos dos 8 aos 14 anos, com o objetivo de desenvolver a compreensão do conceito média, recorreram a tarefas com distintos tipos de dados e formas de apresentação, apelando a diferentes propriedades da média. Os resultados obtidos evidenciaram que a compreensão das várias propriedades melhora com o avanço da idade, não tendo sido clara a evidência, de que o tipo de dados e as diferentes formas de apresentação, influenciem a compreensão do conceito *média*. A generalidade dos alunos manifestou melhor desempenho nas tarefas que envolviam a compreensão das propriedades: (i) a média é um valor compreendido entre os extremos da distribuição; (ii) a média é influenciada pelo valor de cada um dos dados; e (iii) a média não tem que ser igual a um dos valores dos dados. As tarefas com menor desempenho corresponderam à compreensão das propriedades: (i) a soma dos desvios dos dados relativamente à média é zero; (ii) tem que se ter em conta os valores nulos no cálculo da média; e (iii) o valor da média é um "representante" dos dados a partir dos quais foi calculada.

As dificuldades que os alunos sentem em torno do conceito média, muitas das vezes não estão relacionadas com a determinação, mas com a compreensão do seu significado. Boaventura e Fernandes (2004) verificaram que muitos alunos do 12.º ano sentiam dificuldades na determinação da média de um conjunto de dados sendo conhecido as médias parciais de dois subconjuntos de dimensão distinta e que se complementam. A maioria das respostas incorretas corresponde ao cálculo da média simples das médias parciais, isto é, não consideraram a dimensão dos subconjuntos de dados. Cai (1995) ao solicitar a determinação de um dado desconhecido, num pequeno conjunto de dados apresentado sobre a forma de um pictograma, de modo a satisfazer o valor conhecido da média, obteve resposta correta por parte de metade dos alunos (com idades entre os 12-13 anos), sendo que cerca de 20% dos alunos recorreu a uma estratégia de tentativa e erro e pouco mais de um quarto dos alunos se mostrou capaz de inverter o algoritmo do cálculo da média. Boaventura e Fernandes (2004) propuseram a um conjunto de alunos do 12.º ano, a apresentação de um conjunto de quatro dados sendo conhecido o valor das medidas estatísticas (média, mediana e moda), e constataram que mais de 80% dos alunos se mostrou incapaz de responder corretamente, a maioria não teve em consideração o valor da mediana, indicando quatro dados que apenas satisfaziam a moda e a média.

Carvalho e César (2000) ao analisar as respostas de 84 díades, de alunos do 7.º ano, aos itens da figura 4, constataram que no primeiro item, 73% das díadas recorrem a argumentos que apelam ao significado matemático da média, envolvem uma das três propriedades da média apontadas por Strauss e Bichler (1988) como sendo as de mais fácil compreensão. As restantes díadas recorrem a argumentos que não apelam ao significado matemático do conceito média. No segundo item, a maioria das díades (56%) recorre apenas a argumentos que apelam ao significado matemático dos conceitos, mas pouco elaborados do ponto de vista concetual. Argumentam, recorrendo à comparação da grandeza dos valores das duas medidas estatísticas e não procedem à análise da distribuição de salários (assimétrica com cauda à esquerda).

Numa empresa escolheram-se, ao acaso, cinco dos seus empregados para se fazer um estudo acerca dos salários. Obtiveram-se os seguintes resultados:

Empregado	A	B	C	D	E
Salários (escudos/mês)	54 000	42 000	60 000	48 000	180 000

1. Achas que os cinco empregados estão de acordo quando se disser que a maioria dos empregados dessa empresa tem um salário igual à média? Porquê?
2. O que achas que representa melhor os salários nesta empresa, a média ou a mediana? Porquê?

Figura 4. Tarefa a realizar em díade proposta por Carvalho e César (2000).

Com o mesmo conjunto de alunos, Boaventura e Fernandes (2004) salientam que ao solicitar-se a posição das medidas centrais em distribuições simétricas e assimétricas, os alunos mostraram-se incapazes de recorrer à ideia “centro de massa” para posicionar a média em relação à mediana ou recorrer à informação fornecida no histograma para estimar o valor da média. Nas distintas distribuições, os alunos mostram-se capazes de identificar a classe modal. Face a distribuições simétricas, nomeadamente com forma de sino, existe uma maior percentagem de alunos que identifica corretamente a posição das medidas, do que em distribuições assimétricas. Para estes investigadores a maioria das dificuldades manifestadas advém do trabalho superficial realizado no âmbito da Estatística e da inexistência de situações que apelem ao desenvolvimento concetual dos conceitos estatísticos, nos manuais escolares, recurso primordial na preparação do ensino.

Novaes (2004) reconhece que os alunos, de um Curso Superior de tecnologia em Turismo, mobilizam erroneamente os conceitos de média e mediana, nomeadamente quando trabalham com distribuições simétricas, e considera que a utilização dos termos no quotidiano e nos meios de comunicação com vários significados distintos pode ser a causa deste erro.

Dos estudos anteriores conclui-se que os alunos manifestam significativas dificuldades quando as situações propostas exigem que sejam aplicadas as medidas de tendência central a novas situações (inversão do algoritmo da média, distribuição que satisfaça determinadas requisitos, seleção da estatística que melhor represente uma distribuição, etc.) ficando a maioria a um nível de conhecimento que se designa por processual.

2.3.2. Dificuldades ao nível de tabelas e gráficos estatísticos

Representação tabelar. Embora não abundem estudos relativos às dificuldades inerentes à construção, leitura de tabelas de frequência e interpretação das diferentes frequências utilizadas, Batanero (2001) considera que a sua construção, constitui uma primeira redução, pois perde-se a visão individualizada dos dados obtendo uma distribuição de frequências (absolutas, relativas ou acumuladas) o que constitui uma visão agregada dos dados. A autora considera que a generalidade das dificuldades dos alunos advém do facto de ser necessária a compreensão do que é uma distribuição de frequências, objeto matemático complexo que aborda o conjunto (população ou amostra).

No estudo realizado por Carvalho (1996) os alunos de 7.º ano envolvidos não tiveram dificuldades na determinação das frequências absolutas, mas apenas um quarto destes se mostrou capaz de determinar as respetivas frequências relativas, tendo a maioria dos alunos considerado a frequência absoluta como denominador da fração. Num conjunto de alunos do 7.º ano, Fernandes, Carvalho e Ribeiro (2007) reconheceram que um dos erros recorrente na construção de tabelas de frequências foi assumir os dados de valor numérico zero como elemento neutro no cômputo total das frequências.

Reading (1999) reconheceu que os alunos, do ensino básico e secundário, perante o mesmo conjunto de dados, apresentado de modo tabelar e de modo gráfico, produzem descrições distintas dos dados em análise. Perante a representação gráfica dos dados as descrições tendem a ser mais corretas à medida que aumenta o nível de escolaridade e recorrem menos a juízos de valores na descrição, valorizando mais a utilização de noções estatísticas.

Representações gráficas. Quando os alunos constroem determinado gráfico, realizam uma série de ações e usam conceitos e propriedades que variam mediante o tipo de gráfico (Ruiz & al., 2009), pelo que existem dificuldades e equívocos específicos a cada representação. Diferentes estudos constataam que a construção, compreensão e interpretação da generalidade destas representações não é alcançada durante o ensino básico, Fernandes, Carvalho e Ribeiro (2007) verificaram que alunos do 7.º ano revelam dificuldades na definição da escala a utilizar na construção dos gráficos de barras e na leitura e interpretação de pictogramas; Morais (2010) constatou que muitos dos alunos do 9.º ano, envolvidos no estudo, optam por uma representação desadequada face aos dados apresentados, tal como Ruiz e al. (2009) o tinham constatado num grupo de alunos (futuros professores).

A interpretação de pictogramas apresenta desafios aos alunos: (i) pictograma que envolve a repetição de um ideograma, quando a legenda é definida *a priori*, sendo então necessário estabelecer a relação entre cada ideograma e a frequência absoluta que representa, acresce-lhe o facto de na maioria das vezes o ideograma ser uma fração (Curcio, 1989) e (ii) pictograma cujo ideograma varia proporcionalmente à frequência, ocorre muitas vezes a alteração do ideograma nas suas várias dimensões (altura, largura, comprimento), o que conduz a uma representação incorreta. A construção de um pictograma quando envolve outras competências matemáticas, tais como, noção de área, volume, semelhança de figuras, torna-se mais complexa (Carvalho, 2009). A título de exemplo, no estudo desenvolvido por Wu (2004) o que envolveu alunos entre os 13-15 anos, 72,2% dos alunos tiveram dificuldades com esta representação, os alunos desconheciam que o tamanho do ideograma representa proporcionalmente frequência/quantidade, e um pictograma tridimensional poderia exagerar a informação dramaticamente caso se altere proporcionalmente em mais do que uma dimensão.

A construção e compreensão de um gráfico circular está intimamente relacionada com a compreensão do raciocínio proporcional, da noção de ângulo, da capacidade de trabalhar com números fracionários, percentagens e da correta manipulação de material de desenho e medida (Carvalho, 2001b; Curcio, 1989; Espinel & al, 2009), pelo que, os erros cometidos pelos alunos na construção, estão associados à marcação dos ângulos e à não proporcionalidade entre o sector circular e a frequência que representa (Carvalho, 2001b; Espinel & al, 2009; Morais, 2010).

Wu (2004) constata que cerca de 90% dos alunos não conhecia uma ou mais das seguintes convenções relativas à escala: (i) inicia-se na origem do referencial, (iii) res-

peita-se a unidade escolhida ao longo de todo o eixo e (iii) em qualquer gráfico estatístico se recorre a uma escala; reconhece ainda que a generalidade dos alunos não indica título e/ou não procede à etiquetagem dos eixos. Para Konold e Higgins (2003), o facto de os alunos não respeitarem a unidade escolhida ao longo de todo o eixo ocorre porque muitas vezes a construção da escala é feita ao mesmo tempo que a representação dos dados, pelo que, muitos alunos têm dificuldades em estabelecer a diferença entre os valores ao longo da escala e os dados que esta está a representar. As dificuldades manifestadas na construção e interpretação de diferentes representações gráficas persistem para além do ensino básico, a escolha de uma representação desadequada face aos dados apresentados também foi observada por Ruiz e al. (2009) num grupo de alunos (futuros professores primários).

No estudo realizado por Wu (2004) o habitual equívoco entre o gráfico de barras e o histograma foi cometido por cerca de 45% dos alunos. Reconhece ainda que à medida que se avança no nível de ensino a percentagem de alunos que confunde as representações diminui, pelo que a capacidade dos alunos para diferenciar gráficos com aparências similares aumentar com a idade. Lee e Meletiou (2003) ao propor questões, como a apresentada na figura 5, a alunos do ensino superior, identificaram as seguintes dificuldades, na construção de histogramas: (i) o “pé” da barra corresponde a um dado individual em vez de corresponder a um conjunto de dados agrupados numa classe e (ii) apresentam um diagrama de dispersão assumindo duas variáveis no contexto da situação em estudo.

Ao construir um histograma que descrever a distribuição de salários para os indivíduos com 40 ou mais anos, que ainda não se aposentaram:

- a) Que grandeza associa ao eixo vertical?
- b) Que grandeza associa ao eixo horizontal?
- c) Qual seria a forma adequada da distribuição dos salários? Explica a tua resposta.

Figura 5. Questão proposta por Lee e Meletiou (2003).

Na construção de histogramas, Espinel e al. (2009) constataram que os alunos (futuros professores primários) cometeram os seguintes erros, para além dos habituais problemas relacionados com as escalas: (i) barras em separado, (ii) etiquetagem incorreta das barras e (iii) omissão dos intervalos de frequência nula. No estudo realizado por

Morais (2010) os alunos cometam apenas os dois primeiros erros. Segundo Shaughnessy (1996) como o histograma está associada a variáveis quantitativas contínuas, as dificuldades de construção e interpretação são consequência do conceito de eixo contínuo.

No diagrama de caule-e-folhas, segundo Espinel e al. (2009) as dificuldades manifestadas pelos alunos relacionam-se com a noção de número e com a eventualidade do caule não possuir dados ou a estes ser atribuído o valor zero.

Quanto à representação gráfica, diagrama de extremos e quartis, os investigadores Bakker, Biehler e Konold (2005) reconhecem que alunos do 7.º ano, 8.º ano e universitários, manifestam dificuldades, na sua construção, inerentes à determinação das medidas de localização necessárias. Constataram que existe um conjunto significativo de alunos que ao interpretar um diagrama de extremos e quartis não referencia qualquer medida de dispersão, nomeadamente a amplitude interquartil ou o intervalo interquartil, o que entra em contradição aparente com a pertinência visual do intervalo interquartil na representação. As dificuldades de interpretação resultam de três aspetos (i) é necessário recorrer a uma visão agregada dos dados, tal como nas restantes representações; (ii), apresenta-se a densidade relativa (dispersão e variação dos dados) em vez da frequência; (iii) a densidade relativa relaciona-se de modo inverso com o comprimento da componente do diagrama. Estes investigadores sugerem uma transição gradual da visão individual dos dados para uma visão agregada, propondo que se apresente, em paralelo com o diagrama de extremos e quartis, o respetivo diagrama de pontos com a localização dos quartis, o que fomenta a compreensão da relação entre a densidade e o comprimento de cada componente. Watson, Fitzallen, Wilson e Creed (2008) propõem uma representação prévia “hat plot” (diagrama chapéu) (Figura 6), disponível no *software* do Tinker-Plots, como meio de passagem para o diagrama de extremos e quartis. Defendem que esta representação permite aos alunos o desenvolvimento intuitivo de distribuição de dados, pois tendem a observar e a usar a expressão “aglomerado modal” para a zona central da representação, e identificam três zonas da distribuição – centro/meio, abaixo e acima, tal como tinha acontecido no estudo de Konold & al. (2002). Consideram que a análise da copa da representação em forma de chapéu “*hat plot*” (o conjunto dos 50% dos dados centrais) em relação ao conjunto dos dados é uma medida central. Concluíram que os alunos consideram útil utilizar a informação relativa à copa da representação na comparação de dois ou mais conjuntos de dados, ao resumir os dados, a descrever a dispersão e a concentração dos dados, o que desenvolveu uma compreensão mais ade-

quada das medidas centrais (média e mediana). A posterior inclusão de um marcador vertical correspondente ao valor da mediana, permitirá a passagem para o diagrama de extremos e quartis, com um maior grau de consciência e de compreensão.

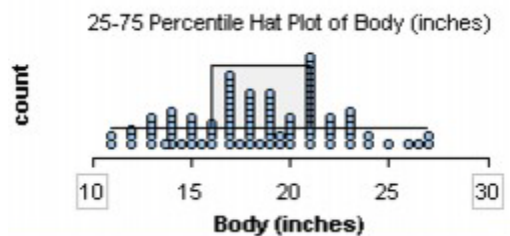


Figura 6. Exemplificação de um “*Hat plot*” / Diagrama chapéu.

Para Espinel e al. (2009) a construção das várias representações, apresenta dificuldades específicas ao estar associada à construção de tabelas e ao envolver variados conceitos, como escalas, origem dos eixos, variável independente e dependente, coordenadas, variáveis discretas e contínuas e distribuição de frequências. Artega, Batanero e al. (2009) consideram que uma possível explicação para todas estas dificuldades é o facto da simplicidade da linguagem gráfica ser apenas aparente, uma vez que qualquer gráfico é, por um lado, um modelo matemático baseado na ideia de distribuição de frequências, e por outro lado, uma mesma representação gráfica; por exemplo, um gráfico de barras ou um histograma pode recorrer a distintos objetos matemáticos (frequências absolutas, relativas, acumuladas, medidas estatísticas de resumo).

2.4. Avaliação em Estatística

O relatório de avaliação, *Measuring What Counts* (MSEB, 1993) indica três princípios básicos para a avaliação em Matemática. Segundo Shaughnessy, Garfield e Greer (1996) quando se trabalha em Estatística, deverão ser tidos em consideração dois deles:

- *O Princípio do Conteúdo:* A avaliação deve reflectir o conteúdo de Estatística que é mais importante para os alunos aprenderem.
- *O Princípio da Aprendizagem:* A avaliação deve melhorar a aprendizagem da Estatística e ser apoio à prática de um bom ensino.

Esses princípios implicam o uso de formas alternativas de avaliação a fim de fornecer informações mais completas sobre o que os alunos aprenderam e sobre o que são capazes de fazer com o seu conhecimento. A sua concretização pressupõe uma visão da avaliação como processo dinâmico, que continuamente fornece informação sobre o progresso do aluno para a realização dos objetivos de aprendizagem (NCTM, 1991). Esta visão da avaliação reconhece que, quando as informações obtidas são consistentes com as metas de aprendizagem e são usadas para extrair ilações para ajustar o processo de ensino-aprendizagem, podem contribuir para a melhoria das aprendizagens bem como documentá-las. A avaliação tem assim como objetivo captar o modo como os alunos pensam, argumentam, e aplicam a sua aprendizagem, ou seja “criar uma «biografia» sobre a aprendizagem dos alunos, constituindo assim uma base para melhorar a qualidade do ensino” (NCTM, 1991, p. 240). Deste modo, a avaliação assume-se como parte integrante do ensino-aprendizagem, e não apenas como uma atividade a realizar no culminar deste processo.

Para Garfield (1994), o principal objetivo da avaliação reside na melhoria das aprendizagens do aluno. Face a esse objetivo a autora enumera os seguintes seis efeitos secundários da recolha de informação para avaliação: fornecer informações específicas aos alunos sobre como aprenderam um tópico em particular e onde estão a revelar dificuldades; fornecer informações ao professor sobre o quanto o grupo turma parece entender um determinado tema e quais as tarefas adicionais que poderá ser necessário introduzir, ou se é adequado avançar para outro tópico; fornecer informações de diagnóstico aos professores sobre a compreensão de cada aluno ou dificuldades na compreensão de um novo conteúdo; fornecer informações aos professores sobre as percepções dos alunos e as reações da turma, quanto a um assunto, ou uma tarefa particular; fornecer um indicador geral de sucesso dos alunos na realização dos objetivos do tópico em análise e para ajudar os alunos a determinar globalmente os seus pontos fortes e fracos quanto aos conteúdos do tópico em análise.

Uma preocupação expressa por Garfield (1994) e Holmes (2002) é a necessidade de elaborar itens e tarefas que avaliem os níveis mais profundos de compreensão estatística. Garfield constata que:

À medida que os objetivos da educação estatística mudam para objetivos mais abrangentes e ambiciosos, como o desenvolvimento de pessoas capazes de pensar em termos da estatística, que podem aplicar o seu

conhecimento na resolução de problemas reais, é revelada a incompatibilidade entre a avaliação tradicional e o que se deseja como resultado por parte dos alunos. (p. 2)

Garfield (1994), em colaboração com Gal (Garfield & Gal, 1999), sugere um conjunto de cinco etapas a fim de facilitar uma abordagem mais ponderada da avaliação de conceitos e competências estatísticas:

- O que vou avaliar? Que aspetos pretendo avaliar? Estes podem ser divididos em conceitos, capacidades, aplicações, atitudes e crenças.
- Qual é o propósito desta avaliação? Pode contribuir predominantemente para recolher fornecer informação para professor ou para o aluno.
- Que método de avaliação vou utilizar? Os exemplos incluem teste, relatório, projeto individual ou em grupo, trabalho escrito ou portefólio.
- Quem vai realizar a avaliação?
- Qual será (ou quais serão) a(s) ação(ões) tomada(s) como resultado da avaliação e do *feedback* aos alunos?

Analisando um dos quatro exemplos apresentados por Garfield é possível apercebermo-nos da utilização deste conjunto das cinco etapas (Quadro 6).

Quadro 6. Exemplificação do processo de avaliação (Garfield, 1994).

O que vou avaliar?	Capacidade do aluno para aplicar técnicas básicas de análise exploratória de dados.
Propósito	Determinar se os alunos são capazes de aplicar as suas capacidades na recolha, análise e interpretação de dados.
Método	Um projeto, onde são dadas instruções quanto ao tamanho da amostra, o formato do relatório, e assim por diante.
O que é para avaliar?	Primeiro, o aluno conclui uma auto ponderação usando uma cópia da ficha de avaliação que o professor irá usar, e que foi distribuída, antes de completar o projeto. Em seguida, o professor avalia o projeto de acordo com cada uma das seis seguintes categorias: comunicação, interpretação, representação dos dados, cálculos estatísticos, tomada(s) de decisão, resultados e conclusões.
Ações / feedback	As pontuações são atribuídas a cada categoria e dadas a conhecer aos alunos, juntamente com observações escritas, para que eles possam aprender com esse <i>feedback</i> e incluir tais considerações no seu projeto seguinte.

Holmes (2002) defende que o uso de dados reais, como por exemplo *CensusAtSchool*, torna possível elaborar perguntas e tarefas que avaliam os níveis mais profundos de entendimento. Por exemplo, apresenta a tarefa da figura 7 como adequada para a avaliação do seguinte objetivo: Especificar um problema e seu planejamento e identificar questões que possam ser abordadas através de métodos estatísticos.

Célia quer comparar as alturas de todos os alunos com a mesma idade no Reino Unido. Descreve um esboço de um plano para o seu projeto.

Deve incluir:

- Como irá obter os dados de sua própria turma e como irá recolher a informação adequada a partir do *site* "CensusAtSchool";
- Como irá representar os dados;
- Que medida(s) estatística(s) irá utilizar para descrever os dados;
- Que tipo de conclusão poderá estabelecer.

Figura 7. Tarefa proposta por Holmes (2002).

Depois da realização da tarefa será sempre possível incluir, como trabalho de aula, o seguinte questionamento: “Face ao esboço do plano que apresentaste para a Célia, consideras que é adequado? O que mudavas no que apresentaste? Esqueceste-te de alguma coisa?...” Para Holmes este momento tem naturalmente mais impacto nas aprendizagens dos alunos, no caso deste trabalho ter subjacente a avaliação formativa e ser possível incorporar no trabalho do aluno alterações após este questionamento.

No âmbito da construção de itens de resposta fechada para a avaliação do pensamento estatístico, Cobb (1998) oferece cinco princípios para a sua conceção: (i) aumentar o número de opções; (ii) pedir julgamentos comparativos, e não apenas de uma determinada categoria; (iii) envolver várias características dos dados simultaneamente; (iv) envolver dois ou mais modos de pensar a estatística (por exemplo, representações distintas, informação verbal, numérica, tabelar, etc.) e (v) pedir comparações dinâmicas (por exemplo, face a uma investigação estatística descrita, avaliar como: (1) essencial, (2) desejável, (3) muito dispendioso e (4) impossível, diferentes descrições de alteração da investigação).

Para Gal e Garfield (1999) e Garfield e Change (2000) devem valorizar-se os seguintes instrumentos:

- Testes, que incluam cálculos, gráficos ou questões de desenvolvimento e argumentação;
- Produções escritas em que os alunos resumam o que aprenderam e o que foi mais difícil numa aula em particular;
- Projetos individuais ou de grupo;
- Provas abrangendo um vasto leque de material;
- Observação periódica da atitude;
- Relatório escrito das atividades desenvolvidas em sala de aula;
- Questões abertas/problema estatístico (análise crítica de uma situação ou notícia de foro estatístico);
- Mapas de conceitos;
- Portefólios que incluam uma seleção de diferentes trabalhos realizados.

Gal e Garfield (1999) e Garfield e Change (2000) consideram que, embora estes instrumentos de avaliação possam ser utilizados para atribuir uma classificação. A sua utilização deverá ter sempre presente as aprendizagens, isto é, ajudar os alunos a aprender e a melhorar seu desempenho. Os dados recolhidos através destes instrumentos informam o professor sobre o que os alunos estão a aprender, as suas competências e as dificuldades sentidas.

Na década de 90, para a avaliação em larga escala do conhecimento estatístico de alunos do secundário, surgiram nos Estados Unidos diferentes instrumentos. Um deles é o Advanced Placement (AP in Statistics), exame facultado pela instituição College Board, composto por 35 itens de escolha múltipla e 6 itens de desenvolvimento incluindo uma tarefa investigativa, envolvendo os seguintes tópicos: Exploração de dados – observação de regularidades e inícios de tendências; Planeamento de um estudo - decidir o que recolher e como recolher; Antecipar regularidades – produção de modelos (usando probabilidade e simulação) e Inferência estatística - modelos de confirmação. Os itens de desenvolvimento são avaliados de modo holístico tendo em consideração os seguintes critérios: (i) O aluno demonstra conhecimento dos conceitos estatísticos envolvidos? (ii) O aluno apresenta uma explicação clara do que foi feito no âmbito da análise e o porquê dessa decisão? e (iii) O aluno expressa de modo claro as suas conclusões? Deste modo valoriza-se o conhecimento estatístico e a comunicação das ideias estatísticas de modo claro em detrimento da correta realização de cálculos estatísticos,

tal como defendido por Scheaffer (2000). Outro instrumento é o Statistical Reasoning Assessment (SRA), desenvolvido no âmbito do projeto ChancePlus para avaliar a eficácia do novo currículo em estatística e o grau de consecução das suas metas de aprendizagem ao nível do secundário. É composto por 20 itens de escolha múltipla. Cada item descreve um problema estatístico ou probabilístico e apresenta várias opções de resposta, tanto corretas como incorretas. A maioria das opções de resposta inclui uma explicação de raciocínio e/ou a explicação de uma decisão em particular. Cada aluno deverá escolher a opção de resposta que melhor corresponde ao seu próprio raciocínio para cada problema. Cada item é então categorizado face a oito tipos capacidades de raciocínio correto ou a oito tipos de equívocos/raciocínios incorretos (Quadro 7).

Quadro 7. Categorias de raciocínio correto e de equívocos no SRA.

Tipos de raciocínio corretos	Tipos de equívocos/raciocínios incorretos
1. Interpreta corretamente probabilidades	1. Equívocos envolvendo média
2. Determina a média adequadamente (com a existência ou não de valores atípicos)	a. Confunde com a moda
3. Calcula corretamente uma probabilidade	b. Não tem em consideração a existência de valores atípicos no cálculo da média
a. Probabilidade como rácio	c. Compara grupos de dados com base nas suas médias
b. Utiliza o raciocínio combinatório	d. Confunde com a mediana
4. Compreende a independência de acontecimentos	2. Orientação para um resultado esperado (dificuldade em ver o resultado como uma série de eventos)
5. Compreende a variabilidade do processo de amostragem	3. Boas amostras representam necessariamente uma percentagem elevada da população
6. Distingue entre correlação e causalidade	4. Lei dos pequenos números (utilização de pequenas amostras)
7. Interpreta corretamente tabelas de dupla entrada	5. Equívoco da representatividade (amostra como fotografia da população)
8. Compreende a importância de amostras grandes	6. Correlação implica causalidade
	7. Engano da equiprobabilidade de acontecimentos
	8. Grupos de dados são comparáveis apenas se forem da mesma dimensão

“Ajudar os alunos a colocar questões que lhes interessam e que possam prosseguir de forma produtiva é um desafio para o professor.”

Konold e Higgins (2003)

Capítulo 3

Unidade de Ensino

A unidade de ensino sobre a qual incide este estudo, relativa ao tópico “Planeamento estatístico”, tem por base tarefas elaboradas pela equipa de professores das turmas-piloto, no âmbito da experimentação do novo programa de Matemática. Assim, começarei por apresentar o enquadramento curricular e os objetivos gerais a atingir, incluindo a hipótese de ensino-aprendizagem subjacente à unidade de ensino, seguindo-se as tarefas, a dinâmica da aula e a avaliação.

3.1. Enquadramento curricular e objetivos

A educação estatística, segundo Martins e Ponte (2010), deve seguir os seguintes princípios metodológicos: (i) usar contextos significativos, se possível a utilização de dados reais recolhidos pelos alunos; (ii) fomentar a comunicação oral e escrita – discussão das tarefas em pequeno e grande grupo bem como a elaboração de relatórios finais e (iii) propor investigações ou projetos estatísticos investigativos desenvolvidos por grupos de alunos. Tais princípios metodológicos pretendem levar o aluno realizar a análise e a interpretação de conceitos estatísticos em relação aos respetivos contextos. A negociação de significados e a reformulação de raciocínios incorretos possibilita que o aluno entenda o significado do valor estatístico encontrado na situação proposta. A discussão em grande grupo “fomenta a partilha e o debate de ideias, a sistematização dos conceitos e a institucionalização de conhecimentos” (p. 16).

A unidade de ensino sugerida pelos professores das turmas-piloto assenta num conjunto de quatro tarefas de carácter exploratório e investigativo que visam proporcionar aos alunos experiências diversificadas e significativas tendo em vista o desenvolvimento das suas capacidades de planeamento estatístico, integrando os conhecimentos desenvolvidos no âmbito do tema “Tratamento de dados”, abordado no 7.º ano.

As tarefas de carácter exploratório e investigativo, segundo Cobb (1993) têm a vantagem dos alunos serem chamados a participar na produção dos dados, pelo que os dados resultantes são “frescos”, não são as “sobras” de outra pessoa. Cria-se potencialmente uma situação em que existem quatro vantagens: (i) os alunos têm “em mão” um conjunto de dados, criado por si o que quase sempre é sinónimo de motivação para a sua análise; (ii) o contato com a variabilidade é imediato e concreto; (iii) contacto com a fiabilidade dos instrumentos de recolha – limpeza de dados; e (iv) contato com problemas que surgem ao nível do planeamento. A desvantagem é, por vezes, alguns dos conjuntos de dados produzidos tenderem a ser semelhantes entre si, nomeadamente ao nível da complexidade.

A unidade de ensino deste estudo insere-se no âmbito de cinco aspetos da competência matemática do *Currículo Nacional do Ensino Básico* (ME, 2001), que todos os alunos devem desenvolver, ao longo dos três ciclos: (i) a predisposição para recolher e organizar dados relativos a uma situação ou a um fenómeno e para os representar de modos adequados, nomeadamente através de tabelas e gráficos e utilizando as novas tecnologias; (ii) a aptidão para ler e interpretar tabelas e gráficos, à luz das situações a que dizem respeito e para comunicar os resultados das interpretações feitas; (iii) a tendência para dar resposta a problemas com base na análise de dados recolhidos e de experiências planeadas para o efeito; (iv) a aptidão para realizar investigações que recorram a dados de natureza quantitativa, envolvendo a recolha e análise de dados e a elaboração de conclusões; e (v) o sentido crítico face ao modo como a informação é apresentada (p. 64).

No que se refere aos aspetos específicos do 3.º ciclo, a unidade de ensino procura promover: (i) a compreensão das noções de moda, média aritmética e mediana, bem como a aptidão para determiná-las e para interpretar o que significam em situações concretas; (ii) a sensibilidade para decidir quais das medidas de tendência central são mais adequadas para caracterizar uma dada situação; (iii) a aptidão para comparar distribuições com base nas medidas de tendência central e numa análise da dispersão dos dados; e (iv) o sentido crítico face a afirmações baseadas em amostras não representativas

(p.65). No âmbito desta unidade de ensino, tendo em atenção o programa de matemática do ensino básico, pretendo que os alunos se envolvam na realização de estudos estatísticos com o objetivo de: (i) compreender informação de natureza estatística e desenvolver uma atitude crítica face a essa informação; (ii) interpretar os resultados obtidos; e (iii) resolver problemas e comunicar em contextos estatísticos (ME, 2007, p. 59) e desenvolver as capacidades transversais: (i) resolver problemas em contextos matemáticos e não matemáticos, adaptando, concebendo e pondo em prática estratégias variadas, discutindo as soluções encontradas e os processos utilizados; (ii) raciocinar matematicamente, formulando e testando conjecturas e generalizações, e desenvolvendo e avaliando argumentos matemáticos incluindo cadeias dedutivas; e (iii) comunicar oralmente e por escrito, recorrendo à linguagem natural e à linguagem matemática, interpretando, expressando e discutindo resultados, processos e ideias matemáticas (ME, 2007, p. 62).

Hipótese de ensino-aprendizagem Sendo o objetivo principal do ensino da Estatística a compreensão do seu papel nos processos de pesquisa e análise de dados, nomeadamente a perceção da variabilidade dos dados, considero fundamental a vivência de situações que permitam aos alunos trabalhar com dados recolhidos em função de questões significativas. Durante a realização da unidade de ensino espero que os alunos recorram aos seus conhecimentos prévios e ao seu conhecimento do contexto de cada situação proposta para resolverem as tarefas apresentadas e deste modo desenvolverem a sua capacidade de planeamento e de análise de dados, e a compreensão da natureza da Estatística como ciência, assim como do seu contributo para resolução de situações que envolvam incerteza.

3.2. Tarefas

A unidade de ensino é constituída por um conjunto de tarefas de natureza exploratória e investigativa, em contextos de semirealidade e realidade, dando ênfase à construção do conhecimento com base no desenvolvimento da compreensão, partindo do conhecimento informal e intuitivo dos alunos para os levar a construir conhecimento formal. Tomou-se a decisão de reajustar o conjunto de tarefas correspondente ao tópico planeamento estatístico proposta pelos professores das turmas-piloto¹, atendendo à tipologia de alunos e às condições de trabalho na escola, nomeadamente procedendo-se à

¹ Disponível no sítio da DGIDC - Direção-Geral de Inovação e de Desenvolvimento Curricular.

inclusão da tarefa “Limpar os dados” e à análise de um texto “As vendas estão a correr bem?” adaptados de Martins e Ponte (2010).

A unidade de ensino é composta por seis tarefas com contextos diversos e que apelam a distintas representações e cumprem os diferentes objetivos. A planificação detalhada da unidade de ensino (Quadro 8) é apresentada no anexo 2.

A tarefa 1 “População e amostra” (Anexo 7) é constituída por três partes. As duas primeiras são discutidas no grupo turma, sendo a parte A relativa à importância das amostras e razões da sua utilização e a parte B referente à importância da representatividade das amostras. Com a discussão destas duas partes da tarefa espera-se que a maioria dos alunos compreenda que no momento da recolha de dados se procura recorrer a amostras representativas e como se deve proceder para a sua obtenção. A parte C é constituída por seis questões, cada uma relativa à análise da representatividade de uma amostra face a uma dada situação. O objetivo da tarefa é distinguir população de amostra e ponderar a representatividade de uma amostra em relação à respetiva população.

Na tarefa 2 “Limpar dados” (Anexo 10) espera-se que os alunos, com a análise dos dados propostos, tomem consciência da necessidade de olhar criticamente para o conjunto de dados de modo a detetar erros/imprecisões antes de os começar a tratar. É fundamental que tomem conhecimento da pertinência desta análise crítica, dada a natureza e tipologia dos dados, para evitar pôr em causa a análise posterior. A utilização de valores com imprecisões põe em causa a fiabilidade das conclusões que se venham a extrair. No entanto, a diminuição da dimensão da amostra pode trazer prejuízo à sua representatividade, sendo necessário por vezes continuar a “recolha” de dados.

Com a tarefa 3 “As vendas estão a correr bem?” (Anexo 11) pretende-se que a análise do texto proporcione um momento de discussão e reflexão relativa aos seguintes aspetos: (i) as etapas de uma investigação estatística, (ii) a contextualização dos dados, (iii) a importância da escolha das medidas de resumo e de representações utilizadas, (iv) o facto de a validade e a fiabilidade das conclusões depender da informação disponível, sendo passível que um estudo motive novas questões. A tarefa 3 constitui, neste ponto do ano, o primeiro contacto com a noção de análise exploratória de dados, levando os alunos a tomar os dados como um todo, promovendo uma reflexão sobre a forma como estes se distribuem, apoiando-se nas medidas de localização e em diferentes formas de representação, bem como assumir que a análise que se espera não se será meramente descritiva, sendo valorizado o contexto, a pertinência da informação recolhida, as conje-

turas sobre os dados de que se dispõe, e inclui a discussão quanto à viabilidade das conclusões extraídas.

Na Tarefa 4, “Um estudo na escola”² (Anexo 12), os alunos irão realizar em grupo um estudo estatístico, tendo que interpretar a situação, definir um critério para a escolha de uma amostra representativa e organizar a recolha e o tratamento dos dados. Por fim é-lhes solicitado que relacionem os dados obtidos e produzam um relatório com a informação mais significativa sobre a população escolar. É nesta tarefa que os alunos irão pela primeira vez estar envolvidos num ciclo de investigação, sendo fundamental valorizar que a representação e a interpretação dos dados permitam a identificação de propriedades verificadas pelos dados recolhidos. Uma outra fase importante será a discussão e a reflexão da variabilidade da amostragem, a formulação de conjeturas e a validade da generalização das conclusões para a população.

A Tarefa 5 “A frequência das vogais na língua portuguesa” (Anexo 13) propõe a análise de dois textos e solicita que se estude com base em amostras extraídas dos textos a frequência das vogais na língua portuguesa, alertando assim para a importância de se identificar e minimizar possíveis fontes de enviesamento na recolha dos dados e a variabilidade existente no processo de amostragem.

Finalmente, a sexta e última tarefa “Previsões” (Anexo 14) propõe aos alunos a realização de um pequeno estudo estatístico onde a análise exploratória dos dados volta a assumir um papel essencial no trabalho a desenvolver. Com base na análise dos dados, os alunos são chamados a inferir e avaliar a argumentação relativa a essas mesmas inferências.

O objetivo das três tarefas finais passa pela tomada de consciência do papel dos métodos estatísticos como meios poderosos de tomada de decisão, isto é compreender que, ao realizar um estudo estatístico, nomeadamente, um estudo de mercado com base numa amostra representativa, se podem fazer previsões, isto é, inferir, para a população correspondente com algum grau de incerteza associado. Dado o elevado grau de desafio e a abertura das tarefas propostas, em especial as Tarefas 4 e 6 que propõem o desenvolvimento de uma investigação estatística para a resolução de um problema e a comunicação da informação que foi passível extrair nesses contextos, é expectável o envolvi-

² Dada o elevado número de alunos da escola em causa, optei por restringir a população em estudo aos alunos do 3.º ciclo.

mento ativo dos alunos desde a primeira fase do processo – formulação das questões a resolver (Abrantes, Ponte, Fonseca & Brunheira, 1999).

A unidade de ensino iria decorrer durante cerca de três semanas letivas. Seriam disponibilizados ao todo 9 blocos de noventa minutos (ou 18 meios blocos de 45 minutos). O quadro 8 apresenta a sequência de tarefas que integram a unidade de ensino e o modo de trabalho em cada aula.

3.3. Dinâmica da sala de aula

A minha expectativa era que os contextos das tarefas pudessem motivar os alunos, dado o seu interesse e relevância, bem como fomentar uma dinâmica de aula que visasse o trabalho de grupo e a discussão do grande grupo. Os alunos estavam em grupos de quatro elementos, dado ser esta a organização que considero mais adequada. De acordo com Programa de Matemática do Ensino Básico (ME, 2007), ao trabalhar em grupo no desenvolvimento de pequenos estudos estatísticos, “os alunos desenvolvem o espírito de iniciativa e autonomia, e enriquecem as suas interações com os colegas” (p. 59).

Relativamente ao uso das tecnologias, na concretização da tarefa 4, os professores de Matemática juntamente com os alunos de cada turma do ensino básico da escola, criaram uma base de dados (numa folha de cálculo) e utilizei geração de conjuntos aleatórios disponíveis para gerar amostras distintas para cada grupo utilizar na resolução. Na tarefa 5, usei o computador, nomeadamente o editor de texto, de modo a facilitar a tarefa de contagem. Na tarefa 6, recorri a uma folha de cálculo para a sistematização dos dados recolhidos, para diminuir o tempo de cálculo na determinação das frequências relativas e facilitar a comparação entre os dados recolhidos e os dados reais.

O trabalho desenvolvido na sala de aula durante a unidade de ensino foi semelhante ao realizado ao longo do restante ano letivo. A aula estava organizada em três momentos: (i) proposta da tarefa aos alunos, (ii) desenvolvimentos do trabalho realizado pelos alunos, e (iii) discussão com toda a turma dos resultados obtidos (Ponte, Brocardo & Oliveira, 2003). Durante a resolução das tarefas os alunos deviam registar por escrito o seu trabalho de modo a organizar o seu raciocínio e/ou argumentação e recorrer a este sempre que na discussão da tarefa o necessitem.

Quadro 8. Sequência de tarefas da unidade de ensino.

Tarefa		Duração (em minutos)/ modo de trabalho	Objetivos
<i>Inicial</i>	<i>Teste inicial / formativo</i>	90/ individual	Identificar os conhecimentos e as capacidades dos alunos antes da realização da unidade de ensino.
1	“População e amostra”	90/ em grupo	Distinguir população de amostra e ponderar a representatividade de uma amostra em relação à respetiva população.
2	“Limpar dados”	45/ a pares	Analisar criticamente um conjunto de dados de modo a detetar imprecisões antes de os começar a tratar, pondo em causa a fiabilidade da análise posterior.
3	“As vendas estão a correr bem?”	45/ em grupo	Discutir e refletir quanto aos seguintes aspetos: (i) etapas de uma investigação estatística, (ii) contextualização dos dados, (iii) importância da escolha das medidas de resumo e de representações utilizadas, (iv) validade e fiabilidade das conclusões e sua dependência da informação disponível, sendo possível que um estudo motive novas questões.
4	“Um estudo na escola”	180/ em grupo	Desenvolver um estudo envolvendo às várias etapas do ciclo investigativo: (i) formular questões e hipóteses; (ii) recolher dados (iii) analisar dados, e (iv) comunicar e interpretar resultados. Discutir e refletir quanto à variabilidade da amostragem e a validade das inferências para a população correspondente.
5	“A frequência das vogais na língua portuguesa”	90/ em grupo	Discutir e refletir sobre o papel dos métodos estatísticos como meios poderosos de tomada de decisão, e sobre a ponderação do grau de incerteza associado a previsões, para a população correspondente.
6	“Previsões”	180/ em grupo	
<i>Final</i>	<i>Teste final / sumativo</i>	90/ individual	Identificar os conhecimentos e as capacidades dos alunos após a realização da unidade de ensino.

A comunicação matemática foi a capacidade transversal predominante na unidade de ensino. Com o envolvimento dos alunos durante a realização das tarefas pretendia que, em pequeno grupo ou no grupo turma, os alunos fossem capazes “de expressar as suas ideias, mas também de interpretar e compreender as ideias que lhes fossem apresentadas e de participar de forma construtiva em discussões sobre ideias, processos e resultados matemáticos” (ME, 2007, p. 8). Em duas das seis tarefas, a comunicação

escrita assumiu um papel primordial, nomeadamente na análise e elaboração de relatórios associados à realização de pequenas investigações estatísticas. Constatei que a unidade de ensino proposta pelos professores das turmas-piloto está de acordo com as orientações curriculares do programa de Matemática ao tomar como objetivo curricular o desenvolvimento da capacidade de comunicação por parte do aluno e ao criar distintas oportunidades de comunicação no trabalho que se realiza ao longo da unidade. Esperava que os alunos evidenciassem progressão “na fluência e no rigor com que se exprimem, oralmente e por escrito, tanto na linguagem natural como na linguagem matemática, usando as notações e simbologia específica [da Estatística] e [que] desenvolvam a sua capacidade de interagir num grupo e na turma” (ME, 2007, p. 62).

A discussão com toda a turma é muito importante pois, segundo Bishop e Goffree (1986), a fase reflexiva de uma atividade é a ocasião mais apropriada para que sejam expostas conexões e significados. Um bom momento de reflexão pode permitir aos alunos relacionar ideias sobre vários temas, mostrando como as ideias matemáticas se interligam. Para Ponte (2005) o professor tem a oportunidade de gerir as intervenções de modo a clarificar os conceitos e procedimentos, sendo através da partilha que se estabelece na fase de discussão, que se promove a avaliação do valor dos argumentos e o estabelecimento de conexões no âmbito da Matemática ou fora desta. Este autor refere ainda que “os momentos de discussão constituem, assim, oportunidades fundamentais para negociação de significados matemáticos e construção de novo conhecimento”(p. 26).

No caso das tarefas 4 e 6, houve dois momentos de discussão, um após a realização da parte I – escolha da amostra – e o outro quando os grupos terminaram a resolução da tarefa. Nos momentos de discussão os grupos partilharam ideias e argumentos distintos, o que decerto enriqueceu as aprendizagens dos alunos, uma vez que se promoveu: (i) o esclarecimento de dúvidas; (ii) a clarificação de aspetos associados a uma interpretação incorreta; (iii) a identificação de erros, possíveis causas e reconstrução da aprendizagem; e (iv) a sistematização de conclusões e conceitos fundamentais.

O meu papel no decorrer da realização da unidade de ensino foi o de orientar a atividade desenvolvida pelos alunos, questionar de modo a clarificar a atividade que decorria ou auxiliar na superação de impasses que surgiram, e de moderar as discussões que ocorreram. Em suma, procurei gerir o discurso convidando à participação, justificação e reflexão por parte dos alunos (Ponte & Sousa, 2010), esperando que os alunos na generalidade das situações fossem agentes ativos da sua própria aprendizagem.

3.4. Avaliação dos alunos

A avaliação foi semelhante à realizada ao longo do restante ano letivo, procurando contemplar os diferentes modos de trabalho utilizados em sala de aula, a natureza do trabalho desenvolvido na unidade e os critérios definidos em grupo e aprovados pelo conselho pedagógico. Em termos de métodos e instrumentos de avaliação, recorri à observação, ao questionamento e discussão, bem como à análise das produções escritas dos alunos em grupos/ou pares em cada tarefa, nomeadamente o relatório escrito realizados em grupo na tarefa 4 e do teste final individual. Tive em consideração a participação, o empenho, a autonomia e cooperação dos alunos na resolução das tarefas e a sua capacidade de argumentação, parâmetros contemplados no domínio dos “princípios e valores”. Utilizei a informação recolhida na realização do teste inicial para identificar os conhecimentos e as capacidades dos alunos antes da realização da unidade de ensino com vista à melhoria das aprendizagens no tópico tratamento de dados e planeamento estatístico, com o intuito de ter um indicador geral do sucesso dos alunos e de saber que dificuldades persistiam no âmbito das aprendizagens, com a intenção de lhes dar feedback quanto aos seus pontos fortes e fracos ao nível concetual e processual.

Face à inexperiência dos alunos na elaboração de relatórios, optei por incluir a tarefa “As vendas estão a correr bem?” para enfatizar os seguintes aspetos inerentes à divulgação de um ciclo investigativo: comunicação, tomada de decisão, interpretação, representações, medidas estatísticas, resultados e conclusões. No final da unidade de ensino, a informação recolhida das várias realizações individuais do teste final possibilitou a identificação dos conhecimentos e das capacidades dos alunos após a realização da unidade de ensino e permitiu compilar dados significativos para a definição da avaliação sumativa de cada aluno.

“Os dados são a matéria-prima de que a informação estatística é feita, e a fim de obter uma boa informação estatística são necessários bons dados.”

Tannenbaum e Arnold (1997)

Capítulo 4

Metodologia de investigação

Neste capítulo justifico a opção por uma metodologia de índole qualitativa, de cunho descritivo e interpretativo, bem como a opção pelo *design* de experiência de ensino recorrendo ao estudo de casos. Apresento os procedimentos metodológicos adotados e fundamento as escolhas efetuadas.

4.1. Opções metodológicas gerais e *design*

Investigação com paradigma interpretativo e abordagem qualitativa. As opções metodológicas de uma investigação são condicionadas fundamentalmente pelas características do seu próprio objeto de estudo, isto é, pelas questões de investigação e pelo grau de profundidade com que se lhes pretende responder. Para Yin (2010) existem três condições a considerar na escolha da abordagem a seguir numa investigação: (i) o tipo de questão de investigação; (ii) o controlo que o investigador tem sobre os acontecimentos; e (iii) a atualidade, ou não, dos acontecimentos. O objetivo deste estudo foi orientado para uma melhor compreensão do processo de aprendizagem de alunos do 8.º ano no quadro de uma experiência de ensino que visa desenvolver a capacidade de planeamento estatístico e de análise de dados, em que o aluno é constructo ativo do seu conhecimento, o professor assume o papel de facilitador e crítico cooperante. O interesse central deste estudo são os significados atribuídos pelos alunos às situações vivenciadas (Erickson, 1986). Assim, atendendo ao objetivo de compreender processos de aprendizagem, nomeadamente através dos significados atribuídos pelos alunos à experiência de ensino,

tornou-se para mim claro que a melhor opção seria seguir um paradigma de caráter interpretativo, enquadrado numa abordagem qualitativa e na forma de estudo de caso.

O estudo apresenta as principais características de uma investigação qualitativa referidas por Bogdan e Biklen (1994): (i) o ambiente natural é a fonte direta de dados, constituindo a investigadora o principal instrumento de recolha de dados; (ii) os dados recolhidos são essencialmente de natureza descritiva e interpretativa; (iii) para a investigadora os processos são mais importantes que os resultados ou produtos; (iv) a análise dos dados é feita de um modo indutivo e exploratório; e (v) a investigadora pretende conhecer o significado que os participantes atribuem às suas experiências.

Investigar a própria prática. Tendo feito a escolha de ser simultaneamente investigadora e professora da turma em que se realiza a experiência de ensino, encontro-me numa situação privilegiada para conhecer alguns dos problemas da sala de aula e, também, para os procurar compreender. Tal como referem Bogdan e Biklen (1994):

Os professores, ao agirem como investigadores, não só desempenham os seus deveres mas também se observam a si próprios, dão um passo atrás e distanciam-se dos conflitos imediatos, tornam-se capazes de ganhar uma visão mais ampla do que se está a passar. (p. 286)

Para Ponte (2002) a investigação dos profissionais sobre a sua prática é fundamental dado que (i) contribui para o esclarecimento e resolução dos problemas; (ii) proporciona o desenvolvimento profissional dos mesmos; (iii) promove a melhoria das organizações em que eles se inserem, e, por vezes, (iv) contribui para o desenvolvimento da cultura profissional nesse campo de prática.

Face às atuais mudanças curriculares senti interesse em investigar e procurar respostas para os problemas da minha prática no tema Organização e tratamento de dados, nomeadamente no tópico Planeamento estatístico. Procuo compreender as dificuldades que os alunos sentem e os processos que mobilizam no âmbito do raciocínio estatístico sobre amostras, dados e representações estatísticas, no contexto de pequenas investigações estatísticas, para proporcionar experiências de aprendizagem significativas e que conduzam a uma evolução positiva da capacidade de planeamento e de análise de dados. A opção metodológica de ser investigadora e professora pode trazer algumas dificuldades ao nível metodológico, nomeadamente no registo de dados e na sua análise. No entanto, julgo terem sido ultrapassadas tendo em consideração que o estudo não se refere à experiência realizada como um todo, mas foca a atenção no processo de ensino-

aprendizagem, tendo conseguido gerar um *corpus* de material empírico que permite estudar essa situação de modo descomprometido, com o necessário distanciamento (Ponte, 2006).

Observação participante. Como investigadora, observei os alunos em sala de aula e, em alguns casos fora desta, por exemplo em situação de entrevista, recolhendo diretamente os dados. Recorro também a gravações vídeo/áudio e a documentos elaborados pelos alunos, estive presente em todos os momentos, observando as ações no contexto escolar. Assim, as ações dos participantes foram observadas no seu ambiente natural, facilitando a compreensão dos fenómenos de interesse (Bogdan & Biklen, 1994). Os dados recolhidos são, essencialmente, registos escritos dos alunos, transcrições de momentos de discussão em sala de aula ou de entrevistas, sendo a sua natureza descritiva e interpretativa. Têm a riqueza de serem relatos de situações na linguagem própria dos participantes, o que permite o acesso aos conceitos que estes usam no âmbito do fenómeno em observação.

Estudo de caso. Como o estudo se baseia no desenvolvimento de uma experiência de ensino e é relativamente escasso o controlo da investigadora sobre os acontecimentos, a modalidade escolhida é o estudo de caso. Segundo Yin (2010), um estudo de caso é uma investigação empírica que se debruça sobre um determinado fenómeno no seu próprio contexto, quando a fronteira entre ambos é pouco evidente. Além disso, é uma modalidade de investigação adequada quando se pretende compreender em profundidade uma situação, num registo descritivo e evidenciar os aspetos singulares mais relevantes que a caracterizam. Este tipo de estudo é desenvolvido com o recurso a informações provenientes de múltiplas fontes. Para André (2009), o estudo de caso é um instrumento propício ao estudo de questões de investigação em contexto educativo:

Se o interesse é investigar fenómenos educacionais no contexto natural em que ocorrem, os estudos de caso podem ser instrumentos valiosos, pois o contacto direto e prolongado do pesquisador com os eventos e situações investigadas possibilita descrever ações e comportamentos, captar significados, analisar interações, compreender e interpretar linguagens, estudar representações, sem desvinculá-los do contexto e das circunstâncias especiais em que se manifestam. Assim, permitem compreender não só como surgem e se desenvolvem esses fenómenos, mas também como evoluem num dado período de tempo. (p. 65)

A escolha deste *design* de investigação prende-se como o facto do estudo a realizar satisfazer as três características essenciais indicadas por Ponte (2006): (i) ser uma investigação de natureza empírica, recorrendo a múltiplas fontes de dados; (ii) a investigadora procurar compreender em maior profundidade a situação; e (iii) os resultados assumirem a forma de descrição cujo objetivo é narrar uma história que contribua com informação significativa, interessante e esclarecedora para o conhecimento existente. Tendo esta investigação por base uma unidade de ensino adaptada à minha turma de 8.º ano, o estudo incide, num nível macro, nessa mesma turma e, num nível micro, em dois alunos dessa turma.

Fases do estudo. Neste estudo existem quatro fases distintas. A primeira inclui a preparação do estudo, o trabalho de construção do quadro concetual e a adaptação da unidade de ensino. Inclui ainda um diagnóstico da capacidade prévia de planeamento estatístico e capacidade de análise de dados, recorrendo a um teste inicial. A segunda fase corresponde à lecionação da unidade de ensino Planeamento estatístico. A unidade tem a duração de dezoito segmentos de quarenta e cinco minutos, sendo privilegiado o trabalho a pares e em pequeno grupo. A dinâmica de aula inclui momentos de discussão coletiva e em duas das tarefas os alunos realizam um pequeno relatório sobre a investigação estatística realizada. Durante todas as aulas da unidade de ensino são feitas gravações vídeo e/ou áudio. A terceira fase é subordinada à aferição das aprendizagens e dificuldades sentidas na apreensão da unidade de ensino, utilizando um teste final e uma entrevista individual, durante e após a unidade de ensino, com os dois alunos que constituem os casos de estudo. Os dados são analisados pela investigadora e a sua interpretação constitui o instrumento primordial de análise, dando voz às perspetivas que os alunos têm sobre os acontecimentos e procurando compreender o modo com eles raciocinam e agem. Por último, a quarta fase inclui o aperfeiçoamento do quadro conceptual, a análise dos dados e a redação do texto dos restantes capítulos da presente dissertação.

4.2. Participantes

A escola. A investigação decorre numa escola do distrito de Lisboa, concelho de Sintra, na qual sou professora há doze anos. A escola situa-se numa freguesia urbana e com elevada densidade populacional, com significativo crescimento da população dado o fluxo migratório. A escola tem 56 turmas desde o 7.º ao 12.º ano, cursos profissionais

e de Educação e Formação de Adultos, com aproximadamente 1600 alunos, distribuídos por três turnos (manhã, tarde e noite).

A turma. Leciono a disciplina de Matemática à turma onde se efetua o estudo, tendo no 7.º ano sido professora desta mesma turma quer de Matemática quer de Estudo Acompanhado. Trata-se de uma turma de 8.º ano e é constituída por 29 alunos, 14 raparigas e 15 rapazes, com idades compreendidas entre os 13 e os 15 anos. A turma manteve-se na totalidade do 7.º para o 8.º ano. No entanto, uma aluna solicitou transferência de escola no final do 1.º período, pelo que no momento do estudo a turma em Matemática compõe-se de 28 alunos. A turma tem uma aluna de nacionalidade cabo-verdiana que está em Portugal desde o ano letivo anterior e apresenta dificuldades a nível do domínio da Língua Portuguesa. Apenas uma aluna da turma teve ao longo do seu percurso escolar uma retenção, no 7.º ano (ano letivo de 2008/09), dada o seu elevado absentismo. A turma está integrada no projeto nacional “Turma mais sucesso escolar” pelos que à exceção das disciplinas Matemática, Educação Visual, Educação Física e das áreas curriculares não disciplinares de Área Projeto e Estudo Acompanhado, existe um outro espaço físico em que grupos de nível são aglutinados com os de outras duas turmas participantes no projeto. Na generalidade das disciplinas intervencionadas no projeto o número de alunos em aula ronda os 12 a 21 alunos.

Alunos casos de estudo. A escolha dos alunos, de entre os que se disponibilizaram, decorreu de modo a que tivessem desempenhos distintos e atitudes distintas face à Matemática. Os critérios considerados partem do pressuposto que assim será possível conseguir uma maior diversidade de modos de ver a Estatística, dado o carácter diferente do raciocínio estatístico face ao raciocínio matemático. Dos alunos da turma escolhi Ana e Rui, pois todos se dispuseram a participar no estudo. No caso da Ana, a sua escolha deve-se ao desempenho acima da média na generalidade dos tópicos lecionados até à data; embora raramente se disponha a intervir oralmente, mostra-se capaz de o fazer em termos de registo escritos e tem por hábito explicitar com clareza e correção científica as suas ideias. A escolha de Rui, teve em consideração o seu interesse e desempenho em Organização e Tratamento de Dados, em contraponto à relativa apatia que revela nos restantes tópicos lecionados, sendo o seu desempenho irregular. Por sua vez, o seu hábito de expressar ideias sem receio de se expor oralmente, e a sua capacidade de fundamentação, apesar das dificuldades na organização e estruturação do discurso e na explicitação escrita das ideias, foram fatores determinantes para a sua escolha. Os critérios

subjacentes à escolha destes alunos foram, assim, a existência de evidência e a abrangência.

4.3. Recolha de dados

4.3.1. Pedidos de autorização

Bogdan e Biklen (1994) consideram que existem duas questões éticas fundamentais relativas à investigação em educação: (i) adesão voluntária ao projeto de investigação, conhecendo a natureza do estudo e os seus deveres e direitos decorrentes do seu envolvimento, e (ii) não correr riscos superiores aos ganhos que poderão advir da participação. Tendo isto em conta, no início do ano letivo solicitei a autorização expressa da direção da escola para a realização do estudo e no início do 2.º período informei o coordenador de departamento e do grupo disciplinar de Matemática da realização do estudo. Por último pedi autorização aos encarregados de educação (anexo 1) para a realização da investigação com os seus educandos tendo incluído uma descrição da investigação e dos seus objetivos, o modo como seriam recolhidos os dados (gravação em vídeo/áudio das aulas e entrevistas), a que se destinavam os resultados e outras informações pertinentes, como a garantia do anonimato dos alunos.

4.3.2. Processo de recolha de dados

A recolha de dados realiza-se em três momentos: antes, durante e após a realização da unidade de ensino “Planeamento estatístico”. Num primeiro momento é aplicado um teste inicial (com características de diagnóstico) aos alunos da turma. O seu objetivo é conhecer, antes da unidade de ensino, a forma como os alunos mobilizam os conhecimentos desenvolvidos no tema “Tratamento de dados” e que noções têm quanto ao planeamento estatístico e análise de dados. Num segundo momento decorre a realização da unidade de ensino durante oito blocos de aula consecutivos. Nesta fase, a recolha de dados envolve gravações áudio, e gravações vídeo, entrevista e recolha de documentos (as resoluções das tarefas e os relatórios escritos produzidos pelos alunos). Após a conclusão da unidade de ensino surge o terceiro momento de recolha de dados, com a realização de um teste final e a realização de uma segunda entrevista. Com este conjunto diversificado de dados, com acentuado carácter descritivo e suscetível de fornecer

informação diversificada, espero recolher evidências que possam contribuir para responder às questões de investigação.

4.3.3. Instrumentos de recolha de dados

Teste inicial (de avaliação diagnóstica). O teste inicial foi aplicado antes da realização da unidade de ensino tendo em vista obter informação mais detalhada acerca de cada aluno quanto à capacidade de análise de dados e às suas noções intuitivas sobre o processo de amostragem e sobre o desenvolvimento e avaliação de inferências. Estes dados serviram para criar uma ideia genérica da turma, ter a noção do que os alunos sabiam ou não fazer e das suas dificuldades quanto ao tópico precedente “Tratamento de dados”. Para a elaboração do teste recorri à adaptação de algumas situações que encontrei em manuais escolares, documentos consultados no decorrer da revisão de literatura, e, para o item do grupo 3, recorri a um item (“Resultados de um teste”) do projeto PISA 2003 (Programme for International Student Assessment). O teste inclui questões de completamento de tabela de frequências, leitura direta de representações gráficas, transformação de uma representação numa outra forma de representação, comparação de distribuições, aferição de conclusões e inferências e escolha de processos de amostragem aleatórios. O teste e a respetiva matriz encontram-se no anexo 3.

Observação auxiliada pelas gravações áudio e vídeo. Para Bogdan e Biklen (1994), a observação é uma ferramenta fundamental pois permite um contacto direto e pessoal com o facto que se está a observar. Os autores consideram que é importante tomar notas das observações realizadas logo após estas ocorrerem. Neste estudo privilegiou-se a observação participante pois o objetivo foi compreender de que modo a unidade de ensino, lecionada com base em tarefas de carácter exploratório e investigativo, desenvolveu a capacidade de planeamento estatístico e de análise de dados. As gravações áudio e vídeo foram fundamentais na observação participante e um suplemento importante a outros métodos de recolha de dados de natureza qualitativa (observação, entrevista, testes e análise de documentos), dado que ao longo das aulas acompanhei o trabalho de toda a turma, foi possível registar aspetos que decerto poderiam ter passado despercebidos. Flick (2004) refere que existem diversos sistemas de transcrição, porém não há um padrão estabelecido, e relembra que o aspeto central dos procedimentos de transcrição é a proteção de anonimato.

Entrevistas. Sendo a entrevista uma das fontes de informação mais importantes para um estudo de caso (Yin, 2010) e tendo em conta o objetivo do estudo, o recurso a este instrumento pareceu-me adequado. Segundo Bogdan e Biklen (1994) este instrumento é utilizado “para recolher dados descritivos na linguagem do próprio sujeito, permitindo ao investigador desenvolver intuitivamente uma ideia sobre a maneira como os sujeitos interpretam aspetos do mundo” (p. 134), permitindo “obter os dados desejados com a máxima eficácia e a mínima distorção” (Tuckman, 2002, p. 348). As entrevistas realizadas foram individuais tendo sido gravadas em vídeo e, posteriormente, observadas e transcritas. O recurso à gravação em vídeo teve a vantagem de incluir ações do sujeito que poderiam ser importantes por exemplo, na clarificação da interpretação da situação, da argumentação utilizada, etc.. De modo a garantir o anonimato a gravação centrou-se na folha de resposta. Dado o carácter semiestruturado das entrevistas, foi seguido um guião previamente elaborado que assumiu uma função orientadora (anexo 4). Os guiões foram constituídos por algumas questões gerais e por uma tarefa (anexo 5, 1.^a entrevista e anexo 6, 2.^a entrevista) que os alunos deviam resolver expondo as suas capacidades no âmbito do planeamento estatístico e da análise de dados. Os alunos tinham liberdade para exprimir a sua interpretação, expor os seus raciocínios e manifestar a sua opinião nas diferentes questões. No desenrolar das entrevistas foram colocadas novas questões tendo em vista esclarecer as suas respostas.

Para este trabalho foram recolhidos dados verbais (gravações áudio) e dados visuais (gravações vídeo). Flick (2004) refere diversos motivos para o uso de dados visuais em substituição de dados verbais ou a utilização conjunta de dois tipos de dados:

Primeiro, existe um desejo, por parte do investigador de ultrapassar os limites da palavra e do relato sobre ações a favor da análise das próprias ações que ocorrem naturalmente. Segundo, a vantagem proporcionada pelo fato de algumas formas de observação funcionarem sem a necessidade de o investigador realizar qualquer intervenção no campo em estudo. Por fim, a possibilidade de adquirir conhecimento através da observação, participando e intervindo no campo e, depois, observando as consequências neste. (p. 174)

Recolha documental. Recolhi e analisei documentos de natureza diversa, fichas com os dados biográficos dos alunos, bem como o projeto curricular de turma, com o objetivo de obter informação que me permitisse a caracterização mais detalhada dos alunos selecionados para o estudo de caso. A caracterização da escola teve por base a

análise do respetivo Projeto Educativo. Para a análise da evolução no âmbito das capacidades de planeamento estatístico e de análise de dados durante a unidade de ensino, tomei por base as resoluções escritas dos testes e das tarefas produzidas pelos alunos, incluindo sempre que a tarefa o solicitar o relatório escrito.

Teste final. O teste final foi aplicado depois da concretização da unidade de ensino e pretendia analisar a evolução do desempenho dos alunos. Assumo assim que a diferença entre o nível de proficiência evidenciado nas respostas dos alunos a um conjunto de itens concebidos para o efeito, no teste inicial e no teste final, reflete o efeito das aprendizagens realizadas. Para a elaboração do teste final, à semelhança do que aconteceu com o teste inicial, recorri à adaptação de algumas situações de manuais escolares e doutros documentos consultados. O teste incluiu questões de leitura direta de representações tabelares e gráficas, transformação de uma representação numa outra forma de representação, comparação de distribuições, conhecimento do ciclo investigativo (formulação de questões, processo de amostragem, análise de dados, aferição de conclusões e inferências). O teste e a respetiva matriz encontram-se no anexo 13.

Em suma, para a recolha de dados recorri aos seguintes instrumentos:

- Observação auxiliada pelas gravações áudio e vídeo;
- Teste inicial e final;
- Entrevistas semiestruturadas;
- Recolha documental, nomeadamente documentos caracterizadores dos alunos (registos biográficos, registo de assiduidade, atas do conselho de turma) e resoluções escritas das tarefas.

A figura 8 descreve a sequência dos processos de recolha de dados ao longo da experiência de ensino.



Figura 8. Processos de recolha de dados ao longo da experiência de ensino.

4.4. Análise de dados

Como indicam Bogdan e Biklen (1994), a análise de dados é o processo de busca e organização de entrevistas, notas de campo e outros materiais recolhidos com o objetivo de ampliar a compreensão desses mesmos dados com vista a poder apresentar-se aquilo que se encontrou. A análise qualitativa dos dados procura ter em consideração, aspetos relevantes das questões de investigação.

Neste estudo há momentos de recolha de dados que são simultaneamente momentos de análise. Se, por um lado, há uma análise de dados posterior à recolha que incide sobre produções escritas e comparações entre teste inicial e teste final, por outro lado, também há uma análise temporalmente coincidente com a recolha de dados, como é o caso nas entrevistas e nas participações orais no âmbito das discussões no final de cada tarefa, assim como no diálogo que se estabelece durante a sua execução que ocorre durante o processo de ensino-aprendizagem da unidade em estudo. Por estas razões, considero estar a seguir uma das orientações de Bogdan e Biklen (1994) quando consideram recomendável que se proceda a alguma análise de dados durante a recolha de dados, pois pode-se correr o risco da recolha de dados não ter orientação e/ou de os dados recolhidos não serem suficientemente completos para realizar a análise desejável.

A interpretação dos dados é um ponto essencial da investigação qualitativa, variando de acordo com as técnicas utilizadas. De acordo com Flick (2004):

A interpretação de textos pode visar dois objetivos opostos: um é revelar, desvendar ou contextualizar as afirmações feitas no texto, o que conduz normalmente à ampliação do material textual (...); o outro visa reduzir o material, parafraseando-o, resumindo-o ou categorizando-o. As duas estratégias são aplicadas, quer em alternativa, quer sucessivamente. Em suma, podem distinguir-se duas estratégias no tratamento do texto: uma é a codificação do material, com objetivos de categorizar ou de elaborar uma teoria; a outra é a análise sequencial, mais ou menos estrita do texto, visando a reconstituição da estrutura do texto e do caso. (p. 192)

Com vista a sistematizar os dados, organizo as respostas dos alunos aos testes, por questão, bem como as respostas dos grupos de alunos ou pares, a cada uma das questões das tarefas, de modo a ficar com uma visão mais global do material recolhido por meio de cada instrumento. Relativamente às entrevistas, começo por efetuar as

transcrições e complementar a minha observação com o auxílio das gravações vídeo. Depois, face aos objetivos do estudo e às questões de investigação inicio então o processo de identificação e classificação dos diferentes tipos de dados, tendo por base os aspetos teóricos revistos na literatura sobre o tema em estudo. Por fim, analiso o percurso de dois alunos (Ana e Rui) no sentido de aferir a sua evolução durante a experiência de ensino, identificando comportamentos/atitudes que evidenciem o desenvolvimento da capacidade de planeamento e de análise de dados.

Há que não esquecer que a Estatística é a ciência dos dados e os dados não são números, mas números em contexto.

Batanero (2001)

Capítulo 5

Desempenho da turma

Este capítulo descreve o desempenho da minha turma de 8.º ano, alvo da experiência de ensino, durante a realização da unidade de ensino que está na base desta investigação. Começo por fazer uma breve descrição da estrutura das aulas e dos materiais de apoio disponibilizados aos alunos. Apresento para cada tarefa da unidade uma breve descrição envolvendo os seguintes aspetos: apresentação da tarefa, modo de trabalho, produções dos vários grupos/ pares aos distintos itens da tarefa, situações inesperadas/ episódios significativos, dificuldades sentidas pelos alunos, o modo como decorreu a discussão coletiva, análise e interpretação do trabalho desenvolvido e das dificuldades sentidas à luz de alguns dos resultados de investigação em educação estatística e referências teóricas, e um balanço da experiência de aplicação da tarefa em ambiente de aula. No último ponto são apresentados os resultados de toda a turma baseados nos testes realizados antes e depois da unidade de ensino. Finalmente, efetuo um balanço final da unidade de ensino.

5.1. Estrutura das aulas e materiais de apoio aos alunos

Estrutura das aulas. Ao longo das aulas, a turma do 8.º ano que participou neste estudo e que era constituída por 28 alunos, teve uma postura participativa e ativa. Apesar de um grupo de 7 a 8 alunos apresentar bastantes dificuldades na disciplina de Matemática, e de um grupo de 6 a 7 alunos apresentar falta de concentração, a maioria trabalhou com entusiasmo e teve vontade de aprender. Estes alunos, desde o início do 7.º ano tinham o hábito de trabalhar aos pares, nos temas de organização e tratamento de

dados, dada a natureza dos contextos e a importância da troca de ideias na construção de uma cidadania mais ativa, os alunos elaboraram as tarefas propostas em grupo de 4 alunos.

Na realização das tarefas distinguiram-se três momentos: (i) *Apresentação*: a professora introduziu oralmente a situação de trabalho e fomentou o questionamento sobre o contexto, de seguida foi feita a leitura do corpo principal da tarefa distribuída e dos itens a serem resolvidos no âmbito do trabalho autónomo dos grupos, quando necessário a professora complementou/esclareceu qual o trabalho a desenvolver; (ii) *Trabalho autónomo dos alunos*: estes resolveram a tarefa em pequeno grupo (ou a pares) e, quando surgiram dificuldades, a professora esclareceu pontualmente os alunos, em alturas em que as dúvidas foram sentidas por parte de um grupo significativo de alunos a professora promoveu um momento de discussão em turma; e (iii) *Discussão final*: Um grupo de alunos começou por apresentar a sua proposta de resolução explicando a resposta e esclarecendo eventuais dúvidas. Seguidamente, os outros grupos apresentaram as suas propostas de resolução à turma e sempre que estas acrescentaram um novo aspeto e/ou uma outra explicação, no final da discussão das várias resoluções dos itens que compunham a tarefa, fez-se uma síntese das aprendizagens aportadas durante a sua realização.

Materiais de apoio aos alunos. Face à decisão do grupo de Matemática da escola, de implementar o programa do ensino básico homologado em 2007, no início do ano letivo 2009/2010, um ano antes da sua generalização, optou-se pela não adoção de manual escolar, pelo que, à semelhança do que ocorreu nos anteriores tópicos, os professores do grupo de matemática a lecionar o 8.º ano de escolaridade disponibilizaram um documento de natureza teórico-prática para apoio às aprendizagens dos alunos, no âmbito do “Planeamento estatístico”.

Para a sua elaboração recorreu-se a outros manuais escolares, nomeadamente “Matemática 10” (Lopes, Bernardes, Loureiro, Varandas, Viana & Bastos, 1997). Num conjunto de dezasseis páginas, as primeiras dez são teóricas e foram organizadas do seguinte modo:

- Noções gerais: Estatística no mundo atual e suas origens; conceitos – população, amostra, censo, sondagem, seleção de amostras: representatividade e dimensão; técnicas de amostragens aleatórias: sistemática, simples e por extrato e; fases de um estudo estatístico.

- Organização e representação: variáveis aleatórias e sua natureza; tabelas de frequências e representações gráficas; determinação de medidas estatísticas de tendência central, localização e dispersão nomeadamente a moda, a média, a mediana, os quartis, a amplitude total e a amplitude interquartil.
- Utilização das medidas estatísticas na interpretação de dados: Adequação das medidas a utilizar face a diferentes condições, e de que informação aporta sobre o conjunto de dados; características específicas de distribuições de dados simétricas e assimétricas, utilização das medidas estatísticas na comparação de duas ou mais distribuições de dados.

As restantes seis páginas do material disponibilizado aos alunos contêm, ao todo, vinte e quatro itens, sendo que os dez primeiros itens abordam questões de planeamento: processo de seleção de uma amostra e reconhecimento de amostras representativas, e identifica ainda aspetos que podem gerar enviesamentos na escolha de uma amostra, bem como a análise crítica de estudos estatísticos face ao uso de amostras não representativas. Os últimos catorze itens apelam à prática no âmbito da análise de dados, tais como a importância de reconhecer características gerais, como a forma, o centro e a dispersão da distribuição, utilizar essas mesmas características para proceder a comparações entre duas ou mais distribuições ou obter outras representações da mesma distribuição.

5.2. A realização das tarefas

Ao longo da unidade de ensino surgiram muitas situações de aprendizagem e de partilha de conhecimentos, pelo que não apresento aqui uma descrição exaustiva de todos os momentos vividos nas aulas. Apresento apenas as resoluções distintas e as situações que levaram a novos conhecimentos ou episódios especialmente marcantes no desenvolvimento das capacidades de planeamento e análise de dados, por exemplo questões que se revelam fundamentais para a compreensão da variabilidade existente nas situações de interesse para a estatística, diferentes elementos que os alunos ponderaram quando analisaram a representatividade de uma amostra, entre outras. Começo por fazer uma apresentação breve do modo de trabalho na aula, o momento da apresentação da tarefa, as produções dos alunos, os momentos que se destacaram na realização e/ou na discussão e termino com uma breve referência ao momento final de cada tarefa.

5.2.1. Tarefa 1 – População e amostra

A tarefa (Anexo 7) era constituída por três partes, nas duas primeiras partes, tal como foi projetado analisaram-se as duas situações, em grande grupo (Figura 9). Na parte A discutiu-se a importância das amostras e as razões da sua utilização e na parte B a importância da representatividade das amostras. A parte C, foi constituída por seis itens, sendo os cinco primeiros (Figura 10), relativos à análise da representatividade de uma amostra face a uma dada situação, e o sexto (Figura 19) relativo à preparação de um estudo estatístico. Na terceira parte da tarefa, os alunos da turma organizaram-se em sete grupos de trabalho com quatro elementos e responderam por escrito aos itens propostos.


<p>Parte A - Porquê uma amostra?</p> <p>Situação 1 Um empresário de uma fábrica de fósforos precisava fazer o controlo de qualidade dos fósforos produzidos pela sua fábrica. Como o poderia fazer? <u>Naturalmente que não vai acender todos os fósforos produzidos...</u></p> <p>População: fósforos produzidos Amostra: um conjunto de fósforos Unidade estatística: o fósforo</p>  <p>Situação 2 10% não sabem quem é o Presidente da República (<i>Sondagem</i> in Expresso 15/03/97) <u>Não é possível inquirir todos os habitantes ...</u></p> <p>Ficha Técnica Sondagem efectuada entre os dias 6 e 31 de Janeiro. A população é constituída pelos habitantes de Portugal Continental, com idade entre os 18 e os 74 anos. A amostra é de 1964 indivíduos, entrevistados directamente, nas suas residências.</p> <p>População: habitantes de Portugal Continental, com idade entre os 18 e os 74 anos Amostra: 1964 indivíduos com idade entre os 18 e os 74 anos Unidade estatística: um indivíduo com idade entre os 18 e os 74 anos</p>	<p>Parte B - O que é uma boa amostra?</p> <p>Situação 1</p> <ul style="list-style-type: none">Conjunto de fósforos produzidos num dia?Conjunto de fósforos produzidos na mesma máquina? <p>Situação 2</p> <ul style="list-style-type: none">Conjunto dos alunos da turma?Conjunto de indivíduos que trabalham na Presidência da República?Conjunto de indivíduos que trabalham em jornais?
--	--

Figura 9. Tarefa “População e amostra” – Parte A e B.

A aula e as produções dos alunos. A tarefa foi introduzida com a seguinte frase: “Até agora trabalhamos sempre com a população, porque esta era de pequena dimensão. Em certos casos trabalhámos com um conjunto de elementos da população, pretendendo apenas descrever o que se passava nesse conjunto. Este ano, vamos procurar trabalhar sempre com o objetivo de tirar conclusões, expandir as nossas considerações relativamente ao estudo estatístico para a população, mesmo quando não nos seja pos-

sível trabalhar com todos os elementos dela. Vamos analisar duas situações em que isso acontece... ”

Iniciou-se então a discussão da parte A, com a apresentação da situação 1 tendo questionado os alunos sobre de que modo iriam proceder para responder à questão, um aluno levantou a hipótese de analisar todos os fósforos de uma caixa, um outro sugeriu testar um fósforo ou um pequeno conjunto de fósforos em cada caixa. Os alunos aperceberam-se que seria impossível indagar cada elemento da população, por que a testagem da qualidade passava pela perda das unidades estatística em estudo.

Na situação 2 a professora leu a frase “10% não sabe quem é o presidente da república (PR)” e, de seguida, colocou a seguinte questão aos alunos, “*De que modo terão [os autores do estudo] encontrado o valor de 10%?*”. Um aluno propôs: “*foram a cada localidade perguntar a algumas pessoas se sabiam quem era o PR... a certo número de pessoas e depois fizeram a percentagem das que responderam [corretamente].*” A maioria dos alunos aceitou o recurso a uma parte da população – amostra – pois afirmaram que a população é enorme, num primeiro momento assumiram tratar-se de todos os residentes em Portugal.

Da análise correspondente à parte A, os alunos identificaram que se recorre ao estudo de uma amostra por razões de tempo, de custo, e quando a recolha de informação implica a destruição da unidade estatística. Numa aula posterior abordou-se o uso de amostras no caso de a população ser infinita.

Prof^a – Se eu tivesse que fazer um resumo da razão pela qual se usam amostras, alguém me consegue indicar uma razão?... Há três razões principais... Quem é que me indica essas três razões principais?... (...) Estas duas situações de alguma forma tentaram ilustrar essas razões?

A1 – Havia um grande número de pessoas... Não conseguíamos entrevistar todos (situação 2).

Prof^a – Não conseguíamos inquirir esse grande número de pessoas, porquê?

A2 – Demorávamos muito...

A3 – Levávamos muito tempo...

Prof^a – Eram questões de tempo, sim...

A4 – Gastávamos muito dinheiro para ir até cada (pessoa)... (situação 2)

Prof^a – Por questões financeiras... Custos ...

(...)

Prof^a – Então, estão todos de acordo que se recorre a amostras por questões de tempo e de custo, certo?

Alunos – Sim.

Prof^a – Qual será a terceira razão... entre a situação 1 e 2, usou-se uma amostra pelas mesmas razões?... O que era diferente?

A5 – Nas duas situações as populações são enormes... O tempo e o custo serve para as duas...

Prof^a – Sim é verdade... Mas porque é que na [situação] 1 não fazia sentido “riscar” os fósforos todos?

A6 – Ficávamos sem fósforos para vender.

Prof^a – Exatamente... Enquanto na situação 2 perguntar à pessoa se saber o nome do PR, não tira pedaço, no caso da situação 1, quando se vai analisar a qualidade de um fósforo...

Alunos – Arde, pega fogo, queima-se.

Prof^a – ... Destroí a unidade estatística, e essa é a terceira razão importantíssima.

A ideia de que uma amostra deve ser representativa de uma população foi trabalhada ao longo da discussão da parte B. Na situação 1, a amostra – conjunto de fósforos produzidos num dia – foi rejeitada pelo primeiro aluno que intervém “*um dia pode ter corrido bem, noutro pode ter corrido mal*”. Ao ser questionado porque tal acontece e qual a pertinência deste facto, responde “*por causa das máquinas, umas podem ter problemas... E os fósforos não prestam e noutros são todos bons...*”. Outro aluno acrescenta “*a madeira porque é constituído um fósforo pode ser diferente, num dia a madeira é boa noutro não presta*”. Uma aluna aborda as condições climáticas do dia em que se realiza o teste dos fósforos, o que não é aceite por parte dos colegas, nomeadamente um colega que alega que numa fábrica deve haver controlo das condições de modo a que o teste ocorra nas mesmas condições. Para outros alunos será melhor ter este aspeto em consideração, apesar do controlo, pois nem sempre tudo está em perfeitas condições de funcionamento, sendo que um outro aluno refere a possível avaria do equipamento. Ainda para a situação 1, a amostra – conjunto de fósforos produzidos na mesma máquina – foi de imediato refutada por alguns alunos, com argumentos como “*há mais máquinas e as máquinas podem não produzir todas da mesma forma*”, “*a máquina podia ter defeito e pensávamos que estavam todas mal*”. Os que consideravam a amostra representativa consideraram que por uma mesma máquina iriam naturalmente passar diferentes condições materiais e que tal aspeto seria suficiente “*... uma caixa de fósforos produzidos na mesma máquina pode ter dois tipos de madeira e outras diferenças pelo que chega...*”.

Para a situação 2, o conjunto dos alunos da turma foi inicialmente rejeitado como “boa” amostra, dado não ser um subconjunto da população definida pelos autores do estudo “*uma turma não deveria ter [pessoas] entre os 18 e os 74 anos*”. Posteriormente, assumindo como população do estudo os residentes em Portugal, a amostra foi novamente rejeitada pela sua pequena dimensão “*normalmente as turmas são pequeninas, assim é demasiado pequena para tirar conclusões*”. Outro argumento que passou

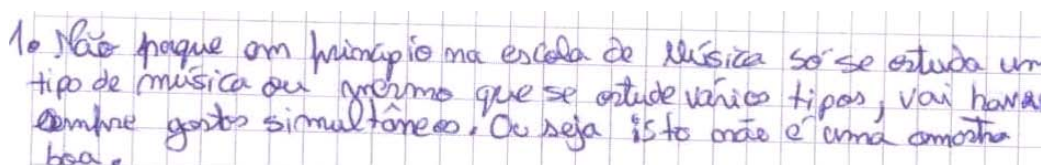
pela discussão, mas não reuniu consenso foi: “os alunos são jovens e não têm interesse pela vida política como a maioria das pessoas”. Para muitos deles, tal não era verdadeiro, dependia da idade dos alunos e do tipo de turma. O conjunto de indivíduos que trabalham na Presidência da República e o conjunto de indivíduos que trabalham em jornais foram de imediato rejeitados, como amostras, o primeiro porque “quem trabalha lá sabe o nome do presidente” e o segundo “quem trabalha nos jornais está sempre a contactar com o nome ou a pessoa [do PR]”. Deste modo, ambas as amostras seriam pouco representativas do conhecimento do nome no PR na população portuguesa, onde naturalmente também existem pessoas que desconhecem esta informação geral.

Parte C – Escolha de amostras de uma população

1. Numa escola, para se investigar as preferências musicais dos alunos, entregou-se um questionário aos alunos que frequentam também uma Escola de Música.
Faz um comentário sobre a representatividade da amostra.
2. Uma empresa de publicidade pretendia perceber quais os anúncios da televisão que mais facilmente eram recordados pelas pessoas, tendo inquirido uma amostra de pessoas à saída de um supermercado num determinado dia.
Faz um comentário sobre a representatividade da amostra.
3. A Directora de uma escola pretendia saber se os alunos estavam satisfeitos com a alimentação fornecida pela cantina da escola. Inquiriu todos os alunos com número ímpar.
Faz um comentário sobre a representatividade da amostra.
4. Para se estudar a característica "cor dos olhos" de uma determinada população, escolheu-se uma amostra constituída apenas por médicos que trabalhavam num hospital da zona.
Faz um comentário sobre a representatividade da amostra.
5. O que terias a dizer sobre a representatividade de uma amostra constituída apenas por médicos para se estudar os conhecimentos de Biologia de uma determinada população?

Figura 10. Tarefa “População e amostra” – Parte C.

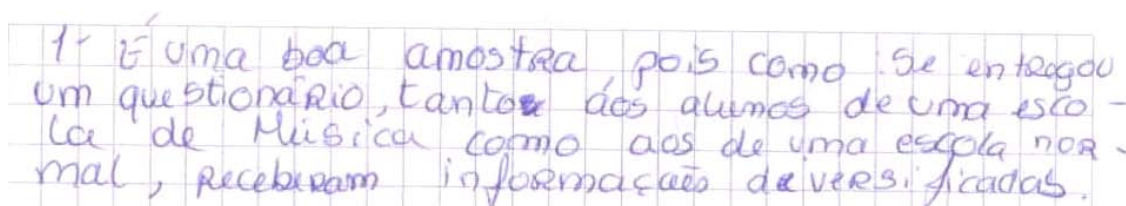
No item 1, dos sete grupos de trabalho, seis indicaram que a amostra não era “boa”. Dois grupos argumentaram que no caso de se questionar alunos de uma escola de música, as preferências musicais iriam ser condicionados pelo que tocam ou aprendem nessa escola, reconhecendo que constitui um contexto muito especial, tal como se pode observar na figura 11.



1o Não porque em principio na escola de música só se estuda um tipo de música ou porque que se estude vários tipos, vai haver sempre gostos simultâneos. Ou seja isto não é uma amostra boa.

Figura 11. Resposta apresentada por um grupo ao item 1 da Parte C.

Um grupo de alunos argumentou que, se devia ter uma amostra de alunos da escola pois tal como está definida a amostra naturalmente que não teriam respostas negativas face ao gosto pela música. Dois grupos consideraram que a amostra não era representativa pois os alunos que a compõem têm conhecimentos de música que os outros não têm. Apenas um grupo considerou a amostra representativa, tendo procedido a uma interpretação errada do item e, por sua vez, também não teve em consideração o facto da escola de música poder dedicar-se a uma tipologia específica de música tal como foi considerado por dois outros grupos. (Figura 12).



1- É uma boa amostra pois como se entregou um questionário, tanto aos alunos de uma escola de Música como aos de uma escola normal, receberam informação diversificada.

Figura 12. Resposta apresentada por um grupo ao item 1 da Parte C.

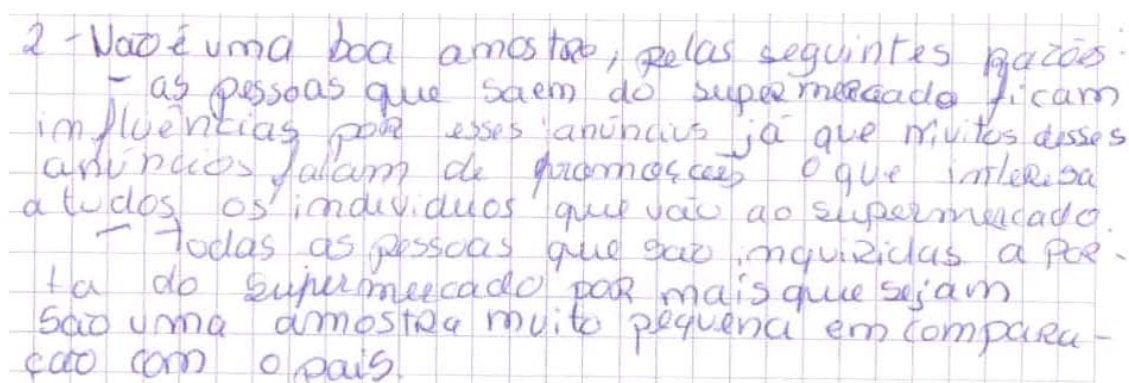
Durante a discussão final, os alunos manifestaram unanimidade no facto desta amostra ser representativa da população de alunos que frequenta a escola de Música.

Quadro 9. Respostas apresentadas pelos alunos ao item 1 da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente - amostra enviesada	24	86
Resposta incorreta	4	14

Relativamente ao item 2, apenas dois dos grupos responderam incorretamente considerando a amostra representativa. Um grupo argumentou a favor da escolha da amostra por se concentrar num determinado dia, pois consideraram que, em termos de publicidade, há uma repetição diária das mesmas campanhas, alegando “*pois os anúncios mais frequentes que passaram na televisão ontem são os mesmos que passarão amanhã*”, argumento que foi refutado na discussão final pois, segundo Ricardo, existem anúncios sazonais “*No Natal é só anúncios de brinquedos e telemóveis ...*” o que foi corroborado por outros colegas ao acrescentarem que nessa época também existem muitas campanhas publicitárias a chocolates e a perfumes. O segundo grupo que considerou a amostra representativa salientou o facto de que seria necessário garantir que as pessoas inquiridas vissem televisão, pois deste modo, considerando os gostos individuais e

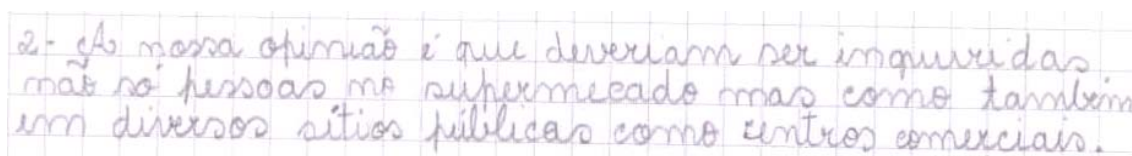
a frequência de cada campanha publicitaria teríamos a garantia da diversidade da amostra. Daí a resposta: “desde que essas pessoas vejam televisão, os anúncios são sempre os mesmos se passam muitas vezes são mais fáceis de recordar, cada pessoa tem os seus gostos nos anúncios”. Na discussão final, outro grupo opôs-se, dado que consideraram a amostra não representativa, ao assumirem que as pessoas inquiridas à saída do supermercado teriam a tendência de indicar anúncios relativos ao supermercado ou promoções que naturalmente influenciaram a sua ida às compras (Figura 13).



2 - Não é uma boa amostra, pelas seguintes razões:
- as pessoas que saem do supermercado ficam influenciadas por esses anúncios já que muitos desses anúncios falam de promoções o que interessa a todos os indivíduos que vão ao supermercado.
- Todas as pessoas que são inquiridas a saída do supermercado por mais que sejam são uma amostra muito pequena em comparação com o país.

Figura 13. Resposta apresentada por um grupo ao item 2 da Parte C.

Os argumentos de dois outros grupos também evidenciaram ter ponderado a influência que o local de inquérito teria nas respostas dos inquiridos, ao considerarem necessário inquirir também noutros locais (Figura 14).



2- A minha opinião é que deveriam ser inquiridas mais as pessoas no supermercado mas como também em diversos sítios públicas como centros comerciais.

Figura 14. Resposta apresentada por um grupo ao item 2 da Parte C.

Um grupo argumentou a não representatividade considerando que ao limitar-se a recolha de dados a um local, teríamos menos diversidade social entre os inquiridos, tendo na discussão final fundamentado do seguinte modo “eles não podem ir a um só local porque.... (...) deveriam ir por exemplo a um bairro pobre, a um bairro rico, e um bairro médio e um assim ficava... Era mais diversificado... Já era mais diversificado as ideias, do que se fores só a um bairro rico ou a bairro pobre”.

Quadro 10. Respostas apresentadas pelos alunos ao item 2 da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente - amostra enviesada	20	72
Resposta incorreta	8	28

No item 3, apenas um grupo considerou a amostra não representativa, respondendo que os gostos variam de pessoa para pessoa. Durante a discussão defenderam que se era viável inquirir todos os alunos com número ímpar então também seria viável inquirir toda a população, o que não foi apoiado pelos outros grupos, pois duplicava-se o tempo e o custo associado. Quatro grupos consideraram a amostra representativa, pois recorreu-se a um processo aleatório, garantindo a diversidade de opiniões (Figura 15).

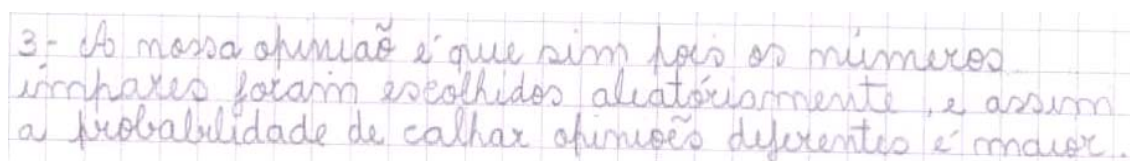


Figura 15. Resposta apresentada por um grupo ao item 3 da Parte C.

Houve dois grupos que, embora tenham considerado a amostra representativa da população, não justificaram a sua resposta. Um dos grupos considerou que a dimensão da amostra era excessiva “*Achamos que será uma amostra muito grande. Nós faríamos uma amostra menor.*”

Quadro 11. Respostas apresentadas pelos alunos ao item 3 da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente - amostra representativa	24	86
Resposta incorreta	4	14

Nos itens 4 e 5, três dos sete grupos não apresentaram qualquer resposta a estes itens, por dificuldades associadas à compreensão do item 4 e/ou dificuldades na gestão do tempo de análise para a parte C. Relativamente ao item 4, dos quatro grupos restantes, três evidenciaram dificuldades ao nível da compreensão. Eis a resposta de um des-

ses grupos que interpretou que a cada médico da amostra se iria pedir a sua opinião profissional sobre a característica “cor dos olhos” na população em estudo (Figura 16).

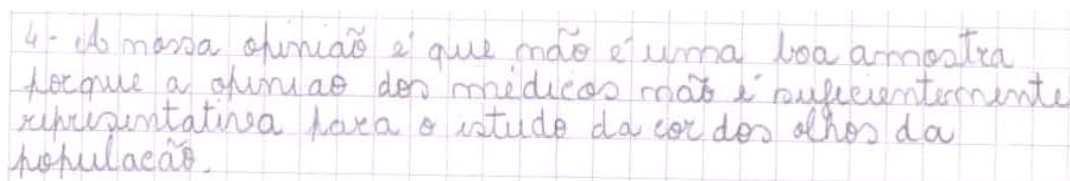


Figura 16. Resposta apresentada por um grupo ao item 4 da Parte C.

Um grupo solicitou os seguintes esclarecimentos sobre o item 4 “*Que população? Que tipo de zona? A sua dimensão? Que relação existe entre a população e o hospital?...*” a professora sugeriu que assumissem como população os residentes do seu conselho e o respetivo hospital. Face a esta contextualização, o grupo considerou que a amostra composta pelos médicos do hospital não seria representativa pois decerto que alguns deles não pertenceriam à população por não serem moradores do conselho (Figura 17).

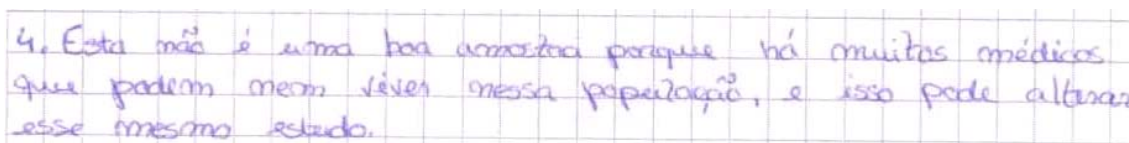


Figura 17. Resposta apresentada por um grupo ao item 4 da Parte C.

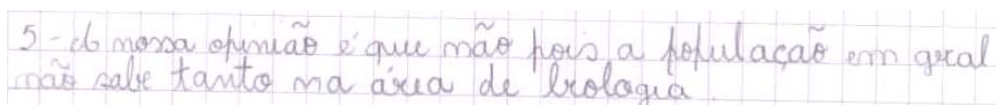
Posteriormente, na discussão final, que decorreu na aula seguinte, um aluno do mesmo grupo ainda acrescentou que a amostra não seria representativa da população para a característica “cor dos olhos” pois, estaria intimamente relacionada o grupo étnico a que cada indivíduo pertence e no conjunto dos médicos do hospital a representatividade dos vários grupos étnicos existentes no conselho não era satisfeita. De imediato a generalidade dos alunos apoiou o argumento, embora tenha instigado os alunos a questionar a veracidade da premissa “determinado grupo étnico está necessariamente mais associado a determinadas cor de olhos do que outro grupo”, argumentei que biologicamente a cor dos olhos é determinada por múltiplos caracteres genéticos, não sendo uma situação em que um caráter é dominante versus outro, pelo que a representatividade não estaria em causa. Os alunos mostraram-se céticos, pois para eles no grupo étnico/ grupo de linhagem negroide, embora existam olhos de cor verde, azul, o que observavam quase sempre era preto ou castanho-escuro, como no conselho existe uma percentagem significativa de pessoas desse grupo, enquanto no hospital entre os médicos, não conside-

ravam que tal extrato tivesse proporção semelhante, pelo que a amostra não era representativa. Face a alguma resistência dos alunos expliquei-lhes que embora não sentisse necessidade da amostra respeitar proporcionalmente os grupos étnicos da população, pois em termos biológicos não existe tal referência, na verdade haveria forma de validar se o que supunham era ou não válido, através de um estudo estatístico com o objetivo de aferir se a cor dos olhos está ou não associada ao grupo étnico do indivíduo, mas as técnicas estatísticas necessárias eram matéria dos cursos da área de matemática aplicada, no ensino superior.

Quadro 12. Respostas apresentadas pelos alunos ao item 4 da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	12	43
Responde corretamente – argumentos válidos	16	57
Resposta incorreta	0	0

No item 5, os quatro grupos que responderam foram unânimes em considerar a amostra não representativa, dado o evidente desfasamento entre o conhecimento da população em geral e o de uma comunidade médica, na área de biologia (Figura 18), o que foi aceite unanimemente pelos restantes alunos da turma.



5 - a minha opinião é que não pois a população em geral não sabe tanto na área de biologia

Figura 18. Resposta apresentada por um grupo ao item 5 da Parte C.

Quadro 13. Respostas apresentadas pelos alunos ao item 5 da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	12	43
Responde corretamente – amostra enviesada	16	57
Resposta incorreta	0	0

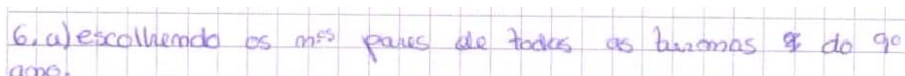
Quanto ao item 6 (Figura 19), dois grupos responderam a ambas as alíneas, e dois outros discutiram o modo como obter a amostra (alínea a)).

Parte C – Escolha de amostras de uma população

6. Uma escola tem 523 alunos do 9.º ano. Pretende-se fazer um estudo sobre os seus projetos quanto ao prosseguimento de estudos. Para isso resolveu fazer-se um inquérito que abranja uma amostra representativa.
- Como obter essa amostra?
 - Elabora três ou quatro perguntas que consideres fundamentais estar no inquérito para conhecer a opinião dos alunos quanto ao prosseguimento dos seus estudos.

Figura 19. Item 6 da Parte C.

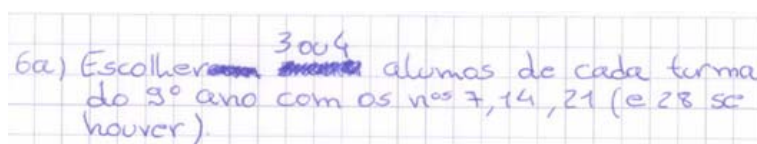
Os quatro grupos que reponderam ao item a), todos consideraram uma amostra aleatória, dois grupos escolheram um processo semelhante ao do item 3, seleção dos alunos com número par /impar (Figura 20).



6.a) escolhendo os nºs pares de todas as turmas 9º do 9º ano.

Figura 20. Resposta de um grupo à alínea a) do item 6 da Parte C.

Um grupo optou por considerar preferível selecionar a amostra por sorteio não tendo indicado a dimensão da mesma e o outro grupo defendeu a escolha de 3 a 4 alunos por cada turma do 9.º ano, sendo por exemplo os alunos cujo número de turma fossem múltiplo de sete (Figura 21).



6a) Escolheram 3 ou 4 alunos de cada turma do 9º ano com os nºs 7, 14, 21 (e 28 se houver).

Figura 21. Resposta de um grupo à alínea a) do item 6 da Parte C.

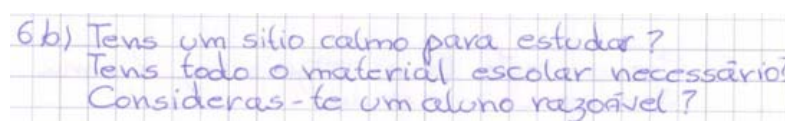
Quadro 14. Respostas apresentadas pelos alunos ao item 6 a) da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	12	43
Responde corretamente – Seleção aleatória da amostra	16	57
Resposta incorreta	0	0

Face às produções dos grupos, ao iniciar a discussão do item 6 a professora leu o corpo do item e com a colaboração dos alunos procurou esclarecer a expressão “projetos quanto ao prosseguimento de estudos” e optou por apresentar, a título de exemplo o documento da escola “Oferta formativa” (Anexo 8). Durante a discussão fica claro que

nem todos os alunos concordam com uma escolha aleatória simples, pois defendem que no caso de se sortear entre todos dos alunos do 9.º ano, existe a possibilidade da amostra ter mais alunos de uma turma do que de outra, defendem que tal possibilidade poderá influenciar o estudo dado que depois dos alunos estarem juntos durante os três anos do 3.º ciclo terão opções em termos de estudo semelhantes “*acabam por ser parecidos, umas turmas são muito boas e outras não são...*”, “*Por exemplo, na nossa turma quase todos vão para ciências...*”, dado que a população está organizada em subconjuntos turmas, consideram que é preferível uma amostra sistemática (com estratificação) como no caso da seleção de todos os alunos com número de turma par/impar em todas as turmas ou a escolha de alguns alunos em cada uma das turmas (4 a 5 alunos por turma). A dimensão da amostra a considerar, também não foi pacífica, pois tal como tinha acontecido na discussão do item 3, alguns alunos consideraram que inquirir metade da população em estudo era excessivo, por outro lado, alguns dos alunos também consideram que no caso de inquirirem apenas 3 a 4 alunos de cada turma teriam uma amostra que consideravam de pequena dimensão “*não são muitos ... 3 × 20 [estimaram que face a 523 alunos teriam cerca de 20 turmas] dá 60 a 80 inquéritos em 523 alunos*”. Perante o impasse gerado esclareci que um dos principais aspetos a ter em consideração é a diversidade que a característica em análise apresenta na população em estudo, pelo que poderíamos optar por inicialmente considerar amostras com dimensão correspondente a cerca de 10% da dimensão da população e depois ajustar face ao que considerarmos ser mais adequado dada a variabilidade presente na população e a precisão que se pretende. Face à proposta de se considerar como amostra o conjunto de 5 a 6 alunos de cada turma de 9.ºano, escolhidos de modo sistemático ou por sorteio (aleatoriedade simples), a generalidade dos alunos anuiu.

Relativamente à alínea b), os dois grupos que responderam interpretaram mal a expressão “projetos quanto ao prosseguimentos de estudos”, um grupo considerou que a questão solicitava um estudo cujo objetivo era estudar os procedimentos para o estudo (Figura 22) e o outro grupo associou o objeto do estudo à qualidade e tipologia do estudo que cada aluno realiza (Figura 23).



6.b) Tens um sítio calmo para estudar?
Tens todo o material escolar necessário?
Consideras-te um aluno razoável?

Figura 22. Resposta de um grupo à alínea b) do item 6 da Parte C.

b) Quantas horas por dia estudas?
 Quais são os teus métodos de estudo?
 Com quantos dias de antecedência?
 Quantos dias é que estudas por semana?

Figura 23. Resposta de um grupo à alínea b) do item 6 da Parte C.

Quadro 15. Respostas apresentadas pelos alunos ao item 6b) da tarefa 1.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	20	71
Formulação de um conjunto de questões adequadas ao objeto de estudo	0	0
Formulação de um conjunto de questões desadequadas ao objeto de estudo	8	29

Durante a discussão da alínea b), instiguei os alunos a responder ao inquérito proposto por cada um dos grupos, tendo-se constado de imediato a falta de coerência da questão “Com quantos dias de antecedência?”, as muitas interpretações do que seria “material escolar necessário”, “aluno razoável”, “métodos de estudo”, “sítio calmo” e o quanto as respostas à questões do tipo “Quantas horas por dia estudas?” dependia do dia da semana, do tipo de semana escolar com/sem testes agendados, com/sem trabalhos de grupo, etc.

Depois, com os contributos dos alunos, com base nos esclarecimentos quanto às ofertas escolares e profissionais após a conclusão do 9º ano, e uma pequena discussão sobre as vantagens e desvantagens de cada tipologia de questões construiu-se o inquérito (Quadro 9/16).

Como os alunos não tiveram formalmente qualquer contacto com experiências aleatórias, considerei pertinente “ilustrar” o conceito amostra representativa de uma população com alguns esquemas visuais, tendo complementado a sistematização das aprendizagens relativas ao processo de amostragem e avaliação de inferências, com a análise dos acetatos “O que é uma boa amostra?” “A tipologia da população também importa?” (Anexo 9).

Quadro 16. Inquérito construído pelo grupo turma.

Inquérito sobre o prosseguimento de estudos No final do 9.º ano pretendes ...				
<input type="checkbox"/> estudar em que tipo de curso...				<input type="checkbox"/> trabalhar em que profissão...
Científico humanístico <input type="checkbox"/> Ciências e tecnologia <input type="checkbox"/> Ciências socioeconómicas <input type="checkbox"/> Línguas e humanidades <input type="checkbox"/> Artes visuais	Tecnológico <input type="checkbox"/> Desporto <input type="checkbox"/> Multimédia <input type="checkbox"/> Outro	Profissional <input type="checkbox"/> Eletrónica, automação <input type="checkbox"/> Manutenção industrial (Auto) <input type="checkbox"/> Turismo <input type="checkbox"/> Outro	Artístico-especializado <input type="checkbox"/> Artes plásticas <input type="checkbox"/> Música <input type="checkbox"/> Dança	Estarias interessado num curso de educação e formação (CEF) ... <input type="checkbox"/> sim <input type="checkbox"/> não Qual a área?
No final do 12.º ano gostarias de continuar a estudar?				
<input type="checkbox"/> sim <input type="checkbox"/> não				

Análise e interpretação. De forma geral, os alunos conseguiram realizar com sucesso a tarefa proposta, revelaram compreender os motivos que implicam a utilização de amostras em estudos estatísticos e reconheceram vantagem em reduzir a quantidade de dados recolhidos (Garfield & Gal, 1999). Em contraponto com os resultados de Jacobs (1999), o facto de os alunos terem proposto de imediato o recurso a amostragem em detrimento do estudo da totalidade da população, nas situações 1 e 2, deve-se, à enorme dimensão das populações em causa, o que tornou evidente as dificuldades da realização de um censo.

Ao longo da realização e discussão da tarefa, os alunos identificaram vários elementos que afetavam a representatividade da amostra (parte B, itens 1, 2) e mostraram-se capazes de reconhecer amostras representativas (itens 1 e 3), competências que integram o raciocínio correto sobre amostras (Garfield, 2002; Martins & Ponte, 2010) e referidas no programa do ensino básico (ME, 2007). Os alunos revelaram ainda sensibilidade quanto à dimensão de uma amostra, considerando que 30 elementos no caso da situação 2 (Nome do PR) eram insuficientes, que no item 3, a amostra com metade da dimensão da população seria demasiado grande, na alínea a) do item 6, embora tenha havido algum desacordo entre os alunos da turma, por fim aceitaram uma amostra com

cerca de 15% da dimensão da população, não cometendo alguns dos erros indicados por Garfield (2002). A amostra escolhida para os itens 4 e 5 tinha como finalidade levar os alunos a reconhecerem que uma mesma amostra pode ser representativa de uma população face ao estudo de uma determinada característica e não o ser no estudo de uma outra característica, o que se revelou difícil de conseguir, pois os alunos revelaram dificuldades na interpretação do item 4 e posteriormente, face à contextualização sugerida pela professora, os alunos em grande grupo tomam-na como não representativa (não ser um subconjunto da população e as características principais do perfil étnico da população não serem respeitadas na amostra). No item 6 aos alunos confrontaram dois processos de amostragem – amostragem simples e amostragem estratificada – optaram por uma amostra estratificada em detrimento de uma amostra aleatória simples não tanto porque pretendiam garantir que todas as tipologias de elementos populacionais fossem contempladas na amostra (Jacobs, 1999) mas porque consideraram que a característica em estudo poderia ter uma menor variabilidade em alguns dos grupos turmas face a outros grupos turmas com maior variabilidade que poderiam não ser tidos em consideração, o que poderia inviabilizar a representatividade da amostra.

A formulação de questões de inquérito foi comprometida pela falta de tempo e pela incorreta interpretação do objeto de estudo (questão de investigação), ainda assim ficou claro que os alunos optaram por questões abertas ou dicotómicas, pouco claras e com possíveis interpretações distintas pelos diferentes inquiridos. Durante a construção do inquérito, os alunos foram gradualmente sentindo a necessidade de formular questões fechadas ou semifechadas com a apresentação das várias hipóteses de resposta em escolha múltipla, de modo a tornar mais claro o que se pretendia em cada questão.

Balanço. Em conclusão, as situações propostas na tarefa revelaram-se pertinentes e interessantes e permitiram o envolvimento entusiástico dos alunos no trabalho que conduziu à aprendizagem dos conceitos de população, amostra e amostra representativa. Apenas dois dos sete grupos terminaram integralmente a tarefa. A discussão da tarefa correspondente aos itens 4, 5 e 6 e a sistematização dos conceitos não foi possível durante o bloco de 90 minutos previsto, pelo que a dimensão da mesma poderá ter de ser reajustada. Perante a enorme diversidade de argumentos e aspetos que os vários grupos apresentaram, e dado que ainda se analisou os documentos dos anexos 8 e 9, no caso desta turma houve a necessidade de mais um bloco letivo. Considero que as situações propostas e o modo como foram trabalhadas na aula contribuiu para o desenvolvimento da noção de representatividade e de variabilidade da amostragem.

5.2.2. Tarefa 2 – Limpar os dados

A tarefa (Anexo 10) propunha a análise de uma pequena base de dados fornecidos (Figura 24), com o objetivo de que os alunos tomassem consciência da necessidade de olhar criticamente para o conjunto de dados recolhidos de modo a detetar erros/imprecisões (dada a natureza e tipologia dos dados) antes de os começar a tratar. O equilíbrio entre a “limpeza” e dimensão da amostra faz-se tendo em consideração a fiabilidade das conclusões extraídas e representatividade da amostra (a diminuição da dimensão da amostra pode trazer prejuízo à sua representatividade).

Sexo	Data de nascimento	Ano de escolaridade	Naturalidade	Altura	Tamanho do pé	Disciplina ou atividade preferida	Distância de casa à escola
M	12-04-1991	5	Portugal	143	26	Educação musical	de 1 a 2 km
F	31/02/92	4	Portugal	132	22	Estudo do Meio	menos de 2 km
F	14-01-1991	5.00	Portugal	14.2	2.3	Educação Física	2.5423 km
M	07-09-1989	6	Portugal	136	25	Matemática	de 1 a 2 km
M	13-12-1991	4	Angola	128	24	Língua Portuguesa	de 1 a 2 km
M	14-03-2001	5	Portugal	140	67	Matemática	menos de 1 km
F	06-05-1989	7	Moçambique	142	24	Língua Portuguesa	de 3 a 5 km
F	15-08-1990	6	Portugal	138	21	Língua Portuguesa	85km
M	20-02-1990	6	Portugal	192	23	Matemática	de 1 a 2 km
M	19-05-1990	6	Portugal	140	20	Educação Física	de 1 a 2 km
O	29-06-1992	7	Lua	128	21	Estudo do Meio	3000km
M	09-10-1991	4	Cabo Verde	48	21	Língua Portuguesa	menos de 1 km
F	18-12-1990	5	Angola	135	21	Matemática	menos de 1 km

Figura 24. Excerto da base de dados da tarefa 2.

A aula e as produções dos alunos. A tarefa foi introduzida do seguinte modo: “Até agora, têm trabalhado em estatística com situações em que os dados de uma ou mais características são-vos fornecidos, este ano serão vocês a recolhê-los. Na maioria das vezes... ao recolher informação cometemos erros de transcrição ou, ao preencher, podemos não compreender as instruções e acabar por responder “ao lado”, e não corresponder ao que se esperava recolher ... também se pode dar o caso de num inquérito ou questionário uma pessoa responder na brincadeira ... por estes motivos, nem sempre é possível trabalhar logo nos dados... há que limpar! ... Ou seja, encontrar os erros e, se possível, corrigir quando nos apercebemos que foram causados por má transcrição ou por não corresponderem à unidade de medida indicada ... evitamos eliminar dados ... pois se inicialmente tínhamos decidido que uma boa amostra tinha que ter 20 ele-

mentos então, se possível convém não eliminar 2 ou 3 [dados] ... ficávamos com uma amostra pouco representativa... teríamos que recolher mais dados...então é importante ponderar bem o que fazer, quando se está nesta fase...”

Os alunos aderiram entusiasticamente e, aos pares, foram procurando erros e imprecisões, propondo algumas correções. Inicialmente identificaram erros e imprecisões dentro de cada característica em estudo, apercebem-se que os registos das características altura e tamanho estavam em centímetros, “*Stora, nenhum aluno que ande no 4.º ano, calça o 26 ou o 24, certo?, é 26 centímetros...*” tendo sido interpelado pelo colega “*Não viste que não é numero [de calçado]... é tamanho...*”. Constataram, com base nas idades e data de nascimento, que os dados não eram atuais e concluíram que os dados teriam sido recolhidos em 2001, o que segundo um aluno, na discussão em grande grupo, se confirmava através a data de nascimento de um dos inquiridos que tinha escrito 14-03-2001 e “*na data de nascimento às vezes enganamo-nos no ano e pomos o ano em que se está*”.

A generalidade das correções inicialmente propostas passou pela alteração da forma de apresentação, no caso das variáveis “data de nascimento” e “ano de escolaridade”, e correções atendendo à grandeza do valor numérico, no caso das variáveis “altura” e “tamanho do pé” (Figura 25).

M	25-03-1992	41	Portugal	132	245	Matemática	menos de 1 km
F	07-08-1991	6	Portugal	150	23	Educação musical	2 saltos
F	14-01-1991	5.9m	Portugal	14.2	2.3	Educação Física	2.5423 km 2 Km
M	31-11-1983 31/11/87	11	Marte	142	22	Informática	km de 5 a 10 km
F	28-04-1988	8	Portugal	145	26.5	Educação Física	1-kg 1.6m

Figura 25. Excerto de duas resoluções de dois pares de alunos.

Mais de metade dos pares de alunos, tiveram em consideração as categorias ou classes consideradas e apresentaram correções adequadas à natureza das variáveis - data de nascimento e distância de casa à escola - “... fevereiro não tem 31 dias ...”, “*Corrigimos o 2.5km e qualquer coisa por de 2 a 3 km, ... os 2 saltos por menos de 1 km, agora o [aponta para 30000 km] não sabemos*” (Figura 26).

2	F	02/92	4	Portugal	132	22	Estudo do Meio	menos de 2 km
3	F	14-01-1991	5	Portugal	142	23	Educação Física	2.5423 km
	F	07-08-1991	6	Portugal	150	23	Educação musical	2.5423 km
	F	03-03-1992	4	Portugal	135	21	Informática	menos de 1 km

Figura 26. Excerto de duas resoluções de dois pares de alunos.

Um par de alunos considerou ter identificado uma inconsistência, ao relacionar a informação relativa ao ano de escolaridade com a disciplina/atividade preferida e argumentou que havia uma incompatibilidade entre as duas informações no caso de alguns alunos, pois as escolhas feitas não correspondiam ao plano de estudos ou às ofertas da escola: “no 4.º ano não há informática e o do 7.º [ano] não tem estudo do meio” (Figura 27), o que foi aceite por alguns dos alunos, não tendo, no entanto, reunido unanimidade. Uma aluna alega “mas [os alunos do 4.º ano] podiam ter [informática] a minha irmã mais velha teve inglês e eu não tive” e outro acrescenta “a disciplina preferida pode ainda ser estudo do meio, mesmo estando no 7.º, certo?!”.

6	29-06-1992	7	Lus	128	21	Estudo do Meio	3000km
M	03-06-1994	4	Portugal	129	22.5	Informática	menos de 1 km
F	03-03-1992	4	Portugal	135	21	Informática	menos de 1 km

Figura 27. Excerto da resolução de um par de alunos.

Quanto à eliminação de dados, nem todos os pares de alunos procederam à eliminação da totalidade dos dados recolhidos junto de um elemento da população, quando a correção não era viável, na discussão final defendem que no caso de a informação recolhida junto de um aluno estar incorreta, por exemplo no tamanho do pé ou quanto ao género, não invalida os restantes dados recolhidos nas outras variáveis.

Profª – Então no caso da resposta [do aluno] da linha 6, encontraram algum erro?

A1 – O 67 ... ninguém tem esse tamanho

A2 – Tem cá um pé ... tem mal a data [de nascimento] tiramos a linha.

A3 – Nós não... o resto está bem... dá para ficar com o resto?

Profª – que vos parece?

A2 – E a data de nascimento, os outros [de alunos do 5.º ano] são de 1991...

A1 – Mudava-se 2001 para 1991... não tínhamos dito que [2001] era o ano.

A4 – Deixa-se o resto... tínhamos menos alunos quando se estuda o tamanho [do pé].

Prof^a – Então o que vos parece ... muda-se o ano da data [de nascimento] e fica-se com os dados todos menos o tamanho do pé ?

(...)

A1– O resto parece bem ... na linha 21 só tem mal o sexo.

A4 – No da Lua, o resto parece bem mas pode ter inventado também o resto... e depois a [altura] média e os outros cálculos estão mal ... mais vale, arranjar mais [respostas] ...

Prof^a – Estando na dúvida quanto à correção de um ou mais dados é preferível eliminar por inteiro, pois pode-se dar o caso ... de no final ter valores que não fazem sentido. A decisão é sempre de quem faz o estudo ... escolher entre recolher mais dados ou usar o que os que têm desde que não contrariem o resto.

Na discussão em torno da informação relativa à 8ª linha /8º aluno do estudo, metade dos pares, considerou que o valor da variável distância de casa à escola – 85 km - estaria incorreto, apenas um par sugeriu a alteração para “8,5 km”, alteração ao nível da grandeza respeitando a natureza da variável. Durante a discussão em grande grupo, outros alunos alegaram que a correção deveria ser “de 5 a 10 km”, alteração tendo em consideração a organização da variável em classes implícita na leitura da coluna correspondente (menos de 1 km, de 1 a 2 km, de 2 a 3 km, de 3 a 5 km, de 5 a 10 km). Um par de alunos adiantou a hipótese do valor – 85 km - estar correto e tratar-se de uma situação em que o aluno que vem com algum familiar para a escola pois deste modo em vez de ficar perto de casa ficava perto do local de emprego do familiar, tal como ocorria com uma colega da turma. Depois de alguma discussão, em que parte dos alunos concorda com a hipótese e outros consideram que não tinha viabilidade pois 85 km é um valor elevado para uma viagem diária de ida e volta, tendo sido contra argumentado que se poderia tratar de uma zona rural e como tal demorar menos que poucos quilómetros em zona urbana. Não sendo possível o consenso, esclareci que para muitas das opções que se fazem na fase da limpeza dos dados é importante ter mais alguma informação sobre os mesmos, nomeadamente: “*Como foram recolhidos?, Onde foram recolhidos?, Quando foram recolhidos?, etc.*”, o que de futuro iria acontecer, pois seriam os alunos a recolher os dados em muitas das tarefas seguintes.

Ao longo da discussão, outras situações foram discutidas, impossibilidade de 11 corresponder ao ano de escolaridade de um aluno inquirido no estudo pois tratava-se apenas de alunos do ensino básico, se seria viável considerar “ir para casa” como atividade preferida e “chinês” como disciplina em vez da correta designação mandarim, diferentes anos de nascimento em alunos de um mesmo ano de escolaridade (teria havi-

do retenção, integração tardia na escola por causa da naturalidade) ou se “2 saltos” seria corrigido por “menos de 1 km”, etc.

Após a discussão solicitei aos alunos, a classificação das variáveis e as representações que seriam mais adequadas aos dados recolhidos em cada uma das variáveis do estudo, sendo que os aspetos mais pertinentes foram: (i) recordar a diferença entre variável qualitativa nominal e ordinal; (ii) possibilidade de determinar as medidas de ordem no caso de a variável ser ordinal – variável ano de escolaridade, (iii) identificar que entre as representações gráficas – gráfico de barras e gráfico circular – era preferível o recurso ao gráfico de barras no caso das variáveis – ano de escolaridade e disciplina/atividade preferida, dada o elevado número de categorias, e (iv) identificar o histograma como forma de representação das variáveis quantitativas contínuas – altura, tamanho e distância – sendo que no caso dos dados das variáveis, altura e tamanho do pé também poderiam usar o diagrama de caule e folha e o diagrama de extremos e quartis.

Por fim indaguei que questões poderiam ser respondidas com base nos dados apresentados na base de dados, dando como exemplo “*A altura média das raparigas em estudo é inferior à [altura média] dos rapazes?*”, após algum silêncio surgiram as seguintes questões por parte dos alunos: “*Que é a atividade/ disciplina preferida?*”, “*Qual a média do tamanho [do pé]?*”, “*Qual é a altura média?*”, “*Qual a nacionalidade mais frequente?*”, “*Há mais rapazes do que raparigas?*” e “*Qual a distância de casa à escola da maioria dos alunos?*”, tendo acrescentado que sendo a variável - ano de escolaridade uma variável qualitativa ordinal, também se poderia questionar “*Qual a percentagem de alunos com habilitações superiores ao 6º ano?*”.

Análise e interpretação. Os alunos revelaram-se capazes de sugerir a alteração de erros que ocorrem com frequência tais como dígitos repetidos, uma virgula mal posicionada, uma data trocada, etc., tendo em consideração a natureza dos dados recolhidos em cada variável do estudo. A discussão em grande grupo foi fundamental para alguns dos alunos compreenderem que a opção de se eliminar informação resulta da impossibilidade de correção e tem como custo a redução da amostra. Os alunos mostraram-se capazes de mobilizar conhecimento prévio (por exemplo, a natureza dos dados) e informal (questionar os dados face a algumas das suas vivências) para perspetivar os dados em contexto, tal como defende Gal (2002) o conhecimento do contexto foi determinante para que identificassem fontes de variação e erro. Sentiram também, a importância de se ter um olhar crítico sobre os dados, e a pertinência de detetar atempadamen-

te erros/imprecisões que poderão comprometer o tratamento e análise subsequente (Martins & Ponte, 2010). Os alunos evidenciaram capacidade no reconhecimento da tipologia dos dados e souberam argumentar que face a um tipo específico de dados, se adequa mais um tipo específico de gráfico ou de medida estatística (Garfield, 2000). Quanto à formulação de questões que pudessem ser respondidas pelos dados, tomando como referencia o quadro de análise de questões de Arnold (2008), os alunos formularam essencialmente questões da categoria resumo e de nível pré-questão (*Que é a atividade/ disciplina preferida?*), surgindo ainda algumas formulações de nível questão da categoria resumo (*Qual é a altura média?*).

Balanço. Em conclusão, na resolução da tarefa os alunos mostraram-se satisfeitos e interessados, puderam recorrer a alguns aspetos da sua experiência, não tendo caído no exagero de personalizar em demasia a análise, o que pode ter sido consequência do facto do contexto lhes ser familiar e, no entanto, estarem desfasados temporalmente. O tempo definido para a resolução da tarefa mostrou-se adequado. A atividade desenvolvida em torno da tarefa trouxe à aprendizagem a necessidade de analisar criticamente um conjunto de dados de modo a detetar imprecisões antes de os começar a tratar pois deste modo evita-se pôr em causa a fiabilidade da análise posterior.

5.2.3. Tarefa 3 – As vendas estão a correr bem?

A tarefa (Anexo 11) propunha a análise de uma pequena investigação estatística (Figura 28) e respetivo relatório, com o objetivo de que os alunos tomassem consciência do que se designa por investigação estatística, dos avanços e recuos que podem ocorrer ao longo de um ciclo investigativo, da tomada de decisões quanto ao tratamento e análise dos dados, bem como da pertinência da validade e fiabilidade das conclusões.

A aula e as produções dos alunos. A tarefa foi apresentada do seguinte modo: *“Hoje vão ler e analisar um estudo estatística realizado por alunos do básico e o relatório final que apresentaram. Com base na leitura pretendo que identifiquem as fases de um estudo estatístico e três aspetos que considerem fundamentais na realização do estudo.”*

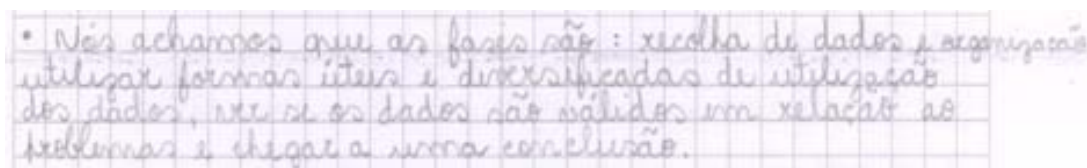
<p>Lê o texto atentamente.</p> <ul style="list-style-type: none"> • Identifica as fases de um estudo estatístico. • Identifica três aspetos que do teu ponto de vista foram importantes na realização deste estudo. <p>As vendas estão a correr bem?</p> <p>A proprietária da livraria LION, montou, num canto da sala, uma máquina de café, a título de experiência. Os lucros com a venda dos livros estavam a baixar, de modo que a venda do café talvez ajudasse a equilibrar o negócio. Passado algum tempo, decidiu adquirir e a instalação da máquina de café era vantajosa.</p>	<p style="text-align: right;"><i>Relatório</i></p> <p><i>Questão</i> Foi-nos pedido que elaborássemos um estudo, a livraria estaria a resultar. A proprietária da livraria, fôme 20 últimos dias, não nos tendo fomecido mais nenhuma: com a manutenção da máquina, os custos do grão de ca neste pequeno negócio.</p> <p><i>Metodologia utilizada</i> Cálculo de algumas medidas esta</p>
---	--

Figura 28. Excerto da tarefa.

Depois de terem feito uma primeira leitura individual, os alunos organizaram-se em grupo de 4 elementos, voltando-se os dois alunos da carteira da frente para trás, como habitualmente acontece, dispunham de 15 a 20 minutos.

Relativamente ao primeiro ponto solicitado, fases de um estudo estatístico, as produções da generalidade dos grupos apontou para a identificação do que entenderam ser as fases do estudo descrito no texto, apenas três dos sete grupos referiram tratar-se de um qualquer estudo.

Mais de metade dos grupos considerou que a primeira etapa de um estudo estatístico é a recolha de dados, seguida da organização de dados, recorrendo a expressões como “*calcular as três medidas*”, “*construção de gráficos*”, “*utilizar formas úteis e diversificadas de utilização dos dados*”, “*calcular algumas medidas estatísticas*”, “*representação dos dados*” ou “*elaboração de gráficos*”, “*organizar os dados em gráficos*”. Sugeriram ainda uma outra etapa que se poderá designar por análise de resultados, recorrendo a expressões tais como “*analisar a informação*”, “*ver se os dados são válidos em relação ao problema*”, “*depois de calcular as medidas, os alunos verificaram que os resultados da moda, da média e da mediana eram iguais*”. Por último apontam o que se poder chamar de conclusão, valendo-se de expressões como por exemplo “*chegar a uma conclusão*”, “*por fim tiram várias conclusões*”, “*fazer previsões, escrever as conclusões obtidas e todos os passos*” (Figura 29 e 30).



• Nas achamos que as fases são: recolha de dados, organização, utilizar formas úteis e diversificadas de utilização dos dados, ver se os dados são válidos em relação ao problema e chegar a uma conclusão.

Figura 29. Resposta de um grupo ao primeiro ponto.

Existe 4 fases nesto estudo:

- 1ª - Recolher ~~os~~ ^{os} dados - A proprietária da livraria forneceu aos alunos a quantidade aferida com a venda do café nos últimos 20 dias.
- 2ª - ~~Organizar~~ ^{Organizar} ~~os~~ ^{os} dados e calcular as 3 medidas (moda, média e mediana)
- 3ª - Análise ~~dos~~ ^{dos} resultados - Depois de calcular as 3 medidas, os alunos verificaram que os resultados da moda, da média e da mediana eram iguais.
- 4ª - Construção de gráficos - Depois da análise dos dados, os alunos construíram um gráfico de cake e folhos e um gráfico de linhas.
Por fim tiraram várias conclusões.

Figura 30. Resposta de um grupo ao primeiro ponto.

Dois dos sete grupos de alunos, depois de incitados pela professora, durante a discussão em pequeno grupo, com questões “Neste caso porque a senhora guardou os dados?” ou “Recolha de dados... sim, mas por que se foram recolher [os dados]?” apresentam uma fase anterior à da recolha de dados, a definição de um problema/questão (Figura 31).

- Passos essenciais de um estudo estatístico:
 - Vêter um problema (situação em análise)
 - Vêter os dados
 - Calcular as medidas estatísticas (moda, média e mediana)
 - Organizar os dados em gráficos, histogramas
 - Analizar a informação
 - Tirar conclusões
 - Fazer Previsões
 - Estremer as conclusões obtidas e todos os passos feitos

Figura 31. Resposta de um grupo ao primeiro ponto.

A discussão em grande grupo passou pela compilação de todos os aspetos apontados pelos diferentes grupos, questionando para cada aspeto se “Ainda não foi dito por outro grupo?”, “Onde pretendem que acrescente? ... Deverá vir antes de ... depois de

...?” e pela aglutinação das várias ações sugeridas pelos alunos em quatro grupos de ações e ainda se o conjunto de ações descritas era específico do estudo em análise ou se era passível de generalização a outros estudos estatísticos. Discutiu-se também se as fases do estudo eram sequenciais ou se estariam interligadas.

A1- Ter um problema ou uma situação em análise ...

Prof^a – e onde é que eu vou por essa fase?

Alunos dos grupos que tinham discutido a questão em pequeno grupo – Em primeiro!

Restante alunos anuem.

Prof^a - ...é primeira ... pois só assim se justifica a necessidade de termos dados... prever a resposta a uma questão que envolve o estudo de uma característica que varie, é o que tem interesse para a estatística.

(...)

A2 - Antes da recolha de dados ... Achamos que tínhamos que ter uma amostra...Obter uma amostra...

Prof^a – Pois à fase inicial que é a questão ou problema, chamamos identificação do objeto de estudo que envolve estudar uma característica comum [a todos os elementos da população], temos que necessariamente pensar, vou usar toda a população ou apenas uma parte [representativa] ... amostra.

(...)

A3- Ali entre o cálculo das medidas e a elaboração de gráficos não pode haver análise de resultados?

Prof^a – Que vos parece?

Os alunos concordam.

Prof^a - Pois na verdade embora se esteja a considerar uma sequência, muitas destas fases estão interligadas ... vejamos a análise e interpretação dos dados é antes da recolha de dados? ...

Alguns alunos – Não

Prof^a - Depois?

Alunos – sim!

Prof^a – e durante?

Alguns alunos – Também pode ser ...

Prof^a – É que estas fases nem sempre são lineares ... nem sempre quando uma acaba começa a outra ... por vezes quanto eu estou a pensar no tipo de dados que irei recolher , também se pensa na sua organização, antecipo o que será mais habitual obter ... ao fazer a organizar... já estou ao mesmo tempo a pensar “se eu olhar para isto desta forma, eu tiro esta conclusão”, quando se está a organizar apercebemos de padrões, regularidades, tendências ou então de irregularidades .. ou seja embora sejam momentos [fases] diferentes estão relacionados.

Depois de explorar as várias inter-relações entre as fases do ciclo estatístico passou-se à sistematização (Figura 32).

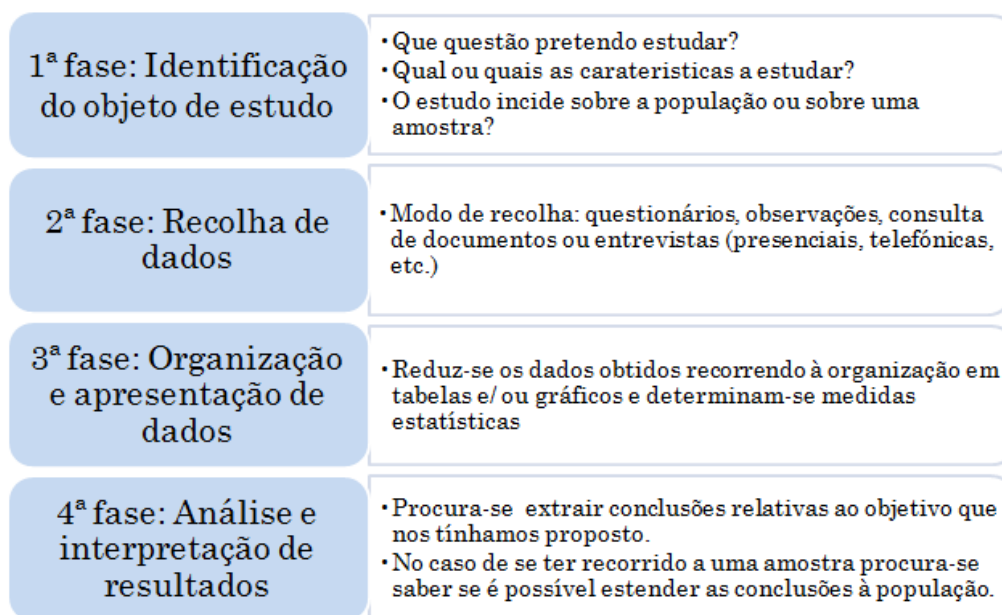


Figura 32. Excerto dos apontamentos teóricos disponibilizados aos alunos.

Quanto ao segundo ponto, “aspetos que foram determinantes na realização deste estudo”, a maioria dos grupos indicou a “recolha de dados” e a “representação gráfica”, seguida de “determinação das medidas estatísticas centrais” (Quadro 10/16), três grupos consideraram determinante “tirar conclusões” e apenas dois grupos referiram o facto de no estudo em análise, os dados serem uma série cronológica e se ter recorrido a diversas representações (Figura 33) ou alegaram que se utilizaram “representações que facilitavam a organização devido à ordem, ver semelhanças e a simetria”.

Quadro 17. Resumo das respostas ao ponto dois.

Aspeto considerado num conjunto de três solicitados	Nº de alunos	% de alunos
Recolha de dados	24	86
Determinação das medidas centrais	16	57
Representação gráfica	20	71
Tirar conclusões	12	43
Múltiplas representações	8	28,5
Ordem cronológica dos dados	8	28,5

• Os três aspectos são: diversas formas de representação de dados, ordem dos dados,

Figura 33. Resposta de um grupo ao segundo ponto.

Ao redigir a questão tinha em mente promover junto dos alunos a discussão dos seguintes aspetos:

- Natureza dos problemas/questões passíveis de serem resolvidos/ respondidas com o recurso à estatística;
- Natureza cronológica dos dados - gráfico de linhas;
- Conclusão circunscritas à informação disponível;

Decidi por isso, iniciar a discussão pelo ponto dois, recorrendo ao facto de seis grupos terem apontado como aspeto relevante a recolha de dados, indaguei que particularidade dos dados teria sido determinante para a análise da evolução do negócio e para que a conclusão tivesse sido “aufere uma quantidade média diária de 300 euros, havendo uma tendência para este valor aumentar” ao que os alunos responderam “*a ordem*” complementando com “*a ordem pela qual os dados foram recolhidos*”. Salientou-se que o facto de os dados serem uma série cronológica permitiu o recurso à representação designada por gráfico de linhas, constatei que os alunos tinham pouca familiaridade com a representação no âmbito da Matemática e não se recordavam de nenhuma situação em que tivessem recorrido à representação. Acrescentei que em vários meios de comunicação social, por exemplo em revistas semanais, quando se apresenta a evolução de share e do número de espetadores de cada canal televisivo, usa-se gráficos de linhas, em jornais para apresentar a evolução do custo da gasolina, tendo então alguns dos alunos referido o uso de gráficos de linhas em jornais/telejornais e no manual da disciplina de geografia para a apresentação da evolução de índices de natureza económica – dados sobre a evolução do emprego/desemprego, exportações ou consumo de um artigo específico. Como a generalidade dos alunos referiu as representações gráficas, como aspeto primordial no estudo em análise, procurei explorar os motivos subjacentes ao uso complementar das duas representações, tendo-se concluído que são ferramentas poderosas no âmbito da análise e da comunicação de uma visão global dos dados.

De modo a suscitar a reflexão sobre o facto de as conclusões serem circunscritas à informação disponível, e que perante novas informações, ou face a informações complementares, as mesmas conclusões poderão deixar de ser válidas, referi que os alunos de uma outra turma de 8.º ano tinham dado importância a uma frase do relatório que se encontrava nas linhas 3, 4 e 5, tendo-se gerado uma pequena discussão em torno do significado de “não nos tendo fornecido mais nenhuma informação, nomeadamente os cus-

tos com a manutenção da máquina, os custos do grão de café, ou com a mão-de-obra envolvida neste pequeno negócio.”

Prof^a - Porque é que os colegas deram importância a esta frase? O que significa?

A1- Porque que não lhes foi dada mais dados, só podiam falar das vendas ...do dinheiro que ganharam

(Procurando uma maior explicitação e apoio dos alunos à ideia da aluna)

Prof^a - Então o que é que eles estão a querer dizer?

A1 – Como só sabem o dinheiro das vendas do café só conseguiram obter a evolução

(Algum burburinho)

Prof^a – Todos concordam com esta apreciação?

(Silêncio)

A2 e outros alunos - não...

Prof^a – Vamos lá ver, eles [alunos] chegam à conclusão que o negócio – a venda de café dentro da papelaria - está bem ou mal?

(Silêncio)

A Prof^a relê a conclusão “com a informação disponível ...”

A2 - Está bem... nos últimos cinco dias está acima de 300 euros.

Prof^a – O estudo foi feito apenas com o dinheiro que entrou em caixa ... mas este dinheiro é dinheiro que a senhora pode usar ... isto é ela guarda estes cerca de 300 euros diários logo no seu bolso?

Alunos – Não...

Prof^a – Então o que acontece?

A3- Ela tem despesas...

Prof^a – Ela gasta em ...

A2- Na compra do café

A3- em água para lavar tudo e para fazer o café

A4 – luz

A1- manutenção da máquina

Prof^a – Ou seja, aparentemente o dinheiro que está a entrar é bom ... é algum dinheiro mas ... sabemos o que ela [proprietária] está a desembolsar para este negócio esteja na sua papelaria?

Alunos – Não.

Prof^a – O que é necessário para nós termos ... Este estudo foi importante por que aparentemente lhe disse que o negócio é bom, tem vindo a aumentar, mas pode ser aprofundado se consideramos outras informações ... o balanço entre o que ela gasta e o que ela recebe ainda não foi ...definido, estudado. Muitas vezes um estudo estatístico é isto mesmo, é apenas uma porta para um problema, dá-nos apenas uma visão de uma área e traz-nos ainda outras perguntas para a frente que podem ser analisadas, ponderadas e se necessário refazer as nossas conclusões ou não.

Recorrendo às duas situações concretas abordou-se a natureza de problemas/questões que são passíveis de serem resolvidos/respondidas com recurso a ferramentas da Estatística.

Prof^a – Quando fazem um estudo estatístico, investigação estatística qual é a fase fundamental, sem “isso” não há uma investigação estatística?

Generalidade dos Alunos - dados, a recolha, os dados ...

Prof^a – São os dados, o fundamental?

A1 – Stora, é a questão, a primeira a ter.

Prof^a - É a questão, o problema, aquilo que quero estudar. Esta questão tem que ser de natureza que possa ser tratada pela Estatística ou seja nem todos os problemas/ situações que ocorrem no nosso dia-a-dia são de natureza Estatística. Por exemplo se eu for estudar o rendimento escolar dos alunos do 8.º ano de escolaridade a Matemática será uma situação/ problema de natureza Estatística?

A2 – sim

Prof^a – Porquê?

A3 – porque se usa a média

A4- porque se é nota mais frequente...

Prof^a - É verdade que se poderão usar essas medidas estatísticas há um conjunto de alunos a inquirir sobre as notas ...

A2- É os dados

Prof^a – as notas, nós sabemos que são valores iguais ou há variabilidade no rendimento [escolar] de aluno para aluno, de turma para turma?

Alunos concordam com a existência de variabilidade.

Prof^a- Sempre que temos um problema em que haja alguma variabilidade da característica que pretendemos estudar e não há certeza da resposta quanto à característica em estudo antes da recolha dos dados então, estamos perante um problema estatístico.

Prof^a- Agora, saber a quantidade gasta nos livros/ manuais escolares adquiridos pelos pais na entrada da escola, dos alunos do 8.º ano da nossa escola, é um estudo estatístico? ... Saber quanto é que eles gastaram?

No meio do burburinho

A3 – não

A4 – não tem variabilidade

Prof^a- e por que não há variabilidade?

A5 – não há dados

Alguns alunos – não há dados!? Há é tudo igual...

A6 – os livros são todos iguais.

A3 - têm todos o mesmo preço.

Prof^a – Pois os livros são todos iguais pois a matriz aprovado pelo Conselho pedagógico é a mesma, os preços estão fixos pelas empresas... então o custo que o pai do menino [número] um do 8.º A e que o pai do menino número] 20 do 8.º H vai gastar se comprarem o conjunto dos livros todos, certo? ...Gastam exatamente o mesmo.

Os alunos parecem concordar, há exceção de A1.

A1- Ó stora isso está mal, em alguns sítios se encomendamos por lá [os] livros, eles fazem 10% de desconto ou se pagam com cartão...

Gera-se um novo burburinho.

Prof^a- Ah então se a gente entrar em consideração com estas situações de 10% de desconto, pais que pagam por exemplo a prestações que poderão ser assim acrescidas de juros...

A7 – Se formos comprar às editoras ou que trabalham nas editoras comprem mais barato...

Os alunos concordam que afinal existe uma variedade de situações que implicam a aquisição do conjunto de todos os livros recomendados com custos distintos.

Prof^a- Então afinal a nossa questão já tem uma natureza estatística porque já compreendemos que afinal não existe um valor único ... existe variabilidade ... assim que nos apercebemos que a situação não é estanque, fechada a um único caso, não sendo igual para todos, passa a ter interesse estatístico, entendido? ...Têm que ser situações em que haja aleatoriedade e variabilidade.

Por fim, ainda fiz referência à estrutura do relatório apresentado e o modo como os autores do relatório procuraram relatar todo o trabalho desenvolvido ao longo da investigação estatística realizada.

Análise e interpretação. Os alunos mostraram-se capazes de identificar várias ações correspondentes ao ciclo estatístico, nomeadamente as chamadas fases finais do ciclo – Dados, Análise e Conclusões, manifestando mais dificuldade na identificação da fase inicial – Identificação do objeto de estudo (correspondente às fases Problema e Plano do ciclo de Wild e Pfannkuch (1999)), o que seria expectável dado que até a este momento, as situações que vivenciaram solicitavam a descrição de um conjunto de elementos face a uma característica comum, para a qual lhes era disponibilizado um conjunto de dados. A discussão gerada à volta de situações para a qual se antecipa uma resposta determinística e situações em que se reconhece à prior a existência de variabilidade da resposta, para a qual se prevê a resposta com base em dados que variam foi pertinente, a maioria dos alunos acabou por reconhecer a diferença entre os dois tipos de questões, desenvolveram a capacidade de reconhecer a existência de variabilidade da resposta, atribuindo assim também maior peso à fase inicial do ciclo investigativo (GAISE, 2005). A generalidade dos alunos referiu as representações gráficas como aspeto primordial no estudo em análise, uma vez que o diagrama de caule e folha confirmava as medidas estatísticas centrais e o gráfico de linhas transmitia a tendência dos dados ficou patente o papel das representações no âmbito da análise e da comunicação dos dados (Friel, Curcio & Bright, 2001). A conversa em torno da conclusão extraída e a confiança na mesma, face aos dados disponíveis, deixa patente que não existe uma resposta certa ou errada, mas sim uma resposta condicionada à qualidade dos meus dados, característica fundamental da Estatística (NCTM, 1991).


Balanço. Durante a resolução da tarefa os alunos estiveram interessados e o tempo definido para a resolução foi adequado. A atividade desenvolvida em torno da tarefa aportou aprendizagens no âmbito da natureza das questões estatísticas, da importância do contexto, pertinência da escolha de medidas estatísticas e representações adequadas à natureza dos dados e que a validade e fiabilidade de uma conclusão depende

da informação disponível, sendo passível que um estudo motive novas questões de investigação.

5.2.4. Tarefa 4 – Um estudo na escola

Na tarefa “Um estudo na escola”³ (Anexo 12 - Figura 34), os alunos foram convidados a realizar em grupo um estudo estatístico, envolvendo os alunos nas diferentes etapas de uma investigação estatística. A primeira etapa consistiu na formulação de questões de pesquisa e definição de características a estudar tendo por base o objetivo de caracterizar os alunos do ensino básico da sua escola o que envolveu interpretar a situação e participar na recolha de dados. Na segunda etapa os alunos definiram um critério para a escolha de uma amostra representativa e identificaram o tipo de dados a analisar, na terceira etapa, face à amostra aleatória obtida, procederam à limpeza e ao tratamento dos dados. No final redigiram um relatório com a informação mais significativa do seu estudo.

Queremos caracterizar os alunos do ensino básico da escola no que diz respeito a algumas medidas. Para isso vamos recolher dados referentes às variáveis indicadas na tabela numa amostra representativa dos alunos da escola.



Parte I – Escolha da amostra

1. Qual é a população em estudo? E qual é a sua dimensão?
2. Qual é a dimensão da amostra com que vamos trabalhar?
3. Indica como proceder para escolher uma amostra representativa.
4. Indica um procedimento que leve à escolha de uma amostra enviesada (não representativa).

Parte II - Recolha de dados

	Aluno(M/F)	Altura (cm)	Envergadura (cm)	Nº sapato
5. Quais as variáveis a estudar? São qualitativas ou quantitativas? Contínuas				

Figura 34. Excerto da tarefa 4.

A aula e as produções dos alunos. Como pretendia proceder à recolha dos dados junto de todos os elementos da população escolar do 3.º ciclo da escola, de modo a que no final da resolução da tarefa fosse possível a comparação das estatísticas calculadas na amostra com os parâmetros da população, solicitei a ajuda dos colegas de matemática

³ Como a proposta inicial tinha como população todos os alunos da escola e dado o elevado número de alunos da escola em causa, optou-se por restringir a população em estudo aos alunos do 3.º ciclo.

e de estudo acompanhado de cada turma para que dinamizassem a recolha dos dados dos alunos das respetivas turmas.

Na turma em estudo, a recolha foi dinamizada por mim cerca de uma semana antes da realização da tarefa na aula de matemática, tendo introduzido a situação do seguinte modo, “*Dentro em breve iremos realizar um estudo estatístico na escola, com o objetivo de caracterizar ... conhecer melhor os alunos do ensino básico, o que gostariam de perguntar de modo a os conhecer melhor?*” de imediato choveram questões de vários alunos “*Qual o teu clube de futebol preferido?*”, “*O que gosta de fazer nos tempos livres?*”, “*Gostas de música? De que tipo?*”, “*Tens namorado(a)?*”, “*Tens Facebook?*”, “*Praticas algum desporto? Qual?*”, “*Costumas ver muita televisão?*”, “*Quanto tempo costumavas estar na net?*”, “*Tens animais de estimação? Que tipo de animais são?*”, “*Moras perto da escola?*”, “*Quantas mensagens, costumavas escrever por dia? E quantas recebes?*”. Depois de se discutir o entendimento da palavra “*costumas*” em relação às várias questões e se definir como sendo o comportamento mais frequente num período recente ao inquirido, “*ver muita televisão*” corresponder a ver mais de 2 horas diárias para o aluno que a formulou, pelo que a questão foi reformulada para “*Na maioria dos dias, do período escolar, vês mais de 2 horas de televisão?*” e o que se deveria entender por “*morar perto*” o que para alguns alunos seria “*demorar menos de 10 minutos [no percurso de casa-escola]*”, o que foi contestado por um aluno “*estás a pensar só nos que vêm a pé, certo ... de carro dá para ser muito*”, foi sugerido considerar “*morar perto*” como uma distância de cerca de 1 km da escola, mas a maioria considerava que os alunos a inquirir poderiam não conhecer a distância percorrida o percurso casa-escola, pelo que prevaleceu o entendimento que “*morar perto*” corresponde a “*demorar menos de 10 minutos quando vens a pé*”, mesmo que isso significa-se que o percurso casa-escola de umas fossem menor que a de outros alunos com o mesmo tempo de demora, pois iria depender do ritmo de andamento de cada um. Como a maioria das questões formuladas tinham o intuito de conhecer preferências e hábitos dos colegas, sugerir que formulassem algumas questões que abordassem características quantitativas, pelo que foram surgindo as seguintes questões “*Quantos irmãos tens?*”, “*Quantas pessoas moram contigo?*”, “*Quantas negativas tiveste no período passado?*”, “*Que nota tens a matemática?*”, “*Recebes mesada? Se sim, quanto recebes?*”, “*Quanto pesas?*”, “*Quantas TV há em casa? E computadores?*”, entre outras.

De seguida esclareci que embora as suas questões fossem pertinentes, e envolvessem muitas variáveis estatísticas interessante, iríamos concentrar um pequeno grupo

de características físicas de modo a não aumentar muito o trabalho envolvido no estudo de caracterização dos alunos do 3.º ciclo da escola – género, altura, envergadura e número de calçado - de imediato foi necessário esclarecer o que se entendia por envergadura, face ao esclarecimento rapidamente se aperceberam que embora nem todos os alunos não saibam exatamente a altura, têm a noção do seu valor, pois no início do ano em educação física todos se mediram em altura e peso, no caso da envergadura seria necessário andar a medir. Informei que de modo a evitar que ocupassem tempo extra aulas na recolha dessa informação junto de alguns colegas, iriam participar na recolha da informação respondente à sua turma, pelo que se iriam organizar em pares e que com o apoio de fitas métricas que iria colocar (verticalmente e horizontalmente) na parede, iriam conhecer a sua altura e envergadura e a dos colegas da turma. Aderiram com entusiasmo e sobe a vigilância a tenta dos restantes colegas “*abre mais os braços*”, “*está a inclinar o livro, assim a altura não fica bem*” os alunos foram-se medindo. Um grupo de alunos depois de alguns registos comenta que se tinham apercebido que “*não há muita diferença, a altura é quase sempre maior*”, remeti o comentário aos restantes alunos enquanto de terminava a o preenchimento da tabela de recolha dos dados (Figura 35), a generalidade dos alunos constatou que no seu caso tal ocorria, um conjunto de 6 alunos reconhece que a sua envergadura toma um valor ligeiramente mais alto que o da altura, teria havido erro de medição, confirma-se as medidas para dois dos alunos e a aula acaba. Um grupo de alunos mais curiosos ainda me questionam “*na próxima aula [de matemática] continuamos?*” ao que responde que seria em breve.

Género (M/F)	Altura (cm)	Envergadura (cm)	Nº sapato
F	162 cm	157 cm	37,5
F	166 cm	161 cm	38 2/3
F	166 cm	161 cm	38
F	157 cm	149 cm	39
F	162 cm	157 cm	39
M	1,57 cm	1,59	38
F	168 cm	169 cm	40
M	176 cm	179 cm	47,5
M	168 cm	164 cm	39
F	161 cm	162 cm	39,5
F	164 cm	153 cm	36
F	1,59 cm	1,49 cm	38

Figura 35. Excerto da tabela de recolha de dados da turma - tarefa 4.

Após uma semana, na aula de Matemática, reiterei que iríamos fazer um estudo estatístico na escola com o objetivo de caracterizar os alunos do 3.º ciclo quanto a algumas características físicas, e que em grupo iriam pensar e responder aos itens 1 a 5 da tarefa 4, depois da sua discussão iríamos por em prática o estudo.

No item 1, todos os grupos identificaram corretamente a população em estudo, no entanto quanto à questão da dimensão, um grupo não respondeu, e outro grupo respondeu incorretamente que a dimensão era o conjunto de todas as turmas do ensino básico (Figura 36), um outro grupo apresentou como dimensão 1000 alunos o que corresponde a um valor excessivo em relação ao valor real – cerca de 660 alunos (Figura 37). Os restantes quatro grupos consideraram que a dimensão da população estaria compreendida entre 600 e 720 face ao número médio de alunos por turma e o número de turmas existente no ensino básico (Figura 38).

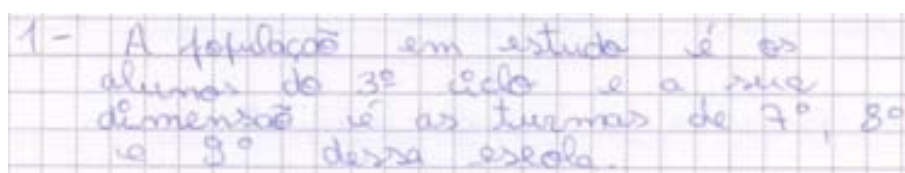


Figura 36. Resposta de um grupo ao item 1 da tarefa 4.

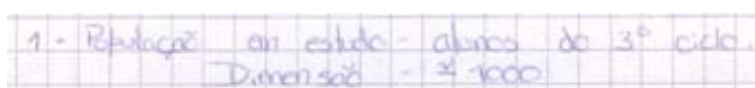


Figura 37. Resposta de um grupo ao item 1 da tarefa 4.

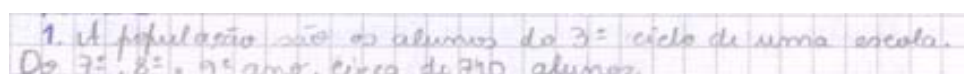


Figura 38. Resposta de um grupo ao item 1 da tarefa 4.

Quadro 18. Respostas apresentadas pelos alunos ao item 1 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Identifica a população em estudo corretamente não refere a sua dimensão	4	14,3
Identifica a população em estudo corretamente e responde à dimensão de modo incorreto	8	28,6
Identifica a população em estudo corretamente e apresenta uma estimativa válida da sua dimensão	16	57,1

Nos itens 2 e 3, os alunos debatem-se com a questão do equilíbrio entre quantidade e qualidade de uma amostra representativa, sendo que um grupo responde “o tamanho da amostra não deve ser muito grande, mas não poder ser pequena, para haver uma boa representatividade” no item 2 e no item 3 indicam que “para haver uma amostra representativa podíamos escolher 4 alunos por turma sorteando-os na calculadora”, um outro grupo responde que iria “trabalhar com uma amostra com alunos de todos os anos” no item 2 e posteriormente no item 3 acrescenta “escolher alguns números por sorteio ou escolher todos os ímpares”, quatro grupos optaram por evitar indicar em específico a dimensão, sendo que dois dos grupos indicam que iriam escolher de cada turma seis alunos cujo número de turma fosse por 8, 22, 24, 28, 15 e 3 ou um conjunto de 4 alunos cujo número de turma fosse por 1, 8, 17 e 25 (Figura 39), um grupo indica que escolheria 5 alunos de cada turma e no item 3 esclarece que “vamos trabalhar com os números 2, 12, 18, 20 e 28 de cada turma” e um outro grupo pensa que a amostra deverá ter uma dimensão correspondente a 30% da população tendo adicionado no item 3 que “temos de escolher alunos aleatoriamente e tem de representar uma boa parte da população”.

Figura 39. Respostas de dois grupos ao item 2 da tarefa 4.

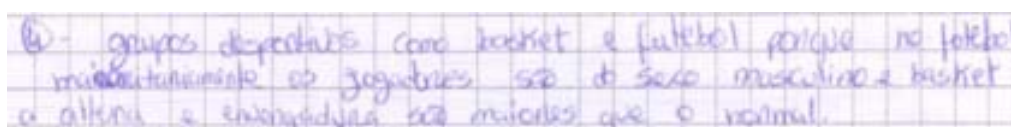
Apenas um grupo se comprometeu com um valor específico indicando que a amostra que iriam trabalhar teria 120 alunos, como tinham estimado que a população tinha de dimensão 600 alunos a amostra considerada teria um tamanho correspondente a 20% da dimensão da população, sendo que consideram que a sua seleção deverá ser aleatória (Figura 40).

Figura 40. Resposta de um grupo aos itens 2 e 3 da tarefa 4.

Quadro 19. Respostas apresentadas pelos alunos aos itens 2 e 3 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Seleção aleatória de uma amostra, sem especificar a dimensão sendo implícito que representa entre 10 e 20% da população	16	57,1
Seleção aleatória de uma amostra cuja dimensão implícita ou explícita represente entre 30 a 50%	8	28,6
Seleção aleatória de uma amostra, com explicitação da dimensão correspondente a 20% da população	4	14,3

No item 4, que solicitava a indicação de um procedimento que implicasse a escolha de uma amostra enviesada, as respostas dos alunos revelaram que ponderaram distintos elementos que poderiam afetar a representatividade da amostra em relação à respetiva população. Dois grupos consideraram que ao limitar-se a escolha da amostra a alunos de um único ano de escolaridade, entre os três anos do 3.º ciclo, teriam “*alunos quase da mesma idade então não podem representar as medidas de alunos dos 11 e os 16/17 anos*” argumento análogo foi utilizado pelo grupo que respondeu “*Escolher os alunos de uma única turma do 3.º ciclo*” e pelo grupo que optou por “*Questionar 160 alunos que sejam amigos/conhecidos das pessoas que estejam a estudar*”, pois no seu caso mais de 90 alunos dos 160 inquiridos seriam do 8.º ano pois este é o número de colegas que pertence ao projeto “Turma Mais”, ao qual a turma pertence. Dois grupos optaram por considerar “*Escolher de alunos de mesmo sexo*”, pois argumentam que existe uma relação estreita entre a estatura e o género, e deste modo teriam informações validas apenas para essa parte da população. Um grupo responde que um procedimento que leva à escolha de uma amostra enviesada, passa pela seleção de alunos que pertença a grupos específicos do desporto escolar, pois determinada prática desportiva está maioritariamente associada a um género em particular ou por que implica uma estatura específica, o que seria pouco representativo da variabilidade da população para as características em estudo (Figura 41).



b) - grupos desportivos como basquet e futebol porque no futebol maioritariamente os jogadores são do sexo masculino e basquet a altura e estatura são maiores que o normal.

Figura 41. Resposta de um grupo ao item 4 da tarefa 4.

Quadro 20. Respostas apresentadas pelos alunos ao item 4 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Estatura <i>versus</i> idade	16	57,1
Estatura <i>versus</i> género	12	42,9
Estatura <i>versus</i> grupos de uma prática desportiva específica	4	14,3

Quanto ao item 5, identificação das variáveis em estudo e sua classificação, a generalidade dos grupos identificou as quatro variáveis em estudo e procedeu à sua correta classificação (Figura 42), apenas um grupo errou a classificação da variável género e não especificou no caso das variáveis quantitativas se se tratava de uma variável discreta ou contínua. Dois grupos embora tenham procedido à correta classificação, omitiram a identidade de uma ou duas das quatro variáveis em estudo.

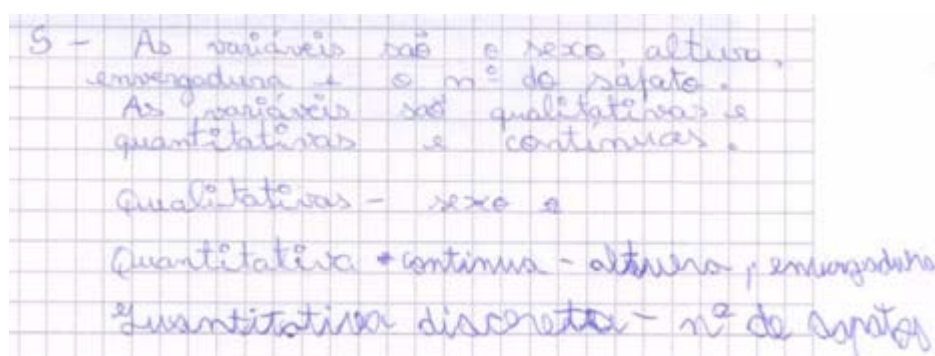


Figura 42. Resposta de um grupo ao item 5 da tarefa 4.

Quadro 21. Respostas apresentadas pelos alunos ao item 5 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Identifica duas ou três das variáveis e classifica corretamente	8	28,6
Identifica as variáveis em estudo mas tem erros na classificação	4	14,3
Identifica e classifica corretamente as várias variáveis em estudo	16	57,1

Terminada esta fase procedeu-se à discussão em grande grupo, das resoluções dos grupos aos cinco primeiros itens, destacando-se que todos os alunos apresentaram como critério de escolha da amostra representativa a seleção aleatória, mas sentiram dificuldade na definição de uma dimensão razoável, atendendo à variabilidade existente

na população, consideraram amostras de dimensão excessiva. Durante a discussão dos itens 2 e 3 os alunos concordaram que no presente estudo seria indiferente um sorteio entre a totalidade dos alunos ou sortear um pequeno número de alunos em cada turma, sendo que lhes seria mais fácil e funcional a recolha no caso da segunda opção, os alunos tiveram alguma dificuldade em abdicar da escolha de 5/6 alunos por turma e apenas a perspectiva de proceder ao tratamento de 120 a 140 dados para cada variável os levou a fazer uma pequena conceção - considerar a escolha aleatória de 3/4 alunos por turma, tendo uma amostra de dimensão a rondar os 70 a 100 elementos. A escolha aleatória de 3 alunos por turma ainda não reunia consenso pois “[o aluno sorteado] *pode estar a faltar ou haver erros nas medidas e ficarmos com muito menos* [elementos na amostra]”.

Na discussão do item 4, os alunos constataram que existiam diferentes elementos a ponderar (como por exemplo a idade, o género) de modo a obter uma amostra representativa pelo que, indicar um procedimento que não tenha em consideração pelo menos um desses elementos, implicava a obtenção de uma amostra enviesada. A identificação das variáveis e sua classificação foi consensual, e uma vez mais se recordou a natureza dos dados recolhidos para cada tipo de variável.

De seguida recordou-se o trabalho de recolha realizado numa das aulas de estudo acompanhado e informei-os de que tinha sido replicado em todas as turmas do 3.º ciclo, pelo que estávamos na posse de toda a informação sobre cada uma das variáveis em estudo, para a população do ensino básico (Figura 43) e com base numa amostra iriam procurar caracterizar os alunos do 3.º ciclo da escola, por sua vez dado que estávamos na posse da informação de toda a população, à exceção dos alunos que faltaram no dia em que se realizou a recolha, iríamos no final do seu estudo, constatar a validade das suas conclusões no âmbito da população. Com base no que tinha exemplificado quanto à geração de números aleatórios na calculadora científica, mostrei a função de geração aleatória de um conjunto de n elementos na folha de cálculo e com base na dimensão escolhida pelos alunos geraram-se quatro amostras aleatórias de 83 elementos com o recurso à função $\text{INT}(\text{ALEATÓRIO}(\text{“nº da população”}) * 641) + 1$ (Figura 43). Os alunos mostraram-se fascinados com o processo e antes de ser possível facultar aos grupos os dados das respetivas amostras, a aula terminou.

Na aula seguinte, face a cada amostra distribuída ao grupo, os alunos identificaram imprecisões e procederam a ajustes de modo a garantir a qualidade dos dados e a não comprometer a análise posterior, tal como tinha sido sugerido na aula anterior por uma das alunas. Outras opções foram feitas, por exemplo, perante valores de altura

como 146,5 cm, houve grupos que mantiveram o valor tal como apresentado, em quanto que outros grupos optaram por considerar o arredondamento ao inteiro superior 147 cm por questões de facilidade na representação informação da altura no diagrama de caule e folha, ou face à omissão de um valor optar por ter n-1 dados numa das variáveis e n dados nas outras variáveis “*trabalhamos com 81 elementos de estudo expeto com o calçado que trabalhamos com 80, porque uma pessoa não respondeu a essa questão.*”.

População						
nº população	Turma	nº de turma	Género	Altura	Envergadura	nº calçado
1	7A	1	F	153,0	147,0	36
2	7A	2	M	150,0	157,0	37
3	7A	3	M	150,0	150,0	38
4	7A	4	F	155,0	152,0	38
5	7A	5	M	150,0	159,0	38
6	7A	6	M	150,0	151,0	41
7	7A	7	M	150,0	149,0	38
8	7A	8	F	150,0	150,0	34
9	7A	9	F	147,0	151,0	37
10	7A	10	M	153,0	159,0	40

Amostra 4						
Nº amostra	Nº população	Turma	Género	Altura	Envergadura	Nº de calçado
1	251	8A	F	163	157	38
2	31	7B	F	153,5	150	37
3	305	8C	F	170	171	41
4	510	9B	M	174	172	42
5	288	8B	F	146	140	36
6	403	8G	F	154	154	38
7	603	9F	M	181	194	45
8	381	8E	F	155	152	34,5
9	260	8A	M	162	163	40
10	299	8B	F	162	168	40

Figura 43. Excerto da base de dados da população em estudo e de uma das amostras utilizadas pelos grupos na tarefa 4.

Dado a enorme dimensão da amostra e a imensidão de trabalho envolvido no tratamento dos dados, a maioria dos grupos não responde ao item 6, apenas dois grupos apresentaram para cada variável, os dados em tabelas de frequências, sendo que um dos grupos, para a variável género apresentou as frequências absolutas e relativas em percentagem, organizou as variáveis altura e envergadura em classes com amplitude de 5 cm e apresentou a tabela de frequências para a variável nº de calçado, tendo errado alguns dos arredondamentos das frequências relativas, e não sendo consistente na escolha do número de casas decimais a que recorrer, pelo que o total não corresponde a 100% em duas das variáveis (Figura 44); o outro grupo apresentou apenas a frequência absoluta no caso da variável género e número de calçado, optou por classes com amplitude de 10 cm para as variáveis contínuas (Figura 45).

6.

Variável	Frequência absoluta	Frequência relativa
Gênero		
Masculino	38	$38:83 \times 100\% = 45,8$
Feminino	45	$45:83 \times 100\% = 54,2$
Total	83	100%

Variável	Frequência absoluta	Frequência relativa
Envergadura (cm)		
< 135	1	$1:83 \times 100\% = 1,2$
135 - 140	1	$1:83 \times 100\% = 1,2$
140 - 145	1	$1:83 \times 100\% = 1,2$
145 - 150	4	$4:83 \times 100\% = 4,82$
150 - 155	12	$12:83 \times 100\% = 15$
155 - 160	18	$18:83 \times 100\% = 21,69$
160 - 165	19	$19:83 \times 100\% = 23$
165 - 170	18	$18:83 \times 100\% = 21,69$
170 - 175	4	$4:83 \times 100\% = 4,82$
175 - 180	3	$3:83 \times 100\% = 3,61$
180 - 185	1	$1:83 \times 100\% = 1,2$
185 - 190	1	$1:83 \times 100\% = 1,2$
Total	83	100,63

Variável	Frequência absoluta	Frequência relativa
Altura (cm)		
145 - 150	4	$4:83 \times 100\% = 4,82$
150 - 155	9	$9:83 \times 100\% = 10,84$
155 - 160	16	$16:83 \times 100\% = 19,28$
160 - 165	24	$24:83 \times 100\% = 28,91$
165 - 170	18	$18:83 \times 100\% = 21,69$
170 - 175	8	$8:83 \times 100\% = 9,64$
175 - 180	2	$2:83 \times 100\% = 2,41$
180 - 185	2	$2:83 \times 100\% = 2,41$
Total	83	100%

Variável	Frequência absoluta	Frequência relativa
Nº de sapato		
34	1	$1:83 \times 100\% = 1,2$
35	3	$3:83 \times 100\% = 3,61$
36	12	$12:83 \times 100\% = 15$
37	12	$12:83 \times 100\% = 15$
38	19	$19:83 \times 100\% = 23$
39	13	$13:83 \times 100\% = 15,6$
40	9	$9:83 \times 100\% = 10,8$
41	5	$5:83 \times 100\% = 6,02$
42	6	$6:83 \times 100\% = 6,02$
43	3	$3:83 \times 100\% = 3,61$
Total	83	99,86

Figura 44. Resposta de um grupo ao item 6 da tarefa 4.

Os restantes grupos, com exceção de um, optaram por passar diretamente à determinação das medidas de tendência central, tendo recorrido à representação de caule-e-folhas, no caso das variáveis quantitativas, representação que a generalidade dos alunos prefere, pois consideram que facilita a determinação da mediana e a identificação da moda (Figura 46).

Quanto ao item 7, quanto à variável género, todos os grupos identificaram a moda, mas apenas três grupos apresentaram a frequência da modalidade mais frequente, dois grupos fizeram referência à impossibilidade de determinar a média e a mediana. Na discussão ficou patente que os restantes grupos de alunos não sentiram necessidade de justificar/referir a impossibilidade da determinação da amplitude da distribuição e as medidas centrais – média e mediana, uma vez que se tratava de uma variável quantitativa não ordinal. Para as variáveis altura e envergadura, todos os grupos identificaram a moda e determinaram a média corretamente, apenas um grupo cometeu erros de contagem dos elementos à esquerda e à direita dos valores centrais da distribuição e determina a mediana incorretamente, para a variável altura. Nem todos os grupos procederam à determinação da amplitude total das distribuições da altura e da envergadura, sendo que

esses valores foram apresentados apenas por quatro dos grupos. Relativamente à variável número de calçado, todos os grupos identificaram corretamente as medidas de tendência central, no entanto apenas três dos grupos determinaram a amplitude da distribuição.

Grupo	Freq. Absoluta
F	42
H	39
Total	81

Intervalo	Freq. Absoluta	Freq. Relativa
130 - 140	1	$1/81 \approx 0,01 = 1\%$
140 - 150	4	$4/81 \approx 0,05 = 5\%$
150 - 160	27	$27/81 = 0,33 = 33\%$
160 - 170	33	$33/81 \approx 0,41 = 41\%$
170 - 180	13	$13/81 \approx 0,16 = 16\%$
180 - 190	2	$2/81 \approx 0,02 = 2\%$
190 - 200	1	$1/81 \approx 0,01 = 1\%$
Total	81	100%

Variáveis	Freq. Absoluta	Freq. Relativa
A 140 - 150	3	$3/81 \approx 0,04 = 4\%$
B 150 - 160	25	$25/81 \approx 0,31 = 31\%$
C 160 - 170	33	$33/81 \approx 0,41 = 41\%$
D 170 - 180	18	$18/81 = 0,22 = 22\%$
E 180 - 190	1	$1/81 \approx 0,01 = 1\%$
F > 190	1	$1/81 \approx 0,01 = 1\%$
Total	81	

Figura 45. Resposta de um grupo ao item 6 da tarefa 4.

Tamanho de calçado

3	66666666	77777777	88888888
3	88888888	99999999	
4	00000000	11111111	22222222
4	222222		

3	66666666	77777777	88888888
3	88888888	99999999	
4	00000000	11111111	22222222
4	222222		

Moda = 38
 Média = 39
 Mediana = 39

Figura 46. Organizaram dos dados para a resolução do item 7 da tarefa 4.

Quadro 22. Respostas apresentadas pelos alunos ao item 6 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não recorre a qualquer forma de redução dos dados	4	14,3
Recorre a tabelas de frequências para uma redução dos dados qualitativos e quantitativos	8	28,6
Recorre à representação diagrama de caule e folhas para uma redução dos dados quantitativos	16	57,1

Quadro 23. Respostas apresentadas pelos alunos ao item 7 (género) da tarefa 4.

Resposta – variável qualitativa - género	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Identifica a moda	4	14,3
Identifica a moda e justifica o seu valor	12	42,85
Identifica a moda e argumentam a impossibilidade da determinação das restantes medidas estatísticas	12	42,85

Quadro 24. Respostas apresentadas pelos alunos ao item 7 (altura) da tarefa 4.

Resposta – variável quantitativa contínua – altura	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Determina corretamente as medidas de tendência central (moda e média) e a amplitude da distribuição	4	14,3
Determina corretamente as três medidas de tendência central	8	28,6
Determina corretamente as três medidas de tendência central e a amplitude da distribuição	16	57,1

Quadro 25. Respostas apresentadas pelos alunos ao item 7 (envergadura) da tarefa 4.

Resposta – variável quantitativa contínua – envergadura	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Determina apenas as medidas de tendência central	12	42,9
Determina as medidas de tendência central e a amplitude da distribuição	16	57,1

Quadro 26. Respostas apresentadas pelos alunos ao item 7 (nº calçado) da tarefa 4.

Resposta – variável quantitativa discreta - nº de calçado	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Determina apenas as medidas de tendência central	16	57,1
Determina as medidas de tendência central e a amplitude da distribuição	12	42,9

No item 8, construção do diagrama de caule e folha para a variável envergadura, dado que o caule 15 ou 16 têm, em todas as amostras, uma elevada frequência e dado que a generalidade dos grupos optaram por usar como escala de representação a colocação de cada algarismo das unidades (folha do dado) em cada quadrícula da folha de resposta, dois grupos cometem o erro de replicar o caule com as folhas que não couberam na largura da folha de resposta (Figura 47), dois grupos cometem o erro de não ser consistentes na escolha da escala de representação (Figura 48). Os restantes três grupos não cometeram nenhum erro e um dos grupos que optou por construir um diagrama de caule e folha comparativo da envergadura em relação ao género masculino e feminino, não apresentando erros de escala (Figura 49), posteriormente esclareci os alunos que dado que se recorre às frequências absolutas só poderiam estabelecer comparações quanto à forma da distribuição, nomeadamente sobre o centro, a concentração dos dados em torno deste e dispersão global.



Figura 47. Resposta de um grupo ao item 8 da tarefa 4.

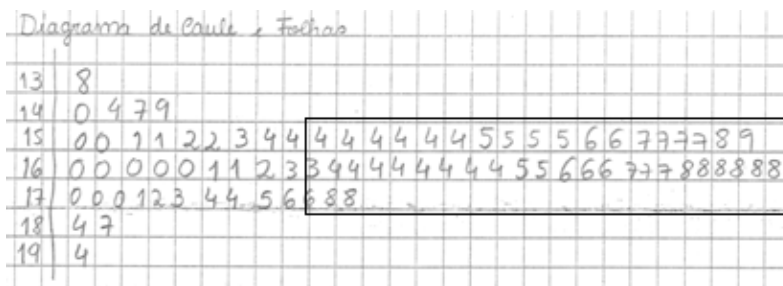


Figura 48. Resposta de um grupo ao item 8 da tarefa 4.



Figura 49. Resposta de um grupo ao item 8 da tarefa 4.

Quadro 27. Respostas apresentadas pelos alunos ao item 8 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Representação do diagrama de caule e folha com incorreções	16	57,1
Representação do diagrama de caule e folha sem incorreções	12	42,9

Item 9, determinação dos quartis e construção do diagrama de extremos e quartis para a variável altura; um grupo, perante o facto de ter detetado a falta de dois elementos, um à esquerda e outro à direita da mediana, o que não iria afetar o seu valor, esqueceu-se de refazer a determinação dos quartis, que, evidentemente, poderia ser afetado pela inclusão, acabou por apresentar valores incorretos para o 1.º e 3.º quartil (Figura 50), face aos valores apresentados para o resumo dos cinco números, a construção do diagrama de extremos e quartis não teve qualquer falha. O grupo que errou a determinação da altura mediana, apresentou valores incorretos para o 1.º e 3.º quartil “enganámo-nos a contar e riscamos de um lado e depois do outro ficou muito confuso” o que terá sido consequência da falta de organização dos dados relativos à altura numa tabela ou num diagrama de caule e folhas, como sugerido no item 6, por sua vez na construção do diagrama de caule e folhas, ao usarem como escala cada quadrícula correspondente a 5 cm, tiveram dificuldades em marcar 147 (mínimo), 156 (1.º quartil), 166 (3.º quartil) (Figura 51). Um grupo de alunos evidenciou dificuldades na escolha do valor mínimo e máximo do eixo métrico, tendo escolhido como valor mínimo 0 cm e máximo 220 cm,

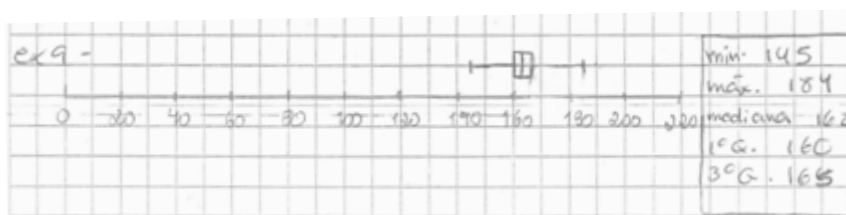


Figura 52. Resposta de um grupo ao item 9 da tarefa 4.

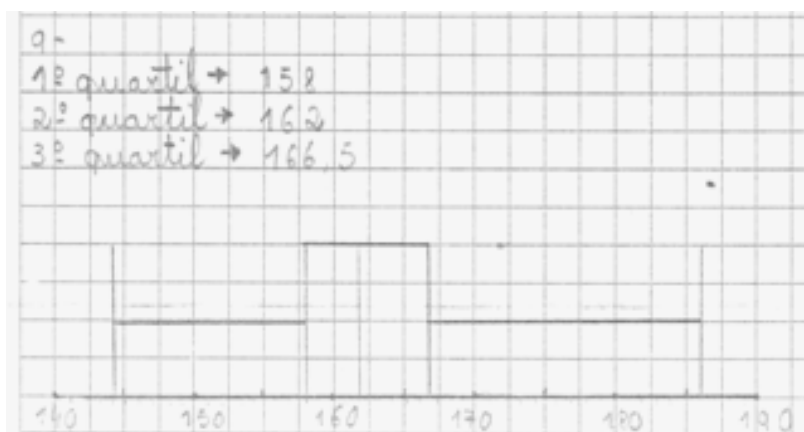


Figura 53. Resposta de um grupo ao item 9 da tarefa 4.

Quadro 28. Respostas apresentadas pelos alunos ao item 9 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Determinação dos quartis incorreta com construção correta do diagrama de extremos e quartis	4	14,3
Determinação dos quartis correta com construção incorreta do diagrama de extremos e quartis	8	28,6
Determinação dos quartis correta com construção correta do diagrama de extremos e quartis	16	57,1

Item 10, construção de um histograma para a variável altura, todos os grupos respeitaram a unidade escolhida ao longo de todo o eixo (horizontal e vertical), e iniciar a escala vertical (correspondente às frequências) na origem do referencial, apenas um grupo não respeitou o início da escala vertical (correspondente aos dados da altura) na origem do referencial, três grupos apresentam um espaçamento entre a origem e a primeira classe, os restantes apresentam um serrilhado para evidenciar o fato de se estar a contrair o eixo horizontal entre 0 e o limite inferior da primeira classe. A generalidade dos grupos optou por construir classes de amplitude 10 cm, dado que a maioria partiu do diagrama de caule e folha para a construção, sendo de salientar ainda que não indica-

ram título e/ou procederam à etiquetagem dos eixos, com exceção de um grupo que designou a representação e de dois grupos que legendaram os eixos, no entanto, todos os grupos apresentaram uma correta etiquetagem das barras. Um dos grupos optou por uma escala múltipla de 2 no eixo vertical o que auxiliou a marcação da frequência de muitas das classes e aumentou a precisão da leitura da frequência de cada classe (Figura 54), um outro grupo optou por usar a altura de duas quadrículas como unidade, o que corresponde a 1cm, e com auxílio da régua procederam a uma construção rigorosa (Figura 55), os restantes grupos optaram por uma escala múltipla de 5, pelo que existem pontuais imprecisões (Figuras 56, 57 e 58).

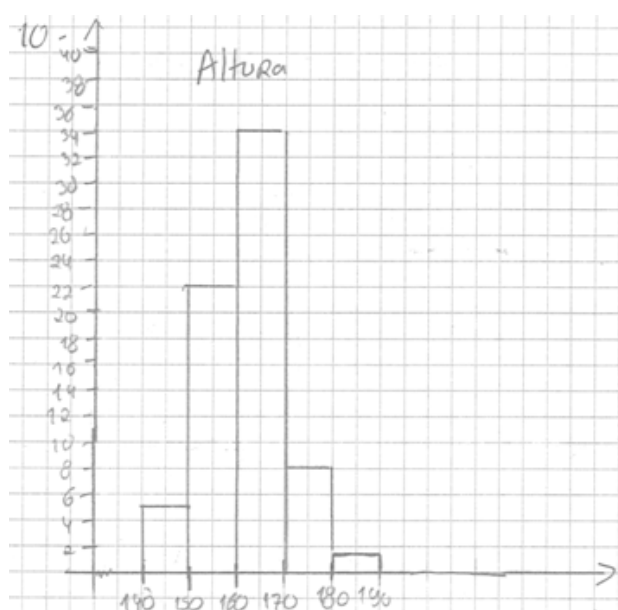


Figura 54. Resposta de um grupo ao item 10 da tarefa 4.

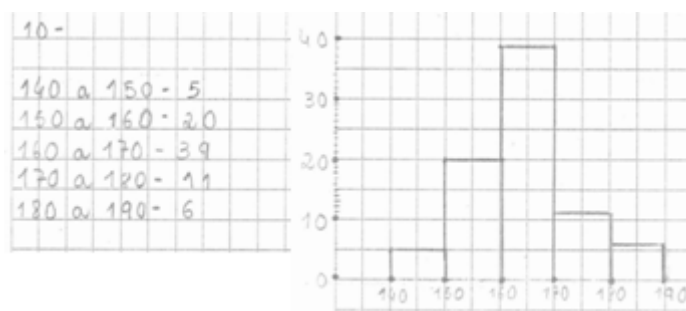


Figura 55. Resposta de um grupo ao item 10 da tarefa 4.

Um grupo colocou por baixo do histograma uma outra representação, um “gráfico de pontos” (valor da variável, frequência), tendo unido os pontos com segmentos de reta tendo obtido como que um perfil da distribuição da altura dos alunos, “dá para ver que

os retângulos menores têm poucas alturas [existem menos alunos com altura nessa classe] e o maior é onde há mais gente com altura aí...” (Figura 56). Um outro grupo pretendeu complementar a sua representação com a inclusão do polígono de frequências, mas não uniu a marca da classe anterior ao ponto médio superior da primeira classe e também não uniu o ponto médio superior da última classe à marca da classe seguinte (Figura 57). O grupo de alunos que nos itens 8 e 9 optou por comparar os dados do género masculino com o feminino para a variável envergadura e altura, para além de apresentar o histograma solicitado, incluiu um “histograma” comparativo (Figura 58), posteriormente esclareci que embora na maioria das publicações surgisse como tal o modo mais correto de proceder à comparação de dois grupos com distinta dimensão, requeria a construção de dois histogramas em separado com a mesma escala e unidade e, utilizando às frequências relativas ou então recorre à sobreposição dos polígonos de frequência relativa.

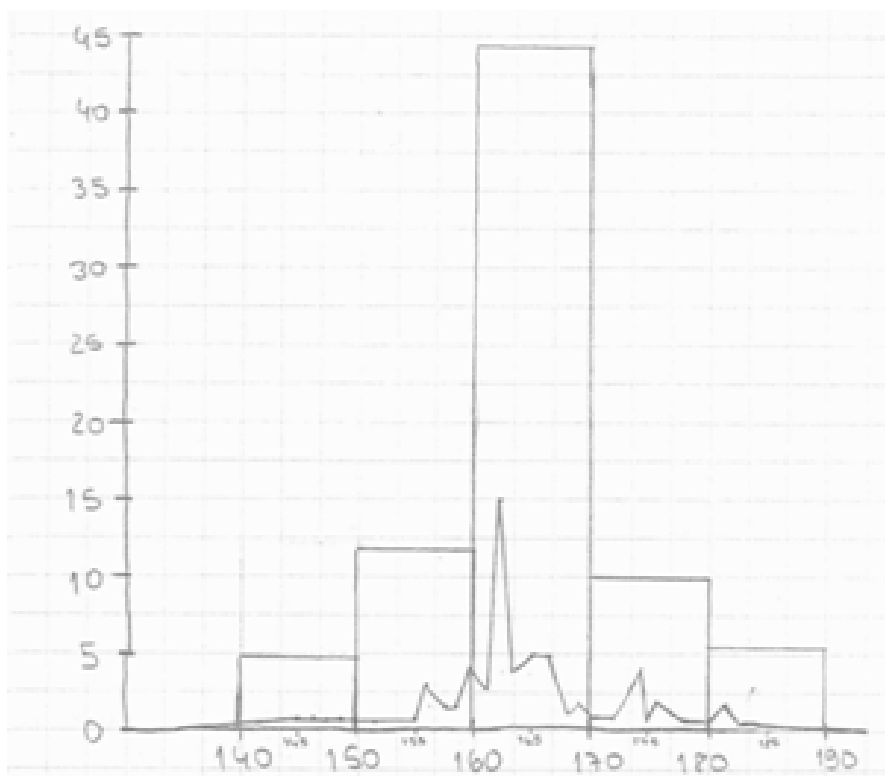


Figura 56. Resposta de um grupo ao item 10 da tarefa 4.

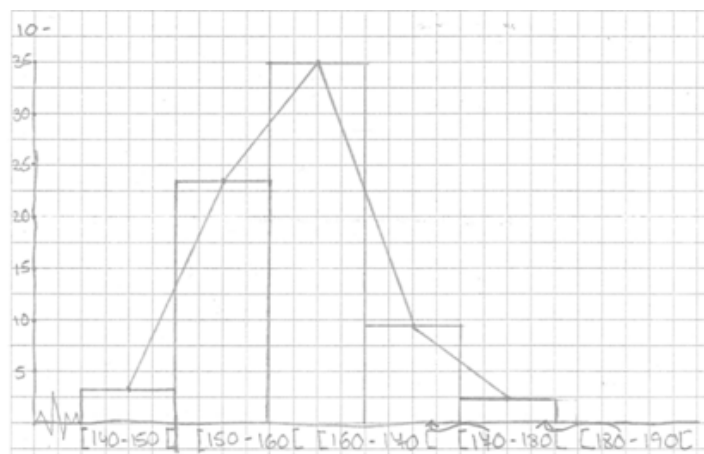


Figura 57. Resposta de um grupo ao item 10 da tarefa 4.

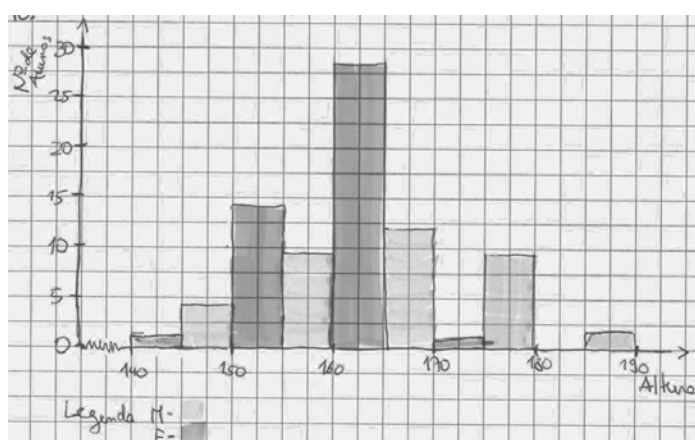


Figura 58. Resposta de um grupo ao item 10 da tarefa 4.

Quadro 29. Nível de cumprimento de alguns aspetos relativos à construção do histograma - item 10 da tarefa 4.

Resposta		Nº de alunos (N= 28)	% de alunos
Legen- das	Titulo, eixos e barras legendadas	4	14,3
	Eixos e barras legendadas	4	14,3
	Barras legendadas	20	71,4
Escala	Construção do histograma com recurso a uma escala pouco adequada aos valores das frequências – imprecisões na altura das barras	16	57,1
	Construção do histograma com utilização correta da escala – refinamento da unidade escolhida mesmo que desadequada	12	42,9
Natureza continua	Correta utilização do eixo horizontal (variável continua)	20	71,4
	Incorreta utilização do eixo horizontal (variável continua) (iniciar o eixo no limite inferior da 1ª classe ou comparação de duas informações para a mesma classe)	8	28,6

O item 11 foi realizado na aula, mas os alunos tiveram a oportunidade de usarem um tempo de estudo acompanhado para reanalisar o trabalho feito e melhorar o relatório que seria entregue apenas na aula seguinte, em que iríamos discutir a resolução dos itens (6 a 11). Um grupo optou por rever o trabalho e passar a limpo, um outro grupo reanalisou os dados do ponto de vista de cada ano de escolaridade e um terceiro grupo compilou as várias informações e pediu para ir à biblioteca redigir o relatório no processador de texto. No âmbito das conclusões, item 11 (parte III), um dos grupos apresenta um pequeno texto contendo informações descritivas da amostra que tratou, sendo que a maioria destas informações é a moda, a média e a mediana para as variáveis quantitativas (Figura 59), um outro grupo começou por detalhar alguns aspetos da amostra, referenciou a moda, no caso da variável qualitativa, as medidas centrais para a altura e a envergadura, incluindo referencia aos valores mínimos e máximos das duas distribuições, sem no entanto proceder a uma análise comparativa (Figura 60).

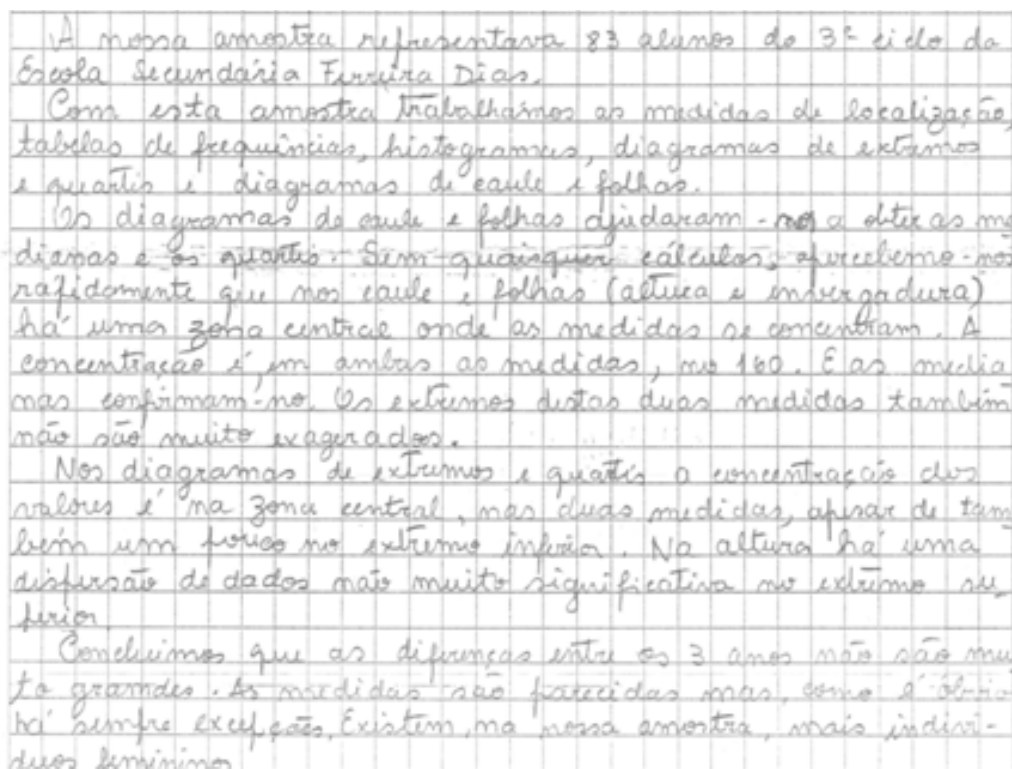
ex 11
 Conclusões:
 - Na altura, a média é 164,3, a moda e a mediana são 162.
 Com estes resultados, chegámos à conclusão que a maioria destes alunos têm estatura média.
 - No número do calçado, a média é 38 e a moda 36 e a mediana 38.
 - Na envergadura, a média e a mediana são 164 e a moda é 166.
 No calçado e na envergadura reparámos que as médias e as medianas são iguais.
 E na altura a moda e a mediana são iguais.

Figura 59. Resposta de um grupo ao item 11 da tarefa 4.

Até neste estudo trabalhamos com 31 elementos de estudo excepto com o calendário trabalhámos apenas com 30, porque uma pessoa não respondeu a esta questão.
 Ao fazer este estudo descobrimos que responderam mais pessoas da sexo feminino do que do sexo masculino, descobrimos que a mais alta tem 1,93m e a mais baixa tem 1,16m, a moda é 1,55m, a média 1,72m e a mediana 1,62m.
 Na envergadura a máxima é 1,94m e o mínimo é 1,33m, a moda é 1,54m, a média é 1,62m e a mediana é de 1,63.

Figura 60. Resposta de um grupo ao item 11 da tarefa 4.

Dois grupos para além das habituais descrições das estatísticas, incluíram um breve comentário quanto à concentração dos dados com base na análise do histograma ou do diagrama de extremos e quartis, num dos grupos afirmaram que “Na variável altura a mediana está muito perto dos quartis visto que a maior parte das medidas está entre 150 e 170” e no outro grupo “Com a observação do diagrama de extremos e quartis tiramos a conclusão que 50% das pessoas tem altura compreendida entre 1,56 m e 1,66 m”. O relatório apresentado por um dos grupos é um pouco mais extenso, começa por referir a dimensão da amostra, as ferramentas de organização dos dados e sua representação, bem como o modo como isso auxiliou a determinação das medidas de localização e a análise da concentração/ dispersão dos dados “Os diagramas de caule e folhas ajudaram-nos a obter as medianas e os quartis. Sem qualquer cálculo, apercebemo-nos rapidamente que nos [diagramas de] caules e folhas (altura e envergadura) há uma zona central onde as medidas se concentram. A concentração é, em ambas as medidas [variáveis], no 160. E as medidas [estatísticas] confirmam-no. Os extremos destas duas medidas [variáveis] também não são muito exagerados. Nos diagramas de extremos e quartis a concentração dos valores é na zona central, nas duas medidas [variáveis], apesar de também um pouco no extremo inferior. Na altura há uma dispersão de dados não muito significativa no extremo superior.” (Figura 61).



A nossa amostra representava 83 alunos do 3º ciclo da Escola Secundária Ferreira Dias.
Com esta amostra trabalhamos as medidas de localização, tabelas de frequências, histogramas, diagramas de extremos e quartis e diagramas de caule e folhas.
Os diagramas de caule e folhas ajudaram-nos a obter as medianas e os quartis. Sem qualquer cálculo, apercebemo-nos rapidamente que nos caule e folhas (altura e envergadura) há uma zona central onde as medidas se concentram. A concentração é, em ambas as medidas, no 160. E as medianas confirmam-no. Os extremos destas duas medidas também não são muito exagerados.
Nos diagramas de extremos e quartis a concentração dos valores é na zona central, nas duas medidas, apesar de também um pouco no extremo inferior. Na altura há uma dispersão de dados não muito significativa no extremo superior.
Concluimos que as diferenças entre os 3 anos não são muito grandes. As medidas são parecidas mas, como é óbvio, há sempre exceções. Existem na nossa amostra, mais indivíduos femininos.

Figura 61. Resposta de um grupo ao item 11 da tarefa 4.

Um grupo estruturou o seu relatório de acordo com os tópicos do relatório analisado na tarefa “Como vamos de vendas?” começa por identificara a questão “*Foi-nos pedido que elaborássemos um estudo para caraterizar os alunos do 3.º ciclo de uma escola relativamente às características que nós procuramos género, altura, envergadura e nº de sapato*” de seguida passaram a explicitar a metodologia utilizadas; para a variável género, na identificação da moda alegaram a inexistência da média, para as variáveis quantitativas, determinaram a média e a mediana, e apresentaram o respetivo diagrama de caule e folhas. Para a variável altura determinaram os quartis da distribuição, uma tabela de frequências e construíram as representações solicitadas no item 9 e 10, finalizaram o relatório com as conclusões, onde apresentam as habituais descrições das medidas estatísticas, incluem um breve comentário relativo a concentração dos dados e um outro, quanto à possível relação entre as variáveis envergadura e altura. Apresentaram as conclusões como extrapolações para a população “(...) *Em média a altura [dos alunos] do 3.º ciclo é 161,5 cm e a mediana 165 cm, vendo o histograma a maior parte da nossa população está entre os 160 cm e os 170 cm. A envergadura está muito próxima da altura, sendo que a envergadura em média é ...*”.

Um dos grupos produziu um texto um pouco mais extenso, começaram por contextualizar a situação; referencia à população, amostra e objetivo do estudo, comentários relativos à análise complementar entre as medidas estatísticas (incluindo referencias aos extremos das distribuições) e as representações gráficas. Ao longo do relatório, embora não assumam que estão a generalizar as suas conclusões à população afirmam: “*apercebemo-nos de que na nossa amostra existiam mais raparigas do que rapazes, o que poderá ser a realidade da nossa população em estudo.*” Concluíram que, regra geral, existiu correspondência ente a altura e a envergadura e que a análise comparativa da altura *versus* género confirma que os rapazes são mais altos que as raparigas (Figura 62).

Apenas numa terceira aula, foi possível discutir os itens 6 a 11. Como nem todos os grupos tinham a mesma amostra, e no caso de terem a mesma poderiam ter feito opções distintas no momento da “limpeza dos dados”, a discussão centrou-se em algumas das considerações e sugestões sobre as várias incorreções/imprecisões detetadas que constavam do *feedback* escrito que produzi para cada grupo, e que referia alguns aspetos tais como:

11.RELATÓRIO: A população em estudo são os alunos do 3º ciclo da Escola Secundária Ferreira Dias, onde se estuda a altura, envergadura e nº de calçado dos mesmos.

Durante a realização do trabalho apercebemo-nos de que na nossa amostra existiam mais raparigas do que rapazes, o que poderá ser a realidade da nossa população em estudo.

Na variável altura, a classe com maior frequência absoluta é a classe 160 – 165, e a sua moda e mediana é igual tendo como valor 162. Ao fazer o diagrama de extremos e quartis chegamos à conclusão que os rapazes em média são mais altos e que as raparigas não ultrapassam 1,7m valor a que vários rapazes ultrapassam. Mas o aluno mais baixo é um rapaz com 1,47m

Na variável envergadura, a classe mais frequente é também a classe 160 – 165 como na altura, a moda e a mediana são próximas da variável altura, concluindo que no geral a nossa altura, corresponde à envergadura, existindo exceções. É também a variável com maior amplitude, seguindo – se da altura.

No nº de calçado a média, a mediana e a moda coincidem, e correspondem de novo à classe mais frequente, a do nº 38 de calçado.

Consegue-se ver que apartir do nº36 para baixo, os valores descem bruscamente mas os números apartir do 39 para cima vão descendo quase uniformemente.

Conclui-se que a moda e a mediana encontram-se geralmente nas classes com maior frequência absoluta, podendo existir exceções.

Os rapazes em média são mais altos que as raparigas mas são mais de extremos e que a altura e envergadura estão relacionadas uma com a outra, porque os valores são geralmente muito próximos.

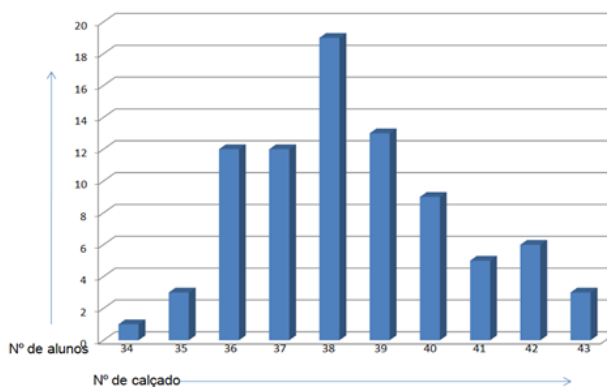


Figura 62. Resposta de um grupo ao item 11 da tarefa 4.

Quadro 30. Nível de cumprimento de alguns aspetos de análise na redação do relatório - item 11 da tarefa 4.

Resposta	Nº de alunos (N= 28)	% de alunos
Recurso apenas a medidas de tendência central para a descrição das distribuições	4	14,3
Recurso a medidas de tendência central e de dispersão para a descrição das distribuições	20	71,4
Esboços de articulação entre as medidas estatísticas e as representações gráficas	20	71,4
Não estabelecem qualquer relação	8	28,6
Estabelecem relações entre as variáveis quantitativas e o ano de escolaridade que está associada à idade (problemas com o tamanho de cada “amostra”)	4	14,3
Estabelecem relação entre as variáveis quantitativas e o género	4	14,3
Estabelecem relação entre as variáveis altura e envergadura	12	42,9
Conclusões circunscritas à amostra	20	71,4
Conclusões generalizadas à população	8	28,6

- Utilizar tabelas de frequência ou diagramas de caule e folha, nas variáveis quantitativas, para uma primeira redução dos dados;
- Utilizar o diagrama de caule e folha para uma maior facilidade na determinação das medidas de localização;
- Utilizar os diagramas de caule e folha e de extremos e quartis para comparar a forma de duas distribuições (centro e dispersão);
- Recordar a construção de um polígono de frequências;
- Recorrer às frequências relativas, caso se pretende comparar grupos de dimensão distinta nas representações - gráfico de barras e histograma – nomeadamente no caso de uma variável contínua, recorrer à comparação dos polígonos de frequências relativas;
- Utilizar escalas adequadas aos valores numéricos que se representam, escolha de uma unidade de fácil partição para maior precisão na construção e leitura da representação;
- Utilizar descritores nos eixos e na representação para facilitar a comunicação dos dados representados;
- Utilizar as medidas estatísticas determinadas em articulação com as representações construídas;
- Utilizar medidas de dispersão em articulação com as medidas de centro para a descrição das distribuições.

A discussão em torno do relatório (item 11), passou pela leitura à turma dos dois relatórios em que houve “esboços” de generalização e que tinham avançado com a conclusão de que as variáveis altura e envergadura se relacionam, tomando valores próximos, sendo que um aluno que se destaque na altura (muito pequena ou muito grande) tenha comportamento similar na envergadura (toma um valor baixo ou muito alto, respetivamente). Perante as leituras, os alunos concluíram que poderiam ter integrado mais as informações recolhidas, expressar mais as relações entre as medidas estatísticas e as representações e, que tal como inicialmente tinham suspeitado, se podia confirmar que a altura e a envergadura se relacionavam. Com o objetivo de fomentar a confiança dos alunos no âmbito do desenvolvimento e avaliação de inferências, apresentei um resumo dos parâmetros populacionais e das medidas estatísticas apuradas junto das quatro

amostras distribuídas, para duas das variáveis, depois de ter procedido à respetiva limpeza dos dados na população e amostras (Figura 63).

Caracterização das amostras e da população *

*Após a exclusão de elementos repetidos ou por falta de informação relativa ao nº de calçado

Variável	Amostra 1		Amostra 2		Amostra 3		Amostra 4		População	
	f.a	f.r%	f.a	f.r%	f.a	f.r%	f.a	f.r%	f.a	f.r%
Género										
F	42	54,5%	44	53,7%	47	61,0%	42	52,5%	339	53,2%
M	35	45,5%	38	46,3%	30	39,0%	38	47,5%	298	46,8%
Total	77		82		77		80		637	

Variável		Amostra 1	Amostra 2	Amostra 3	Amostra 4	População
Altura						
Mínimo		147	144	145	146	140
Máximo		181	186	184	190	190
Amplitude	Quest 7	34	42	39	44	50
Média		161,95	163,35	164,26	163,34	162,98
Moda		162	165	162	155	162
Mediana		162	162,5	163	162,5	162
1º Q	Quest 9	156	158	160	157,75	158
3º Q		166	166	168	169,625	168
Amplitude interquartil		10	8	8	12	10

Figura 63. Quadro resumo comparativo das medidas nas amostras com os parâmetros populacionais da tarefa 4.

A reação dos alunos foi de entusiasmo puro “*Olha ali... a média e a mediana estão parecidas...*”, “*a moda na amostra 4 é muito diferente das outras*”, “*o aluno mais baixo não é muito diferente nas amostras mas é maior do que o da população, e o maior é mais baixo ... é ao contrário*”, “*os quartis não estão mal... estão próximos*”, “*A amplitude interquartil fica ora abaixo ora acima mas não é muito diferente...*”, “*A amplitude [total] nas amostras foi sempre menor que a verdadeira...*”, “*Houve sempre mais raparigas do que rapazes o que bate certo...*” Rapidamente se aperceberam que embora as amostras fossem diferentes existia alguma consistência entre as medidas estatísticas determinadas, nomeadamente no caso da variável qualitativa em que a proporção do

género feminino era maior do que a do género masculino, compatível com a população em que se registava uma diferença de 6 pp entre os dois géneros. Quando questionados sobre se poderiam ter uma amostra em que a percentagem do género masculino fosse superior à do feminino, alguns dos alunos concordaram que podia existir mas o mais comum seria o contrário. No caso das variáveis quantitativas afirmaram que a média e a mediana são preferíveis para a generalização da informação, e que a amplitude interquartil no caso das variáveis em estudo também era um bom indicador, esclareci que isso se devia, neste caso, ao facto das variáveis terem um comportamento normal, em que a distribuição tem a forma de um sino, a generalidade das estaturas centra-se em torno de um valor central e que os extremos da distribuição ocorrem com pouca frequência. Por fim coloquei a seguinte questão aos alunos “*A realização do estudo feito e as conclusões a que se chegaram a quem poderiam interessar?*” de imediato uma aluna respondeu “*aos investigadores ... que comparam as coisas de agora com o passado*”, “*antropólogos*” digo, “*sim, stora isso.*” Um outro aluno sugere investigadores da área das ciências forenses que estimam a estatura, e um outro acrescenta o interesse para a produção de têxteis e calçado para adolescentes, finalizou-se o trabalho com a leitura de um pequeno texto relativo à antropometria e a importância do estudo de características mensuráveis no âmbito da indústria, adaptado de “*Funções no 3.º ciclo com tecnologia*” (p. 68 a 71).

Análise e interpretação. Com a realização da tarefa os alunos puderam envolver-se pela primeira vez num estudo estatístico, a população em estudo era-lhes familiar, o que trouxe vantagens em termos da discussão e ponderação de elementos que afetam a representatividade de uma amostra a extrair. A recolha de parte da informação envolveu os próprios alunos e recorreu-se a mais do que o tradicional questionamento, os alunos adoraram medir-se. O trabalho realizado pelos alunos revela que consideram a escolha aleatória da amostra como o processo mais correto e adequado para que seja representativa da população (Martins & Ponte, 2010). Os alunos ao anteviram a variabilidade da população quanto às características a estudar e optaram por amostras de significativa dimensão. Apesar de anteciparem que perante uma amostra de grande dimensão iriam ter mais trabalho no momento do tratamento dos dados, sentiram necessidade de amostras que representassem uma elevada porção da população (Garfield, 2002). Os alunos perante a fase de produção de dados tiveram que identificar elementos da população que se repetiam na amostra (geração aleatória de números inteiros compreendidos entre 1 e 647) e tiveram que tomar decisões quando os dados recolhidos tinham imprecisões e/ou

erros, o que teve o aspeto positivo de os envolver na fase de planeamento, nomeadamente na tomada de consciência do sistema de medida intrínseco às variáveis em estudo e na gestão dos dados. A generalidade dos alunos dispensou o uso de tabelas de frequência em detrimento do diagrama de caule e folha o que poderá ser consequência do facto de ter existido um hiato de tempo entre a realização do item 5 e do item 6, ou, do facto de no ano transato os alunos terem associado a representação do diagrama de caule e folhas à determinação das medidas de localização e tal associação ainda não ter sido feita com as frequências relativas acumuladas ou então advir da dificuldade na passagem de uma visão individual para uma visão agregada dos dados que esta primeira redução implica, tal como defende Batanero (2001).

Quanto à determinação das medidas de tendência central, os alunos não apresentaram dificuldades na identificação da moda em qualquer das variáveis em estudo, nem na determinação da média no caso das variáveis quantitativas, no caso da mediana, alguns alunos cometeram erros no âmbito da contagem, mas o trabalho realizado manifestou que conheciam o conceito, pois procuraram o valor central da distribuição após a terem ordenado, e no relatório final referiram o intervalo interquartil e que 50% das pessoas tinham uma altura em torno da altura mediana. O sucesso dos alunos na determinação das medidas centrais e de localização poderá ser consequência da maioria ter mantido uma visão ainda individualizada dos dados uma vez que recorreram a diagramas de caule e folha, pelo que a aplicação dos procedimentos processuais foi realizada sem dificuldades de maior. No relatório, a maioria dos alunos procurou indicar as medidas estatísticas obtidas relacionando-as com as representações construídas.

Relativamente às representações, no diagrama de caule e folhas, a dificuldade que foi mais frequente foi a utilização inconsistente da unidade escolhida ao longo dos vários caules. No diagrama de extremos e quartis e no histograma os alunos manifestaram dificuldades na escolha de uma escala mais eficiente e adequada face aos valores a representar, pelo que em alguma das representações a marcação dos cinco números de resumo ou a frequências de cada classe, poderia ter sido um pouco mais precisa e de mais fácil leitura. A generalidade dos alunos não deu título às representações e/ou não procedeu à etiquetagem dos eixos, o que corresponde a uma das convenções que 90% dos alunos envolvidos nos estudos de Wu (2004) manifestaram desconhecer ou não utilizar. É ainda de salientar que na construção dos histogramas, embora nem todos os alunos tenham usado a convenção de cortar o eixo ou o serrilhado de contração dos valores compreendidos entre 0 e o valor mínimo de altura registado, apenas um grupo de quatro

alunos “colou” a barra correspondente à primeira classe ao eixo vertical/ das frequências, os alunos não confundiram a representação solicitada com um gráfico de barras, e o único grupo que se deparou com uma classe de frequência nula procedeu à sua correta representação e não a omitiu. Ainda no âmbito das representações, três grupos de alunos recorreram à interpretação do diagrama de extremos e quartis, um dos grupos referenciou o intervalo interquartil, o outro grupo de alunos optou por salientar a concentração do intervalo interquartil face à concentração/ dispersão dos intervalos [mínimo, 1.ºQ] e [3.º Q, máximo] e o terceiro grupo optou por fazer uma análise comparativa dos diagramas de extremos e quartis *versus* o género, e um outro grupo recorreu à análise da concentração em torno do centro da distribuição expressa no histograma. Tal análise, embora possa ser melhorada em termos da clareza e correção científica, manifesta que um grupo significativo de alunos se mostrou capaz de recorrer a uma visão agregada dos dados e compreender que, no caso do diagrama de extremos e quartis, se apresenta a densidade relativa (dispersão e variação dos dados) em vez da frequência, bem como de apreender que a densidade relativa se relaciona de modo inverso com o comprimento da componente do diagrama, aspetos que Bakker, Biehler e Konold (2005) consideraram necessários para a sua correta interpretação.

No âmbito do desenvolvimento e avaliação de inferências, a generalidade dos alunos mostrou-se pouco à vontade a generalizar as considerações estabelecidas no âmbito da amostra representativa para a população em estudo, pois não sentiam legitimidade para tal, uma vez que a sua amostra era apenas uma pequena parte da população. Tal comportamento corresponde a uma das dificuldades identificadas por Rubin, Bruce e Tenney (1991) quando os alunos trabalham com base em amostras aleatórias. Rubin, Bruce e Tenney (1991) identificaram um outro erro usual no trabalho com amostras aleatórias, a crença de que uma amostra aleatória deve ser um retrato fiel da população. No caso dos alunos desta turma foi possível constatar que este erro não foi cometido, pois, com relativa facilidade, aceitaram que na população, o género feminino predominasse, mas que pudessem existir amostras em que tal não acontecesse, ou que no caso de uma amostra, a moda da altura não correspondesse à altura mais frequente da população, e que naturalmente, o mínimo e o máximo da altura numa amostra pudessem ser respetivamente maiores e menores/ou iguais aos correspondentes valores na população. Este resultado poderá ter sido consequência da análise comparativa feita com as medidas estatísticas das quatro amostras com os parâmetros populacionais, procedimento que

segundo GAISE (2005) e NCTM (2007) ampliar a capacidade de testar a generalização das suas conclusões.

O equilíbrio entre o reconhecimento da variabilidade do processo de amostragem e o reconhecimento de comportamentos que se poderão inferir para a população (Rubin, Bruce & Tenney, 1991) parece-me ter sido conseguido pois, embora os alunos tivessem reconhecido que as conclusões de um estudo estatístico estão associadas a um certo grau de incerteza, foi possível tomar decisões com base neste estudo, nomeadamente ao nível da inferência da forma /centro e da dispersão das distribuições das variáveis quantitativas para a população. Perante as produções e discussão dos vários itens constata-se que o envolvimento dos alunos em todas as fases do ciclo investigativo teve como consequência o entendimento do papel da estatística na construção de modelos de análise de algumas características humanas que são decisivos, nomeadamente para a otimização da produção industrial (Martins& Ponte, 2010).

Balanço. Os alunos envolveram-se na resolução da tarefa com agrado, no entanto, ao longo do tratamento de dados houve momentos de frustração e cansaço, dada a elevada dimensão das amostras. O tempo de realização previsto – 2 blocos – revelou-se insuficiente pois utilizaram-se 45 minutos de estudo acompanhado (EA) para a definição do objeto de estudo, formulação de questões e recolha de dados junto dos alunos da turma e os alunos puderam ainda dispor de outro meio bloco de EA para passar a limpo ou melhorar o relatório do estudo estatístico realizado. Na aula de Matemática utilizaram-se 2,5 blocos sendo meio bloco para a resolução, discussão dos itens 1 a 5 e geração dos quatro conjuntos amostrais; um bloco para a realização dos itens 6 a 11 e um bloco para a discussão dos itens 6 a 11, comparação das estatísticas das amostras com os parâmetros da população em estudo e síntese da tarefa.

Penso que a tarefa poderá sofrer algumas alterações de pormenor, tais como: (i) incluir no item 6, “Para cada variável, organiza os dados em tabelas de frequência”; na parte III pois trata-se de uma primeira redução dos dados o que se deveria incluir no tratamento dos dados; e (ii) incluir guias quanto ao tipo e formato de um relatório no item 11, (Figura 64), considero importantes esta alteração, pois a discussão oral desses elementos na tarefa 3 “Como vamos de vendas?” não foi suficiente para que a generalidade dos grupos tomasse o relatório apresentado como ponto de partida do que iriam construir nesta tarefa “Um estudo na escola”.

Relatório – deve ser dirigido à professora ou aos colegas, que entende o assunto tratado, embora possam não se ter envolvido diretamente no estudo que irão abordar.

- Deve ser escrito na 3.^a pessoa, com verbos em tempo *passado* porque descreve um estudo que já se realizou.
- Deve ter a seguinte forma

Introdução – Objetivo do estudo – questão/ problema a responder formulada de modo claro. Podendo incluir hipóteses ou conjeturas que considerem pertinentes face a questão/problema em estudo;

Metodologia – Procedimentos adotados – amostra escolhida (procedimento de seleção e dimensão); natureza dos dados recolhidos e de que modo se procedeu à sua gestão; tabelas de resumo dos dados; medidas estatísticas determinadas; representações gráficas consideradas pertinentes para a análise e comunicação dos dados e resultados/ considerações imediatas da análise das medidas estatísticas e das representações.

Conclusões – A partir dos dados analisados e dos resultados obtidos dar resposta à questão/ problema se tal for possível, verificar a validade das hipóteses ou conjeturas inicialmente colocadas ou argumentar porque não é possível responder à questão/ problema inicial.

Figura 64. Relatório – Guias orientadoras quanto ao estilo e formato.

5.2.5. Tarefa 5 – A frequência das vogais na língua portuguesa

Esta tarefa (Anexo 13) tinha como propósito o estudo da frequência das vogais na língua portuguesa com base em amostras extraídas de textos da língua portuguesa, tendo subjacente a identificação e a minimização de possíveis fontes de enviesamento na recolha dos dados, bem como a variabilidade existente no processo de amostragem (Figura 65).

Uma editora de jogos vai introduzir no mercado um jogo cujo objetivo é construir palavras. As letras são colocadas em fichas que os jogadores colocam em cada jogada. Para construir um número adequado de fichas é preciso conhecer a frequência relativa com que ocorre cada uma das letras na língua portuguesa. Como algumas das vogais são as letras mais frequentes, a empresa começou por organizar um estudo sobre a frequência de cada uma das vogais.

1. Vamos fazer o mesmo estudo a partir do texto “A Rádio escola” ...
 Escolhe quaisquer 5 linhas consecutivas do texto.
 Considerando as linhas que escolheste preenche a tabela com a frequência com que aparece cada uma das vogais.

Variável/ vogal	Frequência absoluta na amostra do grupo A	Frequência absoluta na amostra do grupo B	Frequência absoluta na amostra do grupo C	Frequência absoluta na amostra do grupo D	Frequência absoluta na amostra do grupo E	Frequência absoluta na amostra da turma	Frequência relativa na amostra da turma
a							

Figura 65. Excerto da tarefa 5.

A aula e as produções dos alunos. A tarefa foi proposta aos alunos do seguinte modo: leitura em grande grupo do primeiro parágrafo, seguida de algumas chamadas de atenção por parte da professora: *“Reparem como um estudo estatístico pode ter pertinência ... para uma empresa, no caso deste jogo com letras, as fichas colocadas tem que respeitar, digamos, a proporção que aquela letra tem na língua... que vogal vos parece ser a mais habitual no português? Os alunos respondem a vogal “a”. Será que na língua inglesa a letra “a” é usada com a mesma frequência que em Portugal? A generalidade dos alunos discorda. E para a língua francesa e ... então como deve proceder a empresa?... Pois bem, pensou em fazer um levantamento do número de vogais na língua portuguesa, cabe a vós ajudar a empresa, e como o vão fazer? ... recorrendo a um texto “A Rádio escola” passível de ser representativo pois à medida que vai sendo escrito recorre a palavras em que intervêm as vogais, tal como ocorre habitualmente no uso da [nossa] língua. Que vos parece? Estão prontos para fazer o estudo?...”*.

Os elementos de cada grupo repartiram as contagens entre si e rapidamente se obtiveram os dados e procedeu-se ao preenchimento da tabela do item 1 (Figura 66), a utilização do processador de texto revelou a desvantagem de não incluir na contagem as vogais acentuadas, ainda assim foi útil na diminuição do tempo de confirmação da contagem. Como eram 7 grupos optei por distribuir as amostras A e E a dois grupos de alunos, em vez de se seleccionar dois ovos conjuntos de 5 linhas, mantendo assim a tabela de registo proposta na tarefa.

Variável/ vogal	Frequência absoluta na amostra do grupo A	Frequência absoluta na amostra do grupo B	Frequência absoluta na amostra do grupo C	Frequência absoluta na amostra do grupo D	Frequência absoluta na amostra do grupo E	Frequência absoluta na amostra da turma
a	54	65	59	57	53	288
e	45	38	45	42	47	217
i	19	30	23	26	23	121
o	45	34	42	37	38	196
u	20	14	23	25	15	97
Total de vogais encontradas	183	181	192	187	176	919

Figura 66. Preenchimento da tabela do item 1 da tarefa 5.

Quanto ao item 2, todos os grupos responderam corretamente que a vogal mais frequente era a letra a, tendo argumentado na discussão final que surgiu 288 vezes entre as 919 vogais contabilizadas, o que confirmava as suas suspeitas iniciais.

Quadro 31. Respostas apresentadas pelos alunos ao item 2 da tarefa 5.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente – vogal a	28	100,0

No item 3, os alunos tiveram dificuldades em recorrer ao raciocínio proporcional, a professora acabou por intervir junto de três grupos de modo a auxiliar.

A1- Nós passamos a 3 e estamos na 4

A2 - não percebemos a 3...

Prof^a – Então vamos esclarecer... ou seja ente as 200 peças vocês acham que serão 40, 40, 40,40 e 40 respetivamente para cada vogal, repartir as 200 peças de igual modo entre as vogais?

Alunos- Não / Claro que não ...

Prof^a – Então vocês não acham que será de modo, 40, 40, 40... para cada uma das vogais por que assim o que aconteceria ... iríamos ter falta de que peças?

A2 – há mais a do que i ...

Prof^a - então iríamos ter falta de a e sobravam is, certo?

Os alunos anuem.

Prof^a – Então aqui estão a estudar as vogais num texto [amostra] com 919 vogais e tu agora ... queres construir uma distribuição ou seja ... isto (indica a primeira coluna) é uma distribuição das várias vogais numa amostra com 183 vogais .. e isto a distribuição das vogais numa amostra de 198 (3ª coluna) e por ai fora ... aqui é uma distribuição das vogais numa amostra já muito grande (coluna da amostra turma) .. agora temos um conjunto de 200 peças como perceber que parte será de as, que parte será de es ...

A1 – com percentagem ..

Prof^a – percentagem ... como utilizavam essa ideia das percentagens?

A1 – por exemplo 919 está para 200 é isso e 183 está para x?

Prof^a- e por quê 183? .. 183 é o total da tua amostra?..

A1 – não era o total de as 288 está para x, e veria assim as peças de a e assim sucessivamente...

Prof^a – (para os colegas do grupo) concordaram com o processo apresentado ... e está a A1 a usar percentagens?...

Antes dos colegas responderem ...

A1- não, mas uma regra de três simples.

Prof^a – pois é um raciocínio proporcional como o das percentagens... concordam com esta estratégia? ...

A3 – pois é aquela regra (faz um movimento em X)

Prof^a – regra de três simples ou proporção caso escrevam a relação de outro modo... Então concordam com a estratégia da A1 ...

Alunos anuem ...

Prof^a – E porque é que é aceitável ... por que fará sentido?

Silêncio

Prof^a – pois percebemos que estas amostras têm este padrão de comportamento, então a amostra que resulta da junção de todas torna-se um conjunto mais próximo da realidade da língua portuguesa, ... a proporção de vogais a na totalidade de vogais que terá a [utilização da] língua portuguesa, será dada aproximadamente por 288/919, basta agora redimensionar para um conjunto de 200 peças, perceberam?

Outros alunos debateram-se com a questão de recorrer a um raciocínio proporcional com base nos dados da sua amostra ou com base nos dados da amostra da turma, sendo que dois grupos recorreram às frequências absolutas da respetiva amostra.

A1 - Temos que fazer uma regra de três simples. É o mais fácil de fazer então 289 está para 919 assim como 200 está para x .

A2 - ... mas isto é fazer sobre a parte ou sobre tudo.

A1 - não é para fazer com a frequência total [de cada vogal].

A2 - Da turma ou da amostra?

A1 - da turma!.

A2 -... então porque é que a professora disse que éramos o grupo A?

A1 - foi para recolher os dados das nossas linhas...

A3 - tá bem ... dá 62,6

A1 - 63 [peças com a vogal a]

A2 - já... Seja 63, pedem o número de fichas, o número é inteiro ...

A4 - eu não fiz esta cena [explicitação da regra de três simples], só ... [e aponta para o cálculo $x = \frac{289}{919} \times 200$].

Apenas um grupo de alunos determinou a frequência relativa, em percentagem de cada vogal na amostra turma e recorreu a esses valores para determinar o número de fichas de cada vogal no conjunto das 200 vogais.

Um grupo estimou a distribuição do número de peças de cada vogal por tentativa considerando a referência de dimensionar uma amostra tomando 10% da população, utilizaram 10% da frequência absoluta na amostra turma para cada vogal, sendo que a soma depois de arredondar ao valor inteiro não chegava a totalizar 100, experimentaram então 20%, como a aproximação era melhor mas a soma do número de peças ainda era inferior a 200, aumentaram ligeiramente, consideraram 22%, sendo que neste caso a soma do arredondamento inteiro de 22% de 288, 22% de 217, 22% de 212 com 22% de 196 e 22% de 97 somava 200. Não tendo sido capaz de justificar o recurso ao valor 22%, posteriormente em grande grupo e após a exposição da sua resolução, esclareci que se tratava da relação entre a dimensão do conjunto a estimar e a dimensão da amostra turma $\frac{200}{919}$ que toma o valor de 21,76 % aproximadamente.

A1 – se fosse 20% [de 288] dava 57,6 ou seja 58, e a seguir 43, 24, 39 e 19 mas não dava 200 (somam 183) e ... nós pensamos ou arredondávamos os números ainda para cima ou fazíamos com uma percentagem maior...

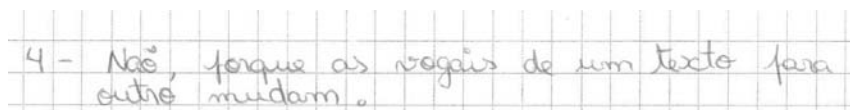
A1 – fomos experimentando, deu certo com 22%

Prof^a – Neste momento não estou a ver porque está correto ... deixem-me pensar um pouco mais no que disseram ... será que não é apenas uma coincidência numérica... não mas funcionou perfeitamente para todas as vogais... esperem o conjunto das 200 vogais em relação à nossa amostra ...(recorro à calculadora) a relação entre os tamanhos é que é 22%, isto é 200 é cerca de 22% de 919 logo o número de peças [com vogal] a nas 200 é também 22% das que existem na amostra turma e assim sucessivamente.

Quadro 32. Respostas apresentadas pelos alunos ao item 3 da tarefa 5.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente	28	100,0

Relativamente ao item 4, sempre que os alunos manifestaram necessidade de concretizar a experiência num outro texto, sugeri o uso de uma página específica de qualquer outro manual escrito em língua portuguesa que tivessem na mochila. Os alunos nos vários grupos responderam que ao repetir a experiência não iriam ter os mesmos resultados, mas a generalidade dos grupos não recorre à frequência relativa de cada vogal, alegam apenas que o uso de outras palavras iria motivar quantidades distintas para cada uma das vogais, pois varia de um texto para outro (Figura 67).



4 - Não, porque as vogais de um texto para outro mudam.

Figura 67. Resposta de um grupo ao item 4 da tarefa 5.

Quadro 33. Respostas apresentadas pelos alunos ao item 4 da tarefa 5.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente	28	100,0

Na discussão do grande grupo, o aluno insistiu no facto do resultado de uma nova experiência ser necessariamente distinto, pois o número de vogais varia de texto para texto. Um outro grupo de alunos reconhece que, embora os resultados de um texto

para outro sejam distintos, existe um padrão de comportamento que se observa na generalidade dos textos representativos da língua portuguesa.

Prof^a – Vocês só disseram que era não, tem mais alguma parte para além de não...

A1 – sim

Prof^a – Então diga-la...

A1 – porque um outro texto pode ter outras palavras com outro número de vogais...

Prof^a – E isso chega... comparar o número... nem sequer vai ser próximo quando se pensa na percentagem de cada vogal no conjunto de todas?

A1 - Talvez...

A2 – Acho que depende do texto aqui neste pedaço do livro de história há mais “e” do que “a”.

A3 – Mas então o a, o e e o o vão ser sempre as mais utilizadas ...

Prof^a – Por isso estão chocados por ter aparecido ligeiramente mais “e” do que “a” ... chocamos?

Os alunos agitam-se mas parecem concordar.

Prof^a – se tivesse aparecido mais “i” do que “a”... chocava-nos?

Alunos – sim /choca

Prof^a – e nessas linhas do texto de história tal acontece?...

Alunos que recorreram ao texto de história – não acontece.

Prof^a - Então vamos lá ver ... depende do texto ...

A4 – Se for um texto em rima ão, vão aparecer mais a e o do que das outras todas...

Prof^a – Perceberam o que A4 disse ... esclarece lá o que disseste.

Alguns alunos – não.

A4 – Se o poema tiver o fim dos versos em ão, nós podemos ter mais a e “o” do que das outras [vogais].

Prof^a - Do que a proporção habitual dessas vogais... certo!... E um poema é uma amostra representativa da língua portuguesa?

Vários alunos, mesmo A4 – Não.

Prof^a – Então chocar-me ia que no caso de um poema, contrarie a tendência que a gente analisou nestas nossas amostras ...

Alunos – Não / não choca

A5 – Normalmente não iria variar muito quando uma pessoa vai escrever normalmente, só se estiver a escrever de maneira que não apareça [a vogal] a ou que apareça uma rima é que vai dar muito diferente.

Prof^a – Não estou há espera que sejam exatamente aqueles valores, certo? Mas estou à espera que sejam...

A3 - parecido

Prof^a – porquê parecido como diz a A3?

A5 – porque quando se escreve normalmente a pessoa não vai pensar em que palavra vai escrever para

A6 – antes de se escreve tu pensas sempre...

A5 - sim mas se quiser escrever agulha não vou pensar que outra coisa [palavra] que não tenha “aa”.

Prof^a – Perceberam o que A5 disse, mas no caso de um poema... a pessoa tem uma intensão, certo? ... utilizar a língua portuguesa ao serviço de uma rima, pode-se dar o caso de não ser um texto representativo da língua portuguesa.

Prof^a – e no texto de [livro] história estava-se a usar a língua portuguesa com alguma intencionalidade literária ou usou-se a língua portuguesa de modo corriqueiro... digamos assim, chocou-vos aparecerem mais “e” do que “a”?

Alunos (reiteram) - Não.

Prof^a – Então como seria uma resposta completa à questão 4? (e relê a questão).

A7 – Posso dizer a nossa resposta de grupo... “*Não, mas as vogais a, e e o seriam as mais usadas e i e u seriam as menos usadas apesar do número de vogais ser diferente de texto para texto.*”

Quanto ao item 5, os grupos optaram na maioria por observar as primeiras cinco linhas completas e face às contagens (e – 75, i – 32, o – 61 e u – 16), concluem que os resultados tinham que ser necessariamente distintos pois a vogal que tem mais representação na nossa língua foi propositadamente excluída da escrita deste texto, tal como indica o título (Figuras 68 e 69).

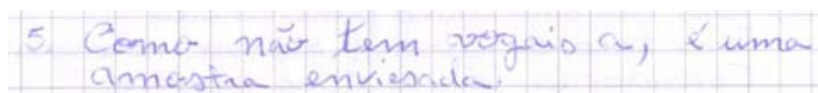


Figura 68. Resposta de um grupo ao item 5 da tarefa 5.

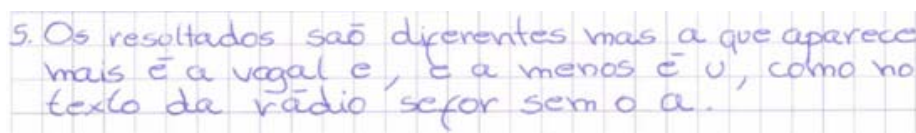


Figura 69. Resposta de um grupo ao item 5 da tarefa 5.

Quadro 34. Respostas apresentadas pelos alunos ao item 5 da tarefa 5.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente	28	100,0

No item 6, dois dos grupos responderam por escrito e recorreram à mesma estratégia de comparação da ordem de grandeza das frequências absolutas/ relativas na amostra turma, com a frequência da ocorrência das vogais entre as diferentes letras da língua portuguesa. Dos restantes grupos, três deles discutiram o item, e na discussão final concordam que a amostra turma era representativa, não tendo, no entanto, conseguido justificar a qualidade da amostra turma.

A1 – Olhando para o texto e para os resultados obtidos, não podemos dizer que os 288 “a” seja 14,63% por que no nosso texto só estamos a observar as vogais e na folha [frequência de ocorrência de letras na língua portuguesa] está o alfabeto todo...

Prof^a – Desculpa interromper mas esta ideia, toda a gente percebeu?

Alguns alunos – Sim

Outros alunos – Mais ou menos

Prof^a – Como é que estes valores [da folha] foram obtidos?

A2 – Sendo 100% as letras todas ...

Prof^a – utilizando, digamos assim, as várias palavras que se escrevem em língua portuguesa, foi-se contabilizar todas as letras do alfabeto.

A3 - no nosso caso o a dá 31,3% em vez de 14,63%.

Prof^a – Os alunos que ainda não calcularam... podem rapidamente fazer o cálculo da frequência relativa de cada vogal da amostra turma, 217 a dividir por 919, etc..

(...)

A2 - ... nesta lista nós podemos ver que os 100% englobam mais de 23 letras e quanto que aqui só englobam cinco.

Prof^a- o que o grupo da A1 e da A2 querem dizer é que aqui [texto “A Rádio escola”] o universo de análise foi todas as letras?

Alunos – não

A1 e outros alunos – apenas as vogais.

Prof^a – então acaba lá a vossa resposta, como é que compararam?

A1 – Comparamos com a ordem das percentagens, por ordem decrescente, como batia certo é uma boa amostra.

Prof^a – Vem pôr no quadro o que fizeram...

Quadro 35. Respostas apresentadas pelos alunos ao item 6 da tarefa 5.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	8	28,5
Responde corretamente sem argumentação	12	43
Responde corretamente e argumenta	8	28,5

Análise e interpretação. Com a realização da tarefa os alunos puderam envolver-se novamente num estudo estatístico, embora considere que a questão da população a estudar, conjunto das palavras da língua portuguesa que tinha a particularidade de não ser um conjunto de pessoas, tenha passado despercebida aos alunos. Ao recorrer-se à observação para obter os dados, os alunos puderam acrescentar mais um método de recolha ao seu repertório (entrevista, questionário). Os alunos envolveram-se na discussão e ponderação de elementos que afetam a representatividade de uma amostra. Um outro aspeto positivo passou pelo reconhecimento da variabilidade do processo de amostragem e o reconhecimento de comportamentos que se poderão inferir para a população (Rubin, Bruce & Tenney, 1991), embora em nenhuma das amostras as percentagens para cada vogal a, e e o tenham sido idênticas, em quase todas as amostras as percentagens correspondentes a essas vogais eram as mais elevadas. A atividade desenvolvida em torno da tarefa acabou por ser importante na tomada de consciência do papel dos métodos estatísticos como meios poderosos de tomada de decisão – definição da distribuição das 200 vogais para o jogo que a empresa pretendia lançar no mercado - e ponderação do grau de incerteza associado à distribuição obtida (Martins& Ponte, 2010).

Balanço. Os alunos participaram com agrado na realização da tarefa, tendo sido adequado o tempo previsto para a sua resolução. O facto de alguns grupos se terem debatido com o item 3: O que se pretende? Como se deve proceder para obter a distribuição?, teve como consequência apenas dois grupos terem realizado a totalidade da tarefa. Considero que a tarefa poderá sofrer algumas alterações: a inserção na tabela do item 1 de mais uma coluna à direita da frequência absoluta, para a determinação da frequência relativa, em percentagem; o item 1 passe a integrar a parte 1 – Recolha de dados; os itens 3 a 6 sejam a parte 2 – Representatividade da amostra turma; e por último, o item 2 corresponda a uma parte 3 – Tomada de decisão.

5.2.6. Tarefa 6 – Previsões

A tarefa “Previsões” (Anexo 14 - Figura 70) tinha como propósito o estudo da frequência do consumo de cada uma das bebidas vendidas no bar da escola, com base numa amostra de alunos questionado na fila do bar, com o objetivo de elaborar uma recomendação do número de unidades de cada tipo de bebida que o bar da escola deve encomendar mensalmente.

A aula e as produções dos alunos. Após uma breve revisão dos objetivos dos dois últimos estudos realizados (tarefa 4 e 5) e do interesse das conclusões obtidas de cada estudo para diferentes grupos industriais (produção de jogos, têxtil, calçado, investigação na área da antropometria), introduziu-se a última tarefa, e em diálogo com os alunos, foi-se clarificando o objetivo do estudo estatístico nomeadamente qual a questão de investigação e que questão de pesquisa seria a mais adequada.

Com esta tarefa pretende-se fazer recomendações ao bar da escola quanto ao número de unidades de cada uma das bebidas que deveria ser encomendado por mês.

Parte I – Escolha da amostra

Nos dois intervalos da manhã e nos dois intervalos da tarde cada grupo questiona um total de 20 alunos da fila do bar sobre se vão comprar alguma bebida e no caso afirmativo qual a bebida que vão comprar. Cada grupo escolhe um dia da semana para realizar a recolha dos dados.

1. Qual é a população em estudo? E qual é a sua dimensão?
2. Qual é a dimensão da amostra com que vamos trabalhar?
3. Indica como proceder para escolher uma amostra representativa.
4. Indica um procedimento que leve à escolha de uma amostra enviesada.

Parte II – Recolha de dados

Figura 70. Excerto da tarefa 6.

Prof^a - agora o que vamos fazer nesta última tarefa é um outro tipo de ajuda que a estatística dá ... quando temos uma gestão de um determinado estabelecimento comercial que já está em funcionamento, há interesse em fazer uma estimativa para os nossos estoques ... quando abro o estabelecimento sei quanto vai ser consumido de cada produto que tenho à venda?

(Alunos respondem que não)

Prof^a - Por vezes os estoques quebram antes do revendedor vir, outras vezes há sobras de um determinado produto que afinal não é tanto do agrado dos consumidores ... então de que maneira vos parece que a estatística pode ajudar em problemas deste género?"

A1 - fazemos um inquérito sobre que produtos preferem.

Prof^a - E junto de quem se aplica o inquérito?

A2 - a alunos

A3 - as pessoas que lá vão comprar

Prof^a - Sim aos consumidores do estabelecimento comercial ... no caso de se tratar de um negócio dentro da escola tem sentido questionar alunos. Vamos agora ler o parágrafo introdutório da nossa tarefa.

(Uma aluna lê em voz alta)

Prof^a - Então perante o objetivo deste estudo que questões iriam incluir no vosso inquérito? E a quem iriam questionar?

A4 - Os alunos da escola sobre "Qual a tua bebida preferida?"

A5 - "As bebidas que costumam beber?"

Prof^a - ... então um aluno responde, costumo beber sumo X, leite com chocolate e sumo néctar, um outro responde gosto de beber leite com chocolate ... e de que modo as respostas dos alunos vos ajudam a recomendar ao gestor do bar da escola o número de bebidas a encomendar por mês?"

A6 - ele vai encomendar mais coisas daquelas que as pessoas mais bebem e do resto menos.

Prof^a - Pois dá para uma ideia da grandeza com as questões de preferência, certo? ...mas na tarefa trata-se de uma questão de tentarmos que não haja prejuízo de ter guardado em estoque, o que ocupa espaço, tempo de descarga ... daí procurarmos fazer uma recomendação o mais ajustada possível. Que questões poderiam nos dar um pouco mais de precisão?

A7 - Que bebida vai beber?

Prof^a - Que vós parece? Concordam ... que se questione aos alunos da fila do bar sobre que bebida vai beber?

Alunos anuem...

Passou-se à leitura do segundo parágrafo.

Prof^a - Então para estimar a encomenda mensal que questões levariam aos alunos?

A4 - "Vais comparar alguma bebida?" Se disserem que sim "Qual a bebida?"

Prof^a - Como podem constatar, deste modo temos informações mais precisas sobre o consumo de cada bebida e não tanto a preferida ou as mais habituais ...

Os alunos não se apercebem que as questões inicialmente formuladas "Qual a tua bebida preferida?" e "As bebidas que costumam beber?" irão fornecer informações distintas e que em ambos os casos a informação recolhida não é adequada para proceder à estimativa do número de unidades a encomendar mensalmente de cada um das bebidas à venda no bar na escola. Tendo optado por deixar tais chamadas de atenção para o final da tarefa no momento de síntese, sugeri que iniciassem a resolução dos itens correspondentes à parte I e depois da sua discussão iríamos dedicar à recolha dos dados. Os alu-

nos organizaram-se nos grupos habituais e passaram à resolução dos primeiros quatro itens.

No item 1, todos os grupos identificaram corretamente a população em estudo, no entanto, quanto à questão da dimensão, um grupo apresentou como dimensão, mais de 2000 alunos, o que corresponde a um valor excessivo que ultrapassa o valor real – cerca de 1600 alunos (Figura 71). Os restantes grupos consideraram que a dimensão da população estaria compreendida entre 1400 e 1800, face ao número médio de alunos por turma e o número de turmas existente na escola ou tomado em consideração a existência de ensino básico e secundário, bem como ensino noturno na escola (Figura 72).

Figura 71. Resposta de um grupo ao item 1 da tarefa 6.

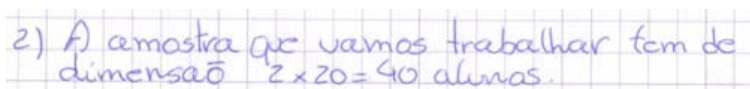
Figura 72. Resposta de um grupo ao item 1 da tarefa 6.

Quadro 36. Respostas apresentadas pelos alunos ao item 1 da tarefa 6.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Identifica a população em estudo corretamente e estima a dimensão muito acima do valor real	4	14,3
Identifica a população em estudo corretamente e apresenta uma estimativa válida da sua dimensão	24	85,7

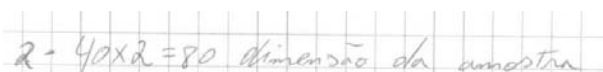
Item 2, identificar a dimensão da amostra. A generalidades dos grupos consideraram apenas a dimensão do conjunto de dados de que iriam ser responsáveis por recolher respondendo 40 alunos pois existem 2 turnos de intervalos (Figura 73) e dois grupos reponderam 80 alunos pois consideram que devem questionar 20 alunos da fila do bar em cada um dos 4 intervalos (Figura 74) as dificuldades de interpretação da questão advém da interpretação da frase “Nos dois intervalos da manhã e nos dois intervalos da tarde cada grupo questiona um total de 20 alunos das fila do bar ...” e de conjugar o

facto de que “Cada grupo escolhe um dia da semana para realizar a recolha dos dados” pelo que cada grupo questiona um total de 20 alunos durante cada turno de intervalos ou seja 40 alunos e como na turma se formavam habitualmente 7 grupos, a dimensão da amostra que vão trabalhar será de $40 \times 7 = 280$ alunos.



2) A amostra que vamos trabalhar tem de dimensão $2 \times 20 = 40$ alunos.

Figura 73. Resposta de um grupo ao item 2 da tarefa 6.



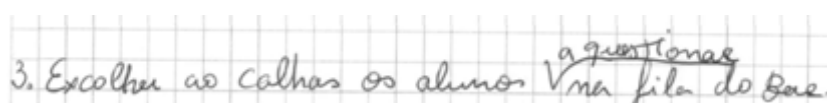
2 - $40 \times 2 = 80$ dimensão da amostra

Figura 74. Resposta de um grupo ao item 2 da tarefa 6.

Quadro 37. Respostas apresentadas pelos alunos ao item 2 da tarefa 6.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Responde corretamente	0	0
Responde incorretamente	28	100,0

No item 3, os vários grupos optaram por um seleção aleatória simples dos alunos que iriam constituir a amostra (Figura 75), complementando na discussão que a questão de “ao calhas” deveria respeitar que cerca de metade dos alunos inquiridos fossem do sexo feminino e os restantes do sexo masculino, pois “*podem ter gostos diferentes, por exemplo os rapazes é mais coca e as raparigas talvez ice tea*” o que não reuniu consenso entre os colegas, pelo que um grupo apresenta como sugestão a escolha da amostra de modo a incluir alunos de diferentes faixas etárias porque por exemplo no caso dos alunos do secundário e do ensino noturno “*bebem café, os do básico não*”, o que reuniu consenso dado que tinha um elemento que alguns dos grupos tinha discutido quando ponderaram a representatividade da amostra. Apenas um grupo apresenta o modo como irá por em prática a seleção aleatória sugerem a escolha de alunos da fila que se encontram afastados entre si (Figura 76).



3. Escolher ao calhas os alunos a questionar. Uma fila do Bae.

Figura 75. Resposta de um grupo ao item 3 da tarefa 6.

3) Questionar alunos da fila afastados uns dos outros.

Figura 76. Resposta de um grupo ao item 3 da tarefa 6.

Quadro 38. Respostas apresentadas pelos alunos ao item 3 da tarefa 6.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Seleção aleatória da amostra	28	100,0

As respostas dos alunos ao item 4, identificação de um procedimento que conduza à escolha de uma amostra enviesada, complementam as respostas ao item 3, três grupos ponderaram o facto das preferências se poderem relacionar com a faixa etária dos alunos questionados (Figura 77) argumentando que “*alunos do 8º ano devem ter gostos diferentes dos mais velhos 11º / 12º ano como é o caso de beber café ou não*”. Três grupos refletiram sobre a questão da influência do conjunto de alunos da fila em que se insere o aluno questionado, explicitaram que se devia evitar escolher mais do que um aluno de um mesmo grupo de amigos na fila do bar pois “*vejo o A a comprar o sumo X e penso podia beber X em vez de sumo Y*”. Um outro sugere que se fosse desvalorizada a recolha da amostra no bar, fazendo, por exemplo, o questionário à porta da escola, onde também entram os alunos que compram bebidas na mercearia e não o bar, a informação recolhida não seria representativa dos consumos de bebidas no bar da escola (Figura 78). Durante a discussão os alunos ponderaram a questão da escolha de uma amostra apenas num turno, sendo que uma aluna defendeu: “*faz mais sentido que se beba mais leite com chocolate de manhã do que cola e de tarde é natural que seja ao contrário*” pois considerou que, pela manhã, os consumos de bebidas estão relacionados com a toma do pequeno-almoço ou seu reforço, daí a opção por produtos lácteos, enquanto de tarde poderão dar prioridade a outros tipo de bebidas, alguns alunos não concordaram, pois nem sempre ponderam esse aspeto no seu próprio consumo pessoal, no entanto acabaram por aceitar que globalmente no caso de se proceder ao inquérito exclusivamente nos intervalos da manhã, ou só da tarde, poderia ter influencia, pois a maioria das pessoas relaciona o tipo de consumo com a hora do dia.

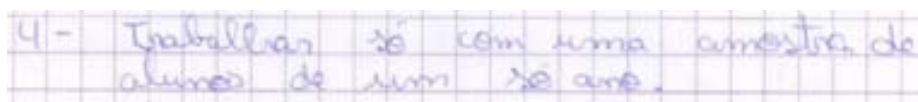


Figura 77. Resposta de um grupo ao item 4 da tarefa 6.

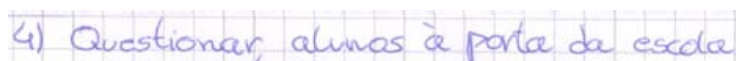


Figura 78. Resposta de um grupo ao item 4 da tarefa 6.

Quadro 39. Respostas apresentadas pelos alunos ao item 4 da tarefa 6.

Resposta	Nº de alunos (N= 28)	% de alunos
Não responde	0	0
Consumo de bebidas <i>versus</i> faixa idade	12	42,85
Consumo de bebidas <i>versus</i> influências do grupo da fila	12	42,85
Consumo de bebidas <i>versus</i> local de realização do questionário	4	14,3

Face a algumas questões apresentadas pelos pais/encarregados de educação o conselho de turma foi informado que deveria evitar a ocupação dos alunos nos seus tempos de estudo nos momentos de maior intensidade de avaliações escritas, pelo que optei por solicitar aos funcionários do bar que guardassem o maior número possível de talões das compras dos alunos em ambos os turnos, durante um período de sete dias em sete caixas respetivamente. Assim foi possível passar de imediato à recolha dos dados - parte II. Quanto à opção de escolher cerca de 40 talões de cada caixa por sorteio, alguns grupos optaram por recolher mais de 40 talões pois poderiam ter que inviabilizar a informação de algum que estivesse impercetível ou danificado. Para cada talão de compra os alunos constatavam se havia ou não compra de bebida e que tipo de bebida, o que permitiu o preenchimento das frequências absolutas da tabela do item 5. Com uma folha de cálculo determinaram-se rapidamente as frequências relativas (Figura 79).

Perante a projeção da tabela os alunos constataram que um número significativo de alunos não consome bebidas (47%) e entre os que consomem bebidas a favorita é o sumo Y (12,5%), a segunda bebida mais consumida é o sumo néctar (9,7%) seguida do leite com chocolate da marca W (8,7%). Um grupo de alunos sugeriu-se que se analisasse por tipo de bebida sendo que são os sumos os mais consumidos (27%) seguidos pelo consumo de água (pouco mais de 11%) e de leite com chocolate (pouco menos de 11%).

Nome da bebida	Grupo 1		Grupo 2		Grupo 3		Grupo 4	
	f. a.	f. r.	f. a.	f. r.	f. a.	f. r.	f. a.	f. r.
Água 0,33 cl	4	9,5%	1	2,5%	4	10,3%	2	5,0%
Água 0,5 l	5	11,9%	2	5,0%	3	7,7%	0	0,0%
Água 1,5 l	1	2,4%	0	0,0%	0	0,0%	0	0,0%
Café	1	2,4%	2	5,0%	0	0,0%	0	0,0%
Galão Máquina	0	0,0%	0	0,0%	0	0,0%	1	2,5%
Leite Choc Z	0	0,0%	3	7,5%	0	0,0%	0	0,0%
Leite Choc W	4	9,5%	3	7,5%	4	10,3%	3	7,5%
Sumo Néctar	0	0,0%	3	7,5%	7	17,9%	5	12,5%
Sumo X	7	16,7%	2	5,0%	1	2,6%	0	0,0%
Sumo Y	3	7,1%	6	15,0%	4	10,3%	9	22,5%
Iogurte Líquido	2	4,8%	0	0,0%	0	0,0%	0	0,0%
Total de consumos	25	54,3%	24	54,5%	23	65,7%	20	60,6%
Sem consumo de bebidas	17	37,0%	16	36,4%	16	45,7%	20	60,6%
Total	42	91,3%	40	90,9%	39	111,4%	40	121,2%

Nome da bebida	Grupo 5		Grupo 6		Grupo 7		Amostra turma	
	f. a.	f. r.	f. a.	f. r.	f. a.	f. r.	f. a.	f. r.
Água 0,33 cl	0	0,0%	1	2,3%	1	2,4%	13	4,5%
Água 0,5 l	3	7,0%	2	4,7%	3	7,3%	18	6,3%
Água 1,5 l	0	0,0%	0	0,0%	0	0,0%	1	0,3%
Café	3	7,0%	2	4,7%	1	2,4%	9	3,1%
Galão Máquina	0	0,0%	0	0,0%	0	0,0%	1	0,3%
Leite Choc Z	0	0,0%	2	4,7%	1	2,4%	6	2,1%
Leite Choc W	2	4,7%	3	7,0%	6	14,6%	25	8,7%
Sumo Néctar	6	14,0%	2	4,7%	5	12,2%	28	9,7%
Sumo X	2	4,7%	1	2,3%	1	2,4%	14	4,9%
Sumo Y	3	7,0%	8	18,6%	3	7,3%	36	12,5%
Iogurte Líquido	1	2,3%	0	0,0%	0	0,0%	3	1,0%
Total de consumos	20	43,5%	21	48,8%	21	51,2%	154	53,5%
Sem consumo de bebidas	23	50,0%	22	51,2%	20	48,8%	134	46,5%
Total	43	93,5%	43	100,0%	41	100,0%	288	100,0%

Figura 79. Preenchimento da tabela do item 5 da tarefa 6.

Os alunos experimentaram alguma frustração na resolução do item 6 tiveram dificuldade em arranjar uma estratégia de modo a projetar os valores apurados na sua amostra recolhida ao longo de sete dias para um consumo mensal pois anteciparam as seguintes limitações: a tendência de consumo de algumas bebidas ser sazonal “*as vezes há sumo de laranja e agora não há*”, muitos dos meses do ano letivo não existe consumo ou existe um consumo parcial, dada a existência de algumas interrupções escolares e decerto que qualquer recomendação que optassem por fazer estaria incorreta, pois exis-

tiam outras bebidas que não tinham ocorrido entre os consumos registados. Ainda assim a maioria dos grupos optou por considerar que os dados da amostra turma representa o consumo médio de 1 ou 2 dias e posteriormente multiplicam com 20 ou 10 as frequências absolutas de cada bebida no que se designou amostra turma, respetivamente de modo a “cobrir” um mês de consumos (Figura 80).

Nome da bebida	f. a.	Amostra como consumo × 20	
		médio de 2 dias	médio de 1 dia
Água 0,33 <i>cl</i>	13	130	260
Água 0,5 <i>l</i>	18	180	360
Água 1,5 <i>l</i>	1	10	20
Café	9	90	180
Galão Máquina	1	10	20
Leite Choc Z	6	60	120
Leite Choc W	25	250	500
Sumo Néctar	28	280	560
Sumo X	14	140	280
Sumo Y	36	360	720
Iogurte Líquido	3	30	60

Figura 80. Previsões apresentadas no item 6 da tarefa 6.

Nome da bebida	Média f.a. por excesso	1/4 do consumo diário × 20 dias	1/5 do consumo Diário × 20 dias
Água 0,33 <i>cl</i>	2	160	200
Água 0,5 <i>l</i>	3	240	300
Água 1,5 <i>l</i>	1	80	100
Café	2	160	200
Galão Máquina	1	80	100
Leite Choc Z	1	80	100
Leite Choc W	4	320	400
Sumo Néctar	4	320	400
Sumo X	2	160	200
Sumo Y	6	480	600
Iogurte Líquido	1	80	100

Figura 81. Previsões apresentadas no item 6 da tarefa 6.

Dois grupos consideraram a média das frequências absolutas de consumos como representativo de $\frac{1}{4}$ (os quatro intervalos mais representativos do dia de funcionamento do bar) e $\frac{1}{5}$ (inclusão do turno noturno cinco intervalos mais representativos do funcio-

namento do bar) do consumo diário pelo que por fim multiplicaram a média obtida por 4x 20 e 5x20 respetivamente (Figura 81).

A realização dos itens 7 e 8, foi feita em grupo turma com a análise conjunta das várias previsões e do consumo real, de bebidas no bar da escola, nos quatro primeiros meses do ano de 2011 (Figura 82).

A comparação entre as previsões e os dados reais corroborou as suspeitas da generalidade dos alunos, as suas previsões não tinham qualidade:

- Existem valores de consumo distintos de mês para mês e a existência de interrupções escolares afeta significativamente o consumo como constataram na comparação entre abril e os restantes três meses “*o bar não esteve sempre aberto por causa da páscoa*”;
- Existem consumos de bebidas que não foram previstas pois na amostra, não ocorreu o seu registo “*como não houve ninguém a beber leite [de pacote], nós nem pensamos nisso*”;
- Existem consumos de bebidas sazonais – sumo de laranja natural, uma vez que o bar tem esta opção quando a laranja é mais abundante; diminuição do consumo de sumo néctar *versus* aumento do consumo de sumos X e Y (refrigerantes); e aumento do consumo de água, essencialmente pela escolha de garrafas de maior capacidade “*vende-se cada vez mais garrafas de litro e meio*”.

Ainda questionei os alunos de que modo, poderíamos melhorar a qualidade da previsão, não havendo sugestões, questionei se seria um bom procedimento tomar como referencia para a previsão de determinado mês a informação de consumo relativo a esse mês registada nos 4/5 últimos anos, a generalidade dos alunos concordou que tal procedimento seria preferível ao nosso estudo no entanto ainda assim segundo uma aluna a previsão poderia ainda assim ao não ter em consideração novas tendências de consumo ser de má qualidade “*no ano passado apareceu o sumo Y e o bar não tinha este ano é o que mais bebemos*”, face ao argumento os alunos concordaram que a introdução de um produto que passe a ser “moda” tem efeitos no consumo dos restante produtos similares. Constataram que a tarefa de prever consumos com base em dados por mais precisos que sejam estará sempre associada a um elevado grau de incerteza. Acrescentei que existem técnicas específicas para o estudo de mercado, sendo feitos inúmeros estudos do impacto de um produto face ao seus similares e que permite reduzir um pouco mais a incerteza associada à previsão.

Nome da bebida	Previsão 1	Previsão 2	Previsão 3	Previsão 4
	C. médio de 1 dia × 20 dias	C. médio de 2 dias × 20 dias	1/4 do consumo diário × 20 dias	1/5 do consumo Diário × 20 dias
Água 0,33 <i>cl</i>	130	260	160	200
Água 0,5 <i>l</i>	180	360	240	300
Água 1,5 <i>l</i>	10	20	80	100
Café	90	180	160	200
Galão Máquina	10	20	80	100
Leite Choc Z	60	120	80	100
Leite Choc W	250	500	320	400
Sumo Néctar	280	560	320	400
Sumo X	140	280	160	200
Sumo Y	360	720	480	600
Iogurte Líquido	30	60	80	100

Nome da bebida	Amostra turma		Jan-11	Fev-11	Mar-11	Abr-11	Méd mensal arred	
	f. a.	f. r.	f. a.	f. a.	f. a.	f. a.	f. a.	f. r.
Água 0,33 <i>cl</i>	13	8,4%	473	287	243	279	321	11,3%
Água 0,5 <i>l</i>	18	11,7%	33	219	288	246	197	6,9%
Água 1,5 <i>l</i>	1	0,6%	23	40	64	73	50	1,8%
Café	9	5,8%	115	101	117	58	98	3,5%
Café descafeinado	-	-	11	6	22	4	11	0,4%
Galão Máquina	1	0,6%	58	36	41	12	37	1,3%
Garoto	-	0,0%	22	7	7	1	9	0,3%
Leite Choc Z	6	3,9%	36	63	72	37	52	1,8%
Leite Choc W	25	16,2%	765	732	683	299	620	21,9%
Pacote de leite	-	-	111	21	0	2	34	1,2%
Copo de leite	-	-	16	11	21	16	16	0,6%
Chá	-	-	14	25	24	14	19	0,7%
Sumo Néctar	28	18,2%	608	486	454	259	452	15,9%
Sumo X	14	9,1%	270	376	375	274	324	11,4%
Sumo Y	36	23,4%	704	633	665	299	575	20,3%
Copo de sumo de laranja	-	-	43	7	0	0	13	0,5%
Iogurte Líquido	3	1,9%	6	16	5	3	8	0,3%
Total de consumos de bebidas	154	100,0%	3308	3066	3081	1876	2836	100,0%

Figura 82. Previsões versus dados reais correspondentes aos últimos 4 meses - item 7 da tarefa 6.

Análise e interpretação. Com a realização da tarefa os alunos puderam envolver-se uma vez mais num estudo estatístico, sendo que a situação em causa tinha características comuns com os habituais estudos de mercado. Perante a questão de investigação, os alunos continuam a revelar dificuldade na formulação de questões de pesquisa de modo que os dados recolhidos permitam responder à questão de investigação. Os alunos envolveram-se na discussão e ponderação de elementos que afetam a representatividade de uma amostra representativa para a situação em estudo, tendo privilegiado um processo de seleção aleatória simples, pois sendo necessário aceder aos alunos utilizadores do bar da escola, o habitual procedimento de seleção aleatória por extratos não se adequava. Perante o estudo foi possível uma vez mais, o reconhecimento da variabilidade no processo de amostragem e o reconhecimento de comportamentos que se poderão inferir para a população (Rubin, Bruce & Tenney, 1991), as bebidas que reuniram uma maior frequência na amostra turma teriam que ter necessariamente uma recomendação de encomendas em maior quantidade, enquanto para as bebidas com menor frequência seria recomendada uma menor quantidade a encomendar mensalmente. Todos os grupos optaram por um raciocínio proporcional para construir a sua previsão. A atividade desenvolvida em torno da tarefa revelou que a generalidade dos alunos recorreu ao seu conhecimento do contexto e foi sensível à variabilidade pelo que sentiram que a previsão produzida com a informação disponível seria necessariamente de má qualidade. Perante a questão sobre se reformular o processo de recolha de dados serviria para a análise de períodos homólogos num conjunto de 4 a 5 anos, a discussão em torno da introdução de novos produtos e os seus efeitos no consumo de produtos similares, reforçou a incerteza associada a construção de previsões e a importância de existirem métodos estatísticos que apoiem as tomadas de decisão (Martins& Ponte, 2010).

Balanço. Os alunos participaram com agrado na realização da tarefa, tendo sido adequado o tempo previsto para a sua resolução. O facto de os alunos conhecerem o contexto envolvido na situação teve como consequência que a generalidade dos grupos não acreditava ser possível produzir uma “boa” recomendação do número de unidades de cada uma das bebidas, não colocaram em causa a dimensão da amostra turma. Consideraram que o grau de incerteza associado à previsão solicitada se devia a múltiplos fatores como o consumo sazonal de algumas bebidas, tempo de inatividade do bar da escola e alterações da tendência de consumo face à introdução de novos produtos.

5.3. Desempenho dos alunos nos testes

Neste ponto apresento uma análise comparativa dos resultados apresentados (Anexo 16), pelos 27 alunos que realizaram o teste inicial (Anexo 3) e os 28 alunos que realizaram o teste final (Anexo 15), no que diz respeito às capacidades de planeamento estatístico - formulação de questões, população e amostra, amostra representativas e elementos a ponderar, e de análise de dados - representações, medidas estatísticas e sua utilização na interpretação dos dados, nomeadamente na comparação de distribuições e na transformação de uma representação numa outra.

5.3.1. Planeamento estatístico

Formular questões. Item 3.2 (teste final), partir da questão de investigação “estudo sobre os hábitos de sono de jovens entre os 12 e os 16 anos” um grupo significativo de alunos (71%) conseguiu formular pelo menos duas questões de pesquisa que permitissem a recolha de dados adequados ao problema em estudo, sendo que 43% dos alunos apresenta um conjunto de três questões de pesquisa adequado ao objeto em estudo. As respostas mais habituais foram a apresentação de pelo menos duas das questões seguintes: “A que horas, costumás ir para a cama/ te deitas?” (75%), “Quantas horas dormes? (em média...)” (64%) e “A que horas costumás acordar / habitualmente acordas?” (39%), em discussão com os alunos, foi considerado que os inquiridos deveriam ser informados que o entendimento das expressões “costumas” ou “habitualmente” se referia ao comportamento mais frequente em período escolar. Na maioria das vezes, a terceira questão formulada tinha uma interpretação subjetiva, o que iria comprometer os dados recolhidos, como por exemplo “Costumas deitar-te muito tarde?”, “Estás habituado desde criança a deitar-te cedo?”, “Achas que dormes o suficiente por dia?”, “Quando acordas, sentes-te cansado?”. Três alunos (11%) optaram por inquirir os jovens quanto à qualidade do sono, formulando uma das seguintes questões “*Acordas durante a noite e ficas com insónias?*” ou “*Acordas durante a noite com pesadelos?*”. Como não surgiram itens relativos a este objetivo no teste inicial, opto por proceder à comparação com os desempenhos dos alunos na realização do item 6b) da tarefa 1 (71% dos alunos não formula qualquer questão e 29% dos alunos formulam questões desadequadas ao objeto de estudo), pelo que o desempenho global dos alunos em termos individuais foi superior ao desempenho em pequeno grupo.

Distinguir entre população e amostra. Item 5.1 (teste final), mais de metade dos alunos (64%) foi capaz de identificar a população como sendo o conjunto dos 1560 alunos da escola em estudo (admitindo que todos frequentam os serviços com a mesma assiduidade), um dos alunos especifica que a população em estudo é o conjunto de alunos que recorrem aos serviços em estudo. O erro mais frequente entre 32% dos alunos foi tratar a amostra como população. No item 6.1 (teste final) a generalidade dos alunos (82%) deu como resposta “A população são as maçãs de Alcobaça” não sendo claro se se referem ao conjunto de todas as maçãs produzidas na região de Alcobaça, se apenas as que foram entregues na cooperativa no dia em estudo, ou se, se trata apenas do conjunto de maçãs escolhidas para analisar (amostra), posteriormente, ao inquirir os alunos na entrega dos testes, referiram trata-se do conjunto de todas as maçãs produzidas na região e processadas pela cooperativa agrícola de Alcobaça. A troca entre os conceitos população e amostra apenas fica clara na resposta de uma aluna “*São as maçãs escolhidas ao acaso entre as que foram entregues na cooperativa agrícola de Alcobaça num determinado dia*”. O erro mais recorrente foi associar o conceito de população à variável em estudo “calibre” cometido por três alunos (11%). No item 6.2 (teste final), um número significativo de alunos (71%) identifica corretamente a dimensão da amostra - 51 maçãs, os erros mais frequentes foram: i) tomar como dimensão o resultado da soma dos vários calibres (14%) e ii) descrever por palavras suas o processo de recolha da amostra considerada (11%).

Amostras representativas e ponderação de elementos que afetam a representatividade. Face ao item 5.1 (teste inicial) tendo como objetivo o reconhecimento de amostras representativas, os desempenhos dos alunos foram bastante positivos, cerca de metade (48%) indicou como opção válida uma e uma só das seguintes alternativas ou as duas alternativas – “Questionar 160 alunos escolhidos por sorteio entre os 1560” ou “Questionar os alunos nº 1,8, 15, 22 de cada uma das quarenta turmas da escola”. Apenas 14% dos alunos optou por selecionar opções que conduziam a enviesamento da amostra, indicando como escolha ou uma das suas escolhas: “Questionar 160 alunos das seis turmas do 12º ano da escola” ou “Questionar 160 alunos que sejam amigos/ conhecidos dos alunos responsáveis pelo estudo.” No item 3.1 (teste final), tendo como objetivo a identificação de um procedimento que implique a escolha de uma amostra enviesada, cerca de três quartos dos alunos (71%) indicou um procedimento que envolveu a ponderação de um ou mais elementos que afetasse a representatividade, nomeadamente a seleção de elementos exclusivamente de uma determinada idade, e/ou género ou de

uma determinada região geográfica (grandes centros urbanos *versus* pequenos aglomerados populacionais rurais). Dois alunos (7%) optaram por indicar a escolha de amostras de pequena dimensão como fator de falta de representatividade. Quanto ao item 5.2 (teste final) mais de três quartos dos alunos (82%) indicou com procedimento a escolha aleatória dos elementos da amostra. Mais de um terço dos alunos (36%) optou por uma seleção aleatória simples, sendo que metade procedia à recolha dos dados junto dos alunos que frequentam os serviços em estudo, e metade optava por recolher os dados entre a população escola, pós sorteio; dois alunos (7%) consideraram um procedimento de seleção estratificada (grupos turma). Mais de um terço dos alunos (39%) embora indique a escolha aleatória dos elementos inclui a condição que metade dos elementos sejam do género feminino, o que poderia inviabilizar a representatividade da amostra caso esta não seja a proporção do género feminino na população escolar.

Em ambos os testes os contextos das situações eram familiares aos alunos, sendo que, em contexto escolar, o desempenho dos alunos é mais positivo. De um modo geral, a turma melhorou o seu desempenho neste tipo de itens, privilegiando a seleção aleatória de uma amostra e argumentando o papel de distintos elementos que poderão afetar a representatividade de uma amostra.

5.3.2. Análise de dados

Construir, analisar e interpretar representações dos dados

Representação tabelar. Numa fase inicial os alunos, mostram-se capazes de trabalhar com as representações tabelares (item 1.1 do teste inicial), todos os alunos determinam o valor em falta na coluna “nº de alunos” mas cerca de um quarto dos alunos não procede ao completamento da coluna “percentagem de alunos”, pelo que apenas 56% dos alunos completa corretamente a tabela de frequências. Um dos alunos (4%) demonstra claramente não ter interpretado a representação tabelar ao não reconhecer os valores da coluna idades como sendo os valores que a variável em estudo toma. No item 1.1 (teste final) 86% dos alunos determina todos os valores em falta, quer na coluna das frequências absolutas, quer na das frequências relativas em percentagem. Dois alunos (7%) iniciam a resolução tendo determinada um dos quatro valores em falta. Dois alunos (7%) não respondem à questão. Os alunos melhoraram substancialmente o seu desempenho na construção e interpretação de representações tabelares.

Histograma. Quanto à construção, análise e interpretação do histograma (grupo 2 do teste inicial e teste final), no item 2.1 que pretendia aferir a identificação de algumas das características do histograma, nomeadamente número de classe e amplitude das classes, os alunos melhoraram o seu desempenho, pois inicialmente 37% dos alunos não respondeu e 48% identifica as duas características, no teste final apenas 4% não responde sendo que 74% identifica corretamente ambas as características. Os itens 2.2, 2.3 e 2.4 que tinham como objetivo o reconhecimento de alguns aspetos da representação e mobilização do conhecimento para uma leitura literal da representação, os desempenhos dos alunos evoluíram do seguinte modo: item 2.2 (número total de efetivos) inicialmente 74% responde corretamente depois 93%; item 2.3.1 (número de efetivos com um tempo maior ou igual a x) com uma taxa de resposta correta de 57%, passou para 76%, item 2.3.2 (número de efetivos com um tempo inferior a x) com uma taxa de resposta correta de 74%, passou para 93%; item 2.4 (indicar valor possível do vencedor) inicialmente 33% dos alunos apresenta um valor correto, 11% indica a classe a que o tempo vencedor pertence, no teste final 50% dos alunos responde corretamente e 4% indica a classe a que o tempo vencedor pertence. O erro mais comum entre os alunos foi escolher para tempo vencedor o limite superior da classe a que o tempo vencedor pertence, apresentando como resposta 4 minutos e 30 segundos, que corresponde a 4,5 minutos, o erro cometido pode ser consequência de uma incorreta utilização da unidade tempo associando 4,3 minutos a 4 minutos e 30 segundos.

Diagrama de caule e folhas. Relativamente à leitura e interpretação do diagrama de caule-e-folhas, nos item 4.1 (teste inicial) e item 6.2 (teste final) solicitava-se a identificação do número total de efetivos apresentados na representação, sendo a questão formulada no âmbito do contexto da situação (número de pessoas que responderam à questão 4 do inquérito e dimensão da população, respetivamente no teste inicial e final), no primeiro item, 67% dos alunos responderam corretamente e no segundo item 71% dos alunos responderam corretamente, não sendo nítido uma melhoria do desempenho, o que se pode ter ficado a dever ao contexto em que a solicitação foi feita pois, no item do teste final, os erros mais frequentes foram: i) tomar o número total de efetivos (dimensão) como o resultado da soma dos vários calibres (14%) e ii) descrever por palavras suas o processo de recolha da amostra considerada (11%). Os itens 4.2 (teste inicial) e 6.3 (teste final) solicitavam a leitura dos extremos, inicialmente 70% dos alunos respondeu corretamente e no teste final a percentagem de respostas corretas subiu para 82%. O erro mais comum entre os alunos que responderam incorretamente (19%)

no teste inicial, foi considerarem o valor mínimo e máximo de contactos respetivamente 0 (folha de menor valor) e 9 (folha de maior valor) mínimo 0. No teste final o erro mais comum entre os alunos que erraram (14%) foi afirmarem que os calibres variavam entre 50 e 70 mm. Denota-se que houve uma melhoria na leitura e interpretação da representação caule| folha. Os alunos tiveram um bom desempenho, na situação apresentada no teste inicial, e no teste final conseguiram melhorar o seu desempenho.

Determinação de medidas estatísticas e sua utilização na interpretação dos dados

Moda. No item 1.2 (teste inicial) é notório que neste conjunto de alunos a generalidade reconhece corretamente (93%), no âmbito da situação, o valor da moda, não sendo de constatar erros tipicamente associados a este conceito, como por exemplo, indicar o valor da frequência absoluta correspondente ao dado mais observado. No item 1.3 (teste final) o desempenho continuou a ser muito bom, com 96% a identificar corretamente a moda, e apenas um aluno a não responder ao item. Possivelmente o bom desempenho e a inexistência de equívocos decorre da questão ter sido formulada sem a utilização do termo “moda” mas sim “Qual a idade/o número de faltas mais usual no conjunto dos alunos em análise?”.

Mediana. No item 4.3 (teste inicial) cerca de dois terços dos alunos mostra-se capaz de indica o valor da mediana, e no item 4.4 (teste final) cerca de três quartos procedeu à correta identificação do valor da mediana. Naturalmente que, o facto de os dados estarem ordenados e os alunos disporem de uma visualização quase individual dos dados, facilitou a sua determinação. No teste inicial os alunos revelaram desempenho satisfatório e melhoram-no no teste final.

Comparar distribuições e validar conclusões. No âmbito da comparação de duas distribuições recorrendo à análise conjunta de medidas estatísticas centrais e de dispersão, no item 3 (teste inicial), cerca de metade dos alunos (48%) revelou-se capaz de encontrar, pelo menos, um argumento válido na defesa do grupo A, um grupo significativo de alunos (15%) apresentou mais do que um argumento válido e 37% dos alunos não respondeu ao item. Registe-se que uma aluna procedeu apenas há comparação das frequências absolutas ao indicar como argumento “há mais negativa em B do que em A” não tendo sentido necessidade de referir que a comparação era viável, dado o facto de os conjuntos terem o mesmo número de elementos. Por outro lado, penso que o facto de se tratar de um histograma comparativo e de os dados estarem agrupados em classes 0-9,

10-19, ..., em vez das típicas classes $[0, 10[$, $[10, 20[$, ... influenciou a leitura da representação pelo que houve um número significativo de alunos a referir "... um dos membros do grupo A teve 0% é por que não fez o teste...". No item 4 (teste final) 61% dos alunos mostrou-se capaz de apresentar um argumento para o facto de a conclusão do empresário não ser válida, por se basear exclusivamente na média. De referir que 18% dos alunos interpretam os valores médios das duas distribuições face à dispersão do conjunto dos dados de menor média e/ou à concentração dos dados em torno da média de maior valor, e acrescentaram, tal como os restantes colegas que responderam correctamente, que a percentagem de resultados elevadas (acima de 105 pontos) era significativamente superior à do grupo com maior média, pelo que o empresário ao pretender pessoas com melhor desempenho deveria escolher as melhores do grupo com menor média global. Os alunos tiveram um desempenho satisfatório, na situação simples apresentada no teste inicial. No teste final conseguiram um melhor desempenho, numa situação mais complexa, pois tratava-se de uma conclusão que fora generalizada para a população com base numa amostra representativa.

Transformar uma representação numa outra representação

Diagrama de caule e folhas - diagrama de extremos e quartis. Item 4.4 (teste inicial) e item 6.4 (teste final), a generalidade dos alunos que procede a uma correta leitura e interpretação do diagrama de caule-e-folhas mostra-se capaz de o transformar no respetivo diagrama de extremos e quartis, por vezes cometendo um erro processual na determinação dos quartis. No item 4.4, 44% dos alunos procedeu à correta determinação dos quartis face à mediana e posterior representação do diagrama, 19% dos alunos determina de modo incorreto os quartis e perante os valores determinadas apresenta o diagrama de extremo e quartis correspondente, correctamente. 37% dos alunos não respondeu. No item 6.4., 71% dos alunos determina correctamente os quartis e procede à correta representação dos cinco valores de resumo, 18% dos alunos comete erros processuais na determinação do 1.º e 3.º quartil mas procede à correta representação do diagrama de extremos e quartis e a percentagem de alunos que não responde diminui para 18%. Depois da unidade de ensino os alunos evidenciam uma melhoria de desempenho no âmbito da construção de um diagrama de extremos e quartis.

Após a unidade de ensino, o desempenho dos alunos no âmbito da construção, análise e interpretação de diferentes representações de dados, melhorou, pois nos vários

itens realizados, diminuiu, não só a percentagem de alunos que responde com incorreções, como a percentagem de alunos que não respondeu.

Representação tabular – gráfico de barras. No item 1.2, que solicitava aos alunos a interpretação da escala de construção do gráfico de barras (teste final), 61% dos alunos mostrou-se capaz de descobrir a unidade da escala utilizada na construção do gráfico de barras, 18% dos alunos compreenderam a questão, no entanto, não se tendo apercebido que o eixo vertical correspondia à percentagem de alunos e não ao número de alunos, encontraram um valor errado para a unidade; os restantes alunos (21%) não responderam.

Histograma – diagrama de extremos e quartis. Os item 5.2. (teste inicial) e item 5.3 (teste final) têm como objetivo aferir a capacidade de os alunos procederem à análise da informação contida num histograma e a transformarem no correspondente diagrama de extremos e quartis, apelando ao reconhecimento das características gerais de uma distribuição (forma, centro e dispersão da distribuição). No item 5.2. (teste inicial) cerca de um quarto dos alunos (26%) mostra-se capaz de reconhecer características gerais, como a forma, centro e dispersão expressas na representação inicial da distribuição e converter tal informação na respetiva localização das cinco das medidas estatísticas – mínimo, quartis e máximo. Houve ainda 15% dos alunos capazes de reconhecer globalmente o enviesamento à esquerda, indicando uma concentração mais acentuada à direita da mediana do que à sua esquerda, no entanto, na localização do 1º e 3º quartil, não há evidência da análise da concentração dos dados entre a mediana e o 3º quartil, face aos dados compreendidos entre o 1º quartil e a mediana. O item 5.3. (teste final) ao solicitar a associação do respetivo diagrama de extremos e quartis, a distribuições com evidente enviesamento à direita, à esquerda, e quase simétrica, entre um conjunto de cinco possíveis diagramas de extremos, constitui uma situação mais complexa. Ainda assim 36% dos alunos mostrou-se capaz de estabelecer as devidas associações, 11% dos alunos estabeleceram corretamente duas das associações, e 28% dos alunos só estabelece corretamente a associação da distribuição quase simetria. O erro mais comum entre os alunos (14%) foi emparelhar o histograma da distribuição enviesada à esquerda com o diagrama de extremos e quartis enviesado à direita e o histograma da distribuição enviesada à direita com o diagrama de extremos e quartis enviesado à esquerda, o que corresponde à dificuldade mobilizar o conhecimento de que a densidade relativa dos dados se relaciona de modo inverso com o comprimento da componente do diagrama. Após a unidade de ensino, o desempenho dos alunos da turma embora tenha continuado

a não ser satisfatório, melhorou consideravelmente, considerando o grau de complexidade do item do teste final.

Generalizar conclusões para a população

Perante o item 4.5 (teste inicial) tendo como objetivo o reconhecimento de conclusões passíveis de serem generalizadas para a população em estudo, com base em amostras, 41% dos alunos optou pela generalização da interpretação da mediana no contexto em estudo, para o conjunto de todos os alunos do 8.º ano de escolaridade, pois embora a amostra fosse composta apenas por 5 alunos de cada turma de 8.º ano, estes tinham sido escolhidos ao acaso. Por sua vez, 36% dos alunos optou por selecionar a opção de generalizar a proporção de alunos que não usam o Messenger, para o conjunto de todos os alunos da escola, com base em convicções pessoais “*quase todos usam a internet como meio de comunicação*”. No item 1.4 (teste final) 39% dos alunos não responde e 21% dos alunos interpreta erradamente a questão. Face à decisão tomada pela direção de encomendar certificados com base nos dados da amostra, 14% dos alunos considera-a incorreta pois sentem que a amostra pode não refletir necessariamente o nível de assiduidade da população, dada a variabilidade da assiduidade entre os alunos. Os restantes alunos (21%) aceitam a decisão com algum grau de incerteza, pois embora se trate de uma amostra representativa, existe variabilidade na assiduidade dos alunos “uma boa amostra, pois para cada turma temos alunos suficientes para a amostra, mas nunca se sabe se vai ser exatamente $x\%$ ”. Itens 6.5 e 6.6 (teste final), identificação de duas conclusões sobre o estudo, um quarto dos alunos não responde aos itens e 14% tece considerações de natureza não matemática sobre o estudo. Entre os alunos que indicam possíveis considerações sobre o estudo, 25% fez referência ao centro e à dispersão, identificando a mediana e a amplitude, 14% refere a dispersão da amostra (amplitude) e indica o valor dos extremos, 14% refere a comparação da concentração e/ou dispersão de dois segmentos do diagrama de extremos e quartis e 7% refere o centro da distribuição recorrendo à indicação da moda e da mediana. As considerações feitas com base na amostra são passíveis de extrapolar para a população, para 46% dos alunos sendo que apenas uma pequena parte dos alunos (18%) fundamenta que a generalização é viável porque a amostra é representativa, 18% dos alunos optou por não generalizar as suas considerações pois considera que embora a amostra seja representativa da produção de um determinado dia, pode ter particularidades específicas que não

correspondem à população dada a sua pequena dimensão, os restantes alunos recorreu a argumentos não matemáticos. De um modo geral, a turma melhorou o seu desempenho neste tipo de itens. Entre a situação apresentada no teste inicial e as situações do teste final, existe um maior grau de abstração, uma parte significativa de alunos redigiu duas considerações respeitantes à amostra (item 6.5) e refletiu sobre a sua extrapolação à população, apresentando argumentos válidos para a sua tomada de decisão. O reconhecimento da variabilidade das características em estudo condiciona os alunos quanto à extrapolação das suas considerações à população da qual se extraiu uma amostra representativa.

5.3.3. Síntese

De um modo geral, os dados permitem inferir sobre a evolução dos alunos em relação à capacidade de análise de dados e verificar as aprendizagens realizadas durante a unidade de ensino no que se refere à capacidade de planeamento estatístico. Globalmente, a turma apresentou um bom desempenho no teste final, tendo em consideração o grau de complexidade de alguns itens e a necessidade de interpretação de um número significativo de textos informativos que introduziram as situações em cada um dos grupos de itens. Um outro ponto positivo do desempenho dos alunos, entre o teste inicial e o teste final, foi a diminuição da percentagem de alunos que não respondeu aos itens.

Relativamente às aprendizagens no âmbito do planeamento estatístico, os alunos evidenciaram um bom desempenho. No entanto a formulação de questões manteve-se um desafio para alguns alunos. Na verdade, num conjunto de três questões de pesquisa solicitadas, 25% dos alunos apresenta apenas uma questão, conseguindo uma formulação clara e interpretação única.

Quanto à seleção de uma amostra representativa, a maioria dos alunos prefere um processo de amostragem aleatório por extratos (45%) em detrimento de um processo aleatório simples, tal como os alunos de 5.º ano do estudo de Jacobs (1999). No presente estudo a escolha de uma amostra estratificada pode ser consequência do facto de se pretender estudar uma característica na população escolar que tem uma organização vertical, a cada ano de escolaridade por regra geral corresponde à uma idade específica, dada a reduzida retenção no sistema educativo e uma organização horizontal, a cada turma corresponde por regra geral uma cultura de grupo, dada a manutenção de generalidade dos alunos no mesmo grupo turma ao longo dos vários anos de escolaridade.

Em relação à ponderação de elementos que podem afetar a representatividade de uma amostra em relação à respetiva população, os alunos evidenciam ter em consideração vários e distintos elementos, embora os mais habituais sejam a restrição da amostra a uma faixa etária específica (amostra apenas com adolescentes mais jovens ou só mais velhos ou numa proporção necessariamente incorreta face à população, como por exemplo 20% de jovens com 12 anos e 80% de jovens com 16 anos) e a restrição a um género ou a combinação de ambas as restrições. Surgiram outros elementos na ponderação da representatividade, nomeadamente dimensão da amostra, com a escolha de amostras com menos de 15 elementos da população, restringir a seleção da amostra a uma área geográfica do país (exclusivamente de um centro urbano), restringir a seleção da amostra a grupos específicos de jovens com condicionamentos em termos dos hábitos de sono – jovens de um colégio interno, jovens que frequentem o instituto do sono, jovens de uma determinada escola cujo turno predominante seja manhã, etc.. A existência de um grande leque de elementos ponderados demonstra a elevada sensibilidade à variabilidade da característica em estudo, por parte da maioria dos alunos, capacidade essencial para o reconhecimento de situações de interesse para a estatística, para ser crítico das inúmeras informações estatísticas veiculadas pelos meios de informação e das decisões tomadas face a resultados de estudos de natureza estatística.

Quanto à distinção entre os conceitos de população e de amostra, os alunos ainda sentem dificuldades, sendo que 36% dos alunos identificou a amostra como população, o que corresponde aos resultados de outros estudos (Fernandes, Carvalho e Ribeiro, 2007; Novaes, 2004).

Quanto à capacidade de análise de dados, ao nível das representações os alunos melhoram o seu desempenho em todos os aspetos apreciados, sendo de salientar o progresso no âmbito da interpretação e construção de uma representação tabelar e do diagrama de extremos e quartis. A realçar que cerca de um quinto dos alunos embora cometa erros processuais na determinação do 1.º e 3.º quartil, completou a construção do diagrama de extremos e quartis face ao seu respetivo conjunto de resumo de cinco números. Ao nível da interpretação de um diagrama de extremos e quartis mais de metade dos alunos recorreu a uma medida de dispersão (a maioria optou por referir a amplitude ou o intervalo dos extremos) ou referiu a concentração dos dados centrais face à outra metade de dados que se encontra nas caudas ou referiu a concentração dos dados abaixo da mediana face à metade de dados que se encontra acima da mediana. Sendo que um quarto dos alunos opta por redigir considerações recorrendo quer a medi-

das centrais, quer a medidas de dispersão. Para metade dos alunos a interpretação de um diagrama de extremos e quartis ainda constitui um problema, pois ou se mostram incapazes de tecer qualquer comentário sobre a representação ou recorreram exclusivamente a medidas de tendência central, o que corresponde os resultados do estudo de Bakker, Biehler e Konold (2005).

No que toca às medidas de tendência central, na identificação da moda e na determinação da mediana, os alunos da turma não cometeram a maioria dos erros referidos na revisão de literatura, nomeadamente nenhum dos alunos identificou a moda como sendo o valor da frequência absoluta mais elevado e nenhum aluno considerou que a mediana era o dado mais comum. Em ambos os testes, se solicitou a determinação da mediana para uma distribuição de dados apresentada num diagrama de caule e folhas, pelo que os alunos os alunos não sentiram necessidade de ordenar os dados e interpretar a frequência de cada dado para a ordenação. O à-vontade que os alunos demonstram na utilização do diagrama de caule e folhas, na determinação da mediana, é um sinal claro de um bom desempenho ao nível de compreensão e interpretação do diagrama e caule e folhas. Quanto à média, um dos equívocos habituais segundo Garfield e Chance (2000) é a comparação de dois ou mais grupos de dados assente exclusivamente na análise da diferença das respetivas médias. Assim em ambos os testes se apresentou uma situação em que tal equívoco foi cometido e solicitou-se a apresentação de pelo menos uma aspeto que invalidasse a conclusão assente exclusivamente na análise da diferença das respetivas médias. No teste inicial cerca de metade dos alunos constatam que existe pelo menos um elemento de análise, que refuta a conclusão extraída. No teste final, embora a situação tenha sido mais complexa e tenha requerido uma maior capacidade de interpretação, o desempenho dos alunos melhorou, visto o aumento da percentagem de alunos que indica um elemento de análise e da percentagem de alunos que indica dois elementos de análise para rebater a conclusão deduzida.

Ainda ao nível do raciocínio estatístico sobre representações, mais de metade dos alunos mostrou-se capaz de reconhecer características gerais, como a forma, o centro e a dispersão da distribuição, isto é ter uma visão do conjunto de dados, um dos três aspetos do raciocínio sobre representações segundo Garfield (2002), ao estabelecer corretas associações entre o histograma e com o respetivo diagrama de extremos e quartis, para uma distribuição simétrica em U e duas distribuições assimétricas com enviesamento à esquerda e a outra à direita. A capacidade de comparar dois tipos de representa-

ção para uma mesma distribuição, o que segundo Martins e Ponte (2010) promove o desenvolvimento de uma atitude crítica relativa à utilização das representações.

Retirar conclusões a partir dos dados e indicar a confiança que se atribui a tais conclusões é uma fase indispensável na aprendizagem da Estatística, no caso dos alunos da turma em análise, no teste final, cerca dois terços dos alunos foi capaz de extrair, pelo menos, duas considerações, sendo que dois quintos dos alunos recorre a, pelo menos, uma medida central e a uma medida de dispersão. Quanto à validade das conclusões, o desempenho dos alunos ainda fica aquém do desejado, dependendo da situação, um quinto a metade dos alunos aceitam que face a amostras representativas e de dimensão razoável, as conclusões considerada sejam extrapoladas à população conscientes de que deste modo estão a aceitar/tomar decisões com algum grau de incerteza. Cerca de um quinto dos alunos independentemente da situação, e embora reconheçam que a amostra seja representativa, no sentido de ter “boa” qualidade, não aceitam a extrapolação das conclusões pois a variabilidade da amostra estará aquém da variabilidade da população ou a variabilidade da amostra em análise poderá ser distinta da de uma outra amostra. Para estes alunos uma amostra deve ser uma porção significativa da população para que se possa proceder a extrapolações, tal como nos estudos de Rubin, Bruce e Tenney (1991) e Metz (1999), a variabilidade do processo de amostragem e a variabilidade da população que poderá não ser expressa em toda a sua amplitude numa amostra representativa induz os alunos a assumir que uma amostra não nos fornece informação útil sobre uma população.

O desempenho dos alunos, em ambos os testes, em itens como 3, 4.4, 4.5 e 5.2 (teste inicial) e 4, 5.3, 6.4, e 6.5, permite concluir que um grupo significativo de alunos, evidencia mobilizar dois dos três aspetos que a investigadora Garfield (2002) considera fundamentais ao nível do raciocínio estatístico no âmbito das medidas estatísticas, que são (i) compreender por que razão as medidas centrais, de dispersão, nomeadamente as de ordem, e a sua respetiva localização, fornecem informação sobre um conjunto de dados, e (ii) saber que um resumo descritivo dos dados deve incluir uma medida central, bem como uma medida de dispersão.

Em itens que envolveram uma maior necessidade de interpretação de textos informativos, a elaboração de um comentário e/ou a justificação de um raciocínio, o desempenho dos alunos também reflete algumas dificuldades ao nível da expressão escrita.

Em termos globais, a turma apresentou um desempenho satisfatório no teste final, tendo evoluído positivamente do teste inicial para o teste final. Os gráficos apresentados (Figuras 83, 84 e 85), referem-se à média de respostas corretas em cada aspeto analisado nos testes.

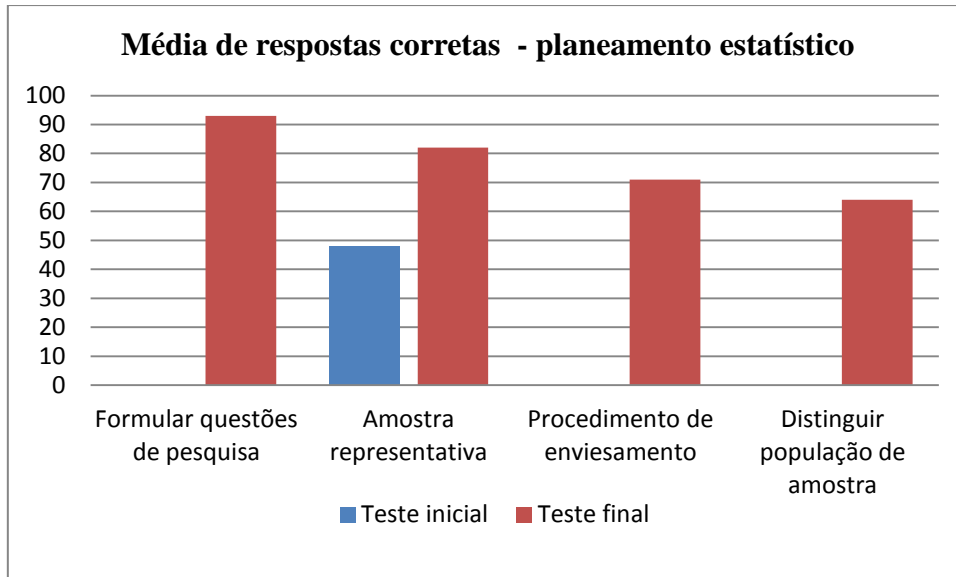


Figura 83. Comparação dos desempenhos no âmbito do planeamento estatístico

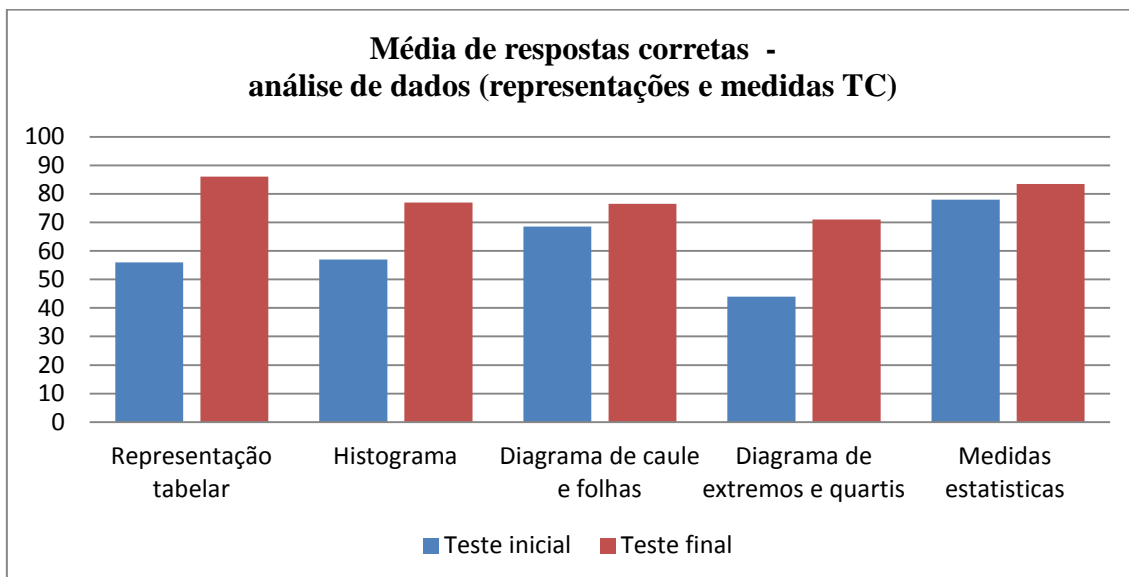


Figura 84. Comparação dos desempenhos no âmbito da construção e interpretação de representações e cálculo de medidas estatísticas de tendência central.

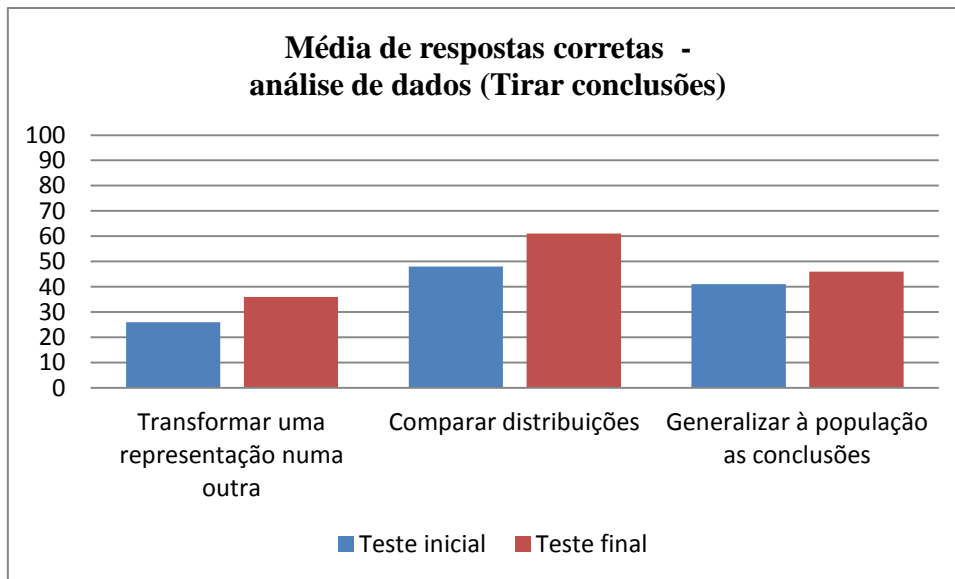


Figura 85. Comparação dos desempenhos no âmbito da extração de conclusões.

5.4. Balanço global das aprendizagens realizadas

A unidade de ensino teve como pressuposto a valorização dos conhecimentos prévios dos alunos, intuitivos e informais e ao seu conhecimento do contexto de cada situação proposta para a partir daí construir conhecimento com significado ao nível do planeamento estatístico e da análise de dados, pretendia-se que os alunos fossem capazes de:

- Formular questões e planear adequadamente a recolha de dados tendo em vista o estudo a realizar.
- Identificar e minimizar possíveis fontes de enviesamento na recolha dos dados.
- Distinguir entre população e amostra e ponderar elementos que podem afetar a representatividade de uma amostra em relação à respetiva população.
- Construir, analisar e interpretar representações dos dados (incluindo o histograma) e tirar conclusões.
- Compreender e determinar a mediana, os quartis e a amplitude interquartis de um conjunto de dados, e utilizar estas estatísticas na sua interpretação.
- Escolher as medidas de localização mais adequadas para resumir a informação contida nos dados.

- Comparar as distribuições de vários conjuntos de dados e tirar conclusões.
- Responder às questões do estudo e conjecturar se as conclusões válidas para a amostra serão válidas para a população. (ME, 2007, p. 60),

Pretendia-se, assim, favorecer a compreensão da natureza da Estatística como ciência, assim como do seu contributo para resolução de situações que envolvam incerteza.

A unidade de ensino tinha sido inicialmente planificada para 9 blocos de 90 minutos, mas posteriormente durante a realização da tarefa 1 detetei a necessidade de complementar a sua realização com a análise do documento “oferta educativa da escola” (Anexo 8) de modo a esclarecer o objeto de estudo do item 6 e dado que o conceito de amostra representativa assenta em noções de probabilidades e os alunos ainda não tinham tido qualquer contato escolar com fenómenos aleatórios, uma vez que apenas o 7.º ano de escolaridade, a disciplina de Matemática foi norteada pelo programa de 2007, procedi à análise dos acetatos “O que é uma boa amostra?” e “A tipologia da população também importa?” (Anexo 9). Na tarefa 4, dado que existia alguma limitação na utilização do tempo não letivo dos alunos, optei por dinamizar a recolha dos dados dos alunos da turma, em análise, numa aula de estudo acompanhado e de modo idêntico foi recolhida a informação dos restantes alunos das várias turmas do ensino básico, estando na posse dos dados da generalidade dos alunos do 3.º ciclo, considerei vantajoso proceder à análise comparativa das estatísticas das amostras selecionadas com os parâmetros da população. Assim, a unidade de ensino foi realizada num total de 10 blocos e meio de 90 minutos.

No decurso da unidade de ensino, os alunos trabalharam essencialmente em grupo, mas também, individualmente, a pares e em grande grupo. De um modo geral, os alunos mostram-se entusiasmados, empenhados e participativos no trabalho desenvolvido. No entanto, algumas produções foram condicionadas por dificuldades na leitura e interpretação do enunciado dos respetivos itens.

Formular questões. No início da unidade de ensino (tarefa 1), os alunos apresentam dificuldades na interpretação do objeto de estudo o que comprometeu o processo de formulação de questões, posteriormente após esclarecimento e discussão em grande grupo optaram por formular questões dicotómicas e de escolha múltipla em detrimento

de uma formulação mais aberta. Na tarefa 2, os alunos mostram-se um pouco mais à vontade na formulação de questões que poderiam ser respondidas com base nos dados apresentados, sendo essencialmente questões de resumo – questões que envolvem a distribuição total dos dados ou que pretendem identificar o que é típico. Ainda durante a realização da unidade, os alunos formulam questões, com o intuito de caracterizar a população escolar do ensino básico da escola (tarefa 4), essencialmente subjetivas, e depois de discutir modo de interpretação como poderiam influenciar os dados – minimização do enviesamento da recolha – os alunos persistem na formulação de questões com expressões com múltiplas interpretações, pois consideram que poderão fazer os devidos esclarecimentos aquando da aplicação do inquérito. Na tarefa 6, no momento da apresentação e discussão do objeto de estudo e modo como iriam por em prática um estudo com o intuito de estimar o estoque do bar da escola, os alunos começam por formular questões, tais como “Qual a tua bebida preferida?” cuja resposta poderia recair numa bebida que o bar não pode disponibilizar ou “[Quais] as bebidas que costumavas beber?” cuja resposta conduz a uma preferência de consumo geral e/ou em quanto utilizador do bar, o que permitia ter uma ter uma visão geral do consumo e pouco precisa do real consumo no bar. No final da unidade de ensino, ainda permanecem algumas dificuldades, pois mais de metade dos alunos, num conjunto de três questões apresentadas, numa delas recorre a expressões com múltiplas interpretações.

Distinguir entre população e amostra. Ao longo da unidade de ensino, os alunos analisaram ou desenvolveram estudos estatísticos, entre as várias situações apresentaram-se surgiram populações de elementos inanimados (fósforos, palavras) e populações de seres vivos, nomeadamente pessoas particularmente em contexto escolar. Os alunos mostraram-se capazes de identificar corretamente a população envolvida no estudo e no caso de se ter recorrido a uma amostra, procediam à sua correta identificação. No final da unidade de ensino, quase um terço dos alunos face, a uma situação em contexto escolar, toma a amostra como população, equívoco recorrente em outros estudos (Fernandes, Carvalho e Ribeiro, 2007; Novaes, 2004).

Identificar e minimizar possíveis fontes de enviesamento na recolha dos dados. Durante a unidade de ensino, em várias situações a formulação de questões foi analisada como um aspeto que poderá introduzir possível enviesamento, os alunos discutiram a importância da clareza, da formulação o mais isenta possível e adequada ao objeto de estudo das questões de pesquisa, bem como a diferença da natureza dos dados recolhidos por “uma mesma questão” segundo duas tipologias (aberta, escolha múltipla). Na

tarefa 4, os alunos puderam constatar a importância de se uniformizar o processo de medição, de modo que a recolha seja o mais precisa possível.

Amostra representativa e ponderação de elementos que afetam a representatividade. Quanto à representatividade, antes da realização da unidade de ensino, mais de metade dos alunos, opta por uma opção que envolve uma amostra enviesada. Durante a unidade de ensino os alunos compreendem a importância de recorrer a um processo aleatório para a seleção de uma amostra representativa, mas como na maioria das situações em que lhes foi pedido um procedimento de amostragem a população era escolar, os alunos preferiram na maioria das vezes recorrer a uma seleção aleatória estratificada em detrimento de uma seleção aleatória simples. No início da unidade, a maioria dos alunos opta por amostras com uma dimensão correspondente a uma proporção significativa da população, e embora acabem por aceitar o estudo de amostras de menor dimensão, cerca de metade dos alunos ao atribuir um grande peso à variabilidade da população no âmbito da característica em estudo e do processo de amostragem, mostra-se incapaz de extrapolar as conclusões extraídas para a população.

Em vários momentos da unidade de ensino, nomeadamente durante a resolução das tarefas 1, 4, 5 e 6 discutiu-se a importância da seleção de uma amostra representativa, e possíveis enviesamentos dos dados face a um procedimento não representativo da população. Nas várias situações discutidas, os alunos extremamente sensíveis à variabilidade da característica em estudo na população e desde o primeiro momento, foram capazes de ponderar variados elementos que poderiam condicionar a representatividade de uma amostra. No teste final cerca de três quartos dos alunos identifica pelo menos um elemento que afeta a representatividade de uma amostra em relação à respetiva população. Ainda, no final da unidade os alunos revelam uma evolução expressiva, sendo que mais de 80% dos alunos opta por um processo de amostragem aleatório.

Construir, analisar e interpretar representações de dados/ Escolher as medidas estatísticas mais adequadas para resumir a informação contida nos dados.

Antes da unidade de ensino, os alunos evidenciavam dificuldades no completamento de uma tabela de frequências, na leitura da informação contida num histograma e na construção de um diagrama de extremos e quartis. As dificuldades ao nível da representação tabelar relacionaram-se com a noção de frequência relativa e/ou a mobilização da informação conjunta da coluna da frequência absolutas e frequência relativa dada de modo parcial. No início da unidade, os alunos quando confrontados com uma questão

de leitura direta da informação no histograma falham especialmente por desconhecimento da convenção (habitualmente assumida limite inferior fechado, limite superior aberto) e/ou dificuldade de interpretação do contexto da situação representada. No caso do diagrama de extremos e quartis, no início da unidade, os erros mais comuns foram: (i) determinação incorreta dos quartis e (ii) utilização da escala de modo incorreto ao não respeitar o valor da unidade ao longo do eixo.

Durante a unidade, no item 10 da tarefa 4, os histogramas construídos pelos grupos apresentam, uma escala horizontal adequada, e no eixo vertical uma escala iniciada em zero, com reprodução da unidade escolhida ao longo do eixo, em alguns casos a unidade escolhida mostra-se desadequada face ao valor das frequências a representar. Ao longo da realização da unidade de ensino, recorreram à correta determinação de frequências relativas nas tarefas 4, 5 e 6, pelo que no teste final se registou uma melhoria significativa o desempenho dos alunos. No item 9 da tarefa 4, mais de metade dos alunos apresenta um diagrama de extremos e quartis incorreto, a discussão em torno da determinação dos quartis e das várias incorreções cometidas pelos grupos, conduziu a uma melhoria do desempenho dos alunos. O facto da maioria dos alunos ter recorrido ao diagrama de caule e folhas para proceder a uma primeira redução dos dados e determinar as medidas de centrais e de localização, nos itens 6 e 7 da tarefa 4, permitiu consolidar a interpretação da simbologia caule|folha e atribuir significado às medidas de localização, fomentada pela análise comparativa das representações – diagrama de caule e folha e diagrama de extremos e quartis.

No decorrer da unidade de ensino o erro mais comum na representação do diagrama de caule e folhas foi a utilização inconsistente da unidade escolhida ao longo dos vários caules. No diagrama de extremos e quartis e no histograma os alunos manifestaram dificuldades na escolha de uma escala mais eficiente e adequada face aos valores a representar, pelo que em alguma das representações a marcação dos cinco números de resumo ou a frequências de cada classe, poderia ter sido um pouco mais precisa e de mais fácil leitura.

Na redação do relatório final da tarefa 4, um conjunto significativo de alunos mobilizou informação sobre o centro, a forma e a dispersão da distribuição recolhida junto das representações – histograma e diagrama de extremos e quartis. A generalidade dos grupos utilizou as medidas estatísticas na interpretação das várias variáveis quantitativas discretas ou contínuas. Um conjunto menor de alunos, inclui no relatório referência a medidas de dispersão. Globalmente a maioria dos alunos, ao longo da unidade

de ensino mostrou-se capaz de as medidas estatísticas mais adequadas para resumir a informação contida nos dados.

Ainda no âmbito das representações, antes da unidade de ensino, apenas um quarto dos alunos se mostrou capaz de construir o diagrama de extremos e quartis tendo como referência de análise o histograma da distribuição. No final da unidade, este tipo de representação continua a ser um desafio para os alunos, dada a complexidade de observar o conjunto de dados representado sob o ponto de vista agregado e ser necessário associar a transformação das frequências absolutas do histograma em densidade no diagrama e assumir que uma maior densidade relativa corresponde a um menor comprimento da componente no diagrama, aspetos que Bakker, Biehler e Konold (2005) consideram ser os maiores obstáculos à interpretação de um diagrama de extremos e quartis. No final da unidade, os alunos melhoraram o seu desempenho na construção, análise e interpretação das várias representações, sendo de salientar a melhoria o desempenho de mais de um quarto de alunos, na representação tabelar e do diagrama de extremos e quartis e na utilização da medidas de centro e dispersão para descrever uma distribuição.

Tirar conclusões e analisar a sua validade para a população. Antes da unidade de ensino, menos de metade dos alunos se mostra capaz de analisar os dados de uma distribuição, correspondente a uma amostra e face a algumas considerações identificar uma conclusão correta e passível de ser generalizada à população em estudo. E mais de um terço dos alunos dispõem-se a generalizar uma conclusão a uma população que contém a população em estudo, com base em convicções pessoais. No decurso da unidade de ensino, nomeadamente nas tarefas 4, 5 e 6, os alunos tiram conclusões no âmbito do seu conjunto amostra, e comprometem-se pontualmente com a sua generalização à população, na tarefa 4, a comparação das estatísticas das quatro amostras e com os parâmetros da população, fomenta a compreensão da variabilidade no processo de amostragem, que por sua vez se equilibra pela similaridade de muitos dos “comportamentos” nas várias amostras, designadamente no âmbito da forma, do centro e da distribuição global das várias amostras que por sua vez são uma “boa percepção” dos mesmos comportamentos na população. Na tarefa 5, a generalidade dos alunos compreende na prática o que foi discutido em termos teóricos no momento de síntese da tarefa 4, ao reconhecer que um estudo idêntico tendo por base um outro texto representativo da língua portuguesa, se obteriam porções de cada vogal distintas das obtidas face ao texto “A rádio escola”, mas assumem que no grupo das vogais, mais usuais, estariam a vogais a,

e e o e as restantes vogais seriam menos corriqueiras. Na tarefa 6, os alunos sentiram, desde o seu início, ser pouco viável obter uma boa previsão do estoque do bar com base na amostra aleatória recolhida pelos vários grupos turma, perante vários condicionantes do consumo, nomeadamente a sazonalidade e as tendências de consumo. No final da unidade de ensino, os alunos continuam a evitar aceitar a tomada de decisões de previsão, com base em estatísticas de uma amostra representativa, sendo que apenas um quinto dos alunos, argumenta que a decisão de estimação era válida embora estivesse associada a algum grau de incerteza, face à variabilidade real da característica em estudo na população. Quanto à análise dos dados e posterior comunicação de conclusões, no final da unidade, a maioria dos alunos redige um pequeno texto com a indicação de um ou duas medidas de tendência central, e inclui um comentário quanto à dispersão, com base em pelo menos uma das seguintes medidas - amplitude, amplitude interquartil, os extremos, o intervalo interquartil. E cerca de metade considera que as conclusões são passíveis de serem extrapoladas à população pois a amostra foi selecionada de modo aleatório, em contrapartida cerca de um quinto dos alunos ponderou a dimensão da amostra face a dimensão da população, o facto de a população não ser homogénea e ter elevada dimensão, considerou que as conclusões não são passíveis de serem extrapoladas à população. No final da unidade, a maioria dos alunos revela-se capaz de mobilizar termos, expressões de teor estatístico com significado, para argumentar a sua posição e que têm em consideração que a validade das conclusões para a população, estudada com base numa amostra, dependem de um equilíbrio entre a qualidade do processo de amostragem e a dimensão da amostra.

Capítulo 6

O Caso de Ana

6.1. Apresentação

Ana, à data do estudo, tinha 14 anos, vivia com a mãe perto da escola e participava no grupo de escalada e canoagem no âmbito do projeto Desporto Escolar. No seu tempo livre costumava ver televisão e recorrer às redes sociais para estar em contato com os amigos. É uma jovem bastante sociável e com boas relações com os colegas e, enquanto aluna, empenhada, interessada e organizada nas várias disciplinas, com um desempenho bastante satisfatório ao longo do seu percurso escolar e sem qualquer retenção. Quando questionada sobre as disciplinas que preferia, indicou Inglês e Ciências Naturais. Afirmou gostar de Matemática mas considerou-se uma aluna mediana e sentindo que poderia ter mais sucesso se estudasse mais do que apenas na véspera dos testes. Quanto à progressão de estudo, indicou que pretendia seguir o curso de Ciências e Tecnologias, e no final do 12.º ano candidatar-se à faculdade num curso da área das Ciências.

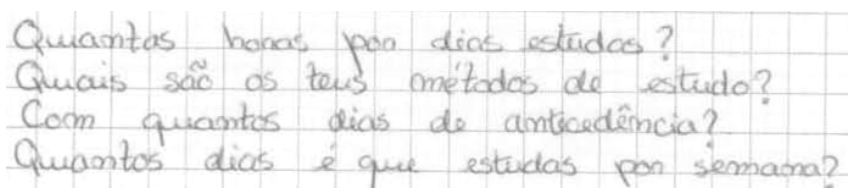
Enquanto sua professora, considerei que o seu desempenho nas aulas de matemática estava acima da média dos alunos da mesma faixa etária, principalmente nos tópicos de Álgebra e Geometria. Os critérios que levaram à escolha de Ana para estudo de caso foram a existência de evidência e a abrangência, dado que a aluna, embora poucas vezes se disponibilizasse para intervir oralmente, mostrava-se capaz de o fazer em registos escritos, expressando as suas ideias de modo claro, com correção científica e com boa capacidade de fundamentação.

6.2. Capacidades de planeamento estatístico e análise de dados no início da unidade de ensino

Nesta secção refiro o desempenho da aluna no âmbito do teste inicial (Anexo 3) e o seu desempenho no âmbito do trabalho de grupo durante a realização das tarefas 1 e 2 (Anexos 7 e 10).

6.2.1. Planeamento estatístico

Formular questões. A análise do trabalho de grupo na questão 6b da tarefa 1 mostrou que Ana e os colegas formulam questões bastante generalistas e dissociadas do objeto em estudo “projetos quanto ao prosseguimento de estudos”. Durante a discussão constatou-se que tanto ela como a generalidade dos colegas não tinha compreendido este objeto. No caso do seu grupo, os alunos interpretaram o objetivo do estudo como sendo a análise dos procedimentos de estudo (Figura 86).



Quantas horas por dias estudas?
Quais são os teus métodos de estudo?
Com quantos dias de antecedência?
Quantos dias é que estudas por semana?

Figura 86. Resposta do grupo de Ana à questão 6b da tarefa 1.

Posteriormente na tarefa 2, Ana participou na fase final da tarefa propondo a formulação de algumas questões passíveis de serem respondidas a partir dos dados da base de dados, “Qual é a altura média?”, “Qual a distância de casa à escola mais frequente?”.

Distinguir entre população e amostra. Ainda na realização da tarefa 1, ao discutir com os colegas de grupo a questão 6, relativa à realização de um estudo estatístico, Ana considerou que em primeiro lugar se deveria decidir o tipo de estudo, e no caso de se optar por uma sondagem, de que modo se deveria proceder à seleção da amostra. Sendo perceptível que a aluna mobilizou corretamente os conceitos de população e de amostra.

Ana - Sim, mas como? Por exemplo, fazes o inquérito a quem? ... a uma pessoa de cada turma ou a dois ou três, ... a toda a gente? Escolhes como?

D - Escolhes turmas...

Rui - Não, escolhes três... escolhes cinco números de cada turma, ... (pausa), vocês não se lembram do teste [inicial], havia lá uma pergunta que tinham feito assim ...

Ana - Temos que fazer aleatório...

J e D - Fazemos um sorteio ...

Amostras representativas e ponderação de elementos que afetam a representatividade. No teste inicial, face a uma questão relativa à escolha de uma amostra, item 5.1, Ana selecionou um processo de amostragem aleatório simples – “questionar 160 alunos escolhidos por sorteio entre os 1560”, rejeitando os processos de amostragem que considerou passíveis de enviesamento, como por exemplo “questionar 160 alunos que sejam amigos/conhecidos dos alunos responsáveis pelo estudo”. Durante o trabalho de grupo, interveio oportunamente constatando que na questão 2 da tarefa 1 – “População e amostra”, a amostra não era representativa dado que o inquérito se realizara à porta de um supermercado e este facto poderia influenciar um grupo significativo dos inquiridos no momento da escolha, devido à publicidade que mais recordavam.

Ana - Ok, isto não pode ser, porque estão num supermercado as pessoas vão logo ... lembra-se logo ... lembra-se logo dos anúncios desse supermercado.

Ainda na realização da questão 5 da tarefa 1, Ana interveio oportunamente constatando que a amostra não era representativa dado o elevado conhecimento na área de Biologia que a generalidade dos médicos domina, por questões de formação académica, comparativamente à restante população.

D - não (está correta) ... porque os médicos não são biólogos!

Ana - Isso não tem nada a ver, (riso) que coisa até tu “dás” biologia ...

Rui - ... porque os médicos têm mais cultura que a maioria das pessoas, são mais espertos

D - Então?

Ana - Tu (D) disseste que eles não são biólogos e isso é o contrário ... os médicos têm um curso...

D - sim, tem um curso mas tiraram um curso de ... medicina

Rui - ... e a medicina é baseada nas ciências...

Ana - ... e que é baseada na biologia!

Breve síntese. Ana, no teste inicial e durante a realização das tarefas 1 e 2, mostrou que, no âmbito do planeamento estatístico, distingue os conceitos de população e amostra, pondera elementos que podem afetar a representatividade de uma amostra em relação à respetiva população e, perante um conjunto de dados, formula questões estatísticas de resumo/descrição válidas e passíveis de serem respondidas pelos dados. Glo-

balmente, o trabalho desenvolvido pela aluna no início da unidade de ensino, no âmbito do planejamento estatístico, foi satisfatório, evidenciando uma compreensão dos conceitos população e amostra e um reconhecimento significativo da variabilidade natural.

6.2.2. Análise de dados

Construir, analisar e interpretar representações dos dados

Representação tabular. No teste realizado no início da unidade de ensino, Ana não demonstrou dificuldades na leitura e interpretação de tabela, tendo completado os espaços em branco corretamente. No entanto, é de salientar que procedeu ao completamento recorrendo exclusivamente à manipulação de regras de “três simples” (Figura 87), não havendo evidência de associação entre a coluna “percentagem de alunos” e o conceito de frequência relativa.

Idades	Nº de alunos	Percentagem de alunos
11	5	4
12	15	12
13	30	24
14	32	25,6
15	30	24
16	11	8,8
17	2	1,6
Total	125	100%

Figura 87. Resposta da Ana à questão 1.1 do teste inicial.

Histograma. Quanto à construção, análise e interpretação do histograma (grupo 2 do teste inicial), no item 2.1 que pretendia aferir a identificação de algumas das características do histograma, nomeadamente número de classe e amplitude das classes, Ana identificou corretamente o solicitado. Nos itens 2.2, 2.3 e 2.4 que tinham como objetivo o reconhecimento de alguns aspetos da representação e mobilização do conhecimento para uma leitura literal da representação, o desempenho de Ana foi muito bom, não tendo falhado nenhuma das questões, e não tendo cometido o erro mais comum entre os colegas de turma que indicaram que o tempo do vencedor da prova seria 2 minutos, dado que Ana respondeu corretamente que o valor estava “entre 0 a 2 minutos por exemplo 1,5 min”.

Diagrama de caule e folhas. Quanto à leitura e interpretação do diagrama de caule-e-folhas, o item 4.1 (teste inicial) solicitava a identificação do número total de efetivos apresentados na representação, sendo a questão formulada no âmbito do con-

texto da situação (número de pessoas que responderam à questão 4 do inquérito), Ana respondeu corretamente. No item 4.2 (teste inicial) que solicitava a leitura dos extremos, Ana procedeu à sua correta identificação. Evidenciou ainda ter conhecimento que a representação apresenta os dados ordenados, recorrendo à identificação dos quartis sobre o mesmo, embora tenha determinado a mediana e restantes quartis incorretamente (Figura 88).

Determinação de medidas estatísticas e a sua interpretação

Moda e mediana. Tal como a generalidade dos seus colegas, Ana não mostrou dificuldades na identificação da moda da distribuição de idades. Procedeu a uma incorreta identificação da mediana e a uma incorreta interpretação pois no completamento da frase do item 4.3 (teste inicial) indicou o valor 14 e no diagrama de extremos e quartis a mediana toma o valor 17 (Figura 88).

Comparar duas distribuições e validar conclusões. No âmbito da comparação de duas distribuições recorrendo à análise conjunta de medidas estatísticas centrais e de dispersão, item 3 (teste inicial), Ana recorreu à comparação de frequências absolutas das classes, não sentiu necessidade de explicitar que se tratava de conjuntos de igual dimensão e, como tal, a comparação é válida. Ana indicou como argumento, “*No grupo A houve mais alunos na casa do 80/89 do que o grupo B, a média é menor por causa do aluno com nota na casa 0/9*”. Ana reconheceu a influência desproporcionada da nota do aluno mais fraco do grupo A, e o facto de a média ser sensível a valores atípicos. Reconheceu também que uma análise comparativa dos grupos deve recorrer a mais informações estatísticas e não somente à comparação das médias, indicando a existência de um maior número de alunos com resultados altos no grupo A.

Transformar uma representação numa outra representação

Diagrama de caule e folhas - diagrama de extremos e quartis. Ana, tal como cerca de 20% dos alunos da turma, perante aos valores obtidos para o resumo da distribuição de cinco números, procedeu à sua correta representação no diagrama de extremos e quartis (Figura 88).

ser questionada quanto ao facto de não ter recorrido a um argumento análogo e seleccionado a opção “*Cerca de 25% dos alunos da escola contactam, via Messenger, com 20 ou mais pessoas.*”, uma vez que assinalou 20 amigos como 3.º quartil, alegou que a informação recolhida junto dos alunos do 8.º ano não seria válida no caso de se pensar em alunos mais velhos, dado que conheciam mais pessoas e tinham mais liberdade de os contactar.

Breve síntese. No teste inicial, realizado antes da unidade de ensino, no que diz respeito à construção, análise e interpretação de distintas representações dos dados, Ana evidenciou um bom nível de análise. Não se limitando a uma leitura literal dos dados, conseguiu recorrer à interpretação da forma e à identificação do centro de uma distribuição de dados numa determinada representação gráfica para a transformar numa outra representação. Manifestou dificuldades na determinação da mediana e dos quartis, bem como na interpretação da mediana, ao completar incorretamente a frase do item 4.3 do teste inicial, o que poderá estar associada ao facto do número de efetivos da amostra ser ímpar, e não tanto à ordenação da distribuição pois reconhece que o diagrama de caule e folhas é um resumo ordenado dos dados. Quanto à utilização da média na análise comparativa de duas distribuições, a sua resposta mostrou que reconhece que a média não é uma medida robusta e a análise comparativa de duas distribuições deve atender a outros indicadores estatísticos. Não foi consistente no recurso a argumentos matemáticos, utilizando na maioria dos casos a sua opinião pessoal quando desenvolveu, avaliou inferências e previsões baseadas numa amostra representativa para a população respetiva. Globalmente, o trabalho desenvolvido pela aluna antes da unidade de ensino, no âmbito da análise de dados, foi satisfatório tendo sido, no entanto possível, melhorar a capacidade de desenvolver e avaliar inferências e previsões baseadas nos dados.

6.3. Capacidades de planeamento estatístico e análise de dados durante a unidade de ensino

Nesta secção refiro o desempenho da aluna no âmbito do trabalho de grupo/par, durante a realização das tarefas 3 a 6 (Anexos 11 a 14) e o seu desempenho na tarefa proposta na primeira entrevista (Anexo 5).

6.3.1. Planeamento estatístico

Formular questões. Na primeira entrevista, perante uma primeira fase de planeamento de um estudo na escola, sobre o cumprimento da recomendação de consumo de água, Ana considerou pertinente o estudo das variáveis “Quantidade de líquidos consumidos nos últimos dois dias (em litros)”, “Número médio de vezes que consome líquidos por dia” e “Quantidade de chá/água consumida nos dois últimos dias em litros” sendo recolhidas com base em questões de pesquisa que solicitavam ao aluno inquirido a estimativa do seu consumo de líquidos, de chá/água e a média diária do número de ingestões de líquidos. Ainda durante a realização da tarefa 4, Ana participou na formulação de questões para a caracterização dos alunos do 3.º ciclo da escola, propondo as seguintes questões “*O que gosta de fazer nos tempos livres?*” e “*Tens telemóvel? Quanto costumas gastar por mês?*”, depois de incentivar os alunos a refletir sobre a interpretação de algumas questões, e fomentar a formulação isenta de múltiplas interpretações, opta por reformular uma das questões “*Em média, quanto gastas por mês em telemóvel?*”

Uma vez que o reconhecimento da existência de variabilidade é fundamental no processo de formulação de questões estatísticas, importa referir que Ana se revelou capaz de reconhecer a variabilidade natural existente num conjunto de dados que incluiu naturalmente elementos distintos e a variabilidade decorrente do processo de amostragem. Na tarefa 5, “Frequência das vogais na língua portuguesa”, reconheceu que podiam existir amostras sem a presença das vogais “a” e “e” embora a tendência sentida no conjunto de diferentes amostras representativas fosse que na maioria persistia um domínio destas vogais, mas não foi capaz de argumentar junto dos restantes colegas que a inexistência das vogais “a” e “e” ou a desproporção de uma vogal face a outra, num determinado texto literário pode resultar de um exercício literário artificial.

Rui - Num texto pode-se usar muito o “a”, noutro não ... de um texto para outro as vogais mudam ...

Ana - Sim, mas o “a”, o “e” aparece muitas vezes nos textos ... aqui o “a” foi tirado de propósito ... não acontece muito...

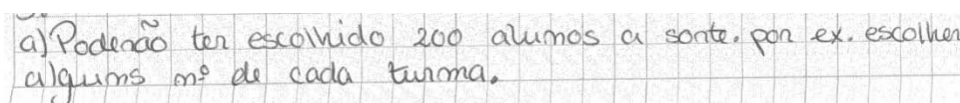
Rui - ... e os poemas com rimas, usam uma letra ...vogal por causa do som!

Ana - ... tá bem, então as letras mudam muito de um texto para outro é isso que vamos dizer? (Restantes elementos anuem) ...

Distinguir entre população e amostra. Nos vários estudos estatísticos realizados, nomeadamente nas tarefas 4 e 6 e; nos grupos 2 e 5 da primeira entrevista, Ana não manifestou dificuldades em identificar a população como sendo o conjunto de alunos da comunidade escolar do ensino básico, na tarefa 4, e os alunos da escola, na tarefa 6 e grupos 2 e 5. Reconheceu que a amostra seria composta por um conjunto de alunos do ensino básico escolhidos aleatoriamente, no caso da tarefa 4 e, na situação do grupo 5 optou por considerar uma amostra com 350 alunos, cerca de $\frac{1}{4}$ dos alunos da escola, escolhidos aleatoriamente no bar da escola.

Amostras representativas e ponderação de elementos que afetam a representatividade

Amostras representativas. Perante a necessidade de escolher uma amostra representativa, nos itens 2a) e 5.2 da primeira entrevista, para o estudo do consumo semanal de refrigerantes e o estudo sobre o cumprimento do consumo de água, respetivamente, Ana optou prioritariamente por uma amostragem aleatória simples. Deparando-se com dificuldades em pôr em prática esse processo de escolha aleatória no item 2a) e dado tratar-se da população “comunidade escolar” optou, então, por uma amostra estratificada segundo os subgrupos turmas (Figura 90).



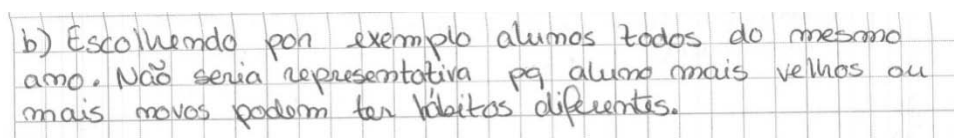
a) Poderão ter escolhido 200 alunos a sorte, por ex. escolher alguns nº de cada turma.

Figura 90. Resposta ao item 2a) da 1.ª entrevista.

No item 5.2, tomou a sua escola como referência e considerou que deveria inquirir cerca de um quarto dos alunos (aproximadamente 350 alunos), tendo explicitado oralmente que iria passar um dia no bar da escola, para dar hipótese aos alunos do horário da manhã e aos da tarde, reunindo assim os dados correspondentes às variáveis escolhidas, junto de vários alunos escolhidos ao acaso, de entre os que entravam no bar.

Ponderação de elementos que afetam a representatividade. Quando inquirida sobre um procedimento que conduza a uma amostra enviesada, Ana indicou, à semelhança do que o seu grupo de trabalho indicou na tarefa 4 e 6, a escolha de uma amostra restrita a um subgrupo da população (faixa etária, género ...). Justificou que deste modo a amostra não seria representativa da população pois a variável em estudo – consumo

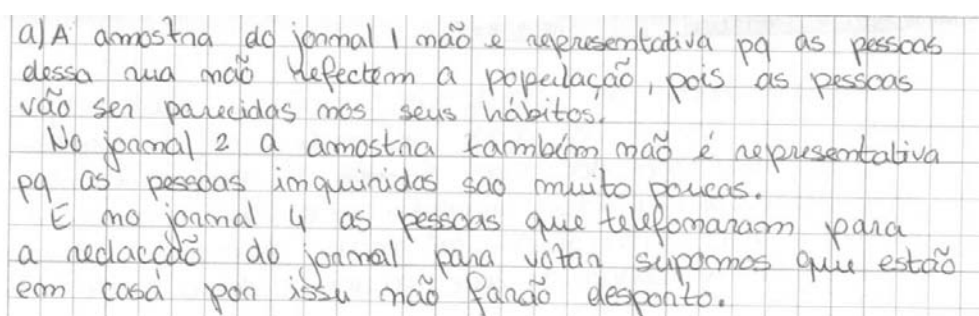
semanal de refrigerantes – relaciona-se com os vários substratos existentes na população em análise, relativamente à característica idade (Figura 91).



b) Escolhendo por exemplo alunos todos do mesmo ano. Não seria representativa pq alunos mais velhos ou mais novos podem ter hábitos diferentes.

Figura 91. Resposta ao item 2b) da 1.^a entrevista.

Análise crítica de estudos estatísticos. Ana mostrou-se capaz de fazer uma análise crítica de estudos estatísticos. No item 4 a) da primeira entrevista apresentou argumentos para a rejeição de estudos realizados com base em amostras não representativas e face a dois estudos com amostras recolhidas com base num mesmo procedimento aleatório, escolheu a amostra de maior dimensão – amostra do jornal 3 (Figura 92).



a) A amostra do jornal 1 não é representativa pq as pessoas dessa não refletem a população, pois as pessoas vão ser parecidas nos seus hábitos. No jornal 2 a amostra também não é representativa pq as pessoas inquiridas são muito poucas. E no jornal 4 as pessoas que telefonaram para a redacção do jornal para votar supomos que estão em casa por isso não farão desporto.

Figura 92. Resposta ao item 4 a) da 1.^a entrevista.

Breve síntese. Durante a lecionação da unidade de ensino, Ana foi capaz de identificar a população em estudo e a amostra em análise nas várias situações propostas. Face a um conjunto de sondagens apresentadas, a aluna mostrou-se crítica, rejeitou os estudos que recorreram a amostras não representativas e justificou o possível enviesamento em cada situação. Em diferentes contextos, identificou e ponderou elementos que afetavam a representatividade de uma amostra em relação à respetiva população e recorreu a exemplos para ilustrar os diferentes elementos que ponderou. Tendo em consideração as categorias do desenvolvimento do conceito de amostra de Watson e Moritz (2000), a aluna encontra-se na categoria mais elevada, designada por grandes amostras sensíveis a enviesamento. Quanto à formulação de questões, optou por formular questões de pesquisa com o intuito de proceder a uma descrição e manifestou dificuldades em formular questões isentas de dupla interpretação.

6.3.2. Análise de dados

Natureza dos dados e representações adequadas. Ana, ao ser inquirida quanto à categorização das variáveis em estudo no item 5 b) (1.^a entrevista), identificou corretamente a categoria das sete variáveis e, perante questionamento da professora, identificou a escolha de uma representação gráfica para algumas das variáveis. Ana, embora não tivesse escolhido as representações mais adequadas à natureza dos dados, revelou conhecer os dados específicos de cada variável, clarificou os motivos da escolha das representações, quando indicou a escolha de medidas estatísticas que iria apresentar e que considerações pensava extrair dos dados recolhidos; “*a maioria consome menos de duas peças*”, “*o normal é beber mais de 1 e menos de 2 litros, contando com pequeno-almoço, almoço, lanche e jantar*”, “*podia dizer que metade [dos alunos] bebe mais de ...*”.

Ana - A n^o 1 é uma variável quantitativa porque fala do n^o de peças de fruta consumidas diariamente ou seja dá para contar por isso é quantitativa, sim ... e é quantitativa discreta porque o n^o de peças que se consome não é seguido, por exemplo não se pode consumir 2,3 maçãs então é discreta ... a [variável] n^o 2 também é quantitativa, mas não é discreta, é contínua ... por causa que os litros é um n^o contínuo pode-se consumir 1,1 litros, por isso ... O n^o de vezes que consome líquidos por dia também é quantitativa mas é como a 1^a é discreta ... A quantidade de água e chá consumida nos últimos dias é como a 2 é contínua ... a 5 acompanha o consumo de água (sim ou não?) é qualitativa porque aqui não está a perguntar quantas vezes acompanha ou não acompanha ... A [variável] 6 também é discreta ... e a 7 é como a [variável] 5 também é qualitativa.

Prof^a - Porquê neste caso?

Ana - Porque ao perguntar uma bebida preferida não se está a contar nada, é mais uma qualidade ...

(...)

Prof^a - ... por exemplo aqui para o n^o de peças de fruta consumidas estavas a pensar em que tipo de gráficos?

Ana - Por exemplo um histograma.

Prof^a - para o n^o de peças consumidas? ... e para a quantidade de líquidos consumido? ...

Ana - Sim ...um diagrama de caule e folhas

Prof^a - Então para a 1^a variável, n^o de peças de fruta, que tipo de respostas esperas ter?

Ana - 1, 2, 4, 6, 1, nenhuma, ...

Prof^a - Então sendo um histograma que estavas a pensar usar... Que classes irias considerar?

Ana - Por exemplo, de nenhuma a duas, de duas a quatro e assim mais ou menos ...

Prof^a - E para a quantidade de líquido, dá-me dois ou três exemplos de respostas possíveis?

Ana - Um litro, meio litro, dois litros (Silêncio)

Prof^a - Tinhas-me dito, um diagrama de caule e folhas não foi? Então quais seriam os teus caules?

Ana - Sim ... então poderia ser o 0, 1 e 2

Prof^a - E as folhas refletiam o quê neste caso?

Ana - Seria 5, 0 e 0

Prof^a - Pois tinha dito meio litro, um litro, dois litros.

Construir, analisar e interpretar representações dos dados

Gráfico de barras. No item 1a) da primeira entrevista, Ana reconheceu que o recorte do eixo vertical é o motivo da leitura errada por parte do coordenador (Figura 93).

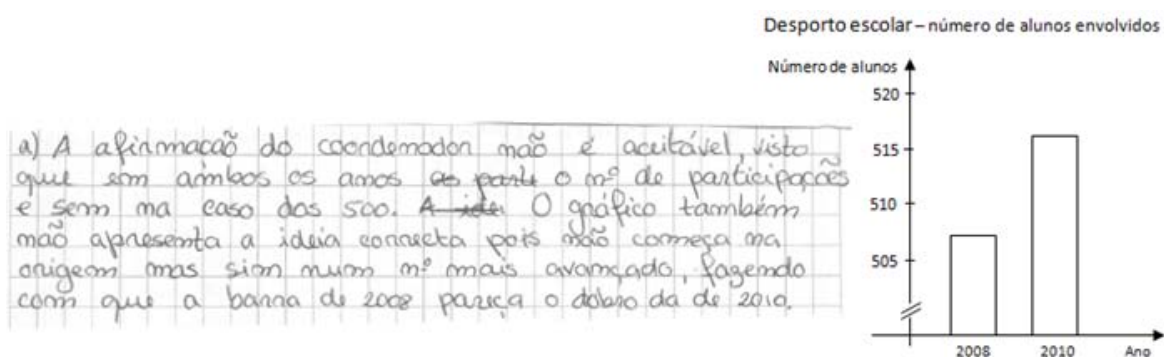


Figura 93. Resposta ao item 1 a) da 1.^a entrevista.

Sectograma. Na primeira entrevista (item 2 c)), Ana reconheceu o erro processual, dos autores do estudo, na determinação da amplitude do ângulo correspondente a uma determinada “fatia” do gráfico, determinou a correspondente amplitude recorrendo a um raciocínio proporcional e alega que o aluno em causa “estava e pensar em percentagem em vez de pensar na amplitude total do círculo”.

Histograma. Ana, ao ser inquirida quanto às decisões que tomaria na construção de um histograma (item 4b da primeira entrevista), para representar 500 dados relativos aos gastos mensais com a prática desportiva que variam entre 5,7 e 52 euros, considerou que devia construir classes com amplitude de 5 euros, perfazendo um total de 11 classes, o que no contexto da situação corresponde a um opção adequada. Durante a realização da tarefa 4, Ana participou na construção do histograma para a variável altura e propôs a inclusão do polígono de frequências, cometendo a incorreção de, no início do polígono, não ter colocado a união do segmento de reta entre o ponto médio da classe anterior e o ponto médio do topo da primeira classe com frequência não nula, o que vol-

tou a ocorrer entre o ponto médio do topo da última classe de frequência não nula e o ponto médio da classe seguinte.

Diagrama de caule e folhas. Na tarefa 4, Ana e os seus colegas de grupo, recorreram ao diagrama de caule e folha para uma primeira redução dos dados para as várias variáveis quantitativas em estudo e para proceder à determinação das medidas de localização da variável altura, tendo no entanto, cometido a incorreção de duplicar um caule que apresentava uma grande quantidade de folhas.

Diagrama de extremos e quartis. Durante a realização da tarefa 4, Ana e os seus colegas determinaram os cinco valores de resumo corretamente e procederam à construção do diagrama de extremos e quartis da variável altura, tendo escolhido uma escala adequada. Apesar disso, no seu relatório não integraram qualquer referência relativa à concentração ou dispersão dos dados da variável altura. No decorrer da primeira entrevista (item 3a) da 1ª entrevista) Ana mostrou-se capaz de ler e interpretar um diagrama de extremos e quartis, quando procede à comparação de duas distribuições de dados.

Determinação de medidas estatísticas e a sua interpretação

Moda, média e mediana. Ana, no item 3b) da primeira entrevista, reconheceu a existência do valor atípico (16 mensagens) no conjunto de dados, procedeu à correta determinação da mediana e defendeu que esta é a medida estatística central mais adequada para resumir a informação contida nos dados, porque “16 é um valor baixo, iria baixar muito a média” o que veio a confirmar ao determinar a média, cerca de 75 mensagens, valor que fica abaixo de sete dos oito dados recolhidos. Como no presente caso a distribuição é bimodal (83 e 85 têm ambos frequência 2), considera que não deve escolher nenhum destes valores.

Comparar distribuições e validar conclusões. Em diferentes situações Ana revelou ser capaz de proceder à comparação de duas distribuições. Na primeira entrevista, ao comparar duas distribuições de desempenhos de 10 atletas, antes e depois de duas semanas de treino (item 1 b)), Ana considerou que a conclusão óbvia do estudo seria assumir que “o treino teve efeitos ao nível do desempenho dado que 8 dos 10 alunos melhoraram os seus resultados”. No item 3 a) da primeira entrevista, sendo conhecidos os respetivos diagramas de extremos e quartis, Ana identificou a operadora cujos utilizadores recorreram mais ao envio de mensagens de texto, e argumentou que a amplitude total e a amplitude interquartil eram maiores nessa distribuição.

Ana - Acho que foi na Edfon porque... os extremos... o máximo e o mínimo são maiores tem uma maior amplitude do que na Telepat, por isso vai haver mais mensagens e também (n)os quartis... também tem uma maior amplitude que os quartis da Telepat...

Fica implícito que a escolha não foi feita apenas porque a amplitude do intervalo interquartil era maior, mas porque se inicia acima do valor mediano da empresa Telepat. Ana argumentou ainda dizendo “*por último porque pudemos ver que quase 50% das pessoas inquiridas na Edfon manda mais mensagens que na Telepat.*”

Generalizar as conclusões para a população

No item 1b) (1ª entrevista) ao ser questionada sobre a validade da conclusão no âmbito do conjunto dos vários alunos que frequenta as distintas modalidades desportivas no projeto Desporto Escolar, Ana considerou que a generalização era possível com base na sua convicção pessoal de que após duas semanas de treinos, independentemente da modalidade e do tipo de atletas, o desempenho global teria melhorado. Ana não questionou a representatividade da amostra (10 atletas de atletismo), quanto à quantidade – pequena dimensão menos de 30 elementos, - nem quanto à qualidade, uma vez que se restringiu a um grupo de atletas de uma das muitas modalidades praticadas. Posteriormente no item 5.2 (1ª entrevista), Ana fundamentou que as conclusões extraídas deveriam ser generalizadas à população, no facto de 350 alunos já ser uma amostra de bom tamanho, o processo de recolha ter sido aleatório e como tal teremos uma “mostra” da potencial diversidade quanto aos hábitos de consumo de água. Durante a elaboração da tarefa 4, Ana e os seus colegas de grupo não fizeram qualquer referência à possibilidade das considerações indicadas serem ou não passíveis de generalização à população. Na tarefa 6, Ana e os seus colegas de grupo mostraram-se céticos quanto ao facto de ser possível prever o estoque mensal do bar com base nos dados recolhidos, pois sentiram que o consumo de bebidas pode ser bastante variável, dependendo da temperatura do dia, da estação do ano, do poder económico e dos gostos pessoais, pelo que a amostra recolhida poderia não ser representativa dos consumos da população em estudo.

Breve síntese. Durante a lecionação da unidade de ensino, Ana revelou sentido crítico face às representações gráficas e a possíveis interpretações erróneas. Além disso, revelou-se capaz de categorizar dados de distinta natureza e saber a razão pela qual um tipo específico de dados se adequa a uma dada representação e/ou medida estatística. A

aluna mostrou-se capaz de construir, analisar e interpretar as várias representações gráficas propostas. Mostrou-se ágil na determinação das medidas de tendência central e capaz de privilegiar a mediana em detrimento da média ou da moda em distribuições com valores atípicos. No âmbito da comparação de distribuições e validação das suas conclusões, procedeu à comparação e argumentou com base em argumentos estatísticos válidos. No que se refere ao desenvolvimento e avaliação de inferências, Ana, embora tenha sido capaz de identificar possíveis fatores de enviesamento em muitos dos estudos apresentados, no momento de extrair conclusões, perante algumas situações, recorreu às suas próprias opiniões e optou por não refletir sobre a qualidade da amostra. Noutras situações, considerou que a diversidade de dados recolhidos numa amostra representativa da população estaria aquém da variabilidade existente na população no âmbito da variável em estudo, pelo que considerou que não era possível proceder à generalização. Globalmente, o trabalho desenvolvido pela aluna durante a unidade de ensino revelou o domínio de vários dos aspetos que Garfield (2002) considera serem fundamentais para o raciocínio estatístico sobre dados e sobre representações.

6.4. Capacidades de planeamento estatístico e análise de dados depois da unidade de ensino

Nesta secção refiro o desempenho da aluna no âmbito da tarefa proposta na segunda entrevista (Anexo 6) e no teste final (Anexo 15).

6.4.1. Planeamento estatístico

Formular questões. No final da unidade de ensino, tendo sido pedido um esboço do estudo para conhecer as práticas de higiene oral (item 4 da 2.^a entrevista), a Ana começou por formular as seguintes questões: “*Quantas vezes escovas o(s) dentes por dia? Costumas utilizar fio dentário? Com que frequência vais ao dentista?*” Ao ser interpelada pela professora sobre o significado atribuído a “costumas utilizar” e constando que o dado “SIM” devolvido na segunda questão poderia significar para algumas pessoas inquiridas uma única utilização semanal, optou por reformular a questão. Na terceira questão, considerou que os dados devolvidos poderiam ser “*regularmente*”, “*com pouca frequência*”. Perante a situação hipotética de um jovem que considera uma ida por ano ao dentista uma frequência regular, optou pela sua reformulação. Com a

nova redação da questão iria recolher o tempo entre idas e deste modo poderia organizar os dados num histograma cujas classes seriam o tempo compreendido entre duas idas ao dentista (Figura 94).

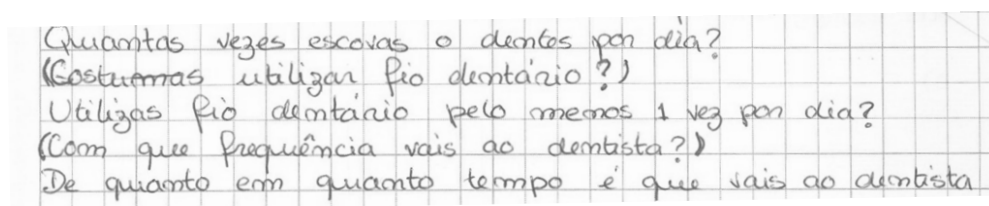


Figura 94. Questões de inquérito propostas por Ana no item 4 da 2.^a entrevista.

No item 3.2 (teste final), cujo objeto de estudo era a qualidade do sono dos adolescentes portugueses entre os 12 e 16 anos, Ana sugeriu a formulação das seguintes perguntas de pesquisa: “*Quantas horas (em média) é que dormes?*”, “*Deitas-te sempre e levantas-te sempre à mesma hora?*” e “*Costumas deitar-te muito tarde?*”, uma vez mais se constata que Ana formulou questões pouco claras, como é o caso da segunda questão, cujos dados recolhidos seriam pouco viáveis, uma vez que uma resposta negativa poderia ocorrer em três cenários possíveis – na maioria das vezes deita-se em horários distintos; ou, na maioria das vezes levanta-se em horários distintos; ou ainda, na maioria das vezes o deitar e o levantar não têm um horário estável. No caso da terceira questão redigida, houve um aspeto subjetivo de interpretação, que comprometeu os dados recolhidos, tendo sido preferível uma questão que indicasse, em específico, uma hora de deitar com a seguinte formulação “*Durante a semana, em três ou mais dias, deitas-te depois ou antes das 22h.30m?*”.

Distinguir entre população e amostra. No item 4 da segunda entrevista, Ana esclareceu que para estudar os hábitos de higiene dentária dos alunos da escola, não necessitava de realizar um censo, mas apenas uma sondagem. Com a realização de uma sondagem considera que se poupa tempo e por sua vez, como pensou que não haveria uma grande variabilidade no âmbito dos hábitos de higiene dentária, a recolha de informação junto de um conjunto razoável de alunos permitiria aceder a uma parte significativa da diversidade dos comportamentos. Ana evidenciou distinguir, no contexto da situação em estudo, os conceitos população e amostra. No teste final, nos itens 5.1 e 5.2 é evidente que Ana identificou corretamente a população e a amostra em estudo, ao responder que a população eram “*os 1560 alunos da escola*” e ao indicar como procedimento de seleção da amostra a escolha do acaso de 160 alunos. Ainda no teste final, nos

itens 6.1 e 6.2, Ana considerou que a população em estudo eram as maçãs de Alcobaça e a amostra era o conjunto das 51 maçãs que foram medidas.

Amostras representativas e ponderação de elementos que afetam a representatividade

Amostras representativas. Perante a necessidade realizar um estudo estatístico sobre os hábitos de higiene dentária, no item 4 da segunda entrevista, Ana considerou que deveria recorrer a uma amostra de alunos da escola em vez de proceder ao estudo de toda a população, pois entre um conjunto razoável de alunos decerto iria obter diferentes respostas que correspondessem à diversidade de hábitos existentes na população em análise. A seleção da amostra passou pelo sorteio de três números de turma e por inquirir todos os alunos com esses números, em cada turma, tendo em consideração que o total de inquiridos deveria ser cerca de 100 elementos. No item 5.2 do teste final, Ana optou por uma amostragem aleatória simples “*escolher 160 alunos da escola ao acaso*”, embora não tenha esclarecido de que modo iria por em prática a seleção aleatória.

Ponderação de elementos que afetam a representatividade. Quando inquirida sobre um procedimento que conduzisse a uma amostra enviesada, Ana, no item 3.1 do teste final, indicou a escolha de uma amostra restrita a um subgrupo da população “*Escolher jovens de um só género*”. Argumentou que deste modo a amostra não seria representativa da população pois a variável em estudo – hábitos de sono dos jovens – relaciona-se com o género, deu como exemplo a possibilidade de em determinadas faixas etárias dos adolescentes, os jovens do sexo masculino poderem ter mais possibilidade de sair à noite que os elementos femininos, pelo que ao restringir o estudo a um único género poderia enviesar os resultados do estudo.

Análise crítica de estudos estatísticos. A Ana continuou a demonstrar-se capaz de fazer uma análise crítica de estudos estatísticos. No item 2 a) da segunda entrevista, num contexto de natureza menos familiar, para alunos da sua faixa etária – estudo relativo às condições de trabalho, apresentou argumentos para a rejeição de amostras não representativas, incluindo possíveis enviesamentos dos dados recolhidos (Figura 95).

a) A melhor previsão da percentagem será a do jornal 2 pois no jornal 4 foram escolhidas ao acaso pessoas de 18 anos numa rede social mas quem costuma usar as redes sociais são pessoas jovens que por vezes tem poucas condições de trabalho o que influenciara o estudo no jornal 3 é a mesma razão e mas em relação ao envio de sms e no jornal 1 como a ~~amostra~~ ^{sondagem} é feita numa determinada hora as pessoas que aí passaram poderão ser muito poucas, podendo até trabalhar no mesmo lugar o que influenciara a sondagem pois poderá haver muitas pessoas com as mesmas condições de trabalho

Figura 95. Resposta ao item 2 a) da 2.^a entrevista.

Breve síntese. No final da unidade de ensino, Ana continuou a mostrar-se capaz de decidir, face a um estudo estatístico, se deveria realizar um censo ou uma sondagem. A ponderação do tempo de realização de um censo e a capacidade de captar a diversidade populacional da variável em estudo foram o mote para a decisão de recorrer ao estudo da população ou a uma parte desta. Perante uma situação de maior complexidade e abstração, continuou a ser crítica face a estudos envolvendo amostras não representativas, identificou e ponderou elementos que afetavam a representatividade de uma amostra em relação à respetiva população e recorreu a algumas situações concretas para ilustrar os diferentes elementos que ponderou. Quanto à formulação de questões, continuou a manifestar dificuldades na formulação de questões isentas de dupla interpretação, mas quando interpelada e instigada a refletir sobre as questões formuladas foi capaz de propor novas redações mais claras, com menor possibilidade de dupla interpretação por parte dos inquiridos e mais adequadas à recolha da informação pretendida, possibilitando a posterior resposta ao problema estatístico em análise.

6.4.2. Análise de dados

Natureza dos dados e representações adequadas. Depois do final da unidade de ensino, tendo sido pedido um esboço do estudo para conhecer as práticas de higiene oral (item 4 da 2.^a entrevista), Ana optou por analisar as seguintes variáveis: (i) Número de escovagens diárias; (ii) utilização de fio dentário pelo menos uma vez por dia (sim/não) e (iii) tempo de corrido entre duas idas ao dentista, tendo procedido à sua correta categorização. No âmbito da organização e tratamento de dados, considerou que para apresentar os dados recolhidos com base na primeira questão, recorreria a um gráfico de barras (para os valores 0, 1, 2, 3, 4, mais de 4 vezes), para os dados da segunda utilizaria

um gráfico circular, pois apenas seriam necessários dois sectores (sim/ não) e no caso dos dados recolhidos pela terceira questão, iria recorrer a um histograma (cujas classes seriam menos de 3 meses, entre 3 e 6, de 6 meses a 1 ano, mais de um ano). Quando questionada sobre quais as medidas estatísticas que utilizaria para descrever os dados, indicou a média e a moda, para as variáveis quantitativas e, para a variável qualitativa, indicou a moda e a frequência relativa de uma determinada modalidade.

Construir, analisar e interpretar representações dos dados

Representação tabelar. No teste final, Ana continuou a não manifestar dificuldades em proceder à leitura e interpretação da informação veiculada pela tabela de frequências absolutas e relativas (em percentagem), tendo completado os valores em falta corretamente.

Gráfico de barras. Ana, na segunda entrevista, face a dois gráficos de barras (Item 2b)) argumentou que o motivo principal da diferença entre as duas representações gráficas, dos mesmos dados, era a escolha de uma unidade de escala distinta.

Ana – Estou a ver os gráficos e no 1.º gráfico o número de queixas vai de 1000 em 1000, faz que pareça que há muitas queixas e no 2.º gráfico como vai de 2000 em 2000 faz com que pareça que há poucas queixas, porque as barras ficam pequeninas.

Como Ana não recorreu ao argumento apresentado na primeira entrevista, perante um situação similar, mas de menor complexidade, optei por lhe propor a construção do gráfico de barras que representasse os dados recolhidos de modo adequado. Começou por reproduzir o gráfico 1, mas constatou que a altura da barra correspondente ao número de queixas de 2003 tinha metade da altura da barra correspondente ao número de queixas de 2004, e a diferença entre o número de queixas registadas entre 2003 e 2004 era apenas de 1000, face a isso apercebeu-se que a escala da representação deveria iniciar-se em 0, em vez de ser 7 000 (Figura 96). Não referiu o espaçamento entre as barras quando discutiu a construção dos gráficos.

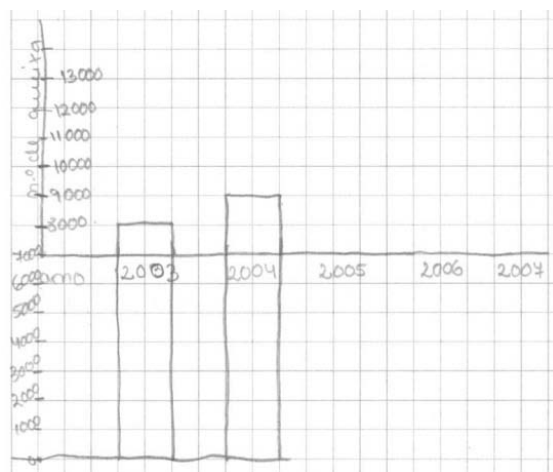


Figura 96. Resposta ao item 2b) da 2.^a entrevista.

Sectograma. Na segunda entrevista, Ana mostrou-se capaz de fazer uma leitura literal da informação representada pelos setogramas ao responder ao item 3 a1) “A prova preferida em ambas as escolas é a de Bloco e que na escola A a menos preferida é a Dificuldade, mas na escola B é a de Velocidade.”. Quando questionada, no item 3 a2), se poderia afirmar que o número de alunos da escola A que prefere competir em provas do tipo Bloco é superior ao número de alunos da escola B a preferir o mesmo tipo de prova, Ana reconheceu ser impossível afirmar que um sector de um sectograma está associado a um maior número de elementos da amostra/população que um sector de outro sectograma dado que desconhece a dimensão dos dois conjuntos de dados (Figura 97).

a2) Não poderemos afirmar que o n.º de alunos na A que preferem a prova Bloco é superior pois não sabemos n.º total de alunos inquiridos em nenhuma das escolas.

Figura 97. Resposta ao item 3 a2) da 2.^a entrevista.

Histograma. No teste final, perante um conjunto de questões (grupo 2) de leitura e interpretação de um histograma com idêntica complexidade ao apresentado no teste inicial, Ana mostrou-se, uma vez mais capaz de responder corretamente a todas as questões solicitadas. Ainda no âmbito da construção de um histograma adequado à representação de um conjunto de dados, Ana na segunda entrevista, item 3 c), perante 35 dados compreendidos entre 8 e 29 minutos, pensou inicialmente na construção de classe de amplitude 2 minutos iniciando a primeira em 8. Quando constatou que iria construir 12 classes, alegou que seriam muitas classes para o número de dados e optou por duplicar a

amplitude das classes para que o número de classes fosse menor e mais adequado “*Opto por construir um gráfico em que a amplitude de cada classe é de 4, [8,12[, [12, 16[, ...*”.

Determinação de medidas estatísticas e sua interpretação

Moda, média e mediana. Ana, no item 3b) da segunda entrevista, foi ágil na determinação das medidas da média e mediana e mostrou-se capaz de privilegiar o uso da média em detrimento da mediana em distribuições com dados atípicos (*outliers*) (Figura 98).

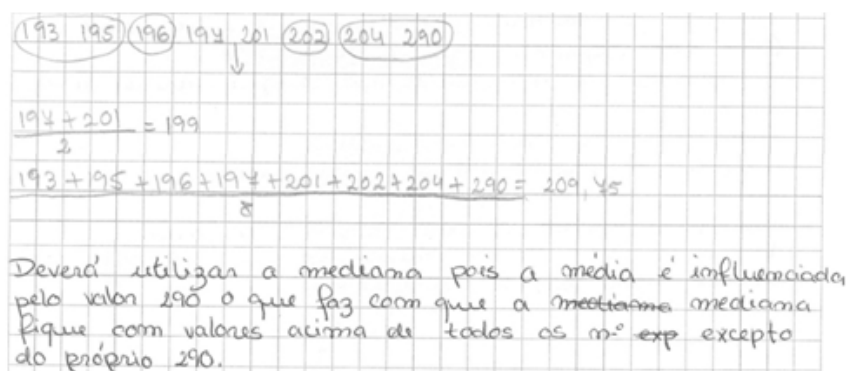


Figura 98. Resposta ao item 3b) da 2ª entrevista.

Comparar distribuições e validar conclusões. Uma vez mais, Ana mostrou-se capaz de proceder à comparação de duas distribuições, recorrendo a conceitos e ideias matemáticas para fundamentar as suas conclusões. Na segunda entrevista, relativamente à comparação de dois diagrama de extremos, Ana considerou que são as salsichas de aves as que contêm menos sódio, dado que “*o valor do mínimo e máximo do primeiro diagrama de extremos e quartis são pequenos comparados com os do 2º diagrama, visto que o 1º quartil também é menor que no 2º diagrama e como a mediana e o 3º quartil também são menores*”. Face à necessidade de comparar dois conjuntos de dados recolhidos em duas fases de uma experiência (Item 1b) da 2.ª entrevista), Ana optou por recorrer apenas ao uso de medidas de tendência central (mediana e média) não tendo feito qualquer referência à amplitude total e/ou interquartil para fundamentar a sua resposta (Figura 99).

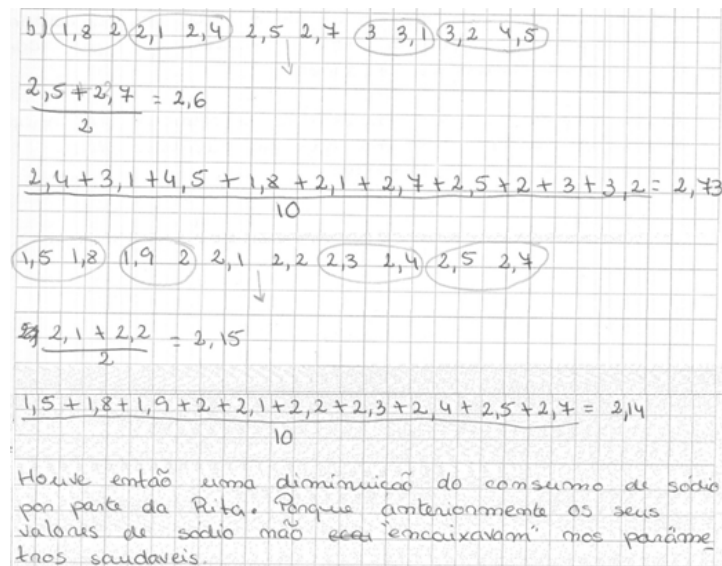


Figura 99. Resposta ao item 1b da 2ª entrevista.

No teste final, item 4, Ana apresentou um argumento para o facto de a conclusão do empresário ser incorreta, dado ter-se baseado exclusivamente na comparação das médias das distribuições. Considerou que se deveria concluir que “O grupo azul não é necessariamente o mais criativo pois há uma maior percentagem de pessoas no grupo verde com a criatividade acima dos 105 pontos, se o empresário pretende trabalhar com pessoas mais criativas deve contratar os melhores do grupo verde.”.

Transformar uma representação numa outra representação

Representação tabelar – gráfico de barras. No item 1.2 (teste final), que solicitava a interpretação da escala de construção do gráfico de barras, Ana mostrou-se capaz de identificar a unidade de escala utilizada na construção.

Diagrama de caule e folhas – histograma. No item 3 d) da segunda entrevista, Ana procedeu a uma associação incorreta, pois não fez a devida leitura da informação que foi registada no eixo vertical dos histogramas – frequência relativa – e assumiu que caules com o mesmo número de folhas nos diagramas de caule e folhas correspondem a classes com igual altura (Figura 100).

e) É o conjunto A que corresponde a representação pois por exemplo nos diagramas de caule e folhas na classe dos 170 há 4 dados em ambos os diagramas como no conjunto B a altura dessas barras é diferente implica que não há o mesmo nº de pessoas com aquelas alturas.

Figura 100. Resposta ao item 3d) da 2ª entrevista.

Diagrama de caule e folhas – diagrama de extremos e quartis. No item 6.4 do teste final, Ana determinou corretamente quer a mediana, quer o 1.º e 3.º quartil, escolheu uma escala adequada aos cinco números de resumo e procedeu à correta representação do diagrama de extremos e quartis correspondente.

Histograma – diagrama de extremos e quartis. No item 5.3 do teste final, Ana estabeleceu uma correspondência correta entre os histogramas e os diagramas de extremos e quartis, ao indicar a seguinte associação $A \rightarrow 1$, $B \rightarrow 3$ e $E \rightarrow 2$ tendo fundamentado a sua escolha com base na forma e na localização da mediana da distribuição “*pois o gráfico 1 vai ficando com as barras mais pequenas, o gráfico 2 pois a barra do meio é a mais pequena então é o meio [da distribuição e é onde se localiza] o [2.º] quartil, o gráfico 3 vai a subir*”.

Generalizar as conclusões para a população.

No item 4 (2ª entrevista), à semelhança da primeira entrevista, Ana argumentou que as conclusões se deveriam generalizar à população, uma vez que 100 alunos era uma amostra de dimensão razoável e, dado que o processo de recolha tinha sido aleatório, se poderia assumir que os dados recolhidos eram uma “mostra” da potencial diversidade dos hábitos de higiene dentária. No teste final, Ana não responde aos itens 1.4 e 6.6, quando questionada pela professora sobre o facto, alegou que é mais fácil explicitar oralmente os motivos da aceitação ou da generalização das conclusões extraídas da amostra para a população, do que proceder à sua redação.

Breve síntese. No final da unidade de ensino, perante situações de maior complexidade e abstração, Ana continuou a mostra-se capaz de ler, interpretar e avaliar diferentes representações. Face a um problema estatístico, identificou variáveis passíveis de mobilizar a recolha de um conjunto de dados adequados para responder à situação, depois refletiu sobre a natureza dos dados que poderia recolher, o modo como os deveria representar e resumir tendo em consideração a sua comunicação. Recorreu à inter-

pretação da forma, à identificação do centro e à dispersão da distribuição dos dados apresentados numa determinada representação para a transformar, na maioria das situações, numa outra representação. Ana continuou a ter um bom desempenho na comparação de duas distribuições, reconheceu características gerais tais como forma e centro de uma distribuição, e recorreu menos à utilização de medidas de dispersão. Quanto à escolha da medida estatística mais adequada para resumir a informação contida nos dados, Ana optou pela mediana, quando se apercebe da existência de valores atípicos. Ana reconheceu que o valor da média seria maior que a generalidade dos dados do conjunto quando o valor atípico era o máximo da distribuição. Ana continuou a manifestar dificuldade em extrapolar as conclusões de sondagens à população em estudo e, embora no decorrer da unidade tenha sido capaz de argumentar oralmente a validade das extrapolações em cerca de metade das situações, o seu desempenho no teste final, é muito fraco, o que se explica, por um lado, pela dificuldade em articular a quantidade e a qualidade da amostra versus a variabilidade populacional, que intuiu para a característica em estudo, e, por outro lado, pela dificuldade em explicitar por escrito os seus argumentos.

6.5. Síntese global

De forma geral, a Ana realizou com sucesso as tarefas propostas, revelando empenho durante o estudo do tema. Mostrou-se capaz de esboçar um plano de desenvolvimento de um estudo estatístico, tomando a decisão de realizar um censo ou uma sondagem após ter ponderado o tempo disponível para a realização do estudo e a variabilidade das características em análise. Revelou ainda dificuldades na formulação de questões de pesquisa claras e que conduzissem a uma interpretação única por parte dos inquiridos. Argumentou os seus pontos de vista recorrendo por vezes a ilustrações concretas, quer para dar ênfase ao(s) elemento(s) que pondera quanto à afetação da representatividade, quer para explicitar a sua opção de amostragem de modo a minimizar possíveis fontes de enviesamento. Aceitou e recorreu à utilização de amostras aleatórias simples, embora sugerissem em algumas situações o processo de amostragem estratificado, em parte por sentir dificuldades em por em prática a seleção aleatória simples e em parte pela visível organização vertical dos alunos, ao longo dos anos de escolaridade e pela organização horizontal em grupo turma da população “alunos da comunidade escolar”. Segundo as categorias do desenvolvimento do conceito da amostra de Watson

e Moritz (2000), a aluna enquadra-se na categoria mais elevada, designada por “Grandes amostras sensíveis a enviesamento”.

Ana mostrou-se capaz de reconhecer e categorizar conjuntos de dados e de escolher uma forma de representação adequada à sua análise. Na análise de dados identificou características gerais como a forma, o centro e a dispersão da distribuição, evidenciando uma visão agregada dos dados. Perante o quadro de análise do raciocínio estatístico sobre os dados e as representações de Garfield (2002), mostrou-se capaz de identificar as medidas estatísticas centrais e de dispersão mais adequadas à análise de um conjunto de dados e de compreender que determinada medida central ou de dispersão fornece informação sobre um conjunto de dados, embora nem sempre incluísse no resumo descritivo, pelo menos uma medida estatística de cada tipo. Procedeu com alguma destreza à comparação de duas distribuições reconhecendo medidas centrais e de dispersão pertinentes a cada situação, privilegiando no entanto o uso de medidas centrais. Utilizou medidas de dispersão na comparação de duas distribuições quando existe uma representação em que predomina uma visualização ordenada dos dados.

Nos vários estudos estatísticos, no momento de extrair conclusões, as respostas de Ana tiveram as seguintes características: (i) recorre às suas próprias opiniões e ignora os dados dos estudos realizados, (ii) considera possível generalizar as conclusões face a uma amostra de dimensão razoável e de seleção aleatória, quando intui que o processo de amostragem terá capturado a variabilidade populacional e (iii) recusa-se a tirar conclusões ou a desenvolver previsões, uma vez que a variabilidade populacional não terá sido “retratada” por meio do processo de amostragem.

Tendo em atenção o desempenho de Ana, a unidade de ensino, valorizando a realização de pequenas investigações estatísticas e análise crítica de estudos estatísticos (Martins & Ponte, 2010), mostrou-se exequível. O desenvolvimento que este trabalho suscitou em Ana foi determinante para a sua compreensão da importância do planeamento estatístico, nomeadamente no progressivo reconhecimento de variabilidade, aspecto fundamental para a compreensão da natureza e do papel da Estatística e para a possibilidade de consolidar e articular conceitos e representações estatísticas. No entanto, tal como defendem Konold e Higgins (2003), no tema do planeamento estatístico, mantém-se a necessidade de propor mais situações de discussão de questões estatísticas e sua conversão em questões de inquérito/pesquisa.

Capítulo 7

O Caso de Rui

7.1. Apresentação

No momento em que o estudo se realizou, Rui tinha 14 anos, vivia com a mãe e a irmã, maior de idade, perto da escola. Adorava a prática desportiva, tendo começado por participar no grupo de basquetebol da escola, e integrava a equipa de iniciados de um grupo desportivo da zona. No restante tempo livre, gostava de jogar jogos na consola ou computador. Estava integrado na turma, tinha uma boa relação com a generalidade dos colegas de turma. Tinha sido bem-sucedido ao longo de todo o seu percurso escolar e as suas disciplinas preferidas eram Inglês, Ciências Naturais e Ciências Físico-Químicas. Afirmou que embora raramente estudasse para os testes, obtinha classificações finais de “Bom” ou “Muito bom”. Quanto à progressão de estudo, indicou que pretendia seguir o curso de Ciências e Tecnologias, e no final do 12.º ano candidatar-se à faculdade num curso da área da Química. No âmbito da disciplina de Matemática, considerou que começou por ser um dos melhores alunos da turma, no 1.º ciclo, mas o seu desempenho foi decrescendo ao longo do 2.º ciclo, sendo à data do estudo um aluno irregular. Afirmou que, tratando-se de tópicos da Álgebra, o seu desempenho era fraco, noutras matérias era satisfatório. Relativamente ao seu modo de estar na aula, considerou-se um aluno conversador e pouco organizado, e admitiu que se distrai facilmente. Como sua professora, reconheci que tinha muita facilidade em se expressar oralmente, embora o fizesse de modo pouco estruturado, constituindo, entre os alunos que voluntariaram, um bom candidato ao nível dos critérios de evidência e abrangência.

7.2. Capacidades de planeamento estatístico e análise de dados no início da unidade de ensino

Nesta secção refiro o desempenho do aluno no âmbito do teste inicial (Anexo 3) e o seu desempenho no trabalho de grupo durante a realização das tarefas 1 e 2 (Anexos 7 e 10).

7.2.1. Planeamento estatístico

Formular de questões. No início da lecionação da unidade de ensino, a análise do trabalho de grupo na questão 6b da tarefa 1, mostrou que Rui procedeu a uma interpretação incorreta do objeto em estudo “projetos quanto ao prosseguimento de estudos”, tendo assumido que se tratar de um estudo relativo aos procedimentos de estudo e os colegas de grupo sustentaram essa interpretação.

Rui - Quantas vezes por dia é que estudas? Não, quantas horas estudas por dia?

J - Que disciplinas preferes?

Ana - Quais [são os] dias em que estudas por semana?

Rui - [Quais] ... os teus métodos de estudo?

Rui - Com quantos dias de antecedência estudas?

D - Quantos dias é que estudas por semana?

Durante a discussão da tarefa 2, Rui interveio e propôs a formulação de duas questões que poderiam ser respondidas com base nos dados da tarefa, “*Qual é a nacionalidade estrangeira mais comum?*” e “*Qual é a idade média?*”, ao ser interpelado por um colega que considera não ser possível responder à segunda questão, Rui argumentou que tendo a data de nascimento e assumindo que os dados se tinham recolhido em 2001 era possível conhecer os dados da variável idade.

Distinguir entre população e amostra. Durante a realização da tarefa 1, na discussão, com os colegas de grupo, sobre o dimensionamento da amostra para o estudo estatístico solicitado (Item 6), Rui evidenciou a compreensão dos conceitos população e amostra.

D - Não que são muitos $6 \times 12 = 60$, dá pouco inquéritos ...

Rui - Pode ser 10 pessoas se forem 10 dá 120.

D - 120 das 523 é pouco.

Rui - Não é nada, mas se achas pouco fazes [o inquérito] a todos.

Ana - 60 assim é pouca gente ...

Rui – Cerca de 500 alunos ... há 20 turmas

Amostras representativas e ponderação de elementos que afetam a representatividade. No item 5.1 do teste inicial, face a uma questão, relativa à escolha de uma amostra, Rui selecionou um processo aleatório de amostragem estratificada – “Questionar 160 alunos que são os números 1, 8, 15, 22 de cada uma das quarenta turmas da escola”, rejeitando outros processos de amostragem passíveis de enviesamento tais como “questionar 160 alunos à porta da escola no início da manhã” ou “questionar 160 alunos das 6 turmas de 12.º ano”.

Durante o trabalho de grupo, constatou que na questão 2 da tarefa 1 – “População e amostra”, a amostra não era representativa dado que o inquérito se realizara à porta de um determinado supermercado e este facto poderia enviesar os dados recolhidos, dado que na sua opinião a publicidade estaria associada aos desejos e às expectativas, pelo que os autores do estudo deveriam diversificar os locais de inquérito de modo a obter dados de distintos contextos socioeconómicos.

Rui - Eles não podem ir a um só local porque Imagina eles deveriam ir por exemplo a um bairro pobre, a um bairro rico, e um bairro médio e um assim ficava ... era mais diversificado ... já era mais diversificado as ideias, do que se tu fores só a um bairro rico ou a [um] bairro pobre.

Na realização do item 3 da tarefa 1, o grupo de Rui constatou que a amostra em causa era representativa dada a sua diversidade e significativa dimensão. Rui tomou a iniciativa de assumir a representatividade da amostra sendo apoiado por Ana, colega de grupo que apresentou o argumento da diversidade da amostra. Rui mostrou-se sensível à questão da significativa dimensão da amostra, concluindo que seria impensável que a amostra não contivesse elementos que frequentavam o refeitório.

Rui - Eu até acho que esta está certa.

Ana e J - Sim esta está ...

Ana - Tipo, tem muita diversidade...

D - Não mas olhem só, pode haver [alunos] números ímpar que não vão almoçar ao refeitório.

Ana - Está bem, mas assim tem-se uma grande diversidade...

Rui - Assim tem-se metade da escola inteira, achas que metade da escola muita gente não vai comer ao refeitório...

Ana - Assim está bom porque tens alguns [que] vão, outros [que] não vão.

Perante a necessidade do grupo indicar um processo de amostragem para um estudo sobre os projetos quanto ao prosseguimento de estudos dos alunos de 9.º ano de uma escola (item 6a da tarefa 1), Rui defendeu uma seleção aleatória estratificada dos elementos a inquirir, por sua vez, a colega Ana defendia uma seleção aleatória simples, o grupo chegou a um impasse, a como nenhum conseguiu convencer o outro. Um terceiro elemento do grupo referiu a dificuldade de por em prática um sorteio entre os alunos da escola e sugeriu um seleção aleatória sistemática, análoga à do item 3, seleção de todos os alunos com número par.

Rui - Não pode ser aleatório... porque se não, pode calhar muitas pessoas de uma turma só.

Ana - Pode... está bem, mas como é muita gente isso não faz diferença ...

Rui - Faz, faz, ...

Ana - Não, não faz!

Rui - Então porque é que tu não ... se escolhe 4 de cada turma por exemplo Porque não se escolhe 5, 8, 17, 24 e 20 escolhes esses cinco números ... escolhes de cada turma e fazes questionários a essas pessoas esses cinco números de cada turma.

Ana - Mas isso é a mesma coisa Rui, é a mesma coisa do que se tu fizeres aleatoriamente.

Rui - Eu sei, mas assim é mais fácil e temos a certeza que vai passar em todas as turmas...

Ana - É a mesma coisa que escolhes aleatoriamente 200 pessoas do 9º ano aleatoriamente ...

Rui - Não, assim está a escolher 5 pessoas de uma turma mas aleatoriamente... não do 9º ano inteiro.

Na discussão do item na turma, Rui interveio e argumentou que no caso de se recorrer a um processo de amostragem aleatória simples, haveria a possibilidade de se inquirir mais alunos de uma turma do que de outra, ou obter uma amostra que não incluísse alunos de determinada turma, o que iria por em causa o estudo porque “*Se tu estás numa turma as pessoas influenciam-se umas as outras...por exemplo na nossa turma quase todos vão para ciências...*”. Considerou que um meio de evitar o enviesamento da amostra passava pela escolha de uma amostra que preservasse a representação proporcional dos vários subgrupos turma.

Breve síntese. O trabalho de Rui, no teste inicial e durante a realização das tarefas 1 e 2, mostrou que, no âmbito do planeamento estatístico, o aluno reconheceu possíveis fontes de enviesamento na recolha de dados e ponderou uma variedade de elementos que poderiam afetar a representatividade de uma amostra em relação à respetiva população (ME, 2007, p.60). Rui mostrou-se capaz de analisar um conjunto de dados e de formular questões aos dados, válidas e pertinentes, no âmbito da categoria resu-

mo/descrição (Arnold, 2008). Durante a realização da tarefa 1, as intervenções de Rui evidenciaram uma compreensão dos conceitos população e amostra e, um reconhecimento significativo da variabilidade natural e da variabilidade do processo de amostragem (GAISE, 2005).

7.2.2. Análise de dados

Construir, analisar e interpretar representações dos dados

Representação tabelar. No teste realizado no início da unidade de ensino, Rui demonstrou dificuldades na leitura e interpretação da tabela, tendo completado apenas o espaçamento correspondente ao “número de alunos” ou seja não manipulou a informação relativa às frequências relativas da variável idade.

Histograma. Quanto a alguns aspetos de construção, análise e interpretação de um histograma, Rui mostrou-se incapaz de identificar o número de classes e a respetiva amplitude (Item 2.1), conseguiu fazer uma leitura literal da representação ao identificar o número total de efetivos (Item 2.2) e o número de participantes que realizaram um tempo de prova inferior a x ou superior ou igual a y (Itens 2.3.1 e 2.3.2), mas respondeu incorretamente ao item 2.4 que solicitava a indicação de um possível tempo de prova para o vencedor ao responder “2 minutos”.

Diagrama de caule e folhas. No que diz respeito à leitura e interpretação de um diagrama de caule e folhas (Itens 4.1, 4.2), Rui não manifestou qualquer dificuldade, identificou corretamente o número de dados representados, os extremos da distribuição e evidenciou ter conhecimento de que a representação apresenta os dados ordenados, ao ter recorrido a esta para a identificação do valor que permitia o correto completamento da fase do item 4.3 (Figura 101).

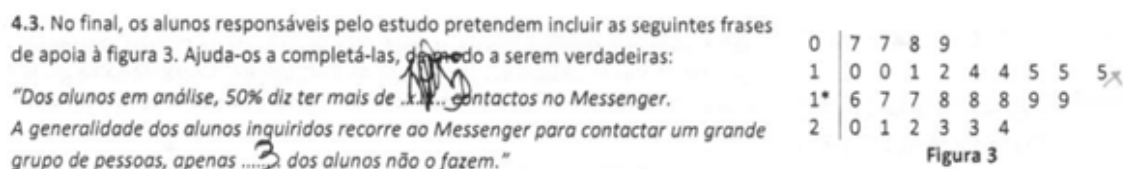


Figura 101. Resposta ao item 4.3 do teste inicial.

Determinação de medidas estatísticas e a sua interpretação

Moda e mediana. Rui identificou, a medida estatística moda, sem dificuldades (Item 1.2 do teste inicial). Procedeu à correta identificação da mediana e à correta interpretação do seu significado quando responde ao item 4.3 (teste inicial).

Comparar duas distribuições e validar as conclusões. Perante a necessidade de comparar duas distribuições, recorrendo à análise conjunta de medidas estatísticas centrais e de dispersão, item 3 do teste inicial, a resposta de Rui evidenciou ter tido dificuldades na interpretação da questão, dado ao ser-lhe pedida a apresentação de um argumento matemático que provasse que o grupo A, com menor valor médio, globalmente tinha reunido melhores resultados Rui argumentou a favor do grupo com maior valor de média (grupo B). Constatou-se ainda que Rui fez uma leitura precipitada de ambos os eixos ao ter escrito “a única negativa que tiveram foi um 49 ... que no grupo A (h)ouve um 0.” (Figura 102).

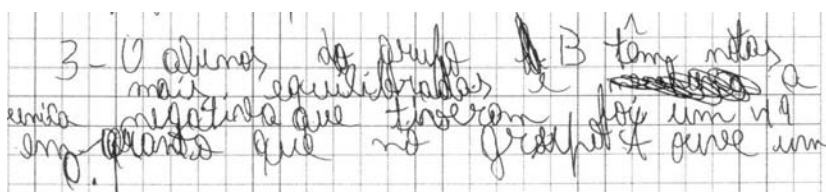


Figura 102. - Resposta ao item 3 do teste inicial.

Transformar uma representação numa outra representação

Diagrama de caule e folhas e diagrama de extremos e quartis. Num item que pretendia aferir a capacidade de determinar as medidas de localização e de utilizá-las para apresentar uma outra representação de um mesmo conjunto de dados, Rui, tal como um terço dos seus colegas mostrou-se incapaz de determinar o 1.º e 3.º quartil, não tendo apresentado qualquer elemento do diagrama de extremos e quartis (Item 4.4 do teste inicial).

Histograma – diagrama de extremos e quartis. Rui não respondeu ao item 5.2 do teste inicial que requeria o reconhecimento das características gerais de uma distribuição (forma, centro e dispersão), representada por um histograma e proceder à respetiva conversão da densidade dos dados nos comprimentos dos vários elementos do diagrama de extremos e quartis.

Generalizar conclusões para a população

No item 1.3 do teste inicial, que se refere à necessidade de estimar a idade dos alunos do 8.º ano de uma escola com base numa amostra de 36 alunos (cuja dimensão corresponde a cerca de 14% da população), Rui considerou que se deveria recorrer ao valor modal ao responder “13 anos por que é idade que tem a maioria das pessoas no 8.º ano”, dada o modo como formula a sua resposta, não fica claro se teve em consideração a sua experiência pessoal ou se teve em consideração a tendência etária da amostra. No item 4.5, considerou que no estudo sobre a utilização do Messenger, com base numa amostra de 30 alunos do 8.º ano, uma conclusão válida e passível de ser generalizada à população escolar seria que “cerca de 10% dos alunos da escola não utilizam o Messenger”, pois defendeu que uma grande parte dos alunos usa as tecnologias e que é possível que sejam apenas cerca de 10% a não utilizar o Messenger para contactar os colegas. Ao ser questionado quanto ao facto de não ter recorrido a seleção da opção “Cerca de 50% dos alunos do 8.º ano contactam, via Messenger, com mais de 15 pessoas.”, uma vez que no item 4.3 tinha concluído que entre os alunos em análise cerca de 50% diz ter mais de 15 contactos no Messenger, alegou que não se apercebeu que poderia ter feito tal associação, ao refletir sobre o assunto considerou ser possível assinalar essa opção dado que se escolheram alunos em todas as turmas de 8.º ano.

Breve síntese. No teste inicial, no que diz respeito à construção, análise e interpretação de distintas representações dos dados, Rui não foi capaz de completar a coluna das frequências relativas numa representação tabelar, mas foi capaz de fazer uma leitura direta de um histograma e de um diagrama de caule e folhas. Nas situações em que se solicitou uma leitura indireta, Rui manifestou dificuldades de interpretação. Quando lhe foi pedida a comparação de duas distribuições e a transformação de uma representação num outra, mostrou-se incapaz de mobilizar as características gerais das distribuições (forma, centro e medidas de dispersão). Quanto às medidas estatísticas, identificou corretamente a moda e a mediana, mas mostrou-se incapaz de determinar os quartis da distribuição. Face a uma distribuição de dados com elevada amplitude, dada a existência de um valor atípico, Rui considerou que a média seria uma medida estatística adequada e representativa da distribuição, bem como ser o único elemento de comparação da distribuição com uma outra de menor amplitude e com maior valor médio. No âmbito do raciocínio estatístico sobre as medidas, cometeu dois dos equívocos mais comuns envolvendo a média. Perante a necessidade de avaliar inferências e previsões,

na maioria das situações, recorreu à sua experiência pessoal para aceitar ou recusar a inferência ou a previsão.

7.3. Capacidades de planeamento estatístico e análise de dados durante a unidade de ensino

Nesta secção refiro o desempenho do aluno, no âmbito do trabalho de grupo/par, durante a realização das tarefas 3 a 6 (Anexos 11 a 14) e o seu desempenho na tarefa proposta na primeira entrevista (Anexo 5).

7.3.1. Planeamento estatístico

Formular questões. Na primeira entrevista, perante uma primeira fase de planeamento de um estudo na escola, sobre o cumprimento da recomendação de consumo de água, Rui começou por escolher como variável o “número de peças de fruta consumidas”, mas para uma melhor clarificação dos dados recolhidos considerou que deveria questionar cada aluno quanto ao “Número de peça de fruta consumidas nos últimos dois dias? E quais as peças de fruta consumidas?”. Ainda durante a realização da tarefa 4, Rui participou na formulação de questões, para a caracterização dos alunos do 3.º ciclo da escola, propondo as seguintes questões de descrição “Gostas de basquetebol?” e “Praticas algum desporto? Qual?”, a segunda questão poderia ter uma reformulação de modo que todos os inquiridos interpretassem o termo “praticas” do mesmo modo, mas para Rui tal não era necessário pois todos os colegas perceberiam que se tratava da prática desportiva após o período letivo, e que deveria ocorrer pelo menos uma vez por semana.

Durante a lecionação da unidade de ensino, num primeiro momento em que se discutiram situações que são objeto de estudo Estatística (síntese da tarefa 3), Rui revelou-se capaz de reconhecer a variabilidade natural, tendo por base a análise da pertinência de “Estudar o valor gasto pelos encarregados de educação na aquisição dos manuais escolares do 8.º ano da nossa escola”, a generalidade dos alunos considerou que a situação não era pertinente e de interesse estatístico dado que o conjunto de manuais escolares era idêntico para cada educando, pelo que o valor gasto era fixo. Rui não concordou, alegou que o valor gasto não era necessariamente fixo, pois nem todos os pais compram no mesmo sítio, havendo locais que fazem um desconto e/ou permitem o pagamento a

crédito pelo que o valor gasto na compra dos manuais escolares do 8.º ano no contexto de uma determinada escola, varia e, como tal, era passível de constituir uma situação de interesse estatístico.

Distinguir entre população e amostra. Tendo em consideração os vários estudos estatísticos realizados, nomeadamente nas tarefas 4 e 6 e nos grupos 2 e 5 da primeira entrevista, Rui evidenciou compreender os conceitos de população e amostra, reconhecendo que nestes estudos, a população era composta pelos alunos elementos da comunidade escolar e que a amostra era um subconjunto de alunos dessa mesma população. No final da primeira entrevista solicitei ao aluno que me indicasse o que significava para si amostra representativa, tendo respondido que era uma parte da população que conseguisse representar a população total, que se constituía como uma pequena mostra da população, e que, no caso de serem alunos, deveria incluir indivíduos dos dois géneros, das várias faixas etárias.

Amostras representativas e ponderação de elementos que afetam a representatividade

Amostras representativas. Durante a realização dos 2 a) e item 5.2 da primeira entrevista, nos quais se solicita a definição de uma amostra representativa, para o estudo do consumo semanal de refrigerantes e o estudo sobre o cumprimento da recomendação do consumo de água, respetivamente, Rui considerou que para obter uma amostra representativa da população em estudo deveria selecionar os seus elementos de modo aleatório. Para o item 2 a) optou por indicar como resposta “*Se houver(em) cerca de 100 turmas, eles podiam escolher dois alunos de cada turma escolhido alheatoriamente, mas um rapaz e uma rapariga de cada turma*” e, no item 5.2 optou pela escolha aleatória de 5 rapazes e 5 raparigas de cada turma. Quando questionado sobre o facto de ter especificado que as amostras deveriam ter igual número de alunos de cada género, alegou que é possível que o género feminino se preocupe mais com questões de bem-estar e saúde, pelo que o seu consumo de refrigerantes pudesse ser inferior ao do sexo masculino, no primeiro caso, e o cumprimento da recomendação de consumo de água pudesse ser maior, no segundo caso. Assim ao não respeitar a proporção dos géneros existentes na população, estar-se-ia a por em causa ambos os estudos.

Ponderação de elementos que afetam a representatividade. Quando inquirido sobre um procedimento que conduzisse a uma amostra enviesada, Rui indicou, à semelhança do que o seu grupo de trabalho apontou na tarefa 4 e 6, a escolha de um conjunto

de alunos de determinado subgrupo populacional - amostra apenas composta por alunos do sexo masculino. Considerou que amostra que não seria representativa da população dado que as características em estudo - consumo semanal de refrigerante e cumprimento da recomendação do consumo de água – se relacionam com género do inquirido e não se estaria a respeitar a proporção populacional de cada género, que Rui assumiu ser em qualquer circunstância cerca de metade do género feminino e metade do género masculino.

Análise crítica de estudos estatísticos. Durante a lecionação da unidade de ensino, Rui mostrou-se capaz de fazer uma análise crítica de estudos estatísticos (Item 4a) da primeira entrevista), atendendo ao uso de amostras não representativas, e apresentou para cada caso um argumento para a sua rejeição. Face o facto de existirem dois estudos com o recurso a um mesmo processo de amostragem válido, argumentou que se deveria considerar que a melhor previsão está associada ao estudo que utilize a amostra aleatória de maior dimensão pois deste modo, aceder-se-ia a uma maior porção da população, o que era preferível.

Rui - Na 1, eu rejeitei... Porque ai não é ao acaso, é mesmo... É na rua mesmo ao pé do jornal... Por exemplo, se houve um ginásio mesmo ao lado do jornal... Há mais probabilidade das pessoas que eles vêm lá... Andarem grande parte no ginásio do que estarem a escolher ao acaso. Depois a 4 não escolhi porque... Foi feita em base de 1000 leitores que ligaram para lá... Então se ligaram para lá... Por exemplo, se isto for um jornal de desporto, os leitores que ligaram para lá são todas pessoas interessadas em desporto e grande parte pode fazer desporto por isso também acho que não é uma boa... Está enviesada.

Rui - No jornal 3 tanto como no jornal 2... Não há grande diferença entre estes dois jornais, tanto um como o outro pode ser bem feito ... Estão os dois estão bem feitos... Só o 1 e 4 é que estão, para mim, completamente errados.

Prof^a - Mas como te dizem para escolher qual é o jornal, em singular entre o 2 e o 3... Optavas por um deles?

Rui - Eu só optava pelo 3 por que tem maior número de pessoas... Assim está... Abrange uma maior área maior ... mais gente.

Breve síntese. Durante a lecionação da unidade de ensino, Rui mostrou-se capaz de identificar a população e a amostra em análise, nas várias situações propostas. Identificou e ponderou elementos que afetam a representatividade de uma amostra em relação à respetiva população em diferentes contextos. Nas diferentes situações, à falta de uma linguagem formal, recorreu a exemplos para ilustrar os elementos que ponderou. Perante a análise de um conjunto de sondagens, revelou reconhecer os estudos que recorreram a amostras não representativas e justificou o possível enviesamento. Reconheceu ainda

que o aumento da precisão de uma sondagem estaria associado a uma maior dimensão da amostra aleatória. Face ao desempenho de Rui na análise e discussão de distintos processos de amostragem, é possível considerar que se encontra na categoria mais elevada da hierarquia do desenvolvimento do conceito de amostra de Watson e Moritz (2000), dado que reconheceu amostras não representativas, sugeriu amostras com dimensão superior a 20 elementos, selecionadas por um processo de distribuição estratificada com a preocupação de evitar enviesamento. Quanto à formulação de questões de pesquisa, o aluno optou por formular questões de descrição e manifestou alguma dificuldade em evitar a formulação de questões isentas de múltiplas interpretações.

7.3.2. Análise de dados

Natureza dos dados e representações adequadas. Rui ao ser inquirido quanto à categorização das variáveis em estudo no item 5 b) (1.ª entrevista), conseguiu diferenciar as variáveis quantitativas das qualitativas. Entre as variáveis quantitativas foi incapaz de identificar as discretas e as contínuas. Quando reflete no modo como iria proceder para organizar e tratar os dados, admitiu que para as variáveis qualitativas iria construir um gráfico circular e para cada uma das variáveis quantitativas iria construir um diagrama de extremos e quartis, considerou ainda que iria recorrer à moda no caso das variáveis qualitativas e à média no caso das variáveis quantitativas. Acrescentou ainda que embora tenha que determinar a mediana para proceder à construção do diagrama de extremos e quartis, para as variáveis quantitativas, a medida de resumo preferível em termos de comunicação dos resultados seria a média.

Construir, analisar e interpretar representações dos dados

Gráfico de barras. No item 1a) da primeira entrevista, Rui rapidamente se apercebeu que a representação era enganosa, respondendo “*não é aceitável porque entre 2008 e 2010 só participam mais 10 pessoas.*”. Num primeiro momento, não referiu o motivo da leitura errada por parte do coordenador, e após ser questionado porque tal teria ocorrido, comentou “*Não é o dobro porque se fosse o dobro tinha que ter aqui assim (eixo do número de alunos) uma marca com 1010.*”, ao ser novamente questionado em relação ao motivo da leitura errada, Rui respondeu “*Ah, aqui assim (eixo do*

número de alunos) *eles cortaram, começaram logo a partir de 505, não começaram a partir do zero por que se não isto não ia dar o dobro .. ia dar 10*".

Sectograma. Na primeira entrevista (Item 2c)), Rui identificou o erro processual, dos autores do estudo na determinação da amplitude do ângulo correspondente a uma "fatia" do gráfico, tendo feito o seguinte comentário "*ele está a fazer isto para a percentagem ou não ... ah, já sei, ele está a fazer como se fosse 100, ele tinha que fazer 360*" e a sua resposta "*A incorreção que foi cometida foi que o aluno estava a determinar a percentagem e não a amplitude.*"

Histograma. Relativamente à tomada de decisões no âmbito da construção de um histograma (Item 4.b da primeira entrevista), Rui decidiu organizar os dados em 11 classes de amplitude 5, pois na eventualidade de optar por uma amplitude de 10, iria construir menos de seis classes e se escolhesse um outro valor, como por exemplo amplitude 7, iria dificultar a leitura do histograma por parte dos leitores do jornal. Durante a realização da tarefa 4, Rui considerou que o eixo horizontal do histograma da distribuição das alturas, poderia ter as classes etiquetadas do seguinte modo: 140-149; 150-159; 160-169; 170-179; 180-189, alegou qualquer pessoa refere a sua altura em centímetros às unidades, sendo contrariado pelos colegas que lhe apresentam dados que tem valores tais com 178,5 cm, aceita uma designação continua das classes do histograma. Ainda durante a realização da tarefa 4, Rui sugeriu ao grupo a inclusão de um comentário relativo ao aglomerado modal do histograma da distribuição da altura no item 11 (relatório do estudo estatístico).

Diagrama de caule e folhas. Na tarefa 4, Rui e os seus colegas de grupo, recorrem ao diagrama de caule e folhas para uma primeira redução dos dados das variáveis quantitativas em estudo e para proceder à determinação das medidas de localização da variável altura, no entanto cometeram a incorreção de duplicar o valor do caule que apresenta uma grande quantidade de folhas.

Diagrama de extremos e quartis. No item 5.2 (primeira entrevista) Rui defendeu a opção da escolha de um diagrama de extremos e quartis, para as variáveis quantitativas, em detrimento do diagrama de caule e folhas e do histograma, por que considera que deste modo era possível reconhecer o centro da distribuição, visualizar a amplitude total da distribuição, e o reconhecer o "meio" ou seja a concentração ou dispersão de 50% dos dados em torno da mediana.

Rui – Para além das pessoas conseguirem ver mais facilmente, mostra o centro mais ou menos das pessoas [dos dados], também acho que o diagrama mostra desde as pessoas que bebem menos, às que bebem mais, e depois o meio, mais ou menos as pessoas [os dados] do meio e acho que assim as pessoas podem ver mais ou menos a média, e não acho que mais nenhuma, ... nenhuma representação mostra a média, bem esta também não mostra a média, mas a mediana.

Determinação de medidas estatísticas e a sua interpretação

Moda, média e mediana. Rui, face ao item 3.b) da primeira entrevista, que solicitava a escolha da medida estatística adequada e que representasse o conjunto dos dados recolhidos, começou por considerar que a média (75,375 sms) e a mediana (83,5 sms), seriam as medidas estatísticas de resumo mais adequados aos dados recolhidos, embora tenha hesitado na utilização da mediana porque a sua determinação não envolvia o valor atípico (16 sms). Rui antecipou o efeito do dado 16 no valor da média (seria inferior a 80 mensagens), mas considerou que não deveria negligenciar qualquer dado estatístico recolhido. Posteriormente, refletiu ainda se deveria utilizar a moda, ao reconhecer que a distribuição era bimodal (83 e 85 com frequência dois). Concluiu que os valores da moda também poderiam ser adequados dado que tomam valores próximos dos restantes da distribuição, acrescentou ainda que noutras distribuições a moda poderia não ser adequada (se a moda tomasse um valor mais díspar que os restantes dados). Ainda, durante a análise e interpretação dos diagramas de extremos e quartis (Item 3.a)), Rui cometeu a incorreção de se referir à mediana como sendo o número médio de variável em estudo.

Comparar duas distribuições e validar conclusões. Rui, quando se viu confrontado com a comparação de duas distribuições de desempenho de 10 atletas, antes e depois de duas semanas de treino (Item 1b), considerou que o treino teve efeito no desempenho dos atletas, pois em 80% dos atletas o desempenho tinha melhorado, alega que não sentiu a necessidade de calcular a média da altura alcançada antes e depois porque mais de metade dos atletas revelaram uma melhoria bastante satisfatória. No item 3a) (1ª entrevista), face a diagramas de extremos e quartis de duas distribuições relativas ao número de mensagens, Rui identificou a operadora Edfon, cujos utilizadores recorrem mais a mensagens, e argumentou que tal conclusão se deve ao facto dos valores assumido pelos quartis (1.º, 2.º e 3.º quartis) na distribuição da Edfon serem superiores

aos correspondentes valores na distribuição da Telepat (Figura 103) e acrescentou oralmente que “só o primeiro quartil da Edfon é maior que a mediana da Telepat”.

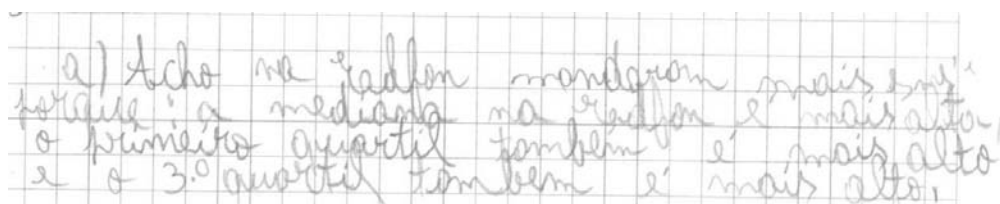


Figura 103. Resposta ao item 3 a) da primeira entrevista.

Generalizar conclusões para a população

No item 1b) da primeira entrevista, Rui ao ser questionado sobre a validade da conclusão no âmbito do conjunto dos vários alunos que frequentam o projeto Desporto Escolar (objetivo do coordenador), começou por dizer que se poderia extrapolar a informação recolhida junto dos 10 atletas aos restantes participantes do desporto escolar, tendo respondido “*Sim, bastava dizer por exemplo que 50% melhora a sua performance desportiva, em vez de 80%*” quando lhe sugeri que refletisse se o conjunto de alunos era representativo, sugeriu uma nova amostragem e com base nesses dados considerou ser válido comunicar uma extrapolação da informação recolhida “*ele podia tirar por exemplo 10 pessoas de cada ano do 7.º ano, 10 pessoas do oitavo, 10 pessoas do nono, 10 pessoas ... do decimo segundo ... 10 mais ou menos e depois fazer... a percentagem, mais ou menos do número das pessoas de todos os anos*”. Ainda na primeira entrevista, no item 5.2, Rui considerou que poderia extrair a percentagem de alunos da escola que cumprem com a recomendação do consumo de água e a percentagem de alunos que fica aquém da recomendação de consumo de água, por que a amostra que considerou era representativa da população.

Breve síntese. Durante a lecionação da unidade de ensino, Rui reconheceu representações enganosas, procedimentos incorretos de construção e interpretações erróneas de diferentes representações dos dados. Mostrou-se capaz de distinguir variáveis segundo as categorias quantitativa/qualitativa, no entanto, foi incapaz de diferenciar o facto dos dados quantitativos serem ou não discretos, mas, ainda assim, optou por representações e medidas estatísticas adequadas à natureza dos dados. Mostrou-se destro na determinação das medidas de tendência central e reconheceu o efeito de um valor atípico no valor da média de uma distribuição, no entanto, apesar de reconhecer que a média

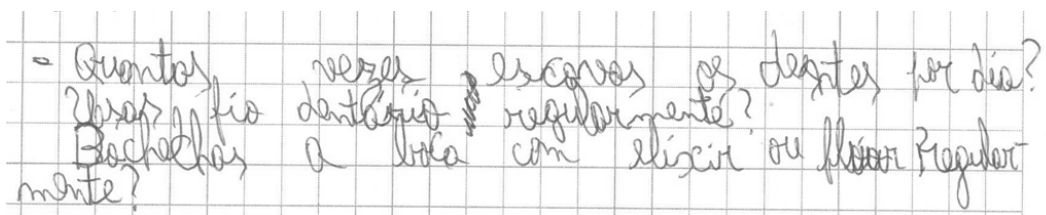
nem sempre é robusta, mostrou-se hesitante em privilegiar outra medida estatística. No que se refere à comparação de distribuições e validação de conclusões, procedeu à comparação e argumentou com base em argumentos estatísticos válidos. Relativamente ao desenvolvimento e avaliação de inferências, aceitou generalizar as conclusões extraídas com base numa amostra representativa da população em estudo.

7.4. Capacidades de planeamento estatístico e análise de dados depois da unidade de ensino

Nesta secção refiro o desempenho do aluno, no âmbito da tarefa proposta na segunda entrevista (Anexo 6) e no teste final (Anexo 15).

7.4.1. Planeamento estatístico

Formular questões. Rui ainda revelou dificuldades na formulação de questões que pudessem especificar o problema “Os alunos da escola têm práticas corretas quanto à higiene oral?” (Item 4 da segunda entrevista) indicou a formulação das seguintes questões de inquérito: “*Quantas vezes escovas os dentes por dia? Usas fio dentário? Bochechas a boca com elixir ou flúor?*” Quando questionado sobre o significado que o inquirido deveria atribuir ao termo “usas”, optou por incluir na formulação das duas últimas questões o termo “regularmente” (Figura 104). Uma vez mais, ao ser questionado sobre o significado que se deveria atribuir ao termo “regularmente”, alegou que correspondia a uma utilização de pelo menos 2 a 3 vezes por semana e, mencionou que seria ele próprio a aplicar o inquérito pelo que as dúvidas seriam esclarecidas na aplicação, caso surgissem, pois esta foi a prática corrente a que assistiu, quando solicitaram a participação dos alunos da sua turma na realização de um inquérito.



Quantas vezes escovas os dentes por dia?
Usas fio dentário regularmente?
Bochechas a boca com elixir ou flúor Regularmente?

Figura 104. Excerto do esboço do plano proposto no item 4 da 2ª entrevista

No teste final, Rui face ao problema “estudar os hábitos de sono dos jovens portugueses entre os 12 e 16 anos” (item 3.2) formulou quatro questões de pesquisa. Embora se possa considerar que existem alguns aspetos formais a melhorar na formulação, é de salientar que a análise conjunta da resposta à segunda e terceira questões (hora de deitar e de acordar) permite validar os dados recolhidos com a primeira questão (número médio de horas de sono diário). Um outro aspeto interessante do conjunto de questões formuladas prende-se com o facto de ser ainda possível estudar a relação entre as variáveis - hora de deitar e cansaço às primeiras horas do dia (Figura 105).

3.2 - Perguntas, horas dormes por noite normalmente?
- Acordas a que horas?
- Deitas-te a que horas?
- Quando acordas sentes-te cansado?

Figura 105. Resposta ao item 3.2 do teste final

Distinguir entre população e amostra. No item 4 da segunda entrevista, Rui esclareceu que para estudar os hábitos de higiene dentária dos alunos da escola, não haveria necessidade de realizar um censo, apenas uma sondagem, no entanto considerou que a amostra deveria ter uma maior dimensão do que a que considerou em outros estudos, dado que antecipava uma maior variabilidade populacional para as variáveis em estudo. Evidenciou distinguir, no contexto da situação em estudo, os conceitos de população e de amostra. No teste final, nos itens 5.1 e 5.2 foi evidente que Rui identificou corretamente a população e a amostra em estudo, ao responder que a população era “os 1560 alunos daquela escola” e ao indicar como procedimento de seleção da amostra a escolha ao acaso de 160 alunos sendo que 80 de cada género. Ainda no teste final, nos itens 6.1 e 6.2, Rui considerou que a população em estudo eram as maçãs de Alcobaça e que a amostra era o conjunto das 51 maçãs, cujas dimensões constavam do diagrama de caule e folhas.

Amostras representativas e ponderação de elementos que afetam a representatividade

Amostras representativas. Na segunda entrevista, ao ser solicitado a Rui um esboço do estudo a realizar para responder ao problema “Os alunos da escola têm práti-

cas corretas quanto à higiene oral?” (Item 4), considerou que a variabilidade inerente a situação deveria ser tida em consideração na tomada de decisão relativa o tipo de estudo e/ou dimensão da amostra no caso de optar por uma sondagem.

Rui - Então eu acho que deveria ir a todas as turmas... Da escola com um inquérito, como é normal fazer... Stora eu estou a pensar... Escolher toda a gente da escola... Mas eu também penso que assim...

Prof^a - Levantar informação de todas as pessoas...

Rui - Sim, mas eu acho que assim ia ficar... Com uma dimensão demasiado grande... Ia ser mais difícil para analisar, mas se formos escolher, por exemplo cinco pessoas da turma pode não ficar tão bem... Eu acho mesmo que fazer um censo é o melhor.

Prof^a - Então neste caso uma sondagem... Não te dá precisão suficiente...

Rui - Eu acho que não, porque neste tipo de... Por exemplo, qual é a bebida que se bebe mais ou uma coisa assim a maioria dos jovens bebem todos as mesmas bebidas ice tea ou coca-cola e assim, mas aqui acho que ficava melhor fazer um censo... Porque... Os métodos para tratar dos dentes, da boca não são iguais para todos os jovens, há pessoas que escovam mais os dentes por dia, há pessoas que escovam depois de todas as refeições, há pessoas que escovam só à noite, há pessoas que escovam à noite e de manhã... Depende muito.

Prof^a - Então e em vez de considerar toda a gente, se considerasses uma amostra um pouco maior do que a amostra que consideras ao estudar a bebida preferida... Achas que não daria informação suficiente?

Rui - Se calhar metade da turma, 15 pessoas por turma mais ou menos acho que já dava... 14 por turma, 7 raparigas e 7 rapazes.

No item 5.2 (teste final), Rui optou por uma amostragem estratificada ao considerar que deveria “*Escolher 80 rapazes e 80 raparigas de diversos anos e turma*”, não tendo especificado o modo como iria por em prática a seleção, e no item 4 da segunda entrevista continuou a defender que no âmbito do problema em estudo “As raparigas gastam menos dinheiro na escola que os rapazes?” se deveria salvaguardar a proporção dos géneros existente na população que será sempre de cerca de metade para cada um.

Ponderação de elementos que afetam a representatividade. Rui no item 3.1 do teste final, indicou que um procedimento passível de conduzir a uma amostra enviesada, passaria pela seleção de um conjunto de pequeno de elementos da população restritos a um género de uma área específica do país (Figura 106). Alegou que com um conjunto tão pequeno de elementos femininos e de uma área rural, iria obter respostas com menor diversidade que a existente na população, o que poderia implicar a extração de más conclusões.

Figura 106. Resposta ao item 3.1 do teste final

Análise crítica de estudos estatísticos. No âmbito da análise crítica de estudos estatísticos, perante um contexto de natureza menos familiar, para alunos da sua faixa etária – estudo relativo às condições de trabalho – Rui escolheu o jornal 2 embora tenha hesitado entre duas sondagens, argumentou a rejeição de duas das três amostras não representativas. A rejeição da sondagem do jornal 3, devia-se ao facto da amostra ter sido seleccionado entre os assinantes do jornal e no caso do jornal 4, devia-se ao facto da amostra ter sido seleccionado entre pessoas naturalmente mais jovens, pelo que em ambos os casos a amostra seria composta exclusivamente por pessoas com determinado perfil, em vez de conter elementos da população com perfil diversificado.

Rui – O 3 não dá usaram só os assinantes que ... podem ser uma parte diferente da população em geral.

(...)

Rui – Entre este e este (jornal 1 e 2) não sei... só há diferença na quantidade de pessoas, e que são eleitores e que os outros não são ... mas também são ao acaso....

(...)

Rui - As redes sociais são mais usadas pelos pessoas 18 .. 20 mais novas Não quer dizer que não haja outras pessoas nas redes, mais velha ... mas é mais os jovens, logo não dá vá ... por que era de uma parte [específica] da população.

Rui – Eu não consigo explicar o jornal 1...

Breve síntese. No final da unidade de ensino, Rui continuou a mostrar-se capaz de decidir, face a um estudo estatístico, se deveria realizar um censo ou uma sondagem. A capacidade de captar a diversidade populacional da variável em estudo foi o mote para a decisão de recorrer ao estudo da população ou a uma parte desta. Perante uma situação de maior complexidade e abstracção, continuou a ser crítico face a estudos envolvendo amostras não representativas, identificou e ponderou elementos que afetavam a representatividade de duas das três amostras em relação à respetiva população e recorreu a algumas situações concretas para ilustrar os diferentes elementos que ponderou. Quanto à formulação de questões, procurou formular questões mais adequadas à recolha da informação pretendida, possibilitando a posterior resposta ao problema estatístico em análise. No entanto, persistiu alguma dificuldade na formulação de questões,

nomeadamente no âmbito da clareza e objetividade que, segundo o aluno, são ultrapassadas pela presença do autor do inquérito na aplicação.

7.4.2. Análise de dados

Natureza dos dados e representações adequadas. Depois do final da unidade de ensino, tendo sido pedido um esboço do estudo para conhecer as práticas de higiene oral (item 4 da 2.^a entrevista), Rui optou por analisar as seguintes variáveis: (i) número de escovagens diárias; (ii) utilização de fio dentário pelo menos uma vez por dia (sim/não) e (iii) utilização de elixir dentário ou fluor pelo menos uma vez por dia (sim/não), tendo procedido à sua correta categorização. No âmbito da organização e tratamento de dados, considerou que para apresentar os dados recolhidos com base na primeira questão, recorreria a um gráfico de barras (para os valores 1, 2, 3, 4 e 5 vezes). Quanto aos dados recolhidos com a segunda e a terceira questão, correspondente às variáveis (ii) e (iii), considerou que não haveria necessidade de os representar graficamente; dado que existiriam apenas duas modalidades, seria preferível divulgar a percentagem correspondente a cada modalidade. Quando questionado sobre quais as medidas estatísticas que utilizaria para descrever os dados, indicou a média para a primeira variável (quantitativa) e a moda para as restantes (variáveis qualitativas).

Construir, analisar e interpretar representações dos dados

Representação tabelar. No teste final (Item 1.1), Rui não manifestou dificuldades em proceder à leitura e interpretação da informação veiculada pela tabela de frequências absolutas e relativas em percentagem, tendo completado os valores em falta corretamente (Figura 107).

1.1-

$$a = \frac{x}{150} = \frac{10}{100} \quad x = \frac{150 \times 10}{100} = 15$$

$$b = \frac{x}{150} = \frac{12}{100} \quad x = \frac{150 \times 12}{100} = 18$$

$$d = 5 + 7 + 8 + 30 + 12 + 15 + 12 = 144$$

$$150 - 144 = 6$$

$$d = 3 + 8 + 12 + 20 + 8 + 10 + 8 = 96$$

$$150 - 96 = 54$$

$a = 15$
 $b = 18$
 $c = 4$
 $d = 2$

Figura 107. Resposta ao item 1.1 do teste final.

Gráfico de barras. Rui, na segunda entrevista, face a duas representações enganosas (Item 2b)) argumentou que os motivos principais da diferença entre as duas representações gráficas, dos mesmos dados, era a omissão da origem do eixo vertical e a escolha de uma unidade de escala distinta (Figura 108). Tal como a sua colega Ana, Rui não referiu o espaçamento entre as barras no eixo horizontal.

Rui – isto é porque um começa no 7000 e anda de 1000 em 1000 e o outro começa no 7000 e anda de 2000 em 2000, vai ficar com as barras todas muito mais baixas, vai parecer que tem poucas queixas e este [gráfico 1] que tem muitas queixas ... eu acho que eles deviam fazer de 1500 em 1500.

(...)

Rui – Se começar no zero o gráfico vai ficar muito grande...

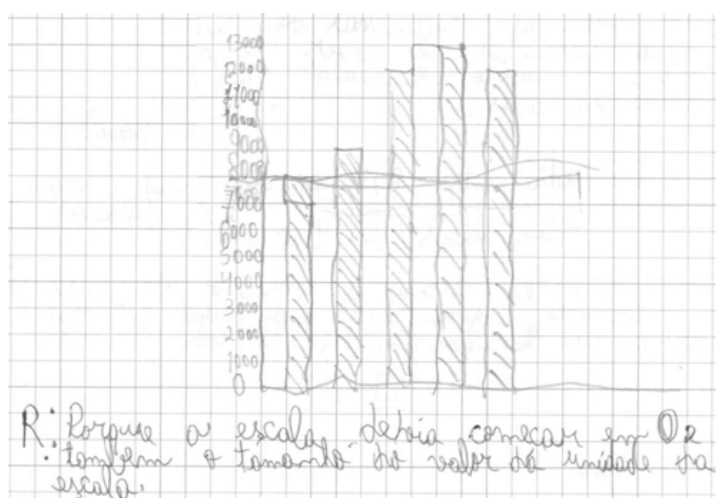


Figura 108. Resposta ao item 2b) da 2.^a entrevista.

Sectograma. Na segunda entrevista, Rui continuou a mostra-se capaz de fazer uma leitura literal da informação representada por um setograma, ao responder ao item 3 a1) “*Na escola A a prova que os alunos gostam menos é a de dificuldade. Nas duas escolas a prova favorita é a bloco*”. Quando questionado no item 3 a2), se poderia afirmar que o número de alunos da escola A que prefere competir em provas do tipo Bloco é superior ao número de alunos da turma B a preferir o mesmo tipo de prova, Rui de imediato reconheceu ser impossível afirmar que um sector de um setograma estaria associado a um maior número de elementos do conjunto em análise, ao responder “*Não podemos afirmar, porque não sabemos a quantidade de alunos nas escolas. Por exemplo haver 100 na A e 500 na B, 50% de 100 é menos do que 45 % de 500 cerca de 200*”.

Histograma. Quanto a alguns aspetos de construção, análise e interpretação de um histograma (Grupo 2 do teste final), Rui no teste final, perante um conjunto de questões de leitura e interpretação de um histograma com idêntica complexidade ao apresentado no teste inicial, mostrou-se capaz de identificar o número de classes e a respetiva amplitude (Item 2.1), conseguiu fazer uma leitura literal da representação ao identificar o número total de efetivos (Item 2.2) e o número de participantes que realizaram um tempo de prova inferior a x ou superior ou igual a y (Itens 2.3.1 e 2.3.2). Rui respondeu incorretamente ao item 2.4 que solicitava a indicação de um possível tempo de prova para o vencedor ao responder “*4 minutos e 30 segundos*” correspondente a 4,5 minutos, quando questionado sobre o equívoco alegou que “*Stora, pensei dizer 4,3 ... então 4 minutos e 30 segundos*”. Ainda no âmbito da construção de um histograma adequado à representação de um conjunto de dados (item 3 c) da segunda entrevista), Rui perante 35 dados compreendidos entre 8 e 29 minutos, considerou que deveria construir o menor número possível de classe, por que existem poucos dados.

Rui – Podia ser de 0 a 5, 5 a 10 ... mas assim a primeira não tinha dados ... de 8 a 10, de 10 a 12, de 12 a 14 ... assim ficam muitas classes, ... de 8 a 12, de 12 a 14, ..., 28 a 32 dá 6 classes, a de 5 em 5 era a mais fácil para quem lê o histograma, mas tínhamos que fazer poucas ... não são muitos [dados].

Determinação de medidas estatísticas e a sua interpretação

Moda, média e mediana. Rui no item 3b) da segunda entrevista, determinou corretamente as medidas estatísticas centrais, e reconheceu que a distribuição era amodal.

Na determinação da mediana, procedeu do seguinte modo particular: Após ter ordenado a distribuição, encontrou os dois valores centrais e em vez de determinar a sua média, optou por listar os números inteiros do intervalo dos valores centrais [197, 201] e deste modo encontrou o valor da mediana. Perante a existência de um valor atípico mostrou-se relutante em optar pela mediana, embora tenha tido consciência que a média pudesse ser afetada por um “mau tempo” de treino, tratando-se de tempos de um atleta, a apreciação do seu desempenho deveria envolver todos os registos ocorridos. Alegou ainda que era esta a medida utilizada para referir os desempenhos físicos dos atletas, por se tratar de um valor que seja o equilíbrio entre os vários registos. Após algum tempo de reflexão, optou por referir oralmente que os valores numéricos que resumiam melhor os tempos dos últimos oito treinos eram a média e a mediana. Acrescentou ainda que aceitou o valor da mediana porque está mais próxima da maioria dos registos, e no caso de se procurar um patrocinador, era um bom número de resumo. Reafirma que a média seria o único valor que representaria “honestamente” o desempenho do atleta. Por escrito, Rui apresentou uma resposta bastante breve, dada a dificuldade em expressar os seus argumentos “no papel” (Figura 109).

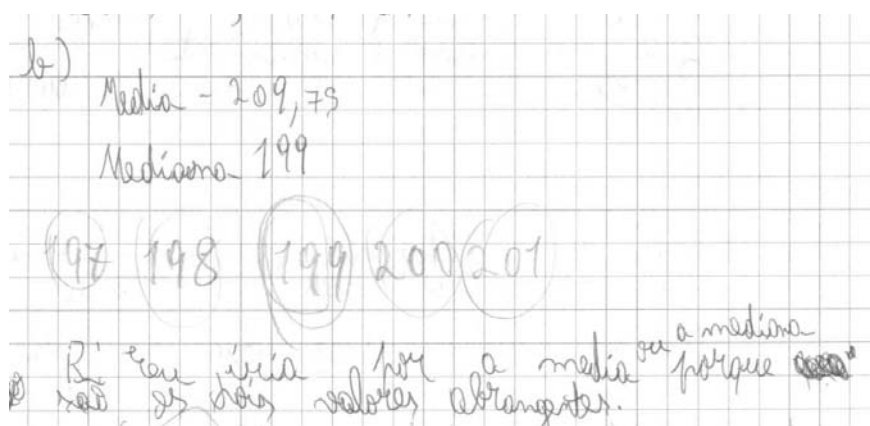
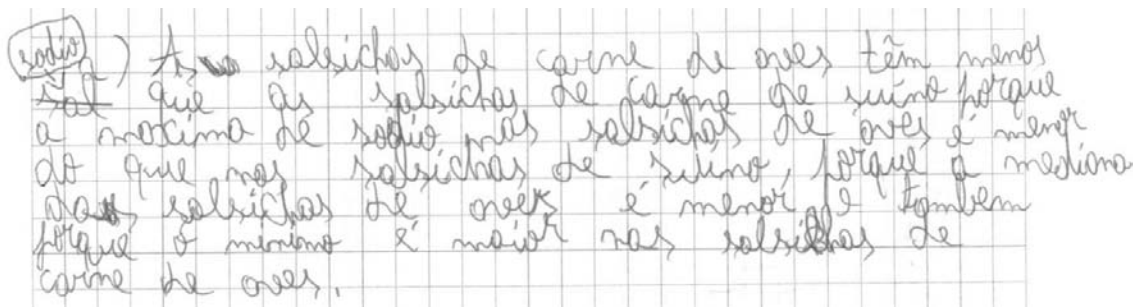


Figura 109. Resposta ao item 3b) da 2.^a entrevista.

Comparar duas distribuições e validar conclusões. No item 1a) (segunda entrevista), face aos diagramas de extremos e quartis de duas distribuições relativas à quantidade de sódio, Rui identificou a carne de aves como sendo o tipo de carne que globalmente teria menos quantidade de sódio por cada 100 g e argumentou que a sua conclusão seria válida pois os valores assumidos pelos três quartis (1.º, 2.º e 3.º quartis) na distribuição da carne de aves eram inferiores aos correspondentes valores na distribuição da carne de suíno (Figura 110), à semelhança do que tinha feito no item 3^a da pri-

meira entrevista. Tendo instigado Rui a refletir sobre a utilização de outros elementos da representação, respondeu que tinha comparado os centros e os valores mínimos e máximos e como tal tinha tudo para comprovar a validade do que afirmava.



(sódio) As ~~as~~ salsichas de carne de ovelha têm menos sal que as salsichas de carne de suíno porque a máxima de sódio nas salsichas de ovelha é menor do que nas salsichas de suíno, porque a mediana das salsichas de ovelha é menor do que a mediana das salsichas de suíno e também porque o mínimo é maior nas salsichas de carne de ovelha.

Figura 110. Resposta ao item 1a) da 2.^a entrevista.

Perante a necessidade de comparar dois conjuntos de dados recolhidos em duas fases de uma experiência (Item 1b) da segunda entrevista), Rui recorreu à comparação dos valores médios dos conjuntos “Ela desceu o consumo de sódio porque antes ela ingeria cerca de 2,73 g de sódio por dia e agora ela ingere cerca de 2,14 g de sódio por dia.”. Quando questionado se existiriam outras ferramentas/medidas estatísticas a que poderia ter recorrido, respondeu, “podia ter calculado a mediana de cada um e comparado ... os valores são mais baixos [mínimo e máximo] ... a mediana deve ser mais baixa”. No item 4 do teste final, alegou que a conclusão do empresário poderia ser refutada porque no grupo com menor média existiam as pessoas com melhores resultados “tem o grupo de pessoas com melhor nota no teste de criatividade”.

Transformar uma representação numa outra representação

Representação tabular – gráfico de barras. No item 1.2 (teste final), que solicitava a interpretação da escala de construção do gráfico de barras, Rui embora tenha compreendido o que se pedia, cometeu um erro de leitura da grandeza representada no eixo vertical, pois considerou que as quatro unidades de escala da barra “1 falta” correspondiam a 18 alunos (leitura da coluna das frequências absolutas), em vez de ter recorrido à correspondente percentagem (leitura da coluna das frequências relativas), tendo respondido “É de 4,5 em 4,5”.

Diagrama de caule e folhas - diagrama de extremos e quartis. No item 6.4 (teste final), Rui mostrou-se capaz de determinar o valor numérico dos quartis recorrendo ao

facto de ter reconhecido que os dados apresentados no diagrama de caule e folhas estavam ordenados, posteriormente escolheu uma escala adequada ao conjunto dos cinco números de resumo e construiu o diagrama de extremos e quartis sem dificuldade (Figura 111).

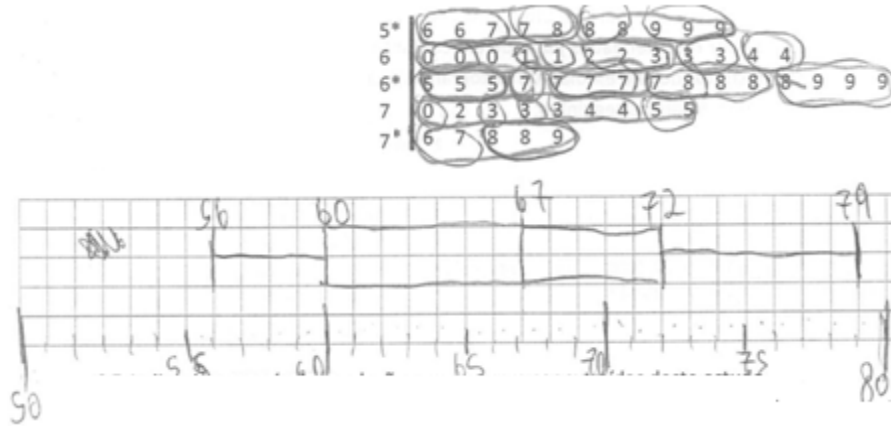


Figura 111. Resposta ao item 6.4 do teste final.

Diagrama de caule e folhas – histograma. No item 3 d) da segunda entrevista, Rui procedeu a uma associação correta, alegou que no grupo II existiam um maior número de efetivos e como tal, no grupo I os 7 alunos em 20, do caule 15, iriam representar 35% (cálculo mental 5 em 20 \rightarrow 25% e 2 em 20 \rightarrow 10% então se 5+2 era 7 então 25+10 = 35%) e no grupo II os 9 alunos em 25 do caule 15 correspondiam a 36% (cálculo mental 5 em 25 \rightarrow 20% então 1 em 25 \rightarrow 4% pelo que 9 = 5+4x1= 36%). Rui reconheceu que a transformação dos diagramas de caule e folhas assentava na conversão da quantidade de folhas (frequência absoluta) em informação percentual (frequências relativas) que estariam próximos caule a caule (Figura 112).

e) R. Porque a quantidade de pessoas no grupo II é maior e por isso as percentagens são parecidas com as do grupo I por isso escolho o conjunto B.

Figura 112. Resposta ao item 3d) da 2ª entrevista.

Histograma - diagrama de extremos e quartis. No item 5.3 do teste final, Rui estabeleceu uma única correspondência correta entre os histogramas e os diagramas de extremos e quartis, ao indicar a seguinte associação $A \rightarrow 3$, $B \rightarrow 1$ e $E \rightarrow 2$ e não fundamentou a sua resposta. Rui cometeu o erro de emparelhar o histograma da distribuição com enviesamento à esquerda com o diagrama de extremos e quartis enviesado à direita,

o que corresponde a uma dificuldade de converter uma grande densidade relativa dos dados, num pequeno comprimento da respetiva componente do diagrama de extremos e quartis.

Generalizar conclusões para a população

No item 4 da segunda entrevista, Rui considerou que as conclusões extraídas do estudo com base na amostra se deveriam generalizar à população porque tinha tido o cuidado de definir uma amostra de boa dimensão (um pouco menos de 50% da população) e cujo processo de seleção era aleatório dentro dos extratos género e turma. No teste final, item 1.4, Rui considerou que a Direção iria falhar, pois, segundo Rui, grande parte dos alunos, ou seja mais de 38% dos alunos não tinha faltas e como tal, iriam faltar certificados de assiduidade. Ainda no teste final, nos itens 6.5. e 6.6, relativos ao estudo do calibre das maçãs produzidas na zona de Alcobaça, Rui indicou como conclusões o facto de a mediana tomar o valor 67 mm e da amplitude da variável em estudo ser 23 mm, considerando que as conclusões deveriam ser extrapoladas à população, mas não justificou a sua posição.

Breve síntese. No final da unidade de ensino, perante situações de maior complexidade e abstração, Rui continuou a mostra-se capaz de ler, interpretar e avaliar diferentes representações. Perante um problema estatístico, identificou variáveis passíveis de mobilizar a recolha de um conjunto de dados adequados para responder à situação, refletiu sobre a natureza dos dados que poderia recolher, e o modo como os deveria representar e resumir tendo em consideração a sua comunicação. Em metade das situações, mostrou-se incapaz de recorrer à interpretação da forma, à identificação do centro e à dispersão da distribuição dos dados apresentados numa determinada representação para a transformar numa outra representação. O aluno continuou a ter um bom desempenho na comparação de duas distribuições, reconheceu características gerais tais como forma e centro de uma distribuição, não tendo recorrido à utilização de medidas de dispersão. Quanto à escolha da medida estatística mais adequada para resumir a informação contida nos dados, continuou a optar pela média mesmo contendo a distribuição dos dados valores atípicos. Rui, uma vez mais, reconheceu que o valor da média seria maior que a generalidade dos dados do conjunto quando o valor atípico era o máximo da distribuição. No que se refere à comparação de distribuições e validação das suas conclusões, continuou a mostrar-se capaz de proceder à comparação e justificou recorrendo a

argumentos estatísticos válidos. Durante a entrevista, relativamente ao desenvolvimento e avaliação de inferências, aceitou generalizar as conclusões extraídas à população em estudo, com base numa amostra representativa, por considerar que a amostra que escolheu era um retrato quase fiel da população. Nas restantes situações propostas após a unidade de ensino, o seu desempenho é fraco, tendo rejeitado o facto de a Direção poder prever o número de certificados, com base em convicções pessoais sobre a assiduidade dos alunos e não tendo fundamentado a sua resposta numa das situações.

7.5. Síntese global

De modo geral, Rui conseguiu realizar com sucesso as tarefas propostas, revelando entusiasmo durante o estudo do tema. Mostrou-se, desde o primeiro contacto com o processo de amostragem, capaz de reconhecer possíveis fontes de enviesamento na recolha de dados e de ponderar elementos que pudessem afetar a representatividade de uma amostra. Argumentou os seus pontos de vista recorrendo a exemplos concretos que assegurou estarem de acordo com as características iniciais da situação, com o objetivo de dar ênfase ao elemento que ponderava quanto à afetação da representatividade ou explicitou a sua opção de amostragem de modo a minimizar as fontes de enviesamento que identificava. O aluno recorreu à utilização de amostras aleatórias simples em detrimento do estudo da população. Analisando o raciocínio estatístico durante o processo de amostragem constata-se que Rui não evidenciou mal entendidos indicados por Garfield (2002). No entanto, quando lhe foi solicitado que indicasse um processo de amostragem, recorreu geralmente à estratificação da amostra, argumentando que, em várias situações, o género é um elemento a ponderar pelo que seria necessário um processo de amostragem que garantisse a proporcionalidade do estrato género. O aluno, ao longo da unidade, revelou-se capaz de identificar a variabilidade/natureza estatística em diferentes situações. Face às categorias do desenvolvimento do conceito de amostra de Watson e Moritz (2000), o desempenho de Rui enquadra-se na categoria mais elevada da hierarquia, denominada por “Grandes amostras sensíveis a enviesamento”. Constata-se que, progressivamente, o aluno passou a formular questões de inquérito tendo em vista a sua pertinência no processo de recolha de dados, procurou formular questões relevantes e significativas face ao problema em estudo, embora nem sempre tenha ponderado se a formulação era suficientemente clara e/ou objetiva.

Progressivamente, Rui mostrou-se capaz de reconhecer e categorizar conjuntos de dados e de escolher uma forma de representação adequada à sua análise. No âmbito da análise de dados identificou características gerais como a forma, o centro e, pontualmente, a dispersão da distribuição, predominando ainda uma visão individual dos dados. Mostrou-se incapaz de identificar as medidas estatísticas centrais mais adequadas à análise de uma distribuição acentuadamente assimétrica. O seu desempenho, quanto à comparação de duas distribuições, foi evoluindo, ainda que em algumas situações o aluno tenha exibido o equívoco de proceder à comparação recorrendo exclusivamente à comparação das medidas centrais, não sendo visível o papel das medidas de dispersão na análise comparativa.

Nos vários estudos estatísticos, no momento de retirar conclusões, as respostas de Rui tiveram as seguintes características: (i) recorre às suas próprias opiniões e ignora os dados dos estudos realizados, (ii) considera possível generalizar as conclusões face a uma amostra de dimensão razoável e de seleção aleatória, quando intui que o processo de amostragem terá capturado a variabilidade populacional e (iii) recusa-se a tirar conclusões ou a desenvolver previsões, uma vez que a variabilidade populacional não terá sido “retratada” por meio do processo de amostragem.

O facto da unidade de ensino valorizar a realização de pequenas investigações estatísticas e a análise crítica de estudos estatísticos, suscitou, em Rui um progressivo reconhecimento da variabilidade natural e da variabilidade do processo de amostragem, aspetos fundamentais para a compreensão da natureza e do papel da Estatística.

In general, if you want to change something you have to understand it and if you want to understand something you have to change.

Bakker (2004)

Capítulo 8

Conclusão

Este capítulo inicia-se com uma síntese do estudo, em que recorro os seus objetivos, as questões de investigação e a metodologia de investigação. Depois, apresento os principais resultados que emergem do trabalho dos alunos no que diz respeito ao desenvolvimento da sua capacidade de planeamento estatístico e de análise de dados. Por fim, concluo com uma breve reflexão pessoal sobre a experiência de investigação, as implicações na minha prática profissional e apresento algumas recomendações para futuros estudos.

8.1. Síntese do estudo

O ano de 2013 foi escolhido para ser o “ano internacional da Estatística”, várias iniciativas nacionais e internacionais procuraram dar ênfase ao importante papel da Estatística, no modo com estados, empresas, organizações e indivíduos singulares tomam decisões ou fazem previsões com base em informações organizadas segundo métodos estatísticos e probabilísticos, tendo em conta que a maior parte da atividade humana assenta em fenómenos aleatórios. Em Portugal, a investigação no âmbito do ensino da Estatística tem sido escassa. O aparecimento de algumas investigações foi motivado pelo surgimento do programa de Matemática (ME, 2007), que procurou romper com uma aprendizagem quase passiva do conhecimento estatístico e promover uma aprendizagem ativa da literacia estatística e do processo de investigação estatística, desde do 1.º ciclo, tendo com propósito principal desenvolver nos alunos a capacidade de compreender e de produzir informação estatística, bem como de a utilizar para resolver problemas e tomar decisões fundamentadas.

Com este estudo pretendi promover o desenvolvimento da capacidade de planeamento estatístico e análise de dados dos alunos do 8.º ano, no âmbito de uma unidade de ensino baseada em tarefas de carácter exploratório e investigativo. Pretendi ainda, identificar potencialidades da didática da Estatística e dificuldades de concretização desta unidade de ensino.

O quadro teórico deste estudo aborda aspetos relacionados com o desenvolvimento do ciclo investigativo: (i) formulação de questões estatísticas, (ii) planeamento estatístico; (iii) análise exploratória de dados; (iv) desenvolvimento e avaliação de inferências e previsões baseadas nos dados. Tive também em atenção as principais dificuldades que os alunos enfrentam na aprendizagem da Estatística, quer ao nível dos conceitos, quer ao nível da leitura, construção e interpretação de tabelas e outras representações.

Este estudo tem por base uma unidade de ensino que visa o desenvolvimento da capacidade de planeamento estatístico e a análise de dados com alunos de 8.º ano. A unidade segue as orientações do programa de Matemática do Ensino Básico (ME, 2007), tendo também sofrido influência da revisão da literatura que realizei sobre o tema e que contempla uma grande diversidade de tarefas, incluindo pequenas análises de estudos estatísticos e a realização de investigações estatísticas. A unidade de ensino foi realizada no ano letivo de 2010/11, sendo os participantes no estudo, os alunos de uma turma de 8.º ano da qual fui professora. Dessa turma, foram selecionados dois alunos que foram objeto de análise aprofundada em estudo de caso. Desta forma, a metodologia utilizada foi de natureza qualitativa e interpretativa, baseada em três estudos de caso (a turma e dois alunos), numa vertente de experiência de ensino. Esta metodologia permitiu conhecer o significado que os participantes, inseridos no seu ambiente escolar natural, atribuem aos aspetos da Estatística visados neste estudo, e também aos seus conhecimentos e às suas experiências.

A principal fonte de recolha de dados foram os testes (inicial e final) dos alunos, as suas produções escritas durante a unidade de ensino e as duas entrevistas individuais realizadas aos dois alunos envolvidos no estudo. Foram também utilizados registos de áudio, do trabalho do grupo em que os alunos-caso se inseriam e dos seus diálogos com a professora, bem como registos de vídeo das aulas.

8.2. Conclusões do estudo

Neste ponto apresento as principais conclusões do estudo no que diz respeito ao desenvolvimento das capacidade de planeamento estatístico e de análise de dados e em função das questões nele formuladas, indicando (i) as capacidade de planeamento estatístico e de análise de dados que os alunos tinham antes da realização da unidade de ensino, (ii) as aprendizagens evidenciadas pelos alunos, no final da unidade de ensino, ao nível do planeamento estatístico, (iii) as aprendizagens evidenciadas pelos alunos, no final da unidade de ensino, ao nível da análise de dados e (iv) as principais dificuldades demonstradas pelos alunos antes, durante e no final da unidade de ensino, quer ao nível dos conceitos, quer ao nível dos procedimentos inerentes ao planeamento e à análise de dados.

8.2.1. Planeamento estatístico

1. Formular questões. No início da unidade de ensino, a generalidade dos alunos, a trabalhar em grupo, foi incapaz de formular questões adequadas ao objeto de estudo, por dificuldade de interpretação desse mesmo objeto.

No entanto, ao longo da unidade, o desempenho dos alunos mostrou que o conhecimento dos vários e distintos contextos (Gal, 2002) foi importante e permitiu que, evoluíssem positivamente. Durante a unidade de ensino foi notória a capacidade dos alunos para proporem questões da categoria resumo, correspondentes aos níveis questão e pré-questão, face à apresentação de um conjunto de dados multivariado, o que corresponde à tendência sentida por Arnold (2008) que envolveu alunos do ensino secundário, numa situação similar.

No final da unidade, cerca de três quartos dos alunos, mostra-se capaz de formular, pelo menos, duas questões de pesquisa suscetíveis de permitir a recolha de dados adequados ao problema em estudo, sendo que mais de metade destes alunos apresentou um conjunto de três questões adequadas.

As discussões em torno da tipologia de questão de pesquisa e o modo como a sua formulação poderia influenciar os dados obtidos foram pertinentes e fomentaram a capacidade dos alunos para se colocarem no papel do inquirido, minimizando o envolvimento associado a uma formulação com juízo de valor, ambígua ou com múltiplas interpretações (GAISE, 2005). Foi evidente a importância da reflexão, em torno da

antecipação dos dados, com base nas questões que cada aluno-caso delineou durante as entrevistas, tendo existido significativos ganhos ao nível da reformulação das questões propostas e ao nível da perceção da variabilidade dos dados a recolher (Konold & Higgins, 2003).

2. *Distinguir população e amostra.* Durante a realização da unidade de ensino, os alunos não manifestaram grandes dificuldades na identificação da população e amostra, nas várias situações apresentadas. Os alunos revelaram também compreender os motivos que implicam a utilização de amostras em estudos estatísticos e reconheceram vantagens associadas à redução do número de dados recolhidos (Garfield & Gal, 1999). Em contraponto com os resultados de Jacobs (1999), o facto de os alunos privilegiarem a realização de sondagens em detrimento de censos, deveu-se à significativa dimensão das populações em análise e possivelmente também à maturidade dos alunos da turma, de uma faixa etária mais elevada que a dos alunos do estudo de Jacobs.

No final da unidade, cerca de um terço dos alunos, numa situação envolvendo a “população escolar”, considerou que a população era o conjunto de elementos da amostra, equívoco recorrente nos estudos de Fernandes, Carvalho e Ribeiro (2007) e de Novaes (2004).

3. *Amostras representativas e ponderação de elementos que afetam a representatividade.*

Antes da realização da unidade de ensino, cerca de metade dos alunos mostrou-se capaz de escolher, entre um conjunto possível de alternativas, uma seleção aleatória de uma amostra. Perante o trabalho desenvolvido pelos alunos ao longo da unidade, constatou-se que a generalidade reconhece que o único processo que garante a representatividade de uma amostra, envolve a seleção aleatória dos elementos que a compõem. No início da unidade de ensino, a maioria dos alunos opta por amostras aleatórias com dimensão significativamente grande, ao longo da unidade, a maioria dos alunos foi-se mostrando capaz de pensar e aferir a dimensão da amostra a selecionar face à variabilidade da característica em análise e precisão do estudo a realizar, sendo tal facto evidente nos desempenhos quer de Ana e quer de Rui nas duas entrevistas. No entanto, dado que a maioria das situações envolveu a população escolar, os alunos utilizam a evidente organização vertical em anos de escolaridade (que associam à característica idade, dada o baixo índice de retenção) e a organização horizontal em turmas (que associam à existência de uma cultura de grupo turma), para justificar o facto de privilegiarem uma seleção estratificada em relação a uma seleção aleatória simples.

Ao longo da unidade de ensino, os alunos nas várias situações propostas, mostraram-se capazes de identificar e ponderar vários elementos passíveis de afetar a representatividade de determinada amostra, bem como de reconhecer amostras representativas de determinada população, competências que integram o raciocínio estatístico sobre amostras, segundo Garfield (2002) e Martins e Ponte (2010). No final da unidade de ensino, a maioria dos alunos foi capaz de indicar um procedimento que envolveu a ponderação de pelo menos um aspeto que afetasse a representatividade de uma amostra e, mais de quatro quintos dos alunos considerou que uma amostra representativa deveria ser obtida por um processo de seleção aleatório, pelo que o desempenho da generalidade dos alunos corresponde às categorias “Grandes amostras com seleção aleatória” ou “Grandes amostras sensíveis a enviesamento”. Assim, os resultados obtidos no presente estudo são similares aos de Watson e Moritz (2000) para os alunos do 9.º ano. Durante a unidade, nomeadamente na discussão final das tarefas 4 e 5, um pequeno grupo de alunos demonstrou compreensão da variabilidade da amostragem, ao reconhecer que perante duas ou mais amostras representativas de igual dimensão seria possível obter frequências relativas distintas para um mesmo acontecimento, sendo no entanto possível identificar uma tendência similar nas várias amostras. Para Pfannkuch (2008) tal desempenho corresponde a uma categoria de nível superior à categoria “Grandes amostras sensíveis a enviesamento” de Watson e Moritz (2000).

É de salientar que os alunos não evidenciaram nenhum dos mal-entendidos compilados por Garfield (2002) no âmbito da escolha de uma amostra. No entanto, um terço dos alunos no teste final, à semelhança de Rui, quando lhe foi solicitado que indicasse um processo de amostragem, recorreu à estratificação da amostra segundo o género, argumentando que, naquela situação, o género era um elemento a ponderar na representatividade da amostra, para a variável em estudo, pelo que seria necessário um processo de amostragem que garantisse a proporcionalidade dos dois géneros. Tal como no estudo de Jacobs (1999), este grupo de alunos evidenciou debilidade na compreensão do processo de seleção aleatório simples, que iria naturalmente gerar uma amostra com tendência a satisfazer a proporcionalidade existente na população e, evidenciou desconhecer que se recorre a um processo de seleção aleatório simples precisamente porque na maioria das populações em estudo é impossível identificar e estratificar os vários traços da população que possam ter peso no estudo da característica em análise.

8.2.2. Análise de dados

1. *Natureza dos dados e representações adequadas.* Durante a realização da unidade os alunos da turma conseguem, sem dificuldades, reconhecer e caracterizar os dados como qualitativos ou quantitativos, discretos ou contínuos e explicitam por que para um determinado tipo de dados se adequa uma determinada representação ou medida estatística, evidenciaram dominar os dois aspetos que Garfield (2002) considera pertinentes no âmbito do raciocínio estatístico sobre os dados. Tal como os restantes alunos da turma, Ana e Rui mostraram um bom desempenho neste âmbito.

2. *Construir, analisar e interpretar representações dos dados.* Antes da unidade de ensino, cerca de metade dos alunos evidenciou dificuldades na leitura, interpretação e construção de uma tabela de frequências, de um histograma e de um diagrama de extremos e quartis. Antes da unidade, os cerca de três quintos dos alunos que se mostraram incapazes de construir o diagrama de extremos e quartis, sendo a distribuição apresentada inicialmente num diagrama de caule e folha, sentiram dificuldades (i) na leitura e interpretação do símbolo caule | folha e, (ii) na determinação da localização dos quartis da distribuição. Antes da unidade de ensino, as dificuldades de leitura e interpretação de um histograma, foram geradas pelo desconhecimento da convenção que se considera como classe um intervalo semifechado com extremo inferior incluído e extremo superior excluído.

Durante a realização da unidade de ensino, nomeadamente na tarefa 4, foi nítida a importância que os alunos atribuem ao diagrama de caule e folhas ao recorrerem a esta representação como primeira redução dos dados e como um meio facilitador para a determinação das medidas de ordem, o que corresponde às vantagens da representação apontadas pelos investigadores Batanero, Estepa e Godino (1991). Ainda assim, na construção do diagrama de caule e folhas, os alunos cometeram as seguintes incorreções: (i) duplicação de um caule sem proceder há separação de $C|$ folhas inferiores a 5 e $C*|$ folhas superiores ou iguais a 5 e, (ii) inconsistência na utilização da unidade de escala escolhida ao longo da representação. Quanto aos histogramas, a generalidade dos alunos compreendeu que se trata de uma representação associada a variáveis quantitativas contínuas pelo que se iria tratar de um eixo contínuo, sendo necessário: (i) iniciar a escala na origem do referencial, e (ii) respeitar a unidade escolhida ao longo de ambos os eixos. Deste modo, os alunos não evidenciaram os habituais erros descritos na literatura.

Quanto à construção de um diagrama de extremos e quartis, foi evidente a melhoria dos desempenhos dos alunos durante e no final da unidade de ensino, motivada pela melhoria das destrezas na localização dos quartis e na escolha de uma escala adequada ao conjunto dos cinco números de resumo. Ainda assim, cerca de um quinto dos alunos manifestou dificuldades, na construção do diagrama de extremos e quartis, no final da unidade, por ainda não saber determinar as medidas de localização necessárias, tal como foi referido no estudo de Bakker, Biehler e Konold (2005), que envolveu alunos de variados anos de escolaridade.

Ainda durante a realização da unidade, foi possível observar que vários grupos de alunos teceram breves comentários quanto à concentração dos dados: (i) em torno das medidas centrais (média e/ou mediana); (ii) no intervalo interquartil ou (iii) no intervalo correspondente ao “aglomerado modal”, o que corresponde à capacidade de reconhecimento das características gerais forma e centro de uma distribuição, bem como a uma visão agregada dos dados. Na generalidade das representações construídas pelos alunos não foi indicado o título e os descritores dos eixos, o que vai de encontro às omissões referidas por Wu (2004).

No que diz respeito à construção, análise e interpretação da representação tabular, o desempenho da turma entre o teste inicial e final melhorou de modo significativo. Os alunos Ana e Rui ainda evidenciaram um excelente desempenho na análise crítica de processos de construção erróneos de gráficos de barras e sectorogramas, e na argumentação da adequação das representações por si, escolhidas para comunicar os dados obtidos, durante as entrevistas, satisfazendo os três aspetos que Garfield (2002) considera pertinentes no âmbito do raciocínio estatístico sobre as representações.

Globalmente a maioria dos alunos da turma, mostrou-se capaz de aceder a diferentes representações e às ideias por estas expressas, demonstrando dominar um conjunto de ferramentas que aumentam significativamente a sua capacidade de pensar matematicamente (NCTM, 2007).

3. *Determinação de medidas estatísticas e sua interpretação (moda, média e mediana).* No que diz respeito à identificação da moda, o desempenho dos alunos foi consistente antes, durante e depois da unidade de ensino, não tendo surgido qualquer dos equívocos associados a este conceito. Quanto à média, durante a unidade de ensino, os alunos não evidenciaram dificuldades na sua determinação. Relativamente à correta identificação da mediana, o desempenho dos alunos foi melhorando, sendo que no final, cerca de um quarto dos alunos não respondeu ou determinou de modo incorreto o seu

valor. Durante a unidade de ensino, na tarefa 4, a maioria dos alunos mostrou-se capaz de argumentar a inexistência de média e mediana caso da variável qualitativa em estudo. A generalidade dos alunos não evidenciou os habituais equívocos de troca dos conceitos média e mediana, moda e média ou moda e mediana. O nível de proficiência dos alunos na identificação/determinação das medidas estatísticas de centro correspondeu ao que Boaventura (2003) constatou no seu estudo – a mediana revelou-se a medida de tendência central mais difícil, seguida da média e, por fim, a moda.

4. *Comparar distribuições e validar conclusões.* Face a um dos habituais equívocos sobre a média identificado por Garfield e Chance (2000), propus a análise de uma situação que referia a comparação de duas distribuições assentes exclusivamente na comparação do valor das respetivas médias. Antes da unidade de ensino, cerca de metade dos alunos evidenciou dificuldades na análise comparativa e na identificação de um argumento matemático relativo à dispersão dos dados que justificasse o facto da comparação assente exclusivamente na média ser incorreto. No final da unidade, face a uma situação de maior complexidade o desempenho dos alunos melhorou, tendo três quintos dos alunos referido um ou mais argumentos para refutar a comparação assente exclusivamente na análise da diferença das médias.

Ana evidenciou um bom desempenho na comparação de duas distribuições, antes, durante e no final da unidade, reconheceu características gerais tais como forma e centro de uma distribuição, tendo recorrido menos à utilização de medidas de dispersão. Quanto à escolha da medida estatística mais adequada para resumir a informação contida nos dados, a aluna optou pela mediana quando se apercebeu da existência de valores atípicos, pois reconheceu, nas várias situações, o efeito do dado atípico no valor da média.

Rui mostrou-se incapaz de identificar as medidas estatísticas centrais mais adequadas à análise de uma distribuição acentuadamente assimétrica, tendo considerado a média como a medida estatística mais adequada apesar de lhe reconhecer a falta de robustez e ter sido capaz de antecipar o efeito de um dado atípico no valor da média, o seu desempenho poderá ter sido influenciado pelo contexto das situações apresentadas e o significativo peso atribuído à média, como medida de resumo, nas situações quotidianas. O desempenho de Rui, quanto à comparação de duas distribuições foi evoluindo, ainda que em algumas situações o aluno exiba o equívoco de proceder à comparação recorrendo exclusivamente à comparação das medidas centrais, não sendo visível o papel das medidas de dispersão na análise comparativa.

Globalmente o desempenho dos alunos da turma, evidenciou que foram capazes de compreender por que razão as medidas centrais, fornecem informação sobre a distribuição e reconhecer que as medidas estatísticas centrais e de dispersão são uteis para comparar duas ou mais distribuições, dois dos aspetos que Garfield (2002) considera pertinentes no âmbito do raciocínio estatístico sobre as medidas estatísticas.

5. *Transformar uma representação numa outra representação.* No início da unidade poucos alunos se mostram capazes de converter um histograma de uma distribuição assimétrica no respetivo diagrama de extremos e quartis. No final da unidade, o desempenho global, na conversão de um histograma no respetivo diagrama de extremos e quartis, continuou a ser pouco satisfatório, embora mais de um terço dos alunos, numa situação de maior complexidade, se tenha mostrado capaz de associar a cada um dos três histogramas os respetivos diagrama de extremos e quartis, mobilizando o reconhecimento de características gerais como forma, centro e dispersão das distribuições.

É de salientar que o erro mais comum entre os alunos, foi o emparelhamento de um histograma de uma distribuição assimétrica, por exemplo à esquerda, com um diagrama de extremos e quartis com assimetria à direita, o que corresponde às dificuldades de interpretação de um diagrama de extremos e quartis que Bakker, Biehler e Konold (2005) considerem ser consequentes de uma falta de visão agregada dos dados, da falta de entendimento que num diagrama de extremos e quartis se indica a densidade relativa, em vez da frequência relativa, e esta se relaciona de modo inverso com o comprimento da respetiva componente do diagrama. A conversão entre um diagrama de caule e folhas e o respetivo diagrama de extremos e quartis não apresentou dificuldades para a maioria dos alunos, sendo que mais de três quartos dos alunos mostrou-se capaz de a realizar.

Ana recorreu à interpretação da forma, à identificação do centro e à dispersão da distribuição dos dados apresentados numa determinada representação para a transformar, na maioria das situações, numa outra representação, evidenciando uma visão agregada dos dados. Quanto à transformação de uma representação numa outra, o desempenho de Rui foi evoluindo positivamente e, no final da unidade, mostrou-se capaz de recorrer à interpretação da forma, à identificação do centro e à dispersão da distribuição dos dados apresentados para proceder à transformação, em metade das situações apresentadas na tarefa da segunda entrevista e teste final.

6. *Generalizar conclusões para a população.* No início da unidade os alunos optaram por extrair conclusões envolvendo o centro da distribuição, e quando consideraram passível a generalização das conclusões à população, fundamentaram essencial-

mente com base nas suas opiniões pessoais. Durante e no final da unidade os alunos revelaram uma progressiva evolução positiva, ao nível da capacidade de extrair conclusões a partir dos dados recolhidos, indicando fundamentalmente um comentário relativo ao centro e à dispersão da distribuição. Quanto à extrapolação das conclusões para a população, a partir do que foi apurado na amostra, um quinto dos alunos aceitou as inferências ou fazer as inferências tendo em consideração a qualidade e a quantidade da amostra, sendo no entanto sensíveis ao facto destas estarem associadas a algum grau de incerteza. No outro extremo, cerca de um quinto dos alunos recusou-se a extrapolar as conclusões extraída à população dada que se mostraram muito sensíveis à variabilidade do processo de amostragem ou consideram que a variabilidade da amostra, mesmo que representativa, estará aquém da variabilidade populacional.

Ana, ao longo da unidade, manifestou dificuldade em extrapolar as conclusões de sondagens à população em estudo e, embora no decorrer da unidade tenha sido capaz de argumentar oralmente a validade das extrapolações, em cerca de metade das situações, o seu desempenho no teste final, é muito fraco, o que se explica, por um lado, pela dificuldade em articular a quantidade e a qualidade da amostra versus a variabilidade populacional, que intuiu para a característica em estudo, e, por outro lado, pela dificuldade em explicitar por escrito os seus argumentos.

Nos vários estudos estatísticos, no momento de extrair conclusões, as respostas de Rui tiveram as seguintes características: (i) recorre às suas próprias opiniões e ignora os dados dos estudos realizados, (ii) considera possível generalizar as conclusões face a uma amostra de dimensão razoável e de seleção aleatória, quando intui que o processo de amostragem terá capturado a variabilidade populacional e (iii) recusa-se a tirar conclusões ou a desenvolver previsões, uma vez que a variabilidade populacional não terá sido “retratada” por meio do processo de amostragem.

No geral o desempenho dos alunos da turma, evidencia um conflito entre a representatividade e a variabilidade do processo de amostragem, semelhante ao que Rubin, Bruce e Tenney (1991) observaram no estudo que envolveu alunos do ensino secundário.

8.3. Reflexão final

A concretização deste estudo contribuiu de modo expressivo para o meu desenvolvimento profissional. Na minha programação, vi-me confrontada com o tópico “Pla-

neamento estatístico” que teria de ser desenvolvido em torno da realização de alguns estudos estatísticos, com o propósito de fomentar, por parte dos alunos: (i) a compreensão de informação de natureza estatística e desenvolvimento de uma atitude crítica face a esta informação, (ii) a interpretação dos resultados obtidos, e (iii) a resolução de problemas e a comunicação em contextos estatísticos. Resolvi pesquisar sobre o ensino e aprendizagem da Estatística, tendo procedido à revisão das orientações curriculares mais recentes, tanto nacionais como internacionais, e à consulta de literatura relativa ao ensino e aprendizagem da Estatística, no que se refere ao desenvolvimento do raciocínio e pensamento estatístico, o que me permitiu aperfeiçoar os meus conhecimentos sobre o modo como ajudar os alunos mais eficazmente a ultrapassar as suas dificuldades e a melhorar o seu desempenho.

Ao desempenhar o papel de investigadora, tive também a possibilidade de viver em “primeira mão” as várias fases e desafios intrínsecos à realização de um ciclo investigativo. Com a realização desta investigação, tive oportunidade de perceber a importância de preparar com pormenor todos os momentos de um estudo desta natureza. Comecei a ter essa consciência durante a realização de três trabalhos de grupo que envolveram a observação e análise de algumas das minhas aulas, para três disciplinas no ano curricular do mestrado. Esses trabalhos foram muito importantes porque me fizeram refletir e reler uma vez mais textos sobre o ensino e a aprendizagem da Matemática e permitiram-me compreender que existem várias situações, que embora não sejam previsíveis, requerem uma resposta imediata e consistente por parte do professor, sendo então fundamental deter um grande leque de estratégias, relacionadas com o modo como os alunos aprendem determinado conceito ou ideia matemática, para além de um bom banco de exemplos e contraexemplos para fazer face aos desafios que os alunos sentem ao longo do processo de ensino-aprendizagem.

As maiores dificuldades que senti na realização deste estudo dizem respeito: (i) à construção de itens para os testes e à realização das entrevistas de modo a que aferissem determinados aspetos do raciocínio estatístico em análise, e (ii) à análise dos dados e sua categorização. Por vezes tornou-se difícil, dadas as múltiplas fontes, a sua gestão e, dada a natureza interdependente dos vários aspetos do raciocínio estatístico, tive dificuldade, por vezes, em definir dimensões de análise, e limitar alguns dados a apenas uma das dimensões em análise. A identificação de artigos e trabalhos de investigação a utilizar como quadro teórico, no âmbito do desenvolvimento da capacidade de planeamento estatístico também levantou problemas, nomeadamente pelo facto de ainda não

haver muitas investigações. Ainda, senti dificuldades na seleção e confronto de ideias de artigos e trabalhos de investigação a utilizar no quadro teórico e no confronto de ideias por eles apresentadas, nomeadamente no âmbito da capacidade de análise de dados. Durante a realização desta investigação, tive a oportunidade de realizar, com o meu orientador, três comunicações, duas em encontros nacionais e uma num encontro internacional de investigação em Educação Matemática. Estas comunicações tiveram como foco cada um dos estudos-caso e foram importantes momentos de reflexão sobre os dados obtidos e sua análise, dando posteriormente origem aos capítulos cinco, seis e sete deste estudo.

Este estudo constituiu uma grade mais-valia para o meu desempenho enquanto professora, pois ao assumir o duplo papel de professora e investigadora estive na situação privilegiada de construir uma sequência de aprendizagem, mas também de analisar e refletir sobre os seus efeitos ao nível das aprendizagens, nomeadamente sobre as dificuldades e os erros e equívocos mais comuns na aprendizagem da estatística. A sua realização tornou-me uma professora mais reflexiva e, sobretudo mais atenta à atividade realizada e às dificuldades sentidas pelos meus alunos.

Apesar dos resultados deste estudo não serem generalizáveis, considero que o estudo realizado contribui para o aumento do conhecimento sobre o desenvolvimento da capacidade de planeamento estatístico e análise de dados, por parte dos alunos, sendo por isso de interesse para a prática docente e para investigadores em Educação Matemática, uma vez que reúne um conjunto de experiências e informações que poderão ser pertinentes para estudos futuros. O estudo sugere uma abordagem que parte das vivências dos alunos na resolução de situações que envolvem incerteza, nomeadamente na perceção da variabilidade dos dados, para a compreensão do papel da Estatística nos processos de pesquisa e análise de dados, em vez de se centrar apenas em aspetos rotineiros e processuais, como tantas vezes acontece. Esta abordagem vivencial tem implicações muito positivas para a aprendizagem dos alunos (NCTM, 2007).

Penso que será importante desenvolver mais estudos à luz das orientações do programa de Matemática do ensino básico (ME, 2007), sobre a capacidade de planeamento e análise de dados dos alunos do 3.º ciclo, considerando a utilização de recursos tecnológicos, como por exemplo o Tinker Plots, o que permitiria um outro tipo de análise de caráter mais exploratório que a do presente estudo. Um outro aspeto a investigar seria a capacidade de planeamento e análise de dados, dos alunos do 3.º ciclo, considerando gerações de alunos com uma maior experiência estatística, ao nível do 1.º ou 2.º

ciclos, diferente da dos alunos deste estudo, pois quando a investigação mostra que, em determinada idade, existem dificuldades em compreender determinado conceito, não é possível saber que tipo de dificuldades teriam nesse mesmo conceito no caso de terem tido experiência prévias em estatística (Konold & Higgins, 2003).

Referências

- Abrantes, P., Ponte, J. P., Fonseca, H., & Brunheira, L. (1999). *Investigações matemáticas na sala de aula e no currículo*. Lisboa: APM e Projecto MPT.
- Abrantes, P., Serrazina, L., & Oliveira, I. (1999). *A Matemática na educação básica*. Lisboa: Ministério da Educação.
- André, M. (2009). Estudo de caso, uma alternativa de pesquisa em educação. Em C. Silva, V. Santos-Wagner, O. Marcilino & E. Foerste (Eds), *Metodologia da pesquisa em educação do campo: povos, territórios, movimentos sociais, saberes da terra, sustentabilidade*. Vitória, ES: UFES, Programa de Pós-Graduação em Educação.
- Australian Educational Council. (1991). *A national statement on mathematics for Australian schools*. Melbourne: Australian Educational Council and Curriculum Corporation.
- Arnold, P. (2008). What about the P in the PPDAC cycle? An initial look at posing questions for statistical investigation. *Proceedings of the 11th International Congress of Mathematics Education*, Monterrey, Mexico, 6-13 July, 2008. (Consultado a 21 de fevereiro de 2010) <http://tsg.icme11.org/tsg/show/15>
- Arnold, P. (2009). *Context and its role in posing investigative questions*. Paper presented at the Sixth International Forum for Research on Statistical Reasoning, Thinking and Literacy, Brisbane, Australia, 10-16 July, 2009.
- Arnold, P., & Pfannkuch, M. (2010). Enhancing students' inferential reasoning: From hands on to "movie snapshots". In C. Reading (Ed.), *Data and context in statistics education: Towards an evidence-based society. Proceedings of the Eighth International Conference on Teaching Statistics, Ljubljana, Slovenia, July 2010*. Voorburg, The Netherlands: International Statistical Institute. (Consultado a 2 de agosto de 2011) http://www.stat.auckland.ac.nz/~iase/publications/icots8/ICOTS8_2D2_ARNOLD.pdf
- Arteaga, P., Batanero, C., Díaz, C., & Contreras, J. M. (2009). El lenguaje de los gráficos estadísticos. *UNIÓN*, 18, p. 93-104.
- Bakker, A. (2004). Reasoning about shape as a pattern in Variability. *Statistics Education Research Journal*, 3(2), 64-83 (Consultado a 2 de fevereiro de 2010) [https://www.stat.auckland.ac.nz/~iase/serj/SERJ3\(2\)_Bakker.pdf](https://www.stat.auckland.ac.nz/~iase/serj/SERJ3(2)_Bakker.pdf)
- Bakker, A., Biehler, R., & Konold, C. (2005). Should young students learn about box plots? Em G. Burrill & M. Camden (Eds.), *Curricular Development in Statistics Education. Proceedings of the International Association for Statistical Education 2004 Roundtable* (pp. 163-173). Auckland: IASE. (Consultado a 4 de abril de 2010) https://www.stat.auckland.ac.nz/~iase/publications/rt04/4.2_Bakker_etal.pdf
- Batanero, C. (2000a). Hacia donde va la educación estadística? *Blaix*, 15, 2-13.(Consultado a 12 de junho de 2010) <http://www.ugr.es/local/batanero>
- Batanero, C. (2000b). Dificuldades de los estudiantes en los conceptos estadísticos elementales: el caso de las medidas de posición central . Em C. Loureiro, F. Oliveira & L. Brunheira (Eds.), *Ensino e aprendizagem da Estatística* (pp. 31-48). Lisboa: Sociedade Portuguesa de Estatística e Associação dos Professores de Matemática.
- Batanero, C. (2001). *Didáctica de la Estadística*. Granada: Universidade de Granada. (Consultado a 12 de junho de 2010) <http://www.ugr.es/local/batanero>
- Batanero, C., Estepa, A., & Godino, J. D. (1991). Análisis exploratorio de datos: Sus posibilidades en la enseñanza secundaria. *Suma*, 9, 25-31

- Barros, P.M., & Fernandes, J. A. (2001). Dificuldades de alunos (futuros professores) em conceitos de estatística e probabilidades. Em I. Lopes, J. Silva & P. Figueiredo (Orgs.), *Actas do Profmat 2001* (pp. 197 - 201). Vila Real: APM.
- Ben-Zvi, D. (2004). Reasoning about data analysis. Em D. Ben-Zvi & J. Garfield (Eds.), *The Challenge of Developing Statistical Literacy, Reasoning and Thinking*, (pp. 121 -145). Dordrecht, the Netherlands: Kluwer Academic Publishers.
- Ben-Zvi, D., Garfield, J., & Zieffler, A. (2006). Research in the statistics classroom: Learning from teaching experiments. Em *Thinking and Reasoning with Data and Chance*, (pp. 467 -479). Burrill, G. (Ed.) Reston VA: National Council of Teachers of Mathematics.
- Ben-Zvi, D., & Friedlander, A. (1997). Statistical thinking in a technological environment. Em J. B. Garfield & G. Burril (Eds.), *Research on the Role of Technology in Teaching and Learning Statistics* (pp. 45-55). Voorburg, The Netherlands: International Statistical Institute.
- Bill, A., Henderson, S., & Penman, J. (2010). Two Test Items to Explore High School Students' Beliefs of Sample Size when Sampling from Large Populations. Em L. Sparrow, B. Kissane, & C. Hurst (Eds.), *Shaping the future of mathematics education: Proceedings of the 33rd annual conference of the Mathematics Education Research Group of Australasia* (pp. 77 - 84). Fremantle: MERGA
- Bishop, A. J. & Goffree, F. (1986). Classroom organization and dynamics. Em B. Christiansen, A. G. Howson & M. Otte (Orgs.), *Perspectives on mathematics education* (pp. 309-365). Dordrecht: D. Reidel.
- Bogdan, R., & Biklen, S. K. (1994). *Investigação qualitativa em educação: Uma introdução à teoria e aos métodos*. Porto: Porto Editora. (Tradução portuguesa do original de 1991).
- Boaventura, M. G., & Fernandes, J. (2004). Dificuldades de alunos do 12.º ano nas medidas de tendência central: O contributo dos manuais escolares. Em *Actas do I Encontro de Probabilidades e Estatística na Escola* (pp. 103-126).
- Branco, J. (2000). Estatística no secundário: o ensino e seus problemas. Em C. Loureiro, F. Oliveira & L. Brunheira (Eds.), *Ensino e aprendizagem da Estatística* (pp. 11-30). Lisboa: Sociedade Portuguesa de Estatística e Associação dos Professores de Matemática.
- Carvalho, C. (1996). Algumas questões em torno de tarefas estatísticas com alunos do 7º ano. Em António Roque e Maria João Lagarto (Orgs.), *Actas do ProfMat 96* (pp.165-171). Lisboa: Associação de Professores de Matemática.
- Carvalho, C. (1998). Tarefas estadísticas e estratégias de resposta. Comunicação apresentada no *VI Encontro de Educação Matemática da Sociedade Portuguesa de Ciências da Educação*. Castelo de Vide, Portugal.
- Carvalho, C. (2001a). *Interação entre pares. Contributos para a promoção do desenvolvimento lógico e do desempenho estatístico no 7º ano de escolaridade*. Tese de Doutoramento. Universidade de Lisboa.
- Carvalho, C. (2001b). Literacia estatística. Comunicação apresentada na mesa redonda Literacia Estatística do *I Seminário de Ensino de Matemática – 14ª Conferência realizada pelo COLE, Campinas (São Paulo)*. (Consultado a 4 de abril de 2010) <http://cie.fc.ul.pt/membrosCIE/ccarvalho/doc8.pdf>.
- Carvalho, C. e César, M. (2000) As aparências iludem: reflexões em torno do ensino da estatística no ensino básico. Em C. Loureiro, F. Oliveira & L. Brunheira (Eds.), *Ensino e aprendizagem da Estatística* (pp. 212-225). Lisboa: Sociedade Portuguesa de Estatística e Associação dos Professores de Matemática.
- Carvalho, C. (2009). Reflexões em torno do ensino e da aprendizagem da Estatística. Em J. A. Fernandes, F. Viseu, M. H., Martinho & P. F. Correia (Orgs), *Actas do II Encontro de*

- Probabilidades e Estatística na escola* (pp. 22-36). Braga: Centro de Investigação em Educação da Universidade do Minho.
- Cobb, G. W. (1993) Reconsidering Statistics Education: A National Science Foundation Conference. *Journal of Statistics Education* v.1, n.1 (Consultado a 12 janeiro de 2010) <https://www.amstat.org/publications/jse/v1n1/cobb.html>
- Cobb, G. W. (1998). The objective-format question in statistics: Dead horse, old bath water, or overlooked baby? Paper presented at the annual meeting of the American Educational Research Association, San Diego, CA. (Consultado a 22 janeiro de 2010) <https://apps3.cehd.umn.edu/artist/articles/Cobb.pdf>
- Cobo, B. & Batanero, C (2000). La mediana en la educación secundaria. Un concepto sencillo? *UNO*, 24
- Cunha, M.H. e Almeida, M.R. (1996). Estatística nos 7º e 10º anos: avaliação de uma experiência. *Educação e Matemática*, 38, (21-28).
- Curcio, F. R. (1989). Developing graph comprehension: Elementary and middle school activities. Reston: National Council of Teachers of Mathematics.
- Erickson, Frederick (1986). *Qualitative methods in research on teaching*. Em M. C. Wittrock (Ed.), *Handbook of research on teaching* (pp. 119-161). Nova Iorque: MacMillan.
- Espinel, M. C., González, M. T., Bruno, A. & Pinto, J. (2009). Las gráficas estadísticas. Em L. Serrano (Ed.), *Tendencias actuales de la investigación en educación estocástica* (pp.57-74). Málaga: Gráficas San Pancrácio.
- Estrada, A., Batanero, C., & Fortuny, J. (2004). Un estudio sobre conocimientos de estadística elemental de profesores en formación. *Education Matemática*, 16 (1), 89-111.
- Fernandes, J. A., Alves, M. P., Machado, E., Correia, P., & Rosário, M. (2009). Ensino e avaliação das aprendizagens em Estatística. Em J. A. Fernandes, F. Viseu, M. H. Martinho & P. F. Correia (Orgs.) *Ensino e Aprendizagem de Probabilidades e Estatística – Actas do II Encontro de Probabilidades e Estatística na Escola* (pp. 52-71). Braga: Centro de Investigação e Educação da universidade do Minho.
- Fernandes, J.A., Carvalho, C. & Ribeiro, S. (2007). Caracterização e implementação de tarefas de Estatística: um exemplo no 7.º ano de escolaridade. *Revista Zetetiké*, 15(28), 27-61.
- Fernandes, J. A., Sousa, M. V., & Ribeiro, S. (2004). O ensino de Estatística no ensino básico e secundário: Um estudo exploratório. Em J. A. Fernandes, M. V. Sousa & S. A. Ribeiro (Orgs.), *Ensino e Aprendizagem de Probabilidades e Estatística – Actas do I Encontro de Probabilidades e Estatística na Escola* (pp. 165-193). Braga: Centro de Investigação e Educação da universidade do Minho.
- Flick, U. (2004). *Introducción a la investigación cualitativa*. Madrid: Ediciones Morata. (Tradução espanhola do original de 2002).
- Friel, S. N., Curcio, F & Bright, G. W. (2001). Making sense of graphs: critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education* 32 (2), 124-158.
- GAISE Report (2005). *Guidelines for Assessment and Instruction in Statistics Education* (GAISE) Report, A Pre-k-12 curriculum framework, August 2005 –American Statistical Association. (Consultado a 13 de junho de 2010) <http://it.stlawu.edu/~rlock/gaise/>
- Gal, I. (2002) Adult's Statistical Literacy: Meanings, Components, Responsibilities. *International Statistical Review*, v. 70, n. 1, 1-25.

- Gal, I., & Ginsburg, L. (1994). The role of beliefs and attitudes in learning statistics: toward and assessment framework . *Journal of Statistics Education*, 2(2). (Consultado a 26 de novembro de 2010) <http://www.amstat.org/publications/jse/v2n2/gal.html>
- Garfield, J., (1994). Beyond Testing and Grading: Using Assessment To Improve Student Learning. *Journal of Statistics Education* 2(1) (Consultado a 12 junho de 2010) www.amstat.org/publications/jse/v2n1/garfield.html
- Garfield, J., (2002). The Challenge of developing statistical reasoning. *Journal of Statistics Education* 10(3). (Consultado a 12 Junho de 2010) www.amstat.org/publications/jse/v10n3/garfield.html
- Garfield, J. & Chance, B. (2000). Assessment in Statistics Education: Issues and Challenges. *Mathematics Thinking and Learning*, 2, 99-125. (Consultado a 4 de abril de 2010) <https://app.gen.umn.edu/artist/articles/Garfield02.pdf>
- Garfield, J., & Gal, I. (1999). Assessment and statistics education: Current challenges and directions. *International Statistical Review*, 67(1), 1-12.
- Garfield, J., & Ben-Zvi, D. (2007). How students learn statistics revisited: A current review of research on teaching and learning statistics. *International Statistical Review*, 75(3), 372-396. (consultado a 12 Junho de 2010) http://www.ma.iup.edu/~tshort/rochester/12c_Joan_Garfield_ISR_paper_July_15_2007.pdf
- Graham, A., (2006). *Developing thinking in statistics*. London: Paul Chapman Publishing.
- Hawkins, A. (1997). Forward to basics! A personal view of developments in statistical education. *International Statistical Review*, 65(3), 280-287.
- Holmes, P. (2000). What sort of statistics should be taught in schools — and why? Em C. Loureiro, F. Oliveira, & L. Brunheira (Eds.), *Ensino e aprendizagem da Estatística* (pp. 49-56). Lisboa: Sociedade Portuguesa de Estatística e Associação dos Professores de Matemática.
- Holmes, P. (2002). Assessment: New Ways of Pupil Evaluation Using Real Data. *Teaching Statistics*, 24(3), 87-89.
- Jacobs, V. R. (1999). How do students think about statistical sampling before instruction? *Mathematics Teaching in the Middle School*, 5(4), 240-246, 263.
- Konold, C. & Higgins, T. (2003). Reasoning about data. Em G. Burrill (Ed.), *Research Companion to Principles and Standards of School Mathematics* (pp. 193-215). Reston, VA: NCTM.
- Lee, C., & Meletiou, M. (2003). Some difficulties of learning histograms in introductory statistics. *Joint Statistical Meetings- Section on Statistical Education*. (Consultado a 12 Junho de 2010) <http://www.statlit.org/PDF/2003LeeASA.pdf>
- Lopes, A. e Bernardes, A., Loureiro, C., Varandas, J., Viana, J. & Bastos, R. (1997). *Estatística Matemática 10*. Edições Contraponto.
- Loura, L. (2009). Organização e Tratamento de Dados no novo Programa do Ensino Básico. *Educação e Matemática*, 105, (46-49).
- Martins, M. E., Monteiro, C., Viana, J. P. & Turkman, M. A. (1997). *Estatística 10º ano de escolaridade*. Lisboa: ME - DES.
- Martins, M.E. & Ponte, J.P. (2010). *Organização e tratamento de dados*. Lisboa: ME-DGIDC.
- ME (1991a). *Organização Curricular e Programas. Ensino Básico 3º Ciclo*. Lisboa: Direção Geral do Ensino Básico e Secundário, Ministério da Educação, vol. I.
- ME (1991b). *Organização Curricular e Programas. Ensino Básico 3º Ciclo*. Lisboa: Direção Geral do Ensino Básico e Secundário, Ministério da Educação, vol. II.

- ME (1990). Organização Curricular e Programas. Ensino Básico 1º Ciclo. Lisboa: Direção Geral do Ensino Básico e Secundário, Ministério da Educação
- ME (2001). *Currículo nacional do ensino básico: Competências essenciais*. Lisboa: ME-DEB.
- ME (2007). *Programa de Matemática do Ensino Básico*. Lisboa: ME-DGIDC.
- Metz, K. E. (1999). Why sampling works or why it can't: Ideas of young children engaged in research of their own design. Em F. Hitt & M. Santos (Eds.), *Proceedings of the 21st annual meeting of the North American Chapter of the International Group for the Psychology of Mathematics Education*, Vol. 2 (pp. 492-498). Cuernavaca, Mexico: PME. (consultado a 27 de agosto de 2011) <http://www.matedu.cinvestav.mx/publicaciones/e-librosydoc/pme-procee.pdf#page=492>
- NCTM (1991). *Normas para o currículo e a avaliação em Matemática escolar*. Lisboa: APM e IIE. (Tradução portuguesa do original de 1989).
- NCTM (2007). *Princípios e normas para a Matemática escolar*. Lisboa: APM. (Tradução portuguesa do original de 1998).
- Morais, P. (2010). *Construção, leitura e interpretação de gráficos estatísticos por alunos do 9.º ano de escolaridade*. Dissertação de Mestrado. Universidade do Minho.
- Novaes, D. V. (2004). *A mobilização de conceitos estatísticos: estudo exploratório com alunos de um curso de tecnologia em turismo*. Dissertação de Mestrado em Educação Matemática, São Paulo. PUC-SP.
- Pfannkuch, M., & Horring, J. (2005). Developing statistical thinking in a secondary school: A collaborative curriculum development. Em G. Burrill & M. Camden (Eds.), *Curricular development in statistics education: International Association for Statistical Education 2004 Roundtable*, (pp. 204-218). Voorburg, The Netherlands: International Statistical Institute.
- Pfannkuch, M., & Wild, C. (2003). Statistical thinking: How can we develop it? *Proceedings of the 54th International Statistical Institute Conference*, Berlin, Germany, 13-20 August, 2003. Voorburg, The Netherlands: ISI. (Consultado a 21 de fevereiro de 2010) <http://www.stat.auckland.ac.nz/~iase/publications/3/3235.pdf>
- Pfannkuch, M. (2006a). Informal inferential reasoning. Em A. Rossman & B. Chance (Eds.), *Working cooperatively in statistics education: Proceedings of the Seventh International Conference on Teaching Statistics*, Salvador, Brazil. Voorburg, The Netherlands: International Statistical Institute. (Consultado a 24 de agosto de 2011) http://www.stat.auckland.ac.nz/~iase/publications/17/6A2_PFAN.pdf
- Pfannkuch, M. (2006b). Comparing box plot distributions: A teacher's reasoning. *Statistics Education Research Journal*, 5(2), 27-45, (Consultado a 24 de agosto de 2011) [http://www.stat.auckland.ac.nz/~iase/serj/SERJ5\(2\)_Pfannkuch.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ5(2)_Pfannkuch.pdf)
- Pfannkuch, M. (2008). Building sampling concepts for statistical inference: A case study. Em *ICME-11 Proceedings*, Monterrey, Mexico, July 2008. (Consultado a 24 de agosto de 2011) <http://tsg.icme11.org/tsg/show/15>
- Ponte, J. P. (2006). Estudos de caso em educação matemática. *Bolema*, 25, 105-132.
- Ponte, J. P. (2002). Investigar a nossa própria prática. Em GTI (Ed.), *Refletir e investigar sobre a prática profissional* (pp. 5-28). Lisboa: APM.
- Ponte, J. P., Brocado, J., & Oliveira, H. (2003). *Investigações matemáticas em sala de aula*. Belo Horizonte: Autêntica.
- Ponte, J. P., & Fonseca, H. (2000). A estatística no currículo do ensino básico e secundário. Em C. Loureiro, F. Oliveira, & L. Brunheira (Eds.), *Ensino e aprendizagem da Estatística*

- (pp. 179-194). Lisboa: Sociedade Portuguesa de Estatística e Associação dos Professores de Matemática.
- Ponte, J. P., & Sousa, H. (2010). Uma oportunidade de mudança na Matemática do ensino básico. Em GTI (Ed.), *O professor e o programa de Matemática do ensino básico* (pp. 11-41). Lisboa: APM.
- Reading, C. (1999) Understanding data tabulation and representation. Em O. Zaslavsky (Ed.), *Proceedings of the 23rd Conference of the International Group for the Psychology of Mathematics Education*, July, Vol 4, Haifa, Israel, pp. 97 — 104. (Consultado a 12 Junho de 2010) <http://www.une.edu.au/ehps/resources/pdfs/creading/b.pdf>.
- Rubenstein, R. & Thompson (2012). Reading visual representations. *Mathematics Teaching in the Middle School*, 17(9), 544-550.
- Rubin, A., Bruce, B., & Tenney, Y. (1991). Learning about sampling: Trouble at the core of statistics. Em D. Vere-Jones (Ed.), *Proceedings of the Third International Conference on Teaching Statistics*. Vol. 1 (pp. 314-319). Voorburg: International Statistical Institute.
- Ruiz, B.; Arteaga, P. & Batanero, C. (2009). Competencias de futuros profesores en la comparación de datos. Em L. Serrano (Ed.), *Tendencias actuales de la investigación en educación estocástica* (pp. 57-74). Málaga: Gráficas San Pancrácio.
- Scheaffer, R. (2000). Statistics for a new century. Em M. J. Burke & F. R. Curcio (Eds.), *Learning mathematics for a new century* (pp. 158-173). Reston, VA: NCTM.
- Shaughnessy, J. M. (1992). Research in probability and statistics: Reflections and directions. Em D. A. Grouws (Ed.), *Handbook of research and mathematics teaching and learning* (pp. 465-494). Nova Iorque: MacMillan.
- Shaughnessy, J. M. (1996). Discussion: empirical research on technology and teaching statistics. Em J. B. Garfield & G. Burrill (Eds.) *Research on the role of technology in teaching and learning statistics* (pp. 217-219). Voorburg, The Netherlands: International Statistical Institute.
- Shaughnessy, J. M., Garfield, J., & Greer, B. (1996). Data handling. Em A. Bishop et al. (Eds.), *International handbook of mathematics education* (v.1, pp. 205-237). Dordrecht: Kluwer Academic.
- Tuckman, B. W. (2002). *Manual de investigação em educação* (2^a Ed.). Lisboa: Fundação Calouste Gulbenkian.
- Tunnenbaum, P., & Arnold, R. (1997). *Excursions in modern mathematics*. Prentice-Hall.
- Turkman, M. A. A., & Ponte, J. P. (2000). Introdução. Em C. Loureiro, F. Oliveira, & L. Brunheira (Eds.), *Ensino e aprendizagem da Estatística* (pp. 5-9). Lisboa: Sociedade Portuguesa de Estatística e Associação dos Professores de Matemática.
- Watson, J., Fitzallen, N. E., Wilson, K., & Creed, J. (2008). The representational value of hats. *Mathematics Teaching in the Middle School*, 14(1), 4-10.
- Watson, J. M., & Moritz, J. B. (2000). Developing concepts of sampling. *Journal for Research in Mathematics Education*, 31(1), 44-70.
- Wild, C., & Pfunnkuch, M. (1999). Statistical thinking in empirical. *Internacional Statistical Review*, 67(3), 223-265.
- Wu, Y. (2004). *Singapore secondary school students' understanding of statistical graphs*. (Consultado a 12 Junho de 2010) <http://iase-web.org/documents/papers/icme10/Yingkang.pdf>

- Yin, R. (2010). *Estudo de caso: Planejamento e métodos*. (4ª Ed.) São Paulo: Artmed Editora. (Tradução portuguesa do original de 2009).
- Zieffler, A., Garfield, J., delMas, R. & Reading, C., (2008). A framework to support research on informal inferential reasoning. *Statistics Education Research Journal*, 7(2), 40-58. [http://www.stat.auckland.ac.nz/~iase/serj/SERJ7\(2\)_Zieffler.pdf](http://www.stat.auckland.ac.nz/~iase/serj/SERJ7(2)_Zieffler.pdf)

ANEXOS

Anexo 1 – Pedidos de autorização para realização do estudo

Exmo(a). Sr(a). Diretora da Escola Secundária
com 3º ciclo de Ferreira Dias

Cristina Maria Martins Roque, professora do 1º grupo, vem solicitar autorização para concretizar, nesta escola, o estudo de investigação em educação intitulado “Planeamento estatístico e análise de dados com alunos de 8.º ano”. Este estudo visa dar novos contributos sobre o modo como a resolução de tarefas com carácter exploratório e investigativo, promove as capacidades de planeamento e de análise de dados, em alunos do 8.º ano de escolaridade, e integra-se no âmbito do curso de Mestrado em Educação, na área de especialização em Didática da Matemática, Instituto de Educação da Universidade de Lisboa.

A concretização deste estudo implicará a recolha de dados de alunos do 8.ºano, referentes à disciplina que leciono. O trabalho empírico terá por base o desempenho dos alunos do 8.º C, ao longo do 2º Período nas diversas tarefas propostas no tópico “Planeamento estatístico”. Serão objeto de análise, nesta investigação: i) materiais produzidos dentro e fora da sala de aula pelos alunos, como, por exemplo, fichas de trabalho e relatórios; ii) transcrições de algumas das interações geradas entre eles; e iii) transcrições de entrevistas que lhes sejam realizadas, fora da sala de aula. A recolha de dados envolverá a gravação em áudio e/ou vídeo de alguns destes momentos.

Em todo o processo serão salvaguardados os direitos de privacidade e anonimato que assistem aos participantes e à própria escola, enquanto instituição. Os encarregados de educação serão informados sobre este estudo, sendo essencial o seu consentimento, para possibilitar a participação dos alunos que nele pretendam vir a colaborar.

12 de Janeiro de 2011

Pede deferimento,

(Cristina Maria Martins Roque)

Exm^o(a) Sr(a) Encarregado(a) de Educação

No âmbito do Mestrado em Educação na Especialidade de Didática da Matemática, no Instituto de Educação da Universidade de Lisboa, estou a proceder a um estudo que tem como objetivo compreender de que modo a resolução de tarefas de carácter exploratório e investigativo promove as capacidades de planeamento e de análise de dados dos alunos. Para o efeito necessito de observar e recolher dados sobre o trabalho dos alunos durante as aulas do tópico “Planeamento Estatístico”, que decorrerão no 2.º período. A recolha de dados basear-se-á na gravação em vídeo e áudio das aulas, bem como em entrevistas a alunos, feitas em horário extracurricular, de modo a compreender melhor como os alunos reagem face às tarefas propostas e tentando clarificar algum aspeto menos explícito das gravações.

Face ao exposto solicito autorização para proceder à recolha de dados, junto do seu educando, comprometendo-me desde já a garantir o anonimato dos alunos e a confidencialidade dos dados obtidos que serão utilizados apenas no âmbito da referida investigação, por mim e pelo meu supervisor, e para divulgação de resultados em encontros de natureza científica.

Agradecendo desde já a atenção dispensada, apresento os meus melhores cumprimentos.
Escola Secundária c/ 3.º ciclo de Ferreira Dias, Agualva, 17 de Fevereiro de 2011

A professora de Matemática do 8.º ---

A Diretora

(Cristina Roque)

(Leonídia Cunha)

✂

DECLARAÇÃO

Declaro que autorizo / não autorizo que meu (minha) educando(a) _____
_____ nº ____ turma --- do 8.º ano da Escola
Sec. c/ 3.º ciclo de Ferreira Dias, a participar na recolha de dados baseada na gravação
em vídeo e áudio das aulas, bem como em entrevistas a alunos, feitas em horário extra-
curricular, de modo a compreender melhor como os alunos reagem face às tarefas pro-
postas e tentando clarificar algum aspeto menos explícito das gravações, dirigida pela
professora Cristina Roque, no âmbito do seu estudo de Mestrado.

Agualva, __ / ____ / ____

Assinatura do Enc. de Educação: _____

Anexo 2 – Planificação da unidade de ensino

Tema	Organização e tratamento de dados	Propósito principal de ensino	Desenvolver nos alunos a capacidade de compreender e de produzir informação estatística bem como de a utilizar para resolver problemas e tomar decisões informadas e argumentadas.		
Tópico	Planeamento estatístico	Objetivos gerais	<ul style="list-style-type: none"> Compreender a informação de natureza estatística e desenvolver uma atitude crítica face a esta informação; Ser capazes de planear e realizar estudos que envolvam procedimentos estatísticos, interpretar os resultados obtidos e formular conjecturas a partir deles, usando linguagem estatística; Ser capazes de resolver problemas e de comunicar em contextos estatísticos. 		
Tópico e subtópicos	Objetivos específicos	Notas	Tarefas e modo de trabalho	Outros instrumentos	Blocos previstos
Planeamento estatístico <ul style="list-style-type: none"> Especificação do problema Recolha de dados População e amostra 	(A) Identificar os conhecimentos e as capacidades dos alunos antes da realização da unidade de ensino.	<p>O planeamento deve contemplar o tipo e o número de dados a recolher.</p> <p>Os dados são informação em contexto.</p> <p>Propor a recolha de dados de fontes primárias e secundárias, incluindo a Internet e publicações periódicas.</p> <p>Diversificar os métodos de recolha de dados: observação, experimentação e questionários.</p> <p>Ciclo investigativo, pertinência e fundamentação da informação extraída.</p>	Teste inicial (individual) (A)	Acetatos (Parte A e Parte B)	1
	(B) Distinguir entre população e amostra e ponderar elementos que podem afetar a representatividade de uma amostra em relação à respetiva população.		Tarefa 1 - Parte A e Parte B (em discussão com a turma) e Parte C (em grupo) (B) (C)		1
	(C) Identificar e minimizar possíveis fontes de enviesamento na recolha dos dados.		Tarefa 2 – Limpar dados (a pares) (E) (F)		0,5
	(D) Formular questões e planear adequadamente a recolha de dados tendo em vista o estudo a realizar.		Tarefa 3 – As vendas estão a correr bem? (em grupo) (B) (D) (G) (I) (K) (L) (M)		0,5
	(E) Distinguir dados de natureza qualitativa de dados de natureza quantitativa, discreta ou contínua.		Tarefa 4 (em grupo) Parte I e Parte II (dados das turmas de 8º ano em sala de aula e restantes recolhido pelos professores de matemática do 3º ciclo) (B) (C) (D) (G) (J) (K) Parte III e IV (B a M)		2
	(F) Analisar criticamente um conjunto de dados de modo a detetar imprecisões antes de os começar a tratar, pondo em causa a fiabilidade da análise posterior.		Tarefa 5 e anexos I e II (em grupo) (B) (C) (J) (K) (L) (M)	1	
	(G) Construir, analisar e interpretar representações dos dados (incluindo o histograma) e tirar conclusões.		Tarefa 6 (em grupo) Parte I e Parte II (B) (C) (D) (G) Parte III (em discussão com a turma) (J) (K) (L) (M)	2	
	(H) Compreender e determinar a mediana, os quartis e a amplitude interquartis de um conjunto de dados, e utilizar estas estatísticas na sua interpretação.		Teste final (individual) (N)	1	
	(I) Escolher as medidas de localização mais adequadas para resumir a informação contida nos dados.				
	(J) Comparar as distribuições de vários conjuntos de dados e tirar conclusões.				
(K) Responder às questões do estudo e conjecturar se as conclusões válidas para a amostra serão válidas para a população.					
(L) Análise da qualidade da previsão por comparação aos dados reais.					
(M) Relevância da resolução de problemas estatísticos					
(N) Identificar os conhecimentos e as capacidades dos alunos após a realização da unidade de ensino.					

Anexo 3 – Teste inicial e respetiva matriz de objetivos

Teste Formativo sobre OTD – 8.º Ano	Data: ____ / ____ / 2011 Duração: 90 minutos Profª Cristina Roque
<ul style="list-style-type: none">• ANTES DE COMEÇARES, LÊ ATENTAMENTE CADA QUESTÃO.• ESCRIVE AS TUAS RESPOSTAS DE FORMA LEGÍVEL, SE A TUA LETRA NÃO FOR LEGÍVEL, ESCRIVE EM LETRAS MAIÚSCULAS.• NO CASO DAS QUESTÕES 1.1., 4.3, 4.4, 4.5, 5.1 E 5.2 RESPONDE NO ENUNCIADO.• DEVES EXPLICAR/ JUSTIFICAR TODAS AS TUAS RESPOSTAS. A PROFESSORA TEM QUE PERCEBER COMO RESPONDESTES A CADA QUESTÃO.	

GRUPO 1 -Elaborou-se um estudo estatístico sobre as idades dos alunos do 3º ciclo de uma escola, seguindo os seguintes passos:

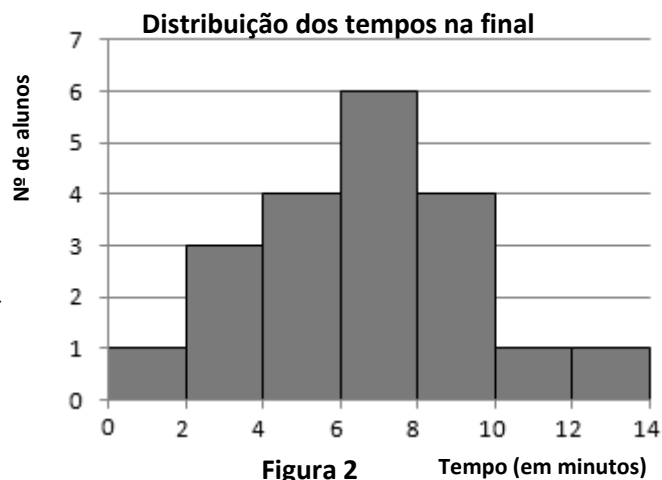
- (A) Recolheu-se junto dos Diretores de Turma o registo das idades dos alunos de todas as turmas do 3.º ciclo que tinham como número de turma 3, 10, 17 ou 24;
- (B) Contabilizou-se o número de alunos que têm cada uma das idades obtidas no passo anterior.
- (C) Calculou-se a percentagem de alunos que têm uma determinada idade em relação ao total de alunos em análise.
- (D) Registaram-se os dados obtidos numa tabela como a que se segue.

Tabela 1

Idades	Nº de alunos	Percentagem de alunos
11	5	
12	15	12
13	30	24
14	32	25,6
15	30	
16		8,8
17	2	
Total	125	100%

- 1.1. Completa os dados em falta na tabela 1 e apresenta os cálculos ou raciocínios efetuados.
- 1.2. Qual a idade mais usual no conjunto dos alunos em análise?

GRUPO 2 - A professora de Matemática propôs a realização de um concurso de *tangram* aos alunos das suas três turmas de 8.º ano. No histograma ao lado (Figura 2), registou-se o tempo de que cada aluno necessitou para completar a figura da fase final do concurso.



2.1. Quando se organizaram os tempos de prova, optou-se por construir quantas classes? Com que amplitude?

2.2. Quantos alunos participaram na final do concurso?

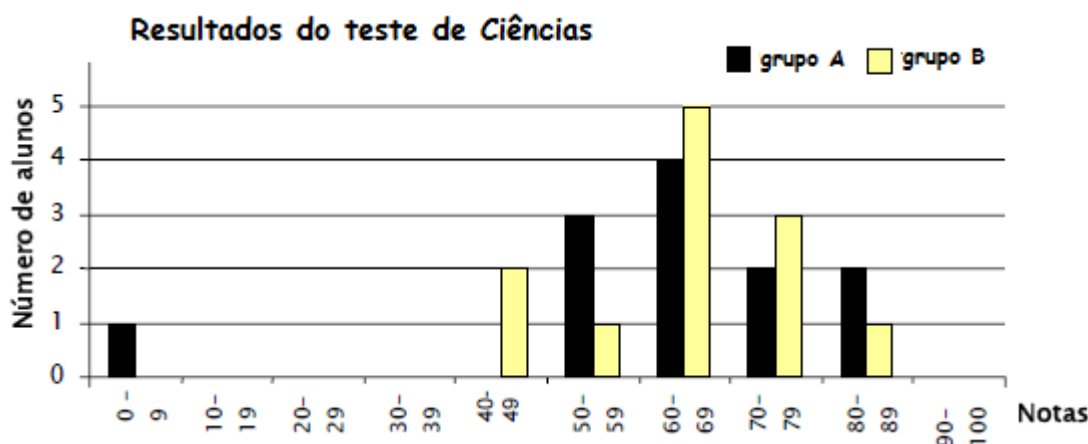
2.3. Quantos alunos demoram

2.3.1 6 ou mais minutos a completar a figura?

2.3.2. menos de 4 minutos?

2.4. Indica um possível tempo de prova para o aluno vencedor.

GRUPO 3 - O gráfico seguinte mostra os resultados de um teste de Ciências obtidos por dois grupos de alunos, designados por «Grupo A» e «Grupo B». A nota média no grupo A é de 62,0 e no grupo B de 64,5. Os alunos ficam aprovados neste teste se tiverem uma nota igual ou superior a 50.



Com base neste gráfico, o professor concluiu que o grupo B teve melhores resultados neste teste do que o grupo A.

Os alunos do grupo A não estão de acordo com o professor. Tentam convencer o professor de que o grupo B não teve necessariamente melhores resultados.

Utilizando o gráfico, apresenta um argumento matemático que possa ser utilizado pelos alunos do grupo A.

GRUPO 4 - Um grupo de alunos está a realizar um estudo estatístico sobre a *utilização do Messenger pelos alunos da escola*, pelo que selecionaram por sorteio 5 alunos de cada uma das 6 turmas do 8.º ano da escola e solicitaram o preenchimento de um inquérito. Todos os alunos selecionados responderam ao inquérito mas alguns responderam apenas à primeira questão pois não utilizam o Messenger.



Na quarta questão “Com quantas pessoas contactas via Messenger?” as respostas recolhidas foram as apresentadas no diagrama de caule e folhas (Figura 3).

0	7	7	8	9					
1	0	0	1	2	4	4	5	5	5
1*	6	7	7	8	8	8	9	9	
2	0	1	2	3	3	4			

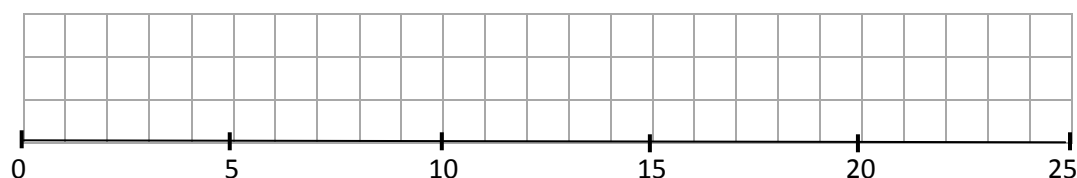
Figura 3

NOTA: Dado que o número de dados com caule é elevado optou-se por considerar a sua subdivisão em dois caules, sendo um 1 para as folhas menores ou iguais a cinco e 1* para as restantes folhas.

- 4.1. Quantos alunos responderam à quarta questão do inquérito?
- 4.2. Entre os alunos inquiridos qual o número mínimo e o número máximo de pessoas contactadas via Messenger?
- 4.3. No final, os alunos responsáveis pelo estudo pretendem incluir as seguintes frases de apoio à figura 3. Ajuda-os a completá-las, de modo a serem verdadeiras:

“Dos alunos em análise, 50% diz ter mais de contactos no Messenger.

A generalidade dos alunos inquiridos recorre ao Messenger para contactar um grande grupo de pessoas, apenas dos alunos não o fazem.”
- 4.4. No relatório final, os alunos responsáveis pelo estudo incluíram o diagrama de extremos e quartis relativos aos dados da figura 3; ajuda-os na sua construção.



4.5. Durante a apresentação do trabalho à turma, o grupo de alunos terminou com as seguintes conclusões: **Assinala com uma cruz a(s) que considera(s) válida(s).**

- “Cerca de 90 % dos alunos da escola utilizam o Messenger.”
- “Cerca de 25 % dos alunos da escola contactam, via Messenger, com menos de 10 pessoas.”
- “Cerca de 50% dos alunos do 8.º ano contactam, via Messenger, com mais de 15 pessoas.”
- “Cerca de 25% dos alunos do 8.º ano contactam, via Messenger, com 20 ou mais pessoas.”

GRUPO 5 - Como parte de um projeto os alunos de uma turma pretendem responder à questão “*As raparigas gastam menos dinheiro na escola do que os rapazes?*”. Para responder, pretendem recolher dados quanto aos gastos semanais na escola (bar, papelaria e refeitório), junto de 160 dos 1560 alunos da escola.



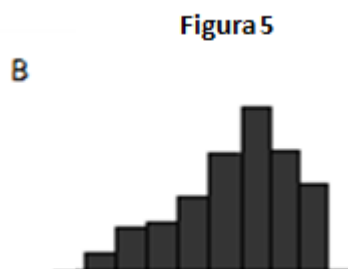
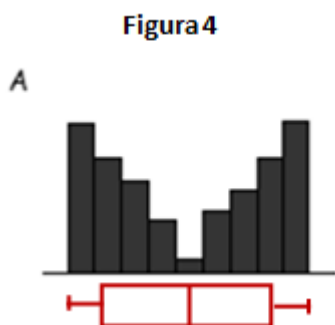
5.1. Qual ou quais das seguintes opções seria uma escolha adequada para obter dados de 160 alunos em representação dos 1560 da escola?

Assinala com uma cruz a(s) que considera(s) válida(s).

- Questionar 160 alunos à porta da escola no início da manhã;
- Questionar 160 alunos que são os nºs 1, 8,15, 22 de cada uma das quarenta turmas da escola;
- Questionar 160 alunos das seis turmas do 12º ano da escola;
- Questionar 160 alunos escolhidos por sorteio entre os 1560;
- Questionar 160 alunos que sejam amigos/conhecidos dos alunos responsáveis pelo estudo;
- Questionar 160 alunos na fila do bar da escola no primeiro intervalo da tarde.

5.2. Os alunos responsáveis pelo estudo optaram por formar dois grupos de trabalho que trabalharam em separado (grupo A e B). O grupo A, com base nos 55 dados que recolheu, elaborou o histograma e o respetivo diagrama de extremos e quartis indicado abaixo (Figura 4).

O grupo B construiu apenas o histograma com os 55 dados que recolheu. Constrói um esboço do diagrama de extremos e quartis associado ao histograma B (Figura 5).



Bom trabalho!
Profª Cristina Roque

Objetivos – teste inicial		Conceitos e procedimentos	Resolução de problemas/ Raciocínio
Representações	<p>Avaliação da capacidade de construção</p> <p>Adequação aos dados e ao que se pretende transmitir</p> <p>Interpretação das representações (leitura nos vários níveis)</p> <p>Sentido crítico face a representações enganosas</p> <p>Transformação de uma representação numa outra representação (reconhecer características gerais, como a forma, o centro e a dispersão da distribuição)</p>	<p>1.1., 2.1.</p> <p>2.2., 2.3, 2,4</p> <p>4.1., 4.2</p>	4.4., 5.2.
Medidas	<p>Avaliação da capacidade de determinação</p> <p>Adequação das medidas a utilizar face a diferentes condições, e de que modo representam ou não um conjunto de dados.</p> <p>Interpretação das medidas</p> <ul style="list-style-type: none"> reconhecer que medidas de ordem, e a sua respectiva localização fornecem informação sobre um conjunto de dados utilização das medidas estatísticas centrais e de dispersão são úteis para comparar conjuntos de dados. 	1.2, 4.3	3
Planeamento	<p>Formulação de questões</p> <ul style="list-style-type: none"> Reconhecimento da variabilidade Reconhecer fase a uma situação a necessidade de estudar a população ou uma amostra <p>Processo de amostragem</p> <ul style="list-style-type: none"> Análise crítica de estudos estatísticos face ao uso de amostras não representativas Reconhecer amostras representativas e conhecer aspetos que poderão gerar enviesamentos na escolha de uma amostra Reconhecer a amplitude das extrapolações: <ul style="list-style-type: none"> - saber o que se pode inferir a partir de uma amostra, - saber de que modo a amostra está relacionada da população de qual foi extraída 	5.1	4.5

Anexo 4 – Protocolo das entrevistas realizadas

INFORMAÇÕES AO ALUNO

- Relembrar o aluno de que o objetivo deste estudo é tentar perceber como os alunos pensam durante a resolução de tarefas estatísticas, por isso será solicitado a resolver tarefas oralmente e por escrito.
- Solicitar o aluno a verbalizar sempre do seu pensamento à medida que vai resolvendo as tarefas propostas, para que se possa perceber os raciocínios que o levam aos procedimentos que vai adotando e registando.
- Informar o aluno de que para resolver as tarefas e justificar os seus raciocínios para além da verbalização poderá recorrer a esquemas e/ou representações estatísticas, palavras e/ou cálculos.
- Comunicar ao aluno que a entrevista será gravada, já que seria muito difícil registar tudo o que se disser ao longo da mesma.
- Comunicar ao aluno que os resultados das resoluções das tarefas, não influenciarão a sua avaliação.
- Informar o aluno que será questionado durante a resolução das tarefas propostas, em função das produções que vai registando, com o objetivo de se perceber claramente como pensa (o porquê dos procedimentos e das opções a que recorre e das dificuldades que tem na resolução de cada uma das tarefas propostas).
- Informar o aluno de que cada tarefa será apresentada numa folha individual, onde deverá fazer os registos da sua resolução.

NOTAS PARA O ENTREVISTADOR

- Deverá entregar-se uma tarefa de cada vez, respeitando a sequência numérica das tarefas.
- Ler o tronco comum da cada tarefa que poderá conter dois ou mais itens e pedir ao aluno que transmita por palavras suas o significado da mesma, não permitindo que a leitura e interpretação sejam motivo de dificuldade na resolução da mesma.
- Colocar questões ao longo da entrevista de forma a tentar perceber o raciocínio do aluno, sempre que o mesmo não fique claro com as produções e verbalizações dos procedimentos que vai apresentando. O objetivo é perceber o porquê dos procedimentos e das opções a que recorre e das dificuldades com que se depara na resolução das tarefas propostas, a fim de se recolherem evidências sobre o seu domínio do planeamento estatístico e da análise de dados.

Em várias das tarefas que resolveste, o estudo estatístico não incidiu sobre a população, recorreram a amostras.

Para ti, o que significa uma amostra?

E amostra representativa?

Tomando alguns exemplos da tarefa, explica porque foi utilizada uma amostra?

- Caso não tenham sido contemplados aspetos relativos à adequação das representações, validade das conclusões, possibilidade de inferência para a população, etc..

Colocar as questões finais seguintes...

Que representação usarias em alternativa a ... Porquê?

Em várias das tarefas que resolveste, no final do estudo estatístico respondeu-se à questão que se estava a investigar?

Tomando alguns exemplos da tarefa, explica porque as conclusões válidas para a amostra serão ou não válidas para a população? Explica a tua resposta.

- Ter em conta a matriz de objetivos das tarefas da entrevista (tabela 1). A entrevista não deverá ultrapassar os 60 minutos. Findos estes, deverá ser interrompida e agendada uma data e hora para a sua continuidade e/ou conclusão.

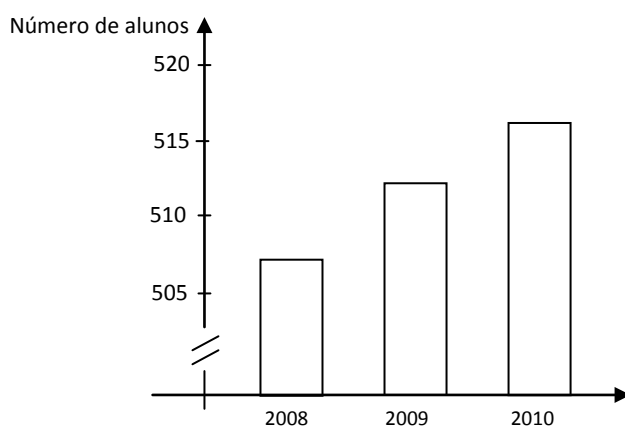
Anexo 5 – Tarefa da 1ª entrevista e respetiva matriz de objetivos

1. O coordenador do desporto escolar da escola ESCFD realiza anualmente um relatório: relativo ao número de participantes, modalidades praticadas, resultados em provas de concelho,.....



- a) No relatório de 2010 apresentou o seguinte gráfico e inclui o comentário lateral seguinte:

Desporto escolar – número de alunos envolvidos



“Como podemos observar entre 2008 e 2010, houve um aumento do número de alunos envolvidos nas atividades do desporto escolar, tendo duplicado o número de participantes.”

Consideras que a afirmação do coordenador é uma interpretação aceitável do gráfico?

Dá uma explicação que justifique a tua resposta.

- b) Pensando no seu próximo relatório, o coordenador pretende recolher informação necessária para demonstrar a evolução do desempenho dos participantes no desporto escolar. Para isso, optou por acompanhar 10 alunos que ingressaram no grupo de atletismo. Ao medir a altura em cm, que esse grupo de alunos consegue saltar na vertical após uma corrida, no início e após duas semanas de treino, obtiveram-se os valores seguintes.

Altura que saltam na vertical (em cm)										
Aluno	Ana	Bela	Carla	Diana	Elsa	Flora	Gina	Hilda	Inês	Joana
Antes	115	112	107	119	115	138	126	105	104	115
Depois	128	115	106	128	122	145	132	109	102	117



Com base nos dados o coordenador poderá constatar que as duas semanas de treino tiveram efeito ao nível do desempenho destes alunos? Explica a tua resposta.

2. Como parte de um projeto “A Saúde e os hábitos alimentares” os alunos de uma turma do 8º ano pretendem recolher dados junto de 200 dos 1025 alunos, quanto ao consumo semanal de refrigerantes.



- a) Indica de que modo, terão procedido para escolher um conjunto de 200 representativos da população em estudo.
- b) Indica um procedimento que leve à escolha de uma amostra enviesada (não representativa).
- c) Face aos dados recolhidos junto dos 25 alunos seleccionados para a sua amostra, quanto ao consumo semanal de refrigerantes construíram a seguinte tabela:

Nº de bebidas	Nº de alunos
Menos de 3 (<3)	62
4	56
5	40
6	24
7	10
8 ou mais (≥8)	8

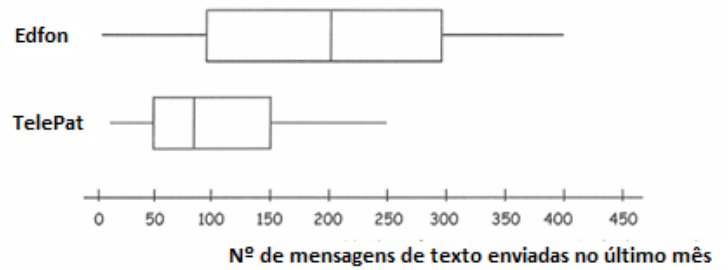
Os alunos querem construir um gráfico circular para representar os dados. E um deles determinou a amplitude dos ângulos para cada sector (fatia) do gráfico, do seguinte modo

$$\frac{62}{200} = \frac{31}{100} = 31^\circ \text{ para a primeira "fatia".}$$

De seguida marcou os ângulos obtidos com um transferidor num círculo à sua escolha e apercebeu-se que lhe sobrava espaço. Que incorreção foi cometida pelo aluno?

3. Há dois principais operadores de telecomunicações. Ambos disponibilizam serviço de mensagens de texto. Na operadora Edfone paga-se 20 cêntimos por cada mensagem de texto enviada. Na TelePat pode enviar no máximo 500 texto mensagens de texto com um custo mensal de 10 euros.

- a) O João está interessado em comparar o número de mensagens de texto enviadas pelos utilizadores da Edfone e TelePat, pelo que inquiriu 100 pessoas (cerca de igual número de utilizadores da Edfone e TelePat). Os resultados obtidos estão apresentados nos seguintes diagramas de extremos e quartis:



Sabendo que a questão a que o João pretendia responder era: "Em qual das operadas de telecomunicações, os utilizadores enviaram mais mensagens de texto no último mês?" Responder à pergunta do João. De que modo os diagramas de extremos e quartis mostram isso? Apresenta três argumentos para a fundamentação da resposta.

- b) O João reuniu informação relativa ao seu próprio envio de mensagens de texto mensalmente, tendo constatado que em cada um dos últimos 8 meses enviou:

85 85 80 83 16 84 87 83

Ficou na dúvida se deveria utilizar o valor numérico da média, da moda ou da mediana para resumir a informação recolhida neste 8 meses, que valor numérico(s) escolherias? Explica porque escolheste esse(s) valor numérico(s).

4. Em Lagutrop, foram efetuadas sondagens relativas à prática desportiva (prática de pelos menos duas horas semanais), os quatro jornais nacionais efetuaram separadamente sondagens a nível nacional.



- a) Quanto à percentagem de população que pratica desporto os resultados dessas quatro sondagens são os seguintes:

Jornal 1: 36,5% (sondagem realizada com base numa amostra de 500 cidadãos, inquiridos na rua em que se localiza o jornal);

Jornal 2: 41,0% (sondagem realizada com base numa amostra de 500 cidadãos, escolhidos ao acaso);

Jornal 3: 39,0% (sondagem realizada com base numa amostra de 1000 cidadãos, escolhidos ao acaso);

Jornal 4: 44,5% (sondagem realizada com base em 1000 leitores do jornal, que telefonaram para a redação para votar).

Qual é o jornal, cujos resultados darão, provavelmente, uma melhor previsão da percentagem de população que pratica desporto?

Indica para cada alternativa não selecionada uma razão da sua rejeição.

- b) O Jornal 2 na semana seguinte pretende publicar os dados relativos aos “gastos mensais com a prática desportiva”. O responsável pelo artigo face aos 500 dados que recolheu e que variam entre os € 5,7 e os € 52, pretende incluir um histograma para representar os dados.

Iniciando a primeira classe em €5, qual a amplitude de cada classe de forma que no mínimo sejam construídas 6 classes? Explica por palavras, cálculos ou esquemas a tua resposta.

5. O Carlos leu um artigo sobre o consumo de água, do qual se apresenta o seguinte excerto:

O inverno faz cair não só a temperatura, como o consumo de água.

Entretanto, especialistas advertem: a sensação de sede diminui, mas a importância da água para o organismo não. O consumo de cerca de dois litros deste líquido nesses dias mais frios é importante para evitar problemas renais, doenças de pele e desidratação.

A nutricionista Edna Garambone afirma que o cuidado deve ser redobrado com idosos e crianças, que desidratam com mais facilidade. Segundo ela, no inverno, o corpo de fato precisa de menos água do que no verão, daí sentirmos menos sede. O problema é que pessoas que já bebem pouco no calor, bebem menos ainda no frio. Ela dá uma receita capaz não só de hidratar, mas aquecer o corpo no friozinho do inverno: “Se não quer tomar água, uma boa saída é tomar chá. Outras fontes de hidratação, são as frutas como exemplo melão, laranja e pêra. Estas frutas são mais ricas em água do que outras frutas típicas do inverno, como abacate, maçã e uva”. A nutricionista avisa “Refrigerante, só se for numa ilha deserta e se não tiver outra escolha!”.



E decidiu estudar o seguinte problema “ Os alunos, da sua escola, cumprem a recomendação de consumo de água?”.

- Ao planear o estudo estatístico para obter resposta ao problema, o Carlos pensou em recolher informação relativa às seguintes variáveis estatísticas:
 - Número de peças de fruta consumidas diariamente;
 - Quantidade de líquidos consumida nos dois últimos dias (em litros);
 - Números de vezes que consome líquidos por dia;
 - Quantidade de água/chá consumida nos dois últimos dias (em litros);
 - Acompanha as refeições com o consumo de água (sim ou não?);
 - Número de refrigerantes consumidos diariamente;
 - Bebida preferida;

a) Quais das variáveis estatísticas anteriores selecionarias para o estudo?

b) Como classificas cada uma das variáveis consideradas pelo Carlos?

- Indica um esboço de como deve desenvolver o seu estudo, incluindo pelo menos os seguintes aspetos:
 - Recolha de dados: onde obter os dados? , como obter os dados? e que quantidade de dados?
 - Organização e tratamento de dados: Como irá representar os dados? Que medidas estatísticas irá usar para descrever os dados?
 - Análise e interpretação: Que tipo de conclusão poderá vir a extrair?

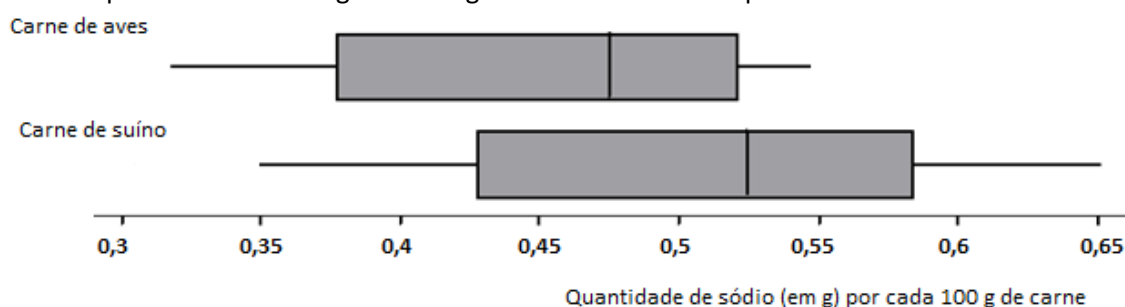
	Objetivos	Conceitos e procedimentos	Resolução de problemas/ Raciocínio
Representações	<p>Avaliação da capacidade de construção</p> <p>Adequação aos dados e ao que se pretende transmitir</p> <p>Interpretação das representações (leitura nos vários níveis)</p> <p>Sentido crítico face a representações enganosas</p> <p>Transformação de uma representação numa outra representação (reconhecer características gerais, como a forma, o centro e a dispersão da distribuição)</p>	2a, 4b	1a
Medidas	<p>Avaliação da capacidade de determinação</p> <p>Adequação das medidas a utilizar face a diferentes condições, e de que modo representam ou não um conjunto de dados.</p> <p>Interpretação das medidas</p> <ul style="list-style-type: none"> • reconhecer que medidas de ordem, e a sua respetiva localização fornecem informação sobre um conjunto de dados • utilização das medidas estatísticas centrais e de dispersão são úteis para comparar conjuntos de dados. 		3b 1b, 3a
Planeamento	<p>Formulação de questões</p> <ul style="list-style-type: none"> • Reconhecimento da variabilidade • Reconhecer fase a uma situação a necessidade de estudar a população ou uma amostra <p>Processo de amostragem</p> <ul style="list-style-type: none"> • Análise crítica de estudos estatísticos face ao uso de amostras não representativas • Reconhecer amostras representativas e conhecer aspetos que poderão gerar enviesamentos na escolha de uma amostra • Reconhecer a amplitude das extrapolações: <ul style="list-style-type: none"> - saber o que se pode inferir a partir de uma amostra, - saber de que modo a amostra está relacionada da população de qual foi extraída 	2a, 2b	5 4a 5 1b, 5
	<p>Conhecimento do ciclo investigativo: (i) formular questões e hipóteses; (ii) recolher dados (iii) analisar dados, e (iv) comunicar e interpretar resultados.</p>		5

Anexo 6 – Tarefa da 2ª entrevista e respetiva matriz de objetivos

1. Os alunos de uma turma de 12.º ano, alertados para o efeito nocivo do sódio, elemento que entra na composição do sal de mesa (cloreto de sódio), decidiram desenvolver um projeto: “O peso do sódio na alimentação”.

Sabe-se que o sal de mesa tem 40% de sódio, ou seja, 1 grama de sal contém 400 miligramas de sódio. O sódio está presente também em vários produtos industrializados que consumimos diariamente, como pão, queijo, cereais, bolachas, enlatados, etc.

- a) Um grupo optou por analisar a quantidade de sódio contida nas embalagens de salsichas enlatadas. Para tal, recolheram a quantidade de sódio/100g, por consulta da rotulagem, num conjunto de 50 embalagens de salsichas (cerca de igual número de embalagens com carne de suíno e carne de aves). Os dados obtidos estão apresentados nos seguintes diagramas de extremos e quartis:



Sabendo que a questão a que o grupo pretendia responder era: “Em que tipo de carne as embalagens de salsichas enlatadas contêm menos sódio?” Responde à questão do grupo. Apresenta três argumentos para fundamentar a tua resposta.

- b) A Rita, aluna da turma, ficou preocupada com o seu consumo diário de sódio, ao ter conhecimento de que “não se deve consumir mais do que 1,8 g a 2,4 g de sódio por dia”. Decidiu monitorizar o seu consumo diário de sódio:

1.º recolheu dados relativos ao consumo diário de sódio em 10 dias, que foram (em g):

2,4 3,1 4,5 1,8 2,1 2,7 2,5 2,0 3,0 3,2

2.º alterou os seus hábitos alimentares durante um mês, tendo evitado consumir conservas/enlatados e salgados variados.

3.º recolheu dados relativos ao consumo diário de sódio nos 10 últimos dias do mês em experiência, tendo obtido os seguintes valores (em g):

2,0 1,8 1,5 2,1 2,3 2,2 2,7 2,5 1,9 2,4

A Rita poderá concluir que a alteração dos seus hábitos alimentares contribuiu para uma diminuição do consumo de sódio? Explica a tua resposta.

2. Em Lagutrop, foram efetuadas sondagens para saber a **opinião da população quanto as condições de saúde e segurança nos locais de trabalho.**

a) Quatro jornais efetuaram separadamente sondagens a nível nacional. Os resultados dessas quatro sondagens, quanto à percentagem de população que consideraram ser más as condições de saúde e segurança nos locais de trabalho foram os seguintes:

Jornal 1	Jornal 2	Jornal 3	Jornal 4
35%	35,5%	39%	32%
Sondagem realizada com base em 1500 entrevistas de rua, entre as 10h- 12h, junto de cidadãos, com mais de 18 anos, escolhidos ao acaso.	Sondagem realizada, com base numa amostra de 1000 cidadãos, com mais de 18 anos, escolhidos ao acaso entre os eleitores.	Sondagem realizada com base em 1500 SMS enviadas por assinantes, com mais de 18 anos, para a redação do jornal.	Sondagem realizada com base em 1000 entrevistas digitais a cidadãos, com mais de 18 anos, escolhidos ao acaso numa rede social digital.

Qual é o jornal cujos resultados darão, provavelmente, uma melhor previsão da percentagem de população que consideram ser más as condições de saúde e segurança nos locais de trabalho, em Lagutrop?

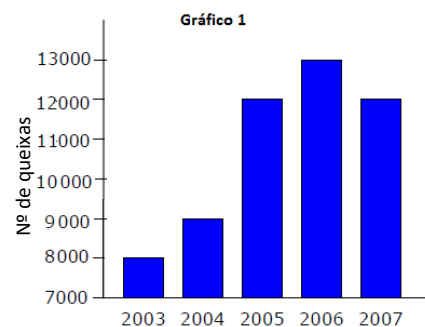
Indica para cada alternativa não selecionada uma razão da tua rejeição.

b) O jornal 2 pretende publicar, na semana seguinte, os dados relativos ao número de queixas, recebidas pelos serviços de inspeção das condições de saúde e segurança. Os dados recolhidos quanto ao número de queixas entre 2003 e 2007 foram os seguintes:

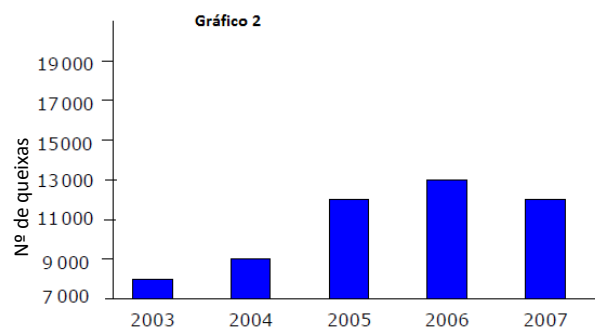
Ano civil	2003	2004	2005	2006	2007
Nº de queixas	8 000	9 000	12 000	13 000	12 000

A equipa gráfica propôs o gráfico 1 para ilustrar os dados.

O editor recusou-o alegando que deste modo se estava a acentuar o aumento do número de queixas, alarmando a população em demasia.



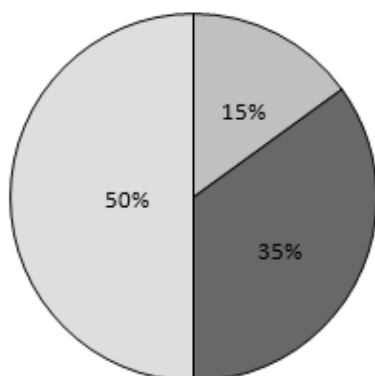
Quando lhe apresentaram o gráfico 2 também o rejeitou por considerar que deste modo se desvalorizava o aumento do número de queixas face à real situação.



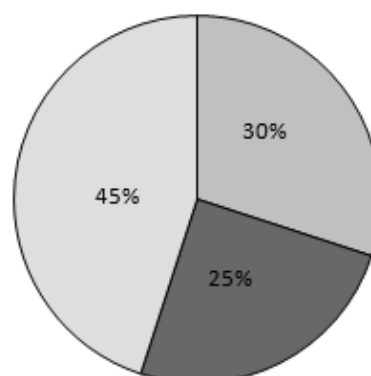
Identifica a(s) razão(ões) pela(s) qual (is) os gráficos 1 e 2 transmitem a informação de modo enviesado.

3. A escalada é uma modalidade cada vez mais apreciada, surgindo clubes em várias escolas. As competições de escalada realizam-se em estruturas artificiais de escalada e existem três tipos de competição – **Dificuldade**, **Velocidade** e **Bloco**.
- a) Quanto ao tipo de competição, os alunos do clube de escalada da escola A e da escola B indicaram o tipo de competição preferido. Os gráficos circulares seguintes ilustram os dados recolhidos.

Tipo de prova de competição preferida –
Clube da escola A



Tipo de prova de competição preferida –
Clube da escola B



□ Dificuldade
■ Velocidade
□ Bloco

a1) Indica duas informações que poderiam ser extraídas da leitura dos gráficos.

a2) Podemos afirmar que o número de alunos da escola A que prefere competir em provas do tipo **Bloco** é superior ao número de alunos da escola B a preferir esse mesmo tipo de prova?

- b) O melhor atleta de escalada da escola B, nos 8 últimos treinos para uma prova de competição do tipo **Dificuldade**, obteve os seguintes tempos (em segundos):

204 290 196 202 195 193 197 201

A professora responsável deverá utilizar que valor(es) numérico(s) para resumir os tempos dos últimos oito treinos? Explica a tua resposta.

- c) Numa competição do tipo **Bloco**, registou-se o tempo, em minutos que cada aluno demorou na resolução de dois problemas de escalada.

Os dados recolhidos foram os seguintes:

23	21	25	24	14	12	8
13	20	20	21	26	9	12
22	17	25	29	15	15	27
9	26	13	17	15	24	29
11	22	16	9	13	19	20

Pretende construir um histograma para representar os dados.

Qual a amplitude de cada classe de forma que no mínimo sejam construídas seis classes?

Explica por palavras, cálculos ou esquemas a tua resposta.

d) No clube de escalada de uma escola A os 45 alunos estão repartidos por dois grupos de 20 e 25 alunos, respetivamente. Os diagramas de caule-e-folhas seguintes contêm a informação recolhida relativa à altura de cada aluno, em centímetros.

Grupo I – 20 alunos

14	2	7	8	9			
15	0	0	1	7	7	7	8
16	3	5	5	6	7		
17	0	2	3	3			

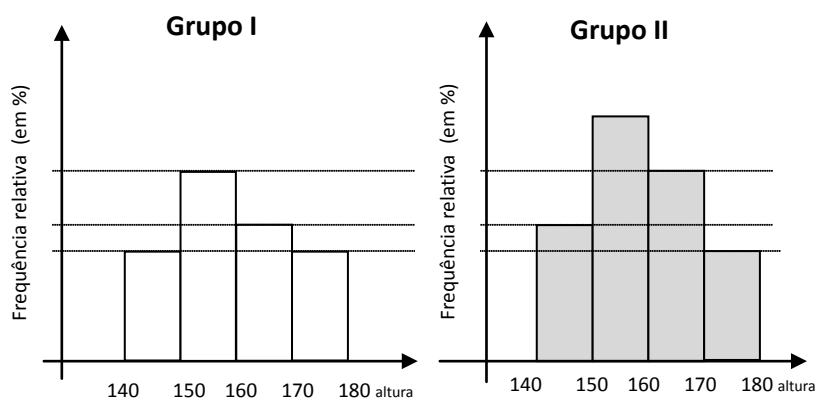
Grupo II -25 alunos

14	5	6	6	7	8				
15	1	2	3	3	4	5	7	8	9
16	0	0	4	5	5	6	6		
17	1	1	4	4					

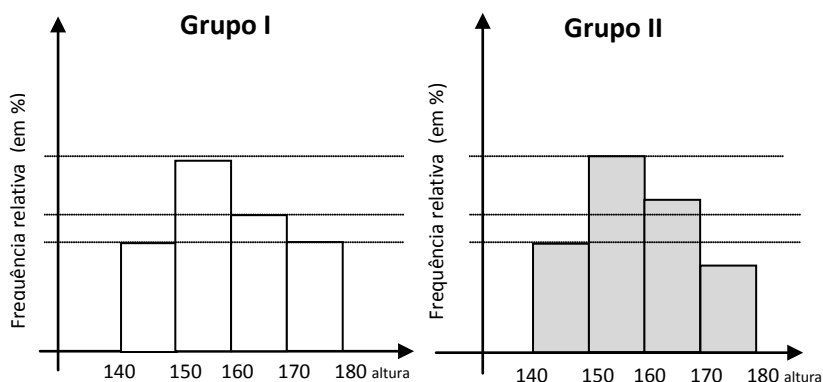
Qual dos seguintes conjuntos de histogramas (conj. A ou conj. B) corresponde à representação dos dados recolhidos nos grupos I e II respetivamente? Explica a tua resposta.

Nota: Em cada conjunto de histogramas a escala utilizada é idêntica para os dois grupos em análise.

Conjunto A – Distribuição das alturas



Conjunto B – Distribuição das alturas


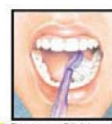





4. Na última consulta dentária a Rita leu um folheto que alertava para alguns cuidados relativos à higiene oral.

* A n o s	Frequência da escovagem dos dentes	Material utilizado na escovagem dos dentes	Execução da escovagem dos dentes	Dentífrico fluoretado
0 - 3	<u>2x dia</u> a partir da erupção do 1º dente: uma obrigatoriamente antes de deitar	<u>Gaze Dedeira</u> <u>Escova macia</u> de tamanho pequeno	<u>Pais</u>	<u>1000-1500 ppm</u> Quantidade idêntica ao tamanho da unha do 5º dedo da criança
3 - 6	<u>2x dia</u> uma obrigatoriamente antes de deitar	<u>Escova macia</u> de tamanho adequado à boca da criança	<u>Pais e/ou Criança</u> a partir do momento em que a criança adquire destreza manual, faz a escovagem sob a supervisão	<u>1000-1500 ppm</u> Quantidade idêntica ao tamanho da unha do 5º dedo da criança
M a i s d e 6 a n o s	<u>2x dia</u> uma obrigatoriamente antes de deitar	<u>Escova macia ou em alternativa média</u> de tamanho adequado à boca da criança ou do jovem	<u>Criança e/ou Pais</u> se a criança não tiver adquirido destreza manual, a escovagem tem que ter a intervenção ativa dos pais	<u>1000-1500 ppm</u> quantidade aproximada de 1 centímetro

Escovagem dos dentes

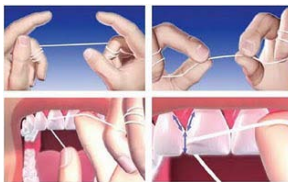
A escovagem dos dentes, para ser eficaz, ou seja, para remover a placa bacteriana, **precisa ser feita com rigor e demora 2 a 3 minutos**. Deve incluir a limpeza da língua.

Uso do fio dentário

A partir do momento em que há destreza manual (a partir dos 8 anos) é **obrigatório o uso do fio dentário**. O fio dentário deve ser **utilizado uma vez por dia**, de preferência à noite antes de deitar.



Idas ao dentista

A **1ª consulta** deve ser realizada quando toda a dentição de leite está completa, de modo a estabelecer regras preventivas de saúde oral e intercalar hábitos que possam ser prejudiciais. Quando existe boa saúde oral a **criança pode ser observada a cada 6 meses**. Em situações de elevado risco de cárie, esta periodicidade deve ser reduzida para **intervalos de 3 meses**.

* Informação do Programa Nacional de Promoção da Saúde Oral, pág. 6.

A Rita, aluna de uma escola Secundária com 3º ciclo, decidiu estudar o seguinte problema:

“Os alunos da escola têm práticas corretas quanto à higiene oral?”.

Para isso resolveu fazer-se um inquérito a um conjunto representativo de alunos da escola.

Indica um esboço de como deve desenvolver o estudo, incluindo pelo menos os seguintes aspetos:

- Planeamento /Recolha de dados:

Como deve proceder para obter uma amostra representativa? Com que dimensão?

Que questões incluir no inquérito Formula três ou quatro questões que consideres fundamentais a incluir no inquérito para conhecer as práticas/hábitos orais.

Que variáveis estão associadas às questões formuladas.

- Organização e tratamento de dados:

Como deve organizar os dados?

Como deve representar os dados?

Que medidas estatísticas irá usar para descrever os dados?

- Análise e interpretação:

Que tipo de conclusões poderá vir a extrair?

Objetivos		Conceitos e procedimentos	Resolução de problemas/ Raciocínio
Representações	<p>Avaliação da capacidade de construção</p> <p>Adequação aos dados e ao que se pretende transmitir</p> <p>Interpretação das representações (leitura nos vários níveis)</p> <p>Sentido crítico face a representações enganosas</p> <p>Transformação de uma representação numa outra representação (reconhecer características gerais, como a forma, o centro e a dispersão da distribuição)</p>	<p>3c</p> <p>3 a1</p>	<p>3a2</p> <p>2b</p> <p>3d</p>
Medidas	<p>Avaliação da capacidade de determinação</p> <p>Adequação das medidas a utilizar face a diferentes condições, e de que modo representam ou não um conjunto de dados.</p> <p>Interpretação das medidas</p> <ul style="list-style-type: none"> reconhecer que medidas de ordem, e a sua respetiva localização fornecem informação sobre um conjunto de dados utilização das medidas estatísticas centrais e de dispersão são úteis para comparar conjuntos de dados. 		<p>3b</p> <p>3b</p> <p>1a, 1b</p>
Planeamento	<p>Formulação de questões</p> <ul style="list-style-type: none"> Reconhecimento da variabilidade Reconhecer fase a uma situação a necessidade de estudar a população ou uma amostra <p>Processo de amostragem</p> <ul style="list-style-type: none"> Análise crítica de estudos estatísticos face ao uso de amostras não representativas Reconhecer amostras representativas e conhecer aspetos que poderão gerar enviesamentos na escolha de uma amostra Reconhecer a amplitude das extrapolações: <ul style="list-style-type: none"> saber o que se pode inferir a partir de uma amostra, saber de que modo a amostra está relacionada da população de qual foi extraída 		<p>4</p> <p>2^a</p> <p>4</p>
	<p>Conhecimento do ciclo investigativo: (i) formular questões e hipóteses; (ii) recolher dados (iii) analisar dados, e (iv) comunicar e interpretar resultados.</p>		<p>4</p>

Parte A- Porquê uma amostra?

Situação 1 Um empresário de uma fábrica de fósforos precisava fazer o controlo de qualidade dos fósforos produzidos pela sua fábrica.

Como o poderia fazer?

Naturalmente que não vai acender todos os fósforos produzidos...

População: fósforos produzidos

Amostra: um conjunto de fósforos

Unidade estatística: o fósforo



Situação 2

10% não sabem quem é o Presidente da República

(*Sondagem* in Expresso 15/03/97)

Não é possível inquirir todos os habitantes ...

Ficha Técnica

Sondagem efectuada entre os dias 6 e 31 de Janeiro. A **população** é constituída pelos habitantes de Portugal Continental, com idade entre os 18 e os 74 anos. A **amostra** é de 1964 indivíduos, entrevistados directamente, nas suas residências.

População: habitantes de Portugal Continental, com idade entre os 18 e os 74 anos

Amostra: 1964 indivíduos com idade entre os 18 e os 74 anos

Unidade estatística: um indivíduo com idade entre os 18 e os 74 anos

Parte B - O que é uma boa amostra?

Situação 1

Conjunto de fósforos produzidos num dia?

Conjunto de fósforos produzidos na mesma máquina?

Situação 2

Conjunto dos alunos da turma?

Conjunto de indivíduos que trabalham na Presidência da República?

Conjunto de indivíduos que trabalham em jornais?

População é um conjunto de unidades individuais, que podem ser pessoas ou resultados experimentais, com uma ou mais características comuns, que se pretendem estudar

Amostra de uma população é um sub-conjunto da população, que se estuda com o objectivo de tirar conclusões para a população.

Parte C – Escolha de amostras de uma população

1. Numa escola, para se investigar as preferências musicais dos alunos, entregou-se um questionário aos alunos que frequentam também uma Escola de Música.
Faz um comentário sobre a representatividade da amostra.
2. Uma empresa de publicidade pretendia perceber quais os anúncios da televisão que mais facilmente eram recordados pelas pessoas, tendo inquirido uma amostra de pessoas à saída de um supermercado num determinado dia.
Faz um comentário sobre a representatividade da amostra.
3. A Diretora de uma escola pretendia saber se os alunos estavam satisfeitos com a alimentação fornecida pela cantina da escola. Inquiriu todos os alunos com número ímpar.
Faz um comentário sobre a representatividade da amostra.
4. Para se estudar a característica "cor dos olhos" de uma determinada população, escolheu-se uma amostra constituída apenas por médicos que trabalhavam num hospital da zona.
Faz um comentário sobre a representatividade da amostra.
5. O que terias a dizer sobre a representatividade de uma amostra constituída apenas por médicos para se estudar os conhecimentos de Biologia de uma determinada população?
6. Uma escola tem 523 alunos do 9.º ano. Pretende-se fazer um estudo sobre os seus projetos quanto ao prosseguimento de estudos. Para isso resolveu fazer-se um inquérito que abranja uma amostra representativa.
 - a. Como obter essa amostra?
 - b. Elabora três ou quatro perguntas que consideres fundamentais estar no inquérito para conhecer a opinião dos alunos quanto ao prosseguimento dos seus estudos.

Anexo 8 – Folheto da oferta formativa da escola – esclarecimento para o item 6 da tarefa 1

Ensino Diurno CURSOS PROFISSIONAIS NÍVEL 4

Técnico de Electrónica, Automação e Comando
Técnico de Electrónica, Automação e Computadores
Técnico de Manutenção Industrial (Mecatrónica Automóvel)
Técnico de Turismo
(Consulte o folheto dos Cursos Profissionais)

Ensino Nocturno ENSINO RECORRENTE POR MÓDULOS – 12º ANO

Ciências e Tecnologias
Ciências Sociais e Humanas

CURSOS DE EDUCAÇÃO E FORMAÇÃO DE ADULTOS - EFA

EFA Básico B 2 – 6º Ano
EFA Básico B 3 – 9º Ano
EFA Secundário Escolar – 12º Ano
EFA Secundário Dupla Certificação - 12º Ano (Certificação escolar e profissional)
Técnico de Auxiliar de Saúde (Sujeito a confirmação pelo Ministério da Educação)
Técnico de Electrónica Automação e Computadores
Técnico de Informação e Animação Turística
Unidades de Formação (EFA) para finalizar o 12º Ano
Unidades de Formação de Curta Duração - UFCD Informática e Inglês

EDUCAÇÃO EXTRA ESCOLAR

Português Para Todos – PPT (Nível A1/A2 e B1/B2 – 150 horas/curso)

INFORMAÇÕES:



Direção
Secretaria
Serviço de Psicologia e Orientação Escolar

INTERNET:

Página da Escola



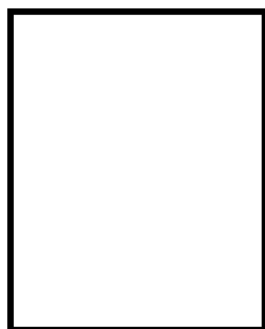
Ministério da Educação
www.min-edu.pt

Agência Nacional para a Qualificação, I.P. (ANQ)

www.anq.gov.pt

Novas Oportunidades

www.novasoportunidades.gov.pt



OFERTA FORMATIVA

ENSINO DIURNO
ENSINO BÁSICO - 3º CICLO
ENSINO SECUNDÁRIO
Cursos Científico-Humanísticos
Cursos Profissionais
ENSINO NOCTURNO
Cursos EFA – Básico e Secundário
Ensino Recorrente por Módulos - 12º Ano
Educação Extra-Escolar

Ensino Diurno ENSINO BÁSICO – 3º CICLO

Línguas Estrangeiras:

Francês, Inglês,
Alemão e Espanhol

(7º Ano - a formação de turmas com estas línguas estrangeiras está dependente da existência do número mínimo de alunos)

ENSINO SECUNDÁRIO CURSOS CIENTÍFICO-HUMANÍSTICOS

Formação Geral:

Português

Língua Estrangeira I / II / III:
Alemão / Espanhol / Francês / Inglês

Filosofia

Educação Física

Artes Visuais

Formação Específica:

Desenho A
(10º, 11º e 12º Ano)

Geometria Descritiva A
História da Cultura e das Artes
Matemática B

(O aluno escolhe duas disciplinas binais)

Ciências e Tecnologias

Formação Específica:

Matemática A
(10º, 11º e 12º Ano)

Física e Química A
Biologia e Geologia
Geometria Descritiva A

(O aluno escolhe duas disciplinas binais)

Línguas e Humanidades

Formação Específica:

História A
(10º, 11º e 12º Ano)

Geografia A
Matemática Aplicada às Ciências Sociais
Língua Estrangeira II / III - Alemão, Espanhol,
Francês ou Inglês (a formação de turmas com estas línguas estrangeiras está dependente da existência do número mínimo de alunos)
(O aluno escolhe duas disciplinas binais)

Ciências Socioeconómicas

Formação Específica:

Matemática A
(10º, 11º e 12º Ano)

Economia A
Geografia A
História B

(O aluno escolhe duas disciplinas binais)

NOTA: O funcionamento de qualquer opção depende do número mínimo de alunos para a formação da turma.



SERVIÇOS TÉCNICO-PEDAGÓGICOS

Serviço de Psicologia e Orientação (SPO)
Educação de Alunos Cegos e com Baixa Visão (Escola de Referência)
Equipa de Avaliação Especializada
Biblioteca (integrada na Rede de Bibliotecas Escolares)
Acção Social Escolar (ASE)

PROJECTOS DE APOIO AO ALUNO

GAPA - Gabinete de Apoio ao Aluno e Provedoria do Aluno
PM - Plano da Matemática
PES - Promoção e Educação para a Saúde
PNL - Plano Nacional de Leitura
Programa de Tutorias
Projecto de Educação pelos Pares
Projecto "Mais Sucesso Escolar"

ACTIVIDADES DE EXTENSÃO E COMPLEMENTO CURRICULAR

Clio - Clube de História e Património
Clube de Francês
Clube Europeu
Clube Multimédia
Clube Origami
Clube de Teatro
Desporto Escolar
Gaia - Ambiente e Desenvolvimento
NCT - Núcleo de Ciências e Tecnologias
Núcleo de Ciências
Programa Eco Escolas
Programa do Prémio Infante D. Henrique
Projecto Mun - Model United Nations
Projecto Radiação Ambiente
Projecto Seguranet

CURSOS PROFISSIONAIS

Os Cursos Profissionais são uma modalidade de educação, inserida no ensino secundário, que se caracteriza por uma forte ligação com o mundo profissional. A aprendizagem valoriza o desenvolvimento de competências para o exercício de uma profissão, em articulação com o sector empresarial local.

DESTINATÁRIOS

Os Cursos Profissionais destinam-se a jovens que concluíram o 9º ano de escolaridade ou formação equivalente, que procuram uma formação mais prática e voltada para o mercado de trabalho.

DURAÇÃO

Um ciclo de formação de 3 anos lectivos com Formação em Contexto de Trabalho (FCT).

CERTIFICAÇÃO ESCOLAR E PROFISSIONAL

A conclusão com aproveitamento de um curso profissional confere um diploma de 12º ano de escolaridade; Um certificado de qualificação profissional de nível 4 do Quadro Nacional de Qualificações.

PROSSEGUIMENTO DE ESTUDOS

A conclusão com aproveitamento de um curso profissional permite:

- acesso a um curso de especialização tecnológica - nível 5;
- acesso ao ensino superior.

MAIS INFORMAÇÕES

Ministério da Educação:

www.min-edu.pt/index.php?&white&pid=575

Agência Nacional para a Qualificação, I. P. (ANQ)

www.anq.gov.pt

Novas Oportunidades:

www.novasoportunidades.gov.pt

CURSOS PROFISSIONAIS



PROMOVER O SUCESSO EDUCATIVO
NUMA ESCOLA DE QUALIDADE

PLANO DE ESTUDOS

FORMAÇÃO SOCIOCULTURAL

Português	320	320	320	320
Língua Estrangeira	220	220	220	220
Área de Integração	220	220	220	220
Tec. de Informação e Comunicação	100	100	100	100
Educação Física	140	140	140	140
Subtotal	1000	1000	1000	1000

FORMAÇÃO CIENTÍFICA

Matemática	300	300	300	100
Física e Química	200	200	200	-
Geografia	-	-	-	200
História e Cultura das Artes	-	-	-	200
Subtotal	500	500	500	500

FORMAÇÃO TÉCNICA

Tecnologias e Processos	-	-	410	-
Organização Industrial	-	-	120	-
Desenho Técnico	-	-	170	-
Práticas Oficiais	-	-	480	-
Electricidade e Electrónica	349	312	-	-
Sistemas Digitais	186	222	-	-
Tecnologias Aplicadas	282	210	-	-
Automação e Comando	363	-	-	-
Automação e Computadores	-	436	-	-
Comunicar em Espanhol	-	-	-	180
Turismo - Informação e Animação Turística	-	-	-	402
Técnicas de Comunicação em Acolhimento Turístico	-	-	-	240
Operações Técnicas em Empresas Turísticas	-	-	-	358
Formação em Contexto de Trabalho (FCT)	420	420	420	420
Subtotal	1600	1600	1600	1600
Total de horas/curso	3100	3100	3100	3100

ESTRUTURA CURRICULAR

Os cursos apresentados têm uma estrutura curricular organizada por módulos.

OBJECTIVOS

Desenvolvimento de competências pessoais e profissionais para o exercício de uma profissão;
Privilegiar as ofertas formativas que correspondem às necessidades de trabalho locais e regionais.

SAÍDAS PROFISSIONAIS

Curso Profissional de Técnico de Automação e Comando

- os conhecimentos teóricos e práticos adquiridos permitem a estes profissionais desempenhar tarefas de carácter técnico relacionadas com a instalação, manutenção, reparação e adaptação de sistemas eléctricos, electrónicos, pneumáticos e hidráulicos de automação industrial.

Curso Profissional de Técnico de Automação e Computadores

- os conhecimentos teóricos e práticos adquiridos permitem a estes profissionais desempenhar tarefas de carácter técnico relacionadas com a instalação, manutenção, reparação e adaptação de equipamentos electrónicos de automação industrial e de computadores.

Curso Profissional de Técnico de Manutenção Industrial (Mecatrónica Automóvel)

- os conhecimentos teóricos e práticos adquiridos permitem a estes profissionais desempenhar funções na área da reparação automóvel.

Curso Profissional de Técnico de Turismo

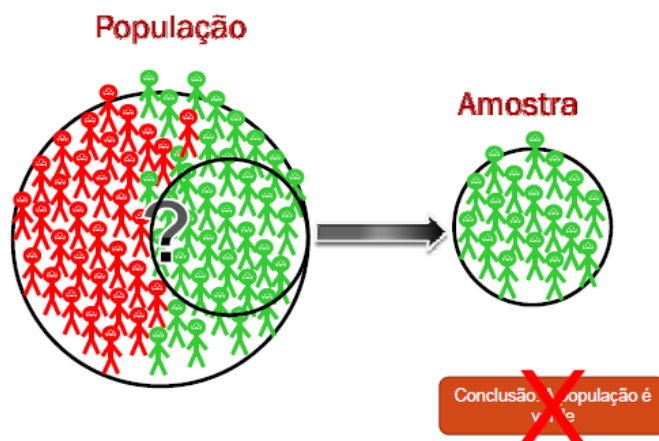
- desenvolver e executar serviços de informação, organização e animação de eventos em empresas de turismo;
- executar serviços em agências de viagens;
- executar os serviços de recepção e acolhimento em unidades turísticas.

O que é uma boa amostra?

É um subconjunto representativo da população (em quantidade e qualidade) sobre a qual incide a observação de um estudo estatístico.

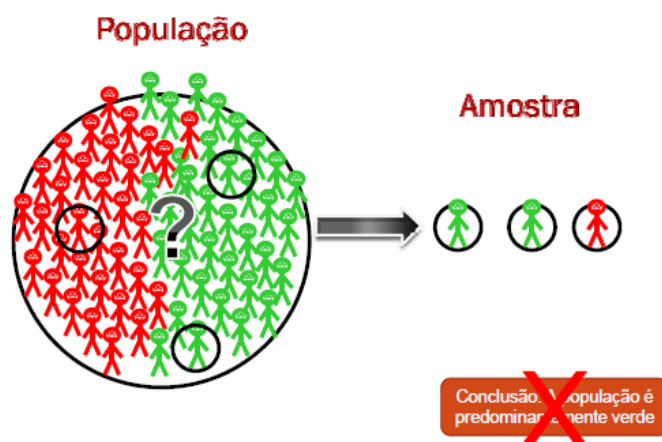
Situação 1

Boa quantidade sem qualidade



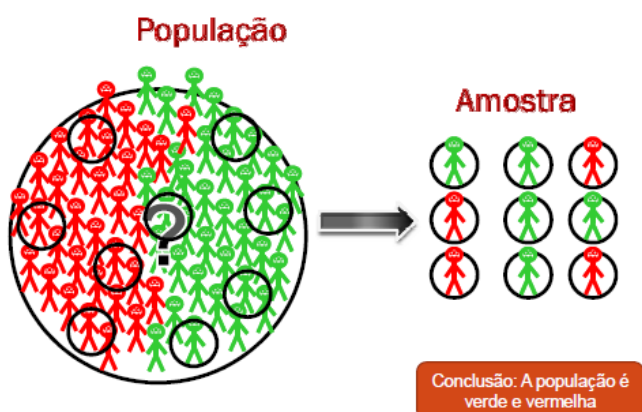
Situação 2

Boa qualidade sem quantidade



Situação 3

Boa qualidade e boa quantidade

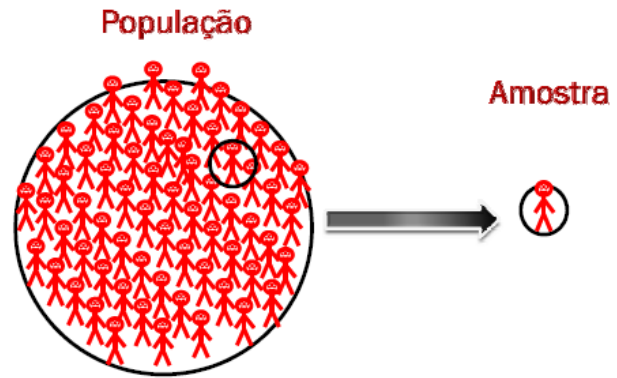


- Quando se recorre ao estudo de uma amostra de determinada população, tem-se com objetivo extrapolar para todos os elementos da população os resultados observados na amostra, pelo que a sua representatividade é fundamental.
- Com base nos resultados observados numa amostra, mesmo em condições ótimas de recolha de dados, permanece sempre alguma incerteza nas conclusões formuladas sobre a população.

O fato de população ser mais ou menos homogênea em relação à característica em estudo deve ser ponderado na “hora” da escolha de uma amostra....

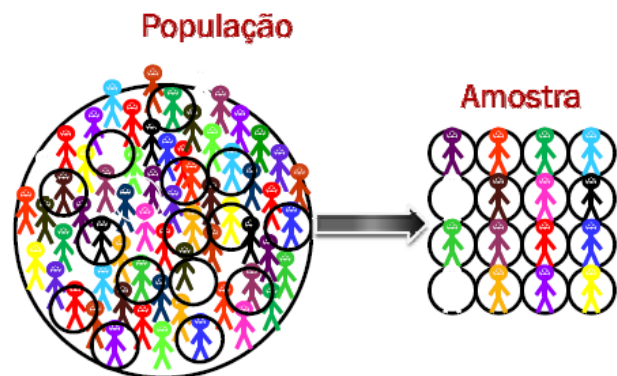
Situação 4

População homogênea



Situação 5

**População heterogênea
(grande variabilidade)**



de http://www.fop.unicamp.br/reuniao/downloads/3dia_Lauris_Calculo_Amostra.pdf

Anexo 10 – Tarefa 2: Limpar dados

É comum, quando se procede a uma análise de dados recolhidos verificar que estes contêm erros, acidentais ou não acidentais. Assim, antes de se proceder ao tratamento dos dados através de tabelas, gráficos ou do cálculo de medidas, deve-se olhar criticamente para os dados recolhidos, com o objetivo de os “limpar” dos erros. Por exemplo, se ao recolher informação sobre o tamanho do pé, se obtiver a informação de 300cm, obviamente que este valor está errado. Este erro pode ser acidental, nomeadamente ao digitar no computador o zero, carregou-se 2 vezes e ficaram 2 zeros. Se numa resposta sobre o ano de escolaridade, aparecer 1,2, também está errado, pois o ano de escolaridade tem de se exprimir na forma de um número inteiro. Estes erros podem ser acidentais, mas há outros que podem resultar de respostas dadas com pouco cuidado ou por brincadeira.

Na tabela que se apresenta a seguir, alguns alunos mais brincalhões entretiveram-se a alterar alguns dos dados de uma tabela que contém respostas de alunos do ensino básico. Procura detetar esses erros e quando possível, sugere alterações de forma a ter dados “limpos”:

Sexo	Data de nascimento	Ano de escolaridade	Naturalidade	Altura	Tamanho do pé	Disciplina ou atividade preferida	Distância de casa à escola
M	12-04-1991	5	Portugal	143	26	Educação musical	de 1 a 2 km
F	31/02/92	4	Portugal	132	22	Estudo do Meio	menos de 2 km
F	14-01-1991	5.00	Portugal	14.2	2.3	Educação Física	2.5423 km
M	07-09-1989	6	Portugal	136	25	Matemática	de 1 a 2 km
M	13-12-1991	4	Angola	128	24	Língua Portuguesa	de 1 a 2 km
M	14-03-2001	5	Portugal	140	67	Matemática	menos de 1 km
F	06-05-1989	7	Moçambique	142	24	Língua Portuguesa	de 3 a 5 km
F	15-08-1990	6	Portugal	138	21	Língua Portuguesa	85km
M	20-02-1990	6	Portugal	192	23	Matemática	de 1 a 2 km
M	19-05-1990	6	Portugal	140	20	Educação Física	de 1 a 2 km
O	29-06-1992	7	Lua	128	21	Estudo do Meio	3000km
M	09-10-1991	4	Cabo Verde	48	21	Língua Portuguesa	menos de 1 km
F	18-12-1990	5	Angola	135	21	Matemática	menos de 1 km
F	18-07-1991	0.5	Portugal	13.7	21	Ir para casa	de 3 a 5 km
M	03-06-1934	4	Portugal	129	22.5	Informática	menos de 1 km
F	13-02-1989	7	Moçambique	148	24	Matemática	de 1 a 2 km
F	15-09-1988	7	Portugal	150	23	Educação Física	de 1 a 2 km
F	07-08-1989	7	Portugal	140	20	Informática	menos de 1 km
M	08-06-1989	7	Angola	1520	24	Matemática	menos de 1 km
M	31/11/87	11	Marte	142	22	Informática	de 5 a 10 km
F/M	16-07-1988	8	Portugal	142	26	Chinês	de 2 a 3 km
F	28-04-1988	8	Portugal	145	26.5	Educação Física	1 kg
M	25-03-1992	4.1	Portugal	132.1	2.4.5	Matemática	menos de 1 km
M	26-02-1992	4	Portugal	130	21	Educação Física	menos de 1 km
F	08-07-1999	6	Portugal	142	22	Língua Portuguesa	de 2 a 3 km
M	23-05-1990	6	Cabo Verde	151	25.5	Matemática	de 2 a 3 km
M	01-03-1987	9	Angola	162	25	Educação Física	menos de 1 km
F	07-08-1991	6	Portugal	150	23	Educação musical	2 saltos
F	03-03-1992	4	Portugal	135	21	Informática	menos de 1 km

Anexo 11 – Tarefa 3: As vendas estão a correr bem?

Lê o texto atentamente.

- **Identifica as fases de um estudo estatístico.**
- **Identifica três aspetos que do teu ponto de vista foram importantes na realização deste estudo.**

As vendas estão a correr bem?

A proprietária da livraria LION, montou, num canto da sala, uma máquina de café, a título de experiência. Os lucros com a venda dos livros estavam a baixar, de modo que a venda do café talvez ajudasse a equilibrar o negócio. Passado algum tempo, decidiu averiguar se a instalação da máquina de café era vantajosa.

Assim, pediu a um grupo de alunos do 3.º ciclo, da escola ao lado da livraria, que costumavam passar por lá para folhear uns livros, para lhe fazerem um pequeno estudo sobre se valeria a pena continuar com a instalação da máquina de café. Prontamente aceitaram ajudar a senhora. Depois de discutirem de que modo se poderia abordar o problema, uma coisa não tinham dúvidas: precisavam de dados!

Felizmente a dona da livraria tinha registado as quantias auferidas com a venda do café, nos últimos 20 dias. Os jovens registaram os dados, tendo a senhora garantido que a ordem apresentada, era a ordem pela qual os dados tinham sido recolhidos: 300, 100, 200, 300, 100, 200, 200, 300, 300, 300, 200, 300, 400, 300, 300, 400, 500, 400, 400, 500

Na posse dos dados, começaram a pensar de forma poderiam extrair alguma informação que pudesse ajudar a proprietária. Nessa discussão, alguns dos alunos decidiram calcular algumas medidas, fazer algumas representações gráficas, ou seja, tentar arranjar alguns processos úteis de conseguir que os dados lhes devolvessem informação ...

Calcularam as medidas estatísticas - a moda, a média e a mediana. No entanto, um deles considerou que o cálculo só desses valores era pouco para caracterizar os dados que tinham, pelo que decidiram fazer também duas representações gráficas, nomeadamente uma representação em caule-e-folhas e gráfico de linhas (para cada dia a quantia auferida). Ao tomarem esta decisão, consideraram que deste modo a proprietária da livraria teria a informações complementares e de fácil compreensão.

Relatório

Questão – Foi-nos pedido que elaborássemos um estudo, para averiguar se a venda de café na livraria estaria a resultar. A proprietária da livraria, forneceu-nos uma tabela com as vendas nos 20 últimos dias, não nos tendo fornecido mais nenhuma informação, nomeadamente os custos com a manutenção da máquina, os custos do grão de café, ou com a mão-de-obra envolvida neste pequeno negócio.

Metodologia utilizada – Cálculo de algumas medidas estatísticas: moda, média e mediana -300 €

O facto de as três anteriores serem iguais, levou-nos a desconfiar que a distribuição dos dados era simétrica, o que se veio a confirmar com o gráfico de caule-e-folhas, que apresentamos a seguir:

10	0	0						
20	0	0	0	0				
30	0	0	0	0	0	0	0	0
40	0	0	0	0				
50	0	0						

A distribuição dos dados é simétrica, razoavelmente concentrada em torno da média de 300 euros, não se tendo registado quantias muito baixas ou muito elevadas.

Como a representação anterior não nos diz nada sobre a forma como o negócio está a evoluir, construiu-se um gráfico de linhas, que se apresenta a seguir:

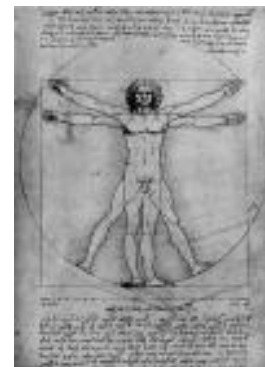


Este gráfico é bem informativo ao mostrar que o negócio tem tendência para crescer. O que não era visível no diagrama de caule-e-folhas.

Conclusão – Com a informação disponível, concluímos que com a venda do café a proprietária auferia uma quantia média diária de 300 euros, havendo uma tendência para este valor aumentar.

Anexo 12 – Tarefa 4: Um estudo na escola

Queremos caracterizar os alunos do ensino básico da escola no que diz respeito a algumas medidas. Para isso vamos recolher dados referentes às variáveis indicadas na tabela numa amostra representativa dos alunos da escola.



Parte I – Escolha da amostra

1. Qual é a população em estudo? E qual é a sua dimensão?
2. Qual é a dimensão da amostra com que vamos trabalhar?
3. Indica como proceder para escolher uma amostra representativa.
4. Indica um procedimento que leve à escolha de uma amostra enviesada (não representativa).

Parte II - Recolha de dados

5. Quais as variáveis a estudar? São qualitativas ou quantitativas? Contínuas ou discretas?
6. Para cada variável, organiza os dados em tabelas de frequência.

Aluno(M/F)	Altura (cm)	Envergadura (cm)	Nº sapato

Parte III – Tratamento dos dados

7. Para cada variável, determina a amplitude de cada uma das distribuições, a média, a moda e a mediana.
8. Constrói, para a variável envergadura o diagrama de caule-e-folhas.
9. Para a variável altura, determina os quartis e faz o diagrama de extremos e quartis.
10. Para a variável altura constrói um histograma.

Parte IV - Conclusões

11. Elabora um relatório com a caracterização da amostra relativamente às variáveis estudadas.

Anexo 13 – Tarefa 5: Frequência das vogais na língua portuguesa

Uma editora de jogos vai introduzir no mercado um jogo cujo objetivo é construir palavras. As letras são colocadas em fichas que os jogadores colocam em cada jogada. Para construir um número adequado de fichas é preciso conhecer a frequência relativa com que ocorre cada uma das letras na língua portuguesa. Como algumas das vogais são as letras mais frequentes, a empresa começou por organizar um estudo sobre a frequência de cada uma das vogais.

1. Vamos fazer o mesmo estudo a partir do texto “A Rádio escola” ...

Escolhe quaisquer 5 linhas consecutivas do texto.

Considerando as linhas que escolheste preenche a tabela com a frequência com que aparece cada uma das vogais.

Variável/ vogal	Frequência absoluta na amos- tra do grupo A	Frequência absoluta na amos- tra do grupo B	Frequência absoluta na amos- tra do grupo C	Frequência absoluta na amos- tra do grupo D	Frequência absoluta na amos- tra do grupo E	Frequência absoluta na amos- tra da turma	Frequência relativa na amostra da turma
a							
e							
i							
o							
u							
Total de vogais encontradas							

2. Qual é a vogal mais frequente?
3. Se a empresa quiser que o seu jogo tenha 200 peças com vogais, qual a quantidade de fichas que deve haver com cada uma das vogais?
4. Se repetíssemos a mesma experiência, mas com um outro texto obteríamos exatamente a mesma percentagem para cada uma das vogais?
5. Escolhe agora quaisquer 5 linhas consecutivas do texto: “Um texto sem a letra A” e analisa a frequência de cada uma das vogais. Compara com os resultados obtidos no texto anterior.
6. Compara os resultados da amostra da turma com os resultados da frequência de cada uma das letras do alfabeto na língua portuguesa. Procura na internet, no *wikipédia* esta informação.

http://pt.wikipedia.org/wiki/Alfabeto_portugu%C3%AAs

O que podes concluir sobre a qualidade da amostra utilizada?

A RÁDIO ESCOLA

Este ano letivo a turma do Pedro queria aventurar-se num novo projeto – a Rádio Escola.

O Pedro queria reunir, entre todas as turmas da escola, pequenos grupos de alunos que alternassem, semanalmente, na condução da Rádio. Iria também precisar da colaboração de um professor que ficasse responsável como coordenador do projeto, talvez um professor de Educação Musical.

Na Rádio Escola podiam divulgar músicas, dar a conhecer as atividades da escola, as opiniões dos alunos, fazer dedicatórias... Numa fase mais avançada, podiam também enriquecer os programas da rádio com entrevistas, notícias, associar-se a campanhas de ajuda e solidariedade, enfim, tudo atividades que caracterizam um verdadeiro programa de rádio. E, claro, não podiam esquecer a dinamização de uma página na Internet. Ora bem, ideias não faltavam! A Clara já sabia trabalhar muito bem com a aparelhagem da escola. Agora, só teriam de lhe dar bom uso e com maior frequência.

A Ana lembrou que para a inauguração da Rádio podiam convidar um grupo de música para vir tocar na escola...

– Mas que tipo de música vamos escolher? – perguntou a Bela.

Bem, mais vale fazer um estudo estatístico para conhecermos melhor os gostos musicais dos nossos colegas.

O que é que vocês acham? – lembrou o Pedro.

Começaram pelos alunos do 3.º ciclo. Seria essa a primeira população a investigar. O estudo ainda deu algum trabalho, mas foi muito giro fazê-lo porque tiveram mesmo de trabalhar em equipa, envolvendo várias disciplinas!

Primeiro, na aula de Matemática, escolheram uma amostra representativa da população em estudo; depois, na aula de Língua Portuguesa, elaboraram um inquérito; seguiu-se a organização dos dados em tabelas de frequências e gráficos, trabalho que também foi feito em Matemática.

Nesta fase, o Pedro comentou que a banda portuguesa com maior probabilidade de ficar em 1.º lugar seria os Da Weasel. Gerou-se uma ligeira discussão e fizeram-se apostas. A turma dividiu-se em dois grupos: o grupo dos Da Weasel e o grupo do Boss AC. Apostaram um lanche no bar da escola!

A parte final do estudo foi a análise dos resultados e o registo das conclusões. Nas aulas de Estudo Acompanhado, usaram o programa Excel e voltaram a fazer os gráficos, desta vez no computador.

Finalmente, afixaram os resultados na sala de convívio dos alunos.

As conclusões foram interessantes, coincidindo até com os gostos da turma do Pedro.

Fantástico! As três bandas portuguesas que os alunos do 3.º ciclo mais ouviam eram os Da Weasel, Boss AC e Buraka Som Sistema. O grupo do Pedro ganhou a aposta. Em relação aos estrangeiros, ficaram no topo da lista os Snow Patrol, Muse e Bob Sinclair.

Agora, só faltava estudar os gostos dos alunos do 2.º ciclo... O mais provável é que fossem diferentes! Tinham muito trabalhinho pela frente, mas o que é certo é que estavam todos animados e a Rádio Escola “tinha pernas para andar”. Só precisava do empenho e dedicação de todos!

Ah! E o Pedro ficou encarregado de escolher a ementa do tal lanche para o seu grupo, “cerimónia” que teria lugar no sábado da semana em que fossem publicadas as preferências musicais dos alunos do 2.º ciclo.



Um texto sem a letra "A".

Isto é possível?
É possível sim...

Sem nenhum tropeço posso escrever o que quiser sem ele, pois rico é o português e fértil em recursos diversos, tudo isso permitindo mesmo o que de início, e somente de início, se pode ter como impossível.

Pode dizer-se tudo, com sentido completo, como se isso fosse mero ovo de Colombo, desde que se tente. Sem se inibir, pode muito bem o leitor empreender este belo exercício dentro do nosso fecundo e peregrino dizer português, puríssimo instrumento dos nossos melhores escritores e mestres do verso, instrumento que nos legou monumentos dignos de eterno e honroso reconhecimento.

Trechos difíceis resolvem-se com sinónimos.

Observe-se bem: é certo que, em se querendo, esgrime-se sem limites com este divertimento instrutivo.

Brinque-se mesmo com tudo.

É um belíssimo desporto do intelecto, pois escrevemos o que quisermos sem o "E" ou sem o "I" ou sem o "O" e, conforme meu exclusivo desejo, escolherei outro, discorrendo livremente, por exemplo sem o "P", "R" ou "F", o que quiser escolher. Podemos, em corrente estilo, repetir um som sempre ou mesmo escrever sem verbos.

Com o concurso de termos escolhidos, isso pode ir longe, escrevendo-se todo um discurso, um conto ou um livro inteiro sobre o que o leitor melhor preferir.

Porém, mesmo sem o uso pernóstico dos termos difíceis, muito e muito se prossegue do mesmo modo, discorrendo sobre o objeto escolhido, sem impedimentos.

Deploro sempre ver moços deste século inconscientemente esquecerem e oprimirem hoje o nosso português, culto e belo, querendo substituí-lo pelo inglês. Porquê?

Cultivemos o nosso polifónico e fecundo verbo, doce e melodioso, porém incisivo e forte, messe de luminosos estilos, voz de muitos povos, escrínio de belos versos e de imenso porte, ninho de cisnes e de condores.

Honremos o que é nosso, oh moços estudiosos, escritores e professores!

Honremos o digníssimo modo de dizer que nos legou um povo humilde, porém viril e cheio de sentimentos estéticos, púgil, de heróis e de nobres descobridores de mundos novos!

Autor: Desconhecido.

Frequência de ocorrência de letras na língua portuguesa.

A lista da esquerda aparece por ordem alfabética e a da direita por ordem decrescente de frequência

Letra Frequência

A	14,63%
B	1,04%
C	3,88%
D	4,99%
E	12,57%
F	1,02%
G	1,30%
H	1,28%
I	6,18%
J	0,40%
K	0,02%
L	2,78%
M	4,74%
N	5,05%
O	10,73%
P	2,52%
Q	1,20%
R	6,53%
S	7,81%
T	4,34%
U	4,63%
V	1,67%
W	0,01%
X	0,21%
Y	0,01%
Z	0,47%

Letra Frequência

A	14,63%
E	12,57%
O	10,73%
S	7,81%
R	6,53%
I	6,18%
N	5,05%
D	4,99%
M	4,74%
U	4,63%
T	4,34%
C	3,88%
L	2,78%
P	2,52%
V	1,67%
G	1,30%
H	1,28%
Q	1,20%
B	1,04%
F	1,02%
Z	0,47%
J	0,40%
X	0,21%
K	0,02%
W	0,01%
Y	0,01%

Anexo 14 – Tarefa 6: Previsões

Com esta tarefa pretende-se fazer recomendações ao bar da escola quanto ao número de unidades de cada uma das bebidas que deveria ser encomendado por mês.

Parte I – Escolha da amostra

Nos dois intervalos da manhã e nos dois intervalos da tarde cada grupo questiona um total de 20 alunos da fila do bar sobre se vão comprar alguma bebida e no caso afirmativo qual a bebida que vão comprar. Cada grupo escolhe um dia da semana para realizar a recolha dos dados.

1. Qual é a população em estudo? E qual é a sua dimensão?
2. Qual é a dimensão da amostra com que vamos trabalhar?
3. Indica como proceder para escolher uma amostra representativa.
4. Indica um procedimento que leve à escolha de uma amostra enviesada.

Parte II – Recolha de dados

5. Preenche a tabela com os resultados recolhidos por cada um dos grupos.

Grupo 1		Grupo 2		Grupo 3		Grupo 4		Grupo 5		...
Nome da bebida	Frequência	Nome da bebida	Frequência	Nome da bebida	Frequência	Nome da bebida	Frequência	Nome da bebida	Frequência	
Alunos que não compram bebida		Alunos que não compram bebida		Alunos que não compram bebida		Alunos que não compram bebida		Alunos que não compram bebida		

Parte III – Previsão

6. A partir dos dados recolhidos prevê a quantidade de bebidas que o bar deve encomendar mensalmente para poder responder aos pedidos dos utentes.
7. Confirma as tuas previsões com os dados reais.
8. Explica por que é que se pode considerar a tua previsão uma boa ou má previsão.

Anexo 15 – Teste final e respetiva matriz de objetivos

Teste Sumativo sobre OTD – 8.º Ano	Data: ____ / ____ / 2011 Duração: 90 minutos Profª Cristina Roque
---	--

- ANTES DE COMEÇARES, LÊ ATENTAMENTE CADA QUESTÃO.
- ESCRIBE AS TUAS RESPOSTAS DE FORMA LEGÍVEL, SE A TUA LETRA NÃO FOR LEGÍVEL, ESCRIBE EM LETRAS MAIÚSCULAS.
- **DEVES EXPLICAR/ JUSTIFICAR TODAS AS TUAS RESPOSTAS.** A PROFESSORA TEM QUE PERCEBER COMO RESPONDESTES A CADA QUESTÃO.

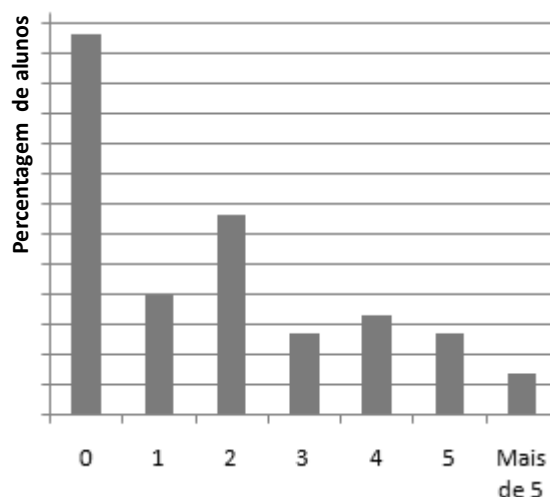
GRUPO 1 - Elaborou-se um estudo estatístico no qual se pretendia estudar o número de faltas dos alunos do 3.º ciclo da Escola, seguindo os seguintes passos:

- (A) Recolheu-se junto da Direção da escola as pautas de final do 3.º período de todas as turmas do 3.º ciclo tendo-se recolhido os dados relativos ao número de faltas nas várias das disciplinas, dos alunos cujo número de turma fosse múltiplo de 4 (n.º de turma 4, 8, 12, 16, 20, 24 e 28)
- (B) Organizaram-se os dados recolhidos quanto à variável **número de faltas** numa tabela e procedeu-se à sua representação gráfica.

Tabela 1

Nº de faltas	Nº de alunos	Percentagem de alunos
0	57	38
1	18	12
2	30	20
3	12	8
4		10
5	12	
Mais de 5		
Total	150	100%

Distribuição do número de faltas



- 1.1. Completa os dados em falta na tabela 1 e apresenta os cálculos ou raciocínios efetuados.
- 1.2. Sabendo que no eixo vertical do gráfico de barras (Figura 1) se indicou a percentagem de alunos, indica a escala utilizada. Justifica a tua resposta.
- 1.3. Qual é o número de faltas mais usual no conjunto dos alunos em análise?
- 1.4. Com base nos dados recolhidos a Direção da Escola está a pensar encomendar numa gráfica certificados de assiduidade* para 38% dos alunos do 3.º ciclo.
(*certificado atribuído aos alunos que não tiveram faltas ao longo do ano)

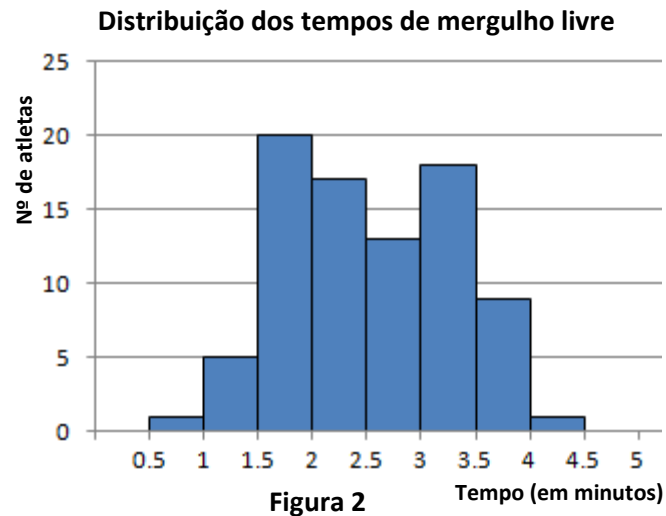
Comenta a decisão da Direção, considerando os seguintes aspetos:

- Qualidade da amostra
- Quantidade/dimensão da amostra.

GRUPO 2 - A professora de Matemática propôs a análise do histograma (Figura 2), no qual se registaram os tempos de mergulho livre de um conjunto de atletas numa prova de mergulho livre.



Mergulho livre - o mergulhador depende exclusivamente de sua capacidade pulmonar, preparação física e controle emocional.



2.1. Quando se organizaram os tempos de mergulho, optou-se por construir quantas classes? Com que amplitude?

2.2. Quantos atletas participaram na prova?

2.3. Quantos atletas realizaram um mergulho

2.3.1. com 3 ou mais minutos?

2.3.2. cuja duração foi inferior a 2 minutos?

2.4. Indica um tempo possível de mergulho para o atleta vencedor da prova.

GRUPO 3 – Poucas horas de sono podem comprometer o desenvolvimento infantil

Data: 03-09-2007

Crianças que dormem pouco podem vir a desenvolver problemas de crescimento, com limitações ao nível cognitivo e comportamental. O alerta surge das conclusões de um estudo recentemente realizado em Inglaterra, que teve como objeto de estudo a problemática do sono infantil.



"Estamos a ver surgir o sono *junk*, que não tem a duração e a qualidade que deveria para alimentar o cérebro com o descanso que precisa", alertam os investigadores ingleses do *Sleep Council*, realçando ainda a importância de reeducar para o sono de qualidade, cujo mínimo recomendado é de 8 horas, sob pena de um mau sono conduzir ao aumento da obesidade, dos problemas de tiroide e das depressões.

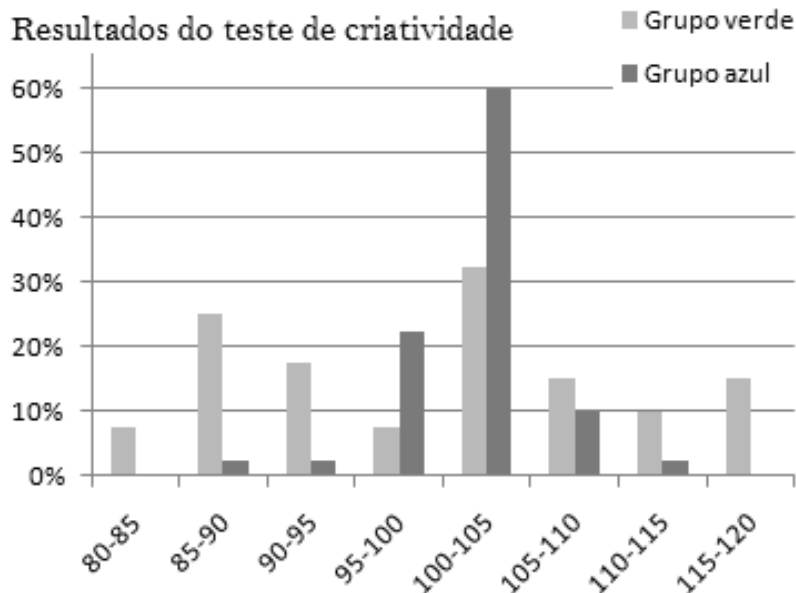
Adaptado de <http://www.portaldacrianca.com.pt/noticiasn.php?id=65>

Um jornal pretende realizar em Portugal um inquérito a jovens entre os 12 e os 16 anos, de modo a recolher alguma informação sobre os hábitos de sono.

3.1. Indica um procedimento que leve à escolha de uma amostra enviesada.

3.2. Formula três questões que possam ser utilizadas para estudar os hábitos de sono dos jovens portugueses.

GRUPO 4 - Um industrial de sucesso escreveu um livro sobre como fazer negócios. No livro afirmou "Se quer as pessoas mais criativas a trabalhar consigo, contrate pessoas Azuis". Baseou a sua conclusão num estudo que realizou com o objetivo de comparar o a Criatividade das pessoas Azuis e Verdes. No estudo utilizou-se uma amostra aleatória de 80 pessoas (40 de cada grupo). Cada pessoa realizou um teste de Criatividade. **A média do grupo Verde foi de 98 e a do grupo Azul foi 101.** Sendo os resultados das 80 pessoas apresentado no gráfico abaixo.



Grupos de ativistas em todo o país ficaram indignados, alegando que os resultados do industrial tinham sido uma coincidência. Cada grupo de ativistas realizou um estudo comparando os resultados de testes de criatividade em grupos de pessoas Azuis e Verdes recorrendo a amostras representativas da população.

Obtiveram resultados quase idênticos aos do empresário, concluindo que os dados apresentados no gráfico eram confiáveis. Mas a conclusão do industrial estava errada!

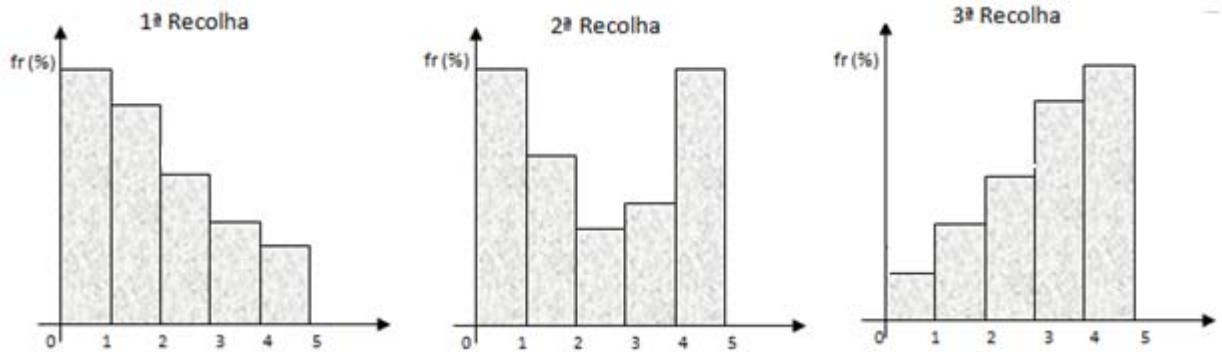
Com base no gráfico de resultados do estudo do industrial, apresenta pelo menos dois argumentos matemáticos que contrariem a conclusão a que chegou.

GRUPO 5 - Como parte de um projeto os alunos de uma turma pretendem responder à questão “As raparigas gastam menos dinheiro na escola do que os rapazes?” Para responder, pretendem recolher dados, quanto aos gastos semanais na escola (bar, papelaria e refeitório), junto de 160 dos 1560 alunos da escola.

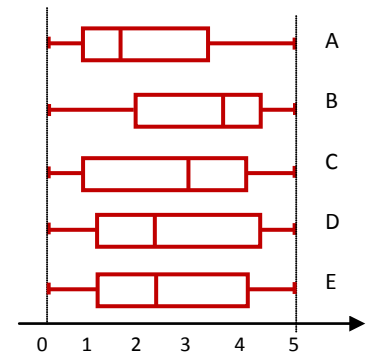
5.1. Indica a população em estudo?

5.2. Indica como proceder para escolher uma amostra representativa.

5.3. Optou-se por recolher dados junto de um conjunto representativo de 160 alunos da escola, em três momentos distintos do ano escolar. Os histogramas relativos à distribuição dos gastos semanais na escola, em cada recolha, foram os seguintes:



Indica para cada histograma a letra do correspondente diagrama de extremos e quartis. Justifica a tua resposta.



Objetivos – teste final		Conceitos e procedimentos	Resolução de problemas/ Raciocínio
Representações	<p>Avaliação da capacidade de construção</p> <p>Adequação aos dados e ao que se pretende transmitir</p> <p>Interpretação das representações (leitura nos vários níveis)</p> <p>Sentido crítico face a representações enganosas</p> <p>Transformação de uma representação numa outra representação (reconhecer características gerais, como a forma, o centro e a dispersão da distribuição)</p>	<p>2.1</p> <p>1.1,2.2,2.3,2.4, 6.3</p>	1.2, 5.3, 6.4
Medidas	<p>Avaliação da capacidade de determinação</p> <p>Adequação das medidas a utilizar face a diferentes condições, e de que modo representam ou não um conjunto de dados.</p> <p>Interpretação das medidas</p> <ul style="list-style-type: none"> reconhecer que medidas de ordem, e a sua respetiva localização fornecem informação sobre um conjunto de dados utilização das medidas estatísticas centrais e de dispersão são úteis para comparar conjuntos de dados. 	1.3, 6.4	<p>6.4, 6.5</p> <p>4</p>
Planeamento	<p>Identificação da população em estudo e amostra a que se recorreu</p> <p>Formulação de questões</p> <ul style="list-style-type: none"> Reconhecimento da variabilidade Reconhecer fase a uma situação a necessidade de estudar a população ou uma amostra <p>Processo de amostragem</p> <ul style="list-style-type: none"> Análise crítica de estudos estatísticos face ao uso de amostras não representativas Reconhecer amostras representativas e conhecer aspetos que poderão gerar enviesamentos na escolha de uma amostra Reconhecer a amplitude das extrapolações: <ul style="list-style-type: none"> - saber o que se pode inferir a partir de uma amostra, - saber de que modo a amostra está relacionada da população de qual foi extraída 	5.1/6.1/6.2	<p>3.2</p> <p>3.1, 5.2</p> <p>1.4, 6.5, 6.6</p>
	<p>Conhecimento do ciclo investigativo: (i) formular questões e hipóteses; (ii) recolher dados (iii) analisar dados, e (iv) comunicar e interpretar resultados.</p>		Grupo 1 e 6

Anexo 16 – Análise dos testes aplicados

		Teste inicial (N= 27)		Teste Final (N= 28)	
	Tipo de questão	Item	% de sucesso	Item	% de sucesso
Planeamento	Formulação de questões	-----	-----	3.2	43% (3 questões adequadas) 29% (2 questões adequadas)
	Amostra representativas	5.1	48%	5.2	82%
	Procedimento de envio-samento	-----	-----	3.1	71%
	Distinguir população e amostra	-----	-----	5.1 6.1	64%
Análise de dados	Representação tabelar	1.1	56%	1.1	86%
	Representação tabelar para gráfico de barras			1.2	61%
	Histograma	2.1	48%	2.1	74%
		2.2.	74%	2.2.	93%
		2.3.1	57%	2.3.1	76%
		2.3.2	74%	2.3.2	93%
		2.4	33%	2.4	50%
	Diagrama de caule e folha	4.1	67%	6.2	71%
		4.2	70%	6.3	82%
	Diagrama de Extremos e quartis	4.4	44% 19% erra a determinação do 1.º e 3.ºQ	6.4	71% 18% erra a determinação do 1.º e 3.ºQ
	Moda (gráfico de barras)	1.2	93%	1.3	96%
	Mediana (diagrama de caule e folhas)	4.3	63%	6.4	71%
	Transformação entre representações (histograma – diagrama de extremos e quartis)	5.2	26% 15% com enviesamento à esquerda	5.3	36% 11% falha uma associação
Comparar distribuições/ Tirar conclusões	3	48% dos quais 15% apresenta dois argumentos	4	61% dos quais 18% apresenta dois argumentos	
Tirar conclusões/ generalização à população	4.5	41% generaliza a todos os alunos do 8.º ano porque a amostra é expressiva face a essa população A generalização à população escolar assentou em própria opinião.	1.4 6.5 6.6	21% aceita a decisão considerando que está associada a algum grau de incerteza. 60% refere pelo menos dois aspetos relativos à concentração dos dados ou à sua dispersão ou aponta um aspeto relativo ao centro e outro relativo à dispersão. 46% considera válida a extrapolação pois para 18% a amostra é representativa (secção aleatória simples). 18% não considera válida a extrapolação pois a amostra tem pequena dimensão.	