

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Análise de dados de desemprego em Cabo Verde

Saturnino Sanches Borges

Mestrado em Estatística

2010

UNIVERSIDADE DE LISBOA

FACULDADE DE CIÊNCIAS

DEPARTAMENTO DE ESTATÍSTICA E INVESTIGAÇÃO OPERACIONAL



Análise de dados de desemprego em Cabo Verde

Saturnino Sanches Borges

Mestrado em Estatística

Orientadora: Prof. Dr^a. Maria Isabel Calisto Frade Barão

2010

RESUMO

Este trabalho baseia-se na análise de dados do desemprego em Cabo Verde nos anos de 2006 e 2008, usando informação da base de dados do INE e IEFP. Partindo da análise dos dados em estudo vai-se procurar descrever e perspectivar metodologias que contemplam as variáveis qualitativas e quantitativas com significado social positivo para a sociedade deste país.

Após a introdução no capítulo 1, fez-se, no capítulo 2, a análise exploratória dos dados do desemprego em Cabo Verde referente aos anos 2006 e 2008.

No capítulo 3 estudam-se associações entre variáveis, usando a metodologia de tabelas de contingência, através da realização de testes de independência e testes de homogeneidade, e análise de medidas de associação. As variáveis usadas, vão ser essencialmente, o escalão etário, o género e o ano.

O capítulo 4 é dedicado ao estudo de modelos Log - lineares em tabela de contingência, finalizando-se o trabalho com a apresentação das principais conclusões.

Palavras-chave: Desemprego, Cabo Verde, Tabelas de Contingência, Medidas de Associação, Modelos Log-lineares.

ABSTRACT

This research paper analyses data of unemployment in Cape Verde from the year 2006 to 2008, using the information from the National Institute of Statistics and Institute of Employment and Professional Training. Starting from the data analysis, the research will follow a methodological description including qualitative and quantitative approaches with positive social values for this country.

After the introduction in chapter 1, an exhaustive analysis was made in second chapter, on the unemployment rates in Cape Verde from 2006 to 2008.

In the third chapter, the association between several variables was discussed using the contingency table methodology through the accomplishment of tests of independence and tests of homogeneity, and analysis of measures of association. The variables used were age, gender and year.

The fourth chapter is dedicated to the study of the log – linear models in contingency tables, the work being finishing with the presentation of the main conclusions.

Key-words: Unemployment, Cape Verde, Contingency Tables, Measures of Association, Log-linear models.

AGRADECIMENTOS

O meu maior agradecimento dirige-se a minha orientadora Professora Doutora Maria Isabel Barão por todo o conhecimento, entusiasmo, espírito crítico e paciência que ofereceu ao acompanhar a realização da minha dissertação de mestrado.

À minha mãe, Francisca Sanches Horta e o meu pai, Tomas Sanches Borges, pelos valores nobres que inculcaram na minha pessoa e por tudo aquilo que por mim fizeram. Serei eternamente grato.

Ao Instituto Nacional de Estatística de Cabo Verde, agradeço a disponibilidade em ceder os dados originais dos inquéritos aos desempregados no ano 2006 e 2008, essenciais ao projecto de dissertação.

Desejo também expressar os meus agradecimentos aos meus amigos, especialmente a José Alves e Justimiano Andrade, bem como, os modos como sempre incentivaram-me.

Por fim, quero expressar todo carinho, aos meus irmãos, contribuindo assim de forma especial, para a conclusão do meu Mestrado.

ÍNDICE GERAL

RESUMO	i
ABSTRACT	ii
AGRADECIMENTOS	iii
ÍNDICE GERAL	iv
ÍNDICE DE FIGURAS	vi
ÍNDICE DE TABELAS	vii
1. INTRODUÇÃO	1
1.1. Tema	1
1.2. Objectivos.....	1
1.3. Plano	1
2. ANÁLISE PRELIMINAR DOS DADOS DE DESEMPREGO EM CABO VERDE NO ANO 2006 E 2008	3
2.1 A nível do domínio	3
2.2 A nível dos ramos de actividades.....	6
2.3 A nível da Situação na Profissão.....	11
2.4 A nível da profissão	15
3. TABELAS DE CONTINGÊNCIA	19
3.1 Tabelas de Contingência Bidimensionais.....	19
3.1.1 Teste de Independência.....	23
3.1.2 Teste de Homogeneidade.....	27
3.2 Medidas de Associação	29
3.2.1 Medidas Baseadas na Estatística χ^2	29
3.2.2 Medidas de Redução Proporcional do Erro de predição	35
3.3 Tabelas de Contingência Tridimensionais.....	47
3.3.1 Medidas de Associação Parcial	48
3.3.2 Medidas de Associação Múltipla.....	56
4. MODELOS LOG-LINEARES	59

4.1 Modelos log-lineares em tabelas bidimensionais	59
4.2 Modelo Log-linear em Tabelas com três variáveis	67
5. CONCLUSÃO.....	73
6. TABELAS DE DADOS REAIS.....	77
7. BIBLIOGRAFIA E REFERÊNCIAS.....	81
7.1. Cabo Verde	81
7.2. Geral.....	81

ÍNDICE DE FIGURAS

Figura 2.1: Diagrama circular da proporção de desemprego a nível de domínio.	3
Figura 2.2: Diagrama circular da proporção de desemprego a nível de domínio.	4
Figura 2.3: Diagrama de barra da comparação do desemprego a nível de domínio.	5
Figura 2.4: Diagrama circular da proporção de desemprego a nível de domínio.	7
Figura 2.5: Diagrama circular da proporção de desemprego a nível do ramo de actividade.	9
Figura 2.6: Diagrama de barra da comparação do desemprego a nível do ramo de actividade.	10
Figura 2.7: Diagrama circular da proporção de desemprego a nível da situação na profissão.	11
Figura 2.8: Diagrama circular da proporção de desemprego a nível da situação na profissão.	13
Figura 2.9: Diagrama de barra da comparação do desemprego a nível da situação na profissão.	14
Figura 2.10: Diagrama circular da proporção de desemprego a nível da profissão	16
Figura 2.11: Diagrama circular da proporção de desemprego a nível da profissão	17
Figura 2.12: Diagrama de barra da comparação do desemprego a nível da profissão..	18
Figura 5.5: Gráficos de normal Q-Q Plot of Adjusted Residuals e de Detrended Normal Q-Q of Adjusted Residuals.	72

ÍNDICE DE TABELAS

Tabela 2.1: Amostra disponível na base de dados a nível do domínio no ano 2006.	3
Tabela 2.2: Amostra disponível na base de dados a nível do domínio no ano 2008.	4
Tabela 2.3: Amostra disponível na base de dados a nível do ramo de actividade no ano 2006.....	6
Tabela 2.4: Amostra disponível na base de dados a nível do ramo de actividade no ano 2008.	8
Tabela 2.5: Amostra disponível na base de dados a nível da situação na profissão no ano 2006.....	11
Tabela 2.6: Amostra disponível na base de dados a nível da situação na profissão no ano 2008.....	12
Tabela 2.7: Amostra disponível na base de dados a nível da profissão no ano 2006 ...	15
Tabela 2.8: Amostra disponível na base de dados a nível da profissão no ano 2008 ...	17
Tabela 3.1: Tabela de contingência dos dados de desemprego em Cabo Verde no ano 2006	20
Tabela 3.2: Tabela em percentagens dos desempregados no ano 2006.....	21
Tabela 3.3: Tabela de contingência dos dados de desemprego em Cabo Verde no ano 2008	21
Tabela 3.4: Tabela em percentagens dos desempregados relativamente ao ano 2008	22
Tabela 3.5: Tabela de frequência esperada do desemprego em Cabo Verde no ano 2006	23
Tabela 3.6: Tabela de contribuição de χ^2 individuais referente aos dados de 2006.....	24
Tabela 3.7: Tabela de frequência esperada do desemprego em Cabo Verde no ano 2008.....	25
Tabela 3.8: Tabela de contribuição de χ^2 individuais referente aos dados de 2008 ...	26
Tabela 3.9: Amostra disponível na base de dados do desemprego no ano 2006 e 2008.	27
Tabela 4.1: Tabela das frequências observados e esperados de desemprego em Cabo Verde no ano 2006.....	61
Tabela 4.2: Tabela de K-Way and Higher-Order Effects para os dados de 2006.	61
Tabela 4.3: Tabela de Step summary para os dados de 2006.....	62

Tabela 4.4: Tabela de Cell Counts and Residuals para os dados de 2006.....	62
Tabela 4.5: Tabela de parameter estimates para os dados de 2006.	63
Tabela 4.6: Tabela de K-Way and Higher-Order Effects referente aos dados de 2008.	64
Tabela 4.7: Tabela de Step Summary referente aos dados de 2008.....	64
Tabela 4.8: Tabela de Cell Counts and Residuals referente aos dados de 2008	65
Tabela 4.9: Tabela de parameter estimates referente aos dados de 2008.	66
Tabela 5.1: Tabela de Step summary para tabela tridimensional.	69
Tabela 5.2: Tabela de K-Way and Higher-Order Effects para tabela tridimensional.	69
Tabela 5.3: Tabela de Parameter Estimates para tabela tridimensional.	70
Tabela 5.4: Tabela de Cell Counts and Residuals para tabela tridimensional.	71
Tabela 6.1: Dados dos desempregados e empregados a nível do domínio em Cabo Verde no ano 2006.....	77
Tabela 6.2: Dados dos desempregados e empregados a nível da profissão em Cabo Verde no ano 2006.	77
Tabela 6.3: Dados dos desempregados e empregados a nível da situação na profissão em Cabo Verde no ano 2006.....	78
Tabela 6.4: Dados dos desempregados e empregados a nível do ramo de actividade em Cabo Verde no ano 2006.....	78
Tabela 6.5: Dados dos desempregados e empregados a nível do domínio em Cabo Verde no ano 2008.....	79
Tabela 6.6: Dados dos desempregados e empregados a nível da profissão em Cabo Verde no ano 2008.....	79
Tabela 6.7: Dados dos desempregados e empregados a nível da situação na profissão em Cabo Verde no ano 2008.....	80
Tabela 6.8: Dados dos desempregados e empregados a nível do ramo de actividade em Cabo Verde no ano 2008.....	80

1. INTRODUÇÃO

O desemprego é um indicador do estado do mercado de trabalho de qualquer país. Cabo Verde cresceu de forma notável desde 1975, mas o desemprego continua sempre presente. A elevada taxa de desemprego é um dos maiores problemas da economia cabo-verdiana, cuja resolução entra como o primeiro ponto da agenda política dos vários governos. A sua medição, compreensão e estratégia de combate fazem parte dos principais documentos de referência do país.

1.1. Tema

O tema em estudo nesta dissertação de mestrado baseia-se na análise de dados do desemprego em Cabo Verde nos anos de 2006 e 2008. Partindo da análise de dados em estudo vai-se procurar descrever e perspectivar metodologias que contemplem as variáveis qualitativas e quantitativas com significado social positivo para a sociedade desse país. A questão central que se coloca é de saber quais as variáveis que estão mais decisivamente associadas ao desemprego em Cabo Verde referente aos anos 2006 e 2008.

1.2. Objectivos

Como o presente estudo pretendemos de um modo geral, analisar os dados do desemprego em Cabo Verde nestes dois anos, com intuito de conhecer a sua real situação. Constitui ainda objectivo deste estudo, em termos mais específicos, identificar as variáveis que estão associadas ao desemprego em Cabo Verde e descrever um modelo para essa relação.

1.3. Plano

Neste relatório de análise, e dada a natureza dos nossos dados no campo do desemprego em Cabo Verde, vamos começar, no capítulo 2, por fazer uma análise exploratória dos dados do desemprego em Cabo Verde referente aos anos 2006 e 2008, incluindo várias representações gráficas dos dados, seguida de uma comparação do desemprego entre o ano 2006 e 2008. No capítulo 3 estuda-se a tabela de contingência,

ilustra-se o conceito, medidas de associação, teste de independência e teste de homogeneidade. O capítulo 4 encarrega-se dos modelos log-lineares em tabela de contingência. Por fim, no capítulo 5 faz-se uma conclusão e indicam-se algumas pistas para o trabalho futuro.

2. ANÁLISE PRELIMINAR DOS DADOS DE DESEMPREGO EM CABO VERDE NO ANO 2006 E 2008

2.1 A nível do domínio

O presente estudo baseia-se na base de dados do inquérito a 33574 indivíduos desempregados a nível de domínio em Cabo Verde no ano 2006, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), os resultados encontram-se representados na seguinte tabela:

<i>Domínio</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>São Vicente</i>	<i>8179</i>	<i>24</i>	<i>24</i>
<i>Santo Antão</i>	<i>4487</i>	<i>13</i>	<i>37</i>
<i>Sal</i>	<i>860</i>	<i>3</i>	<i>40</i>
<i>Interior de Santiago</i>	<i>10364</i>	<i>31</i>	<i>71</i>
<i>Praia</i>	<i>7862</i>	<i>23</i>	<i>94</i>
<i>Fogo</i>	<i>1822</i>	<i>6</i>	<i>100</i>
<i>Total</i>	<i>33574</i>	<i>100</i>	

Tabela 2.1: Amostra disponível na base de dados a nível do domínio no ano 2006.

Para a análise da tabela 2.1, ilustra-se o procedimento de obtenção de uma representação de diagrama circular onde estão representados as proporções de desempregados a nível de domínio.

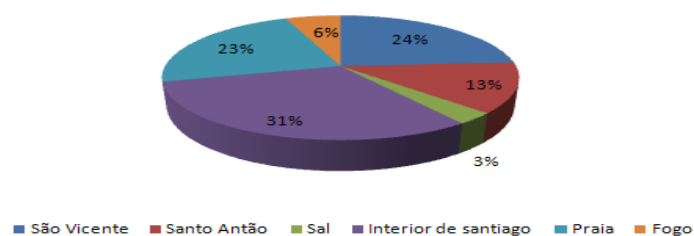


Figura 2.1: Diagrama circular da proporção de desemprego a nível de domínio.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível de domínio em Cabo verde no ano 2006.

Observando o gráfico verifica-se que a nível de domínio no interior de Santiago apresenta uma maior proporção de desempregados (31%) enquanto que na ilha de Sal verifica-se uma menor proporção de desempregados (3%).

Uma outra base de dados do inquérito a 35476 indivíduos desempregados a nível de domínio em Cabo Verde referente ao ano 2008, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), forneceu resultados da seguinte tabela:

<i>Domínio</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>São Vicente</i>	<i>3519</i>	<i>10</i>	<i>10</i>
<i>Santo Antão</i>	<i>9727</i>	<i>27</i>	<i>37</i>
<i>Sal</i>	<i>1430</i>	<i>4</i>	<i>41</i>
<i>Interior de Santiago</i>	<i>7034</i>	<i>20</i>	<i>61</i>
<i>Praia</i>	<i>11498</i>	<i>32</i>	<i>93</i>
<i>Fogo</i>	<i>2268</i>	<i>7</i>	<i>100</i>
<i>Total</i>	<i>35476</i>	<i>100</i>	

Tabela 2.2: Amostra disponível na base de dados a nível do domínio no ano 2008.

Para a análise da tabela 2.2, ilustra-se o procedimento de obtenção de uma representação de diagrama circular onde estão representados as proporções de desempregados a nível de domínio.

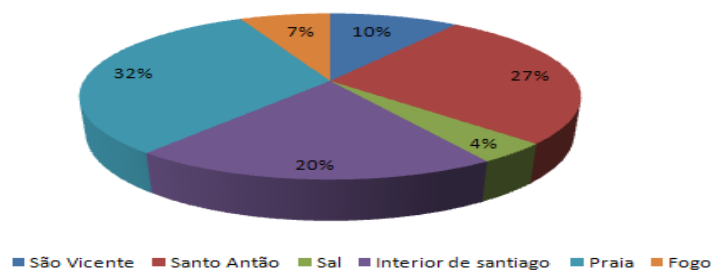


Figura 2.2: Diagrama circular da proporção de desemprego a nível de domínio.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível de domínio em Cabo Verde no ano 2008.

Observando o gráfico verifica-se que a nível de domínio a cidade da Praia apresenta uma maior proporção de desempregados (32%) enquanto que a ilha do Sal verifica-se uma menor proporção de desempregados (4%).

Comparação a nível de domínio

O estudo vai incidir no desemprego a nível de domínio no ano 2006 e 2008. Constrói-se o diagrama de barra na figura 2.3.

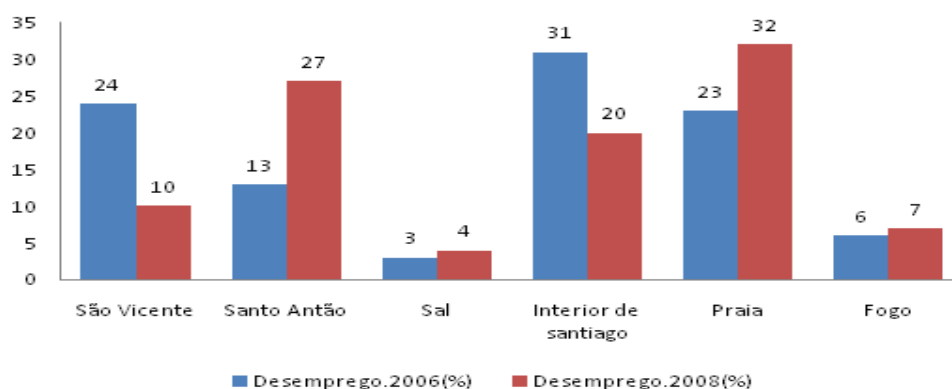


Figura 2.3: Diagrama de barra da comparação do desemprego a nível de domínio.

Da análise do gráfico da figura 2.3 verifica-se uma diferença entre as proporções de desempregados a nível de domínio do ano 2006 em relação ao ano 2008. A nível de domínio verifica-se que na cidade da Praia e na ilha de Santo Antão no ano 2008 obtiveram uma subida da proporção de desempregados em relação a 2006. A proporção de desemprego na ilha de Santo Antão foi de 13% no ano 2006, o que apresenta uma subida de 14% em relação a 2008.

A nível de domínio no ano 2006 a 2008 a taxa de desemprego desceu ligeiramente de 18,3% para 17,8%, uma variação de 0,5 %. (ver tabelas 6.1 e 6.5)

O nível de desemprego em Cabo Verde relativamente ao ano 2006 e 2008 apresenta disparidades acentuadas entre as diversas ilhas e concelhos do arquipélago. Pelo menos em duas ilhas, Sal e Fogo, a taxa de desemprego é inferior a 15%. Em sentido contrário encontra-se os concelhos da Praia, o Interior de Santiago, ilha de Santo Antão e São Vicente onde a taxa de desemprego ronda os 27,6%

2.2 A nível dos ramos de actividades

O outro estudo baseia-se na base de dados do inquérito a 18220 indivíduos desempregados a nível dos diferentes ramos de actividades em Cabo Verde no ano 2006, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), os resultados encontram-se representados na tabela em baixo.

<i>Ramo de actividade</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>Agricultura e Pesca</i>	967	5	5
<i>Indústria Extractiva</i>	217	1	6
<i>Indústria Transf. E Electricidade</i>	1596	9	15
<i>Construção</i>	3922	22	37
<i>Comércio</i>	1752	10	47
<i>Alojamento e Restauração</i>	1130	6	53
<i>Transportes e comunicações</i>	598	3	56
<i>Actividades Financeiras</i>	28	0	56
<i>Serviços às empresas</i>	468	3	59
<i>Administração pública</i>	3307	18	77
<i>Educação</i>	256	1	78
<i>Saúde</i>	33	0	78
<i>Outros serviços</i>	867	5	83
<i>Famílias com Empregados</i>	1986	11	94
<i>Organismo Internacionais</i>	196	1	95
<i>NR</i>	897	5	100
<i>Total</i>	18220	100	

Tabela 2.3: Amostra disponível na base de dados a nível do ramo de actividade no ano 2006.

Para a análise da tabela 2.3, ilustra-se o procedimento de obtenção de uma representação de diagrama circular onde estão representados as proporções de desempregados a nível do ramo de actividade.

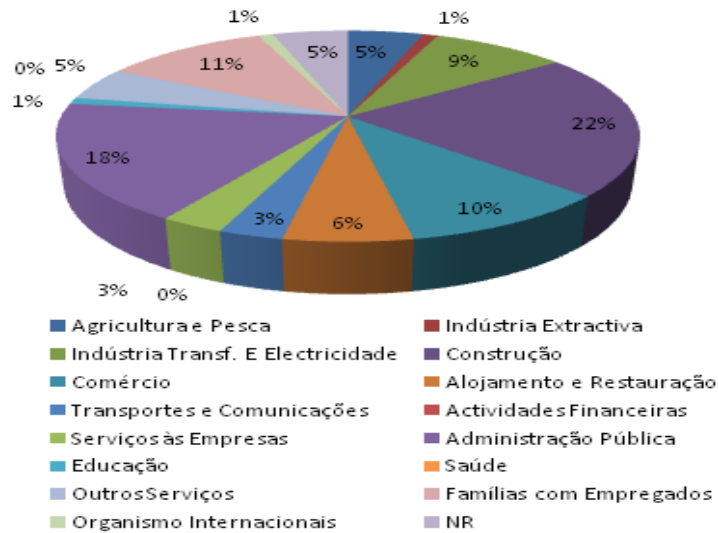


Figura 2.4: Diagrama circular da proporção de desemprego a nível de domínio.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível do ramo de actividade em Cabo Verde no ano 2006.

Da análise do gráfico podemos verificar que as proporções de desempregados a nível do ramo de actividade em Cabo Verde no ano 2006 apresentam uma grande diferença entre cada ramo.

O ramo da construção civil é o ramo que apresenta maior proporção de desempregados enquanto que nos ramos de actividade financeira e da saúde a proporção de desempregados é de 0%.

Uma outra base de dados do inquérito a 16646 indivíduos desempregados a nível dos diferentes ramos de actividades em Cabo Verde referente ao ano 2008, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), conduziu aos resultados que se encontram representados na tabela em baixo.

<i>Ramo de actividade</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>Agricultura e Pesca</i>	699	4	4
<i>Indústria Extractiva</i>	93	1	5
<i>Indústria Transf. E Electricidade</i>	645	4	9
<i>Construção</i>	2894	17	26
<i>Comércio</i>	1690	10	36
<i>Alojamento e Restauração</i>	682	4	40
<i>Transportes e comunicações</i>	558	3	43
<i>Actividades Financeiras</i>	125	1	44
<i>Serviços às empresas</i>	371	2	46
<i>Administração pública</i>	1104	7	53
<i>Educação</i>	236	1	54
<i>Saúde</i>	13	0	54
<i>Outros serviços</i>	131	1	55
<i>Famílias com Empregados</i>	1561	10	65
<i>Organismo Internacionais</i>	12	0	65
<i>NR</i>	5832	35	100
<i>Total</i>	16646	100	

Tabela 2.4: Amostra disponível na base de dados a nível do ramo de actividade no ano 2008.

Para a análise da tabela 2.4, ilustra-se o procedimento de obtenção de uma representação de diagrama circular onde estão representados as proporções de desempregados a nível do ramo de actividade.

Comparação a nível do ramo de actividade

A nível do ramo de actividade ilustra-se o procedimento de obtenção de uma representação de diagrama de barra na figura 2.6.

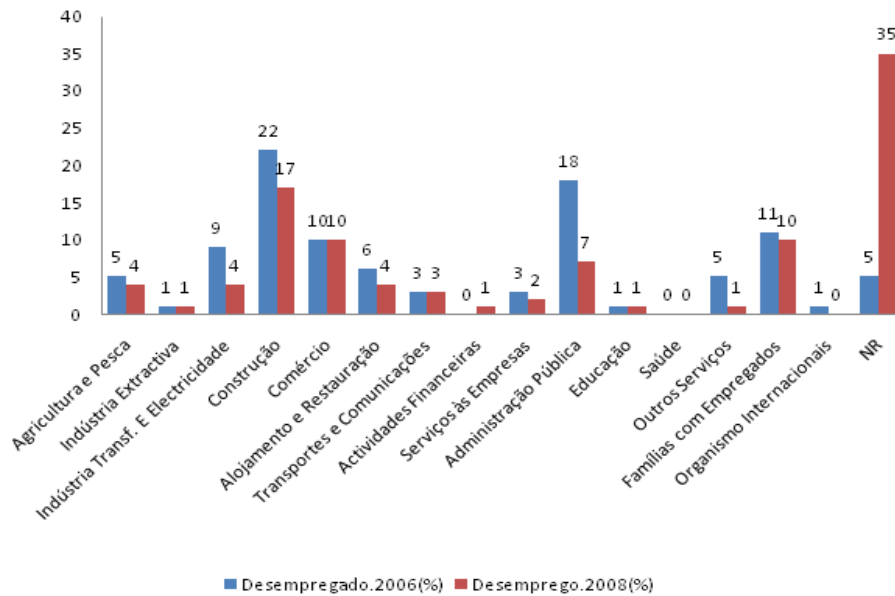


Figura 2.6: Diagrama de barra da comparação do desemprego a nível do ramo de actividade.

Observando o gráfico da figura 2.6 verifica-se que nas áreas das indústria extractiva, no comércio, nos transportes e comunicações, na educação e na saúde as proporções dos desempregados mantêm-se enquanto que no grupo onde os desempregados não estão inseridos em nenhum ramo de actividade (NR), a proporção de desempregados em 2008 teve uma subida em relação a 2006.

Da análise do gráfico verifica-se que a proporção de desempregados no ramo da administração pública em Cabo Verde foi de 7% no ano 2008, o que representou uma redução de 11% em relação a 2006. A nível da construção em 2008 houve também uma redução na proporção de desempregados em relação a 2006.

A nível de ramo de actividades a taxa de desemprego no ano 2006 era de 10.9% enquanto que no ano 2008 foi de 9.3%, o que desceu ligeiramente de 10.9% para 9.3%, uma variação de 1.6%. (ver tabelas 6.4 e 6.8)

2.3 A nível da Situação na Profissão

O seguinte estudo vai incidir na base de dados do inquérito a 18223 indivíduos desempregados a nível da situação na profissão em Cabo Verde no ano 2006, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), os resultados encontram-se representados na seguinte tabela:

<i>Situação na Profissão</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>Administração pública</i>	3730	20	20
<i>Conta Próp.c/Pes.ao Serv</i>	385	2	22
<i>Conta Próp.s/Pes.ao Serv</i>	2135	12	34
<i>Casa de Família</i>	2689	15	49
<i>Sector Empres.Privado</i>	6968	38	87
<i>Sector Empres.Estado</i>	330	2	89
<i>Outra Situação</i>	962	5	94
<i>Trab.Fam.Sem remun.</i>	458	3	97
<i>NR</i>	566	3	100
<i>Total</i>	18223	100	

Tabela 2.5: Amostra disponível na base de dados a nível da situação na profissão no ano 2006.

Para o estudo da tabela 2.5 constrói-se o diagrama circular onde estão representados as proporções de desempregados a nível da situação na profissão.

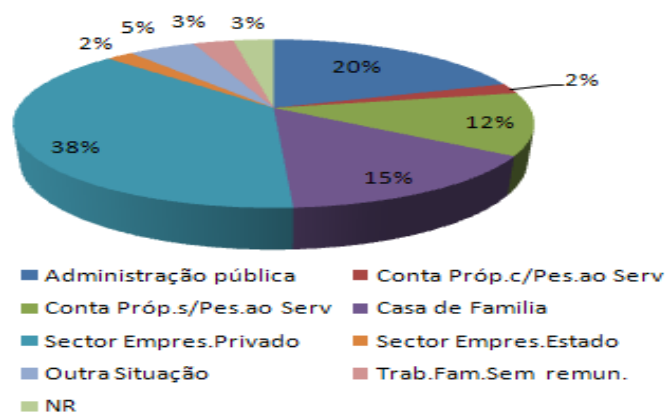


Figura 2.7: Diagrama circular da proporção de desemprego a nível da situação na profissão.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível da situação na profissão em Cabo Verde no ano 2006.

Observando o gráfico verifica-se que a nível da Conta Próp.c/Pes.ao Serv e do Sector Empres. Estado a proporção de desempregados é de 2% enquanto que a nível do Sector Empres. Privado verifica-se uma maior proporção de desempregados (38%).

Os resultados de uma outra base de dados do inquérito a 16721 indivíduos desempregados a nível da situação na profissão em Cabo Verde referente ao ano 2008, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), encontram-se representados na seguinte tabela:

<i>Situação na Profissão</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>Administração Pública</i>	<i>1564</i>	<i>9</i>	<i>9</i>
<i>Conta Próp.c/Pes.ao Serv</i>	<i>646</i>	<i>4</i>	<i>13</i>
<i>Conta Próp.s/Pes.ao Serv</i>	<i>787</i>	<i>5</i>	<i>18</i>
<i>Casa de Família</i>	<i>1634</i>	<i>10</i>	<i>28</i>
<i>Sector Empres.Privado</i>	<i>5543</i>	<i>33</i>	<i>61</i>
<i>Sector Empres.Estado</i>	<i>630</i>	<i>4</i>	<i>65</i>
<i>Outra Situação</i>	<i>278</i>	<i>2</i>	<i>67</i>
<i>Trab.Fam.Sem remun.</i>	<i>242</i>	<i>1</i>	<i>68</i>
<i>NR</i>	<i>5397</i>	<i>32</i>	<i>100</i>
<i>Total</i>	<i>16721</i>	<i>100</i>	

Tabela 2.6: Amostra disponível na base de dados a nível da situação na profissão no ano 2008

Para o estudo da tabela 2.6 constrói-se novamente o diagrama circular onde estão representados as proporções de desempregados a nível da situação na profissão.

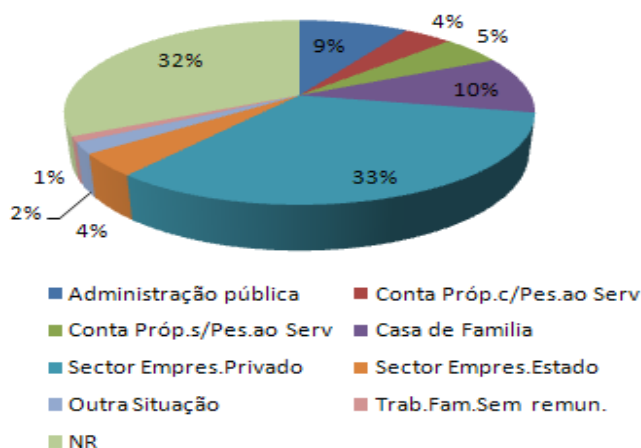


Figura 2.8: Diagrama circular da proporção de desemprego a nível da situação na profissão.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível da situação na profissão em Cabo Verde no ano 2006.

Observando o gráfico verifica-se que a nível da Conta Próp.c./Pes.ao Serv e do Sector Empres.Estado as proporção de desempregados são iguais, uma proporção de 4%, enquanto que a nível de Sector Empres.Privado verifica-se uma maior proporção de desempregados (33%).

Comparação a nível da situação na profissão

A nível da situação na profissão ilustra-se o procedimento de obtenção de uma representação de diagrama de barra na figura 2.9.

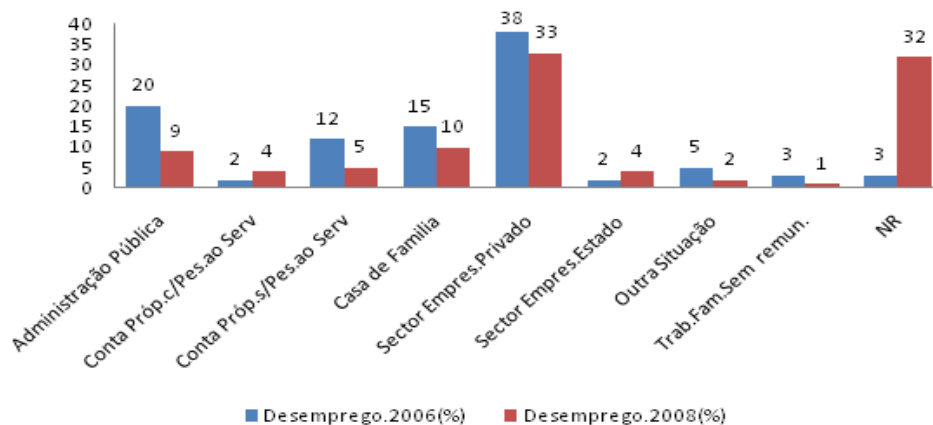


Figura 2.9: Diagrama de barra da comparação do desemprego a nível da situação na profissão.

Da análise do gráfico da figura 2.9 verifica-se uma diferença entre as proporções de desempregados a nível da situação na profissão do ano 2006 em relação ao ano 2008. A nível da administração pública, conta próp.s/pes.ao serv e casa de família verifica-se uma descida da proporção de desempregados no ano 2008 em relação ao ano 2006. A proporção de desemprego no grupo onde os desempregados não estão inseridos em nenhum ramo de actividade (NR) em Cabo Verde foi de 3% no ano 2006, o que apresenta uma subida de 29% em relação a 2008.

A nível da situação na profissão a taxa de desemprego no ano 2006 era de 10.9%, houve uma descida de 1.6% em relação ao ano 2008. (ver tabelas 6.3 e 6.7)

2.4 A nível da profissão

O estudo agora vai incidir na base de dados do inquérito a 18220 indivíduos desempregados a nível da profissão em Cabo Verde no ano 2006, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), os resultados encontram-se representados na seguinte tabela:

<i>Profissão</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>Forças Armadas</i>	<i>0</i>	<i>0</i>	<i>0</i>
<i>Q.sup.da Ap, Dir.e Q.Sup.</i>	<i>145</i>	<i>1</i>	<i>1</i>
<i>Especialistas</i>	<i>33</i>	<i>0</i>	<i>1</i>
<i>Técnico e Prof. De nível Intermédio</i>	<i>246</i>	<i>1</i>	<i>2</i>
<i>Pessoal Administrativo e Similares</i>	<i>911</i>	<i>5</i>	<i>7</i>
<i>Pessoal dos Serviços e Vendedores</i>	<i>1.980</i>	<i>11</i>	<i>18</i>
<i>Agric. e Trab.Qual. Da Agric. e Pescas</i>	<i>621</i>	<i>3</i>	<i>21</i>
<i>Operários, Artífices e Similares</i>	<i>3.947</i>	<i>22</i>	<i>43</i>
<i>Oper. De Inst. De Máq de Trab. Da Montagem</i>	<i>669</i>	<i>4</i>	<i>47</i>
<i>Trabalhadores não Qualificados</i>	<i>9.084</i>	<i>50</i>	<i>97</i>
<i>NR</i>	<i>584</i>	<i>3</i>	<i>100</i>
<i>Total</i>	<i>18220</i>	<i>100</i>	

Tabela 2.7: Amostra disponível na base de dados a nível da profissão no ano 2006.

Para o estudo da tabela 2.7 constrói-se o diagrama circular onde estão representados as proporções de desempregados a nível da profissão.

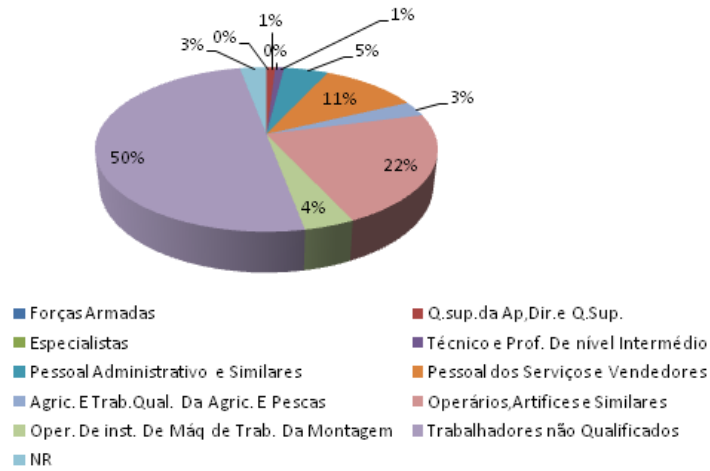


Figura 2.10: Diagrama circular da proporção de desemprego a nível da profissão.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível da profissão em Cabo verde no ano 2006.

Observando o gráfico verifica-se que a nível das forças armadas e especialistas a proporção de desempregados é de 0% enquanto que a nível de trabalhadores não qualificados verifica-se uma maior proporção de desempregados (50%).

Uma outra base de dados do inquérito a 16721 indivíduos desempregados a nível da profissão em Cabo Verde referente ao ano 2008, realizado pelo Instituto Nacional de Estatística (INE) e Instituto de Emprego e Formação Profissional (IEFP), forneceu resultados da seguinte tabela:

<i>Profissão</i>	<i>Frequência</i>	<i>Percentagem</i>	<i>Percentagem Acumulada</i>
<i>Forças Armadas</i>	38	0	0
<i>Q.sup.da Ap,Dir.e Q.Sup.</i>	0	0	0
<i>Especialistas</i>	156	1	1
<i>Técnico e Prof. De nível Intermédio</i>	48	0	1
<i>Pessoal Administrativo e Similares</i>	363	2	3
<i>Pessoal dos serviços e vendedores</i>	1936	12	15
<i>Agric. e Trab.Qual. Da Agric. e Pescas</i>	186	1	16
<i>Operários, Artífices e Similares</i>	1666	11	27
<i>Oper. De Inst. De Máq de Trab. Da Montagem</i>	327	2	29
<i>Trabalhadores não Qualificados</i>	6245	37	66
<i>NR</i>	5756	34	100
<i>Total</i>	16721	100	

Tabela 2.8: Amostra disponível na base de dados a nível da profissão no ano 2008.

Para o estudo da tabela 2.8 constrói-se novamente o diagrama circular onde está representado a proporção de desempregados a nível da profissão.

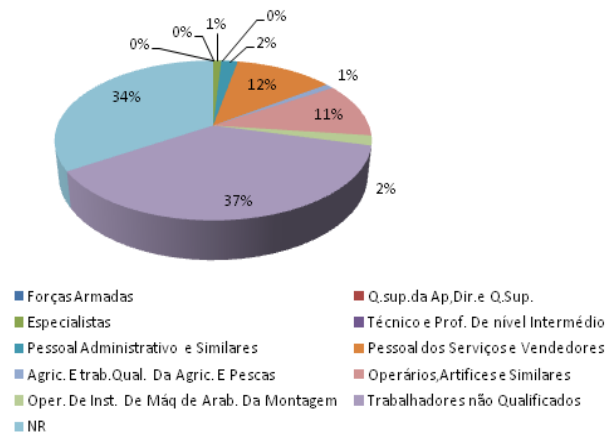


Figura 2.11: Diagrama circular da proporção de desemprego a nível da profissão.

Nesta representação juntamente com a identificação da categoria, indica-se as proporções de desempregados a nível da profissão em Cabo Verde no ano 2008.

Observando o gráfico da figura 2.11 verifica-se que a nível das forças armadas, “Q.Sup.da Ap,Dir.e Q.Sup” e a nível de Técnico e Profissionais de nível Intermédio a proporção de desempregados é de 0%, enquanto que a proporção de desempregados a nível de trabalhadores não qualificados é mais elevada.

Comparação a nível da profissão

O estudo vai incidir agora no desemprego a nível da profissão. Constrói-se novamente o diagrama barra.

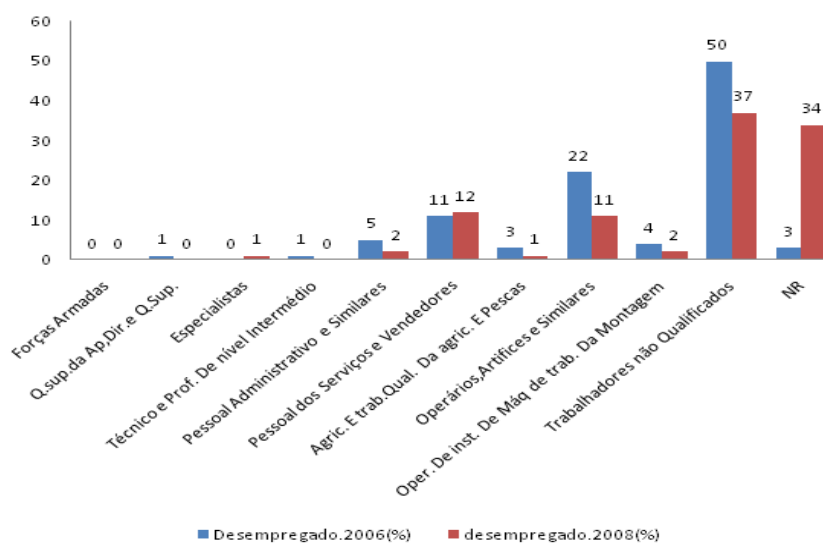


Figura 2.12: Diagrama de barra da comparação do desemprego a nível da profissão.

Da análise do gráfico da figura 2.12 verifica-se que a proporção de desempregados a nível das forças armadas mantém-se enquanto que no grupo onde os desempregados não estão inseridos em nenhum ramo de actividade (NR), a proporção de desempregados continua a ter uma subida em relação a 2006. A proporção de desempregados no ramo de trabalhadores não qualificados em Cabo verde foi de 37% no ano 2008, o que apresentou uma redução de 13% em relação a 2006. A nível de operários, artífices e similares verifica-se também uma redução na proporção de desempregados em relação a 2006.

A nível da profissão a taxa de desemprego no ano 2006 era de 10.9%, houve uma descida de 1.6% em relação ao ano 2008. (ver tabelas 6.2 e 6.6)

3. TABELAS DE CONTINGÊNCIA

3.1 Tabelas de Contingência Bidimensionais

Os indivíduos podem ser classificados por mais que uma variável qualitativa. Nesse caso, interessa estudar a associação entre todas as variáveis que levaram à classificação, sendo portanto necessário condensar toda a informação. Para tal, constroem-se tabelas de contingência.

Uma tabela de contingência é uma tabela de frequências que representa um conjunto de dados que foram classificados simultaneamente segundo várias características. Neste trabalho utilizamos tabelas de contingência bidimensionais para estudar a associação entre as variáveis de interesse para o nosso estudo.

Para os dados do desemprego em Cabo Verde, utilizamos a tabela de contingência bidimensional. Considera-se A e B duas variáveis com r e s categorias, respectivamente. A tabela de contingência resultante da classificação de n indivíduos de uma amostra, segundo estas variáveis, tem o seguinte aspecto:

	B_1	B_2	...	B_s	<i>Total</i>
A_1	n_{11}	n_{12}	...	n_{1s}	$n_{1.}$
A_2	n_{21}	n_{22}	...	n_{2s}	$n_{2.}$
.
.
.	.	.		.	
A_r	n_{r1}	n_{r2}	...	n_{rs}	$n_{r.}$
<i>Total</i>	$n_{.1}$	$n_{.2}$...	$n_{.s}$	n

Onde:

- A é Variável que representa *escalão etário* e B é a variável que representa o *género* com 6 e 2 categorias
- n_{ij} ($i = 1, \dots, 6; j = 1, 2$) representa a frequência observada da célula (i, j) .
- $n_{i.} = \sum_{j=1}^2 n_{ij}$, $i = 1, \dots, 6$ representa o total marginal da linha.

- $N_j = \sum_{i=1}^2 n_{ij}, j = 1,2$ representa o total marginal de coluna.
- $n = \sum_{i=1}^6 \sum_{j=1}^2 n_{ij} = \sum_{i=1}^6 n_i. = \sum_{j=1}^2 n_j$, representa o total dos desempregados.

A nossa tabela de contingência resulta de um inquérito realizado a 33575 indivíduos desempregados no ano 2006 em Cabo Verde, estes tiveram de assinalar o *género*, *masculino* e *feminino*, *escalão etário* -15 a 24 anos, 25 a 34 anos, 35 a 44 anos, 45 a 54 anos, 55 a 64 anos, 65 anos e mais (+). Para resumir a informação contida na amostra, construiu-se a seguinte tabela de contingência 6x2:

<i>Escalão Etário</i>	<i>Masculino</i>	<i>Feminino</i>	<i>Total</i>
<i>15 a 24 anos</i>	<i>7687</i>	<i>9845</i>	<i>17532</i>
<i>25 a 34 anos</i>	<i>2986</i>	<i>5436</i>	<i>8422</i>
<i>35 a 44 anos</i>	<i>1318</i>	<i>2999</i>	<i>4317</i>
<i>45 a 54 anos</i>	<i>869</i>	<i>1832</i>	<i>2701</i>
<i>55 a 64 anos</i>	<i>245</i>	<i>198</i>	<i>443</i>
<i>65 anos e +</i>	<i>104</i>	<i>56</i>	<i>160</i>
<i>Total</i>	<i>13209</i>	<i>20366</i>	<i>33575</i>

Tabela 3.1:Tabela de contingência dos dados de desemprego em Cabo Verde no ano 2006.

Da análise da tabela 3.1 podemos tirar algumas conclusões, tais como:

- 1- O *escalão etário* entre 15 a 24 anos apresenta maior número de desempregados.
- 2- O *género feminino* apresenta maior número de desempregados.
- 3- O *escalão etário* entre 65 anos e mais, apresenta menor número de desempregados.
- 4- Desemprego cresce com a classe etária.

Conforme o problema a estudar, a tabela pode ser completada com as percentagens referentes ao total da linha, coluna ou ao valor total.

Na tabela 3.1 em estudo, fará sentido acrescentar as percentagens para se comparar em cada grupo etário qual é a percentagem do desemprego em Cabo Verde no ano 2006 do género feminino. Apresenta-se uma nova tabela em percentagens:

Escalao.Etario ^ Género Crosstabulation

			Género		Total
			Feminino	Masculino	
Escalao.Etario	15 a 24 anos	Count	9845	7687	17532
		% within Escalao.Etario	56,2%	43,8%	100,0%
	25 a 34 anos	Count	5436	2986	8422
		% within Escalao.Etario	64,5%	35,5%	100,0%
	35 a 44 anos	Count	2999	1318	4317
		% within Escalao.Etario	69,5%	30,5%	100,0%
	45 a 54 anos	Count	1832	869	2701
		% within Escalao.Etario	67,8%	32,2%	100,0%
	55 a 64 anos	Count	198	244	442
		% within Escalao.Etario	44,8%	55,2%	100,0%
	65 anos e +	Count	56	104	160
		% within Escalao.Etario	35,0%	65,0%	100,0%
Total		Count	20366	13208	33574
		% within Escalao.Etario	60,7%	39,3%	100,0%

Tabela 3.2:Tabela em percentagens dos desempregados no ano 2006.

Na tabela 3.2, podemos então observar que no género feminino há uma maior proporção de desempregados no grupo etário entre 35 a 44 anos (69,5%) do que no género masculino (30,5%).

Observando agora uma outra tabela de contingência que resulta de um outro inquérito realizado a 34593 indivíduos desempregados em Cabo Verde referente ao ano 2008. E para resumir a informação contida na amostra, construiu-se a seguinte tabela de contingência 6x2:

<i>Escalão Etário</i>	<i>Masculino</i>	<i>Feminino</i>	<i>Total</i>
<i>15 a 24 anos</i>	<i>7980</i>	<i>10017</i>	<i>17997</i>
<i>25 a 34 anos</i>	<i>2914</i>	<i>5509</i>	<i>8423</i>
<i>35 a 44 anos</i>	<i>1532</i>	<i>3031</i>	<i>4563</i>
<i>45 a 54 anos</i>	<i>565</i>	<i>1811</i>	<i>2376</i>
<i>55 a 64 anos</i>	<i>228</i>	<i>642</i>	<i>870</i>
<i>65 anos e +</i>	<i>187</i>	<i>177</i>	<i>364</i>
<i>Total</i>	<i>13406</i>	<i>21187</i>	<i>34593</i>

Tabela 3.3:Tabela de contingência dos dados de desemprego em Cabo Verde no ano 2008.

Da análise da tabela podemos tirar algumas conclusões, tais como:

- 1- O grupo etário entre 15 a 24 anos apresenta maiores números de desempregados.
- 2- O género feminino apresenta maiores números de desempregado.
- 3- O grupo etário entre 65 anos e mais, apresenta menores números de desempregados.
- 4- Desemprego cresce com a classe etária.

Na tabela 3.3 em estudo, fará sentido acrescentar as percentagens para se comparar em cada grupo etário qual é a percentagem do desemprego em Cabo Verde do género feminino em 2008. Apresentam uma nova tabela em percentagens:

			Género		Total
			Masculino	Feminino	
Escalaio.Etario	15-24 anos	Count	7980	10017	17997
		% within Escalao.Etario	44,3%	55,7%	100,0%
	25-34 anos	Count	2914	5509	8423
		% within Escalao.Etario	34,6%	65,4%	100,0%
	35-44 anos	Count	1532	3031	4563
		% within Escalao.Etario	33,6%	66,4%	100,0%
	45-54 anos	Count	565	1811	2376
		% within Escalao.Etario	23,8%	76,2%	100,0%
	55-64 anos	Count	228	642	870
		% within Escalao.Etario	26,2%	73,8%	100,0%
	65 anos e +	Count	187	177	364
		% within Escalao.Etario	51,4%	48,6%	100,0%
Total		Count	13406	21187	34593
		% within Escalao.Etario	38,8%	61,2%	100,0%

Tabela 3.4: Tabela em percentagens dos desempregados relativamente ao ano 2008

Nesta Tabela 3.4, podemos então observar que há mais desempregados do género feminino em Cabo Verde no grupo etário entre 45 a 54 anos (76,2%) do que no grupo etário que varia entre 65 anos e +(48,6%).

3.1.1 Teste de Independência

Com objectivo de verificar se existe independência entre o *escalão etário* e o *género* dos desempregados em Cabo Verde no ano 2006, e denotando por A o atributo *escalão etário* e por B o atributo *género*, as hipóteses a testar são

$$H_0: A \text{ e } B \text{ são independentes} \quad \text{vs} \quad H_1: A \text{ e } B \text{ não são independentes}$$

Uma vez que A assume 6 modalidades e B assume 2 modalidades, sob H_0 , a estatística teste,

$$\chi_{rs}^2 = \sum_{i=1}^6 \sum_{j=1}^2 \frac{(N_{ij} - E_{ij})^2}{E_{ij}}$$

tem distribuição assintótica do Qui-quadrado com $(r - 1)(s - 1) = 5$ graus de liberdade, em que $E_{ij}(i = 1, \dots, 6; j = 1, 2)$ representa a frequência esperada da célula (i, j) sob a hipótese de independência.

Calculada a tabela de frequências esperadas

<i>Escalão Etário</i>	<i>Masculino</i>	<i>Feminino</i>	<i>Total</i>
<i>15 a 24 anos</i>	<i>6897</i>	<i>10635</i>	<i>17532</i>
<i>25 a 34 anos</i>	<i>3313</i>	<i>5109</i>	<i>8422</i>
<i>35 a 44 anos</i>	<i>1698</i>	<i>2619</i>	<i>4317</i>
<i>45 a 54 anos</i>	<i>1063</i>	<i>1638</i>	<i>2701</i>
<i>55 a 64 anos</i>	<i>174,3</i>	<i>268,7</i>	<i>443</i>
<i>65 anos e +</i>	<i>62,95</i>	<i>97,05</i>	<i>160</i>
<i>Total</i>	<i>13209</i>	<i>20366</i>	<i>33575</i>

Tabela 3.5: Tabela de frequência esperada do desemprego em Cabo Verde no ano 2006.

vem

$$\chi_{rs}^2 = \sum_{i=1}^6 \sum_{j=1}^2 \frac{(N_{ij} - E_{ij})^2}{E_{ij}} = 492.4$$

donde $\chi_{(obs)}^2 = 492,4$ e $p - value \leq 0.0005$. Neste caso rejeita-se a hipótese de independência entre as variáveis *escalão etário* e o *género* do desemprego em Cabo Verde no ano 2006 para todos os níveis de significância.

Este resultado é um resultado global: indica não independência entre as variáveis *A* e *B* do desemprego em Cabo Verde, mas não diz se é em relação a toda a tabela ou se só numa parte da tabela.

Podemos assim estar interessados em fazer comparações suplementares dentro da tabela. A maneira mais simples de estudar a importância dos desvios entre as frequências observadas e as esperadas é determinar as contribuições individuais à estatística χ_{rs}^2 . Essas contribuições vêm na tabela seguinte:

<i>Escalão Etário</i>	<i>Masculino</i>	<i>Feminino</i>
<i>15 a 24 anos</i>	<i>90,39189</i>	<i>58,63</i>
<i>25 a 34 anos</i>	<i>32,34393</i>	<i>20,98</i>
<i>35 a 44 anos</i>	<i>85,19404</i>	<i>55,26</i>
<i>45 a 54 anos</i>	<i>35,27992</i>	<i>22,88</i>
<i>55 a 64 anos</i>	<i>28,69308</i>	<i>18,61</i>
<i>65 anos e +</i>	<i>26,77438</i>	<i>17,37</i>

Tabela 3.6: Tabela de contribuição de χ^2 individuais referente aos dados de 2006.

E apercebemos imediatamente que a maior parte $\chi_{(5)(1)}^2 = 492.4$ tem origem na cela (1,1), daí verifica-se que no grupo etário entre 15 a 24 anos do género masculino é o grupo que mais contribui para associação das variáveis.

Continuando o estudo de independência para o ano 2008, obtiveram-se as frequências esperadas

<i>Escalão Etário</i>	<i>Masculino</i>	<i>Feminino</i>	<i>Total</i>
<i>15 a 24 anos</i>	<i>6974,47</i>	<i>11023</i>	<i>17997</i>
<i>25 a 34 anos</i>	<i>3264,21</i>	<i>5158,8</i>	<i>8423</i>
<i>35 a 44 anos</i>	<i>1768,32</i>	<i>2794,7</i>	<i>4563</i>
<i>45 a 54 anos</i>	<i>920,783</i>	<i>1455,2</i>	<i>2376</i>
<i>55 a 64 anos</i>	<i>337,155</i>	<i>532,84</i>	<i>870</i>
<i>65 anos e +</i>	<i>141,063</i>	<i>222,94</i>	<i>364</i>
<i>Total</i>	<i>13406</i>	<i>21187</i>	<i>34593</i>

Tabela 3.7: Tabela de frequência esperada do desemprego em Cabo Verde no ano 2008.

e

$$\chi_{rs}^2 = \sum_{i=1}^6 \sum_{j=1}^2 \frac{(N_{ij} - E_{ij})^2}{E_{ij}} = 656.2$$

donde $\chi_{(obs)}^2 = 656.2$ e $p - value \leq 0.0005$. Neste caso rejeita-se a hipótese de independência entre as variáveis *escalão etário* e o *género* do desemprego em Cabo Verde no ano 2008 para todos os níveis de significância.

Este resultado é um resultado global: indica não independência entre as variáveis *A* e *B* do desemprego em Cabo Verde no ano 2008, mas não diz se é em relação a toda a tabela ou se só numa parte da tabela.

Podemos assim estar interessados em fazer comparações suplementares dentro da tabela. A maneira mais simples de estudar a importância dos desvios entre as frequências observadas e as esperadas é determinar as contribuições individuais à estatística χ_{rs}^2 .

Essas contribuições vêm na tabela seguinte:

<i>Escalão Etário</i>	<i>Masculino</i>	<i>Feminino</i>
<i>15 a 24 anos</i>	<i>144,9707634</i>	<i>91,7297425</i>
<i>25 a 34 anos</i>	<i>37,57280915</i>	<i>23,7740633</i>
<i>35 a 44 anos</i>	<i>31,58264141</i>	<i>19,9838057</i>
<i>45 a 54 anos</i>	<i>137,471808</i>	<i>86,9848047</i>
<i>55 a 64 anos</i>	<i>35,33954529</i>	<i>22,3609734</i>
<i>65 anos e +</i>	<i>14,95951301</i>	<i>9,46557943</i>

Tabela 3.8: Tabela de contribuição de χ^2 individuais referente aos dados de 2008.

e vemos imediatamente que a maior parte $\chi^2_{(5)(1)} = 656.2$ tem origem na cela (1,1), o que verifica-se que no grupo etário entre 15 a 24 anos do género masculino é o grupo que mais contribui para associação das variáveis.

3.1.2 Teste de Homogeneidade

Com objectivo de efectuar uma comparação entre as proporções do desemprego a nível do domínio entre os anos 2006 e 2008, utilizou-se um inquérito realizado a 33574 indivíduos desempregados relativamente ao ano 2006 e outro inquérito realizado a 35476 indivíduos desempregados correspondente ao ano 2008, os resultados desses inquéritos encontram-se a seguir apresentados na tabela 3.9.

Domínio	Desemprego.2006	Desemprego.2008	Total
<i>São Vicente</i>	8179	3519	11698
<i>Santo Antão</i>	4487	9727	14214
<i>Sal</i>	860	1430	2290
<i>Interior de Santiago</i>	10364	7034	17398
<i>Praia</i>	7862	11498	19360
<i>Fogo</i>	1822	2268	4090
Total	33574	35476	69050

Tabela 3.9: Amostra disponível na base de dados do desemprego no ano 2006 e 2008.

Os dados da tabela correspondem as duas amostras, uma em cada linha da tabela, obtidas de modo independente e de dimensões respectivamente $n_1 = 11698, n_2 = 14214, n_3 = 2290, n_4 = 17398, n_5 = 19360, n_6 = 4090$. O número total de desempregados nesses dois anos, 2006 e 2008 corresponde a $N = 69050$.

Hipóteses a testar:

$$H_0: p_{i/2006} = p_{i/2008} \quad vs \quad H_1: p_{i/2006} \neq p_{i/2008}$$

As proporções do desemprego a nível do domínio no ano 2006 e 2008 são, $\hat{p}_{1/2006} = 0.7, \hat{p}_{1/2008} = 0.3, \dots, \hat{p}_{6/2006} = 0.45, \hat{p}_{6/2008} = 0.55$.

A tabela seguinte apresenta as frequências esperadas do desemprego em Cabo Verde no ano 2006 e 2008 sob o pressuposto de homogeneidade:

<i>Domínio</i>	<i>Desemprego.2006</i>	<i>Desemprego.2008</i>
<i>São Vicente</i>	<i>5687,887791</i>	<i>6010,112209</i>
<i>Santo Antão</i>	<i>6911,235858</i>	<i>7302,764142</i>
<i>Sal</i>	<i>1113,460681</i>	<i>1176,539319</i>
<i>Interior de Santiago</i>	<i>8459,383809</i>	<i>8938,616191</i>
<i>Praia</i>	<i>9413,361912</i>	<i>9946,638088</i>
<i>Fogo</i>	<i>1988,669949</i>	<i>2101,330051</i>
<i>Total</i>	<i>33574</i>	<i>35476</i>

À semelhança do teste de independência, a estatística do teste é

$$\chi^2 = \sum_{i=1}^6 \sum_{j=1}^2 \frac{(n_{ij} - \hat{e}_{ij})^2}{\hat{e}_{ij}} \text{ com } i = 1, \dots, 6 \text{ e } j = 1, 2$$

Que, sob o pressuposto de H_0 ser verdadeira, tem distribuição assintótica do Qui-Quadrado com $(r - 1)(s - 1)$ graus de liberdade em que $r = 6$ e $s = 2$.

As frequências observadas do desemprego em Cabo Verde a nível de domínio no ano 2006 e 2008 e as estimativas das frequências esperadas \hat{e}_{ij} calculadas sob o pressuposto de H_0 ser verdadeira, devem diferir pouco se H_0 for de facto verdadeira.

Assim, valores muito grandes da estatística teste traduzem um grande afastamento dos dados em relação à hipótese nula, conduzindo à rejeição desta. A estatística teste mede o afastamento dos dados em relação à hipótese de homogeneidade.

Cálculo do valor observado da estatística teste do desemprego em Cabo Verde a nível de domínio no ano 2006 e 2008:

$$\chi_{obs}^2 = 5250.43$$

O quantil de probabilidade de 0.95 da distribuição χ_5^2 é 11.071. Como o valor observado da estatística teste é $5250.43 > 11.071$ então, para um nível de

significância 0.05, rejeita-se a hipótese da homogeneidade do desemprego em Cabo Verde a nível do domínio no ano 2006 e 2008 isto é, a distribuição do desemprego a nível do domínio em 2006 não é semelhante à de 2008.

3.2 Medidas de Associação

De seguida serão apresentadas medidas de associação para tabela $r \times s$ no seu aspecto mais geral, ou seja, sem considerar nenhuma tabela em Particular.

3.2.1 Medidas Baseadas na Estatística χ^2

Um valor elevado de χ^2 no teste de independência sugere que as variáveis estão associadas, apesar de não quantificar essa associação procurou-se utilizar esta estatística como medida de associação.

χ^2 atinge o seu valor máximo em $n \times \min\{r - 1, s - 1\}$, este valor é obtida quando todas as frequências não nulas se situam na maior diagonal da tabela (tabelas $r \neq s$), correspondendo ao caso em que existe associação completa. Por exemplo, suponha que tem uma tabela $r \times s$, com $r < s$ e apenas os valores n_{ij} , com $i = j$, são diferentes de zero. Então vem

$$\chi^2 = n \times (r - 1)$$

Se em relação à tabela tiver $r > s$ ve

$$\chi^2 = n \times (s - 1)$$

Este resultado permite concluir que à medida que n aumenta o valor de χ^2 também aumenta, o seu intervalo de variação é $[0, +\infty[$.

Para usar esta estatística como medida de associação é necessário introduzir algumas modificações, procurou-se então alterar χ^2 com o objectivo de construir uma medida de associação com intervalo de variação limitado e que minimizasse a influência da dimensão da amostra.

Assim sendo, foram apontados vários coeficientes resultantes de diferentes alterações feitas sobre χ^2 , de entre os quais citamos alguns.

- **Coefficiente de Contingência em Média Quadrática**

$$\Phi^2 = \sum_{i=1}^r \sum_{j=1}^s \frac{(p_{ij} - p_i \cdot p_j)^2}{p_i \cdot p_j} = \sum_{i=1}^r \sum_{j=1}^s \frac{p_{ij}^2}{p_i \cdot p_j} - 1$$

O seu intervalo de variação depende de dimensionalidade de tabela:

$$0 \leq \Phi^2 \leq \min \{r - 1, s - 1\}$$

No caso de independência vem $\Phi^2 = 0$ e no caso de associação completa

$$\Phi^2 = \min \{r - 1, s - 1\}$$

Pelo facto do seu domínio de variação depender da dimensionalidade da tabela, Φ^2 não é particularmente adequado como medida de associação.

- **Coefficiente de Contingência de Pearson**

Um desses coeficientes é o apresentado por Pearson (1904):

$$P = \sqrt{\frac{2}{2 + 1}}$$

O coeficiente P varia no intervalo $[0,1]$, assumindo o valor 0 no caso de existir independência. Em geral, não atinge o seu valor máximo, falhando portanto um dos requisitos desejáveis dos coeficientes de associação. Na verdade P assume o valor máximo em $\sqrt{\frac{\min \{r-1, s-1\}}{1 + \min \{r-1, s-1\}}}$ no caso de associação completa, ou seja, o valor máximo que ele pode tomar depende da dimensionalidade da tabela, r e s , e só será 1 quando o número de categorias for finito.

Em geral Φ e P são muito semelhantes sendo Φ sempre maior que P .

- **Coefficiente de Contingência de Tshuprow**

Tshuprow (1919) propôs um coeficiente, definido por:

$$T = \sqrt{\frac{2}{\sqrt{(r-1)(s-1)}}}$$

Este coeficiente varia 0 e 1, assumindo o valor 0 no caso de existir independência. Só assume o valor 1 quando o número de linhas da tabela for igual ao número de colunas, $r = s$ e se houver associação completa.

- **Coefficiente de Contingência de Cramér**

Cramér(1946) propôs um coeficiente que atinge o valor 1 na situação de associação completa e 0 no caso de independência:

$$V = \sqrt{\frac{2}{\min(r-1, s-1)}}$$

Em geral, V aparece ao quadrado, ou seja:

$$V^2 = \frac{2}{\min(r-1, s-1)}$$

V^2 varia entre 0 e 1 e quanto mais forte é associação, maior é o seu valor.

- **Modificação de Sakoda ao Coeficiente de Contingência de Pearson**

Um dos pontos fracos do coeficiente de Contingência de Pearson é o facto de depender da dimensionalidade da tabela. Com o objectivo de suplantar esta dificuldade, Sakoda (1977) sugeriu a seguinte modificação ao coeficiente P :

$$P^* = \sqrt{\frac{2 \min(r, s)}{(1 + 2) \min(r-1, s-1)}}$$

P^* assume o valor 0 no caso de independência e o valor 1 no caso de associação completa.

Ao testar a hipótese de independência entre as variáveis *escalão etário* e o *género* do desemprego em Cabo Verde no ano 2006, verifica-se que

$$\chi^2 = 492.4$$

E que $p - value \leq 0.0005$. Portanto, para todos os níveis de significância foi rejeitada a hipótese de independência entre as variáveis em estudo.

Uma vez rejeitada a hipótese de independência convém tentar quantificar a associação existente entre *escalão etário* e o *género* do desemprego em Cabo Verde no ano 2006. Para tal serão calculadas as estimativas das medidas de associação entre *escalão etário* e o *género* do desemprego em Cabo Verde no ano 2006 baseadas na estatística χ^2 :

- **Coeficiente de Contingência em Média Quadrática**

$$\hat{\Phi}^2 = \frac{\chi^2}{n} = 0.014665483$$

Os estimadores de máxima verosimilhança das outras medidas são obtidos substituindo o Φ^2 por $\hat{\Phi}^2 = \frac{\chi^2}{n}$ em cada uma das expressões apresentadas. Portanto vem:

- **Coeficiente de Contingência de Pearson**

$$\hat{P} = \sqrt{\frac{\chi^2}{\chi^2 + n}} = 0.120222775$$

- **Coeficiente de Contingência de Tshuprow**

$$\hat{T} = \sqrt{\frac{\chi^2}{n\sqrt{(r-1)(s-1)}}} = 0.181088425$$

- **Coefficiente de Contingência de Cramér**

$$\hat{V} = \sqrt{\frac{\chi^2}{n \min(r-1, s-1)}} = 0.121101128$$

- **Modificação de Sakoda ao Coeficiente de Contingência de Pearson**

$$\hat{P}^* = \sqrt{\frac{\chi^2 \min(r, s)}{(n + \chi^2) \min(r-1, s-1)}} = 0.170020678$$

Para testar se estes valores são significativos é necessário construir intervalos de confiança e/ou realizar testes de hipóteses H_0 : coeficiente igual a zero. Alguns valores obtidos através do SPSS foram :

Symmetric Measures			
		Value	Approx. Sig.
Nominal by Nominal	Phi	,121	,000
	Cramer's V	,121	,000
	Contingency Coefficient	,120	,000
N of Valid Cases		33574	

O $p - value$, tem-se $P = P[rejH_0] \leq 0.0005 < \alpha = 0.05$, o que levam concluir que apesar de 'fraca', existe associação significativa entre o *escalão etário* e o *género*.

Ao testar novamente a hipótese de independência entre as variáveis *escalão etário* e o *género* do desemprego em Cabo Verde agora referente ao ano 2008, verifica-se que

$$\chi^2 = 656.2$$

e que $p - value \leq 0.0005$. Portanto, para todos os níveis de significância rejeita-se a hipótese de independência entre as variáveis.

Convém tentar quantificar a associação existente entre *escalão etário* e o *género* do desemprego em Cabo Verde referente ao ano 2008. Para tal serão calculadas a

estimação das medidas de associação entre *escalão etário* e o *género* do desemprego em Cabo Verde baseadas na estatística χ^2 :

- **Coefficiente de Contingência em Média Quadrática**

$$\hat{\Phi}^2 = 0.018969041$$

- **Coefficiente de Contingência de Pearson**

$$\hat{P} = 0.136440153$$

- **Coefficiente de Contingência de Tshuprow**

$$\hat{T} = 0.205951611$$

- **Coefficiente de Contingência de Cramér**

$$\hat{V} = 0.137728143$$

- **Modificação de Sakoda ao Coeficiente de Contingência de Pearson**

$$\hat{P}^* = 0.192955515$$

Para testar se estes valores são significativos continuamos a realizar testes de hipóteses. Podemos ver através do quadro em baixo o valor de *p – value*:

Symmetric Measures

		Value	Approx. Sig.
Nominal by Nominal	Phi	,138	,000
	Cramer's V	,138	,000
	Contingency Coefficient	,136	,000
N of Valid Cases		34593	

O *p – value*, *tem-se* $P = P[rejH_0] \leq 0.0005 < \alpha = 0.05$, o que nos leva outra vez a concluir que apesar de ser ‘fraca’, existe associação significativa entre o *escalão etário* e o *género*.

3.2.2 Medidas de Redução Proporcional do Erro de predição

Como já foi referido, a interpretação das medidas baseadas no Qui-quadrado só pode ser feitas em termos de afastamento entre frequências observadas e frequências esperadas, nenhuma delas tem interpretação probabilística.

Goodman e Kruskal consideraram que essas medidas tinham uma fraca interpretação ou nenhuma pelo que propuseram um outro tipo de medidas que contemplam este aspecto. Assim, propuseram a construção de modelos probabilísticos de predição, devendo cada modelo ser escolhido de acordo com a investigação em causa. Cada medida será uma probabilidade, ou uma função de probabilidade, calculada sob um dado modelo.

As medidas consideradas por Goodman e Kruskal são definidas em termos de redução da probabilidade de má classificação, em relação a uma das variáveis, resultante do conhecimento ou não da classificação do indivíduo em relação à outra variável. Cada uma das medidas quantifica a relativa utilidade de uma variável em melhorar a predição da classificação dos elementos de uma população, com respeito à outra variável.

Isto é, suponha que pretende predizer a categoria de uma das variáveis a que pertence um elemento da população, escolhido ao acaso, nas duas situações seguintes:

Caso 1: sem qualquer outra informação;

Caso 2: conhecida a sua categoria em relação à outra variável.

As medidas de associação propostas por Goodman e Kruskal vão quantificar a diferença entre estas duas situações, ou seja:

$$\text{Medidas} = \frac{P[\text{errar no Caso 1}] - P[\text{errar no Caso 2}]}{P[\text{errar no Caso 1}]} = \frac{P_1 - P_2}{P_1}$$

Estas medidas variam entre 0 e 1, onde 0 significa que o conhecimento de uma variável em nada ajuda a prever a outra variável, e 1 significa que o conhecimento de uma variável identifica perfeitamente a categoria da outra.

Medidas Baseadas na Predição Óptima

Estas medidas estão divididos em dois tipos de predição Óptima assimétrica e um tipo de predição simétrica.

Predição Óptima Assimétrica λ_b

Para esta primeira medida será considerado um modelo que se adapte às seguintes condições:

- i.* Duas classificações A e B ;
- ii.* Continuidade subjacente irrelevante;
- iii.* Ordem entre as categorias sem interesse;
- iv.* Existência de assimetria, em que a classificação A precede de algum modo a B .

Este modelo consiste em prever a categoria de B a que pertence um elemento da população, escolhido ao acaso, nas duas situações seguintes:

Caso 1: sem qualquer outra informação;

Caso 2: conhecida a sua categoria de A ;

Esta medida de associação é designada por λ_b e, tal como já foi referido, quantifica a diferença entre estas duas situações, sob a hipótese de que a predição consiste em seleccionar a categoria de B mais provável em cada ocasião.

Suponha que está no **caso 1**, ou seja, apenas conhece a distribuição marginal da variável B e, com esta informação, pretende estimar a que categoria, dessa variável, irá pertencer um elemento escolhido aleatoriamente da população. Uma regra de classificação aceitável diz que o elemento irá pertencer à categoria da variável B com maior probabilidade, isto é

$$\begin{aligned} P[\text{um elemento ser classificado na categoria de } B \text{ mais provável}] &= \max p_j \\ &= p_{\alpha} \text{ com } j = 1, \dots, s \end{aligned}$$

em que

$$P_j = P[\text{um elemento pertencer à categoria } B_j]$$

(probabilidade marginal da j -ésima coluna da tabela de probabilidade)

A classe mais provável é a classe B_j de B tal que $P_{.j} = P_{.α}$ e a probabilidade de erro será dada por:

$$P_1 = P[\text{errar no caso 1}] = 1 - p_{.α}$$

Para o **caso 2** suponha que sabe que o elemento pertence à categoria i da variável A . Aplicando a regra de classificação às probabilidades da linha i da tabela obtém-se

$$P[\text{um elemento ser classificado na categoria } A_i \text{ e na categoria de } B \text{ mais provável}] \\ = \max p_{ij} = p_{iα} \text{ com } j = 1, \dots, s$$

A categoria mais provável de B será B_j tal que $p_{ij} = p_{iα}$ e a probabilidade de erro, neste caso, será dada por

$$P_2 = P[\text{errar no Caso 2}] = 1 - \sum_{i=1}^r P_{iα}$$

Porque pelo cálculo da seguinte probabilidade condicional vem

$$\frac{P[\text{ser classificado na categoria de } B \text{ mais provável e pertencer à categoria } A_i]}{P[\text{pertencer à categoria } A_i]} \\ = \frac{P_{iα}}{P_i}$$

a correspondente probabilidade condicional de errar é $1 - \frac{p_{iα}}{p_i}$ e a respectiva probabilidade não condicional de erro é (recorrendo ao Teorema das Probabilidades Totais):

$$P_2 = \sum_{i=1}^r P[\text{errar sabendo que à categoria } A_i] P[\text{pertencer à categoria } A_i] \\ = \sum_{i=1}^r (1 - \frac{p_{iα}}{p_i}) p_i = 1 - \sum_{i=1}^r p_{iα}$$

Posto isto, a medida λ_b é definida por:

$$\frac{P_1 - P_2}{P_1} = \frac{\sum_{i=1}^r p_{i\alpha} - p_{. \alpha}}{1 - p_{. \alpha}}$$

e representa o decréscimo relativo da probabilidade de errar ao predizer a categoria B_j de um indivíduo quando a sua categoria A_i de A não é conhecida e quando passa a sê-lo, ou seja, dá a proporção de erros que se podem eliminar ao predizer a categoria B_j de um indivíduo quando se tem conhecimento prévio da sua categoria segundo a variável A .

A medida λ_b possui as seguintes propriedades:

1. λ_b assume valores ao intervalo $[0,1]$.
2. λ_b é indeterminado se todos os elementos da população são classificados numa e numa só categoria de B , ou seja, se as probabilidades não nulas apresentam-se numa coluna da tabela.
3. $\lambda_b = 0$ se o conhecimento da categoria de A não auxilia a predição da categoria de B , isto é, se as probabilidades, máximas de cada linha da tabela estão todas na mesma coluna.
4. $\lambda_b = 1$ se o conhecimento da categoria de A a que um indivíduo pertence especifica completamente a sua classe de B , isto é, se cada linha da tabela contém no máximo uma célula diferente de 0.
5. A independência estatística entre A e B implica $\lambda_b = 0$. O recíproco pode não se verificar, isto é, pode ser nulo sem que haja independência.
6. λ_b não se altera com permutações de linhas ou colunas da tabela.

O facto de λ_b poder ser nulo sem que haja independência estatística poderá ser apontado como uma desvantagem desta medida. No entanto, Goodman e Kruskal consideram que λ_b mede a associação em sentido bastante restrito, seguindo uma determinada interpretação probabilística e, o facto de não existir associação neste sentido, não impede que haja noutros. Dai que seja lógico pretender-se que, nesses casos, tome o valor 0, embora não haja independência estatística.

Predição Óptima Assimétrica λ_a

Ainda dentro da predição óptima assimétrica, pode-se considerar um outro modelo probabilístico. Existem situações em que se tem todo o interesse em usar o conhecimento sobre B para melhorar a predição de A . Para estas situações verificam-se as condições (i), (ii) e (iii) apresentadas para medidas λ_b e a condição (iv) é:

(iv) existência de assimetria, em que a classificação B precede de algum modo a A .

Este modelo consiste em prever a categoria de A a que pertence um indivíduo da população, escolhido ao acaso, nos dois casos:

Caso 1: sem qualquer outra informação.

Caso 2: conhecida a sua categoria de B .

Como facilmente se reconhece, este modelo é igual ao anterior a menos de uma troca entre as duas variáveis A e B . Trocando as regras de B para A no cálculo de λ_b , vem:

$$\lambda_a = \frac{\sum_{j=1}^s p_{\beta j} - p_{\beta}}{1 - p_{\beta}}$$

em que λ_a representa o decréscimo relativo na probabilidade de errar ao prever a categoria A_i de um indivíduo quando a sua categoria B_j não é conhecida e quando passa a sê-lo, ou seja, dá a proporção de erros que se podem eliminar ao prever a categoria A_i de um indivíduo quando se tem conhecimento prévio da sua categoria B_j .

Predição Óptima Simétrica λ

Para situações em que se verificam as condições (i), (ii) e (iii) dos dois modelos anteriores e existe simetria entre as variáveis, Goodman e Kruskal (1954) propuseram a medida λ que surge do seguinte problema de predição:

Suponha que metade das vezes o objectivo é estimar a categoria de A (aleatoriamente) e a outra metade das vezes o objectivo é estimar a categoria de B (aleatoriamente) de um indivíduo escolhido de uma população, nas situações seguintes:

Caso 1: sem qualquer informação.

Caso 2: conhecendo a categoria do indivíduo segundo uma das variáveis quando estamos a prever a sua categoria segundo a outra, ou seja, conhecendo A_i quando estamos a prever B_j e vice-versa.

Tendo em conta os modelos de predição óptimo assimétricos, facilmente se compreende que a probabilidade de errar no caso 1 é:

$$P_1 = 1 - \frac{p_{\alpha} - p_{\beta}}{2}$$

Metade das vezes, a categoria mais provável será a B_j de B tal que

$$p_{.j} = p_{\alpha} = \max p_{.j} \text{ com } j = 1, \dots, s$$

E, na outra metade das vezes, a categoria mais provável será a A_i de A tal que

$$p_{i.} = p_{\beta} = \max p_{i.} \text{ com } i = 1, \dots, r$$

Utilizando o mesmo raciocínio, conclui-se que a probabilidade de errar no caso 2 é dada por:

$$P_2 = 1 - \frac{\sum_{i=1}^r p_{i\alpha} + \sum_{j=1}^s p_{\beta j}}{2}$$

A medida de associação λ , que representa o decréscimo da probabilidade de errar quando se passa do caso 1 para caso 2, é dada por:

$$\lambda = \frac{\sum_{i=1}^r p_{i\alpha} + \sum_{j=1}^s p_{\beta j} - p_{\alpha} - p_{\beta}}{2 - p_{\alpha} - p_{\beta}}$$

As propriedades de λ são:

1. λ assume valor no intervalo $[0,1]$, estando sempre entre λ_a e λ_b .
2. λ é indeterminado se todos os elementos da população são classificados numa só célula da tabela.
3. $\lambda = 0$ se houver independência estatística, mas o recíproco pode não se verificar.

4. $\lambda = 1$ se o conhecimento da categoria de A a que um indivíduo pertence especifica completamente a sua classe de B e vice-versa, ou seja, em cada linha e em cada coluna há apenas uma célula diferente de 0.
5. λ não se altera com permutações de linhas ou colunas da tabela.

Como ilustração destas medidas baseadas na predição óptima pretendemos verificar se é possível usar *escalão etário* (variável A) para predizer o *género* (variável B) e vice-versa em relação ao desemprego em Cabo Verde no ano 2006. Para tal serão calculadas as medidas de predição óptima.

Comecemos por calcular a estimativa da medida de predição óptima assimétrica λ_b :

$$L_b = \frac{\sum_{i=1}^6 R_{i\alpha} - R_{\cdot\alpha}}{1 - R_{\cdot i}} = \frac{\sum_{i=1}^6 N_{i\alpha} - N_{\cdot\alpha}}{n - N_{\cdot\alpha}} = 0.0072$$

em que N_{ij} é a variável aleatória que representa numero dos desempregados nas categorias A_i de A e B_j de B . Os N_{ij} são variáveis aleatórias Binomiais de parâmetros n e p_{ij} . R_{ij} é a proporção de desempregados nas categorias A_i de A e B_j de B da tabela, ou seja, $R_{ij} = \frac{N_{ij}}{n}$.

$$e \ n = \sum_{i=1}^6 \sum_{j=1}^2 N_{ij}$$

$$\sum_{i=1}^6 \sum_{j=1}^2 R_{ij} = 1$$

Os totais marginais são:

AS proporções marginais são:

$$\begin{cases} N_{\cdot i} = \sum_{j=1}^2 N_{ij} \\ N_{\cdot j} = \sum_{i=1}^6 N_{ij} \end{cases}$$

$$\begin{cases} R_{\cdot i} = \sum_{j=1}^2 R_{ij} \\ R_{\cdot j} = \sum_{i=1}^6 R_{ij} \end{cases}$$

$$N_{i\alpha} = \max N_{ij} \quad \text{com } j = 1,2$$

$$R_{i\alpha} = \max R_{ij} \quad \text{com } j = 1,2$$

$$N_{\beta j} = \max N_{ij} \quad \text{com } i = 1, \dots, 6$$

$$R_{\beta j} = \max R_{ij} \quad \text{com } i = 1, \dots, 6$$

$$N_{\cdot\alpha} = \max N_{\cdot j} \quad \text{com } j = 1,2$$

$$R_{\cdot\alpha} = \max R_{\cdot j} \quad \text{com } j = 1,2$$

$$N_{\beta\cdot} = \max N_{ij} \quad \text{com } i = 1, \dots, 6$$

$$R_{\beta\cdot} = \max R_{i\cdot} \quad \text{com } i = 1, \dots, 6$$

A estimativa da proporção de erros que se podem eliminar ao predizer o *género* quando se tem conhecimento prévio do *escalão etário* do desemprego em Cabo Verde no ano 2006 é de 0,72%. Como $p_{.α} \neq 1$ e λ_b é diferente de 0 e 1, vem que L_b tem distribuição aproximadamente normal com valor médio λ_b e desvio padrão estimado:

$$\sigma(L_b) = \sqrt{\frac{(n - \sum_{i=1}^r n_{i\alpha})(\sum_{i=1}^r n_{i\alpha} + n_{i\alpha} - 2 \sum_{i=1}^c n_{i\alpha})}{(n - n_{.α})^3}} = 0.0019$$

onde $\sum_{i=1}^c n_{i\alpha}$ é a soma de todas as frequências máximas dos grupos etário que ocorrem no mesmo género.

O intervalo de confiança a 95% para a estimativa da medida de predição óptima assimétrica λ_b do desemprego em Cabo Verde no ano 2006 é [0.00356, 0.01082]. Este resultado leva a concluir que apesar de ‘fraca’ a associação é significativa.

Por outro lado, poderíamos estar interessados em predizer o *escalão etário* tendo conhecimento prévio do *género*. Nesse caso, seria necessário calcular a medida de predição óptima assimétrica λ_a , cuja estimativa é:

$$L_a = \frac{\sum_{j=1}^s n_{\beta j} - n_{\beta.}}{n - n_{\beta.}} = \frac{(687 + 9845) - 17532}{33575 - 17532} = 0$$

Este valor indica que não existe decréscimo relativo na probabilidade de errar ao predizer o *escalão etário* quando o *género* dos desempregados em Cabo Verde não é conhecido e quando passa a sê-lo. Ao contrário das medidas apresentadas anteriormente, este valor não significa que exista independência entre as variáveis (pois, como vimos, essa hipótese é rejeitada).

Como é óbvio, também é possível calcular a medida de predição óptima simétrica, λ , cuja estimativa é:

$$L = \frac{\sum_{i=1}^r n_{i\alpha} + \sum_{j=1}^s n_{\beta j} - n_{\alpha} - n_{\beta}}{2n - n_{\alpha} - n_{\beta}}$$

$$= \frac{(33575 - 20366) \times L_b + (33575 - 17532) \times L_a}{2 \times 33575 - 20366 - 17532} = 0.0033$$

O Valor de L está compreendido entre o valor de L_a e o valor de L_b .

A distribuição de amostragem de L é assintoticamente normal. Contudo, para a estimativa do desvio padrão é necessário introduzir a seguinte notação adicional:

- $n_{*\alpha}$ representa o $n_{i\alpha}$ que se encontra no mesmo grupo etario do desemprego em Cabo Verde que o máximo dos n_i ;
- $n_{\beta*}$ representa o $n_{\beta j}$ que se encontra no mesmo género dos desempregados em Cabo Verde que o máximo dos n_j ;
- n_{**} representa o $n_{i\alpha}$ que se encontra no mesmo grupo etario dos desempregados em Cabo Verde que o máximo dos n_i e no mesmo género do desemprego em Cabo verde que o máximo dos n_j .
- $\sum_{i=1}^* n_{i\alpha}$ representa a soma de todas as colunas (linhas) máximas da tabela do desemprego em Cabo Verde, que também são linhas (colunas) máximas, isto é, é a soma de todos os n_{ij} que verificam $n_{ij} = n_{i\alpha} = n_{\beta j}$.

E considere ainda:

- $U_1 = \frac{1}{n} (n_{\alpha} + n_{\beta})$
- $U_2 = \frac{1}{n} (\sum_{i=1}^r n_{i\alpha} + \sum_{j=1}^s n_{\beta j})$
- $U_3 = \frac{1}{n} (\sum_{i=1}^c n_{i\alpha} + \sum_{j=1}^d n_{\beta j} + n_{\beta*} + n_{*\alpha})$

Desde que λ não seja igual a 0 ou a 1, Goodman e Kruskal (1963) mostraram que para um n suficientemente grande, L é uma variável aproximadamente normal com valor médio λ e variancia estimada por

$$\hat{\sigma}^2(L) = \frac{U}{n(2 - U_1)^4}$$

onde

$$U = (2 - U_1)(2 - U_2)(U_1 + U_2 + 4 - 2U_3) - 2(2 - U_1)^2 \left(1 - \frac{1}{n} \sum_{i=1}^* n_{i\alpha} \right) - 2(2 - U_2)^2 \left(1 - \frac{n_{**}}{n} \right)$$

A estimativa do desvio padrão é:

$$\sigma^2(L) = 0.01108$$

e o intervalo de confiança se 95% para λ é $[-0.0185, 0.02497]$. Mas, como o intervalo de variação desta medida não está compreendido entre 0 e 1, é necessário truncar esse intervalo e passar a considerá-lo:

$$[0, 0.02497]$$

Com base em todos estes resultados pode-se confirmar que a proporção de erros que se podem eliminar ao predizer uma variável quando se tem conhecimento prévio da outra é muito pequena (num dos casos é nula). Este resultado permite concluir que a associação existente entre o *escalão etário* e o *género* dos desempregados em Cabo Verde no ano 2006 (neste caso em termos de predição) é muito ‘fraca’.

Vamos prosseguir para o ano de 2008 pretendendo agora verificar se é possível usar *escalão etário* (variável *A*) para predizer o *género* (variável *B*) e vice-versa em relação ao desemprego em Cabo Verde referente ao ano 2008. Para tal serão calculadas as medidas de predição óptima.

Começemos por calcular a estimativa da medida de predição óptima assimétrica λ_b :

$$L_b = 0.00075$$

A estimativa da proporção de erros que se podem eliminar ao predizer o *género* quando se tem conhecimento prévio do *escalão etário* é de 0,075%. Como $p_{\alpha \neq 1}$, e λ_b é diferente de 0 e 1, no entanto vem que tem distribuição aproximadamente normal com valor médio λ_b e desvio padrão estimado:

$$\sigma(L_b) = 0.00596$$

O intervalo de confiança a 95% para a estimativa da medida de predição óptima assimétrica λ_b do desemprego em Cabo Verde referente ao ano 2008 é $[-0.0109, 0.0124]$.

Mas, como o intervalo de variação desta medida não está compreendido entre 0 e 1, é necessário truncar esse intervalo e passar a considerar como intervalo de confiança a 95% para λ_b o seguinte intervalo:

$$[0, 0.0124].$$

Por outro lado, poderíamos estar interessados em predizer o *escalão etário* tendo conhecimento prévio do *género*. Nesse caso, seria necessário calcular a medida de predição óptima assimétrica λ_a , cuja estimativa é:

$$L_a = \frac{(7980 + 10017) - 17997}{345593 - 17997} = 0$$

Este valor indica que não existe decréscimo relativo na probabilidade de errar ao predizer o *escalão etário* quando o *género* dos desempregados em Cabo Verde no ano 2008 não é conhecido e quando passa a sê-lo. Ao contrário das medidas apresentadas anteriormente, este valor não significa que exista independência entre as variáveis (pois, como vimos, essa hipótese é rejeitada).

Como é óbvio, também é possível calcular a medida de predição óptima simétrica, λ , cuja estimativa é:

$$L = \frac{(34593 - 21187) \times L_b + (34593 - 17997) \times L_a}{2 \times 34593 - 21187 - 17997} = 0.0033$$

O Valor de L está compreendido entre o valor de L_a e o valor de L_b .

A estimativa do desvio padrão é:

$$\sigma^2(L) = \sqrt{\frac{U}{n(2 - U_1)^4}} = 0.01108$$

O intervalo de confiança se 95% para λ é $[-0.00091, 0.00158]$. Mas, como o intervalo de variação desta medida não está compreendido entre 0 e 1, é necessário truncar esse intervalo e passar a considerar como intervalo de confiança a 95% para λ o seguinte intervalo:

$$[0, 0.00158]$$

Com base em todos estes resultados pode-se concluir que a proporção de erros que se podem eliminar ao predizer uma variável quando se tem conhecimento prévio da outra é muito pequena (num dos casos é nula). Este resultado permite também concluir que a associação existente entre as variáveis (neste caso em termos de predição) é muito 'fraca'.

3.3 Tabelas de Contingência Tridimensionais

Considera-se A , B e C três variáveis com r , s e t categorias, respectivamente. A tabela de contingência resultante da classificação dos n indivíduos de uma amostra, segundo estas variáveis, tem o seguinte aspecto:

		C_1			...		C_t		
	B_1	B_2	...	B_s	...	B_1	B_2	...	B_s
A_1	n_{111}	n_{121}	...	n_{1s1}	...	n_{11t}	n_{12t}	...	n_{1st}
A_2	n_{211}	n_{221}	...	n_{2s1}	...	n_{21t}	n_{22t}	...	n_{2st}
.
.
.
A_r	n_{r11}	n_{r21}	...	n_{rs1}	...	n_{r1t}	n_{r2t}	...	n_{rst}

No nosso caso, tem-se uma tabela de contingência tridimensional com 6 linhas, 2 colunas, 2 estratos e $6 \times 2 \times 2$ células. Em que A é a variável que representa *escalonamento*; B é a variável que representa *género*; C é a variável que representa *Ano*; n_{ijk} é a variável que representa o número de elementos da amostra classificados simultaneamente nas categorias A_i de A , B_j de B e C_k de C .

As frequências marginais, ou totais marginais dos nossos dados em estudo, são de dois tipos:

- Marginais simples ou de uma só variável:

$$- n_{i..} = \sum_{j=1}^2 \sum_{k=1}^2 n_{ijk} , \quad i = 1, \dots, 6$$

$$- n_{.j.} = \sum_{i=1}^6 \sum_{k=1}^2 n_{ijk} , \quad j = 1, 2$$

$$- n_{..k} = \sum_{i=1}^6 \sum_{j=1}^2 n_{ijk} \quad , \quad k = 1,2$$

- Marginais duplas ou de duas variáveis:

$$- n_{ij.} = \sum_{k=1}^2 n_{ijk} \quad , \quad i = 1, \dots, 6; \quad j = 1,2$$

$$- n_{i.k} = \sum_{j=1}^2 n_{ijk} \quad , \quad i = 1, \dots, 6; \quad k = 1,2$$

$$- n_{.jk} = \sum_{i=1}^6 n_{ijk} \quad , \quad i = 1, \dots, 6; \quad j = 1,2$$

O total da tabela é igual ao número dos desempregados em Cabo Verde.

$$n = \sum_{i=1}^6 \sum_{j=1}^2 \sum_{k=1}^2 n_{ijk} = 68168$$

Quando se dispõe de uma classificação segundo mais do que duas variáveis, em geral, começa-se por analisar a associação entre as variáveis duas a duas.

Para além da associação entre as variáveis duas a duas, um dos tipos de associação que convém estudar é associação parcial entre duas variáveis tendo em conta o efeito das restantes variáveis. Outro tipo de associação é a associação múltipla que permite estudar a associação entre uma variável e todas as restantes.

3.3.1 Medidas de Associação Parcial

Como medidas de associação parcial, para tabelas de contingência tridimensionais, serão apresentados duas medidas definidas com base nas medidas de associação para tabelas bidimensionais. Ou seja, estas medidas não são originais, são adaptações das medidas anteriormente definidas.

Apenas serão analisadas as medidas de associação entre A e B dada a variável C , em todos os outros casos as medidas serão semelhantes (como é lógico, com as devidas adaptações).

Média das Medidas de Associação

Suponha que tem uma tabela de contingência $r \times s \times t$ e que quer medir a associação parcial entre as variáveis A e B dada a variável C . Para tal é necessário seguir os seguintes passos:

1. Criar uma série de tabelas de contingência bidimensionais $r \times s$, usando cada um dos níveis da variável C . Para cada categoria C_k dada variável C , é construída uma tabela bidimensional formada pelo cruzamento das classificações segundo as variáveis A e B , na qual a probabilidade de cada célula (i, j) é dada por $\frac{p_{ijk}}{p_{..k}}$. Como variável C tem t categorias obtêm-se t tabelas bidimensionais (uma por cada categoria de C).
2. Calcular uma medida de associação para cada uma das t tabelas (calcula-se sempre a mesma medida, pode ser qualquer uma das apresentadas anteriormente para as tabelas bidimensionais), obtendo-se um conjunto de t medidas de associação.
3. Para determinar o valor da medida de associação parcial existem duas possibilidades:
 - (a) calcular simplesmente a média de t medidas anterior;
 - (b) calcular a média das t medidas anteriores ponderada pelas probabilidades marginais da variável C ($p_{..k}$).

Como estes dois tipos de médias variam significativamente, se o número de observações em cada uma das tabelas bidimensionais diferir substancialmente, é preferível calcular a média ponderada.

Para exemplificar estas medidas vamos recorrer à medida proposta por Goodman e Kruskal-medida optimal assimétrica λ_b .

Suponha que tem uma tabela tridimensional $r \times s \times t$ e que para cada categoria da variável C foi construída uma tabela bidimensional. Para cada uma dessas tabelas bidimensionais calculou-se medida de associação λ_b (será designada por $\lambda_b(k)$ a medida de associação que se obtém quando se fixa a categoria k da variável C). A medida de associação parcial entre as variáveis A e B dada a variável C consiste em calcular a média (simple ou ponderada) dos valores das medidas $\lambda_b(k)$, $k = 1, \dots, t$, isto é:

- **Media Simples:**

$$\lambda_b(A, B|C) = \frac{\sum_{k=1}^t \lambda_b(k)}{t}$$

- **Média Ponderada:** ponderada pelas probabilidades marginais da variável C ($p_{..k}$):

$$\lambda_b(A, B|C) = \sum_{k=1}^t p_{..k} \lambda_b(k)$$

Estimação da Medida de Associação

Neste caso é difícil definir a estimativa destas medidas de associação, uma vez que, como é óbvio, essa estimativa depende da medida de associação para tabelas bidimensionais seleccionada.

Mais uma vez, vamos supor que se optou por usar a medida de Goodman e Kruskal. Ou seja, suponha que se pretende estimar a medida $\lambda_b(A, B|C)$ anteriormente definida. Para obter essa estimativa basta substituir as probabilidades p_{ij} pelas suas estimativas de máxima verosimilhança, $\hat{p}_{ij} = \frac{n_{ij}}{n}$, e para cada um dos casos vem:

- **Média Simples:**

$$L_b(A, B|C) = \frac{\sum_{k=1}^t L_b(k)}{t}$$

com $L_b(k)$ a estimativa de $\lambda_b(k)$.

- **Média Ponderada:**

$$L_b(A, B|C) = \sum_{k=1}^t \frac{n_{..k}}{n} L_b(k)$$

com $L_b(k)$ a estimativa de $\lambda_b(k)$.

Medidas Baseadas Directamente em Probabilidades de Erro

Estas medidas foram propostas por Goodman e Kruskal e são simplesmente a generalização das medidas λ 's apresentadas para tabelas $r \times s$.

Tal como aconteceu com a medida anterior, apenas serão definidas as medidas de associação parcial entre as variáveis A e B dada a variável C , sendo os outros casos feitos de modo análogo.

Esta medida consiste em prever a categoria da variável B a que pertence um elemento da população, escolhido ao acaso e sendo conhecida a sua categoria em relação à variável C , nas duas situações seguintes:

Caso 1: sem qualquer outra informação;

Caso 2: conhecida a sua categoria em relação à variável A .

No caso 1, a categoria de B com maior probabilidade de ser escolhida, sabendo que a categoria de C é k , é o $\max p_{.jk}$, ou seja, é a categoria cuja probabilidade marginal de linha, dentro da classe C_k , seja máxima. No caso 2 será a probabilidade máxima que se encontra no cruzamento das categorias i de A e k de C , ou seja, $\max p_{ijk}$.

Com base no que foi dito, vem que, no caso 1, a probabilidade de errar ao prever a categoria de B é

$$1 - \max_{j=1, \dots, s} \frac{p_{.jk}}{p_{.k}}$$

Enquanto no caso 2 é

$$1 - \sum_{i=1}^r \max_{j=1, \dots, s} \frac{p_{ijk}}{p_{.k}}$$

Portanto, ao considerar os indivíduos seleccionados aleatoriamente da população onde a sua categoria da variável C é conhecida, vem que a probabilidade de errar ao prever a sua categoria na variável B é, no caso 1,

$$1 - \sum_{k=1}^t \max_{j=1, \dots, s} p_{.jk}$$

Enquanto que no caso 2 é

$$1 - \sum_{k=1}^t \sum_{i=1}^r \max_{j=1, \dots, s} p_{ijk}$$

O decréscimo relativo na probabilidade de errar ao predizer a categoria de B de um indivíduo quando se passa do caso 1 para o caso 2, conhecida a sua categoria na variável C , é

$$\lambda'_b(A, B|C) = \frac{\sum_{k=1}^t \sum_{i=1}^r p_{i\alpha k} - \sum_{k=1}^t p_{.\alpha k}}{1 - \sum_{k=1}^t p_{.\alpha k}}$$

onde $p_{i\alpha k} = \max_{j=1, \dots, s} p_{ijk}$

$$p_{.\alpha k} = \max_{j=1, \dots, s} p_{.jk}$$

Esta medida baseia-se na predição optimal onde a categoria de C é sempre conhecida (daí o termo de associação “parcial”), é assimétrica no sentido em que apenas a predição das categorias de B são consideradas e não se altera por permutações independentes das classes dentro de cada variável.

Tal como aconteceu com as medidas de predição óptima (nas tabelas bidimensionais) basta trocar o papel da variável B pelo da variável A em relação à medida anterior, e obtém-se a seguinte medida de associação parcial assimétrica:

$$\lambda'_a(A, B|C) = \frac{\sum_{k=1}^t \sum_{j=1}^s p_{\beta jk} - \sum_{k=1}^t p_{\beta.k}}{1 - \sum_{k=1}^t p_{\beta.k}}$$

Com $p_{\beta jk} = \max_{i=1, \dots, r} p_{ijk}$ e $p_{\beta.k} = \max_{i=1, \dots, r} p_{i.k}$.

Esta medida representa o decréscimo relativo na probabilidade de erro ao determinar a categoria de A de um indivíduo, seleccionado aleatoriamente, quando só se conhece a sua categoria de C e se passa a conhecer também a sua categoria de B .

Como é óbvio, tal como aconteceu anteriormente, também existe a possibilidade de obter uma medida simétrica:

$$\lambda(A, B|C) = \frac{(1 - \sum_{k=1}^t p_{.\alpha k})\lambda'_b(A, B|C) + (1 - \sum_{k=1}^t p_{\beta.k})\lambda'_a(A, B|C)}{2 - \sum_{k=1}^t p_{.\alpha k} - \sum_{k=1}^t p_{\beta.k}}$$

Como foi dito no início do capítulo, só foram definidas medidas de associação parcial entre as variáveis A e B dada a variável C , como é lógico de forma análoga seria possível definir medidas de associação parcial baseadas na predição óptima para os outros dois casos:

- Medidas de associação parcial entre as variáveis A e C dada a variável B ;
- Medidas de associação parcial entre as variáveis B e C dada a variável A ;

Estimação de $\lambda_b(A, B|C)$

O estimador de máxima verosimilhança sob a hipótese do modelo de amostragem ser multinomial, é

$$L'_b(A, B|C) = \frac{\sum_{k=1}^2 \sum_{i=1}^6 n_{i\alpha k} - \sum_{k=1}^2 n_{.\alpha k}}{n - \sum_{k=1}^2 n_{.\alpha k}}$$

com $n_{i\alpha k} = \max_{j=1, \dots, s} n_{ijk}$ e $n_{.\alpha k} = \max_{j=1, \dots, s} n_{.jk}$.

No estudo anterior foram analisadas as variáveis *escalão etário* e *género* dos desempregados em Cabo Verde no ano 2006 e 2008, chegou-se à conclusão que apesar de fraca, a associação entre as variáveis é significativa. De seguida serão analisadas as mesmas variáveis em conjunto com a variável *ano*, com o objectivo de tentar verificar se a associação continua significativa.

Para o estudo de mais uma variável construiu-se a seguinte tabela de contingência tridimensional dos desempregados em Cabo Verde:

<i>Ano</i>	<i>C1</i>		<i>C2</i>	
	<i>B1</i>	<i>B2</i>	<i>B1</i>	<i>B2</i>
<i>Género</i>				
<i>A1</i>	7687	9845	7980	10017
<i>A2</i>	2986	5436	2914	5509
<i>Escalão Etário</i>				
<i>A3</i>	1318	2999	1532	3031
<i>A4</i>	869	1832	565	1811
<i>A5</i>	245	198	228	642
<i>A6</i>	104	56	187	177

onde:

- a variável $A = \text{Escalão Etário}$ tem as categorias:

– $A_1 = 15$ a 24 anos;

– $A_2 = 25$ a 34 anos;

– $A_3 = 35$ a 44 anos;

– $A_4 = 45$ a 54 anos;

– $A_5 = 55$ a 64 anos;

– $A_6 = 65$ anos e +.

- a variável $B = \text{Género}$ tem as categorias:

– $B_1 = \text{Masculino}$;

– $B_2 = \text{Feminino}$.

- a variável $C = \text{Ano}$ tem as seguintes categorias:

– $C_1 = 2006$;

– $C_2 = 2008$.

Com base nestes dados, quando é testada a hipótese de independência mútua, recorrendo ao teste do Qui-Quadrado, obtém-se

$$\chi^2 = \sum_{i=1}^6 \sum_{j=1}^2 \sum_{k=1}^2 \frac{(n_{ijk} - \frac{n_{i..}n_{.j.}n_{..k}}{n^2})^2}{\frac{n_{i..}n_{.j.}n_{..k}}{n^2}} = 1366.6$$

e $p - value = 0$. Para todos os níveis de significância rejeita-se a hipótese de independência mútua.

Apesar da hipótese ter sido rejeitada não se pode afirmar que existe associação significativa entre as variáveis, sendo assim necessário quantificar a associação existente.

Para tentar quantificar a associação existente vamos recorrer às medidas de associação parcial (às medidas de associação parcial entre A e B dada a variável C).

Para calcular as primeiras medidas de associação que foram definidas (Média das Medidas de Associação) vamos recorrer, de novo, às medidas de Goodman e Kruskal. Deste modo obtém-se as seguintes estimativas das medidas de associação do desemprego em Cabo Verde:

- Usando a Média Simples vem:

$$L_b(A, B|C) = \frac{\sum_{k=1}^2 L_b(k)}{2} = \frac{0.0072 + 0.0007}{2} = 0.00796$$

- Usando a Média Ponderada vem:

$$L_b(A, B|C) = \sum_{k=1}^2 \frac{n_{..k}}{n} L_b(k) = \frac{33575}{68168} \times 0.0072 + \frac{34593}{68168} \times 0.0007 = 0.00392$$

Calculando a medida baseada directamente em probabilidade de erro vem:

$$L'_b(A, B|C) = \frac{\sum_{k=1}^2 \sum_{i=1}^6 n_{iak} - \sum_{k=1}^2 n_{.ak}}{n - \sum_{k=1}^2 n_{.ak}} = 0.00394$$

Todos os resultados são muito semelhantes, os dois últimos são quase idênticos. Ambos levam a concluir que a associação existente é fraca, neste caso a associação parcial entre o *escalão etário* e o *género* dado o *ano*.

3.3.2 Medidas de Associação Múltipla

Outro tipo de associação, para tabelas de contingência com mais de duas variáveis, é a associação múltipla entre uma das variáveis e todas as outras. Para definir este tipo de associação considere uma tabela tridimensional $r \times s \times t$ formada pelas variáveis A , B e C e suponha que pretende analisar a associação múltipla entre a variável B e a variável (AC) .

A forma mais simples de analisar este tipo de associação consiste em construir uma tabela bidimensional a partir da tabela tridimensional inicial. Ou seja, neste caso em que pretende analisar a associação múltipla entre as variáveis B e (AC) basta construir uma tabela bidimensional cujas linhas representam a variável (AC) e as colunas representam a variável B . A nossa tabela bidimensional do desemprego em Cabo Verde é a seguinte:

	B_1	B_2	
A_1C_1	7687	9845	17532
A_1C_2	7980	10017	17997
A_2C_1	2986	5436	8422
A_2C_2	2914	5509	8423
A_3C_1	1318	2999	4317
A_3C_2	1532	3031	4563
A_4C_1	869	1832	2701
A_4C_2	565	1811	2376
A_5C_1	245	198	443
A_5C_2	228	642	870
A_6C_1	104	56	160
A_6C_2	187	177	364
	26615	41553	68168

Para medir a associação múltipla entre B e (AC) basta aplicar qualquer uma das medidas apresentadas anteriormente para tabelas bidimensionais.

Goodman e Kruskal apresentam o exemplo onde é aplicada a medida λ_b , obtendo-se a seguinte expressão:

$$\lambda_b''(B, (AC)) = \frac{\sum_{k=1}^t \sum_{i=1}^r p_{i\alpha k} - p_{\alpha.}}{1 - p_{\alpha.}}$$

Neste caso o valor de $\lambda_b''(B, (AC))$ representa o decréscimo relativo na probabilidade de errar ao predizer a categoria da variável B , de um indivíduo seleccionado aleatoriamente da população, quando não se dispõe de qualquer informação e se passa a conhecer as suas categorias segundo as variáveis A e C .

É necessário ter em atenção que ao considerar a tabela bidimensional anterior não se está a considerar a associação entre as variáveis A e C .

Apenas é apresentado o caso em que se pretende estudar a associação múltipla entre as variáveis B e (AC) mas, como é lógico, poder-se-ia, com o mesmo procedimento, estudar a associação múltipla entre as variáveis A e (BC) ou entre as variáveis C e (AB) .

Estimação das Medidas de Associação Múltipla

Neste caso é difícil definir a estimativa destas medidas de associação, uma vez que, como é óbvio, essa estimativa depende da medida escolhida.

Vamos supor que se optou por usar a medida de Goodman e Kruskal, ou seja, suponhamos que se pretende estimar a medida $\lambda_b''(B, (AC))$ anteriormente definida.

Para obter uma estimativa desta medida basta substituir as probabilidades p_{ij} pelas suas estimativas de máxima verosimilhança, $\hat{p}_{ij} = \frac{n_{ij}}{n}$, e fica:

$$L_b''(B, (AC)) = \frac{\sum_{k=1}^t \sum_{i=1}^r n_{i\alpha k} - n_{\alpha.}}{1 - n_{\alpha.}}$$

A medida de associação múltipla entre as variáveis B e (AC) é:

$$L'_b(B; (AC)) = \frac{\sum_{k=1}^2 \sum_{i=1}^6 n_{iak} - n_{.a.}}{1 - n_{.a.}} = 0.00394$$

Com base neste resultado pode-se concluir que a associação existente entre a variável *género* e as variáveis *escalão etário* e o *ano* aparenta ser fraca.

É de salientar que ao considerar a tabela bidimensional anterior não se tem em consideração a associação entre as variáveis *escalão etário* e o *ano* do desemprego em Cabo Verde.

4. MODELOS LOG-LINEARES

Anteriormente fomos fazendo referência a hipóteses de interesse a testar em tabelas de contingência do desemprego em Cabo Verde no ano 2006 e 2008. Agora vamos considerar uma abordagem diferente, que passa pelo ajustamento de modelos e estimação dos parâmetros desses modelos. O termo modelo refere-se a uma estrutura conceptual acerca das observações, e os parâmetros do modelo representam os efeitos que variáveis particulares ou combinações de variáveis provocam na determinação dos valores observados da amostra.

4.1 Modelos log-lineares em tabelas bidimensionais

Por simplicidade, consideremos uma tabela de contingência bidimensional e a hipótese de independência entre os factores dessa tabela, ou seja de não existência de interacção entre esses factores

$$p_{ij} = p_i.p_j, \quad i = 1, \dots, 6; j = 1, 2$$

Esta relação estabelece uma estrutura particular ou modelo para os dados, nomeadamente que na população a probabilidade de uma observação surgir (i, j) tem o mesmo valor que o produto das probabilidades marginais.

Já as frequências esperadas para cada célula obtém-se através do produto de p_{ij} pela dimensão da amostra, vindo:

$$\{u_{ij} = np_{ij}\}$$

ou, sob a hipótese de independência

$\{u_{ij} = p_i.p_j \times n\}$, qualquer que seja o número de linha i e de coluna j .

Aplicando logaritmos à equação $\{u_{ij} = np_{ij}\}$, as frequências esperadas permanecem no intervalo $]0, +\infty[$, obtendo-se:

$$\log u_{ij} = \log (np_i \times p_j)$$

Sabendo que logaritmo de um produto é a soma logaritmos, obtém-se o logaritmo das frequências esperadas com a expressão:

$$\log u_{ij} = \log n + \log p_{i.} + \log p_{.j}$$

A primeira parcela ($\log n$) depende do tamanho da amostra;

A segunda parcela ($\log p_{i.}$) baseia-se na probabilidade da linha i ;

A terceira parcela ($\log p_{.j}$) baseia-se na probabilidade da coluna j .

A equação $\log u_{ij} = \log n + \log p_{i.} + \log p_{.j}$ origina uma representação típica do modelo log-linear $\log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B$

A forma mais geral deste modelo é $\log u_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$.

Em que λ é designado por efeito médio geral; λ_i^A é o efeito da linha i da variável A (*Escalão Etário*); λ_j^B é o efeito da coluna j da variável B (*Gênero*); e λ_{ij}^{AB} é o efeito interactivo entre a categoria i de A e j de B por λ_{ij}^{AB} , em que o efeito de uma variável na frequência esperada, depende do nível da outra variável.

i) havendo relação de independência entre as variáveis, o modelo diz-se não saturado, por não incluir o efeito interactivo, isto é, tem a restrição $\lambda_{ij}^{AB} = 0$, sendo dado por:

$$\log \mu_{ij} = \lambda + \lambda_i^A + \lambda_j^B$$

ii) havendo relação de dependência entre as variáveis, o modelo corresponde ao saturado e é dado por:

$$\log u_{ij} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_{ij}^{AB}$$

Quanto maior forem os valores de λ_i^A ou de λ_j^B , maior é a frequência esperada na linha i ou na coluna j da tabela.

Os modelos log-lineares usam as frequências esperadas u_{ij} , em vez das probabilidades, e assumem a distribuição de Poisson para contagem das $I \times J$ células independentes.

Com base na nossa tabela de contingência do desemprego em Cabo Verde no ano 2006, pretende-se agora analisar a relação entre o *gênero* dos desempregados em Cabo Verde no ano 2006 segundo o seu *escalão etário*.

As contagens dos totais marginais são iguais as contagens das frequências esperadas do desemprego em Cabo Verde no ano 2006, como se vê na tabela em baixo.

			Género		Total
			Masculino	Feminino	
Escalao.Etario	15-24 anos	Count	7687	9845	17532
		Expected Count	6897,1	10634,9	17532,0
	25-34 anos	Count	2986	5436	8422
		Expected Count	3313,2	5108,8	8422,0
	35-44 anos	Count	1318	2999	4317
		Expected Count	1698,3	2618,7	4317,0
	45-54 anos	Count	869	1832	2701
		Expected Count	1062,6	1638,4	2701,0
	55-64 anos	Count	244	198	442
		Expected Count	173,9	268,1	442,0
	65 anos e +	Count	104	56	160
		Expected Count	62,9	97,1	160,0
Total		Count	13208	20366	33574
		Expected Count	13208,0	20366,0	33574,0

Tabela 4.1: Tabela das frequências observados e esperados de desemprego em Cabo Verde no ano 2006.

O estudo do melhor modelo a adaptar à nossa tabela em estudo foi feito recorrendo ao SPSS. Um dos outputs fornecidos foi a seguinte tabela, que nos permite concluir que existe interacção significativa entre as variáveis *escalão etário* e *género* ou seja as variáveis *escalão etário* e *género* não são independentes.

K-Way and Higher-Order Effects							
	K	df	Likelihood Ratio		Pearson		Number of Iterations
			Chi-Square	Sig.	Chi-Square	Sig.	
K-way and Higher Order Effects ^a	1	11	39407,105	,000	41282,313	,000	0
	2	5	494,195	,000	491,699	,000	2
K-way Effects ^b	1	6	38912,910	,000	40790,614	,000	0
	2	5	494,195	,000	491,699	,000	0

a. Tests that k-way and higher order effects are zero.

b. Tests that k-way effects are zero.

Tabela 4.2: Tabela de K-Way and Higher-Order Effects para os dados de 2006.

Por exemplo, $K=2$ testa se todas as interacções de ordem dois são nulas. O resultado do teste $\chi^2_5 = 494.195$ com $p - value \leq 0.0005$, mostra que na população pelo menos uma interacção de ordem dois é significativamente diferente de zero. Como só existem duas variáveis, só pode haver uma interacção de ordem dois, o que verifica-se na tabela 4.2.

A tabela seguinte representa um modelo com efeito interactivo.

Step ^a	Effects	Chi-Square ^c	df	Sig.	Number of Iterations
0	Generating Class ^b	Escalao.Etario*Género	,000	0.	
	Deleted Effect 1	Escalao.Etario*Género	494,195	,000	2
1	Generating Class ^b	Escalao.Etario*Género	,000	0.	

a. At each step, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than ,050.

b. Statistics are displayed for the best model at each step after step 0.

c. For 'Deleted Effect', this is the change in the Chi-Square after the effect is deleted from the model.

Tabela 4.3: Tabela de Step summary para os dados de 2006.

O *Step Summary* começa pelo modelo saturado e, na etapa 0, mostra que, eliminando o efeito interactivo *escalao.etario*género*, o modelo se diferenciava significativamente do saturado, com $p - value \leq 0.0005$. Assim, a etapa 1 mostra que o melhor modelo inclui o efeito interactivo.

A tabela seguinte representa as frequências observadas e as esperadas para todas as células e os resíduos.

Escalao.Etario	Género	Observed		Expected		Residual	Standardized Residual	Adjusted Residual	Deviance
		Count	%	Count	%				
15-24 anos	Masculino	7687	22,9%	7687,000	22,9%	,000	,000	,000	,000
	Feminino	9845	29,3%	9845,000	29,3%	,000	,000	,000	,000
25-34 anos	Masculino	2986	8,9%	2986,000	8,9%	,000	,000	,000	,000
	Feminino	5436	16,2%	5436,000	16,2%	,000	,000	,000	,000
35-44 anos	Masculino	1318	3,9%	1318,000	3,9%	,000	,000	,000	,000
	Feminino	2999	8,9%	2999,000	8,9%	,000	,000	,000	,000
45-54 anos	Masculino	869	2,6%	869,000	2,6%	,000	,000	,000	,000
	Feminino	1832	5,5%	1832,000	5,5%	,000	,000	,000	,000
55-64 anos	Masculino	244	,7%	244,000	,7%	,000	,000	,000	,000
	Feminino	198	,6%	198,000	,6%	,000	,000	,000	,000
65 anos e +	Masculino	104	,3%	104,000	,3%	,000	,000	,000	,000
	Feminino	56	,2%	56,000	,2%	,000	,000	,000	,000

a. Model: Poisson

b. Design: Constant + Escalao.Etario + Género + Escalao.Etario * Género

Tabela 4.4: Tabela de Cell Counts and Residuals para os dados de 2006.

O modelo sendo saturado apresenta resíduos nulos, por ter um ajustamento perfeito entre as frequências observadas e as esperadas para todas as células, como se confirma pelos testes do ajustamento do Qui-Quadrado e pela tabela das *Cell Counts and Residuals*.

De seguida, apresenta-se a tabela dos parâmetros estimados:

Parameter Estimates ^a						
Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	4,025	,134	30,123	,000	3,763	4,287
[Escalaio.Etario = 1]	5,169	,134	38,575	,000	4,907	5,432
[Escalaio.Etario = 2]	4,575	,134	34,065	,000	4,312	4,839
[Escalaio.Etario = 3]	3,981	,135	29,515	,000	3,716	4,245
[Escalaio.Etario = 4]	3,488	,136	25,710	,000	3,222	3,754
[Escalaio.Etario = 5]	1,263	,151	8,344	,000	,966	1,560
[Escalaio.Etario = 6]	0 ^a
[Género = 1]	,619	,166	3,735	,000	,294	,944
[Género = 2]	0 ^a
[Escalaio.Etario = 1] * [Género = 1]	-,866	,166	-5,206	,000	-1,193	-,540
[Escalaio.Etario = 1] * [Género = 2]	0 ^a
[Escalaio.Etario = 2] * [Género = 1]	-1,218	,167	-7,281	,000	-1,546	-,890
[Escalaio.Etario = 2] * [Género = 2]	0 ^a
[Escalaio.Etario = 3] * [Género = 1]	-1,441	,169	-8,527	,000	-1,772	-1,110
[Escalaio.Etario = 3] * [Género = 2]	0 ^a
[Escalaio.Etario = 4] * [Género = 1]	-1,365	,171	-7,991	,000	-1,700	-1,030
[Escalaio.Etario = 4] * [Género = 2]	0 ^a
[Escalaio.Etario = 5] * [Género = 1]	-,410	,191	-2,143	,032	-,785	-,035
[Escalaio.Etario = 5] * [Género = 2]	0 ^a
[Escalaio.Etario = 6] * [Género = 1]	0 ^a
[Escalaio.Etario = 6] * [Género = 2]	0 ^a

a. This parameter is set to zero because it is redundant.

Tabela 4.5: Tabela de parameter estimates para os dados de 2006.

A tabela 4.4 mostra as frequências esperadas que se espera obter para cada célula, usando o modelo log-linear. Estas frequências esperadas podem ser obtidas a partir do quadro dos parâmetros estimados, como se mostra para a célula (1,1).

A frequência esperada da célula (1,1), com *grupo etário* entre 15 a 24 anos do *género* masculino dos desempregados em Cabo Verde no ano 2006, é 7687, donde o seu logaritmo é $\log(7687) = 8.95$, que pode ser obtida como:

$$\log u_{11} = \lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB} = 4.025 + 5.169 + 0.619 - 0.866 = 8.95$$

A tabela dos parâmetros estimados mostra para um nível de significância de 0.05, que existe efeito interactivo entre o *género* e *escalão etário* dos desempregados em Cabo Verde no ano 2006.

De seguida analisámos a tabela de contingência do desemprego em Cabo Verde referente agora ao ano 2008.

A tabela seguinte permite nos concluir que existe interacção significativa entre as variáveis *escalão etário* e *género* ou seja as variáveis *escalão etário* e *género* não são independentes.

K-Way and Higher-Order Effects							
K	df	Likelihood Ratio		Pearson		Number of Iterations	
		Chi-Square	Sig.	Chi-Square	Sig.		
K-way and Higher Order Effects ^a	1	11	38147,550	,000	41211,195	,000	0
	2	5	672,879	,000	656,196	,000	2
K-way Effects ^b	1	6	37474,671	,000	40554,999	,000	0
	2	5	672,879	,000	656,196	,000	0

a. Tests that k-way and higher order effects are zero.

b. Tests that k-way effects are zero.

Tabela 4.6: Tabela de K-Way and Higher-Order Effects referente aos dados de 2008.

Por exemplo, $k = 2$ testa se todas as interacções de ordem dois são nulas. O resultado do teste $\chi^2_5 = 672.879$ com $p - value \leq 0.0005$, mostra que na população pelo menos uma interacção de ordem dois é significativamente diferente de zero.

A tabela seguinte compara um modelo com efeito interactivo e outro sem efeito interactivo.

Step Summary						
Step ^a	Effects	Chi-Square ^c	df	Sig.	Number of Iterations	
0	Generating Class ^b	Escalao.Etario*Género	,000	0		
	Deleted Effect 1	Escalao.Etario*Género	672,879	5	,000	2
1	Generating Class ^b	Escalao.Etario*Género	,000	0		

a. At each step, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than ,050.

b. Statistics are displayed for the best model at each step after step 0.

c. For 'Deleted Effect', this is the change in the Chi-Square after the effect is deleted from the model.

Tabela 4.7: Tabela de Step Summary referente aos dados de 2008.

O *Step Summary* começa pelo modelo saturado e, na etapa 0, mostra que, eliminando o efeito interactivo *escalao.etario*género*, o modelo se diferenciava significativamente do saturado, com $p - value \leq 0.0005$. Assim, a etapa 1 mostra que o melhor modelo inclui o efeito interactivo.

A tabela seguinte mostra as frequências que se espera obter para cada célula, usando o modelo log-linear.

Cell Counts and Residuals^{a,b}

Escalao.Etari o	Género	Observed		Expected		Residual	Standardized Residual	Adjusted Residual	Deviance
		Count	%	Count	%				
15-24 anos	Masculino	7980	23,1%	7980,000	23,1%	,000	,000	,000	,000
	Feminino	10017	29,0%	10017,000	29,0%	,000	,000	,000	,000
25-34 anos	Masculino	2914	8,4%	2914,000	8,4%	,000	,000	,000	,000
	Feminino	5509	15,9%	5509,000	15,9%	,000	,000	,000	,000
35-44 anos	Masculino	1532	4,4%	1532,000	4,4%	,000	,000	,000	,000
	Feminino	3031	8,8%	3031,000	8,8%	,000	,000	,000	,000
45-54 anos	Masculino	565	1,6%	565,000	1,6%	,000	,000	,000	,000
	Feminino	1811	5,2%	1811,000	5,2%	,000	,000	,000	,000
55-64 anos	Masculino	228	,7%	228,000	,7%	,000	,000	,000	,000
	Feminino	642	1,9%	642,000	1,9%	,000	,000	,000	,000
65 anos e +	Masculino	187	,5%	187,000	,5%	,000	,000	,000	,000
	Feminino	177	,5%	177,000	,5%	,000	,000	,000	,000

a. Model: Poisson

b. Design: Constant + Escalao.Etario + Género + Escalao.Etario * Género

Tabela 4.8: Tabela de Cell Counts and Residuals referente aos dados de 2008.

A tabela seguinte representa as estimativas dos parâmetros.

Parameter Estimates ^{b,c}						
Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	5,176	,075	68,864	,000	5,029	5,323
[Escalaio.Etario = 1]	4,036	,076	53,226	,000	3,887	4,185
[Escalaio.Etario = 2]	3,438	,076	45,022	,000	3,288	3,588
[Escalaio.Etario = 3]	2,840	,077	36,733	,000	2,689	2,992
[Escalaio.Etario = 4]	2,325	,079	29,529	,000	2,171	2,480
[Escalaio.Etario = 5]	1,288	,085	15,177	,000	1,122	1,455
[Escalaio.Etario = 6]	0 ^a					
[Género = 1]	,055	,105	,524	,600	-,151	,260
[Género = 2]	0 ^a					
[Escalaio.Etario = 1] * [Género = 1]	-,282	,106	-2,665	,008	-,490	-,075
[Escalaio.Etario = 1] * [Género = 2]	0 ^a					
[Escalaio.Etario = 2] * [Género = 1]	-,692	,107	-6,445	,000	-,902	-,481
[Escalaio.Etario = 2] * [Género = 2]	0 ^a					
[Escalaio.Etario = 3] * [Género = 1]	-,737	,109	-6,736	,000	-,952	-,523
[Escalaio.Etario = 3] * [Género = 2]	0 ^a					
[Escalaio.Etario = 4] * [Género = 1]	-1,220	,115	-10,569	,000	-1,446	-,994
[Escalaio.Etario = 4] * [Género = 2]	0 ^a					
[Escalaio.Etario = 5] * [Género = 1]	-1,090	,130	-8,376	,000	-1,345	-,835
[Escalaio.Etario = 5] * [Género = 2]	0 ^a					
[Escalaio.Etario = 6] * [Género = 1]	0 ^a					
[Escalaio.Etario = 6] * [Género = 2]	0 ^a					

a. This parameter is set to zero because it is redundant.

b. Model: Poisson

c. Design: Constant + Escalao.Etario + Género + Escalao.Etario * Género

Tabela 4.9: Tabela de parameter estimates referente aos dados de 2008.

A frequência esperada da célula (1,1), com grupo *etário* entre 15 a 24 anos do género masculino do desemprego em Cabo Verde no ano 2008 é 7980, donde o seu logaritmo é $\log = (7980) = 8.99$, que se pode obter como:

$$\log \mu_{11} = \lambda + \lambda_1^A + \lambda_1^B + \lambda_{11}^{AB} = 5.176 + 4.036 + 0.055 - 0.282 = 8.99$$

A tabela dos parâmetros estimados mostra para um nível de significância de 0.05, que existe efeito interactivo entre o *género* e *escalão etário* dos desempregados em Cabo Verde no ano 2008 e identificamos um caso em que o efeito não é significativo, no que diz respeito a variável género masculino.

4.2 Modelo Log-linear em Tabelas com três variáveis

Examinamos agora os log-lineares disponíveis para uma tabela de contingência tridimensional. O modelo saturado neste caso é:

$$\log u_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ij}^{AB} + \lambda_{ik}^{AC} + \lambda_{jk}^{BC} + \lambda_{ijk}^{ABC}$$

Em que λ é designado por efeito médio geral; λ_i^A é o efeito da linha i da variável A (*Escalão Etario*); λ_j^B é o efeito da coluna j da variável B (*Gênero*); λ_k^C é o efeito da linha k da variável C (*Ano*); λ_{ij}^{AB} é o efeito interactivo entre a categoria i de A e j de B por λ_{ij}^{AB} , λ_{ik}^{AC} é o efeito interactivo entre a categoria i de A e k de C por λ_{ik}^{AC} ; λ_{jk}^{BC} é o efeito interactivo entre a categoria j de B e k de C por λ_{jk}^{BC} ; e λ_{ijk}^{ABC} é o efeito interactivo entre a categoria i de A , j de B e k de C por λ_{ijk}^{ABC} .

Vemos que este modelo inclui parâmetros de efeitos principais para cada variável, parâmetros de efeitos de interacção de segunda ordem para cada par de variáveis, e também parâmetros representando possíveis efeitos de terceira ordem entre as três variáveis.

A hipótese de não existência de interacção de terceira ordem pode ser expressa em termos dos parâmetros . É fácil mostrar que é equivalente a

$$H_0: \lambda_{ijk}^{ABC} = 0, \quad \forall i, j, k$$

Outras hipóteses podem ser expressas com os parâmetros do modelo. Por exemplo, a hipótese de independência mútua, que especifica que não existem associações de qualquer tipo entre as três variáveis, isto é, que não existem interacções de segunda ordem entre qualquer par de variáveis, não existe interacção conjunta das três variáveis, pode agora ser formulada por

$$H_0: \lambda_{ij}^{AB} = \lambda_{ik}^{AC} = \lambda_{jk}^{BC} = \lambda_{ijk}^{ABC} = 0$$

Neste caso o nosso modelo log-linear será simplesmente

$$\log u_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C$$

Que envolve apenas o efeito médio geral e os efeitos principais de cada variável. Se este modelo produz um bom ajustamento dos dados, então isso significa que as diferenças entre as frequências das células apenas reflectem as diferenças entre os totais marginais das variáveis.

Consideremos agora um modelo não saturado, em que $\lambda_{ij}^{AB} = 0$ e $\lambda_{ijk}^{ABC} = 0$. A forma do modelo virá a ser:

$$\log u_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$$

Estabelecer que $\lambda_{ijk}^{ABK} = 0$ é equivalente a estabelecer que a interacção entre as variáveis A e B é a mesma em todos os níveis da variável C ; estabelecer que $\lambda_{ij}^{AB} = 0$ é equivalente a estabelecer que a interacção entre as variáveis A e B é nula. Consequentemente o modelo $\log u_{ijk} = \lambda + \lambda_i^A + \lambda_j^B + \lambda_k^C + \lambda_{ik}^{AC} + \lambda_{jk}^{BC}$ afirma não existe interacção entre as variáveis A e B em cada nível da variável C ou, de outro modo, que as variáveis A e B são condicionalmente independentes dada a variável C . Neste modelo é suposto que as variáveis A e B estão associadas com a variável C uma vez que não especificamos que $\lambda_{ik}^{AC} + \lambda_{jk}^{BC} = 0$.

Se continuarmos a reduzir termos no modelo, ficamos com menos parâmetros do que no modelo de independência mútua, isto é, chegamos a um modelo que não inclui todas as variáveis. Se este modelo se revelar adequado no ajustamento de dados, isso indicará que umas (ou mais) variáveis são redundantes e que a dimensionalidade da tabela pode ser reduzida.

Anteriormente foram analisadas as variáveis *escalão etário* e *género* do desemprego em Cabo Verde, separadamente, em cada ano 2006 e 2008. De seguida serão analisadas as mesmas variáveis em conjunto com a variável *ano*. Pretende-se verificar se o modelo se ajusta bem aos dados com três variáveis em estudo. Um primeiro resultado encontra-se na tabela seguinte.

Step ^a	Effects	Chi-Square ^c	df	Sig.	Number of Iterations
0	Generating Class ^b	Ano*Escalão Etário*Género	0	.	
	Deleted Effect 1	Ano*Escalão Etário*Género	5	,000	3
1	Generating Class ^b	Ano*Escalão Etário*Género	0	.	

a. At each step, the effect with the largest significance level for the Likelihood Ratio Change is deleted, provided the significance level is larger than ,050.

b. Statistics are displayed for the best model at each step after step 0.

c. For 'Deleted Effect', this is the change in the Chi-Square after the effect is deleted from the model.

Tabela 5.1: Tabela de Step summary para tabela tridimensional.

Em termos da tabela *Step Summary*, a escolha do modelo a ajustar aos dados inicia-se neste exemplo com o modelo saturado e verifica-se que o modelo saturado ajusta bem aos dados. A 1ª etapa do *Step Summary*, mostra que a classe gerada é representada simbolicamente por (ano*escalão etário*género).

A tabela seguinte permite nos também concluir que existe interação significativa entre as variáveis *escalão etário*, *género* e *ano*.

	K	df	Likelihood Ratio		Pearson		Number of Iterations
			Chi-Square	Sig.	Chi-Square	Sig.	
K-way and Higher Order Effects ^a	1	23	77569,888	,000	82507,678	,000	0
	2	16	1411,792	,000	1366,801	,000	2
	3	5	165,479	,000	166,417	,000	3
K-way Effects ^b	1	7	76158,096	,000	81140,876	,000	0
	2	11	1246,313	,000	1200,385	,000	0
	3	5	165,479	,000	166,417	,000	0

a. Tests that k-way and higher order effects are zero.

b. Tests that k-way effects are zero.

Tabela 5.2: Tabela de K-Way and Higher-Order Effects para tabela tridimensional.

Da análise da tabela *K-Way and Higher-order*, verifica-se que todo o efeito de ordem é estatisticamente significativo com $p - value \leq 0.0005$.

De seguida fornecemos a tabela das estimativas dos parâmetros.

Parameter Estimates ^{b,c}						
Parameter	Estimate	Std. Error	Z	Sig.	95% Confidence Interval	
					Lower Bound	Upper Bound
Constant	5,176	,075	68,864	,000	5,029	5,323
[Ano = 1]	-1,151	,153	-7,506	,000	-1,451	-,850
[Ano = 2]	0 ^a
[Escalação.Etário = 1]	4,036	,076	53,226	,000	3,887	4,185
[Escalação.Etário = 2]	3,438	,076	45,022	,000	3,288	3,588
[Escalação.Etário = 3]	2,840	,077	36,733	,000	2,689	2,992
[Escalação.Etário = 4]	2,325	,079	29,529	,000	2,171	2,480
[Escalação.Etário = 5]	1,288	,085	15,177	,000	1,122	1,455
[Escalação.Etário = 6]	0 ^a
[Género = 1]	,055	,105	,524	,600	-,151	,260
[Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 1]	1,133	,154	7,361	,000	,832	1,435
[Ano = 1] * [Escalação.Etário = 2]	1,137	,155	7,362	,000	,835	1,440
[Ano = 1] * [Escalação.Etário = 3]	1,140	,155	7,334	,000	,835	1,445
[Ano = 1] * [Escalação.Etário = 4]	1,162	,157	7,410	,000	,855	1,470
[Ano = 1] * [Escalação.Etário = 5]	-,026	,174	-,147	,883	-,368	,315
[Ano = 1] * [Escalação.Etário = 6]	0 ^a
[Ano = 2] * [Escalação.Etário = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 3]	0 ^a
[Ano = 2] * [Escalação.Etário = 4]	0 ^a
[Ano = 2] * [Escalação.Etário = 5]	0 ^a
[Ano = 2] * [Escalação.Etário = 6]	0 ^a
[Ano = 1] * [Género = 1]	,564	,196	2,876	,004	,180	,949
[Ano = 1] * [Género = 2]	0 ^a
[Ano = 2] * [Género = 1]	0 ^a
[Ano = 2] * [Género = 2]	0 ^a
[Escalação.Etário = 1] * [Género = 1]	-,282	,106	-2,665	,008	-,490	-,075
[Escalação.Etário = 1] * [Género = 2]	0 ^a
[Escalação.Etário = 2] * [Género = 1]	-,692	,107	-6,445	,000	-,902	-,481
[Escalação.Etário = 2] * [Género = 2]	0 ^a
[Escalação.Etário = 3] * [Género = 1]	-,737	,109	-6,736	,000	-,952	-,523
[Escalação.Etário = 3] * [Género = 2]	0 ^a
[Escalação.Etário = 4] * [Género = 1]	-1,220	,115	-10,569	,000	-1,446	-,994
[Escalação.Etário = 4] * [Género = 2]	0 ^a
[Escalação.Etário = 5] * [Género = 1]	-1,090	,130	-8,376	,000	-1,345	-,835
[Escalação.Etário = 5] * [Género = 2]	0 ^a
[Escalação.Etário = 6] * [Género = 1]	0 ^a
[Escalação.Etário = 6] * [Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 1] * [Género = 1]	-,584	,197	-2,961	,003	-,971	-,197
[Ano = 1] * [Escalação.Etário = 1] * [Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 2] * [Género = 1]	-,526	,199	-2,648	,008	-,916	-,137
[Ano = 1] * [Escalação.Etário = 2] * [Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 3] * [Género = 1]	-,704	,201	-3,496	,000	-1,099	-,309
[Ano = 1] * [Escalação.Etário = 3] * [Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 4] * [Género = 1]	-,145	,206	-,704	,481	-,549	,259
[Ano = 1] * [Escalação.Etário = 4] * [Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 5] * [Género = 1]	,680	,231	2,938	,003	,226	1,134
[Ano = 1] * [Escalação.Etário = 5] * [Género = 2]	0 ^a
[Ano = 1] * [Escalação.Etário = 6] * [Género = 1]	0 ^a
[Ano = 1] * [Escalação.Etário = 6] * [Género = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 1] * [Género = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 1] * [Género = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 2] * [Género = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 2] * [Género = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 3] * [Género = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 3] * [Género = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 4] * [Género = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 4] * [Género = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 5] * [Género = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 5] * [Género = 2]	0 ^a
[Ano = 2] * [Escalação.Etário = 6] * [Género = 1]	0 ^a
[Ano = 2] * [Escalação.Etário = 6] * [Género = 2]	0 ^a

a. This parameter is set to zero because it is redundant.

b. Model: Poisson

c. Design: Constant + Ano + Escalação.Etário + Género + Ano * Escalação.Etário + Ano * Género + Escalação.Etário * Género + Ano * Escalação.Etário * Género

Tabela 5.3: Tabela de Parameter Estimates para tabela tridimensional.

A tabela *Cell Counts and Residuals* mostra que o modelo log-linear ajusta bem todas as células.

Cell Counts and Residuals^{a,b}

Ano	Escalaão Etário	Género	Observed		Expected		Residual	Standardized Residual	Adjusted Residual	Deviance
			Count	%	Count	%				
2006	15-24 anos	Masculino	7887	11,3%	7887,000	11,3%	,000	,000	,000	,000
		Feminino	9845	14,4%	9845,000	14,4%	,000	,000	,000	,000
	25-34 anos	Masculino	2986	4,4%	2986,000	4,4%	,000	,000	,000	,000
		Feminino	5436	8,0%	5436,000	8,0%	,000	,000	,000	,000
	35-44 anos	Masculino	1318	1,9%	1318,000	1,9%	,000	,000	,000	,000
		Feminino	2999	4,4%	2999,000	4,4%	,000	,000	,000	,000
	45-54 anos	Masculino	889	1,3%	889,000	1,3%	,000	,000	,000	,000
		Feminino	1832	2,7%	1832,000	2,7%	,000	,000	,000	,000
	55-64 anos	Masculino	244	,4%	244,000	,4%	,000	,000	,000	,000
		Feminino	198	,3%	198,000	,3%	,000	,000	,000	,000
	65 anos +	Masculino	104	,2%	104,000	,2%	,000	,000	,000	,000
		Feminino	56	,1%	56,000	,1%	,000	,000	,000	,000
2008	15-24 anos	Masculino	7980	11,7%	7980,000	11,7%	,000	,000	,000	,000
		Feminino	10017	14,7%	10017,000	14,7%	,000	,000	,000	,000
	25-34 anos	Masculino	2914	4,3%	2914,000	4,3%	,000	,000	,000	,000
		Feminino	5509	8,1%	5509,000	8,1%	,000	,000	,000	,000
	35-44 anos	Masculino	1532	2,2%	1532,000	2,2%	,000	,000	,000	,000
		Feminino	3031	4,4%	3031,000	4,4%	,000	,000	,000	,000
	45-54 anos	Masculino	565	,8%	565,000	,8%	,000	,000	,000	,000
		Feminino	1811	2,7%	1811,000	2,7%	,000	,000	,000	,000
	55-64 anos	Masculino	228	,3%	228,000	,3%	,000	,000	,000	,000
		Feminino	642	,9%	642,000	,9%	,000	,000	,000	,000
	65 anos +	Masculino	187	,3%	187,000	,3%	,000	,000	,000	,000
		Feminino	177	,3%	177,000	,3%	,000	,000	,000	,000

a. Model: Poisson

b. Design: Constant + Ano + Escalaão Etário + Género + Ano * Escalaão Etário + Ano * Género + Escalaão Etário * Género + Ano * Escalaão Etário * Género

Tabela 5.4: Tabela de *Cell Counts and Residuals* para tabela tridimensional.

Os resíduos ajustados estandardizados que, em modulo, são superiores a 1.96, com um nível de significância de 0.05, identificam as células em que o modelo tem fraco ajustamento.

A frequência esperada da célula (1,2,1) da tabela *Cell Counts and Residuals*, é 9845, donde o seu logaritmo é $\log(9845) = 9.19$.

Usando a tabela dos parâmetros, substituindo para a célula (1,2,1) vem:

$$\begin{aligned} \log u_{121} &= \lambda + \lambda_1^A + \lambda_2^B + \lambda_1^C + \lambda_{12}^{AB} + \lambda_{11}^{AC} + \lambda_{21}^{BC} + \lambda_{121}^{ABC} \\ &= 5.176 + 4.036 + 0 - 1.151 + 0 + 1.133 + 0 + 0 = 9,19 \end{aligned}$$

Da análise da tabela 5.3 verifica-se que existe efeito interactivo entre as variáveis género, escalão etário e o ano, localizamos dois casos em que não existe efeito interactivo no que diz respeito às variáveis ano.2006*escalão.etário (55 a 64) e ano.2006*escalão.etário (45 a 54)*género.masculino com $p - value = 0.883$ e $p - value = 0.481$. Localizamos também, um caso em que o efeito não é significativo no que diz respeito ao género.masculino com $p - value = 0.600$.

Quando o modelo está bem ajustado os resíduos Q-Q plot têm uma distribuição normal situando-se à volta da linha oblíqua, aproximando-se de um ângulo de 45 graus.

Quando o modelo está bem ajustado os resíduos Detrended Q-Q plot devem distribuir-se aleatoriamente à volta da linha horizontal.

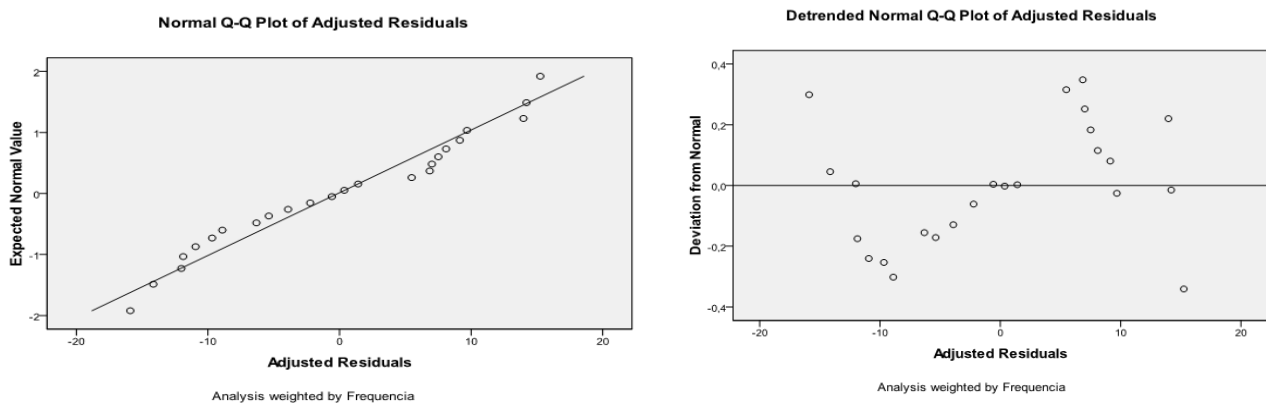


Figura 5.5: Gráficos de normal Q-Q Plot of Adjusted Residuals e de Detrended Normal Q-Q of Ajusted Residuals.

Ambos os gráficos mostram que o modelo se ajusta razoavelmente em todas as células.

5. CONCLUSÃO

Segundo os dados do Instituto de Emprego e Formação Profissional (IEFP) e Instituto Nacional de Estatística (INE), verificamos que a população activa (PA) em 2006 era 183.254 indivíduos, dos quais 33.574 encontravam-se na situação de desemprego. Em 2008 a PA cresceu para 198.855 indivíduos, o que significa que mais 15.601 pessoas passaram a contar para efeitos de cálculo da taxa de desemprego. Dos 198.855 indivíduos estavam desempregados 35.476 indivíduos, o que corresponde a uma taxa de desemprego de 17,8%.

De 2006 a 2008 a taxa de desemprego em Cabo Verde diminuiu ligeiramente de 18,3% para 17,8%, uma variação de 0,5%. No entanto verificou-se um aumento de 1902 indivíduos desempregados no ano 2008 em relação a 2006.

O nível de desemprego em Cabo Verde relativamente ao ano 2006 e 2008 apresenta disparidades acentuadas entre as diversas ilhas e concelhos do arquipélago. Pelo menos em duas ilhas, Sal e Fogo, a taxa de desemprego é inferior a 15%. Em sentido contrário encontra-se os concelhos da Praia, o Interior de Santiago, ilha de Santo Antão e São Vicente onde a taxa de desemprego ronda os 26,7%. No ano 2006, no interior de Santiago apresenta uma maior proporção de desemprego enquanto que no ano 2008 na Cidade da Praia que obteve-se uma maior proporção de desemprego. No interior de Santiago e na cidade da Praia são domínios com maiores contribuições para o desemprego nacional, isto justifica-se pelas suas dimensões populacionais mais elevadas.

A nível do ramo de actividade verifica-se uma grande discrepância do desemprego em Cabo Verde entre cada ramo de actividade nos dois anos em estudo. No ano 2006 apresenta uma maior proporção de desemprego no ramo da construção, enquanto que no ano 2008 obteve-se uma maior proporção de desempregado no grupo onde os desempregados não estão inseridos em nenhum ramo de actividades (NR).

Os ramos de actividades onde o desemprego cresceu são apenas os ramos de actividades financeiras (1%) e RN - Outros (30%). Na agricultura e pesca, indústria extractiva, indústria transformadora e electricidade, construção, comércio, alojamento e restauração, transporte e comunicações, serviços às empresas, administração pública,

educação, outros serviços, famílias com empregados, e organismos internacionais, em todos estes ramos houve uma ligeira descida de desemprego.

A nível da situação na profissão, no ramo do sector empresarial privado apresenta uma maior proporção de desemprego nos dois anos. A nível da profissão verifica-se uma maior proporção de desemprego no ramo de trabalhadores não qualificados nos anos 2006 e 2008.

O nível de desemprego em Cabo Verde relativamente ao ano 2006 e 2008 apresenta disparidades acentuadas entre as diversas ilhas e concelhos do arquipélago. Pelo menos em duas ilhas, Sal e Fogo, a taxa de desemprego é inferior a 15%. Em sentido contrário encontra-se os concelhos da Praia, o Interior de Santiago, ilha de Santo Antão e São Vicente onde a taxa de desemprego ronda os 26,7%.

Em termos do Teste de Homogeneidade verificamos que a distribuição do desemprego em Cabo Verde a nível do domínio em 2006 não é semelhante a de 2008.

Da análise da tabela de contingência em estudo verifica-se que há mais desemprego na camada jovem no ano 2006 do que no ano 2008 com uma diferença de 0,19%.

O desemprego em Cabo Verde afecta sobretudo, a camada jovem com idade compreendida entre os 15 e os 24 anos. Observamos no género feminino, uma maior proporção de desempregados no grupo etário de 35 aos 44 anos (69,5%) do que no género masculino (30,5%) no ano 2006, enquanto que no ano 2008 há uma maior proporção de desempregados do género feminino em Cabo Verde no grupo etário do 45 aos 54 anos (76,2%) do que no género masculino (23,8%).

Com o Teste de Independência entre as variáveis escalão etário e género do desemprego em Cabo Verde no ano 2006 e 2008, concluímos que não existe independência entre as variáveis em estudo.

Com as contribuições individuais de Qui-Quadrado, verificamos que no grupo etário entre 15 a 24 anos do género masculino nos anos 2006 e 2008 são os grupos que mais contribuem para associação das variáveis.

Com o estudo das medidas de associação verificamos de outro modo que existe associação entre o escalão etário e o género do desemprego em Cabo Verde nos anos 2006 e 2008, aparenta ser fraca mas no entanto é significativa.

Da análise de tabela de contingências tridimensionais do desemprego em Cabo Verde concluímos que as medidas de associação parcial e de associação múltipla entre as variáveis escalão etário e o género dado o ano são relativamente fracas.

Com aplicação do modelo Log-lineares nas tabelas bidimensionais dos desempregados em Cabo Verde no ano 2006 e 2008, concluímos que o modelo saturado tem um ajustamento perfeito e que ajusta bem a todas as células da tabela em estudo. A tabela dos parâmetros estimados mostra para um nível de significância de 0.05 que existe efeito interactivo entre o género e o escalão etário dos desempregados em Cabo Verde no ano 2006 e 2008. Concluimos também que o melhor modelo inclui o efeito interactivo. E em relação ao modelo Log-lineares nas tabelas tridimensionais, concluímos que existe efeito interactivo entre as variáveis género, escalão etário e o ano.

Para aplicação da metodologia que escolhemos, utilizamos dois softwares diferentes, Excel 2007 e SPSS Statistics 17.0 vocacionadas para análise de dados em estudo.

Apesar de todas as conclusões que descrevi anteriormente não consegui atingir o meu objectivo principal, enunciado na introdução “ identificar as variáveis que estão mais decisivamente associadas ao desemprego em Cabo Verde e descrever um modelo para essa relação. ”

As variáveis associadas ao desemprego foram estudados individualmente, ao nível univariado, e a análise foi essencialmente descritiva.

A principal dificuldade que encontrámos foi a nível de organização dos dados. Só dispúnhamos de informação conjunto para as variáveis género e escalão etário.

Um modelo para descrever o desemprego em Cabo Verde, teria de ter necessariamente mais variáveis e portanto precisávamos de informação conjunta sobre várias variáveis. Por exemplo, nesse caso poderíamos considerar mais variáveis explicativas nos modelos Log – lineares.

O ideal seria informação ao nível individual, ou seja para cada indivíduo dispor do conhecimento da sua idade, género, domínio, ramo de actividade, profissão, situação na profissão. Nesta situação poderíamos avançar para modelos mais explicativas, por exemplo, Análise de Regressão. Sugerimos, pois, que estas melhorias fossem tomadas em consideração, em futuras recolhas de informação sobre a situação no desemprego em Cabo Verde.

6. TABELAS DE DADOS REAIS

Tabelas dos dados utilizados no estudo:

<i>Domínio</i>	<i>Empregado</i>	<i>Desemprego</i>	<i>Total</i>
<i>Santo Antão</i>	12.065	4.487	16.552
<i>São Vicente</i>	24.696	8.179	32.875
<i>Sal</i>	8.699	860	9.559
<i>Interior de Santiago</i>	48.030	10.364	58.394
<i>Praia</i>	42.781	7.862	50.643
<i>Fogo</i>	13.409	1.822	15.231
Total	149.680	33.574	366.508

Tabela 6.1: Dados dos desempregados e empregados a nível do domínio em Cabo Verde no ano 2006.

<i>Profissão</i>	<i>Empregado</i>	<i>Desempregado</i>	<i>Total</i>
<i>Forças Armadas</i>	190	0	190
<i>Q.Sup.da Ap,Dir.e Q.Sup.</i>	6.704	145	6.849
<i>Especialistas</i>	4.469	33	4.502
<i>Técnico e Prof. De nível Intermédio</i>	6.861	246	7.107
<i>Pessoal Administrativo e Similares</i>	6.758	911	7.669
<i>Pessoal dos Serviços e Vendedores</i>	15.970	1.980	17.950
<i>Agric. E Trab.Qual. Da Agric. E Pescas</i>	21.393	621	22.014
<i>Operários,Artífices e Similares</i>	25.490	3.947	29.437
<i>Oper. De Inst. De Máq de Trab. Da Montagem</i>	6.004	669	6.673
<i>Trabalhadores não Qualificados</i>	54.328	9.084	63.412
<i>NR</i>	1.511	584	2.095
Total	149678	18220	167.898

Tabela 6.2: Dados dos desempregados e empregados a nível da profissão em Cabo Verde no ano 2006.

<i>Situação na profissão</i>	<i>Empregado</i>	<i>Desempregado</i>	<i>Total</i>
<i>Administração Pública</i>	21.139	3.730	24.869
<i>Sector Empres.Privado</i>	37.512	6.968	44.480
<i>Sector Empres.Estado</i>	3.323	330	3.653
<i>Conta próp.c/Pes.ao Serv.</i>	7.699	385	8.084
<i>Conta próp.s/Pes.ao Serv.</i>	46.433	2.135	48.568
<i>Trab.Família.Sem Remun</i>	18.117	458	18.575
<i>Casa de Família</i>	12.005	2.689	14.694
<i>Outra Situação</i>	2.879	962	3.841
<i>NR</i>	573	566	1.139
<i>Total</i>	149.680	18.223	167.903

Tabela 6.3: Dados dos desempregados e empregados a nível da situação na profissão em Cabo Verde no ano 2006.

<i>Ramo de Actividade</i>	<i>Empregado</i>	<i>Desempregado</i>	<i>Total</i>
<i>Agricultura e Pesca</i>	45.125	967	46.092
<i>Indústria Extractiva</i>	5.611	217	5.828
<i>Indústria Transf. E Electricidade</i>	10.662	1.596	12.258
<i>Construção</i>	14.085	3.922	18.007
<i>Comércio</i>	23.931	1.752	25.683
<i>Alojamento e Restauração</i>	4.410	1.130	5.540
<i>Transportes e Comunicações</i>	8.367	598	8.965
<i>Actividades Financeiras</i>	710	28	738
<i>Serviços às Empresas</i>	3.166	468	3.634
<i>Administração Pública</i>	12.771	3.307	16.078
<i>Educação</i>	7.216	256	7.472
<i>Saúde</i>	1.792	33	1.825
<i>Outros Serviços</i>	3.057	867	3.924
<i>Famílias com Empregados</i>	6.581	1.986	8.567
<i>Organismo Internacionais</i>	489	196	685
<i>NR</i>	1.705	897	2.602
<i>Total</i>	149.678	18220	167.898

Tabela 6.4: Dados dos desempregados e empregados a nível do ramo de actividade em Cabo Verde no ano 2006.

<i>Domínio</i>	<i>Empregado</i>	<i>Desemprego</i>	<i>Total</i>
<i>Santo Antão</i>	13.760	3.519	17.279
<i>São Vicente</i>	25.509	9.727	35.236
<i>Sal</i>	8.246	1.430	9.676
<i>Interior de Santiago</i>	61.598	7.034	68.632
<i>Praia</i>	41.294	11.498	52.792
<i>Fogo</i>	12.972	2.268	15.240
<i>Total</i>	163.379	35.476	198.855

Tabela 6.5: Dados dos desempregados e empregados a nível do domínio em Cabo Verde no ano 2008.

<i>Profissão</i>	<i>Empregado</i>	<i>Desempregado</i>	<i>Total</i>
<i>Forças Armadas</i>	185	38	223
<i>Q.sup.da Ap,Dir.e Q.Sup.</i>	1.294	0	1.294
<i>Especialistas</i>	3.263	156	3.419
<i>Técnico e Prof. De nível Intermédio</i>	4.990	48	5.038
<i>Pessoal Administrativo e Similares</i>	5.415	363	5.778
<i>Pessoal dos Serviços e Vendedores</i>	21.783	1.936	23.719
<i>Agric. E Trab.Qual. Da Agric. E Pescas</i>	19.408	186	19.594
<i>Operários,Artifices e Similares</i>	18.734	1.666	20.400
<i>Oper. De Inst. De Máq de Trab. Da Montagem</i>	7.177	327	7.504
<i>Trabalhadores não Qualificados</i>	67.804	6.245	74.049
<i>NR</i>	13.328	5.756	19.084
<i>Total</i>	163381	16721	180.102

Tabela 6.6: Dados dos desempregados e empregados a nível da profissão em Cabo Verde no ano 2008.

<i>Situação na profissão</i>	<i>Empregado</i>	<i>Desempregado</i>	<i>Total</i>
<i>Administração Pública</i>	19.159	1.564	20.723
<i>Sector Empres.Privado</i>	39.226	5.543	44.769
<i>Sector Empres.Estado</i>	4.610	630	5.240
<i>Conta próp.c/Pes.ao Serv.</i>	13.154	646	13.800
<i>Conta próp.s/Pes.ao Serv.</i>	39.508	787	40.295
<i>Trab.Família.Sem Remun</i>	22.979	242	23.221
<i>Casa de Família</i>	12.353	1.634	13.987
<i>Outra Situação</i>	3.108	278	3.386
<i>NR</i>	9.284	5.397	14.681
<i>Total</i>	163.381	16.721	180.102

Tabela 6.7: Dados dos desempregados e empregados a nível da situação na profissão em Cabo Verde no ano 2008.

<i>Ramo de Actividade</i>	<i>Empregado</i>	<i>Desempregado</i>	<i>Total</i>
<i>Agricultura e Pesca</i>	53.563	54262	107.825
<i>Indústria Extractiva</i>	3.077	3169	6.246
<i>Indústria Transf. E Electricidade</i>	9.761	10.405	20.166
<i>Construção</i>	18.413	21.307	39.720
<i>Comércio</i>	24.255	25.945	50.200
<i>Alojamento e Restauração</i>	4.412	5.095	9.507
<i>Transportes e Comunicações</i>	8.791	9.349	18.140
<i>Actividades Financeiras</i>	576	701	1.277
<i>Serviços às Empresas</i>	2.340	2.711	5.051
<i>Administração Pública</i>	11.456	12.560	24.016
<i>Educação</i>	6.197	6.433	12.630
<i>Saúde</i>	1.295	1.308	2.603
<i>Outros Serviços</i>	1.830	1.961	3.791
<i>Famílias com Empregados</i>	6.947	8.508	15.455
<i>Organismo Internacionais</i>	0	12	12
<i>NR</i>	10.195	16.027	26.222
<i>Total</i>	163.108	179753	342.861

Tabela 6.8: Dados dos desempregados e empregados a nível do ramo de actividade em Cabo Verde no ano 2008.

7. BIBLIOGRAFIA E REFERÊNCIAS

7.1. Cabo Verde

Instituto Nacional de Estatística (2006-2008), inquérito aos desempregados – *Manuel do Agente de Terreno*, Praia.

Instituto de Emprego e Formação Profissional (2006-2008), inquérito aos desempregados – *Manuel do Agente de Terreno*, Praia.

Gabinete da Secretaria de Estado para a Luta Contra Pobreza (n.d), *Programa Nacional De Luta Contra a Pobreza: Documento-Quadro*, Praia.

7.2. Geral

Agresti Alan, Finlay Barbara (1997) – *Statistical methods for the Social Sciences*.

Agresti, A. (2002), *Categorical Data Analysis*. (Second Edition) New York: Jonh Wiley and Sons.

Andersen, Erling.B (1997) – *Introduction to the Statistical Analysis of Categorical data*.

Christensen, R. (1997) - *Log-Linear Models and Logistic Regression*. New York: Springer – Verlag

Powers, Daniel. A, Yu Xie (2008) - *Statistical methods for Categorical Data Analysis*.

Everitt.B.S (1992) – *The Analysis of Contingency Tables*: Second Edition.

Lindsey k. Jamas (1989) – *The analysis of Categorical data using GLIM*.

Gomes, P. (1993), *Análise de Dados*, ISEGI, Lisboa.

Murteira, B., (1993) *Análise Exploratória de Dados – Estatística Descritiva*, McGraw – Hill, Lisboa.

Pestana, Dinis Duarte. Velosa, Sílvio Filipe (2002) - *Introdução à Probabilidade e à Estatística*.

Pereira Paula Cristina Sequeira (2001) – *Medidas de Associação*: Tese de Mestrado, Universidade de Lisboa.

Pinto Susana, Isabel Pedrosa (2006) – *Análise de Tabelas de Contingência Ordinais*: Tese de Mestrado, Universidade de Lisboa.

Pestana Maria Helena, Gageiro. João Nunes (2009) – *Análise Categórica, Árvores de Decisão e Análise de Conteúdo*: Em Ciências Sociais e da Saúde com o SPSS.

Paulino, C.D.M. e Singer, J.M. (1997) - *Análise de Dados Categorizados*.

Paulino, C.D. and Singer, J.M. (2006) - *Análise de dados categorizados (Analysis of categorical data*, in Portuguese). São Paulo: Edgard Blücher.

Plackett.R.L (1981) – *The Analysis of Categorical data*.

Peck Roxy, Olsen Chris, Devore L. Jay (2008) - *Introduction to Statistical & Data Analysis*.

Silvestre António Luís (2006) – *Análise de Dados e Estatística Descritiva*.

Simonoff, J.S. (2003), *Analysing Categorical Data*. New York Springer – Verlag.

Walter A.Shewhart, Samuel S. Wilks (2002) – *Categorical Data Analysis*: Wiley Series in Probability and Statistics.

