



## **Perceção parental do peso e estilos de vida dos adolescentes**

Elsa Rute Santos Silva

**Mestrado em Bioestatística**

Dissertação orientada por:  
Prof.<sup>a</sup> Dra. Marília Antunes  
Prof.<sup>a</sup> Dra. Augusta Gama

# Agradecimentos

Às minhas orientadoras, Professora Marília Antunes e Professora Augusta Gama, pela atenção, disponibilidade e por todo o apoio prestado no decorrer deste estudo.

À minha família, pela preocupação demonstrada e pelas palavras de incentivo.

À minha família mais próxima, especialmente à minha Mãe, por ter estado sempre presente nos momentos em que mais precisava e por ter feito das tripas coração para que eu pudesse chegar até aqui e ser a pessoa que sou hoje. Agradeço-te, do fundo do coração, todo o apoio que sempre me deste e por nunca me deixares desistir.

Às meninas que fazem parte da minha vida há 20 anos, obrigada pela amizade e por me apoiarem sempre que preciso.

Ao meu menino, por trazer luz à minha vida.

# Resumo

Nas últimas décadas tem ocorrido um aumento drástico da prevalência do excesso de peso e obesidade em crianças e adolescentes. Para estas idades, a obesidade e o excesso de peso são reconhecidos como os fatores de risco de maior morbidade e mortalidade na vida adulta, sendo principalmente causados por uma alimentação desequilibrada e um estilo de vida sedentário.

A adolescência é considerada um dos períodos mais vulneráveis para o desenvolvimento de excesso de peso e obesidade, em que o género, idade e peso dos jovens, assim como a escolaridade dos pais, são apontados como alguns dos fatores que influenciam a perceção dos pais relativamente ao peso dos filhos. Deste modo, o objetivo desta dissertação foi analisar a perceção parental relativamente ao peso e estilos de vida dos adolescentes, nomeadamente os hábitos alimentares, a atividade física e o sedentarismo.

Para tal, foi usada uma base de dados com dados relativos a 664 pares de inquéritos respondidos por jovens do 7<sup>o</sup> ao 12<sup>o</sup> anos, residentes na Grande Lisboa, e pelos respetivos encarregados de educação.

Visando aspetos relativos ao peso e hábitos dos adolescentes, foram realizadas a harmonização e a quantificação da concordância entre as respostas dadas pelos jovens e pelos encarregados de educação enquanto medida indicadora da perceção parental do peso e estilos de vida dos adolescentes. Para saber que características dos jovens e do meio familiar podem influenciar significativamente a perceção parental em relação ao peso e aos hábitos dos adolescentes foram aplicados modelos de regressão.

Os resultados mostraram que 20,6% dos jovens sofre de excesso de peso e 3,3% de obesidade, sendo a perceção do peso dos jovens com excesso de peso (média = 0.9877) e a perceção do peso dos jovens com obesidade (média = 0.9269) mais baixas comparativamente à perceção do peso dos jovens com peso normal (média = 0.9990). A perceção global do peso e estilos de vida dos adolescentes difere entre os géneros dos jovens (valor-p = 0.039), sendo a melhor perceção referente às raparigas. Quanto às características que influenciam a perceção correta do peso dos jovens, o sexo masculino, o excesso de peso e a obesidade dos jovens são as características que contribuem para que a perceção seja mais baixa. O único fator relacionado com a perceção global foi a idade dos jovens, em que idades superiores contribuem para uma perceção mais correta.

Apesar dos encarregados de educação terem uma boa perceção em relação ao peso dos jovens, esta revela-se mais baixa nos jovens com excesso de peso e obesidade, contribuindo estas características para uma perceção incorreta do peso dos jovens. O peso e estilos de vida dos jovens, no seu conjunto, são melhor percecionados nas raparigas, sendo a idade dos jovens o único fator que influencia essa perceção.

**Palavras-chave:** Perceção parental do peso e estilos de vida dos adolescentes; Medidas de concordância; Modelos de regressão

# Abstract

In the last decades there has been a drastic increase in the prevalence of overweight and obesity in children and adolescents. For these ages, obesity and overweight are recognized as the risk factors of higher morbidity and mortality in adulthood, mainly caused by an unbalanced diet and a sedentary lifestyle.

Adolescence is considered one of the most vulnerable periods for the development of overweight and obesity, in which the gender, age and weight of the young, as well as the schooling of the parents, are pointed out as some of the factors that influence the parental perception of children's weight. Thus, the objective of this dissertation was to analyze the parental perception regarding the weight and lifestyles of adolescents, namely eating habits, physical activity and sedentary lifestyle.

To this end, a database with data of 664 pairs of surveys answered by young people from 7th to 12th years, living in Greater Lisbon, and their guardians was used.

Aiming at aspects related to the weight and habits of adolescents, the harmonization and quantification of the agreement between the responses given by young people and the guardians as an indicative measure of parental perception of weight and lifestyles of adolescents were performed. To know which characteristics of the young people and of the family environment can significantly influence the parental perception in relation to the adolescents' weight and habits, regression models were applied.

The results showed that 20,6% of the young people suffer from overweight and 3,3% of obesity, being the weight perception of young people with overweight (mean = 0.9877) and the perception of the weight of young people with obesity (mean = 0.9269) lower than the weight perception of young people with normal weight (mean = 0.9990). The overall perception of weight and lifestyles of adolescents differs between the sexes of young people (p-value = 0.039), being the best perception regarding girls. Regarding the characteristics that influence the correct perception of the weight of young people, the male gender, overweight and obesity of young people are the characteristics that contribute to the perception being lower. The only factor related to global perception was the age of young people, in which higher ages contribute to a more correct perception.

Although the guardians have a good perception regarding the weight of young people, this is lower in young people with overweight and obesity, contributing these characteristics to an incorrect perception of the weight of young people. The weight and lifestyles of young people, as a whole, are better perceived in girls, being the age of young people the only factor that influences this perception.

**Keywords:** Parental perception of weight and lifestyles of adolescents; Agreement measures; Regression models

# Índice

<b>Índice</b>	<b>iv</b>
<b>Lista de Tabelas</b>	<b>vii</b>
<b>Lista de Figuras</b>	<b>ix</b>
<b>1 Introdução</b>	<b>1</b>
1.1 Objetivos . . . . .	3
1.2 Estrutura da dissertação . . . . .	3
<b>2 Material</b>	<b>4</b>
2.1 Inquéritos . . . . .	4
2.1.1 Inquérito do Jovem . . . . .	4
2.1.2 Inquérito do Encarregado de Educação . . . . .	5
2.1.3 Amostra . . . . .	5
2.1.4 Variáveis em estudo . . . . .	5
<b>3 Metodologia</b>	<b>7</b>
3.1 Harmonização dos inquéritos . . . . .	7
3.1.1 Questões em estudo . . . . .	8
3.1.2 Processo de harmonização . . . . .	9
3.2 Medidas de Concordância . . . . .	11
3.2.1 Coeficiente de Gower . . . . .	11
3.3 Técnicas de Inferência Estatística Não Paramétrica . . . . .	14
3.3.1 Teste de <i>Mann-Whitney-Wilcoxon</i> . . . . .	14
3.3.2 Teste de <i>Kruskal-Wallis</i> . . . . .	15
3.3.3 Coeficiente de correlação de <i>Spearman</i> . . . . .	16
3.4 Modelos de Regressão . . . . .	17
3.4.1 Modelo de Regressão Logística Binária . . . . .	17
3.4.1.1 Formulação do modelo . . . . .	18
3.4.1.2 Estimação dos coeficientes de regressão . . . . .	18
3.4.1.3 Inferência sobre o modelo . . . . .	19
3.4.1.4 Seleção de Modelos . . . . .	20
3.4.1.5 Qualidade de ajustamento do modelo . . . . .	20
3.4.1.6 Capacidade preditiva do modelo . . . . .	24
3.4.1.7 Interpretação dos parâmetros do modelo . . . . .	26
3.4.2 Modelo de Regressão Logística Ordinal . . . . .	28

3.4.2.1	<i>Proportional Odds Model</i> . . . . .	29
3.4.2.2	Pressuposto de <i>odds</i> proporcionais . . . . .	30
3.4.2.3	Estratégias de construção de modelos . . . . .	31
3.4.2.4	Interpretação dos parâmetros do modelo . . . . .	32
3.4.3	Modelo de Regressão Beta . . . . .	33
3.4.3.1	Formulação do modelo . . . . .	33
3.4.3.2	Seleção de modelos . . . . .	34
3.4.3.3	Medidas de diagnóstico . . . . .	34
3.4.3.4	Interpretação dos parâmetros do modelo . . . . .	36
<b>4</b>	<b>Resultados</b> . . . . .	<b>37</b>
4.1	Análise exploratória . . . . .	37
4.2	Análise da concordância parental: aplicação de medidas de concordância . . . . .	40
4.2.1	Percepção do peso . . . . .	41
4.2.2	Percepção dos hábitos alimentares . . . . .	42
4.2.3	Percepção dos hábitos de atividade física . . . . .	43
4.2.4	Percepção dos hábitos sedentários . . . . .	45
4.2.5	Percepção global . . . . .	47
4.3	Análise da concordância parental: aplicação de modelos de regressão . . . . .	50
4.3.1	Percepção do peso . . . . .	51
4.3.1.1	Seleção do modelo . . . . .	52
4.3.1.2	Análise da qualidade de ajustamento do modelo . . . . .	53
4.3.1.3	Capacidade preditiva do modelo . . . . .	56
4.3.1.4	Interpretação do modelo . . . . .	56
4.3.2	Percepção dos hábitos alimentares . . . . .	57
4.3.2.1	Seleção do modelo . . . . .	58
4.3.2.2	Análise da qualidade de ajustamento do modelo . . . . .	61
4.3.2.3	Interpretação do modelo . . . . .	64
4.3.3	Percepção dos hábitos de atividade física . . . . .	64
4.3.3.1	Seleção do modelo . . . . .	65
4.3.3.2	Análise da qualidade de ajustamento . . . . .	66
4.3.3.3	Capacidade preditiva do modelo . . . . .	68
4.3.3.4	Interpretação do modelo . . . . .	69
4.3.4	Percepção dos hábitos sedentários . . . . .	69
4.3.4.1	Seleção do modelo . . . . .	69
4.3.4.2	Análise da qualidade de ajustamento do modelo . . . . .	70
4.3.5	Percepção global . . . . .	72
4.3.5.1	Seleção do modelo . . . . .	72
4.3.5.2	Análise da qualidade de ajustamento do modelo . . . . .	73
4.3.5.3	Interpretação do modelo . . . . .	75
<b>5</b>	<b>Discussão</b> . . . . .	<b>76</b>
5.1	Análise exploratória . . . . .	76
5.2	Análise da percepção parental: medidas de concordância . . . . .	76
5.3	Análise da percepção parental: aplicação de modelos de regressão . . . . .	78

<b>6 Conclusões</b>	<b>79</b>
<b>Bibliografia</b>	<b>81</b>

# Lista de Tabelas

2.1	Descrição das características sociodemográficas dos adolescentes . . . . .	5
2.2	Descrição das características sociodemográficas dos encarregados de educação . . . . .	6
3.1	Identificação das questões e respetivas respostas em relação ao peso dos adolescentes . . . . .	8
3.2	Identificação das questões e respetivas respostas em relação aos hábitos alimentares dos adolescentes . . . . .	8
3.3	Identificação das questões e respetivas respostas em relação aos hábitos de atividade física dos adolescentes . . . . .	8
3.4	Identificação das questões e respetivas respostas em relação aos hábitos sedentários dos adolescentes . . . . .	9
3.5	Harmonização das respostas em relação às questões sobre o peso dos adolescentes . . . . .	9
3.6	Harmonização das respostas em relação às questões sobre os hábitos alimentares dos adolescentes . . . . .	10
3.7	Harmonização das respostas em relação às questões sobre os hábitos sedentários dos adolescentes . . . . .	10
3.8	Harmonização e medidas de concordância das questões sobre o peso e estilos de vida dos adolescentes . . . . .	13
3.9	Tabela de classificação . . . . .	25
4.1	Análise exploratória das características sociodemográficas dos adolescentes . . . . .	38
4.2	Análise exploratória das características sociodemográficas dos encarregados de educação . . . . .	39
4.3	Análise cruzada da Classe de peso dos encarregados de educação vs. Classe de peso dos jovens . . . . .	39
4.4	Análise cruzada da Classe de peso dos encarregados de educação vs. Preocupação dos encarregados de educação quanto ao peso dos jovens . . . . .	40
4.5	Análise cruzada da Classe de peso dos jovens vs. Preocupação dos encarregados de educação quanto ao peso dos jovens . . . . .	40
4.6	Análise descritiva da concordância parental em relação ao peso e estilos de vida dos adolescentes . . . . .	40
4.7	Concordância do peso dos jovens em função das variáveis sociodemográficas dos jovens . . . . .	41
4.8	Concordância do peso dos jovens em função das variáveis sociodemográficas dos encarregados de educação . . . . .	42
4.9	Concordância dos hábitos alimentares dos jovens em função das características sociodemográficas dos jovens e encarregados de educação . . . . .	43
4.10	Valores iniciais e categorização dos valores da concordância do peso . . . . .	51
4.11	Modelo de regressão logística binária da perceção do peso . . . . .	52

4.12	Modelo de regressão logística binária da percepção do peso com as variáveis Sexo e Classe de peso dos jovens . . . . .	53
4.13	Modelo de regressão logística binária da percepção do peso sem as observações 32, 199, 350, 479 e 535 . . . . .	55
4.14	Tabela de classificação para o modelo da percepção do peso . . . . .	56
4.15	Interpretação das variáveis significativas para o modelo da percepção do peso dos jovens .	57
4.16	Categorização dos valores da concordância dos hábitos alimentares dos jovens . . . . .	57
4.17	Modelo de regressão logística ordinal da percepção dos hábitos alimentares . . . . .	58
4.18	Modelo de regressão logística ordinal da percepção dos hábitos alimentares sem a variável Idade dos encarregados de educação . . . . .	61
4.19	Interpretação das variáveis significativas para o modelo da percepção dos hábitos alimentares	64
4.20	Categorização da concordância dos hábitos de atividade física dos jovens . . . . .	65
4.21	Modelo de regressão logística binária da percepção dos hábitos de atividade física . . . .	65
4.22	Modelo de regressão logística binária da percepção dos hábitos de atividade física sem as observações 28, 70, 152, 189, 351 e 404 . . . . .	68
4.23	Tabela de classificação para o modelo da percepção dos hábitos de atividade física . . . .	68
4.24	Interpretação da variável significativa para o modelo da percepção dos hábitos de atividade física dos jovens . . . . .	69
4.25	Modelo de regressão beta da percepção dos hábitos sedentários . . . . .	70
4.26	Modelo de regressão beta da percepção global . . . . .	73
4.27	Interpretação da variável significativa para o modelo da percepção global dos jovens . . .	75

# Lista de Figuras

4.1	Concordância dos hábitos de atividade física dos jovens em função do Sexo dos jovens . . . . .	44
4.2	Concordância dos hábitos de atividade física dos jovens em função da Idade dos jovens . . . . .	44
4.3	Concordância dos hábitos de atividade física dos jovens em função do Nível de atividade física dos jovens . . . . .	44
4.4	Concordância dos hábitos de atividade física dos jovens em função dos Encarregados de educação . . . . .	44
4.5	Concordância dos hábitos de atividade física dos jovens em função da Idade dos encarregados de educação . . . . .	45
4.6	Concordância dos hábitos sedentários dos jovens em função do Sexo dos jovens . . . . .	46
4.7	Concordância dos hábitos sedentários dos jovens em função da Idade dos jovens . . . . .	46
4.8	Concordância dos hábitos sedentários dos jovens em função dos Encarregados de educação . . . . .	46
4.9	Concordância dos hábitos sedentários dos jovens em função da Idade dos encarregados de educação . . . . .	47
4.10	Concordância global em função do Sexo dos jovens . . . . .	47
4.11	Concordância global em função da Idade dos jovens . . . . .	47
4.12	Concordância global em função dos Encarregados de educação . . . . .	48
4.13	Concordância global em função da Idade dos encarregados de educação . . . . .	48
4.14	Faces de Chernoff relativamente ao Sexo dos jovens . . . . .	49
4.15	Faces de Chernoff relativamente aos Encarregados de educação . . . . .	50
4.16	Faces de Chernoff relativamente à Classe de peso dos jovens . . . . .	50
4.17	Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância do peso na escala <i>logit</i> . . . . .	53
4.18	Análise do preditor linear e identificação de observações <i>outliers</i> do modelo da perceção do peso . . . . .	54
4.19	Identificação de observações influentes do modelo da perceção do peso . . . . .	55
4.20	Curva ROC do modelo da perceção do peso . . . . .	56
4.21	Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância dos hábitos alimentares na escala <i>logit</i> . . . . .	59
4.22	Representação gráfica da linearidade entre a variável Idade dos EE e a concordância dos hábitos alimentares na escala <i>logit</i> . . . . .	60
4.23	Gráfico dos resíduos padronizados contra a ordem das observações do modelo da perceção dos hábitos alimentares . . . . .	62
4.24	Gráfico das probabilidades obtidas por validação cruzada do modelo da perceção dos hábitos alimentares . . . . .	62
4.25	Gráfico de <i>leverage</i> do modelo da perceção dos hábitos alimentares . . . . .	63

4.26	Gráfico da estatística D do modelo da percepção dos hábitos alimentares . . . . .	63
4.27	Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância dos hábitos de atividade física na escala <i>logit</i> . . . . .	66
4.28	Análise do preditor linear e identificação de observações <i>outliers</i> do modelo da percepção dos hábitos de atividade física . . . . .	67
4.29	Identificação de observações influentes do modelo da percepção dos hábitos de atividade física . . . . .	67
4.30	Curva ROC do modelo da percepção dos hábitos de atividade física . . . . .	69
4.31	Análise do preditor linear e identificação de observações <i>outliers</i> do modelo da percepção dos hábitos sedentários . . . . .	71
4.32	Gráfico <i>Half-normal</i> dos resíduos da <i>deviance</i> do modelo da percepção dos hábitos sedentários . . . . .	71
4.33	Identificação de observações influentes do modelo da percepção dos hábitos sedentários . . . . .	72
4.34	Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância global na escala <i>logit</i> . . . . .	73
4.35	Análise do preditor linear e identificação de observações <i>outliers</i> do modelo da percepção global . . . . .	74
4.36	Gráfico <i>Half-normal</i> dos resíduos da <i>deviance</i> do modelo da percepção global . . . . .	74
4.37	Identificação de observações influentes do modelo da percepção global . . . . .	75

# Capítulo 1

## Introdução

Nos últimos anos, a obesidade tornou-se um dos principais determinantes da morbidade e mortalidade em muitas regiões do mundo. Segundo a Organização Mundial de Saúde (OMS), a obesidade é definida como uma acumulação anormal ou excessiva de gordura corporal com resultados adversos para a saúde. A prevalência do excesso de peso e obesidade, em crianças e adolescentes, tem vindo a aumentar drasticamente nas últimas décadas. Dados da OMS [1] apontam para que a prevalência tenha aumentado de 4% em 1975 para 18% em 2016, existindo, no mundo, mais de 340 milhões de crianças e jovens (5-19 anos) com excesso de peso ou obesidade. Este aumento ocorreu de forma semelhante entre rapazes e raparigas, i.e., em 2016, 18% das raparigas e 19% dos rapazes tinham excesso de peso. Relativamente aos adolescentes que sofriam de obesidade, 6% eram raparigas e 8% eram rapazes. Nos últimos 10 anos, a prevalência da obesidade entre crianças europeias aumentou de 5% para 10%. Em Portugal, 30% das crianças entre os 7 e os 11 anos são consideradas obesas ou com excesso de peso [2]. De acordo com Sardinha *et al.* [3], 17,0% e 4,6% das raparigas e 17,7% e 5,8% dos rapazes, entre os 10 e os 18 anos, sofrem de excesso de peso e obesidade, respetivamente.

Na idade pediátrica, a obesidade e o excesso de peso são reconhecidos como os fatores de risco de maior morbidade e mortalidade na vida adulta, estando associados ao aparecimento de doenças como: diabetes mellitus tipo II, hipercolesterolemia, distúrbios ortopédicos, hipertensão e outras doenças cardiovasculares, e ao aparecimento de problemas no âmbito psicossocial como a discriminação social e a baixa autoestima [1, 2, 4].

A adolescência é considerada um dos períodos mais vulneráveis para o desenvolvimento de excesso de peso e obesidade [5], devido às diversas mudanças fisiológicas, físicas e psicológicas que ocorrem no indivíduo, juntamente com a constante procura por algo e influência de colegas, escola e sociedade em geral e que afetam os hábitos e as necessidades nutricionais dos jovens [6]. Hábitos como o consumo elevado de alimentos altamente energéticos e ricos em gordura, consumo insuficiente de frutas e vegetais, "saltar" o pequeno-almoço, diminuição da prática de atividade física e elevado tempo despendido a ver televisão ou a jogar no computador, são hábitos pouco saudáveis que promovem o excesso de peso e a obesidade na adolescência [1, 2, 5, 7, 8, 9, 10, 11].

Evidências indicam que os hábitos adquiridos durante a infância e adolescência permanecem durante a fase adulta e que são influenciados significativamente pelo ambiente familiar [8, 12]. Deste modo, para que na adolescência haja obtenção e consolidação de hábitos saudáveis, torna-se importante que o ambiente familiar seja adequado. A disponibilização de alimentos como frutas e vegetais, propicia um aumento de consumo desses alimentos por parte dos jovens, assim como, apoiar e providenciar meios para a prática de atividades desportivas, estimula os jovens a se tornarem pessoas mais ativas [13]. Apesar de os jovens, durante a fase da adolescência, experienciarem maior autonomia e independência

quanto às suas escolhas alimentares e desportivas, continuam a ser bastante influenciados não só pelo ambiente familiar, como referido anteriormente, mas também pelos hábitos que eles percebem de seus progenitores [12, 13].

Sabendo que determinados hábitos dos jovens que proporcionam um aumento de peso prejudicial para a saúde podem ser modificados, uma das formas de prevenir o excesso de peso e a obesidade na adolescência é a alteração desses hábitos. Como os progenitores têm uma grande influência sobre os filhos é possível afirmar que, se eles forem capazes de reconhecer os hábitos menos saudáveis dos jovens, poderão intervir de modo a modificá-los. No entanto, só o conseguirão fazer se reconhecerem a necessidade de tais mudanças ou perceberem o risco que os seus filhos correm ao ter hábitos pouco saudáveis [14]. Assim, se os progenitores forem capazes de perceber corretamente o peso dos filhos, mais propensos estarão em tentar perceber que hábitos eles têm e que influenciam o ganho de peso e dispostos a ajudá-los na obtenção de hábitos mais saudáveis.

Para perceber a capacidade dos pais em reconhecer o peso dos filhos, têm sido feitos alguns estudos nesse sentido. Muitos deles revelam uma considerável percentagem de pais que não identifica corretamente o peso dos filhos, sobretudo quando estes têm excesso de peso ou obesidade, e fatores como o nível de escolaridade dos pais, género, peso e idade dos jovens como fatores que influenciam a capacidade dos pais em reconhecer o peso dos adolescentes [14, 15, 16, 19]. Os hábitos alimentares, práticas desportivas e comportamentos sedentários dos jovens também têm sido alvos de estudos [17, 18]. Estes estudos têm sido feitos com o objetivo de conhecer e caracterizar os hábitos dos jovens, mas não de perceber se os pais percebem ou não corretamente o tipo de hábitos que eles têm.

Estudar o reconhecimento dos pais em relação ao peso dos filhos é uma parte essencial de muitos programas de intervenção e prevenção da obesidade, dado que permite um melhor conhecimento sobre a capacidade dos pais em reconhecer o peso dos filhos. Nos casos em que os pais classificam erroneamente o peso dos filhos, estes programas irão ajudá-los a tomar conhecimento do que é um peso saudável e as consequências físicas e mentais que o excesso de peso pode trazer para a vida dos filhos. Juntamente com o reconhecimento do peso vem o reconhecimento dos hábitos dos filhos visto que, nos casos em que estes são pouco saudáveis ao contribuírem para um ganho de peso prejudicial para a saúde, pode ajudar a motivar e educar famílias no sentido de serem capazes de proporcionar um ambiente familiar saudável e promover mudanças saudáveis nos hábitos dos jovens, de modo a que estes tenham uma vida longa e saudável [14, 19].

## **1.1 Objetivos**

Com o propósito de analisar a percepção parental no reconhecimento do peso e estilos de vida dos adolescentes, a presente dissertação foi desenvolvida com os seguintes objetivos específicos:

1. Conhecer a percepção parental em relação ao peso e estilos de vida dos adolescentes
  - (a) Análise da percepção parental relativamente ao peso e hábitos dos jovens, tais como: hábitos alimentares, sedentários e de atividade física;
  - (b) Análise da percepção parental relativamente ao peso e hábitos dos jovens em função das características sociodemográficas dos jovens e dos encarregados de educação.
2. Conhecer o papel das características sociodemográficas dos jovens e dos encarregados de educação no grau de percepção parental relativamente ao peso e estilos de vida dos adolescentes

## **1.2 Estrutura da dissertação**

Com o intuito de se atingirem os objetivos mencionados, a presente dissertação encontra-se organizada em mais cinco capítulos. No Capítulo 2 é abordado o material usado para dar resposta aos objetivos, sendo explicado o procedimento para a sua recolha e o tipo de informação que disponibiliza. No Capítulo 3 são abordadas as diversas metodologias aplicadas: harmonização, medidas de concordância, técnicas não paramétricas e modelos de regressão. Os resultados obtidos para o conhecimento da percepção parental e do papel de determinadas características na percepção são apresentados no Capítulo 4, sendo apresentada a discussão desses resultados no Capítulo 5. Por fim, no Capítulo 6, são apresentadas algumas conclusões gerais sobre o trabalho realizado.

## Capítulo 2

# Material

No presente trabalho, os dados utilizados foram extraídos de uma base de dados mais abrangente construída a partir de um estudo sobre "Hábitos Alimentares e Estilos de Vida do Adolescente", desenvolvido no ano 2012, em escolas do 2º e 3º Ciclos do Ensino Básico (CEB) e Secundário, do concelho de Sintra, com o objetivo de determinar a prevalência de excesso de peso e obesidade em adolescentes e de conhecer a influência de fatores dos estilos de vida e do ambiente familiar no peso. Os dados foram recolhidos no âmbito do trabalho de mestrado das alunas Graça Sobral e Luísa Gordo do Mestrado de Biologia Humana e Ambiente (MBHA), entre fevereiro e maio de 2012 [20].

A recolha de dados consistiu na avaliação antropométrica dos adolescentes e na realização de dois inquéritos, um aos adolescentes e o outro aos seus encarregados de educação. A avaliação antropométrica foi realizada quanto ao peso e estatura dos jovens, de acordo com a metodologia do *Centers for Disease Control and Prevention* (CDC, 2000) e os inquéritos foram construídos a partir de questionários internacionais e nacionais validados como: *Youth Risk Behavior Survey* (YBRS), *International Physical Activity Questionnaire* (IPAQ), Aventura Social/KIDSCREEN e Estudo Nacional de Prevalência de Obesidade Infantil em Portugal, alterações de 2002 a 2009. No entanto, os inquéritos aplicados não foram validados para a população portuguesa.

O protocolo de estudo foi autorizado pela Direção Geral de Inovação e de Desenvolvimento Curricular, Monitorização de Inquéritos em Meio Escolar e pela Direção de cada uma das escolas. A recolha dos dados foi realizada de forma anónima, tendo sido apenas incluídos no estudo os jovens que aceitaram participar e dos quais se obteve consentimento informado dos respetivos encarregados de educação, i.e., a autorização para participarem no estudo.

### 2.1 Inquéritos

Os inquéritos foram respondidos por jovens inscritos no 3º CEB e Secundário (7º a 12º anos) e pelos respetivos encarregados de educação, podendo estes ser um dos progenitores ou outro adulto. O preenchimento do inquérito relativo aos jovens foi realizado no mesmo dia que a avaliação antropométrica, tendo sido enviado para casa o inquérito referente ao encarregado de educação.

#### 2.1.1 Inquérito do Jovem

O inquérito respondido pelos jovens abordava os seguintes tópicos:

- informação de natureza sociodemográfica do jovem (idade, género e nível de escolaridade) e da respetiva família;

- questões referentes aos estilos de vida, atividade física/sedentarismo, hábitos alimentares e hábitos de sono;
- questões sobre a autopercepção do peso, saúde e apetite.

### 2.1.2 Inquérito do Encarregado de Educação

O inquérito aplicado aos encarregados de educação abordava os seguintes tópicos:

- informação sociodemográfica do encarregado de educação (idade, nível de escolaridade e parentesco com o jovem);
- informação antropométrica autorreportada (peso e altura);
- questões sobre os seus hábitos de atividade física e hábitos sedentários;
- questões sobre os hábitos sedentários, de atividade física e de sono dos jovens;
- questões sobre a percepção e preocupação que o encarregado de educação tem relativamente ao peso, saúde e alimentação do seu educando.

### 2.1.3 Amostra

Para o presente estudo foram analisados 664 pares de inquéritos, a partir dos quais se consideraram apenas as questões relativas ao peso e estilos de vida dos jovens.

### 2.1.4 Variáveis em estudo

Com vista a se atingirem os objetivos propostos, são descritas nas Tabelas 2.1 e 2.2 as variáveis em estudo relativas aos jovens e aos encarregados de educação, respetivamente. As variáveis em estudo correspondem às características sociodemográficas dos indivíduos.

Tabela 2.1: Descrição das características sociodemográficas dos adolescentes

Variável	Classificação
Sexo	Qualitativa nominal 1. Masculino 2. Feminino
Idade (em anos)	Quantitativa
Idade decimal*	Quantitativa
Idade exata	Quantitativa
Ano de escolaridade	Qualitativa ordinal
Classe de peso	Qualitativa ordinal 1. Magro 2. Peso normal 3. Excesso de peso 4. Obeso
Nível de atividade física	Qualitativa ordinal 1. Baixo 2. Moderado 3. Elevado

\*Idade a meio ano, i.e., 12, 12.5, 13, ..., 18

Tabela 2.2: Descrição das características sociodemográficas dos encarregados de educação

Variável	Classificação
Encarregado de educação	Qualitativa nominal 1. Pai 2. Mãe 3. Outro, quem?
Idade (em anos)	Quantitativa
Escolaridade	Qualitativa ordinal
Classe de peso	Qualitativa ordinal 1. Magro 2. Peso normal 3. Excesso de peso 4. Obeso
Preocupação com o peso do educando	Qualitativa ordinal 1. Não está preocupado 2. Relativamente preocupado 3. Preocupado

As variáveis “Classe de peso” e “Nível de atividade física” dos jovens (Tabela 2.1) e a variável “Classe de peso” dos encarregados de educação (Tabela 2.2) são variáveis cujas respostas apresentadas não foram obtidas diretamente da base de dados, tendo sido determinadas por meio de outras informações disponíveis na base de dados.

No caso da variável “Classe de peso” dos jovens, as classes de peso foram definidas de acordo com o critério IOTF (*International Obesity Task Force*) de Cole *et al.* (2000 e 2007). Este critério tem como base a idade decimal (i.e., 12, 12.5, 13, ..., 18), o sexo e o índice de massa corporal (IMC) dos jovens. Para o cálculo do IMC ( $IMC = \text{Peso}(Kg) / \text{Altura}^2(m)$ ), foram usados os valores do peso e altura dos jovens resultantes da avaliação antropométrica realizada.

Relativamente ao “Nível de atividade física” dos jovens, este foi determinado a partir da informação relativa aos tipos de atividade física praticados pelos jovens e pelo respetivo tempo despendido, seguindo as diretivas de processamento de dados e análise do IPAQ em formato curto [23].

Para a variável “Classe de peso” dos encarregados de educação, as classes de peso foram definidas segundo o IMC dos encarregados de educação, calculado a partir do peso e altura autorreportados pelos mesmos.

## Capítulo 3

# Metodologia

Na presente dissertação, com o objetivo de analisar a percepção parental no reconhecimento do peso e estilos de vida dos adolescentes é necessário, em primeiro lugar, harmonizar os inquéritos dos jovens e dos encarregados de educação. Após o processo de harmonização, são aplicados três métodos estatísticos. O primeiro método a aplicar são as Medidas de Concordância, que permitem obter o grau de concordância entre as respostas dos jovens e dos encarregados de educação a questões sobre o peso e hábitos dos adolescentes e, assim, conhecer a percepção dos encarregados de educação em relação ao peso e estilos de vida dos jovens. De modo a saber como a concordância/percepção varia em função de determinadas características sociodemográficas dos jovens e dos encarregados de educação, são aplicadas Técnicas de Inferência Estatística Não Paramétrica. Por fim, para conhecer a influência dessas características no grau de percepção parental em relação ao peso e estilos de vida dos adolescentes, são aplicados Modelos de Regressão.

### 3.1 Harmonização dos inquéritos

De modo a conhecer a percepção parental em relação ao peso e estilos de vida dos adolescentes é necessário saber quais as respostas dadas pelos jovens e encarregados de educação (EE) às questões sobre o peso e os hábitos. Com base nessas respostas é possível determinar a concordância (percepção) que existe entre elas. No entanto, antes de se poder quantificar a concordância, torna-se necessário seguir o seguinte processo:

1. Identificar as questões que são semelhantes (i.e., dizem respeito ao mesmo assunto) nos inquéritos dos jovens e dos respetivos EE;
2. Para cada par de questões semelhantes, identificar o tipo de respostas possíveis (escala e número);
3. Caso o tipo de respostas possíveis seja diferente, harmonizar as respostas e agrupá-las do mesmo modo nos dois inquéritos;
4. Finalizado o processo de harmonização, identificar a natureza das respostas de cada par de questões;
5. Dependendo da natureza das respostas, aplicar a medida de concordância adequada.

### 3.1.1 Questões em estudo

As questões que foram identificadas como sendo semelhantes e as respectivas respostas estão apresentadas nas Tabelas 3.1, 3.2, 3.3 e 3.4. A Tabela 3.1 corresponde às questões e respostas em relação ao peso dos jovens, a Tabela 3.2 aos hábitos alimentares, a Tabela 3.3 aos hábitos de atividade física e a Tabela 3.4 aos hábitos sedentários.

Tabela 3.1: Identificação das questões e respectivas respostas em relação ao peso dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Resposta	Questão	Resposta
Como descreveria o peso atual do seu educando?	<ol style="list-style-type: none"> <li>1. Magro</li> <li>2. Normal</li> <li>3. Tem algum peso a mais</li> <li>4. Tem muito peso em excesso</li> </ol>	Avaliação antropométrica do peso (Kg) e estatura (m)	Resposta escrita

Tabela 3.2: Identificação das questões e respectivas respostas em relação aos hábitos alimentares dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Resposta	Questão	Resposta
O seu educando toma todos os dias o pequeno-almoço antes de sair de casa?	<ol style="list-style-type: none"> <li>1. Sim</li> <li>2. Às vezes</li> <li>3. Não</li> <li>4. Não sei</li> </ol>	Assinala para cada dia da semana o local onde tomaste o pequeno-almoço.	Resposta escrita, com as seguintes opções: escola, restaurante, café, casa, casa de outros ou não (se não fez a refeição)
Em sua casa há sempre fruta e produtos hortícolas?	<ol style="list-style-type: none"> <li>1. Sim</li> <li>2. Às vezes</li> <li>3. Não</li> <li>4. Não sei</li> </ol>	Em casa tens sempre fruta e/ou vegetais?	<ol style="list-style-type: none"> <li>1. Sim</li> <li>2. Às vezes</li> <li>3. Não</li> </ol>

Tabela 3.3: Identificação das questões e respectivas respostas em relação aos hábitos de atividade física dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Resposta	Questão	Resposta
Como motiva o seu educando para que faça atividade física/desportiva?	<ol style="list-style-type: none"> <li>1. Participando com ele em atividades desportivas</li> <li>2. Comprando material para desporto</li> <li>3. Indo com ele para espaços ao ar livre</li> <li>4. Dizendo-lhe que deve fazer atividade física, desporto</li> <li>5. Assistindo às suas atividades desportivas</li> <li>6. Levando-o aos locais onde faz desporto</li> <li>7. Falando na importância do exercício físico na saúde</li> <li>8. Não digo nada sobre se deve ou não fazer atividade física</li> </ol>	De que forma os teus pais influenciam a tua prática de atividade física/desportiva?	<ol style="list-style-type: none"> <li>1. Participam comigo</li> <li>2. Compram material para desporto</li> <li>3. Vamos para espaços: parques naturais/campo/.../praia</li> <li>4. Dizem que devo fazer atividades físicas</li> <li>5. Vão assistir às minhas atividades desportivas</li> <li>6. Levam-me aos locais onde faço desporto</li> <li>7. Falam na importância do exercício físico na saúde</li> <li>8. Não dizem nada sobre se devo ou não fazer atividades físicas</li> </ol>

Tabela 3.4: Identificação das questões e respetivas respostas em relação aos hábitos sedentários dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Resposta	Questão	Resposta
Quantas horas, em média, por dia gasta a ver televisão?	Resposta escrita <sup>1</sup>	Quantas horas, em média, por dia gastas a ver televisão?	Resposta escrita <sup>1</sup>
Quantas horas, em média, por dia gasta a ver no computador ou jogos eletrónicos?	Resposta escrita <sup>1</sup>	Quantas horas, em média, por dia, gastas no computador/internet sem ser a fazer trabalhos da escola?	Resposta escrita <sup>1</sup>
		Quantas horas, em média, por dia, gastas a jogar jogos eletrónicos?	Resposta escrita <sup>1</sup>
Geralmente costuma ver TV enquanto toma as refeições?	1. Raramente 2. Às vezes 3. Quase sempre	Costumas ver TV enquanto tomas as refeições?	1. Sim 2. Não

### 3.1.2 Processo de harmonização

As questões que são identificadas como semelhantes podem ter diferentes respostas. A diferença pode residir na escala das respostas e/ou no número de respostas que as questões apresentam. Quando isto acontece, torna-se necessário harmonizar as respostas de ambas as questões, i.e., encontrar um conjunto de respostas que seja adequado a ambas as questões.

No presente estudo, o processo de harmonização foi realizado de forma arbitrária, estando apresentados nas Tabelas 3.5, 3.6 e 3.7 os processos aplicados às questões sobre o peso, hábitos alimentares e hábitos sedentários dos adolescentes, respetivamente.

Tabela 3.5: Harmonização das respostas em relação às questões sobre o peso dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Harmonização	Questão	Harmonização
Como descreveria o peso atual do seu educando.	Não é necessário aplicar o processo de harmonização	Avaliação antropométrica do peso (Kg) e estatura (m)	Classe de peso (IOTF)
1. Magro 2. Normal 3. Tem algum peso a mais 4. Tem muito peso em excesso		Resposta escrita	1. Magro 2. Peso normal 3. Excesso de peso 4. Obeso

Na Tabela 3.5, é possível observar que a questão sobre o peso dos adolescentes no inquérito dos encarregados de educação é do tipo qualitativo ordinal, ao contrário da questão correspondente no inquérito dos jovens. Visto que foram realizadas avaliações antropométricas dos jovens em relação ao seu peso e altura, foi usada esta informação para o cálculo do IMC e, conseqüentemente, da determinação das classes de peso de acordo com o critério IOTF, como referido na Secção 2.2.

Na Tabela 3.6, no inquérito dos encarregados de educação, a resposta “Não sei” escolhida pelos indivíduos foi retirada de ambas as questões e passou a ser tratada como resposta omissa.

No inquérito dos adolescentes, o processo de harmonização foi apenas aplicado à questão sobre o pequeno-almoço. Nessa questão, os jovens tinham de escrever, para cada dia da semana, o local onde tomavam o pequeno-almoço ou se o não tomavam. Na construção do conjunto de respostas para a nova

<sup>1</sup>Resposta para três períodos da semana: "2ª a 6ª feiras", "sábado" e "domingo"

Tabela 3.6: Harmonização das respostas em relação às questões sobre os hábitos alimentares dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Harmonização	Questão	Harmonização
O seu educando toma todos os dias o pequeno-almoço antes de sair de casa?  1. Sim 2. Às vezes 3. Não 4. Não sei	O seu educando toma todos os dias o pequeno-almoço antes de sair de casa?  1. Sim 2. Às vezes 3. Não	Assinala para cada dia da semana o local onde tomaste o pequeno-almoço  Resposta escrita, com as seguintes opções: escola, restaurante, café, casa, casa de outros ou não (se não fez a refeição)	Tomas todos os dias o pequeno-almoço antes de sair de casa?  1. Sim 2. Às vezes 3. Não
Em sua casa há sempre fruta e produtos hortícolas?  1. Sim 2. Às vezes 3. Não 4. Não sei	Em sua casa há sempre fruta e produtos hortícolas?  1. Sim 2. Às vezes 3. Não	Em casa tens sempre fruta e/ou vegetais?  1. Sim 2. Às vezes 3. Não	Não é necessário aplicar o processo de harmonização

Tabela 3.7: Harmonização das respostas em relação às questões sobre os hábitos sedentários dos adolescentes

Encarregado de Educação		Adolescente	
Questão	Harmonização	Questão	Harmonização
Quantas horas, em média, por dia gasta a ver televisão?  Resposta escrita <sup>1</sup>	Não é necessário aplicar o processo de harmonização	Quantas horas, em média, por dia gastas a ver televisão?  Resposta escrita <sup>1</sup>	Não é necessário aplicar o processo de harmonização
Quantas horas, em média, por dia gasta a ver no computador ou jogos eletrónicos?  Resposta escrita <sup>1</sup>	Não é necessário aplicar o processo de harmonização	Quantas horas, em média, por dia, gastas no computador/internet sem ser a fazer trabalhos da escola  Resposta escrita <sup>1</sup>  Quantas horas, em média, por dia, gastas a jogar jogos eletrónicos?  Resposta escrita <sup>1</sup>	Quantas horas, em média, por dia, gastas no computador ou em jogos eletrónicos?  Resposta escrita <sup>1</sup>
Geralmente costuma ver TV enquanto toma as refeições?  1. Raramente 2. Às vezes 3. Quase sempre	Geralmente costuma ver TV enquanto toma as refeições  1. Sim 2. Não	Costumas ver TV enquanto tomas as refeições?  1. Sim 2. Não	Não é necessário aplicar o processo de harmonização

questão “Tomas todos os dias o pequeno-almoço antes de sair de casa?”, consideraram-se a resposta “Casa” e os dias da semana “2<sup>a</sup>. a 6<sup>a</sup>. feira” da questão original. As respostas escolhidas para o conjunto foram “Sim”, “Às vezes” e “Não”. A resposta “Sim” corresponde aos jovens que escreveram “Casa” em quatro ou cinco dias da semana, a resposta “Às vezes” aos jovens que escreveram “Casa” em dois ou três dias da semana e a resposta “Não” aos jovens que não escreveram “Casa” ou apenas a escreveram em um dia da semana.

Em relação ao tempo despendido pelos jovens ao computador ou em jogos eletrônicos (Tabela 3.7), existem duas questões no inquérito dos adolescentes e apenas uma no inquérito dos encarregados de educação.

As duas questões do inquérito dos jovens “Quantas horas, em média, por dia, gastas no computador/internet sem ser a fazer trabalhos da escola?” e “Quantas horas, em média, por dia, gastas a jogar jogos eletrônicos?” foram transformadas numa só questão definida como “Quantas horas, em média, por dia, gastas no computador ou em jogos eletrônicos?”. A resposta a esta nova questão corresponde à média do número de horas registado para as duas questões atrás referidas e é dada para cada um dos três períodos da semana (2<sup>a</sup> a 6<sup>a</sup> feiras, sábado e domingo).

Em relação à questão sobre se é costume os jovens verem TV às refeições (Tabela 3.7), no inquérito dos encarregados de educação esta questão é do tipo qualitativo ordinal e no inquérito dos jovens é do tipo qualitativo nominal. A questão que sofreu alterações foi a do inquérito dos EE, em que as respostas passaram a ser iguais às respostas da questão dos jovens, i.e., “Sim” e “Não”. A resposta “Sim” corresponde à resposta “Quase sempre” e a resposta “Não” às respostas “Raramente” e “Às vezes”.

As respostas à questão sobre os hábitos de atividade física dos jovens, em ambos os inquéritos, não sofreram qualquer tipo de alteração, pelo que não foi necessário apresentá-las nesta Secção.

## 3.2 Medidas de Concordância

Após realizado o processo de harmonização de ambos os inquéritos, são escolhidas as medidas de concordância adequadas para quantificar a concordância entre as respostas dadas pelos jovens e encarregados de educação. As medidas de concordância a aplicar dependem da natureza das respostas, podendo as respostas ser de natureza quantitativa (contínua ou discreta) ou qualitativa (nominal ou ordinal).

No presente estudo, as medidas de concordância adotam o conceito das medidas de semelhança entre objetos. As medidas de semelhança podem ser definidas como funções que permitem quantificar o valor de semelhança ou proximidade que existe entre dois objetos, através da comparação das suas características. Estas medidas assumem valores no intervalo  $[0,1]$  pelo que, quanto maior for a semelhança entre dois objetos, mais próximos de 1 serão os valores.

No contexto desta dissertação, cada par de objetos corresponde às respostas dadas por cada um dos jovens e pelo respetivo encarregado de educação a cada uma das questões escolhidas. Por sua vez, as medidas de concordância a aplicar vão basear-se no Coeficiente de Semelhança de Gower e numa extensão deste coeficiente proposta por Podani [26].

### 3.2.1 Coeficiente de Gower

O Coeficiente de Gower, na sua forma original, permite combinar diferentes medidas de semelhança aplicadas a dados de natureza qualitativa nominal (duas ou mais categorias) e quantitativa, mas não de natureza qualitativa ordinal. Podani propõe uma extensão deste coeficiente para dados de natureza ordinal, permitindo que o Coeficiente de Gower englobe todos os tipos de dados. Assim, para cada par de respostas ( $pai_j, filho_j$ ), representado por  $(p_j, f_j)$ :

- se questão  $i$ ,  $i = 1, \dots, I$ , for nominal com  $K$  níveis e o inquirido registar uma ou mais escolhas

(resposta múltipla), a medida de concordância é dada por:

$$s_i(p_j, f_j) = \frac{\sum_{k=1}^K s_{ik}(p_j, f_j)}{K}, \quad (3.1)$$

em que, para cada nível  $k$ :

$$s_{ik}(p_j, f_j) = \begin{cases} 1 & \text{se } p_{ikj} = f_{ikj}, \\ 0 & \text{se } p_{ikj} \neq f_{ikj}. \end{cases} \quad (3.2)$$

onde  $p_{ikj}$  representa a resposta ao nível  $k$  da questão  $i$  registada pelo pai, definindo-se  $f_{ikj}$  de forma análoga para o filho.

- se questão  $i$  for quantitativa:

$$s_i(p_j, f_j) = 1 - \frac{|p_{ij} - f_{ij}|}{R_i}, \quad (3.3)$$

em que:

$$R_i = \max(\max_j(p_{ij}), \max_j(f_{ij})) - \min(\min_j(p_{ij}), \min_j(f_{ij})).$$

- se questão  $i$  for ordinal:

$$s_i(p_j, f_j) = 1, \text{ se } r_i(p_j) = r_i(f_j). \quad (3.4)$$

$$s_i(p_j, f_j) = 1 - \frac{|r_i(p_j) - r_i(f_j)| - (T_i(p_j) - 1)/2 - (T_i(f_j) - 1)/2}{\max\{r_i\} - \min\{r_i\} - (T_{i,\max} - 1)/2 - (T_{i,\min} - 1)/2}, \text{ se } r_i(p_j) \neq r_i(f_j). \quad (3.5)$$

em que:

- $r_i(p_j)$  : *rank score* da resposta do *pai* <sub>$j$</sub> , i.e., posição que a resposta do *pai* <sub>$j$</sub>  ocupa na amostra constituída pelas respostas dos pais ordenadas de forma crescente;
- $r_i(f_j)$  : *rank score* da resposta do *filho* <sub>$j$</sub> ;
- $T_i(p_j)$  : n<sup>o</sup> de pais e filhos que têm *rank score* igual ao do *pai* <sub>$j$</sub> ;
- $T_i(f_j)$  : n<sup>o</sup> de pais e filhos que têm *rank score* igual ao do *filho* <sub>$j$</sub> ;
- $T_{i,\max}$  : n<sup>o</sup> de pais e filhos que têm *rank* igual ao *rank* máximo =  $\max\{r_i\}$ ;
- $T_{i,\min}$  : n<sup>o</sup> de pais e filhos que têm *rank* igual ao *rank* mínimo =  $\min\{r_i\}$ .

Por fim, o Coeficiente de Gower:

$$G(p_j, f_j) = \frac{\sum_{i=1}^I w_i(p_j, f_j) \times s_i(p_j, f_j)}{\sum_{i=1}^I w_i(p_j, f_j)}. \quad (3.6)$$

em que:

$$w_i(p_j, f_j) = \begin{cases} 0 & \text{se } p_{ij} \text{ ou } f_{ij} \text{ desconhecidos,} \\ 1 & \text{c.c.} \end{cases}$$

isto é,  $w_i(p_j, f_j)$  é zero se o *pai* <sub>$j$</sub>  ou o *filho* <sub>$j$</sub>  não responderam à questão  $i$ .

O Coeficiente de Gower permite agregar os níveis de concordância obtidos a partir das medidas apresentadas anteriormente e quantificar o grau de concordância global entre as respostas de cada um dos jovens e do respetivo encarregado de educação a todas as questões escolhidas ou apenas a um conjunto de questões.

Estas medidas vão permitir quantificar o quão semelhantes são as respostas dadas, ao tomarem valores entre 0 e 1, correspondendo o valor 0 a ausência de concordância e o valor 1 a concordância total. Valores mais elevados indicam uma melhor percepção parental em relação ao peso e estilos de vida dos adolescentes.

Como resultado da harmonização dos inquéritos e da descrição das medidas de concordância a aplicar, encontram-se na Tabela 3.8 as questões sobre o peso e estilos de vida dos adolescentes, a classificação das respetivas respostas e as medidas de concordância adequadas.

Tabela 3.8: Harmonização e medidas de concordância das questões sobre o peso e estilos de vida dos adolescentes

	Questão	Resposta	Classificação	Medida de concordância
Peso	Descrição do peso do adolescente	1. Magro 2. Peso normal 3. Excesso de peso 4. Obeso	Qualitativa ordinal	Questão ordinal
Hábitos alimentares	Se o adolescente toma, de 2 <sup>a</sup> . a 6 <sup>a</sup> . feira, o pequeno-almoço antes de sair de casa	1. Sim 2. Às vezes 3. Não	Qualitativa ordinal	Questão ordinal
	Se em casa do adolescente há sempre fruta e produtos hortícolas			
Hábitos de atividade física	De que forma os encarregados de educação influenciam a prática de atividade física/desportiva do adolescente	1. Participam com ele 2. Compram-lhe material para desporto 3. Vão com ele para espaços ao ar livre 4. Dizem-lhe que deve fazer atividade física 5. Assistem às suas atividades desportivas 6. Levam-no aos locais onde faz desporto 7. Falam-lhe na importância do exercício físico na saúde 8. Não lhe dizem nada sobre se deve ou não fazer atividade física	Qualitativa nominal com 8 níveis (resposta múltipla)	Questão nominal com K níveis e a possibilidade do inquirido registar uma ou mais escolhas
Hábitos sedentários	Quantas horas, em média, por dia, gasta o adolescente a ver televisão	Resposta escrita para os três períodos da semana "2 <sup>a</sup> . a 6 <sup>a</sup> . feira", "sábado" e "domingo"	Quantitativa (contínua)	Questão quantitativa
	Quantas horas, em média, por dia, gasta o adolescente no computador ou em jogos eletrónicos			
	Se o jovem costuma ver, geralmente, TV às refeições	1. Sim 2. Não	Qualitativa nominal com 2 níveis	Questão nominal (equação 3.2)

Na Tabela 3.8, estão referidas as medidas de concordância a aplicar para cada uma das questões em estudo. No entanto, para o caso dos hábitos que apresentam mais de uma questão e em que as questões são de igual natureza (hábitos alimentares) ou de natureza diferente (hábitos sedentários), é necessário aplicar o coeficiente de Gower para a obtenção da concordância global relativamente a esses hábitos.

### 3.3 Técnicas de Inferência Estatística Não Paramétrica

As técnicas de Inferência Estatística Não Paramétrica surgem como um processo de atenuar problemas de difícil resolução no campo da Estatística Paramétrica. Os métodos paramétricos obrigam a que, muitas vezes, sejam assumidos pressupostos que nem sempre são reflexo do modelo estudado e que, quando não são verificados, podem implicar erros na análise e conclusões.

Quando os pressupostos de normalidade e homocedasticidade falham em relação à distribuição da população em estudo ou se desconhece *a priori* essa distribuição, os testes não paramétricos funcionam como uma alternativa aos testes paramétricos.

Nesta dissertação serão aplicados os testes de *Mann-Whitney-Wilcoxon* e de *Kruskal-Wallis*, com o objetivo de comparar duas ou mais populações em relação às suas distribuições. Com o objetivo de verificar a existência de associação entre duas variáveis, será aplicado o coeficiente de correlação de *Spearman*.

#### 3.3.1 Teste de *Mann-Whitney-Wilcoxon*

O teste de *Mann-Whitney-Wilcoxon* é um teste que permite a comparação entre duas amostras independentes, sendo usado como alternativa ao teste *t* paramétrico para duas amostras independentes. Enquanto o teste *t* compara as médias de duas amostras independentes, o teste de *Mann-Whitney-Wilcoxon* compara o centro da localização das duas amostras, como forma de detetar diferenças entre as duas populações correspondentes.

Considerem-se duas amostras aleatórias independentes de dimensões  $n_1$  e  $n_2$ ,  $(x_1, x_2, \dots, x_n)$ , retirada da população  $X$  e  $(y_1, y_2, \dots, y_n)$ , retirada da população  $Y$ .

Suponha-se, sem perda de generalidade, que  $n_1 \leq n_2$  com um total  $n = n_1 + n_2$ .

As hipóteses a testar são:

$H_0$ : As duas amostras são provenientes de populações com a mesma distribuição *vs.*

$H_1$ : As duas amostras são provenientes de populações com distribuições distintas

Se  $F(x)$  definir a função de distribuição da população  $X$  e  $G(x)$  a função de distribuição da população  $Y$ , que supõem-se contínuas, as hipóteses anteriores podem ser reformuladas e escritas do seguinte modo:

$$H_0 : F(x) = G(x), \forall x \text{ vs. } H_1 : \exists x : F(x) \neq G(x).$$

A estatística de teste tem como base as ordens (*ranks*) das observações das amostras. Assim, as duas amostras são combinadas e ordenam-se as  $n$  observações por ordem crescente, assinalando o grupo a que pertencem. No caso de existirem observações empatadas, atribui-se a média dada pela posição sequencial das observações que lhe correspondem.

A estatística de teste é dada por

$$U = \min(U_1, U_2), \tag{3.7}$$

em que:  $U_1 = n_1 n_2 + \frac{n_1(n_1+1)}{2} - R_1$  e  $U_2 = n_1 n_2 + \frac{n_2(n_2+1)}{2} - R_2$ ,

onde  $n_1$  é a dimensão da amostra menor,  $n_2$  a dimensão da amostra maior,  $R_1$  a soma das ordenações da amostra menor e  $R_2$  a soma das ordenações da amostra maior.

Quando  $n_1$  e  $n_2$  são pequenos ( $\leq 10$ ) compara-se o valor observado da estatística de teste,  $U_{obs}$ , com o valor tabelado,  $U_\alpha$ , usando a Tabela da distribuição *Mann-Whitney-Wilcoxon*. Neste caso, se  $U_{obs} < U_\alpha$ , o valor observado da estatística de teste pertence à região crítica, pelo que se rejeita a hipótese nula de as duas amostras serem provenientes de populações com a mesma distribuição, ao nível de significância  $\alpha$ .

Quando  $n_1$  e  $n_2$  são grandes ( $> 10$ ), considera-se a seguinte estatística de teste:

$$Z = \frac{U - \frac{n_1 n_2}{2}}{\sqrt{\frac{n_1 n_2 (n_1 + n_2 + 1)}{12}}}, \quad (3.8)$$

que tem distribuição assintótica  $N(0, 1)$ , se  $H_0$  verdadeira.

Se o valor observado da estatística de teste pertencer à região não crítica, i.e.,  $z_{\frac{\alpha}{2}} < Z < z_{1-\frac{\alpha}{2}}$ , em que  $z_p$  é o quantil de probabilidade  $p$  da distribuição  $N(0, 1)$ , não se rejeita  $H_0$  ao nível de significância  $\alpha$ .

Alternativamente, se o valor-p, i.e., a probabilidade de se obter sob a hipótese  $H_0$  um valor pelo menos tão extremo quanto o resultado que se observou, for inferior ao nível de significância  $\alpha$  estabelecido, rejeita-se  $H_0$ .

No caso de se verificar a rejeição da hipótese nula, significa que há evidência de que existe diferença significativa entre as distribuições das duas populações.

### 3.3.2 Teste de *Kruskal-Wallis*

O teste de *Kruskal-Wallis* é uma generalização do teste de *Mann-Whitney-Wilcoxon* para comparação de três ou mais amostras independentes.

Considerem-se  $p$  ( $\geq 3$ ) populações,  $X_1, X_2, \dots, X_p$ , a partir das quais são retiradas  $p$  amostras aleatórias, de dimensões  $n_1, n_2, \dots, n_p$ . Deste modo, tem-se, genericamente,  $(X_{i1}, X_{i2}, \dots, X_{in})$  da população  $X_i$ .

Considere-se  $R(X_{ij})$  a ordem (*rank*) atribuída à observação  $X_{ij}$  quando se considera a reunião das amostras e seja  $R_i = \sum_{j=1}^{n_i} R(X_{ij})$ , a soma das ordens das observações da  $i$ -ésima amostra,  $i = 1, 2, \dots, p$ .

O número total de observações é dado por  $N = \sum_{i=1}^p n_i$ .

As hipóteses a testar são:

$H_0$ : As  $p$  populações têm a mesma distribuição vs.

$H_1$ : Pelo menos uma das  $p$  populações não tem a mesma distribuição

A estatística de teste é dada por

$$K = \frac{1}{s^2} \left[ \sum_{i=1}^p \frac{R_i^2}{n_i} - \frac{N(N-1)^2}{4} \right], \quad (3.9)$$

onde:  $s^2 = \frac{1}{N-1} \left[ \sum_{i=1}^p \sum_{j=1}^{n_i} R(X_{ij})^2 - \frac{N(N-1)^2}{4} \right]$ .

Para o caso de não existirem empates (ou de o seu número ser muito pequeno), esta estatística reduz-se a

$$K = \frac{12}{N(N+1)} \sum_{i=1}^p \frac{R_i^2}{n_i} - 3(N+1). \quad (3.10)$$

Quanto maior for o valor observado de  $K$ , menos plausível é a hipótese nula de as  $p$  populações terem a mesma distribuição e, portanto, rejeita-se  $H_0$  para valores grandes da estatística de teste. Assim:

- 1) Se  $p = 3$  e  $n_i \leq 5$ , para  $i = 1, \dots, p$ , a distribuição exata de  $K$  encontra-se tabelada, rejeitando-se  $H_0$  ao nível de significância  $\alpha$  se  $K_{obs} > k_{crit,\alpha}$ ,
- 2) Se  $p > 3$  ou  $n_i > 5$  para algum  $i$ ,  $K$  tem distribuição assintótica  $\chi_{p-1}^2$ , sob a validade de  $H_0$ . Se  $K_{obs} > \chi_{p-1,1-\alpha}^2$ , rejeita-se  $H_0$  ao nível de significância  $\alpha$ .

Alternativamente, se o valor-p for inferior ao valor  $\alpha$  estabelecido, rejeita-se  $H_0$  ao nível de significância  $\alpha$ .

No caso de se verificar a rejeição da hipótese nula, conclui-se que existem diferenças significativas entre as distribuições das  $p$  populações em estudo.

### 3.3.3 Coeficiente de correlação de Spearman

O coeficiente de correlação de *Spearman* é uma estatística que avalia o grau de correlação entre duas variáveis quantitativas. Este coeficiente é utilizado como alternativa ao coeficiente de correlação de *Pearson*, quando os pressupostos subjacentes a este coeficiente (distribuição normal bivariada e homocedasticidade) não são satisfeitos.

Considere-se uma amostra aleatória de dimensão  $n$  de dados bivariados  $(X_i, Y_i), i = 1, \dots, n$ , proveniente de uma população contínua  $(X, Y)$ .

Seja  $R_i$  = ordem da observação  $X_i$  na amostra ordenada dos  $X_i$ 's e  $S_i$  = ordem da observação  $Y_i$  na amostra ordenada dos  $Y_i$ 's.

Seja  $D_i = R_i - S_i$ , a diferença entre as ordens atribuídas ao  $i$ -ésimo indivíduo em cada uma das ordenações  $X$  e  $Y$ .

Admitindo que não existem empates entre os  $X_i$ 's nem entre os  $Y_i$ 's (ou de o seu número ser muito pequeno), o coeficiente de correlação de *Spearman* é dado por

$$r_s = 1 - \frac{6 \sum_{i=1}^n D_i^2}{n(n^2 - 1)}. \quad (3.11)$$

O coeficiente  $r_s$  toma valores no intervalo  $[-1, 1]$ , assumindo o valor 1 quando entre o conjunto das observações existe uma associação direta perfeita (i.e., valores mais altos de uma variável estão associados a valores mais altos da outra variável) e o valor -1 quando se verifica uma associação inversa perfeita (i.e., valores mais altos de uma variável estão associados a valores mais baixos da outra variável). Quando não se verifica qualquer associação entre as variáveis  $X$  e  $Y$ , o coeficiente  $r_s$  assume valores próximos de zero.

O coeficiente de *Spearman* pode ser usado como estatística de teste para testar a hipótese de que não existe associação entre  $X$  e  $Y$  contra a hipótese alternativa conveniente.

Assim, as hipóteses em teste são:

$H_0$ : Não existe associação entre as variáveis  $X$  e  $Y$ , i.e.,  $X$  e  $Y$  são independentes *vs.*

$H_1$ : Existe associação entre  $X$  e  $Y$  / Existe associação direta entre  $X$  e  $Y$  / Existe associação indireta entre  $X$  e  $Y$

Se  $H_0$  verdadeira,  $E(r_s) = 0$  e  $var(r_s) = 1/n - 1$ . Assim, para uma amostra de dimensão elevada ( $n > 30$ ), a distribuição de  $\sqrt{n-1}r_s$  pode ser aproximada pela distribuição  $N(0, 1)$ .

Rejeita-se a hipótese nula de que não existe associação entre as variáveis se o valor-p associado ao valor observado da estatística de teste for inferior ao nível de significância  $\alpha$  estabelecido.

O nível de significância estabelecido para os testes de *Mann-Whitney-Wilcoxon* e de *Kruskal-Wallis* e para o coeficiente de correlação de *Spearman* é  $\alpha = 0.05$ .

### 3.4 Modelos de Regressão

Em muitos estudos estatísticos existem problemas em que o principal objetivo é estudar a relação entre variáveis, ou mais particularmente, analisar a influência que uma ou mais variáveis (explicativas) têm sobre uma variável de interesse, a variável resposta. O modo como o estatístico aborda tal problema é através do estudo de um modelo de regressão que relaciona a variável de interesse com as variáveis explicativas. Quando a relação entre a variável de interesse e as variáveis explicativas é linear e a variável resposta tem distribuição Normal, os modelos de regressão designam-se de modelos lineares.

Os modelos lineares generalizados surgem como uma extensão dos modelos lineares para os casos em que se verifica que a variável resposta possa ter uma distribuição que não seja a Normal e em que a relação entre a variável resposta e as variáveis explicativas não necessite de ser linear. Desta forma, os modelos lineares generalizados assumem para a variável resposta uma distribuição pertencente a uma família de distribuições com propriedades muito específicas, a família exponencial, e uma função de ligação que relaciona o valor médio da variável resposta e a combinação linear das variáveis explicativas (i.e., preditor linear).

Nesta dissertação serão usados dois casos particulares dos modelos lineares generalizados, o modelo de regressão logística binária e o modelo de regressão logística ordinal e um modelo não pertencente à classe dos modelos lineares generalizados, o modelo de regressão beta.

#### 3.4.1 Modelo de Regressão Logística Binária

Na regressão logística binária, a variável resposta (dependente) é dicotômica, tomando valores 1 quando o acontecimento de interesse ocorre (sucesso) e 0 quando não ocorre (insucesso).

Seja  $Y$  a variável resposta que assume o valor 1 ou 0. As probabilidades de ocorrência e de não ocorrência do acontecimento de interesse são dadas, respetivamente, por

$$P(Y = 1) = \pi \quad \text{e} \quad P(Y = 0) = 1 - \pi.$$

Desta forma, a variável  $Y$  tem distribuição de Bernoulli de parâmetro  $\pi$ , cuja função de distribuição de

probabilidade é dada por

$$P(Y = y) = \pi^y(1 - \pi)^{1-y},$$

com  $E(Y) = \pi$  e  $Var(Y) = \pi(1 - \pi)$ .

### 3.4.1.1 Formulação do modelo

Sendo o objetivo dos modelos de regressão estudar a relação entre uma ou mais variáveis explicativas e uma variável resposta considere-se, para o caso em que existe mais de uma variável explicativa, um conjunto  $p$  de variáveis independentes dadas pelo vetor

$$X^T = (X_1, \dots, X_p).$$

A probabilidade da  $i$ -ésima observação da variável resposta,  $Y_i$ , condicionada à observação de um conjunto de variáveis independentes  $\mathbf{x}_i = (x_{i1}, x_{i2}, \dots, x_{ip})^T$ , tomar valor 1 é dada por

$$P(Y_i = 1 | \mathbf{x}_i), \text{ para } i = 1, \dots, n,$$

que corresponde ao valor médio da variável resposta  $Y_i$  dado o vetor de variáveis independentes  $\mathbf{x}_i$ , para  $i = 1, \dots, n$ .

No modelo de regressão clássico o valor médio é dado por

$$E(Y | X_i = x_i) = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n.$$

No caso da regressão logística, para descrever a relação linear existente entre a variável resposta e as variáveis explicativas, faz-se uso da função *logit*. Esta função é definida como o logaritmo da razão entre a probabilidade de sucesso e a probabilidade de insucesso do acontecimento de interesse, sendo a equação correspondente dada por

$$\text{logit}[P(Y_i = 1 | X_i)] = \ln \left[ \frac{P(Y_i = 1 | X_i)}{1 - P(Y_i = 1 | X_i)} \right] = \beta_0 + \beta_1 x_{i1} + \dots + \beta_p x_{ip}, \quad i = 1, \dots, n, \quad (3.12)$$

onde os coeficientes de regressão  $\beta_j, j = 1, \dots, p$ , representam a variação esperada no logaritmo da chance (*logit*) por unidade de variação na variável  $X_j$ , partilhando as restantes variáveis o mesmo valor.

A probabilidade correspondente é, então, dada por

$$P(Y_i = 1 | X_i) = \frac{e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}{1 + e^{\beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}}}, \quad i = 1, \dots, n. \quad (3.13)$$

### 3.4.1.2 Estimação dos coeficientes de regressão

Nos modelos lineares generalizados, o principal objetivo é a estimação dos parâmetros desconhecidos do modelo, no caso  $\beta_j, j = 1, \dots, p$ . Para a estimação destes coeficientes recorre-se ao método da máxima verosimilhança cuja solução é obtida através de métodos iterativos.

### 3.4.1.3 Inferência sobre o modelo

Depois de estimados os coeficientes do modelo de regressão é necessário avaliar a sua significância, i.e., determinar se as variáveis explicativas inseridas no modelo estão associadas significativamente à variável resposta. Para esse efeito recorre-se à estatística de *Wald* e ao teste de razão de verosimilhanças.

#### Teste de *Wald*

O teste de *Wald* é o teste utilizado para verificar se a relação entre cada uma das variáveis explicativas e a variável resposta é significativa. As hipóteses a testar são:

$$H_0 : \beta_j = 0 \text{ vs. } H_1 : \beta_j \neq 0, \quad j = 1, \dots, p.$$

A estatística de teste e a respetiva distribuição assintótica, sob a validade de  $H_0$ , são dadas por

$$W_j = \frac{\hat{\beta}_j}{se(\hat{\beta}_j)} \cap N(0, 1), \quad (3.14)$$

onde  $se(\hat{\beta}_j)$  é o erro-padrão do estimador  $\hat{\beta}_j$  do coeficiente de regressão  $\beta_j$ .

A hipótese  $H_0$  é rejeitada se o valor-p associado ao valor observado da estatística de teste for inferior a um nível de significância  $\alpha$  estabelecido.

#### Teste de razão de verosimilhanças

O teste de razão de verosimilhanças é o teste utilizado quando se pretende comparar dois modelos aninhados, i.e., modelos em que as variáveis de um modelo constituem um subconjunto das variáveis do outro modelo.

Dados dois modelos aninhados,  $M_p$  e  $M_q$ , com um número de variáveis  $p$  e  $q$ , respetivamente, tal que  $p < q$ , o teste de razão de verosimilhanças é aplicado para testar a hipótese

$$H_0: \beta_{p+1} = \beta_{p+2} = \dots = \beta_q = 0 \text{ vs. } H_1: \exists j : \beta_j \neq 0, j = p + 1, \dots, q$$

A estatística de teste e a respetiva distribuição assintótica, sob a validade de  $H_0$ , são dadas por

$$G = -2 \ln \left[ \frac{\max L_{M_p}(\beta)}{\max L_{M_q}(\beta)} \right] \cap \chi_{q-p}^2, \quad (3.15)$$

onde  $\beta = (\beta_0, \dots, \beta_p, \beta_{p+1}, \dots, \beta_q)$ ,  $L_{M_p}(\beta)$  é a função verosimilhança<sup>1</sup> sob  $H_0$  e  $L_{M_q}(\beta)$  é a função verosimilhança<sup>1</sup> sob  $H_1$ .

Rejeita-se a hipótese nula de não diferença estatisticamente significativa entre os dois modelos, para um dado nível de significância  $\alpha$ , se o valor-p associado ao valor observado da estatística de teste for inferior a  $\alpha$ . Se  $H_0$  é rejeitada, o modelo considerado é o especificado na hipótese  $H_1$ .

---

<sup>1</sup>função verosimilhança, como função de  $\beta$ .

#### 3.4.1.4 Seleção de Modelos

A abordagem tradicional na construção de modelos estatísticos consiste em encontrar o modelo mais parcimonioso possível e que se ajuste adequadamente aos dados observados. Na prática, há normalmente um elevado número de variáveis que podem ser importantes para explicar a variabilidade da variável resposta, o que implica na existência de vários modelos com diferentes combinações das variáveis explicativas. Na regressão logística, um dos métodos que facilita o processo de seleção é o método de seleção *Stepwise*.

##### Método *Stepwise*

O método *Stepwise* é um procedimento automático de seleção de variáveis que consiste em três métodos de seleção: *forward*, *backward* e *both*.

O método de seleção *forward stepwise* inicia-se a partir de um modelo nulo (i.e., modelo que contém um único parâmetro) ao qual se vai adicionando, uma de cada vez, as variáveis que podem ser significativas para explicar a variabilidade da variável resposta.

O método *backward stepwise* é o oposto do método *forward stepwise*, em que a construção do modelo se inicia com o modelo maximal (i.e., modelo com todas as variáveis explicativas) ao qual se verifica, a cada passo, se uma variável pode ou não ser excluída do modelo.

Por último, o método *both stepwise* consiste numa combinação dos dois métodos atrás mencionados. Este método funciona como o método *forward*, partindo de um modelo nulo ou completo, com a diferença de que uma variável que foi incluída ou excluída do modelo pode vir a ser excluída ou incluída num dos passos seguintes. Desta forma, depois de adicionada ou retirada uma variável do modelo, o método permite confirmar se algumas variáveis já incluídas ou excluídas podem vir a ser depois excluídas ou incluídas no modelo.

A fase de incluir ou excluir uma variável do modelo é a fase de avaliação da significância das variáveis ou de comparação de modelos. Para isso, são utilizadas medidas estatísticas adequadas para a sua avaliação, podendo ser uma dessas medidas o critério de informação de Akaike (AIC) [32]. Esta medida é uma estatística que tem por base o logaritmo da verosimilhança e penaliza o modelo que contém muitas variáveis. A medida AIC é dada por

$$AIC = -2[\ln(L) - k], \quad (3.16)$$

onde  $k$  é o número de parâmetros do modelo (incluindo o parâmetro de interseção  $\beta_0$ ) e  $L$  o valor da verosimilhança para o modelo estimado.

A medida AIC mede o grau de informação que se perde com o ajuste de um determinado modelo, sendo principalmente útil na comparação de modelos que não partilhem o mesmo subconjunto de variáveis (modelos não aninhados). Quanto menor o valor de AIC de um determinado modelo, menor será a perda de informação e melhor será o ajustamento desse modelo. Deste modo, o modelo a ser selecionado será o que apresenta menor valor AIC.

#### 3.4.1.5 Qualidade de ajustamento do modelo

Através do método de seleção *Stepwise* são selecionadas as covariáveis mais importantes a partir de um conjunto inicial de variáveis explicativas. Apesar de este método selecionar as variáveis mais importan-

tes para o modelo, não garante, do ponto de vista prático, que o modelo seja o melhor. Desta forma, para garantir que o modelo selecionado se ajusta adequadamente aos dados são aplicadas ferramentas estatísticas, como medidas de qualidade e análise de resíduos.

## Medidas de Qualidade

Na regressão logística, para avaliar a qualidade de ajustamento do modelo podem ser usados a *Deviance* (Desvio) e o Teste de *Hosmer e Lemeshow*.

### *Deviance*

Para verificar a qualidade de ajustamento do modelo estimado consideram-se dois modelos, o modelo com as variáveis selecionadas (modelo ajustado- $M$ ) e o modelo saturado- $S$ , com um número de variáveis  $p$  e  $n$  respetivamente ( $p < n$ ), em que se introduz uma medida que dá a distância entre os valores ajustados pelo modelo  $M$  e os valores observados. Essa medida de discrepância entre o modelo  $S$  e o modelo  $M$  à qual se dá o nome de *Deviance* é baseada na estatística de razão de verosimilhanças e é dada por [32]

$$D = -2 \ln \left[ \frac{L(\text{modelo ajustado})}{L(\text{modelo saturado})} \right] \cap \chi_{n-p}^2, \quad (3.17)$$

representando o valor  $D$  o desvio do modelo ajustado em relação ao modelo saturado.

Para um determinado nível de significância  $\alpha$ , rejeita-se a hipótese  $H_0$  de o modelo ajustado ser considerado adequado aos dados se o valor observado de  $D$  for superior a  $\chi_{n-p, 1-\alpha}^2$  ou se valor- $p > \alpha$ . Deste modo, quanto menor o valor de  $D$ , melhor será a adequação do modelo ajustado aos dados observados.

### Teste de *Hosmer e Lemeshow*

O Teste de *Hosmer e Lemeshow* é um teste que avalia, de forma global, a qualidade de ajustamento de um modelo.

As hipóteses a testar são:

$H_0$ : O modelo ajusta-se bem aos dados vs.  $H_1$ : O modelo não se ajusta bem aos dados

Para calcular a estatística de Hosmer e Lemeshow é necessário a divisão da amostra em  $g$  grupos, sendo o número de grupos mais utilizado  $g = 10$ . A distribuição dos grupos é feita com dimensão aproximadamente igual, sendo o primeiro grupo constituído pelas observações que possuem os menores valores para as probabilidades estimadas pelo modelo e o último grupo constituído pelas observações que possuem os valores mais altos para as probabilidades estimadas. A distribuição destes dados encontra-se organizada sob a forma de uma tabela de contingência  $g \times 2$ , que contém as frequências observadas (i.e., número de sucessos entre os indivíduos  $y$  do grupo  $i$ ,  $i = 1, \dots, g$ ) e esperadas. Para o caso da variável resposta  $Y$  tomar valor 1, a frequência esperada é dada pela soma das probabilidades estimadas de todos os indivíduos dentro de cada grupo  $i$ , para  $i = 1, \dots, g$ .

Desta forma, a estatística de teste sob a validade de  $H_0$  é dada por

$$X_{HL}^2 = \sum_{i=1}^g \frac{(z_i - \hat{u}_i)^2}{\hat{u}_i} \cap \chi_{g-2}^2 \quad (3.18)$$

em que:  $z_i = \sum_{j=1}^{n_i} y_j$ ,  $n_i$  é a dimensão do grupo  $i = 1, \dots, g$  e  $\hat{u}_i = \sum_{j=1}^{n_i} \hat{y}_j$ .

A hipótese do modelo se ajustar bem aos dados é rejeitada se o valor-p associado ao valor observado da estatística de teste for menor que o nível de significância  $\alpha$  estabelecido.

### **Análise de Resíduos**

A análise de resíduos é útil para avaliar a qualidade de ajustamento do modelo quanto à escolha da distribuição, da função de ligação e em termos do preditor linear, assim como na identificação de observações mal ajustadas pelo modelo (Turkman e Silva, 2000).

O resíduo bruto ou *raw residual* é definido como a diferença entre o valor observado  $y_i$  e o valor  $\hat{\mu}_i$  ajustado pelo modelo, para uma determinada observação  $i$ .

Diferentes tipos de resíduos podem ser utilizados para avaliar a qualidade de ajustamento do modelo, tais como: resíduos de *Pearson*, resíduos da *Deviance*, resíduos de verosimilhança e resíduos padronizados.

### **Resíduos de Pearson**

O resíduo de *Pearson* para a  $i$ -ésima observação é definido por

$$r_{P_i} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)}}, \quad i = 1, 2, \dots, n. \quad (3.19)$$

O resíduo  $r_{P_i}$ , assim definido, corresponde à contribuição de cada observação para o cálculo da estatística de *Pearson* generalizada, dada pela expressão

$$X^2 = \sum_{i=1}^n \frac{(y_i - \hat{\mu}_i)^2}{V(\hat{\mu}_i)}, \quad (3.20)$$

onde  $V(\hat{\mu}_i)$  é a função de variância estimada para a distribuição do modelo em estudo.

### **Resíduos da Deviance**

O resíduo da *Deviance* correspondente à  $i$ -ésima observação é dado pela seguinte expressão

$$r_{D_i} = \text{sin}(\text{al}(y_i - \hat{\mu}_i) \sqrt{d_i}), \quad (3.21)$$

onde  $d_i$  é a contribuição da  $i$ -ésima observação para a medida *Deviance*.

### **Resíduos padronizados**

Para uma análise adequada dos resíduos é necessário que eles sejam padronizados pelo respetivo desvio padrão.

O resíduo de *Pearson* padronizado é dado por

$$r_{P,E} = \frac{y_i - \hat{\mu}_i}{\sqrt{V(\hat{\mu}_i)(1 - h_{ii})}}, \quad i = 1, 2, \dots, n. \quad (3.22)$$

E o resíduo da *Deviance* é dado por

$$r_{D,E} = \frac{r_{D_i}}{\sqrt{(1 - h_{ii})}}, \quad i = 1, 2, \dots, n, \quad (3.23)$$

onde  $h_{ii}$  é o *leverage* da  $i$ -ésima observação.

### Resíduos de verosimilhança

Os resíduos de verosimilhança são obtidos através da comparação entre o desvio verificado quando se ajusta o modelo às  $n$  observações e o desvio obtido quando se aplica o mesmo modelo a  $n - 1$  observações, i.e., excluindo o  $i$ -ésimo elemento da amostra. Sendo a aproximação usada para o cálculo destes resíduos  $h_{ii}r_{P,E}^2 + (1 - h_{ii})r_{D,E}^2$ , o resíduo de verosimilhança para a  $i$ -ésima observação é dado por

$$r_{L_i} = \text{sign}(y_i - \hat{\mu}_i) \sqrt{h_{ii}r_{P,E}^2 + (1 - h_{ii})r_{D,E}^2}. \quad (3.24)$$

### Análise gráfica dos resíduos

As representações gráficas dos resíduos são um método informal para verificar a qualidade de ajustamento do modelo.

#### 1) Avaliação do preditor linear e identificação de *outliers*

Para avaliação do preditor linear e identificação de *outliers* são utilizados os seguintes gráficos:

- Gráfico dos resíduos padronizados contra a ordem das observações.  
Este gráfico permite detetar as observações consideradas *outliers*, uma vez que permite identificar os resíduos com valores anormalmente elevados ao se encontrarem fora do limite considerado para a distribuição dos resíduos. Permite, também, avaliar a qualidade de ajustamento do modelo através da análise do padrão da distribuição dos resíduos. Se o gráfico apresentar um padrão sistemático significa que o modelo ajustado não é o correto.
- Gráfico das probabilidades obtidas por validação cruzada.  
Este gráfico é indicado para a deteção de *outliers* e consiste no cálculo da probabilidade da  $i$ -ésima observação sabendo as restantes  $n - 1$  observações. Esta probabilidade é designada de probabilidade cruzada e é representada por

$$P(y_i | \mathbf{y}_{(i)}) \approx \exp\left(\frac{-r_{L_i}^2}{2}\right) \sqrt{1 - h_{ii}}, \quad (3.25)$$

onde  $\mathbf{y}_{(i)}$  corresponde ao conjunto de dados que exclui a  $i$ -ésima observação.

Este gráfico indica as observações com probabilidade reduzida de ocorrerem, sendo consideradas

como possíveis *outliers*.

## 2) Avaliação da função de ligação

Um método para verificar se a função de ligação escolhida é adequada consiste em considerar  $Z$  como uma nova variável explicativa a adicionar ao preditor linear [33]. Cada valor da variável  $Z$  é dado por  $z_i = -\{1 + \hat{p}_i^{-1} \ln(1 - \hat{p}_i)\}$ , em que  $\hat{p}_i$  corresponde à probabilidade do valor ajustado  $\hat{\mu}_i$  da  $i$ -ésima observação. Depois de adicionada a variável  $Z$ , verifica-se se há declínio na função desvio do modelo e se este é significativo. Caso o desvio não seja significativo, a função de ligação é considerada adequada.

## 3) Identificação de observações influentes

Uma observação é considerada influente se a sua exclusão do conjunto de dados produz alterações significativas nas estimativas dos parâmetros do modelo.

A identificação de observações influentes pode ser feita através da medida  $h_{ii}$ , cujo valor representa a influência da  $i$ -ésima observação sobre o próprio valor ajustado. Segundo os autores Hoaglin & Welsch (1978), uma observação é considerada influente se o valor  $h_{ii} > 2\frac{p}{n}$ , sendo  $p$  o número de parâmetros no modelo e  $n$  o número total de observações.

Para avaliar de que forma uma dada observação  $i$  pode influenciar as estimativas dos parâmetros do modelo, utiliza-se a medida distância de *Cook* dada por

$$DC_i = \frac{h_{ii}^2 r_{iE}^2}{p(1 - h_{ii})}. \quad (3.26)$$

Uma observação é considerada tão mais influente quanto maior for o seu valor  $DC_i$ .

As representações gráficas úteis na identificação de observações influentes são os gráficos de  $h_{ii}$  e  $DC_i$  contra a ordem das observações, sendo designados como Gráfico de *leverage* e Gráfico da estatística  $D$ , respetivamente.

### 3.4.1.6 Capacidade preditiva do modelo

Quando o objetivo do ajustamento do modelo de regressão logística é a predição, torna-se necessário avaliar a capacidade preditiva do modelo, i.e., a sua capacidade de discriminar os indivíduos que experimentaram o acontecimento de interesse daqueles que não experimentaram.

Hosmer & Lemeshow (2013) sugerem algumas técnicas usadas para a avaliação da capacidade preditiva do modelo, entre elas a Curva ROC e a *Confusion matrix*.

#### Curva ROC

A curva ROC (*Receiver Operating Characteristic*) é uma representação gráfica que permite ilustrar a capacidade de discriminação de um modelo de regressão logística binária, através do estudo da variação da sensibilidade e 1-especificidade para diferentes pontos de corte.

Seja  $\hat{Y} = 1$  se um indivíduo selecionado da população em estudo for classificado como acontecimento de interesse e  $\hat{Y} = 0$  se classificado como não acontecimento. Desta forma, a sensibilidade é definida como

a probabilidade de o acontecimento de interesse ser classificado de forma correta, i.e.,  $P(\hat{Y} = 1|Y = 1)$  e a especificidade como a probabilidade do não acontecimento de interesse ser classificado corretamente,  $P(\hat{Y} = 0|Y = 0)$ .

Para se proceder à classificação dos indivíduos, torna-se necessário estabelecer um ponto de corte que determine a probabilidade de os indivíduos serem classificados como sendo ou não acontecimentos de interesse. Se considerarmos o ponto de corte  $C = 0.5$ , significa que, para um valor  $\hat{Y}$  maior ou igual a 0.5 o indivíduo será classificado na classe 1, caso contrário será classificado na classe 0. No entanto, através da curva ROC, é possível escolher um ponto de corte que maximize simultaneamente a sensibilidade e a especificidade do modelo. Desta forma, a curva ROC é um gráfico que representa a sensibilidade (ou taxa de verdadeiros positivos) *versus* 1-especificidade (ou taxa de falsos positivos) resultantes da variação de um valor de corte, podendo essa representação evidenciar os valores para os quais existe otimização da sensibilidade em função da especificidade, sendo o ponto correspondente o que se encontra mais próximo do canto superior esquerdo do diagrama, uma vez que o índice de verdadeiro positivo é 1 e o de falso positivo 0.

### **Confusion matrix**

A matriz de confusão (*confusion matrix*) é uma tabela  $2 \times 2$  construída para o ponto de corte escolhido, como apresentado na Tabela 3.9.

Tabela 3.9: Tabela de classificação

Classificação	Valores observados		Total
	Classe(0)	Classe(1)	
Classe (0)	$n_{11}$	$n_{12}$	$n_{1.}$
Classe (1)	$n_{21}$	$n_{22}$	$n_{2.}$
Total	$n_{.1}$	$n_{.2}$	$n$

A partir da tabela de classificação, as medidas de desempenho do modelo são dadas por:

$$\text{Sensibilidade} = n_{22}/n_{2.}$$

$$\text{Especificidade} = n_{11}/n_{1.}$$

$$\text{Porcentagem total de acertos} = ((n_{11} + n_{22})/n) \times 100$$

Apesar do gráfico da curva ROC fornecer o melhor ponto de corte, a área sob a curva (AUC), que varia entre 0 e 1, classifica de forma mais precisa a capacidade de discriminação do modelo, i.e., fornece uma medida da capacidade do modelo em discriminar os valores da variável resposta  $Y = 1$  dos valores de  $Y = 0$ . De acordo com Hosmer & Lemeshow (2013), as diferentes classificações para os valores da área sob a curva ROC são:

- se  $AUC = 0.5$ , não há discriminação;
- se  $0.7 \leq AUC < 0.8$ , discriminação aceitável;
- se  $0.8 \leq AUC < 0.9$ , discriminação excelente;
- se  $AUC \geq 0.9$ , discriminação excepcional.

### 3.4.1.7 Interpretação dos parâmetros do modelo

Depois de ajustado o modelo e assumindo o pressuposto de que ele se ajusta adequadamente aos dados e de que os coeficientes estimados são significativos, torna-se necessário interpretar os valores associados aos coeficientes.

Para a interpretação dos coeficientes recorre-se a duas expressões: *odds* (possibilidades) e *odds ratio* (razão de possibilidades). O termo *odds* é definido como o quociente entre a probabilidade de ocorrência de um acontecimento ( $\pi$ ), e a probabilidade de não ocorrência do acontecimento ( $1 - \pi$ ), ou seja

$$Odds = \frac{\pi}{1 - \pi}, \quad (3.27)$$

em que  $A$  e  $B$  são dois cenários de um acontecimento, sendo o *odds* do cenário  $A$  dado por  $\frac{\pi(A)}{1 - \pi(A)}$  e o do cenário  $B$  dado por  $\frac{\pi(B)}{1 - \pi(B)}$ .

O *odds ratio* é definido como

$$OR = \frac{\frac{\pi(A)}{1 - \pi(A)}}{\frac{\pi(B)}{1 - \pi(B)}}, \quad (3.28)$$

que descreve a possibilidade de ocorrência do cenário  $A$  em relação à ocorrência do cenário  $B$  do acontecimento.

A interpretação dos valores dos coeficientes vai depender da natureza das variáveis explicativas que, no caso desta dissertação, são categóricas (nominais e ordinais) e contínuas. No caso das variáveis explicativas categóricas, estas têm duas ou mais categorias.

#### Variável explicativa categórica com duas categorias

Quando uma variável explicativa,  $x$ , tem apenas duas categorias, ela toma os valores 0 e 1.

Admitindo que  $x$  é a única variável explicativa no modelo, uma simplificação do modelo descrito na equação 3.12 assume a forma:

$$\text{logit}[P(Y = 1|x)] = \ln\left(\frac{\pi(x)}{1 - \pi(x)}\right) = \beta_0 + \beta_1 x. \quad (3.29)$$

Quando  $x = 0$ :  $\text{logit}[P(Y = 1|x = 0)] = \ln\left(\frac{\pi(0)}{1 - \pi(0)}\right) = \beta_0$ , pelo que  $\frac{\pi(0)}{1 - \pi(0)} = e^{\beta_0}$

Quando  $x = 1$ :  $\text{logit}[P(Y = 1|x = 1)] = \ln\left(\frac{\pi(1)}{1 - \pi(1)}\right) = \beta_0 + \beta_1$ , pelo que  $\frac{\pi(1)}{1 - \pi(1)} = e^{\beta_0 + \beta_1}$

A medida *odds ratio* é escrita como o quociente do *odds* para  $x = 1$  relativamente ao *odds* para  $x = 0$ , ou seja:

$$OR = \frac{\text{odds}(Y = 1|x = 1)}{\text{odds}(Y = 1|x = 0)} = \frac{e^{\beta_0 + \beta_1}}{e^{\beta_0}} = e^{\beta_1}. \quad (3.30)$$

Este valor representa a variação na chance da variável resposta  $Y$  tomar valor 1 quando a variável explicativa toma valor 1 ( $x = 1$ ) em comparação a quando toma valor 0 ( $x = 0$ ).

#### Variável explicativa categórica com mais de duas categorias

Quando uma variável explicativa,  $x$ , é constituída por mais de duas categorias é necessário utilizar variáveis *dummy*. Estas variáveis funcionam como variáveis dicotómicas, assumindo apenas os valores 0 ou 1. Caso a variável em estudo seja composta por  $k$  categorias, deverão ser criadas  $k - 1$  variáveis *dummy*, que vão corresponder a  $k - 1$  categorias da variável explicativa. A categoria à qual não é indexada nenhuma variável *dummy* corresponde à classe de referência, que assume valor 0. Desta forma, é possível comparar cada uma das  $k - 1$  categorias com a categoria de referência.

O cálculo do *odds ratio* vai ser semelhante ao caso da variável com duas categorias, bem como a interpretação.

### Variável explicativa contínua

Quando um modelo logístico contém uma variável explicativa contínua é necessário, antes de fazer a interpretação do respetivo coeficiente estimado, verificar a linearidade entre a variável resposta e a variável explicativa na escala *logit*.

Uma forma de verificar a linearidade é a representação gráfica dos valores *logit* (ou *ln odds*) da variável resposta em função dos valores da variável explicativa, através de um gráfico de dispersão. Caso a distribuição dos pontos seja linear, significa que existe linearidade entre as variáveis na escala *logit*.

Um outro método que pode ser aplicado para verificar a linearidade é a comparação entre dois modelos lineares, em que um dos modelos tem um termo linear  $x$  e o outro tem o termo linear  $x$  e um termo não linear  $x^2$  [30]. Desta forma, os modelos são dados por

$$\text{Modelo 1 : } \text{logit}[P(Y = 1|x)] = \beta_0 + \beta_1 x \quad (3.31)$$

$$\text{Modelo 2 : } \text{logit}[P(Y = 1|x)] = \beta_0 + \beta_1 x + \beta_2 x^2, \quad (3.32)$$

em que  $Y$  é a variável resposta e  $x$  é a variável explicativa contínua.

As hipóteses a testar são:

$H_0$ : O modelo é linear em  $x$  vs.  $H_1$ : O modelo é quadrático em  $x$

Que equivale a:

$H_0$ :  $\beta_2 = 0$  vs.  $H_1$ :  $\beta_2 \neq 0$

Para um nível de significância  $\alpha$  estabelecido, rejeita-se a hipótese nula se o valor-p associado ao valor observado da estatística de teste for inferior a  $\alpha$ . Se a hipótese nula for rejeitada, significa que o modelo é quadrático em  $x$ , i.e., não existe linearidade na escala *logit* entre a variável resposta  $Y$  e a variável explicativa  $x$ . No caso de não se verificar a condição de linearidade, torna-se necessário categorizar a variável explicativa, em duas ou mais categorias.

Assumindo que existe linearidade na escala *logit* para a variável contínua e que existe apenas uma variável no modelo, a respetiva equação é dada por

$$\text{logit}[P(Y = 1|x)] = \ln \left[ \frac{P(Y = 1|x)}{1 - P(Y = 1|x)} \right] = \beta_0 + \beta_1 x. \quad (3.33)$$

Neste caso,  $\beta_1$  representa o valor da mudança no *logit* quando o valor da variável  $x$  aumenta em uma unidade.

A variação do *odds* correspondente a um incremento unitário é dado por

$$OR = \frac{\text{odds}(Y = 1|x + 1)}{\text{odds}(Y = 1|x)} = \frac{e^{\beta_1(x+1)}}{e^{\beta_1 x}} = e^{\beta_1}. \quad (3.34)$$

Se o valor da variável  $x$  aumentar em  $k$  unidades, a diferença no *logit* será de  $k\beta_1$  e o *odds ratio* será dado por

$$OR = e^{k\beta_1}. \quad (3.35)$$

### Interpretação dos valores de OR

Quanto à interpretação dos valores de *OR* para variáveis categóricas, temos:

- $OR = 1$ : significa que o *odds* do acontecimento  $Y = 1$  para os indivíduos do grupo  $x = 1$  é igual ao *odds* do acontecimento  $Y = 1$  para os indivíduos do grupo  $x = 0$ ;
- $OR > 1$ : significa que o *odds* do acontecimento  $Y = 1$  para os indivíduos do grupo  $x = 1$  é superior ao *odds* do acontecimento  $Y = 1$  para os indivíduos do grupo  $x = 0$ ;
- $OR < 1$ : significa que o *odds* do acontecimento  $Y = 1$  para os indivíduos do grupo  $x = 1$  é inferior ao *odds* do acontecimento  $Y = 1$  para os indivíduos do grupo  $x = 0$ .

No caso das variáveis contínuas, a interpretação dos valores de *OR* é a seguinte:

- $OR = 1$ : significa que o *odds* do acontecimento  $Y = 1$  é igual para os indivíduos com um incremento unitário na variável  $x$  e para os indivíduos sem o incremento unitário na variável  $x$ ;
- $OR > 1$ : significa que o *odds* do acontecimento  $Y = 1$  é superior para os indivíduos com um incremento unitário na variável  $x$  relativamente aos indivíduos sem o incremento na variável  $x$ ;
- $OR < 1$ : significa que o *odds* do acontecimento  $Y = 1$  é inferior para os indivíduos com um incremento unitário na variável  $x$  relativamente aos indivíduos sem o incremento na variável  $x$ .

### 3.4.2 Modelo de Regressão Logística Ordinal

Na Secção 3.3.1 foi descrito o modelo de regressão logística que é aplicado quando a variável resposta  $Y$  assume os valores 0 ou 1. No entanto, quando estamos perante uma variável resposta com mais de duas categorias e em que existe uma ordem subjacente a essas categorias, torna-se necessário aplicar um modelo de regressão específico que incorpore este tipo de informação. O modelo a adotar será um dos modelos que pertence à classe da regressão logística ordinal. Esta classe de modelos é constituída por diferentes modelos, dos quais Hosmer & Lemeshow (2013) destacam *Adjacent categories logit model*, *Continuation ratio logit model* e *Proportional odds model*, sendo este último o que irá ser abordado nesta

dissertação, devido à sua fácil aplicação e interpretação.

### 3.4.2.1 Proportional Odds Model

Considerando a regressão logística ordinal como uma generalização da regressão logística binária:

Seja  $Y$  a variável resposta composta por  $J (J \geq 3)$  categorias ordenadas mutuamente exclusivas e exaustivas;

Seja  $X_i$  um conjunto de  $p$  variáveis independentes para cada  $i$ -ésima observação;

Seja  $\pi_{ij} = Pr(Y_i = j|X_i), i = 1, \dots, n$  e  $j = 1, \dots, J$ ; a probabilidade da variável resposta  $Y$  para  $i$ -ésima observação condicionada a um conjunto de variáveis  $X_i$ , assumir como resposta a  $j$ -ésima categoria.

A probabilidade cumulativa da  $i$ -ésima observação, condicionada a um conjunto de variáveis  $X_i$ , pertencer à categoria  $j$  ou a uma das categorias inferiores a  $j$  é dada por

$$P_{ij} = Pr(Y_i \leq j|X_i) = \pi_{i1}(X_i) + \pi_{i2}(X_i) + \dots + \pi_{ij}(X_i). \quad (3.36)$$

Para descrever a relação existente entre a variável resposta e as variáveis explicativas pode fazer-se uso de diferentes funções de ligação, sendo a mais usual na regressão ordinal a função *logit*, já aplicada na regressão logística binária.

Considere-se  $g(\cdot)$  como a função *logit*. O modelo de *odds* proporcionais (*proportional odds model*) é dado por

$$\begin{aligned} g(P_{ij}) &= \ln \frac{P_{ij}(X_i)}{1 - P_{ij}(X_i)} = \ln \left[ \frac{Pr((Y_i \leq j)|X_i)}{1 - Pr((Y_i \leq j)|X_i)} \right] = \ln \left[ \frac{Pr((Y_i \leq j)|X_i)}{Pr((Y_i > j)|X_i)} \right] = \\ &= \ln \left[ \frac{\pi_{i1} + \pi_{i2} + \dots + \pi_{ij}}{\pi_{ij+1} + \pi_{ij+2} + \dots + \pi_{iJ}} \right] = \theta_j - X_i^T \beta, \quad i = 1, \dots, n \text{ e } j = 1, \dots, J - 1. \end{aligned}$$

Ou equivalentemente

$$Pr(Y_i \leq j|X_i) = \frac{e^{\theta_j - X_i^T \beta}}{1 + e^{\theta_j - X_i^T \beta}}, \quad i = 1, \dots, n \text{ e } j = 1, \dots, J - 1, \quad (3.37)$$

em que  $\theta_j$  varia entre os  $J - 1$  *logits*, permanecendo os valores dos coeficientes de regressão iguais para cada uma das equações.

Assim, para um conjunto de variáveis explicativas  $X$ , o modelo vai comparar a probabilidade de  $Y$  tomar como respostas a categoria  $j$  e inferiores a  $j$  com a probabilidade de  $Y$  tomar como respostas categorias superiores a  $j$ , para  $j = 1, \dots, J - 1$ .

Quando a variável resposta tem apenas duas categorias, o modelo descrito na equação 3.37 equivale ao modelo logístico binário.

A estimação dos coeficientes de regressão é feita com recurso ao método da máxima verosimilhança.

Depois de estimados os coeficientes do modelo é avaliada a sua significância, recorrendo-se à estatística de *Wald* e ao teste de razão de verosimilhanças, descritas na Secção 3.3.1.3.

### 3.4.2.2 Pressuposto de *odds* proporcionais

Um dos pressupostos subjacentes ao modelo de *odds* proporcionais é o de que a relação entre cada par de categorias-resposta que a variável  $Y$  toma é a mesma. Em outras palavras, este modelo assume que os coeficientes que descrevem a relação entre  $Y$  tomar como respostas as categorias inferiores ou iguais à  $j$ -ésima categoria *versus* tomar como respostas as categorias superiores à  $j$ -ésima categoria são iguais aos coeficientes que descrevem a relação entre a variável  $Y$  tomar como respostas as categorias inferiores ou iguais à  $(j + 1)$ -ésima categoria *versus* tomar como respostas as categorias superiores à  $(j + 1)$ -ésima categoria. Este pressuposto é designado de *odds* proporcionais ou de regressão paralela.

Como referido anteriormente, o modelo assume o pressuposto de que a relação entre todos os pares de categorias-resposta de  $Y$  é a mesma, o que resulta em apenas um conjunto de coeficientes. Caso o pressuposto não seja satisfeito torna-se necessário formular diferentes modelos para descrever a relação existente entre cada par de categorias-reposta.

Para saber se o pressuposto de *odds* proporcionais é satisfeito existem dois testes que podem ser aplicados: o teste de razão de verosimilhanças e o teste de *Brant*. Estes testes, em amostras de grande dimensão, são assintoticamente equivalentes e produzem a mesma resposta.

#### Teste de razão de verosimilhanças

No contexto da regressão logística ordinal, o teste de razão de verosimilhanças compara o modelo de *odds* proporcionais (modelo restrito) com o modelo que não assume o pressuposto de as *odds* serem proporcionais (modelo não restrito).

As hipóteses a testar são:

$H_0$ : Não há diferença entre os coeficientes das variáveis independentes que descrevem a relação entre todos os pares categorias-resposta da variável  $Y$  (pressuposto de *odds* proporcionais é satisfeito) *vs.*

$H_1$ : Existe diferença entre os coeficientes que descrevem a relação entre todos os pares categorias-resposta da variável  $Y$

A estatística de teste e a respetiva distribuição, sob a validade de  $H_0$ , são dadas por

$$G = -2\ln \left[ \frac{L(\text{modelo restrito})}{L(\text{modelo não restrito})} \right] \cap \chi^2_{[(J+1)-2] \times p}, \quad (3.38)$$

onde  $L(\text{modelo restrito})$  é a função de verosimilhança do modelo de *odds* proporcionais e  $L(\text{modelo não restrito})$  a função de verosimilhança do modelo que não assume o pressuposto de as *odds* serem proporcionais.  $[(J + 1) - 2] \times p$  corresponde aos graus de liberdade para este teste, em que  $(J + 1)$  é o número de categorias da variável resposta e  $p$  o número de variáveis explicativas em cada modelo.

Para um nível de significância  $\alpha$  estabelecido rejeita-se a hipótese nula, de que o pressuposto de *odds*

proporcionais é satisfeito, se o valor-p associado ao valor observado da estatística de teste for inferior a  $\alpha$ .

### Teste de *Brant*

Para verificar se o pressuposto de *odds* proporcionais é satisfeito, o teste de *Brant* verifica, para cada variável independente do modelo  $k$ ,  $k = 1, \dots, p$ , se o coeficiente associado descreve a mesma relação entre todos os pares de categorias-resposta.

As hipóteses em teste são:

$H_0$ : Não há diferença entre o coeficiente da variável independente  $k$  que descreve a relação entre todos os pares categorias-resposta da variável  $Y$  vs.

$H_1$ : Há diferença entre o coeficiente da variável independente  $k$  que descreve a relação entre todos os pares categorias-resposta da variável  $Y$

Este teste compara as estimativas dos coeficientes obtidas a partir do ajustamento do modelo descrito na equação 3.37 (modelo de *odds* proporcionais) com as estimativas obtidas do ajustamento do seguinte modelo

$$Pr(Y_i \leq j | X_i) = \frac{e^{\theta_j - X_i^T \beta_j}}{1 + e^{\theta_j - X_i^T \beta_j}}, \quad i = 1, \dots, n \text{ e } j = 1, \dots, J - 1, \quad (3.39)$$

em que  $\theta_j < \theta_{j+1}$ .

Neste modelo, os valores dos coeficientes de regressão podem ser diferentes entre os pares categorias-resposta da variável  $Y$ , ao contrário do que acontece no modelo de *odds* proporcionais.

O teste de *Brant* tem distribuição Qui-Quadrado em que o número de graus de liberdade é igual a  $[(J + 1) - 2] \times p$ , onde  $(J + 1)$  é o número de categorias da variável resposta e  $p$  o número de variáveis explicativas em cada modelo.

Para um nível de significância  $\alpha$  estabelecido, rejeita-se a hipótese nula  $H_0$  se o valor-p associado ao valor observado da estatística de teste for inferior a  $\alpha$ .

Para além do teste de *Brant* verificar o pressuposto de *odds* proporcionais para cada uma das variáveis independentes, também o verifica para todas as variáveis independentes em simultâneo, tal como o teste de razão de verosimilhanças.

#### 3.4.2.3 Estratégias de construção de modelos

Algumas das estratégias para a construção de modelos de regressão logística binária podem ser aplicadas na construção de modelos de regressão logística ordinal (Hosmer & Lemeshow, 2013). No caso da seleção de modelos, o método de seleção *Stepwise* descrito na Secção 3.3.1.4 é um dos métodos que pode ser aplicado.

Depois de ajustado o modelo, torna-se necessário verificar se o ajustamento é adequado, podendo ser feita esta verificação através de testes e de análise de resíduos.

Os testes habituais para a verificação da qualidade de ajustamento do modelo são os testes *Lipsitz* e *Hosmer-Lemeshow* para modelos ordinais. Como cada um destes testes deteta diferentes tipos de falta

de ajuste do modelo, Hosmer & Lemeshow recomendam o uso de ambos os testes. As hipóteses em teste são iguais às dos testes aplicados para verificar a qualidade de ajustamento de modelos logísticos binários, em que um valor-p não significativo indica um bom ajustamento do modelo aos dados.

Em relação à análise de resíduos, esta é considerada problemática, devido ao facto de não estar disponível para modelos ordinais em *packages* de *softwares* estatísticos. Desta forma, o que se recomenda é que a análise seja feita para modelos binários separados, sendo os resultados binários dados por

$$\tilde{Y}_{ji} = \begin{cases} 1 & \text{se } Y_i \leq j, \\ 0 & \text{se } Y_i > j. \end{cases} \quad j = 1, \dots, J-1 \quad (3.40)$$

Se tomarmos como exemplo a variável resposta  $Y$  assumir como respostas uma de três categorias ( $J = 3$ ), o número de modelos binários formulados será dois ( $J - 1$ ). Ao aplicar a expressão 3.40 para o primeiro modelo formulado, a variável  $Y$  vai assumir os valores 0 ou 1 se:

$$\tilde{Y}_{ji} = \begin{cases} 1 & \text{se } Y_i \leq 1, \\ 0 & \text{se } Y_i > 1. \end{cases} \quad j = 1, 2 \quad (3.41)$$

Para o segundo modelo, a variável  $Y$  vai assumir os valores 0 ou 1 se:

$$\tilde{Y}_{ji} = \begin{cases} 1 & \text{se } Y_i \leq 2, \\ 0 & \text{se } Y_i > 2. \end{cases} \quad j = 1, 2 \quad (3.42)$$

Desta forma, a análise de resíduos descrita na Secção 3.3.1.5 pode ser aplicada a cada um dos novos modelos binários.

#### 3.4.2.4 Interpretação dos parâmetros do modelo

Depois de ajustado o modelo e assumindo que os pressupostos do modelo ordinal (*odds* proporcionais e adequação do modelo) são satisfeitos e que os coeficientes estimados são significativos, procede-se à interpretação dos valores associados aos parâmetros do modelo.

As medidas a usar são as medidas já definidas na Secção 3.3.1.7: *odds* e *odds ratio*. Desta forma, considere-se uma dada variável explicativa do modelo como sendo dicotómica. O respetivo *odds ratio* é dado por

$$OR = \frac{\text{odds}(Y \leq j | x = 1)}{\text{odds}(Y \leq j | x = 0)} = \frac{e^{\tau_{j+1} - (\beta_0 + \beta_1)}}{e^{\tau_{j+1} - \beta_0}} = e^{-\beta_1}, \quad j = 1, \dots, J-1. \quad (3.43)$$

Este valor representa a chance da variável resposta  $Y$  tomar como resposta as categorias mais baixas quando a variável explicativa toma valor 1 em comparação a quando toma valor 0.

Quando no modelo existem variáveis categóricas com mais de duas categorias e variáveis contínuas, o raciocínio em relação à comparação das diferentes categorias ou dos valores que as variáveis atrás referidas tomam é o mesmo que apresentado na regressão logística binária, sendo a única diferença a forma como se interpreta o valor do *odds ratio*. Na interpretação deste valor, em vez de se considerar unicamente uma resposta binária ( $Y = 0$  ou  $Y = 1$ ), consideram-se diversas respostas binárias resultantes da agregação progressiva das categorias mais baixas,  $Y \leq j$  vs  $Y > j$ , para  $j = 1, \dots, J-1$ .

No caso das variáveis contínuas, para se verificar a linearidade existente entre elas e a variável resposta na escala *logit*, pode recorrer-se aos mesmos métodos referidos no modelo de regressão logística binária.

### 3.4.3 Modelo de Regressão Beta

A distribuição beta é frequentemente usada para modelar variáveis aleatórias que assumem valores no intervalo ]0,1[, como percentagens, taxas e proporções. Caso se pretenda relacionar uma variável de interesse com um conjunto de variáveis explicativas e tendo a variável de interesse distribuição beta, a classe de modelos de regressão a aplicar será a classe de modelos de regressão beta (Ferrari e Cribari-Neto, 2004). Esta classe de modelos, como já referido, é útil para modelar variáveis contínuas  $Y$  que assumem valores no intervalo ]0,1[, mas quando essas variáveis assumem os valores extremos do intervalo, 0 e 1, é necessário aplicar uma transformação aos valores de  $y$ . Essa transformação é dada por  $(y \times (n - 1) + 0.5)/n$ , sendo  $n$  a dimensão da amostra [41].

A função densidade de probabilidade da distribuição beta é dada por

$$f(y; \mu; \phi) = \frac{\Gamma(\phi)}{\Gamma(\mu\phi)\Gamma((1-\mu)\phi)} y^{\mu\phi-1} (1-y)^{(1-\mu)\phi-1}, \quad 0 < y < 1, \quad (3.44)$$

com  $0 < \mu < 1$  e  $\phi > 0$ .  $\Gamma(\cdot)$  é a função gama e  $\phi$  o parâmetro de precisão.

Desta forma, a variável  $y$  segue distribuição beta com parâmetros  $\mu$  e  $\phi$ , em que:

$E(y) = \mu$  e  $Var(y) = \mu(1-\mu)/(1+\phi)$ , sendo  $\mu$  o valor médio da variável resposta.

#### 3.4.3.1 Formulação do modelo

Seja  $Y$  a variável resposta e  $Y_1, \dots, Y_n$  a amostra aleatória de tal modo que  $Y_i \sim B(\mu_i, \phi), i = 1, \dots, n$ . O modelo de regressão beta é dado por

$$g(\mu_i) = x_i^T \beta = \eta_i, \quad (3.45)$$

em que  $\beta = (\beta_1, \dots, \beta_k)^T$  é um vetor de  $k \times 1$  parâmetros de regressão desconhecidos ( $k < n$ ),  $\mathbf{x}_i = (x_{i1}, \dots, x_{ik})^T$  o vetor de  $k$  variáveis explicativas para a  $i$ -ésima observação,  $\eta_i$  o preditor linear da  $i$ -ésima observação (i.e.,  $\eta_i = \beta_1 x_{i1} + \dots + \beta_k x_{ik}$ ) e  $g(\cdot)$  a função de ligação.

No modelo de regressão beta pode fazer-se uso de diferentes funções de ligação como: *logit*, *probit*, *complementary log-log*, *log-log* e *Cauchy*. A função de ligação que irá ser adotada é a função *logit*, dada por

$$\text{logit } g(\mu) = \ln(\mu/(1-\mu)), \quad (3.46)$$

que descreve a relação linear entre a variável resposta e as variáveis explicativas.

Deste modo, o valor médio correspondente à  $i$ -ésima observação é dado por

$$E(y_i) = \mu_i = \frac{e^{x_i^T \beta}}{1 + e^{x_i^T \beta}}. \quad (3.47)$$

A estimação dos parâmetros do modelo é feita com recurso ao método da máxima verosimilhança.

Depois de os parâmetros do modelo serem estimados, torna-se necessário avaliar a sua significância, podendo recorrer-se à estatística de *Wald* e à estatística de razão de verosimilhanças.

### 3.4.3.2 Seleção de modelos

Na classe de modelos de regressão beta, a seleção de modelos pode ser feita com recurso ao critério de informação de Akaike (AIC), descrito na Secção 3.4.1.4, em que o modelo selecionado é o modelo com menor valor AIC.

### 3.4.3.3 Medidas de diagnóstico

Como já referido para os modelos de regressão logística, após o ajustamento de um modelo, torna-se necessário verificar se o modelo estimado se ajusta adequadamente aos dados. Para isso, podem ser usadas medidas de qualidade e análise de resíduos.

#### Medidas de qualidade

Na regressão beta, uma das medidas de qualidade de ajustamento do modelo que pode ser aplicada é a medida designada por pseudo  $R^2$ ,  $R_p^2$ . Esta medida é definida pelo quadrado do coeficiente de correlação da amostra entre os valores observados da variável resposta,  $y_1, \dots, y_n$ , e os respetivos valores estimados pelo modelo,  $\hat{\mu}_1, \dots, \hat{\mu}_n$ .  $R_p^2$  toma valores entre 0 e 1,  $0 \leq R_p^2 \leq 1$ , correspondendo o valor 1 a um perfeito ajustamento do modelo aos dados observados.

Outras medidas pseudo  $R^2$  que podem ser aplicadas são  $R_p^{2*} = 1 - \ln \hat{L} / \ln \hat{L}_0$  e  $R_{LR}^2 = 1 - (\hat{L}_0 / \hat{L})^{2/n}$ , onde  $\hat{L}_0$  e  $\hat{L}$  são os valores de verosimilhança calculados para o modelo nulo e para o modelo ajustado, respetivamente.

#### Análise de resíduos

Nos modelos de regressão beta, diferentes tipos de resíduos podem ser utilizados para verificar a qualidade de ajustamento do modelo estimado, dentre os quais se destacam os resíduos de *Pearson* e os resíduos da *deviance*.

#### Resíduos de *Pearson*

O resíduo de *Pearson* para a  $i$ -ésima observação é definido por:

$$r_i = \frac{y_i - \hat{\mu}_i}{\sqrt{\hat{v}ar(y_i)}}, \quad (3.48)$$

onde  $\hat{\mu}_i = g^{-1}(x_i^T \hat{\beta})$  e  $\hat{v}ar(y_i) = \{\hat{\mu}_i(1 - \hat{\mu}_i)\} / (1 - \hat{\phi})$ .

#### Resíduos da *deviance*

O resíduo da *deviance* correspondente à  $i$ -ésima observação é dado pela seguinte expressão:

$$r_i^d = \text{sign}(y_i - \hat{\mu}_i) \{2(l_i(\tilde{\mu}_i, \hat{\phi}) - l_i(\hat{\mu}_i, \hat{\phi}))\}^{1/2}, \quad (3.49)$$

em que cada uma das observações  $i, i = 1, \dots, n$ , contribui a quantidade  $(r_i^d)^2$  para a medida *Deviance*.

As observações cujos resíduos da *deviance* tomem valores elevados podem ser consideradas *outliers*.

As representações gráficas úteis para a verificação da qualidade de ajustamento do modelo ajustado são:

- Gráfico dos resíduos de *Pearson* contra o índice das observações ou o preditor linear.  
Em relação ao primeiro gráfico referido, este permite detetar não só as observações consideradas *outliers*, devido aos valores elevados dos respetivos resíduos, como também avaliar a qualidade de ajustamento do modelo através da análise do padrão da distribuição dos resíduos. Se o gráfico apresentar um padrão sistemático significa que o modelo estimado não é o correto.  
Em relação ao segundo gráfico referido, este permite detetar, através da análise do padrão da distribuição dos resíduos, se a função de ligação escolhida é a adequada. Se for detetada qualquer tendência na distribuição dos resíduos, pode significar que a função de ligação escolhida não é a mais adequada.
- Gráfico dos resíduos da *deviance* contra o índice das observações.  
Este gráfico, tal como o gráfico dos resíduos de *Pearson* contra o índice das observações, permite detetar as observações consideradas *outliers*.
- Gráfico *Half-normal* dos resíduos da *deviance* (valores absolutos) com envelope simulado.  
Este gráfico permite decidir se os resíduos observados são consistentes com o modelo ajustado. Se existir uma proporção considerável de resíduos fora dos limites do envelope, isto pode ser visto como uma evidência contra a adequabilidade do modelo ajustado.

### Identificação de observações influentes

Para a identificação de observações influentes, Wei et al. (1998) propuseram uma medida *leverage* generalizada,  $GL(\tilde{\theta})$ . O elemento  $(i, u)$  de  $GL(\tilde{\theta})$ , i.e., o valor de *leverage* generalizado do estimador  $\tilde{\theta}$  em  $(i, u)$ , corresponde à taxa instantânea de mudança no  $i$ -ésimo valor ajustado em relação ao  $u$ -ésimo valor observado, sendo  $\theta$  o vetor dos parâmetros do modelo.

Para avaliar de que forma uma dada observação  $i$  pode influenciar as estimativas dos parâmetros do modelo, utiliza-se a medida aproximada da distância de *Cook* dada por

$$C_i = \frac{h_{ii} r_i^2}{k(1 - h_{ii})^2}, \quad (3.50)$$

onde  $h_{ii}$  corresponde ao *leverage* da  $i$ -ésima observação.

Uma observação é considerada tão mais influente quanto maior for o valor  $C_i$  correspondente.

As representações gráficas úteis na identificação de observações influentes são os gráficos de  $h_{ii}$  generalizado e  $C_i$  contra o índice das observações.

#### 3.4.3.4 Interpretação dos parâmetros do modelo

Depois de ajustado o modelo e assumindo o pressuposto de que ele se ajusta adequadamente aos dados e de que os coeficientes estimados são significativos, procede-se à interpretação dos valores associados aos parâmetros do modelo.

As medidas que podem ser usadas são *odds* e *odds ratio*.

Considerando o modelo composto por uma única variável explicativa e sendo esta dicotômica, o respetivo *odds ratio* é dado da mesma forma que na regressão logística binária, ou seja:

$$OR = e^{\beta_1}. \quad (3.51)$$

Quando se compara a variável explicativa tomar valor 1 ( $x = 1$ ) em relação a quando toma valor 0 ( $x = 0$ ), o valor de *odds ratio* representa o aumento ou diminuição (dependendo do sinal de  $\beta$ ) da estimativa da percentagem/taxa/proporção esperada do evento de interesse.

Quando no modelo existem variáveis categóricas com mais de duas categorias e variáveis contínuas, o raciocínio em relação à comparação das diferentes categorias ou dos valores que as variáveis atrás referidas tomam é o mesmo que apresentado na regressão logística binária, sendo a única diferença a forma como se interpreta o valor do *odds ratio*. No caso da regressão beta, o valor de *odds ratio* representa o aumento ou diminuição da estimativa da percentagem/taxa/proporção esperada do evento de interesse.

No caso das variáveis contínuas, para se verificar a linearidade existente entre elas e a variável resposta na escala *logit*, pode recorrer-se aos métodos referidos no modelo de regressão logística binária.

Para todos os testes descritos nos modelos de regressão logística binária, logística ordinal e beta foi estabelecido o nível de significância  $\alpha = 0.05$ .

## Capítulo 4

# Resultados

Neste capítulo todas as análises apresentadas foram implementadas através do *software R (RStudio Version 3.4.3)*. Os *packages* utilizados foram os seguintes: *aplpack*, *gmodels*, *ResourceSelection*, *Epi*, *pROC*, *StatMeasures*, *MASS*, *ordinal*, *brant*, *generalhoslem* e *betareg*.

### 4.1 Análise exploratória

Com o objetivo de compreender melhor as variáveis em estudo descritas em 2.2, procedeu-se à realização de uma análise exploratória. As Tabelas 4.1 e 4.2 dizem respeito à análise exploratória das variáveis referentes aos jovens e aos encarregados de educação, respetivamente.

Na Tabela 4.1, é possível observar que a maioria (58%) dos adolescentes que responderam ao inquérito são do sexo feminino. Em relação à idade, existe um desequilíbrio quanto à distribuição, verificando-se uma percentagem superior de jovens com idades mais baixas. Quanto à classe de peso, o peso normal predomina entre os jovens (71,8%), contudo, as percentagens referentes aos jovens com excesso de peso (20,6%) e obesidade (3,3%), apesar de baixas, são um motivo de preocupação no que respeita à saúde dos jovens. Em relação à atividade física, a maioria dos jovens pratica um nível de atividade física elevado, que se revela como um bom indicador de hábitos desportivos saudáveis.

Quanto aos encarregados de educação (Tabela 4.2), é possível observar que as mães e os pais não foram os únicos a responder, havendo também respostas de outros familiares dos jovens. No entanto, a grande maioria são mães (79,2%). A idade dos adultos que responderam aos questionários varia entre os 17 e os 82 anos, tendo a grande maioria dos adultos idade inferior a 50 anos. Em relação ao peso, cerca de 48% tem peso normal e a maioria não se preocupa com o peso dos adolescentes.

Com o objetivo de perceber se o peso dos encarregados de educação pode influenciar o peso dos adolescentes ou a preocupação que os EE têm relativamente ao peso dos jovens, foram construídas duas tabelas de análises cruzadas entre cada par de variáveis. A Tabela 4.3 mostra a análise cruzada entre o peso dos EE e o peso dos jovens e a Tabela 4.4 a análise cruzada entre o peso dos EE e a preocupação que estes têm em relação ao peso dos jovens.

Na Tabela 4.3, é possível notar que dos encarregados de educação com peso normal, 74,7% tem filhos/educandos com peso normal e dos EE com excesso de peso, 67,9% tem também educandos com peso normal. Apesar de não ter sido efetuado nenhum teste estatístico, é possível perceber que não existe

Tabela 4.1: Análise exploratória das características sociodemográficas dos adolescentes

Variável	<i>n</i> (%)*
<b>Sexo</b>	
Masculino	279 (42,0%)
Feminino	385 (58,0%)
<b>Idade</b>	
12	120 (18,1%)
13	166 (25,0%)
14	156 (23,5%)
15	112 (16,9%)
16	45 (6,8%)
17	51 (7,7%)
18	14 (2,1%)
<b>Ano de escolaridade</b>	
7º	194 (29,2%)
8º	179 (27,0%)
9º	194 (29,2%)
10º	49 (7,4%)
11º	29 (4,4%)
12º	19 (2,9%)
<b>Classe de peso</b>	
Magro	28 (4,2%)
Peso normal	477 (71,8%)
Excesso de peso	137 (20,6%)
Obeso	22 (3,3%)
<b>Nível de atividade física</b>	
Baixo	23 (3,5%)
Moderado	231 (34,8%)
Elevado	395 (59,5%)

\*Resultados obtidos de uma amostra de 664 indivíduos

uma tendência de que o peso dos jovens vai ser semelhante ao peso dos pais/EE, visto que a maioria dos EE tem educandos com peso normal.

Relativamente à preocupação dos encarregados de educação quanto ao peso dos adolescentes, verifica-se na Tabela 4.4 que dos indivíduos com peso normal, 66,1% não se preocupa com o peso dos jovens. Esta mesma situação ocorre nos encarregados de educação com excesso de peso, em que 61,5% não se preocupa com o peso dos educandos. Pelos valores obtidos da análise cruzada, verifica-se que, independentemente do estado de peso em que os EE se encontrem, a maioria não se preocupa com o peso dos jovens.

Na tentativa de perceber se a preocupação dos encarregados de educação relativamente ao peso dos jovens poderia ser diferente consoante o peso dos jovens, foi construída novamente uma tabela de análise cruzada (Tabela 4.5) entre estas duas variáveis.

É possível verificar na Tabela 4.5 que, dos jovens com peso normal, a maioria (70,8%) dos encarregados de educação não se preocupa com o peso deles. Em relação aos jovens com excesso de peso, observa-se que existe um equilíbrio quanto ao número de EE que não se preocupa com o peso dos educandos ou que se preocupa moderadamente. Apesar de o número de jovens obesos ser baixo relativamente ao número

Tabela 4.2: Análise exploratória das características sociodemográficas dos encarregados de educação

Variável	n (%)*
<b>Encarregado de educação</b>	
Pai	99 (14,9%)
Mãe	526 (79,2%)
Outro	38 (5,7%)
Avô	6 (0,9%)
Avó	5 (0,8%)
Tio(a)	8 (1,2%)
Outro	19 (2,9%)
<b>Idade</b>	
< 50	561 (84,5%)
≥ 50	82 (12,3%)
<b>Escolaridade</b>	
Até Ensino básico	257 (38,7%)
Analfabeto	2 (0,3%)
1º Ciclo	60 (9,0%)
2º Ciclo	68 (10,2%)
3º Ciclo	127 (19,1%)
Ensino secundário	184 (27,7%)
Secundário	183 (27,6%)
Curso Profissional	1 (0,2%)
Ensino universitário	166 (25,0%)
Bacharelato/Licenciatura	149 (22,4%)
Mestrado	14 (2,1%)
Doutoramento	3 (0,5%)
<b>Classe de peso</b>	
Magro	10 (1,5%)
Peso normal	320 (48,2%)
Excesso de peso	212 (31,9%)
Obeso	78 (11,7%)
<b>Preocupação com o peso do educando</b>	
Não está preocupado	405 (61,0%)
Relativamente preocupado	172 (25,9%)
Preocupado	79 (11,9%)

\*Resultados obtidos de uma amostra de 664 indivíduos

Tabela 4.3: Análise cruzada da Classe de peso dos encarregados de educação vs. Classe de peso dos jovens

		Classe de peso dos jovens			
		Magro	Peso normal	Excesso de peso	Obeso
Classe de peso dos EE	Magro	1 (10,0%)	8 (80,0%)	1 (10,0%)	0 (0,0%)
	Peso normal	13 (4,1%)	239 (74,7%)	58 (18,1%)	10 (3,1%)
	Excesso de peso	10 (4,7%)	144 (67,9%)	49 (23,1%)	9 (4,2%)
	Obeso	3 (3,8%)	50 (64,1%)	22 (28,2%)	3 (3,8%)

\*Resultados obtidos de uma amostra de 620 pares de indivíduos

de jovens com peso normal ou excesso de peso, é possível perceber que dos jovens obesos, a maioria dos EE preocupa-se com o peso (40,9%) ou que se preocupa moderadamente (50,0%). Estes resultados parecem indicar que quanto mais pesados forem os jovens, maior é a preocupação dos educadores quanto ao peso dos jovens.

Tabela 4.4: Análise cruzada da Classe de peso dos encarregados de educação vs. Preocupação dos encarregados de educação quanto ao peso dos jovens

		Preocupação dos EE quanto ao peso dos jovens		
		Não está preocupado	Está relativamente preocupado	Está preocupado
Classe de peso dos EE	Magro	5 (50,0%)	5 (50,0%)	0 (0,0%)
	Peso normal	211 (66,1%)	78 (24,5%)	30 (9,4%)
	Excesso de peso	128 (61,5%)	53 (25,5%)	27 (13,0%)
	Obeso	42 (54,5%)	21 (27,3%)	14 (18,2%)

\*Resultados obtidos de uma amostra de 614 pares de indivíduos

Tabela 4.5: Análise cruzada da Classe de peso dos jovens vs. Preocupação dos encarregados de educação quanto ao peso dos jovens

		Preocupação dos EE quanto ao peso dos jovens		
		Não está preocupado	Está relativamente preocupado	Está preocupado
Classe de peso dos jovens	Magro	14 (50,0%)	11 (39,3%)	3 (10,7%)
	Peso normal	334 (70,8%)	102 (21,6%)	36 (7,6%)
	Excesso de peso	55 (41,0%)	48 (35,8%)	31 (23,1%)
	Obeso	2 (9,1%)	11 (50,0%)	9 (40,9%)

\*Resultados obtidos de uma amostra de 656 pares de indivíduos

## 4.2 Análise da concordância parental: aplicação de medidas de concordância

Com o objetivo de caracterizar a percepção parental (ou dos EE) em relação ao peso e estilos de vida (hábitos alimentares, sedentários e de atividade física) dos adolescentes, foram realizadas análises descritivas dos valores de concordância obtidos para as questões sobre o peso e hábitos dos adolescentes. De modo a entender como os encarregados de educação percebem os jovens no seu todo, foi realizada também uma análise descritiva da percepção global. A percepção global inclui as percepções do peso e dos estilos de vida dos jovens. Os resultados obtidos encontram-se na Tabela 4.6.

Tabela 4.6: Análise descritiva da concordância parental em relação ao peso e estilos de vida dos adolescentes

Percepção	mínimo	mediana	média	máximo
Peso	0	1	0.9907	1
Hábitos alimentares	0	1	0.9611	1
Hábitos de atividade física	0.1250	0.75	0.7001	1
Hábitos sedentários	0	0.8549	0.8224	1
Global	0.3636	0.8133	0.7929	1

A partir dos resultados da Tabela 4.6, é possível constatar que a concordância média do peso e dos hábitos dos jovens são elevadas, correspondendo o valor mais alto à percepção do peso e o mais baixo aos hábitos de atividade física. Verifica-se, de igual modo, que 50% dos encarregados de educação assume valores de concordância iguais ou superiores a 75%, indicando uma boa capacidade parental no correto reconhecimento do peso e hábitos dos adolescentes.

Para saber se a concordância entre as respostas dos jovens e dos encarregados de educação a questões sobre o peso e hábitos dos adolescentes difere entre determinadas características sociodemográficas dos jovens e dos EE, foram aplicados os testes de *Mann-Whitney-Wilcoxon* e de *Kruskal-Wallis*. Para saber de que modo essa concordância pode estar associada com as características dos jovens e dos EE, foi aplicado o coeficiente de correlação de *Spearman*. Os resultados obtidos permitem caracterizar a percepção parental do peso e estilos de vida dos adolescentes para certas características dos jovens e dos EE.

#### 4.2.1 Perceção do peso

Para saber se a concordância em relação ao peso dos adolescentes difere entre determinadas características dos jovens e dos encarregados de educação, foram escolhidas as seguintes variáveis sociodemográficas:

- Jovens: **Sexo, Idade (em anos) e Classe de peso**
- Encarregados de educação: **Encarregado de educação, Idade, Escolaridade, Classe de peso e Preocupação com o peso do educando**

A análise da diferença entre os valores de concordância em função da variável **Sexo** foi obtida a partir do teste de *Mann-Whitney-Wilcoxon*. Para as restantes variáveis, à exceção da variável **Idade (EE)**, foi aplicado o teste de *Kruskal-Wallis*. Os resultados obtidos estão apresentados nas Tabelas 4.7 e 4.8.

Tabela 4.7: Concordância do peso dos jovens em função das variáveis sociodemográficas dos jovens

	<i>média</i>	<i>valor-p</i>
Sexo		0.0316
Masculino	0.9954	
Feminino	0.9873	
Idade		0.1362
12	0.9998	
13	0.9974	
14	0.9831	
15	0.9851	
16	0.9907	
17	0.9959	
18	0.9425	
Classe de peso		$< 2.2e^{-16}$
Magro	0.9136	
Peso normal	0.9990	
Excesso de peso	0.9877	
Obesidade	0.9269	

Nas Tabelas 4.7 e 4.8 verifica-se que os valores médios da percepção em relação a cada uma das variáveis em estudo são bastante elevados, com valores muito próximos de 1. O sexo masculino e o peso normal são as características dos jovens que revelam uma maior percepção por parte dos encarregados de educação (Tabela 4.7), sendo os pais que apresentam uma maior capacidade média no reconhecimento do peso dos adolescentes (Tabela 4.8). De entre as variáveis em estudo, a percepção difere significativamente entre os géneros ( $\text{valor-p} = 0.03156$ ), as classes de peso dos jovens ( $\text{valor-p} < 2.2e^{-16}$ ) e o tipo de preocupação que os encarregados de educação têm relativamente ao peso dos jovens ( $\text{valor-p} = 4.329e^{-6}$ ).

Tabela 4.8: Concordância do peso dos jovens em função das variáveis sociodemográficas dos encarregados de educação

	<i>média</i>	<i>valor-p</i>
Encarregado de educação		0.4317
Pai	0.9998	
Mãe	0.9898	
Outro	0.9787	
Escolaridade		0.8544
Até Ensino básico	0.9927	
Ensino secundário	0.9867	
Ensino universitário	0.9888	
Classe de peso		0.5946
Magro	0.9999	
Peso normal	0.9903	
Excesso de peso	0.9913	
Obesidade	0.9842	
Preocupação com o peso do educando		$4.329e^{-6}$
Não está preocupado	0.9969	
Relativamente preocupado	0.9787	
Preocupado	0.9843	

Para saber de que forma a concordância em relação ao peso dos jovens está associada com a variável **Idade dos EE**, foi aplicado o coeficiente de correlação de *Spearman*. O resultado obtido ( $r_s = -0.0082$ , valor-p = 0.8363) mostra que não existe associação entre a concordância do peso e a idade dos encarregados de educação, indicando que a percepção parental do peso dos jovens não se altera com a idade dos EE.

#### 4.2.2 Perceção dos hábitos alimentares

Na análise da percepção dos encarregados de educação em relação aos hábitos alimentares dos jovens, foram escolhidas as seguintes variáveis:

- Jovens: **Sexo e Idade (em anos)**
- Encarregados de educação: **Encarregado de educação e Idade**

A análise da diferença entre os valores de concordância em função do **Sexo** foi obtida a partir do teste de *Mann-Whitney-Wilcoxon*. Para as variáveis **Idade (jovens)** e **Encarregado de educação**, foi utilizado o teste de *Kruskal-Wallis*. Os resultados obtidos estão apresentados na Tabela 4.9.

Ao observar os valores da concordância média em relação a cada uma das variáveis em estudo (Tabela 4.9), verifica-se que os encarregados de educação percebem com melhor precisão os hábitos alimentares dos educandos do sexo masculino (média = 0.9650) e dos educandos mais novos. De entre os diferentes encarregados de educação, os que são classificados como "outro"(que inclui os avôs, avós, tios(as) e outros) têm um valor de concordância média de 0.9709, o que reflete uma melhor percepção dos hábitos alimentares dos jovens comparativamente aos pais e mães desses jovens. De entre as variáveis em estudo, a percepção parental difere significativamente entre os sexos e as idades dos jovens e entre os diferentes encarregados de educação.

Relativamente à variável **Idade dos EE**, foi aplicado o coeficiente de *Spearman* para verificar de que forma a variável está associada à concordância dos hábitos alimentares dos jovens. A partir do resultado

Tabela 4.9: Concordância dos hábitos alimentares dos jovens em função das características sociodemográficas dos jovens e encarregados de educação

	<i>média</i>	<i>valor-p</i>
<b>Jovem</b>		
Sexo		0.0074
Masculino	0.9650	
Feminino	0.9582	
Idade		$5.55e^{-4}$
12	0.9822	
13	0.9687	
14	0.9608	
15	0.9493	
16	0.9317	
17	0.9489	
18	0.9261	
<b>Encarregado de educação</b>		
Encarregado de educação		0.0049
Pai	0.9638	
Mãe	0.9598	
Outro	0.9709	

obtido ( $r_s = 0.1155$ , valor-p = 0.0034), verifica-se que a existência de associação entre as variáveis é significativa, indicando que a percepção parental relativamente aos hábitos alimentares dos jovens altera-se à medida que a idade dos EE aumenta.

### 4.2.3 Perceção dos hábitos de atividade física

Na análise da percepção parental em relação aos hábitos de atividade física dos jovens, foram escolhidas as seguintes variáveis:

- Jovens: **Sexo**, **Idade (em anos)** e **Nível de atividade física**
- Encarregados de educação: **Encarregado de educação** e **Idade**

A análise da diferença entre os valores de concordância em função do **Sexo** foi obtida a partir do teste de *Mann-Whitney-Wilcoxon*. Em relação às outras variáveis escolhidas, nomeadamente **Idade (jovens)**, **Nível de atividade física** e **Encarregado de educação**, foi aplicado o teste de *Kruskal-Wallis*. Os gráficos correspondentes à variação dos valores de concordância dos hábitos de atividade física para as variáveis **Sexo**, **Idade (jovens)**, **Nível de atividade física** e **Encarregado de educação** estão apresentados nas Figuras 4.1 a 4.4.

Os hábitos de atividade física dos jovens são os hábitos que os encarregados de educação pior percebem (Tabela 4.6), sendo estes capazes de reconhecer com maior precisão esse tipo de hábitos nas raparigas (figura 4.1) e nos adolescentes mais velhos (figura 4.2). Em relação aos encarregados de educação, os pais e as mães apresentam valores de concordância bastante altos mas também bastante baixos comparativamente aos outros encarregados de educação, indicando que há pais/mães com uma boa e uma má percepção em relação aos hábitos de atividade física dos jovens (figura 4.4). De entre as variáveis em estudo, a percepção não difere significativamente entre os encarregados de educação (valor-p = 0.821)

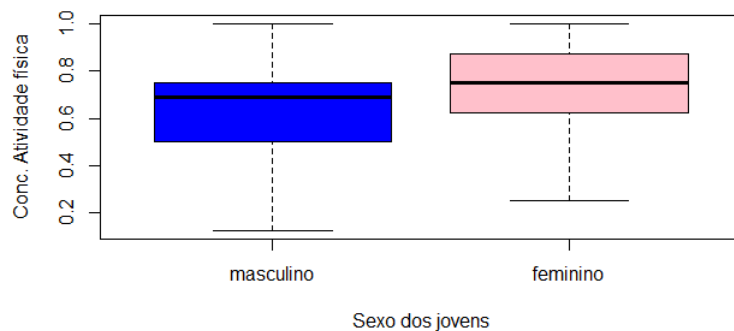


Figura 4.1: Concordância dos hábitos de atividade física dos jovens em função do Sexo dos jovens

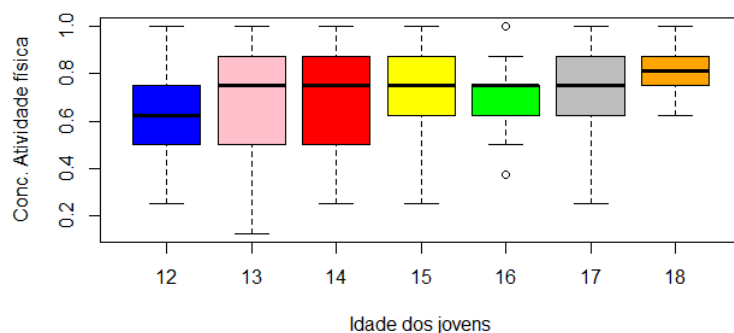


Figura 4.2: Concordância dos hábitos de atividade física dos jovens em função da Idade dos jovens

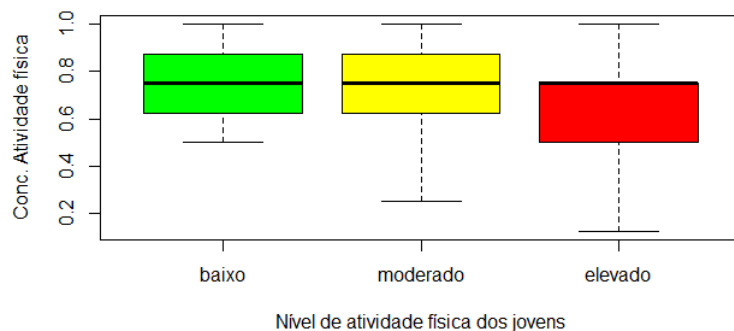


Figura 4.3: Concordância dos hábitos de atividade física dos jovens em função do Nível de atividade física dos jovens

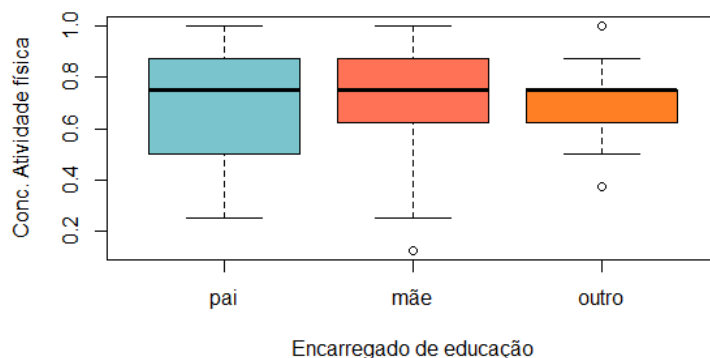


Figura 4.4: Concordância dos hábitos de atividade física dos jovens em função dos Encarregados de educação

mas difere entre os sexos dos jovens (valor-p = 0.0361), as idades (valor-p = 0.0268) e entre os diferentes níveis de atividade física praticados pelos jovens (valor-p = 0.0055), ocorrendo uma maior variação quando a atividade física é elevada (figura 4.3).

Para verificar de que modo a concordância da atividade física dos jovens varia de acordo com a **Idade dos EE**, é apresentada na Figura 4.5 um gráfico de dispersão da concordância em função da idade dos EE.

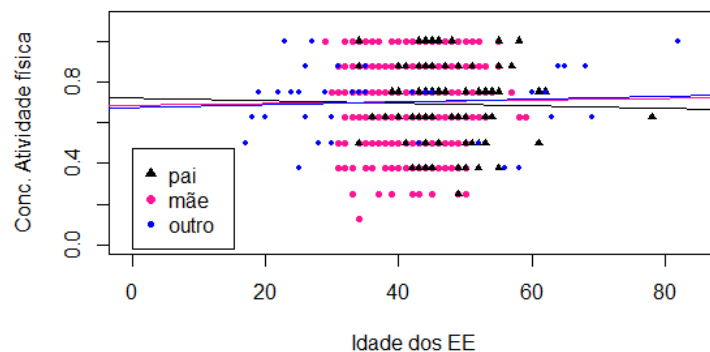


Figura 4.5: Concordância dos hábitos de atividade física dos jovens em função da Idade dos encarregados de educação

Ao observar a figura 4.5, não é possível perceber se a tendência de distribuição dos pontos de concordância em função da idade dos diferentes encarregados de educação (pai, mãe e outro) indica algum tipo de associação entre as duas variáveis. No entanto, ao olhar para os declives das retas presentes no gráfico, estes parecem indicar uma possível associação entre a concordância e a idade dos EE. Ao aplicar o coeficiente de correlação de *Spearman*, o resultado obtido ( $r_s = -0.0118$ , valor-p = 0.7658) demonstra a inexistência de associação entre a concordância e a idade dos EE, mostrando que a percepção parental relativamente aos hábitos de atividade física dos jovens não se altera com a idade dos EE.

#### 4.2.4 Perceção dos hábitos sedentários

Na análise da percepção dos encarregados de educação em relação aos hábitos sedentários dos jovens, foram escolhidas as seguintes variáveis:

- Jovens: **Sexo e Idade (em anos)**
- Encarregados de educação: **Encarregado de educação e Idade**

A análise da diferença entre os valores da concordância em função do **Sexo** foi obtida a partir do teste de *Mann-Whitney-Wilcoxon*. Em relação às variáveis **Idade (jovens)** e **Encarregado de educação** foi utilizado o teste de *Kruskal-Wallis*. Os gráficos correspondentes às funções de distribuição empírica da concordância dos hábitos sedentários em função das variáveis **Sexo, Idade (jovens)** e **Encarregado de educação** estão apresentados nas Figuras 4.6 a 4.8.

As funções de distribuição empírica correspondentes à concordância dos hábitos sedentários em função das variáveis em estudo permitem perceber se há diferenças entre a forma como os encarregados de educação percebem os hábitos sedentários dos adolescentes em função do sexo (figura 4.6) e idade dos jovens (figura 4.7), e se há diferenças quanto à forma como cada um dos EE percebe esses hábitos (figura 4.8). É possível verificar, para cada um dos gráficos, de que não há diferenças visíveis entre as percepções relativamente ao sexo dos jovens, a idade dos jovens nem entre os encarregados de educação. Os valores-p obtidos para as variáveis Sexo (valor-p = 0.0954), Idade dos jovens (valor-p = 0.3949) e

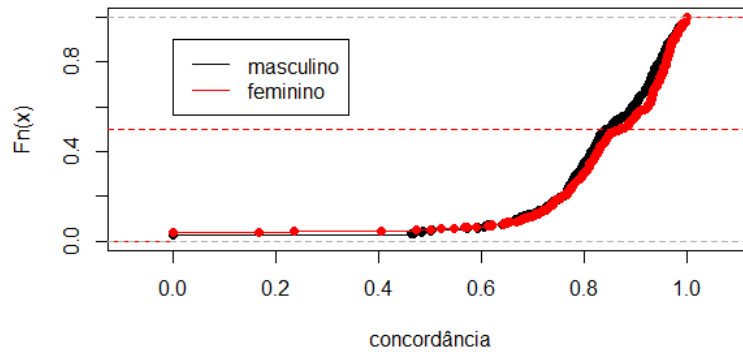


Figura 4.6: Concordância dos hábitos sedentários dos jovens em função do Sexo dos jovens

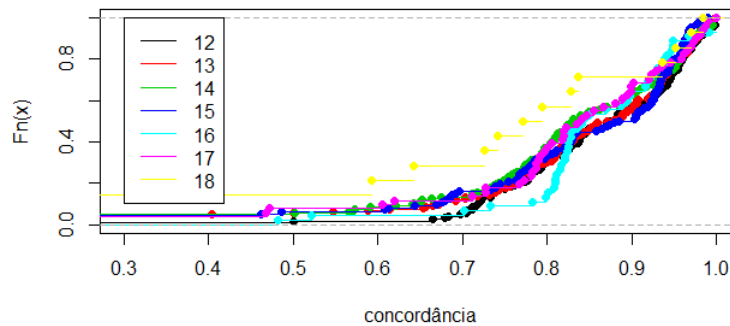


Figura 4.7: Concordância dos hábitos sedentários dos jovens em função da Idade dos jovens

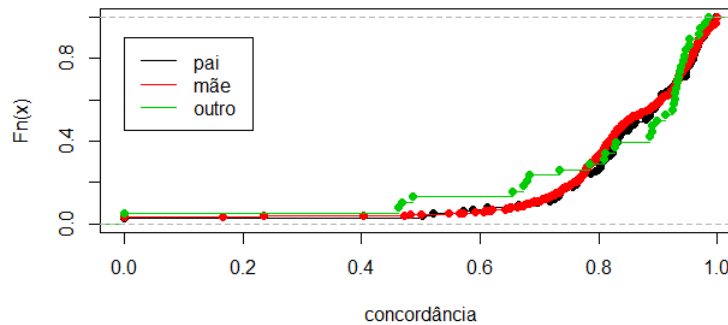


Figura 4.8: Concordância dos hábitos sedentários dos jovens em função dos Encarregados de educação

Encarregado de educação (valor-p = 0.8104), comprovam que a percepção parental não difere significativamente consoante o sexo e a idade dos jovens nem entre os próprios encarregados de educação.

Relativamente à variável **Idade (EE)**, o objetivo foi verificar de que modo a concordância dos hábitos sedentários varia de acordo com ela, através da análise de um gráfico de dispersão apresentado na Figura 4.9.

Como se pode verificar na figura 4.9, o padrão de distribuição dos pontos de concordância em função da idade de cada um dos encarregados de educação (pai, mãe e outro) não segue nenhuma tendência crescente ou decrescente óbvia. No entanto, os declives das retas sugerem uma associação entre as duas variáveis. De modo a verificar se existe associação entre a concordância e a idade dos encarregados de educação, foi calculado o coeficiente de *Spearman* e o respetivo valor-p. A partir do resultado obtido ( $r_s = 0.0899$ , valor-p = 0.0226), comprova-se que existe associação entre a concordância e a idade dos EE,

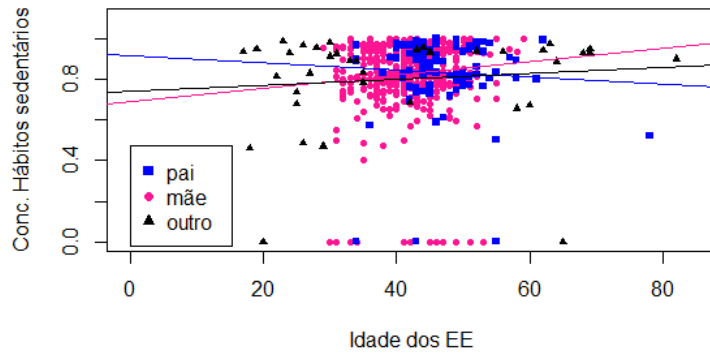


Figura 4.9: Concordância dos hábitos sedentários dos jovens em função da Idade dos encarregados de educação

indicando que a percepção parental dos hábitos sedentários dos jovens altera-se à medida que a idade dos EE aumenta.

#### 4.2.5 Percepção global

Na análise da percepção global, foram escolhidas as seguintes variáveis:

- Jovens: **Sexo** e **Idade (em anos)**
- Encarregados de educação: **Encarregado de educação** e **Idade**

A análise da diferença entre os valores da concordância em função do **Sexo** foi obtida a partir do teste de *Mann-Whitney-Wilcoxon*. Em relação às variáveis **Idade (jovens)** e **Encarregado de educação**, foi utilizado o teste de *Kruskal-Wallis*. Os gráficos da variação dos valores de concordância para as variáveis **Sexo**, **Idade (jovens)** e **Encarregado de educação** estão apresentados nas Figuras 4.10 a 4.12.

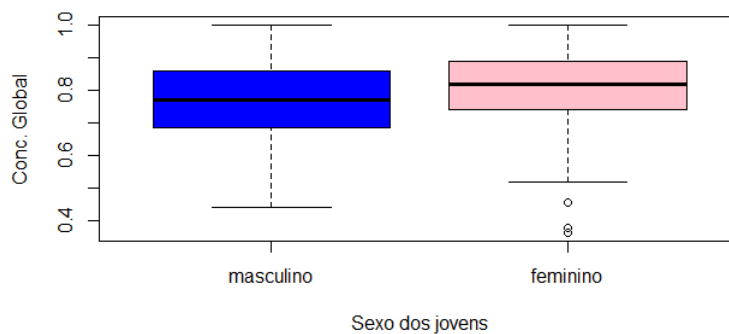


Figura 4.10: Concordância global em função do Sexo dos jovens

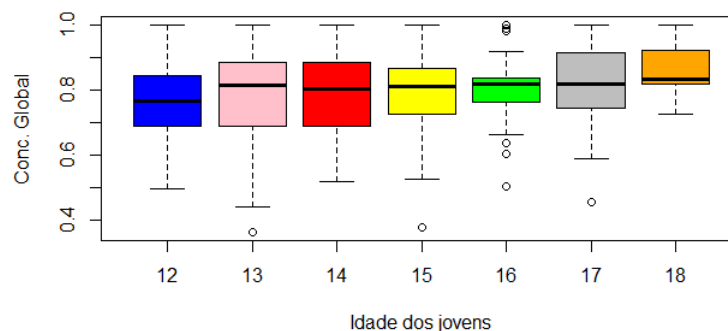


Figura 4.11: Concordância global em função da Idade dos jovens

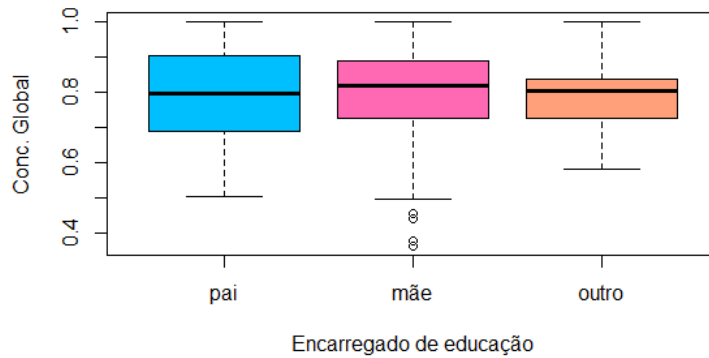


Figura 4.12: Concordância global em função dos Encarregados de educação

Na figura 4.10 é possível observar que, ao nível da mediana, a concordância global difere entre os sexos masculino e feminino (valor-p = 0.039), sendo a concordância superior em relação ao sexo feminino. Isto significa que 50% dos adultos responsáveis pelos jovens tem uma perceção mais correta em relação ao peso e estilos de vida das raparigas do que dos rapazes. Nas figuras 4.11 e 4.12, não se observam diferenças significativas entre as perceções consoante a Idade dos jovens (valor-p = 0.2195) e os Encarregados de educação (valor-p = 0.6998).

De modo a verificar se a concordância global e a **Idade dos EE** estão correlacionadas, foi construído um gráfico de dispersão, que se encontra apresentado na Figura 4.13.

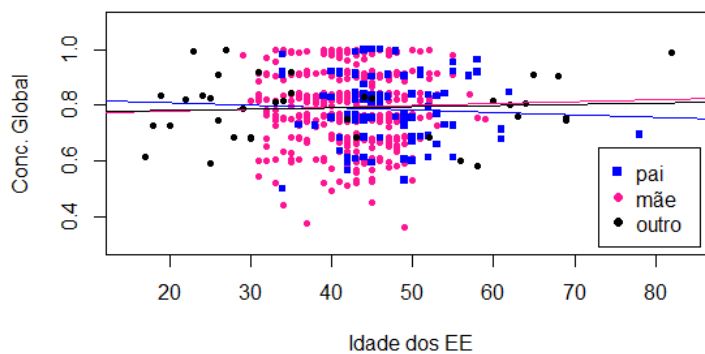


Figura 4.13: Concordância global em função da Idade dos encarregados de educação

Como se pode verificar na figura 4.13, não é possível observar uma tendência óbvia na distribuição dos pontos de concordância em função da idade dos encarregados de educação. No entanto, os declives das retas parecem sugerir uma possível associação entre as duas variáveis. Para verificar a existência de associação entre a concordância e a idade dos encarregados de educação, foi calculado o coeficiente de *Spearman* e o respetivo valor-p. A partir do resultado obtido ( $r_s = 0.0086$ , valor-p = 0.8273), verifica-se que não existe associação entre as variáveis, pelo que se pode concluir que a perceção global não se altera com a idade dos EE.

De forma a complementar e confirmar alguns dos resultados obtidos anteriormente relativamente à forma como a concordância entre as respostas dos jovens e dos EE a questões sobre o peso, hábitos alimentares, hábitos de atividade física e hábitos sedentários dos jovens varia entre indivíduos com diferentes características (ex: entre jovens magros, com peso normal, com excesso de peso e com obesidade), foram utilizadas faces de Chernoff. Estas faces são constituídas por diversos elementos (ex., olhos) e estes ele-

mentos vão ter diferentes tamanhos ou formas consoante os valores que os respetivos parâmetros tomam (ex., os olhos são desenhados a partir dos valores dos parâmetros altura e largura).

Neste caso, os parâmetros vão tomar como valores as médias das concordâncias relativamente às questões sobre o peso e hábitos, i.e.:

- Altura da face: média da concordância global
- Largura da face: média da concordância do peso
- Estrutura da face: média da concordância da atividade física
- Altura da boca: média da concordância da questão sobre o pequeno-almoço (hábitos alimentares)
- Largura da boca: média da concordância da questão sobre a presença de frutas/vegetais em casa (hábitos alimentares)
- Curva do sorriso: média da concordância da questão sobre o tempo gasto ao computador/jogos eletrónicos (hábitos sedentários)
- Altura dos olhos: média da concordância da questão sobre o tempo gasto a ver televisão (hábitos sedentários)
- Largura dos olhos: média da concordância da questão sobre se o jovem vê televisão às refeições (hábitos sedentários)
- Altura do cabelo: média da concordância global
- Largura do cabelo: média da concordância do peso
- Estilo do cabelo: média da concordância da atividade física
- Altura do nariz: média da concordância da questão sobre o pequeno-almoço (hábitos alimentares)
- Largura do nariz: média da concordância da questão sobre a presença de frutas/vegetais em casa (hábitos alimentares)
- Largura das orelhas: média da concordância da questão sobre o tempo gasto ao computador/jogos eletrónicos (hábitos sedentários)
- Altura das orelhas: média da concordância da questão sobre o tempo gasto a ver televisão (hábitos sedentários)

Nas Figuras 4.14 a 4.16 estão apresentadas as faces correspondentes aos jovens do sexo masculino e e feminino, aos encarregados de educação e aos jovens com diferentes pesos, respetivamente.



Figura 4.14: Faces de Chernoff relativamente ao Sexo dos jovens

Relativamente ao sexo dos jovens (figura 4.14) é possível verificar que a largura da face e a largura do cabelo são superiores no sexo masculino, i.e., a média da concordância relativamente ao peso dos jovens é superior no sexo masculino, o que comprova o resultado já obtido na Secção 4.2.1 (Tabela 4.7) de que os encarregados de educação percecionam melhor o peso nos rapazes do que nas raparigas. Outras diferenças evidentes que se podem observar entre os jovens são o nariz e a boca. Em relação à altura do nariz e à altura da boca, é possível observar que são mais baixas no sexo feminino, indicando que os



Figura 4.15: Faces de Chernoff relativamente aos Encarregados de educação



Figura 4.16: Faces de Chernoff relativamente à Classe de peso dos jovens

encarregados de educação têm uma menor percepção quanto ao número de vezes que as raparigas tomam o pequeno-almoço. A questão do pequeno-almoço, que está incluída nos hábitos alimentares, confirma em parte o resultado obtido em relação à concordância dos hábitos alimentares dos jovens (Tabela 4.9), de que esta é mais baixa para as raparigas.

Entre os diferentes encarregados de educação (figura 4.15) é possível observar que a face correspondente às mães é a única com a boca em forma de sorriso e com as orelhas largas. Este resultado demonstra que a média da concordância relativamente ao tempo que os jovens gastam ao computador é superior nas mães, indicando que estas têm uma melhor percepção quanto ao número de horas que os jovens gastam ao computador/jogos eletrónicos. Quanto à estrutura da face e à estrutura do cabelo, estas são muito semelhantes entre as mães e os outros encarregados de educação, o que revela que estes EE têm uma percepção parecida em relação aos hábitos de atividade física dos jovens.

Na figura 4.16, que corresponde à classe de peso dos jovens, é possível verificar que a largura da face e a largura do cabelo são inferiores nos jovens que sofrem de obesidade, o que confirma o resultado obtido na Tabela 4.7 de que os encarregados de educação têm uma percepção mais baixa em relação ao peso de jovens obesos. Relativamente à altura dos olhos e à altura das orelhas, estas são superiores nos jovens com obesidade, revelando que os encarregados de educação têm uma melhor percepção do tempo gasto a ver televisão pelos jovens obesos do que pelos jovens com um peso mais baixo.

### 4.3 Análise da concordância parental: aplicação de modelos de regressão

Com o objetivo de saber que características sociodemográficas dos jovens e dos respetivos encarregados de educação influenciam significativamente a percepção parental em relação ao peso e estilos de vida dos adolescentes, foram aplicados diferentes modelos de regressão consoante os valores da variável resposta, i.e., da percepção.

Durante o processo de seleção de modelos, para que o melhor modelo seja selecionado, é importante que a comparação seja feita entre modelos construídos sobre a mesma amostra. Como algumas das variáveis em estudo (características sociodemográficas) contêm valores omissos, foi necessário selecionar os casos para os quais a percepção parental (variável resposta) e as características sociodemográficas (variáveis explicativas) não apresentassem valores omissos. Deste modo, os modelos ajustados para cada uma das percepções têm como base amostras de diferente dimensão.

### 4.3.1 Percepção do peso

A concordância entre as respostas dos jovens e dos encarregados de educação em relação à questão do peso dos jovens assume valores entre 0 e 1. Teoricamente, o modelo de regressão a aplicar seria o modelo de regressão beta porque, dos modelos apresentados na Secção 3.3, é o que se adequa, dado que a variável de interesse toma valores no intervalo (0,1). No entanto, devido ao facto de a percepção do peso ter sido calculada com base numa só questão, os valores resultantes foram pouco diversos, pelo que foi decidido transformar a concordância do peso numa variável dicotómica, i.e., valor 1 se a concordância é completa e valor 0 se é incompleta. Os valores iniciais da concordância e os valores resultantes da categorização encontram-se apresentados na Tabela 4.10.

Tabela 4.10: Valores iniciais e categorização dos valores da concordância do peso

Valores iniciais	<i>n</i> (%) <sup>*</sup>
0	1 (0,2%)
0.1979	5 (0,8%)
0.8013	5 (0,8%)
0.9992	171 (25,8%)
1	478 (72,0%)
Categorização	<i>n</i> (%)
0	182 (27,4%)
1	478 (72,0%)

<sup>\*</sup>Resultados obtidos de uma amostra de 664 valores de concordância

Correspondendo o valor 0 a uma percepção incorreta do peso dos jovens e o valor 1 a uma percepção correta, verifica-se que 27,4% dos encarregados de educação não percebem corretamente o peso dos jovens (Tabela 4.10).

Como a concordância em relação ao peso dos jovens assume novos valores, 0 ou 1, o modelo de regressão a aplicar é o modelo de regressão logística binária.

Para saber que características (sociodemográficas) dos jovens e dos encarregados de educação podem estar associadas significativamente com uma percepção incorreta do peso dos jovens, são analisadas as seguintes variáveis:

- Jovens: **Sexo, Idade exata e Classe de peso**
- Encarregados de educação: **Encarregado de educação, Idade, Escolaridade, Classe de peso, Preocupação com o peso do educando**

### 4.3.1.1 Seleção do modelo

Com o objetivo de construir o modelo de regressão logística, foi utilizado o método de seleção *Stepwise* descrito na Secção 3.3.1.4.

A partir dos três métodos de seleção aplicados, *backward*, *forward* e *both*, obteve-se o mesmo modelo final com valor AIC = 572.3243 e constituído pelas seguintes variáveis:

1. **Sexo**
2. **Idade exata**
3. **Classe de peso (jovens)**

Os resultados das estimativas dos coeficientes do modelo, da estatística de teste de *Wald* e os respetivos valores-p estão apresentados na Tabela 4.11.

Tabela 4.11: Modelo de regressão logística binária da perceção do peso

Variáveis	Coefficiente	Wald	valor-p
Interseção	3.7796	3.665	$2.47e^{-4}$
Sexo			
Feminino (classe de referência)			
Masculino	-0.4606	-2.170	0.03
Idade exata	-0.1283	-1.861	0.0628
Classe de peso			
Peso normal (classe de referência)			
Magro	-1.9167	-4.501	$6.75e^{-6}$
Excesso de peso	-2.1346	-9.108	$< 2e^{-16}$
Obeso	-4.0053	-5.254	$1.49e^{-7}$

Resultados obtidos de uma amostra de 581 casos

Como se pode observar na Tabela 4.11, os coeficientes associados às categorias das variáveis Sexo e Classe de peso são significativos ao nível de significância  $\alpha = 0.05$ , mas não o que corresponde à variável Idade exata dos jovens. No entanto, como o valor-p = 0.0628 está próximo do  $\alpha$  estabelecido, indica que há uma tendência de que a concordância do peso seja mais baixa há medida que a idade dos jovens aumenta.

No sentido de comparar o modelo com as variáveis selecionadas através do método *Stepwise* e o modelo composto por todas as variáveis, foi aplicado o teste de razão de verosimilhanças descrito na Secção 3.3.1.3. O valor-p obtido foi 0.3454, não sendo significativo ao nível de significância  $\alpha = 0.05$ . Este resultado indica que as variáveis que não foram incluídas no modelo não produzem um acréscimo significativo na sua qualidade, ou seja, as únicas variáveis a manter no modelo são Sexo, Idade exata e Classe de peso dos jovens.

Como a variável Idade exata dos jovens é uma variável contínua, para analisar a linearidade existente entre ela e a variável resposta na escala *logit*, está apresentado na Figura 4.17 a representação gráfica correspondente ao logaritmo da *odds* da concordância do peso em função da idade dos jovens.

Como se pode observar na figura 4.17, a distribuição dos valores *logit* da concordância do peso em função da idade dos jovens não é linear, o que sugere a não existência de linearidade na escala *logit* entre as duas variáveis.

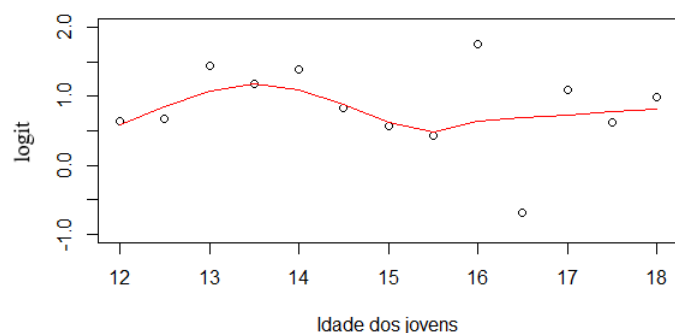


Figura 4.17: Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância do peso na escala *logit*

De modo a verificar a linearidade foi também aplicado o método de comparação de dois modelos lineares cuja variável resposta é a mesma nos dois modelos, i.e., *logit* da concordância do peso. O que difere entre os dois modelos é o preditor linear, ou seja, enquanto um modelo tem como variável explicativa a idade dos jovens, o outro modelo tem como variáveis explicativas a idade dos jovens e a idade dos jovens ao quadrado. O resultado obtido da comparação dos modelos ( $F = 10.2$ ,  $\text{valor-p} = 0.0015$ ) permite concluir que a idade dos jovens ao quadrado é significativa para o modelo, o que comprova a não existência de linearidade na escala *logit* entre a variável resposta e a idade exata dos jovens.

Dado que não se verificou a condição de linearidade entre as variáveis, foi necessário categorizar a variável Idade exata dos jovens. Deste modo, a variável Idade (em anos) dos jovens (i.e., 12, 13, ..., 18 anos) foi considerada como a variável categorizada da Idade exata dos jovens. No entanto, ao ajustar o modelo com a variável categorizada, esta deixou de ser importante em termos estatísticos.

O modelo final que vai ser considerado é o modelo constituído pelas variáveis Sexo e Classe de peso dos jovens, cujos resultados estão presentes na Tabela 4.12.

Tabela 4.12: Modelo de regressão logística binária da percepção do peso com as variáveis Sexo e Classe de peso dos jovens

Variáveis	Coefficiente	Wald	valor-p
Interseção	1.9003	11.124	$< 2e^{-16}$
Sexo			
Feminino (classe de referência)			
Masculino	-0.4489	-2.123	0.0338
Classe de peso			
Peso normal (classe de referência)			
Magro	-2.0073	-4.755	$1.98e^{-6}$
Excesso de peso	-2.0935	-9.048	$< 2e^{-16}$
Obeso	-3.9369	-5.180	$2.21e^{-7}$

Resultados obtidos de uma amostra de 581 casos

#### 4.3.1.2 Análise da qualidade de ajustamento do modelo

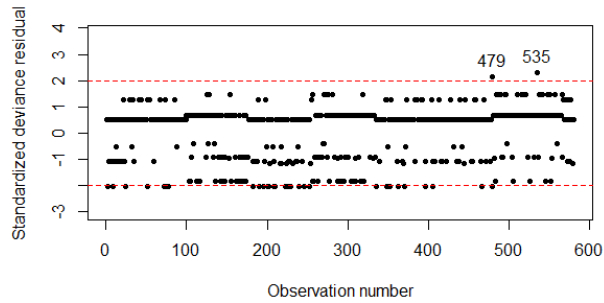
Para verificar a qualidade de ajustamento do modelo escolhido, foram aplicados a medida *Deviance* e o teste de *Hosmer e Lemeshow*.

A *Deviance* permite verificar se o modelo escolhido ajusta-se adequadamente aos dados, ao compará-lo com o modelo saturado. Quanto menor o desvio entre os dois modelos, melhor será o ajustamento do modelo aos dados observados. A partir do resultado obtido ( $D = 563.7806$ ,  $\text{valor-p} = 0.6342$ ) comprova-

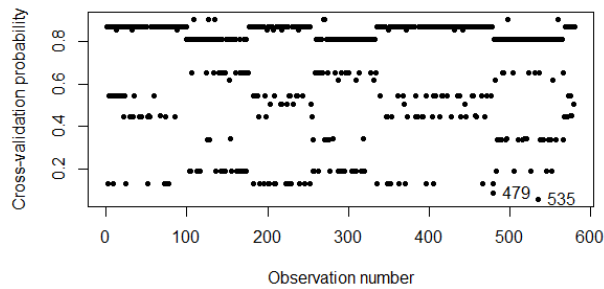
se que o modelo é adequado, o que indica um bom ajustamento do modelo.

Em relação ao resultado obtido do teste de *Hosmer e Lemeshow* ( $X_{HL}^2 = 0.9607$ , valor-p = 0.9985) é possível chegar à mesma conclusão de que o modelo escolhido se ajusta bem aos dados.

Para verificar a qualidade de ajustamento do modelo e identificar as observações consideradas *outliers* ou influentes, estão representadas nas Figuras 4.18 e 4.19 os gráficos correspondentes.



(a) Gráfico dos resíduos padronizados contra a ordem das observações



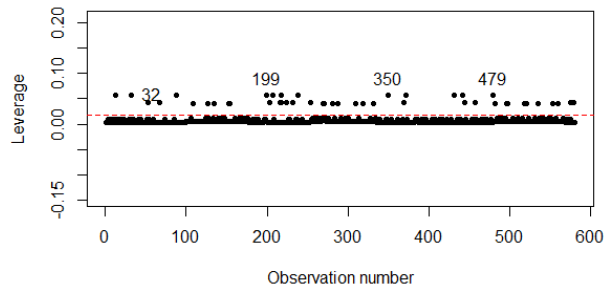
(b) Gráfico das probabilidades obtidas por validação cruzada

Figura 4.18: Análise do preditor linear e identificação de observações *outliers* do modelo da percepção do peso

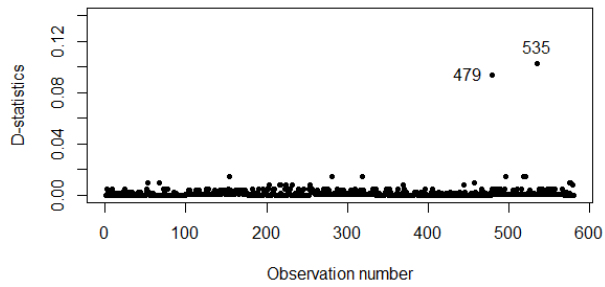
O gráfico dos resíduos padronizados contra a ordem das observações e o gráfico das probabilidades obtidas por validação cruzada, correspondentes às figuras 4.18 (a) e 4.18 (b), permitem a deteção de observações *outliers*. Como se pode verificar na figura 4.18 (a), existem observações cujos resíduos se encontram fora do intervalo estabelecido  $[-2, 2]$ , o que as indicia como possíveis candidatas a *outliers*. Na figura 4.18 (b), as observações cujos resíduos tenham uma probabilidade reduzida de ocorrerem são consideradas *outliers*. As observações 479 e 535 podem ser consideradas candidatas a *outliers* devido aos seus valores de probabilidade serem inferiores a 10% ( $P_{479} = 0.08873$ ,  $P_{535} = 0.05822$ ).

O gráfico dos resíduos padronizados contra a ordem das observações (figura 4.18 (a)) permite também avaliar a qualidade de ajustamento do modelo, através do padrão da distribuição dos resíduos. A distribuição dos resíduos dentro do intervalo não segue nenhum padrão óbvio, o que indica um bom ajustamento do modelo aos dados.

Para a identificação de observações influentes no modelo são analisadas as figuras 4.19 (a) e 4.19 (b). No gráfico de *leverage* (figura 4.19 (a)), as observações são consideradas influentes se os seus valores  $h_{ii}$  são superiores ao valor crítico  $\frac{2p}{n} = 0.01721$ . Como se pode verificar, há muitas observações cujos  $h_{ii}$  são superiores ao valor crítico, destacando-se algumas observações como a 32, 199, 350 e 479, cujos valores  $h_{ii}$  são 0.05740.



(a) Gráfico de *leverage*



(b) Gráfico da estatística D

Figura 4.19: Identificação de observações influentes do modelo da percepção do peso

O gráfico da estatística D (figura 4.19 (b)) permite uma melhor identificação de possíveis observações influentes. As observações 479 e 535 destacam-se, devido aos seus valores  $DC_i$  serem superiores aos das restantes observações ( $DC_{479} = 0.09335$ ,  $DC_{535} = 0.10284$ ).

Na Tabela 4.13 apresentam-se os resultados do modelo ajustado sem as observações 32, 199, 350, 479 e 535.

Tabela 4.13: Modelo de regressão logística binária da percepção do peso sem as observações 32, 199, 350, 479 e 535

Variáveis	Coefficiente	Wald	valor-p
Interseção	1.9003	11.124	$< 2e^{-16}$
Sexo			
Feminino (classe de referência)			
Masculino	-0.4489	-2.123	0.0338
Classe de peso			
Peso normal (classe de referência)			
Magro	-2.0073	-4.755	$1.98e^{-6}$
Excesso de peso	-2.0935	-9.048	$< 2e^{-16}$
Obeso	-3.9369	-5.180	$2.21e^{-7}$

As variações encontradas relativamente aos coeficientes e à estatística de *Wald* das variáveis foram nulas, pelo que todas as variáveis que foram significativas no modelo escolhido continuaram a ser significativas no modelo sem as observações.

Para verificar se a função de ligação escolhida é adequada, foi adicionada ao modelo uma nova variável explicativa, *Z*, e avaliado o declínio na função desvio. O declínio não foi significativo (valor-p = 0.2293), pelo que a função de ligação *logit* é considerada adequada.

### 4.3.1.3 Capacidade preditiva do modelo

Quando existe a possibilidade de utilizar o modelo para fazer previsão, torna-se importante avaliar a sua capacidade preditiva.

Para a avaliação da capacidade preditiva recorreu-se ao estudo das medidas de desempenho através da matriz de confusão e da área sob a curva ROC.

O melhor ponto de corte corresponde a 0.6722321, estando na Tabela 4.14 os resultados possíveis da classificação dos indivíduos em cada classe da variável resposta ( $Y = 0$  e  $Y = 1$ ).

Tabela 4.14: Tabela de classificação para o modelo da percepção do peso

Classificação	Valores observados		Total
	Classe(0)	Classe(1)	
Classe(0)	105	62	167
Classe(1)	64	350	414
Total	169	412	581

Através da Tabela 4.14 é possível estimar as medidas de desempenho, tendo sido obtidos os valores 84,95% para a sensibilidade e 62,13% para a especificidade. Para a percentagem total de acertos, i.e., a proporção de indivíduos corretamente classificados, o valor obtido foi 78,31%. Estes valores mostram que o modelo tem uma boa capacidade em avaliar os verdadeiros positivos.

A área abaixo da curva ROC varia entre 0 e 1 e classifica de forma mais precisa a capacidade do modelo em discriminar os valores da variável resposta  $Y = 1$  dos valores da variável  $Y = 0$ . O gráfico correspondente à curva ROC está apresentado na Figura 4.20.

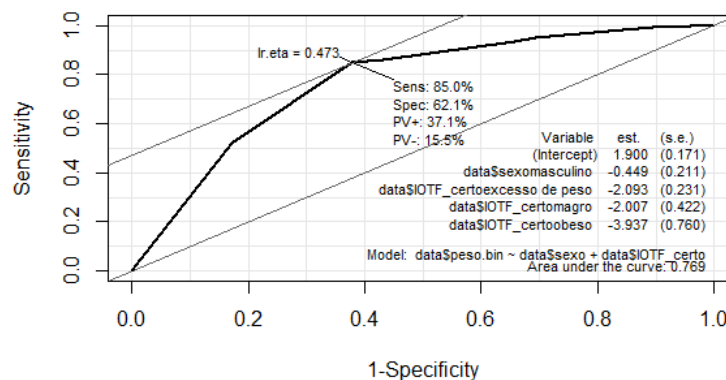


Figura 4.20: Curva ROC do modelo da percepção do peso

O valor obtido da área abaixo da curva foi 0.769, que indica uma discriminação aceitável de acordo com os critérios de avaliação propostos por Hosmer & Lemeshow.

### 4.3.1.4 Interpretação do modelo

Para melhor analisar o modelo em termos de interpretação dos coeficientes, a Tabela 4.15 apresenta os valores de *odds ratio* obtidos para cada uma das variáveis explicativas incluídas no modelo e as respetivas interpretações.

Tabela 4.15: Interpretação das variáveis significativas para o modelo da percepção do peso dos jovens

Variáveis	OR	IC <sub>95%</sub>	Interpretação OR
Sexo: Masculino	0.64	(0.42 ; 0.97)	A chance da percepção parental do peso dos jovens ser correta diminui cerca de 36% se o adolescente for do sexo masculino
Classe de peso: Magro	0.13	(0.06 ; 0.31)	A chance da percepção parental do peso dos jovens ser correta diminui cerca de 87% se o adolescente for magro
Classe de peso: Excesso de peso	0.12	(0.08 ; 0.19)	A chance da percepção parental do peso dos jovens ser correta diminui cerca de 88% se o adolescente tiver excesso de peso
Classe de peso: Obeso	0.02	(0.003 ; 0.07)	A chance da percepção parental do peso dos jovens ser correta diminui cerca de 98% se o adolescente for obeso

### 4.3.2 Percepção dos hábitos alimentares

Tal como os valores de concordância obtidos para a percepção do peso, os valores correspondentes à percepção dos hábitos alimentares são pouco diversos. Deste modo, optou-se por transformar a variável relativa à concordância dos hábitos alimentares numa nova variável. Neste caso, a nova variável vai ser do tipo categórico ordinal, sendo o modelo de regressão a aplicar o modelo de regressão logística ordinal descrito na Secção 3.3.2.

Como os hábitos alimentares são compostos por duas questões, a nova variável vai ser construída do seguinte modo:

- Se as respostas dos EE e dos jovens forem diferentes para as duas questões, a variável assume valor 0;
- Se as respostas dos EE e dos jovens forem diferentes para uma das duas questões, a variável assume valor 1;
- Se as respostas dos EE e dos jovens forem iguais para as duas questões, a variável assume valor 2.

O valor 0 significa falta de concordância, valor 1 concordância moderada e valor 2 concordância total. Os resultados obtidos encontram-se na Tabela 4.16.

Tabela 4.16: Categorização dos valores da concordância dos hábitos alimentares dos jovens

Valores	n (%)*
0	41 (6,2%)
1	142 (21,4%)
2	449 (67,6%)

\*Resultados obtidos de uma amostra de 664 valores de concordância

Pelos valores obtidos na nova variável (Tabela 4.16), é possível observar que 67,6% dos encarregados de educação percecionam corretamente os hábitos alimentares dos jovens.

Para saber que características (sociodemográficas) dos jovens e dos encarregados de educação podem estar associadas significativamente com uma percepção parental mais baixa em relação aos hábitos alimentares dos jovens, são analisadas as seguintes variáveis:

- Jovens: **Sexo e Idade exata**
- Encarregados de educação: **Encarregado de educação e Idade**

#### 4.3.2.1 Seleção do modelo

Com o objetivo de construir o modelo de regressão logística, foi utilizado o método de seleção Stepwise. A partir dos três métodos de seleção aplicados, obteve-se o mesmo modelo final (AIC = 905.8146) constituído pelas seguintes variáveis:

1. **Sexo**
2. **Idade exata (jovens)**
3. **Encarregado de educação**
4. **Idade (EE)**

Os resultados obtidos estão apresentados na Tabela 4.17.

Tabela 4.17: Modelo de regressão logística ordinal da percepção dos hábitos alimentares

Variáveis	Coefficiente	Wald	valor-p
logit 1	-3.9503	-4.0405	$5.33e^{-5}$
logit 2	-2.1367	-2.2147	0.0268
Sexo			
Feminino (classe de referência)			
Masculino	0.4551	2.4059	0.0161
Idade exata	-0.2043	-3.5901	0.0003
Encarregado de educação			
Mãe (classe de referência)			
Pai	0.2748	0.9499	0.3422
Outro	-0.6847	-1.8770	0.0605
Idade (EE)	0.0371	2.8253	0.0047

Resultados obtidos de uma amostra de 613 casos

Como se pode observar na Tabela 4.17, os coeficientes associados às variáveis Sexo, Idade exata dos jovens e Idade dos EE são significativos ao nível de significância  $\alpha = 0.05$ .

Como descrito na Secção 3.3.2.1, quando a variável resposta tem  $J$  categorias, o modelo de *odds* proporcionais será construído para cada uma das  $J - 1$  categorias. Neste caso, a variável correspondente à percepção dos hábitos alimentares dos jovens tem 3 categorias, pelo que serão construídos 2 modelos. Os modelos são:

$$\text{Modelo 1 : } g(P_0) = \ln \left[ \frac{\Pr((Y \leq 0)|X)}{\Pr((Y > 0)|X)} \right] = \theta_0 - X^T \beta$$

$$\text{Modelo 2 : } g(P_1) = \ln \left[ \frac{\Pr((Y \leq 1)|X)}{\Pr((Y > 1)|X)} \right] = \theta_1 - X^T \beta$$

O modelo 1 compara a concordância tomar valor 0 (percepção incorreta) versus tomar valor 1 (percepção correta parcialmente) ou valor 2 (percepção correta total).

O modelo 2 compara a concordância tomar valor 0 ou valor 1 versus tomar valor 2.

Tratando-se de modelos de *odds* proporcionais, os coeficientes associados às variáveis são iguais em ambos os modelos, sendo a única diferença os valores de  $\theta$ , que correspondem aos valores *logit* na Tabela 4.17. Assim, o *logit* 1 corresponde ao  $\theta_0$  do modelo 1 e o *logit* 2 ao  $\theta_1$  do modelo 2.

O modelo selecionado através do método *Stepwise* é constituído por todas as variáveis escolhidas previamente, pelo que não foi necessário efetuar o teste da razão de verosimilhanças, devido ao facto de este ser utilizado para comparar modelos aninhados.

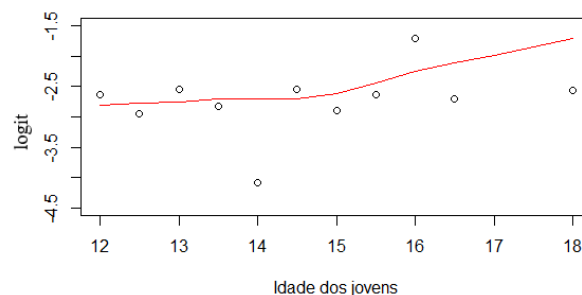
Como a Idade exata dos jovens e a Idade dos encarregados de educação são ambas variáveis contínuas, foi necessário verificar a existência de linearidade na escala *logit* para cada uma das variáveis. No entanto, antes de se proceder à verificação da linearidade, foi necessário construir dois modelos binários a partir do modelo ordinal, i.e., um primeiro modelo em que a variável resposta assume os valores:

$$\tilde{Y}_0 = \begin{cases} 1 & \text{se } Y \leq 0, \\ 0 & \text{se } Y > 0. \end{cases}$$

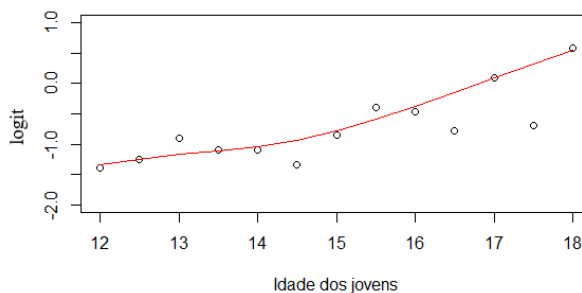
E um segundo modelo em que a variável resposta assume os valores:

$$\tilde{Y}_1 = \begin{cases} 1 & \text{se } Y \leq 1, \\ 0 & \text{se } Y > 1. \end{cases}$$

Finalizado este passo, foram construídas representações gráficas dos valores *logit* da variável resposta (i.e., concordância dos hábitos alimentares) para cada um dos modelos acima referidos em função de cada uma das variáveis contínuas. As representações gráficas estão apresentadas nas Figuras 4.21 e 4.22.



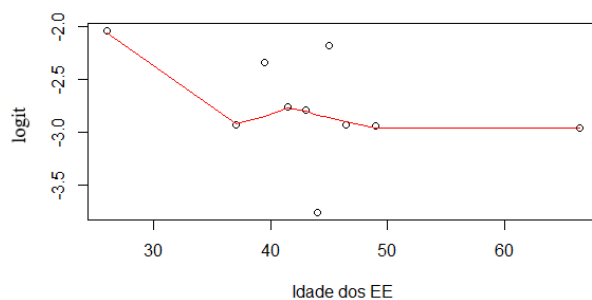
(a) Modelo 1



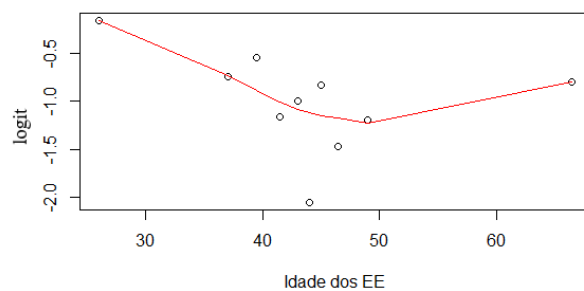
(b) Modelo 2

Figura 4.21: Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância dos hábitos alimentares na escala *logit*

Como se pode observar nas figuras 4.21 (a) e 4.21 (b), a relação entre os valores de *logit* da concordância



(a) Modelo 1



(b) Modelo 2

Figura 4.22: Representação gráfica da linearidade entre a variável Idade dos EE e a concordância dos hábitos alimentares na escala *logit*

dos hábitos alimentares e a idade dos jovens parece seguir uma tendência linear, sendo esta tendência mais evidente na figura 4.21 (b) correspondente ao modelo 2. Deste modo, as figuras 4.21 (a) e 4.21 (b) parecem sugerir a existência de linearidade na escala *logit* entre a concordância dos hábitos alimentares e a idade dos jovens. De modo a verificar a existência de linearidade foi também aplicado o método de comparação de modelos lineares. Os resultados obtidos para o modelo 1 ( $F = 74.419$ , valor- $p < 2.2e^{-16}$ ) e para o modelo 2 ( $F = 100.39$ , valor- $p < 2.2e^{-16}$ ) indicam que a variável idade dos jovens ao quadrado é significativa para o modelo, o que mostra a não existência de linearidade na escala *logit* entre a variável resposta e a idade dos jovens. No entanto, visto que as figuras 4.21 (a) e 4.21 (b) revelam linearidade entre as variáveis, admitiu-se a existência de linearidade para a idade dos jovens.

Relativamente à figura 4.22, em ambos os modelos, a relação entre *logit* da variável resposta e a idade dos encarregados de educação não parece seguir uma tendência linear, sugerindo a inexistência de linearidade na escala *logit* entre as duas variáveis. Neste caso foi também aplicado o método de comparação de modelos lineares. Os resultados obtidos para o modelo 1 ( $F = 105.23$ , valor- $p < 2.2e^{-16}$ ) e para o modelo 2 ( $F = 550.23$ , valor- $p < 2.2e^{-16}$ ) indicam que a variável idade dos EE ao quadrado é significativa para o modelo, o que comprova a inexistência de linearidade na escala *logit* entre a variável resposta e a idade dos EE. Dado que não se verificou a linearidade, a idade dos encarregados de educação foi categorizada em duas categorias, sendo uma referente aos indivíduos com idade  $< 50$  anos e a outra aos indivíduos com idade  $\geq 50$  anos. No entanto, ao ajustar o modelo com a variável Idade categorizada, esta deixou de ser importante em termos estatísticos.

O modelo final que vai ser considerado é o modelo constituído pelas variáveis Sexo, Idade exata dos jovens e Encarregado de educação, cujos resultados estão presentes na Tabela 4.18.

Tabela 4.18: Modelo de regressão logística ordinal da percepção dos hábitos alimentares sem a variável Idade dos encarregados de educação

Variáveis	Coefficiente	Wald	valor-p
logit 1	-5.2846	-6.1544	$7.5378e^{-10}$
logit 2	-3.4955	-4.1666	$3.0923e^{-5}$
Sexo			
Feminino (classe de referência)			
Masculino	0.4761	2.5347	$1.1255e^{-2}$
Idade exata			
	-0.1907	-3.3774	$7.3035e^{-4}$
Encarregado de educação			
Mãe (classe de referência)			
Pai	0.4312	1.5192	0.1287
Outro	-0.7874	-2.2303	0.0257

Tendo sido aplicado o modelo de *odds* proporcionais, é necessário verificar se o pressuposto de *odds* proporcionais é satisfeito. Para isso, são aplicados o teste de razão de verosimilhanças e o teste de *Brant*. Os valores-p obtidos foram 0.5008 para o teste de razão de verosimilhanças e 0.3429 para o teste de *Brant*. Ao nível de significância  $\alpha = 0.05$ , os valores-p não são significativos, pelo que não há evidências para duvidar do cumprimento do pressuposto.

#### 4.3.2.2 Análise da qualidade de ajustamento do modelo

Para verificar a qualidade de ajustamento do modelo composto pelas variáveis Sexo, Idade exata dos jovens e Encarregado de educação, são aplicados os testes *Lipsitz* e *Hosmer-Lemeshow* para modelos ordinais. Os valores-p obtidos para o teste de *Lipsitz* (valor-p = 0.9253) e para o teste de *Hosmer-Lemeshow* (valor-p = 0.6305) não são significativos, o que indica um bom ajustamento do modelo.

Para se proceder à análise de resíduos do modelo ordinal, é necessário construir dois modelos binários de acordo com a expressão 3.40. No primeiro modelo, a variável resposta assume os valores:

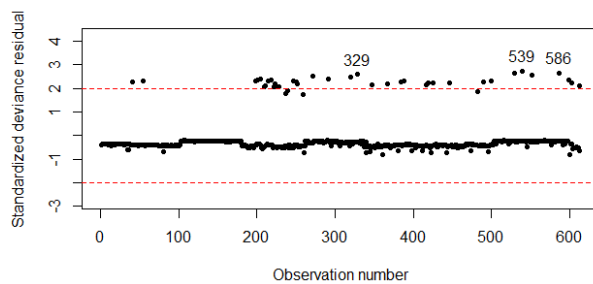
$$\tilde{Y}_0 = \begin{cases} 1 & \text{se } Y \leq 0, \\ 0 & \text{se } Y > 0. \end{cases}$$

E no segundo modelo, a variável resposta assume os valores:

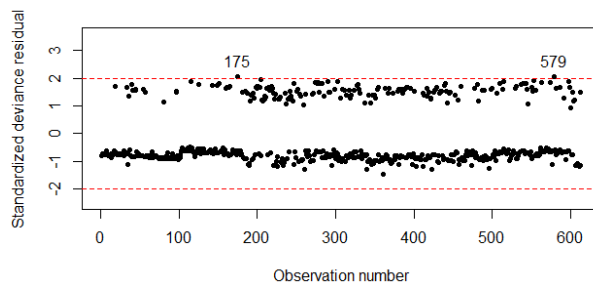
$$\tilde{Y}_1 = \begin{cases} 1 & \text{se } Y \leq 1, \\ 0 & \text{se } Y > 1. \end{cases}$$

Para identificar as observações consideradas *outliers* ou influentes, é realizada a análise de resíduos descrita na Secção 3.3.1.5 a cada um dos modelos referidos acima. Nas Figuras 4.23 a 4.26 estão representados os gráficos correspondentes.

O gráfico dos resíduos padronizados contra a ordem das observações, correspondente às figuras 4.23 (a) e 4.23 (b), permite a deteção de observações *outliers*. Como se pode verificar para o modelo 1 (figura 4.23 (a)), existem observações cujos resíduos se encontram fora do intervalo  $[-2,2]$ , o que as indicia como possíveis candidatas a *outliers*. Já no modelo 2 (figura 4.23 (b)), existem apenas duas observações cujos resíduos se encontram fora do intervalo, sendo essas observações diferentes das identificadas para

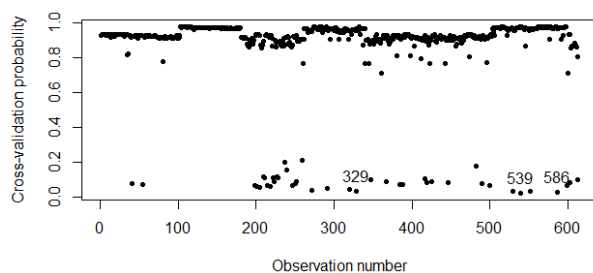


(a) Modelo 1

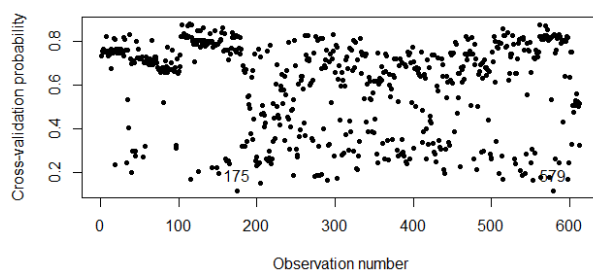


(b) Modelo 2

Figura 4.23: Gráfico dos resíduos padronizados contra a ordem das observações do modelo da percepção dos hábitos alimentares



(a) Modelo 1

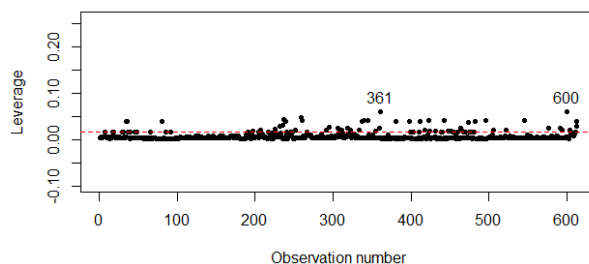


(b) Modelo 2

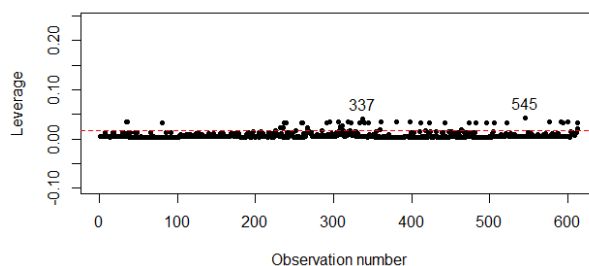
Figura 4.24: Gráfico das probabilidades obtidas por validação cruzada do modelo da percepção dos hábitos alimentares

o modelo 1. Em ambos os gráficos, a distribuição dos resíduos não segue nenhum padrão óbvio, o que indica um bom ajustamento de ambos os modelos.

O gráfico das probabilidades obtidas por validação cruzada que corresponde à figuras 4.24 (a) e 4.24 (b) permite, também, a detecção de observações *outliers*. No modelo 1 (figura 4.24 (a)) as observações 329, 539 e 586, comuns ao gráfico dos resíduos padronizados contra a ordem das observações para o modelo 1, são as observações cujos resíduos têm uma probabilidade de ocorrer inferior a 10% ( $P_{329} = 0.03109$ ,  $P_{539} = 0.02201$ ,  $P_{586} = 0.02781$ ) e, portanto, são consideradas candidatas a *outliers*. No modelo 2 (figura

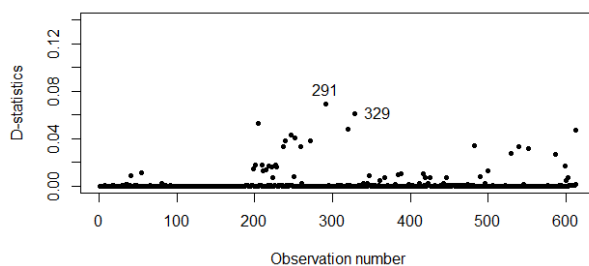


(a) Modelo 1

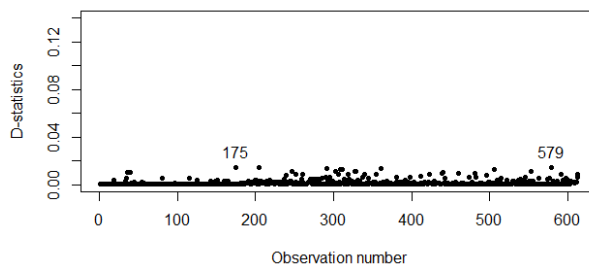


(b) Modelo 2

Figura 4.25: Gráfico de *leverage* do modelo da percepção dos hábitos alimentares



(a) Modelo 1



(b) Modelo 2

Figura 4.26: Gráfico da estatística D do modelo da percepção dos hábitos alimentares

4.24 (b)), é possível identificar as observações 175 e 579 como as observações com maior probabilidade reduzida de ocorrerem, no entanto, essas probabilidades não são inferiores a 10% ( $P_{175} = 0.11457$ ,  $P_{579} = 0.11457$ ), pelo que não são consideradas *outliers*.

Para a identificação de observações influentes nos modelos são analisadas as figuras 4.25 e 4.26. No gráfico de *leverage* (figuras 4.25 (a) e 4.25 (b)), as observações são consideradas influentes se os seus valores  $h_{ii}$  são superiores ao valor crítico = 0.01631. Como se pode verificar, para ambos os gráficos, existem muitas observações cujos valores  $h_{ii}$  são superiores ao valor crítico, pelo que a identificação individual de cada uma das observações torna-se complicada. Desta forma, recorre-se ao gráfico da esta-

tística D (figura 4.26) para uma melhor identificação de possíveis observações influentes. Para o modelo 1 (figura 4.26 (a)), as observações 291 e 329 são as observações cujos valores  $DC_i$  são superiores aos das restantes observações. Contudo, é possível observar que os valores  $DC_i$  correspondentes a essas observações não são tão elevados para que sejam consideradas observações influentes. No modelo 2 (figura 4.26 (b)) verifica-se que não há observações que se destacam em termos de valor  $DC_i$  elevado, o que indica a inexistência de possíveis observações influentes.

Os resultados provenientes da análise de resíduos efetuada a ambos os modelos binários, comprova a conclusão dos testes de qualidade de ajustamento, de que o modelo se ajusta adequadamente aos dados observados.

#### 4.3.2.3 Interpretação do modelo

Para analisar melhor o modelo em termos de interpretação dos coeficientes, a Tabela 4.19 apresenta os valores de *odds ratio* obtidos para cada uma das variáveis explicativas incluídas no modelo e as respetivas interpretações.

Tabela 4.19: Interpretação das variáveis significativas para o modelo da perceção dos hábitos alimentares

Variáveis	OR	IC <sub>95%</sub>	Interpretação OR
Sexo: Masculino	1.61	(1.12 ; 2.24)	A chance da perceção parental dos hábitos alimentares dos jovens ser mais baixa aumenta em 61% se o adolescente for do sexo masculino
Idade exata	0.83	(0.74 ; 0.92)	A chance da perceção parental dos hábitos alimentares dos jovens ser mais baixa diminui cerca de 17% para cada unidade a mais na idade
EE: Outro	0.46	(0.23 ; 0.92)	A chance da perceção parental dos hábitos alimentares dos jovens ser mais baixa diminui cerca de 54% se o encarregado de educação for outra pessoa que não o pai ou a mãe

#### 4.3.3 Perceção dos hábitos de atividade física

Tal como os valores de concordância obtidos para a perceção do peso e dos hábitos alimentares, os valores correspondentes à perceção dos hábitos de atividade física são pouco diversos no intervalo 0 e 1. Deste modo, optou-se por transformar a variável relativa à concordância dos hábitos de atividade física numa nova variável. A nova variável vai ser do tipo dicotómico, sendo o modelo de regressão a aplicar o modelo de regressão logística binária.

Os hábitos de atividade física são compostos por uma questão de escolha múltipla com 8 respostas, em que há a possibilidade de os indivíduos poderem escolher uma ou mais respostas. Neste caso, considera-se que as respostas dos encarregados de educação e as respostas dos jovens são concordantes, se ambos escolherem ou não escolherem as mesmas respostas que compõem a questão. Assim:

- Se o número de respostas concordantes entre os EE e os jovens for igual a 1, 2, 3 ou 4, a nova variável assume valor 0;

- Se o número de respostas concordantes entre os EE e os jovens for igual a 5, 6, 7 ou 8, a nova variável assume valor 1.

O valor 0 corresponde a uma concordância baixa e o valor 1 a uma concordância elevada. Os resultados obtidos encontram-se na Tabela 4.20.

Tabela 4.20: Categorização da concordância dos hábitos de atividade física dos jovens

Categorização	<i>n (%)</i> *
0	152 (22,9%)
1	509 (76,7%)

\*Resultados obtidos de uma amostra de 664 valores de concordância

Pelos valores obtidos a partir da categorização da concordância dos hábitos de atividade física dos jovens (Tabela 4.20), é possível observar que 76,7% dos encarregados de educação têm uma percepção elevada em relação aos hábitos de atividade física dos jovens.

Para saber que características (sociodemográficas) dos jovens e dos encarregados de educação podem influenciar significativamente uma percepção parental mais baixa, são analisadas as seguintes variáveis:

- Jovens: **Sexo, Idade exata e Nível de atividade física**
- Encarregados de educação: **Encarregado de educação e Idade**

#### 4.3.3.1 Seleção do modelo

Com o objetivo de construir o modelo de regressão logística, foi utilizado o método de seleção Stepwise. A partir dos três métodos de seleção aplicados, obteve-se o mesmo modelo final (AIC = 670.5532) constituído pelas seguintes variáveis:

1. **Idade exata**
2. **Nível de atividade física**

Os resultados obtidos estão apresentados na Tabela 4.21.

Tabela 4.21: Modelo de regressão logística binária da percepção dos hábitos de atividade física

Variáveis	Coefficiente	<i>Wald</i>	<i>valor-p</i>
Interseção	-0.906	-0.720	0.4718
Idade exata	0.2126	3.131	0.0017
Nível de atividade física			
Baixo (classe de referência)			
Moderado	-0.7229	-0.946	0.3443
Elevado	-1.1285	-1.496	0.1346

Resultados obtidos de uma amostra de 626 casos

Como se pode observar na Tabela 4.21, a única variável significativa no modelo é a Idade exata dos jovens, ao nível de significância  $\alpha = 0.05$ .

Para comparar o modelo selecionado com o modelo composto por todas as variáveis, foi aplicado o teste da razão de verossimilhanças. O valor-p obtido foi 0.6964, o que indica que ambos os modelos não diferem em termos de qualidade de ajustamento aos dados, devendo-se escolher o modelo com o menor número de variáveis, i.e., Idade exata e Nível de atividade física dos jovens.

Como a variável Idade exata é uma variável contínua, para verificar a linearidade existente entre ela e a variável resposta na escala *logit*, foi construído o gráfico de dispersão correspondente, apresentado na Figura 4.27.

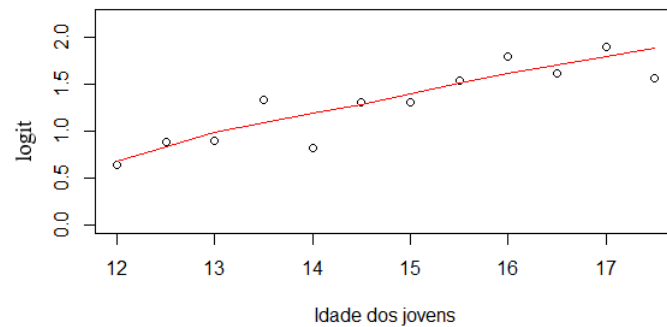


Figura 4.27: Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância dos hábitos de atividade física na escala *logit*

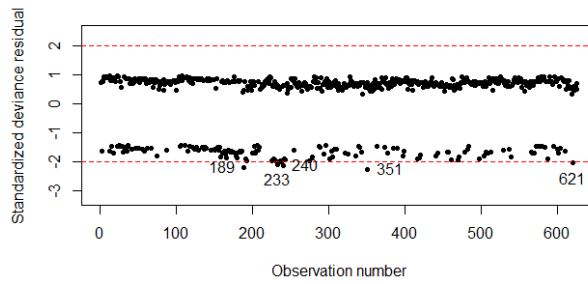
Como se pode verificar na figura 4.27, a relação entre a concordância dos hábitos de atividade física e a idade dos jovens parece linear, o que sugere a existência de linearidade na escala *logit* entre as variáveis. De modo a verificar a linearidade foi também aplicado o método de comparação de modelos lineares. O resultado obtido da comparação dos modelos ( $F = 11.921$ , valor  $p = 5.936e^{-4}$ ) permite concluir que a idade dos jovens ao quadrado é significativa para o modelo, o que indica a existência de não linearidade na escala *logit* entre a variável resposta e a idade dos jovens. No entanto, visto que a figura 4.27 revela linearidade entre as variáveis, considerou-se a existência de linearidade para a Idade exata dos jovens.

#### 4.3.3.2 Análise da qualidade de ajustamento

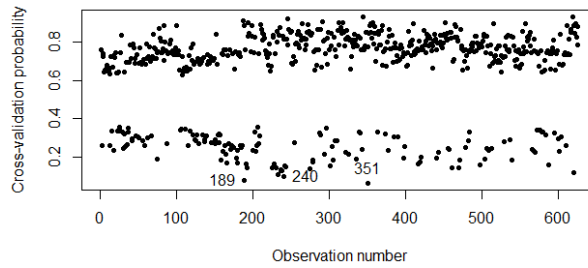
Para verificar a qualidade de ajustamento do modelo escolhido, foram aplicados a medida *Deviance* e o teste de *Hosmer e Lemeshow*. Os valores-p obtidos para a *Deviance* (valor-p = 0.1262) e para o teste *Hosmer e Lemeshow* (valor-p = 0.4801) não são significativos ao nível de significância  $\alpha = 0.05$ , o que indica um bom ajustamento do modelo aos dados.

Para verificar a qualidade de ajustamento do modelo e identificar as observações consideradas *outliers* ou influentes, estão representadas nas Figuras 4.28 e 4.29 os gráficos correspondentes.

O gráfico dos resíduos padronizados e o gráfico das probabilidades obtidas por validação cruzada correspondentes às figuras 4.28 (a) e 4.28 (b), respetivamente, permitem a deteção de observações *outliers*. Como se pode verificar na figura 4.28 (a), existem observações cujos resíduos se encontram fora do intervalo  $[-2,2]$ , nomeadamente as observações 189, 233, 240, 351 e 621, sendo consideradas como possíveis *outliers*. No gráfico das probabilidades obtidas por validação cruzada foram identificadas as observações 189, 240 e 351 como possíveis *outliers*, devido aos valores de probabilidade cruzada serem inferiores a 10% ( $P_{189} = 0.07722$ ,  $P_{240} = 0.09972$ ,  $P_{351} = 0.06496$ ).

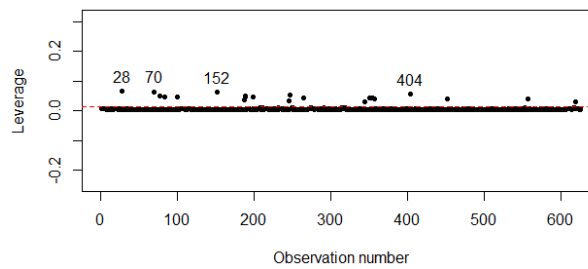


(a) Gráfico dos resíduos padronizados contra a ordem das observações

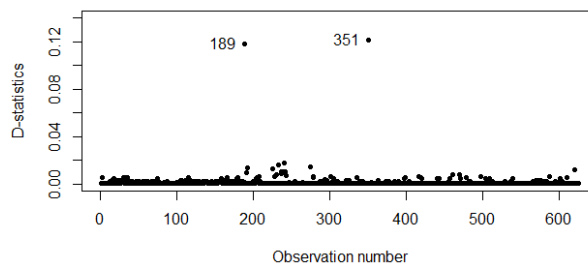


(b) Gráfico das probabilidades obtidas por validação cruzada

Figura 4.28: Análise do preditor linear e identificação de observações *outliers* do modelo da percepção dos hábitos de atividade física



(a) Gráfico de *leverage*



(b) Gráfico da estatística D

Figura 4.29: Identificação de observações influentes do modelo da percepção dos hábitos de atividade física

O gráfico dos resíduos padronizados contra a ordem das observações permite também avaliar a qualidade de ajustamento do modelo, através do padrão da distribuição dos resíduos. A distribuição dos resíduos dentro do intervalo não segue nenhum padrão óbvio, o que indica um bom ajustamento do modelo aos dados.

Para a identificação de observações influentes no modelo são analisadas as figuras 4.29 (a) e 4.29 (b). No gráfico de *leverage* (figura 4.29 (a)), podem ser identificadas algumas observações cujos  $h_{ii}$  são superiores

ao valor crítico = 0.01278, nomeadamente as observações 28, 70, 152 e 404. No gráfico da estatística D (figura 4.29 (b)) é também possível identificar observações influentes, destacando-se as observações 189 e 351, devido aos seus valores  $DC_i$  serem superiores aos das restantes observações ( $DC_{189} = 0.11796$ ,  $DC_{351} = 0.12132$ ).

Na Tabela 4.22 apresentam-se os resultados do modelo ajustado sem as observações 28, 70, 152, 189, 351 e 404.

Tabela 4.22: Modelo de regressão logística binária da perceção dos hábitos de atividade física sem as observações 28, 70, 152, 189, 351 e 404

Variáveis	Coefficiente	Wald	valor-p
Interseção	-0.906	-0.720	0.4718
Idade exata	0.2126	3.131	0.0017
Nível de atividade física			
Baixo	-----	-----	-----
Moderado	-0.7229	-0.946	0.3443
Elevado	-1.1285	-1.496	0.1346

As variações encontradas relativamente aos coeficientes e à estatística de *Wald* das variáveis foram nulas, pelo que todas as variáveis que foram significativas no modelo escolhido continuaram a ser significativas no modelo sem as observações.

Para verificar se a função de ligação escolhida é adequada, foi adicionada ao modelo uma nova variável explicativa, *Z*, e avaliado o declínio na função desvio. O declínio não foi significativo (valor-p = 0.1483), pelo que a função de ligação *logit* é considerada adequada.

#### 4.3.3.3 Capacidade preditiva do modelo

Para a avaliação da capacidade preditiva do modelo recorreu-se ao estudo das medidas de desempenho através da matriz de confusão e da área sob a curva ROC.

O melhor ponto de corte é 1.0473683, estando na Tabela 4.23 os resultados possíveis na classificação dos indivíduos em cada classe da variável resposta ( $Y = 0$  e  $Y = 1$ ).

Tabela 4.23: Tabela de classificação para o modelo da perceção dos hábitos de atividade física

Classificação	Valores observados		Total
	Classe(0)	Classe(1)	
Classe(0)	79	145	224
Classe(1)	67	335	402
Total	146	480	626

Através da Tabela 4.23 é possível estimar as medidas de desempenho, tendo sido obtidos os valores 69,79% para a sensibilidade e 54,11% para a especificidade. Para a percentagem total de acertos, o valor obtido foi 66,13%.

A área abaixo da curva ROC varia entre 0 e 1 e classifica de forma mais precisa a capacidade do modelo em discriminar os valores da variável resposta  $Y = 1$  dos valores da variável  $Y = 0$ . O gráfico correspondente à curva ROC está apresentado na Figura 4.30.

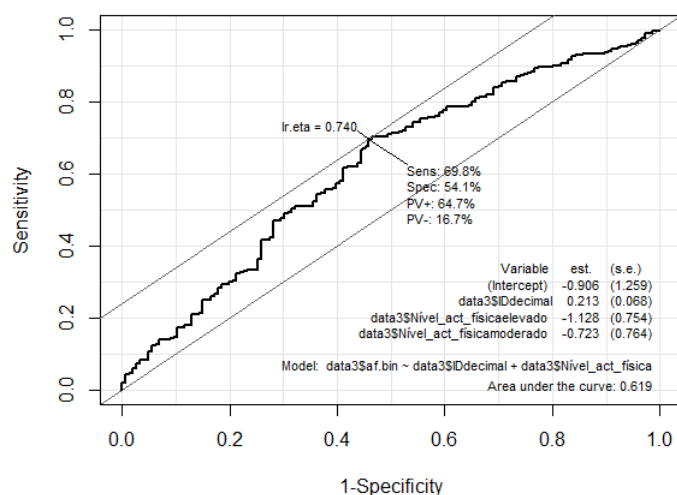


Figura 4.30: Curva ROC do modelo da percepção dos hábitos de atividade física

O valor obtido da área abaixo da curva foi 0.619, que indica um modelo sem poder discriminatório, de acordo com os critérios de avaliação propostos por Hosmer & Lemeshow.

#### 4.3.3.4 Interpretação do modelo

Para analisar melhor o modelo em termos de interpretação do coeficiente significativo, a Tabela 4.24 apresenta o valor de *odds ratio* correspondente e a sua interpretação.

Tabela 4.24: Interpretação da variável significativa para o modelo da percepção dos hábitos de atividade física dos jovens

Variável	OR	IC <sub>95%</sub>	Interpretação OR
Idade exata	1.24	(1.09 ; 1.42)	A chance da percepção parental dos hábitos de atividade física dos jovens ser mais alta aumenta em cerca de 24% para cada unidade a mais na idade

### 4.3.4 Percepção dos hábitos sedentários

Ao contrário dos casos anteriores, os valores de concordância obtidos para a percepção dos hábitos sedentários são bastante diversificados, o que torna possível aplicar o modelo de regressão beta.

Antes da aplicação do modelo, foi necessário aplicar a transformação  $(y \times (n - 1) + 0.5) / n$  aos valores da concordância, devido ao facto de esta assumir como valor mínimo 0 e valor máximo 1.

#### 4.3.4.1 Seleção do modelo

Sabe-se, pela Tabela 4.6, que 50% dos encarregados de educação tem uma percepção superior a 85%, o que indica uma boa percepção parental em relação aos hábitos sedentários que os jovens têm. Assim, para saber que características (sociodemográficas) dos jovens e dos encarregados de educação podem influenciar significativamente a percepção, são analisadas as seguintes variáveis:

- Jovens: **Sexo e Idade exata**

- Encarregados de educação: **Encarregado de educação e Idade**

De entre as variáveis em estudo, é necessário selecionar as que são mais importantes para o modelo. Na regressão beta, a seleção de modelos faz-se através da medida AIC que mede o grau de ajustamento do modelo. Este ajustamento será tanto melhor quanto menor o valor AIC correspondente.

Para saber que variáveis proporcionam ao modelo um menor valor AIC, foi necessário ajustar 16 modelos com diferentes combinações de variáveis, incluindo o modelo nulo e o modelo com todas as variáveis. O modelo com menor valor AIC é o modelo com a variável **Idade dos encarregados de educação** (AIC = -629.7920), estando a análise do modelo apresentada na Tabela 4.25.

Tabela 4.25: Modelo de regressão beta da perceção dos hábitos sedentários

Variáveis	Coefficiente	Wald	valor-p
Interseção	0.6559	2.492	0.0127
Idade (EE)	0.0099	1.633	0.1024

Resultados obtidos de uma amostra de 643 casos

Como se pode observar na Tabela 4.25, a variável Idade dos EE não é significativa ao nível de significância  $\alpha = 0.05$ .

#### 4.3.4.2 Análise da qualidade de ajustamento do modelo

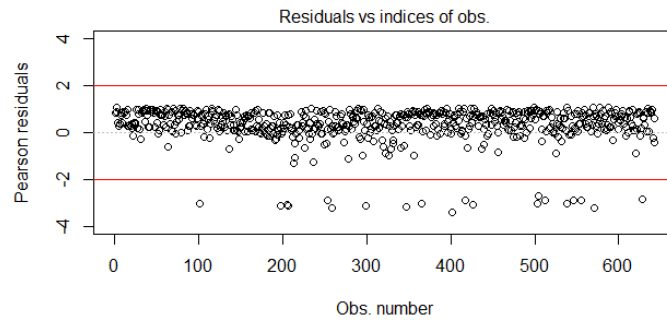
Para verificar a qualidade de ajustamento do modelo foi aplicada a medida pseudo  $R^2$ . Esta medida varia entre 0 e 1, correspondendo o valor 1 a um perfeito ajustamento do modelo. O valor pseudo  $R^2$  obtido para este modelo foi 0.0041, o que indica um péssimo ajustamento do modelo aos dados observados.

Para ajudar na identificação de observações consideradas *outliers* ou influentes, estão representadas nas Figuras 4.31 a 4.33 os gráficos correspondentes.

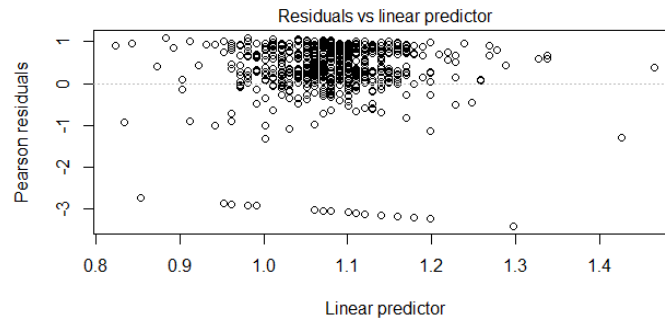
O gráfico dos resíduos de *Pearson* contra o índice das observações (figura 4.31 (a)) e o gráfico dos resíduos da *deviance* contra o índice das observações (figura 4.31 (c)) permitem a deteção de observações *outliers*. Como se pode verificar na figura 4.31 (a), existem observações cujos resíduos se encontram fora do intervalo  $[-2,2]$ , sendo consideradas como possíveis *outliers*. Na figura 4.31 (c) também é possível verificar resíduos fora dos limites do intervalo, o que sugere também a presença de possíveis observações *outliers* no modelo.

O gráfico dos resíduos de *Pearson* contra o preditor linear (figura 4.31 (b)) permite detetar, através da análise do padrão de distribuição dos resíduos, se a função de ligação escolhida é adequada. A distribuição dos resíduos parece não seguir nenhum padrão ou tendência, o que sugere que a função de ligação escolhida é adequada.

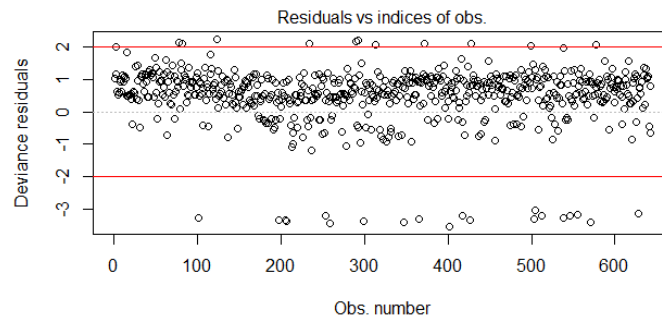
O gráfico *Half-normal* dos resíduos da *deviance* com envelope simulado (figura 4.32) permite verificar a adequabilidade do modelo ajustado, através da presença ou não dos resíduos dentro dos limites do envelope. Como se pode verificar, praticamente todas as observações encontram-se fora dos limites do envelope, indicando que o modelo não se ajusta adequadamente aos dados.



(a) Gráfico dos resíduos de *Pearson* contra o índice das observações



(b) Gráfico dos resíduos de *Pearson* contra o preditor linear



(c) Gráfico dos resíduos da *deviance* contra o índice das observações

Figura 4.31: Análise do preditor linear e identificação de observações *outliers* do modelo da percepção dos hábitos sedentários

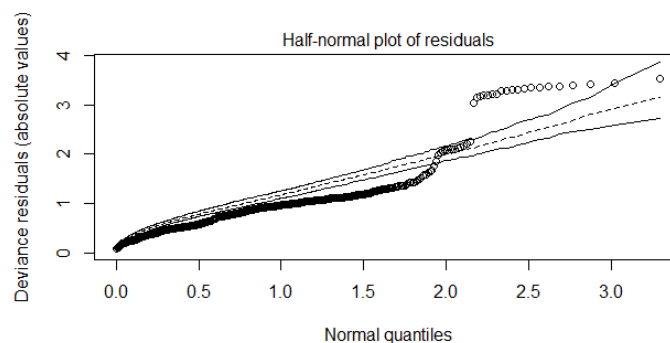
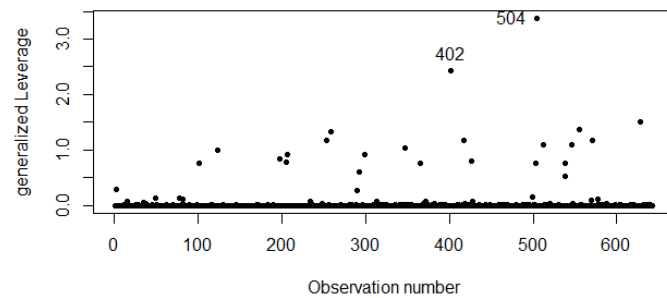
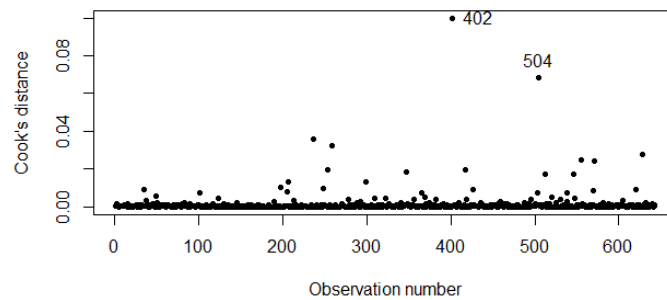


Figura 4.32: Gráfico *Half-normal* dos resíduos da *deviance* do modelo da percepção dos hábitos sedentários

Para a identificação de observações influentes no modelo são analisados os gráficos de *leverage* generalizado (figura 4.33 (a)) e da distância de *Cook* (figura 4.33 (b)). Em ambos os gráficos é possível detetar observações com valores de *leverage* generalizado (figura 4.33 (a)) e valores  $C_i$  (figura 4.33 (b)) superiores aos das outras observações, mas não suficientemente elevados para serem consideradas observações influentes.



(a) Gráfico de *leverage* generalizado



(b) Gráfico da distância de *Cook*

Figura 4.33: Identificação de observações influentes do modelo da percepção dos hábitos sedentários

### 4.3.5 Percepção global

Tal como na percepção dos hábitos sedentários, os valores de concordância para a percepção global são bastante diversificados no intervalo entre 0 e 1, pelo que pode ser aplicado o modelo de regressão beta.

Conforme apresentado na Tabela 4.6, a percepção global assume como valor máximo o valor 1, pelo que se procedeu à transformação dos valores da concordância.

#### 4.3.5.1 Seleção do modelo

Pela Tabela 4.6, sabe-se que 50% dos encarregados de educação tem uma percepção superior a 81%, o que indica uma boa percepção parental em relação ao peso e estilos de vida dos jovens. Assim, para saber que características (sociodemográficas) dos jovens e dos encarregados de educação podem influenciar significativamente a percepção, são analisadas as seguintes variáveis:

- Jovens: **Sexo e Idade exata**
- Encarregados de educação: **Encarregado de educação e Idade**

Para seleccionar o melhor modelo através da medida AIC, foi necessário ajustar 16 modelos com diferentes combinações das variáveis em estudo, incluindo o modelo nulo e o modelo com todas as variáveis. O modelo com menor valor AIC é o modelo com as variáveis **Sexo e Idade exata** dos jovens (AIC = -878.9923), estando a análise do modelo apresentada na Tabela 4.26.

Como se pode observar na Tabela 4.26, a única variável significativa ao nível de significância  $\alpha = 0.05$  é a variável Idade exata dos jovens. No entanto, o facto de o valor-p associado ao coeficiente da categoria

Tabela 4.26: Modelo de regressão beta da percepção global

Variáveis	Coefficiente	Wald	valor-p
Interseção	0.5588	1.813	0.0698
Sexo			
Feminino (classe de referência)			
Masculino	-0.1146	-1.769	0.0769
Idade exata	0.0599	2.844	0.0045

Resultados obtidos de uma amostra de 643 casos

sexo masculino estar próximo do  $\alpha$  estabelecido, indica que há uma tendência de que a concordância global seja mais baixa para os jovens do sexo masculino.

Como a variável Idade exata é uma variável contínua, para verificar a linearidade existente entre ela e a variável resposta na escala *logit*, faz-se uma representação gráfica. Este gráfico está apresentado na Figura 4.34.

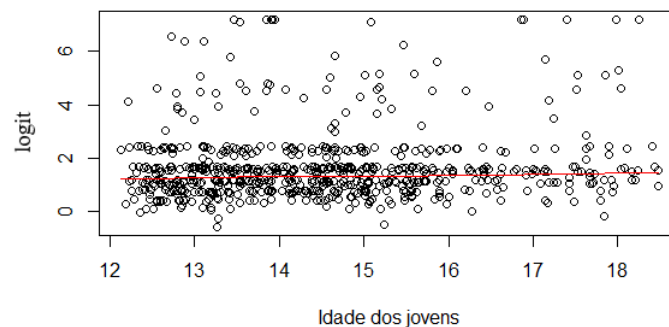


Figura 4.34: Representação gráfica da linearidade entre a variável Idade dos jovens e a concordância global na escala *logit*

Como se pode observar na figura 4.34, a relação entre *logit* da concordância global e a idade dos jovens parece seguir uma tendência linear, indicando a existência de linearidade na escala *logit* entre a variável resposta e a idade dos jovens.

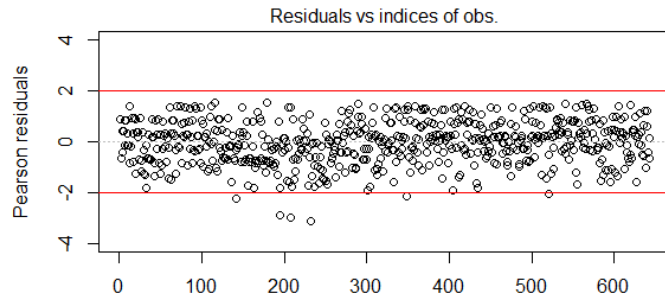
O resultado obtido da comparação entre os dois modelos lineares ( $F = 1.207$ , valor-p = 0.2723) indica que a variável idade dos jovens ao quadrado não é significativa para o modelo, considerando-se a existência de linearidade na escala *logit* entre a variável resposta e a Idade exata dos jovens.

#### 4.3.5.2 Análise da qualidade de ajustamento do modelo

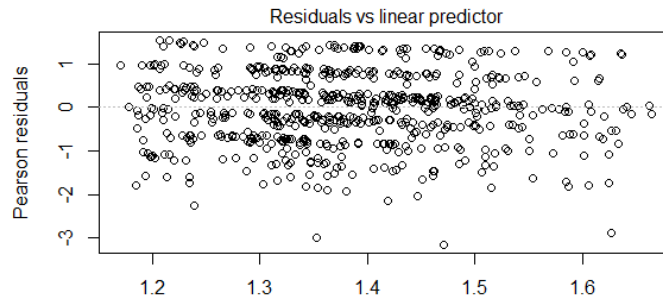
Para verificar a qualidade de ajustamento do modelo foi aplicada a medida pseudo  $R^2$ . O valor pseudo  $R^2$  obtido para este modelo foi de 0.0128, o que indica um péssimo ajustamento do modelo aos dados observados.

Para verificar a qualidade de ajustamento do modelo e identificar as observações consideradas *outliers* ou influentes, estão representadas nas Figuras 4.35 a 4.37 os gráficos correspondentes.

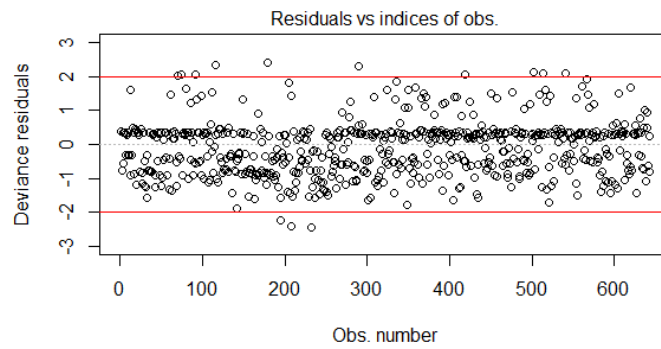
Os gráficos dos resíduos de *Pearson* contra o índice das observações (figura 4.35 (a)) e dos resíduos da *deviance* contra o índice das observações (figura 4.35 (c)) permitem a detecção de observações *outliers*. Como se pode verificar na figura 4.35 (a), existem observações cujos resíduos se encontram fora do intervalo estabelecido  $[-2,2]$ , sendo consideradas como possíveis *outliers*. Na figura 4.35 (c) também é



(a) Gráfico dos resíduos de *Pearson* contra o índice das observações



(b) Gráfico dos resíduos de *Pearson* contra o preditor linear



(c) Gráfico dos resíduos da *deviance* contra o índice das observações

Figura 4.35: Análise do preditor linear e identificação de observações *outliers* do modelo da percepção global

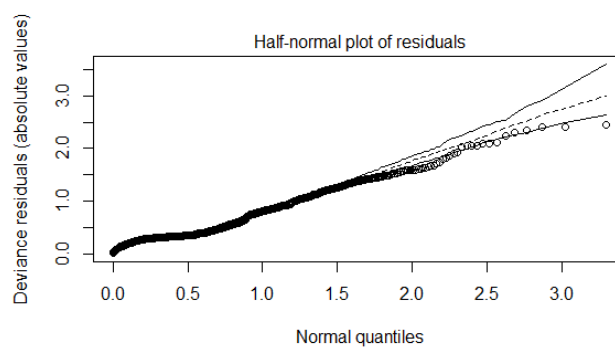
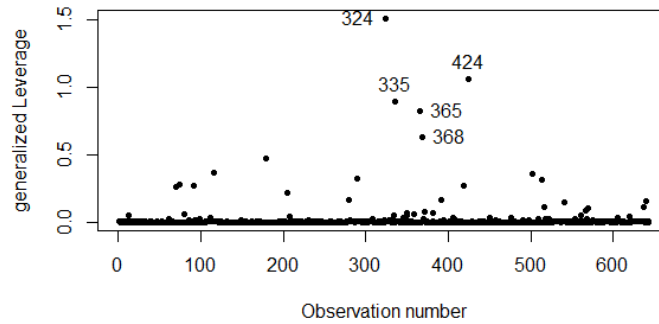


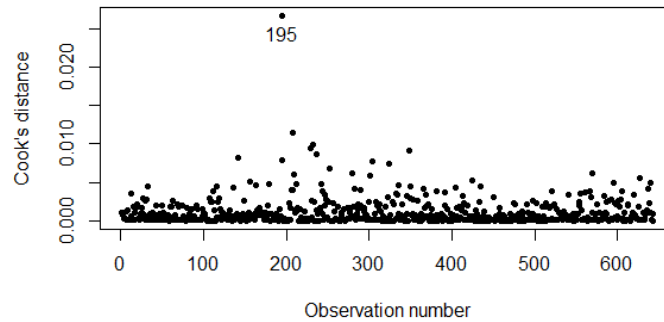
Figura 4.36: Gráfico *Half-normal* dos resíduos da *deviance* do modelo da percepção global

possível verificar resíduos fora dos limites do intervalo, o que confirma a presença de possíveis observações *outliers* no modelo.

O gráfico dos resíduos de *Pearson* contra o preditor linear (figura 4.35 (b)) permite detetar se a função de ligação escolhida é adequada. A distribuição dos resíduos parece não seguir nenhuma tendência que sugira a falta de adequabilidade da função de ligação escolhida, pelo que esta pode ser considerada ade-



(a) Gráfico de *leverage* generalizado



(b) Gráfico da distância de *Cook*

Figura 4.37: Identificação de observações influentes do modelo da percepção global

quada.

O gráfico *Half-normal* dos resíduos da *deviance* com envelope simulado (figura 4.36) permite verificar a adequabilidade do modelo ajustado, através da presença ou não dos resíduos dentro dos limites do envelope. Como se pode verificar, há observações cujos resíduos se encontram fora do envelope. Apesar de a proporção de observações fora do envelope não ser elevada, pode sugerir, mesmo assim, que o modelo não se ajusta adequadamente aos dados.

Para a identificação de observações influentes no modelo são analisados os gráficos de *leverage* generalizado (figura 4.37 (a)) e da distância de *Cook* (figura 4.37 (b)). Em ambos os gráficos, é possível detetar observações com valores de  $h_{ii}$  generalizado (figura 4.37 (a)) e valores  $C_i$  (figura 4.37 (b)) superiores aos das outras observações, mas não suficientemente elevados para serem consideradas observações influentes.

### 4.3.5.3 Interpretação do modelo

Para analisar melhor o modelo em termos de interpretação do coeficiente significativo, a Tabela 4.27 apresenta o valor de *odds ratio* correspondente e a sua interpretação.

Tabela 4.27: Interpretação da variável significativa para o modelo da percepção global dos jovens

Variável	OR	IC <sub>95%</sub>	Interpretação OR
Idade exata	1.06	(1.02 ; 1.11)	A chance da estimativa da percepção global aumenta cerca de 6% para cada unidade a mais na idade dos jovens

## Capítulo 5

# Discussão

### 5.1 Análise exploratória

A análise exploratória realizada para algumas das características sociodemográficas dos jovens e dos encarregados de educação, permitiu tomar conhecimento de que 20,6% dos jovens sofre de excesso de peso e 3,3% sofre de obesidade. Esta situação torna-se ainda mais problemática, quando se constata que 61,0% dos encarregados de educação não se preocupa com o peso do educando, apesar de 40,9% dos encarregados de educação se preocupar com o peso dos educandos quando estes são obesos.

O facto de a maioria dos encarregados de educação não se preocupar com o peso dos jovens, pode levar a que não sejam capazes de o perceber corretamente. Quando isto acontece, principalmente quando os jovens têm excesso de peso ou obesidade, os pais e os outros adultos responsáveis não terão a necessidade de perceber que hábitos dos jovens proporcionam um aumento de peso prejudicial para a saúde nem de os modificar.

### 5.2 Análise da percepção parental: medidas de concordância

Com o objetivo de caracterizar a percepção dos encarregados de educação em relação ao peso e hábitos dos jovens, foram feitas análises descritivas dos valores obtidos para cada uma das percepções. As percepções tomam valores bastante elevados, revelando que metade dos encarregados de educação presentes na amostra têm uma boa capacidade em reconhecer corretamente o peso e os hábitos dos jovens.

Conforme apresentado no Capítulo 3, as questões sobre o peso e hábitos sedentários permitem perceber se os adultos responsáveis percebem corretamente o peso dos jovens e o tempo que os jovens despendem a ver televisão ou no computador. No entanto, em relação aos hábitos alimentares e de atividade física, as questões apresentadas não permitem perceber que tipo de alimentação os jovens têm ou a atividade física que praticam. À exceção da pergunta sobre a frequência com que os jovens tomam o pequeno-almoço, as outras questões dizem respeito à disponibilização de meios para que os jovens consumam frutas/vegetais e pratiquem atividades desportivas. Deste modo, os resultados de concordância obtidos para essas questões não refletem, de facto, a percepção que os encarregados têm em relação à alimentação e atividade física dos jovens. No entanto, no contexto da dissertação, pode-se pensar nessas questões como um reflexo do consumo de frutas/vegetais e de prática de atividades desportivas por parte dos jovens e, assim, concluir que os adultos responsáveis por eles têm uma boa percepção quanto a esses hábitos.

No âmbito da análise da percepção parental em relação ao peso e hábitos dos jovens, pode ser importante perceber como a percepção varia entre algumas características sociodemográficas dos jovens e dos encar-

regados de educação. Com esse intuito, foram feitas análises para a percepção do peso e para cada um dos hábitos.

Relativamente ao peso dos jovens, foi verificado que a percepção parental diferia significativamente entre os géneros e as classes de peso dos jovens, bem como entre os níveis de preocupação dos encarregados de educação quanto ao peso dos educandos. De entre os valores médios da percepção obtidos (Tabela 4.7), é importante destacar que a percepção em relação aos jovens com excesso de peso e obesidade é mais baixa do que em relação aos jovens com peso normal. Apesar dos valores obtidos para as classes de peso estarem próximos de 1, o facto de uma das percepções mais baixas ser referente à classe obesidade, é um indicador de que os encarregados de educação tendem a subestimar o peso dos jovens quando estes têm peso a mais. Esta mesma situação é relatada em alguns estudos do mesmo género, apesar da metodologia aplicada ser diferente. O estudo conduzido por Hearst *et. al* mostra que, dos progenitores que classificam erradamente o peso dos filhos, 64,29% subestimam o peso dos filhos com excesso de peso/obesidade [19]. No estudo de Boutelle *et. al*, 70,5% dos jovens com excesso de peso e 86,2% dos jovens com obesidade têm o seu peso subestimado pelas mães [16]. Apesar de estes dois estudos mostrarem como resultado a percentagem de progenitores que subestimam o peso de jovens e o resultado da dissertação dizer respeito à concordância entre as respostas à questão sobre o peso dos adolescentes, ambos os resultados revelam que há progenitores que tendem a subestimar o peso dos seus filhos quando estes têm peso a mais. Isto pode acontecer devido ao facto de os progenitores não quererem reconhecer que os filhos têm peso a mais com medo de serem responsabilizados pelos profissionais de saúde pelo problema de excesso de peso dos filhos.

O facto de existirem encarregados de educação que não se preocupam com o peso dos educandos poderia levar-nos a considerar que esses encarregados de educação teriam uma menor capacidade em reconhecer o peso correto dos jovens. No entanto, a Tabela 4.8 mostra que não é o caso, sendo os adultos que não se preocupam os mais precisos na classificação correta do peso.

Quanto à percepção dos hábitos alimentares, os encarregados de educação são mais precisos relativamente aos jovens com idades mais baixas. Este resultado faz sentido, visto haver uma maior atenção por parte dos adultos quanto à alimentação dos jovens quando estes são mais novos do que quando são mais velhos e se tornam mais independentes, não só onde fazem as refeições, mas também no tipo de alimentos que consomem. O facto de a mãe poder ser vista como a principal cuidadora do filho em termos da alimentação que disponibiliza, é interessante verificar que é o encarregado de educação que apresenta uma percepção mais baixa em relação aos hábitos alimentares dos jovens (Tabela 4.9).

Acerca dos hábitos de atividade física dos jovens, foi verificado que a percepção parental diferia significativamente entre rapazes e raparigas, entre as idades dos jovens e entre os níveis de atividade física praticados por estes. Dado que estes hábitos não são propriamente hábitos, mas sim formas de incentivo por parte dos encarregados de educação para que os jovens pratiquem atividade física, a interpretação dos resultados não é clara. Neste caso, o que se pode fazer é especular quanto à razão do porquê de a percepção diferir para certas características dos jovens. Visto a questão ser sobre incentivos para a prática de atividade física, as diferenças encontradas podem dever-se, por exemplo, aos jovens acharem que são determinados apoios que incentivam à sua prática desportiva e os adultos acharem que são outros apoios que levam à prática desportiva.

No caso dos hábitos sedentários dos adolescentes, não se verificaram diferentes percepções dos encarregados de educação consoante os géneros e idades dos jovens, nem entre eles próprios. Estas mesmas características foram também utilizadas na análise da percepção global. Esta percepção corresponde à forma como os encarregados de educação percebem, em conjunto, o peso e os hábitos dos jovens. Os resultados encontrados indicam que os adultos percebem de forma mais precisa o peso e os hábitos

nas raparigas do que nos rapazes, sendo esta diferença significativa. Isto sugere que as raparigas continuam a ser alvo de mais atenção por parte dos pais e de outros familiares do que os rapazes, o que não é um bom indício quando os jovens sofrem de excesso de peso ou obesidade e têm hábitos que contribuem para o aumento de peso.

### 5.3 Análise da perceção parental: aplicação de modelos de regressão

No âmbito da análise da perceção parental em relação ao peso e hábitos dos jovens, é igualmente importante perceber se há alguma característica dos jovens ou dos encarregados de educação que influencie a perceção. Com o objetivo de encontrar essas características, foram ajustados diferentes modelos de regressão consoante a perceção em estudo.

No caso da perceção do peso, foi ajustado um modelo de regressão logística binária. Neste modelo, a perceção toma valores 0 e 1, correspondendo o valor 0 a uma perceção errada em relação ao peso dos jovens. A proporção de adultos com uma perceção errada é cerca de 27,4%, sendo este resultado um pouco mais baixo do que os resultados encontrados no estudo de Borges, em que 33,2% dos progenitores não classificam corretamente o peso das crianças e jovens [15] e no estudo de Hearst *et. al*, em que 31% dos progenitores classifica erradamente o peso dos adolescentes ao subestimar o peso [19]. Quanto às características que influenciam significativamente a perceção do peso dos jovens, o sexo masculino, o excesso de peso e a obesidade dos jovens foram identificadas como as características que proporcionam uma diminuição na capacidade dos encarregados de educação em classificar corretamente o peso dos adolescentes, i.e., promovem uma perceção incorreta em relação ao peso dos jovens. Estes resultados estão de acordo com os resultados encontrados num outro estudo, em que a classificação incorreta do peso é 1.65 vezes superior se o adolescente for do sexo masculino (OR = 1.65, IC<sub>95%</sub>: 1.04 - 2.62) e 5.23 vezes superior se o adolescente tiver excesso de peso ou obesidade (OR = 5.23, IC<sub>95%</sub>: 3.04 - 9.01) [19]. Relativamente à perceção dos hábitos alimentares, foi ajustado um modelo de regressão logística ordinal. Neste modelo, a perceção toma os valores 0, 1 e 2, correspondendo o valor 0 a uma incorreta perceção e o valor 2 a uma correta perceção. A maioria dos encarregados de educação, cerca de 67,6%, revelou uma correta perceção. Quanto às características, o sexo masculino e as idades inferiores dos jovens foram identificadas como as características que proporcionam uma perceção mais baixa quanto aos hábitos alimentares dos jovens.

Para modelar a perceção dos hábitos de atividade física, foi ajustado um modelo de regressão logística binária. Neste caso, o valor 0 corresponde a uma perceção baixa e o valor 1 a uma perceção elevada. Das variáveis que foram selecionadas para o modelo, apenas a idade dos jovens foi considerada significativa, sendo uma idade maior a característica que proporciona uma perceção mais alta.

No caso da perceção dos hábitos sedentários e da perceção global, foram ajustados modelos de regressão beta. Nestes modelos, a perceção toma valores entre 0 e 1, correspondendo valores próximos de 0 a uma perceção mais baixa e próximos de 1 a uma perceção mais alta. No caso da perceção dos hábitos sedentários, não foram identificadas características que influenciassem significativamente os valores da perceção. Quanto à perceção global, foi verificado que os encarregados de educação têm um aumento de perceção do peso e estilos de vida dos adolescentes à medida que estes vão ficando mais velhos.

## Capítulo 6

# Conclusões

A prevalência do excesso de peso e obesidade na adolescência tem vindo a aumentar drasticamente nos últimos anos. Dado que certos hábitos alimentares, de atividade física e sedentários podem levar a um aumento peso prejudicial para a saúde, esta dissertação teve como objetivo analisar a percepção parental em relação ao peso e estilos de vida dos jovens.

A partir de inquéritos aplicados a jovens inscritos no 3º Ciclo e Secundário e aos respetivos encarregados de educação, foi possível obter a informação necessária para a realização desta análise.

Para todas as percepções em estudo (peso, hábitos alimentares, hábitos de atividade física, hábitos sedentários e global) foi verificado que metade dos encarregados de educação (numa amostra de 664 indivíduos) tinha uma boa capacidade em reconhecer de forma correta o peso e os hábitos dos jovens.

Com o intuito de verificar se as percepções em estudo variavam em função de uma alguma característica dos jovens ou dos encarregados de educação, foram realizadas análises para cada uma das percepções.

Para a percepção do peso, verificou-se que esta variava significativamente entre os sexos e entre as classes de peso dos jovens, bem como entre os níveis de preocupação dos encarregados de educação quanto ao peso dos educandos. Relativamente à classe de peso dos jovens, uma das percepções mais baixas corresponde aos jovens que sofrem de obesidade. Quanto aos hábitos alimentares, constatou-se que estes eram percecionados de forma mais precisa em jovens com idades mais baixas. Acerca dos hábitos de atividade física, a percepção difere significativamente entre os rapazes e as raparigas, entre as idades dos jovens e entre os níveis de atividade física praticados por estes. No caso dos hábitos sedentários, não se verificaram diferentes percepções consoante as características dos jovens, sexo e idade, nem entre os diferentes encarregados de educação. Para a percepção global, foi verificado que os encarregados têm uma melhor percepção quanto ao peso e hábitos das raparigas do que dos rapazes.

Com o objetivo de perceber que características poderiam influenciar cada uma das percepções foram aplicados modelos de regressão.

No caso da percepção do peso, verificou-se que o sexo masculino, o excesso de peso e a obesidade dos jovens são as características que proporcionam uma percepção correta por parte dos encarregados de educação em relação ao peso dos jovens. Quanto aos hábitos alimentares, as características sexo masculino e idades inferiores dos jovens foram identificadas como as características que resultam numa percepção mais baixa. Para os hábitos de atividade física, verificou-se apenas que a idade dos jovens influenciava significativamente a percepção, sendo uma idade maior a característica que proporciona uma percepção mais alta. No que diz respeito aos hábitos sedentários, não foram identificadas características que influenciassem significativamente a percepção. No caso da percepção global, constatou-se que à medida que os jovens ficam mais velhos, melhor é a percepção parental.

Os resultados obtidos a partir da modelação da percepção do peso, confirmam alguns dos resultados

encontrados em outros estudos do género, de que há uma tendência para a subestimação do peso dos jovens, quando estes têm peso a mais. Já outros resultados obtidos nesta dissertação não vão ao encontro do objetivo proposto, nomeadamente os resultados obtidos para a perceção dos hábitos alimentares e dos hábitos de atividade física. O facto de estes hábitos refletirem disponibilização de meios para consumo de determinados alimentos e para a prática de atividades desportivas, os resultados não permitem ter uma visão clara da capacidade que os pais/EE têm em perceber a alimentação e a atividade física praticada pelos jovens. No entanto, a metodologia usada pode ser vista como uma ferramenta útil no caso de, no futuro, se resolva estudar a perceção dos hábitos alimentares e de atividade física dos jovens, usando para o efeito, as questões apropriadas.

# Bibliografia

- [1] World Health Organization (2018). *Obesity and overweight*. Acedido em 2 de setembro de 2018, em: <http://www.who.int/en/news-room/fact-sheets/detail/obesity-and-overweight>.
- [2] Direção-Geral da Saúde (2018). *Programa de Saúde Escolar e Promoção de uma Alimentação Saudável em meio escolar*. Acedido em 2 de setembro de 2018, em: <https://www.dgs.pt/promocao-da-saude/saude-escolar/areas-de-intervencao/educacao-alimentar.aspx>.
- [3] Sardinha, L., Santos, R., Vale, S., Silva, A., Ferreira, J., Raimundo, A., Moreira, H., Baptista, F. & Mota, J. (2011). Prevalence of overweight and obesity among Portuguese youth: A study in a representative sample of 10–18-year-old children and adolescents. *International Journal of Pediatric Obesity*, 6, (2-2): e124-8.
- [4] Reilly, J. J. & Kelly J. (2011). Long-term impact of overweight and obesity in childhood and adolescence on morbidity and premature mortality in adulthood: systematic review. *International Journal of Obesity*, 35: 891-898.
- [5] Lobstein, T., Baur, L. & Uauy, R. (2004). Obesity in children and young people: a crisis in public health. *obesity reviews*, 5 (Suppl. 1), 4-85.
- [6] Qidway, W., Ishaque, S., Shah, S. & Rahim, M. (2010). Adolescent Lifestyle and Behaviour: A Survey from a Developing Country. *PLoS ONE*, 5(9): e1214.
- [7] Boone, J., Gordon-Larsen, P., Adair, L. & Popkin, B. (2007). Screen time and physical activity during adolescence: longitudinal effects on obesity in young adulthood. *International Journal of Behavioral Nutrition and Physical Activity*, 4:26.
- [8] Campbell, K., Crawford, D., Salmon, J., Carver, A., Garnett, S. & Baur, L. (2007). Associations between the home food environment and obesity-promoting eating behaviors in adolescence. *Obesity*, 15: 719-730.
- [9] Croezen, S., Visscher, T.L.S., ter Bogt, N.C.W., Veling, M.L. & Haveman-Nies, A. (2009). Skipping breakfast, alcohol consumption and physical inactivity as risk factors for overweight and obesity in adolescents: results of the E-MOVO project. *European Journal of Clinical Nutrition*, 63, 405-412.
- [10] Enes, C.C. & Slater, B. (2010). Obesidade na adolescência e seus principais fatores determinantes. *Revista Brasileira de Epidemiologia*, 13(1): 163-171.
- [11] Fonseca, V. Sichieri, R. & Veiga, G. (1998). Fatores associados à obesidade em adolescentes. *Revista Saúde Pública*, 32 (6): 541-9.

- [12] Hanson, N., Neumark-Sztainer, D., Eisenberg, M., Story, M. & Wall, M. (2005). Associations between parental report of the home food environment and adolescent intakes of fruits, vegetables and dairy food. *Public Health Nutrition*, 8 (11), 77-85.
- [13] DuBois, M. (2010). Parental Attitudes and Perceptions Related to Their Children's Physical Activity and Eating Patterns. *All Graduate Theses and Dissertations*. 839.
- [14] Hudson, E., MCCannon, A. & McGloin, A. (2011). Parental weight (mis)perceptions: factors influencing parent's ability to correctly categorise their child's weight status. *Maternal and Child Health Journal*, 16, 1801-1809.
- [15] Borges, C. P. (2015). *Parental perception of their children's weight status*. Tese de Mestrado em Nutrição Clínica. Faculdade de Ciências da Nutrição - Universidade do Porto, Porto.
- [16] Boutelle, K., Fulkerson, J., Neumark-Sztainer, D. & Story, M. (2004). Mother's perceptions of their adolescent's weight status: are they accurate?. *Obesity Research*, 12, 1754-1757.
- [17] Gouveia, C., Pereira-da-Silva, L., Virella, D., Silva, P. & Amaral, J. (2007). Actividade física e sedentarismo em adolescentes escolarizados do concelho de Lisboa. *Acta Pediátrica Portuguesa*, 38(1), 7-12.
- [18] Burusic, J., Sakic, M. & Koprta, N. (2014). Parental perception of adolescent health behaviors: experiences from croatian high schools. *Health Education Journal*, 73(3), 351-360.
- [19] Hearst, M., Sherwood, N., Klein, E., Pasch, K. & Lytle, L. (2011). Parental perceptions of their adolescent's weight status: the ECHO study. *Am J Health Behav.*, 35(2): 248-255.
- [20] Sobral, G. M. R. (2013). *Excesso de Peso e Estilos de Vida na Adolescência*. Dissertação de Mestrado em Biologia Humana e Ambiente. Faculdade de Ciências da Universidade de Lisboa, Lisboa. 107p.
- [21] Cole, T., Bellizzi, M., Flegal, K. & Dietz, W. (2000). Establishing a standard definition for child overweight and obesity worldwide: international survey. *BMJ*, 320(7244): 240-3.
- [22] Cole, T., Flegal, K., Nicholls, D. & Jackson, A. (2007). Body mass index cut offs to define thinness in children and adolescents: international survey. *BMJ*, 335:194.
- [23] The IPAQ Group (2015). *Guidelines for Data Processing and Analysis of the International Physical Activity Questionnaire*.
- [24] Rocha, C. DEIO-FCUL. *Apontamentos de Fundamentos de Bioestatística - Estatística não paramétrica*.
- [25] Antunes, M. (2010). *Prospecção de Dados: princípios e métodos. Texto de apoio às aulas de CRM e Prospecção de Dados*. Lisboa: Centro de Estatística e Aplicações da UL.
- [26] Podani J. (1999). Extending Gower's General Coefficient of Similarity to Ordinal Characters. *Taxon*, Vol. 48, No. 2, 331-340.
- [27] McCullagh, P. and Nelder, J.A. (1989). *Generalized Linear Models*. 2nd edition. London: Chapman and Hall.

- [28] Alvarenga, A. M. (2015). *Modelos lineares generalizados: aplicação a dados de acidentes rodoviários*. Dissertação de Mestrado em Gestão de Informação. Faculdade de Ciências da Universidade de Lisboa, Lisboa.
- [29] Cabral, S. M. DEIO-FCUL. *Modelo Linear Generalizado. Regressão Logística*.
- [30] Harrell, E. F. (2001). *Regression Modeling Strategies: With Applications to Linear Models, Logistic Regression, and Survival Analysis*. New York: Springer Science & Business Media.
- [31] Ferreira, M. C. (2013). *Modelos de Regressão: uma aplicação em Medicina Dentária*. Dissertação de Mestrado em Matemática, Estatística e Computação. Universidade Aberta, Lisboa.
- [32] Turkman, M. A., & Silva, G. L. (2000). *Modelos Lineares Generalizados - da teoria à prática*. Lisboa: Edições SPE.
- [33] Collett, D. (2003). *Modelling Binary Data* (2<sup>a</sup> ed.). Boca Raton, FL: CRC Press.
- [34] Hoaglin, D. & Welsch, R. (1978). The Hat Matrix in Regression and ANOVA. *The American Statistician*, 32(1): 17-22.
- [35] Hosmer, D. W., Lemeshow, S. & Sturdivant, R. (2013). *Applied Logistic Regression* (3<sup>a</sup> ed.). Hoboken, New Jersey: John Wiley & Sons, Inc.
- [36] Dobson, A. (2002). *An Introduction to generalized linear models* (2<sup>a</sup> ed.). Boca Raton, FL: Chapman and Hall/CRC.
- [37] Ari, E. & Yildiz, Z. (2014). Parallel lines assumption in ordinal logistic regression and analysis approaches. *International Interdisciplinary Journal of Scientific Research*, Vol. 1, No. 3, 8-23.
- [38] Lipsitz, S., Fitzmaurice, G. & Molenberghs, G. (1996). Goodness-of-fit tests for ordinal response regression models. *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, Vol. 45, No. 2, pp. 175-190.
- [39] Kieschnick, R. & McCullough, B.D. (2003). Regression analysis of variates observed on (0,1): percentages, proportions and fractions. *Statistical Modelling*, 3, 193-213.
- [40] Ferrari, S. & Cribari-Neto, F. (2004). Beta Regression for Modelling Rates and Proportions. *Journal of Applied Statistics*, Vol. 31, No. 7, 799-815.
- [41] Cribari-Neto, F. & Zeileis, A. (2010). Beta regression in R. *Journal of Statistical Software*, 34(2), 1-24.
- [42] Espinheira, P., Ferrari, S. & Cribari-Neto, F. (2008). Influence diagnostics in beta regression. *Computational Statistics and Data Analysis*, Vol. 52, No. 9, 4417-4431.
- [43] Wei, B.C., Hu, Y.Q. & Fung, W.K. (1998). Generalized leverage and its applications. *Scandinavian Journal of Statistics*, Vol. 25, No.1, 25-37.