Instituto Superior de Economia e Gestão
UNIVERSIDADE TÉCNICA DE LISBOA

Mestrado em Marketing

Análise Quantitativa de Dados em Marketing

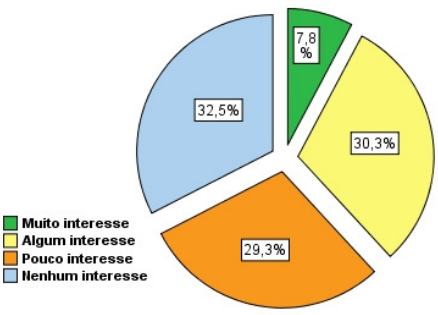
Análise de Dados univariada
Análise de Dados Bivariada
(o essencial)

Qual o seu interesse pela política

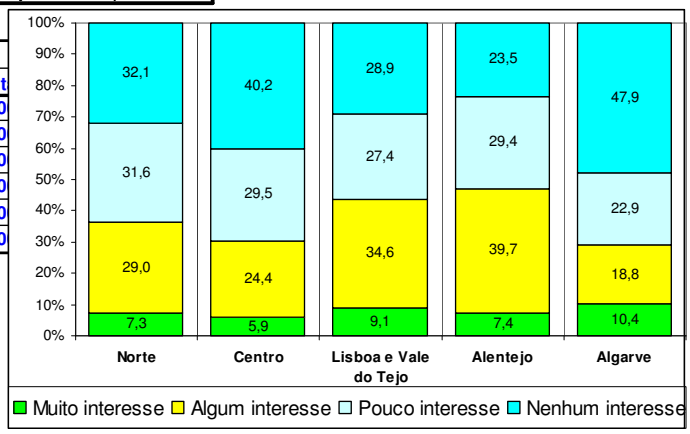
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Muito interesse	117	7,7	7,8	7,8
	Algum interesse	456	30,2	30,3	38,1
	Pouco interes				
	Nenhum inter				
	Total				
Missing	Recusa				
	Não sabe				
	Total				
Total					

		N	%
Qual o seu interesse pela política	Muito interesse	117	7,8
	Algum interesse	456	30,3
	Pouco interesse	441	29,3
	Nenhum interesse	489	32,5
	Total	1503	100,0

Qual o seu interesse pela política



Região	Qual o seu interesse pela política				
	Muito	Algum	Pouco	Nenhum	Tot
Norte	7,3	29,0	31,6	32,1	10
Centro	5,9	24,4	29,5	40,2	10
Lisboa e Vale do Tejo	9,1	34,6	27,4	28,9	10
Alentejo	7,4	39,7	29,4	23,5	10
Algarve	10,4	18,8	22,9	47,9	10
Total	7,8	30,3	29,3	32,5	10



Docente: ***Rui Brites***
e-mail: rui.brites@iscte.pt
Telef.: 217903485/969073534
Skype Name: britesrui

Setembro 2010

ÍNDICE

1. Análise de Dados Univariada ……. 1
1.1. Frequências ……..... 2
1.1.1, Procedimento *Frequencies* ……... 2
1.1.2. Procedimento *Tables* …….. 3
1.1.3. Quadros síntese de frequências de várias variáveis ……. 4
1.2. Estatísticas descritivas e gráficos de perfil ……..... 5
1.2.1. Procedimento *Descriptives* ……... 5
1.2.2. Procedimento *Tables* ……. 6
1.2.3. Gráfico de Perfil ……... 7
1.3. Testes de aderência não paramétricos…… 9
1.3.1. χ^2 (Qui-quadrado) ……..... 9
1.3.2. *Kolmogorov-Smirnov* …….. 10
1.3.3. Teste de aderência paramétrico: *t-Student* ……..... 11
2. Análise de questões de resposta múltipla ……..... 12
2.1. Criar uma variável agregada com a resposta aos diversos items …….. 13
2.2. Frequência de uma questão de resposta múltipla ……. 14
2.3. Cruzamento de uma questão de resposta múltipla …….. 15
3. Análise de Dados Bivariada ……... 16
3.1. Cruzamentos e teste de independência χ^2 (Qui-quadrado) ……..... 17
3.1.1. Procedimento *Crosstabs* ……... 17
3.1.2. Procedimento *Tables* ……. 18
3.2. Testes não paramétricos – procedimento *Non Parametric Tests* …….. 19
3.2.1. Duas amostras independentes: Teste de *Mann-Whitney* ……..... 19
3.2.2. K amostras independentes: Teste de *Kruskal-Wallis* ……... 20
3.3. Testes paramétricos ……... 21
3.3.1. Duas amostras emparelhadas (*t-Student* para amostras emparelhadas) …. 21
3.3.2. Duas amostras independentes (*t-Student* de independência) …… 22
3.3.3. k amostras independentes (Análise de Variância Simples Paramétrica - ANOVA) …. 23
4. Correlação …….. 26
4.1. Correlação não linear: *rho de Spearman* ……..... 26
4.2. Correlação linear simples: *r de Pearson* ……..... 28
5. Regressão linear simples ……... 29
Bibliografia …….. 31
Links úteis relacionados com a matéria deste manual ……... 31
Anexo 1: Testes e procedimentos estatísticos mais utilizados em análise bivariada 32
Anexo 2: Extracto do Questionário *European Social Survey* (*round* 1- 2002) ……... 35

Observação muito importante:

Este manual não dispensa a consulta de obras de referência da análise estatística. Recomenda-se, para o efeito:

Maroco, J. (2007) *Análise Estatística Com o PASW Statistics (ex-SPSS)*, Lisboa, Wook.

Análise de Dados Univariada

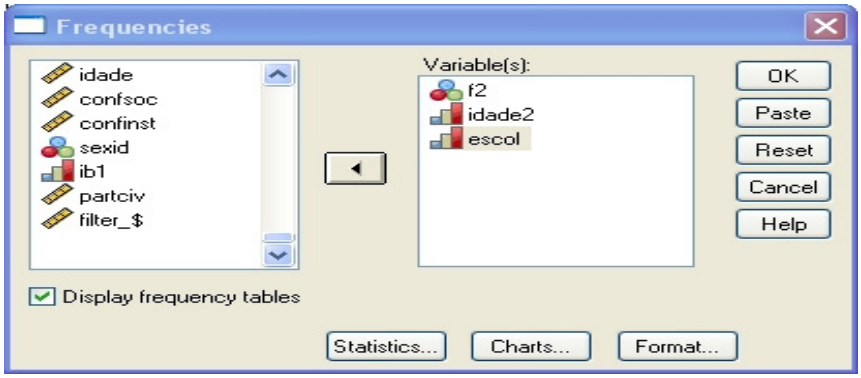
1. Análise de Dados univariada

Nota: A base de dados que vamos utilizar neste capítulo é um extracto da base de dados do *European Social Survey* (*round* 1)[1] com os resultados da aplicação do questionário em Portugal (ficheiro **ESS Portugal 2002 (TPAUB1).sav**)

1.1. Frequências

1.1.1. Utilizando o procedimento *Frequencies*[2]

Variáveis de caracterização social:



Os resultados são os seguintes:

Sexo

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Masculino	630	41.7	41.7	41.7
	Feminino	881	58.3	58.3	100.0
	Total	**1511**	**100.0**	**100.0**	

Idade

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Até 30 anos	341	22.6	22.6	22.6
	31 - 50 anos	505	33.4	33.4	56.0
	51 - 65 anos	315	20.8	20.8	76.8
	> 65 anos	350	23.2	23.2	100.0
	Total	**1511**	**100.0**	**100.0**	

Anos de escolaridade concluídos

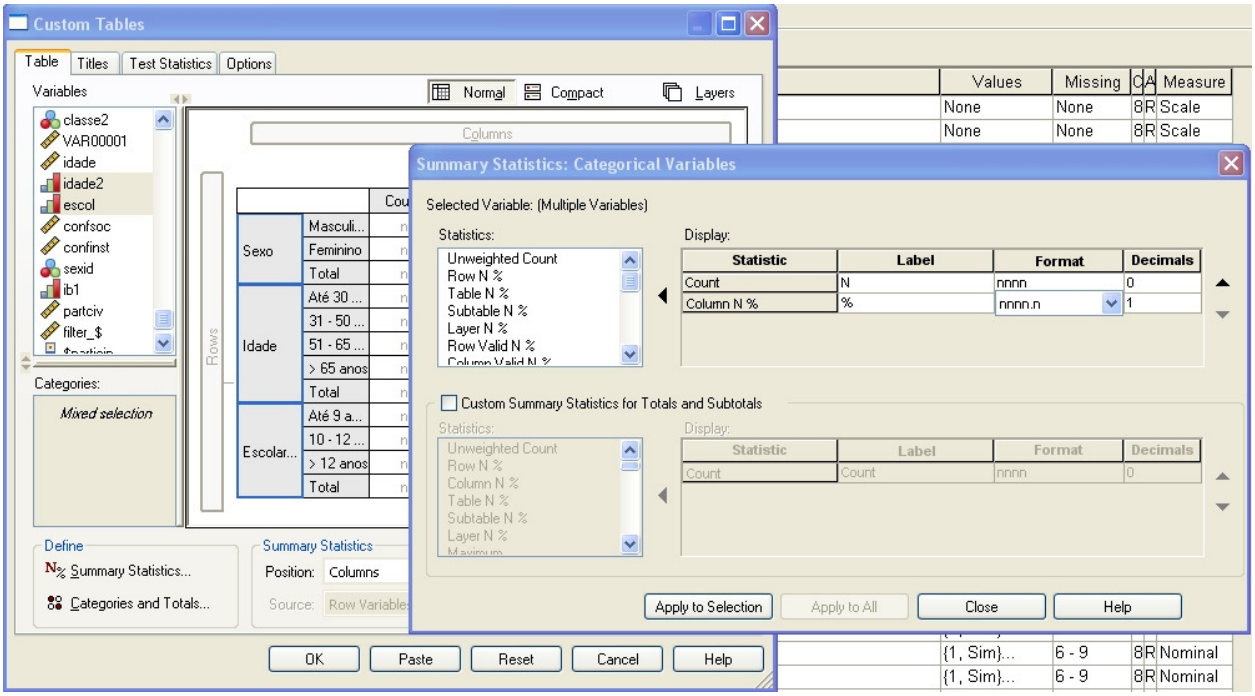
		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	Até 9 anos	1046	69.2	69.3	69.3
	10 - 12 anos	251	16.6	16.6	86.0
	> 12 anos	212	14.0	14.0	100.0
	Total	**1509**	**99.9**	**100.0**	
Missing	NR	2	.1		
Total		**1511**	**100.0**		

[1] A base de dados original está disponível em http://www.europeansocialsurvey.org/.

[2] Nota: as tabelas geradas pelo procedimento *Frequencies* são em formato "rascunho" destinando-se apenas ao controlo e validação da base de dados, com o objectivo de eliminar erros de introdução de dados.

1.1.2. Utilizando o procedimento *Tables*

Variáveis de caracterização social:

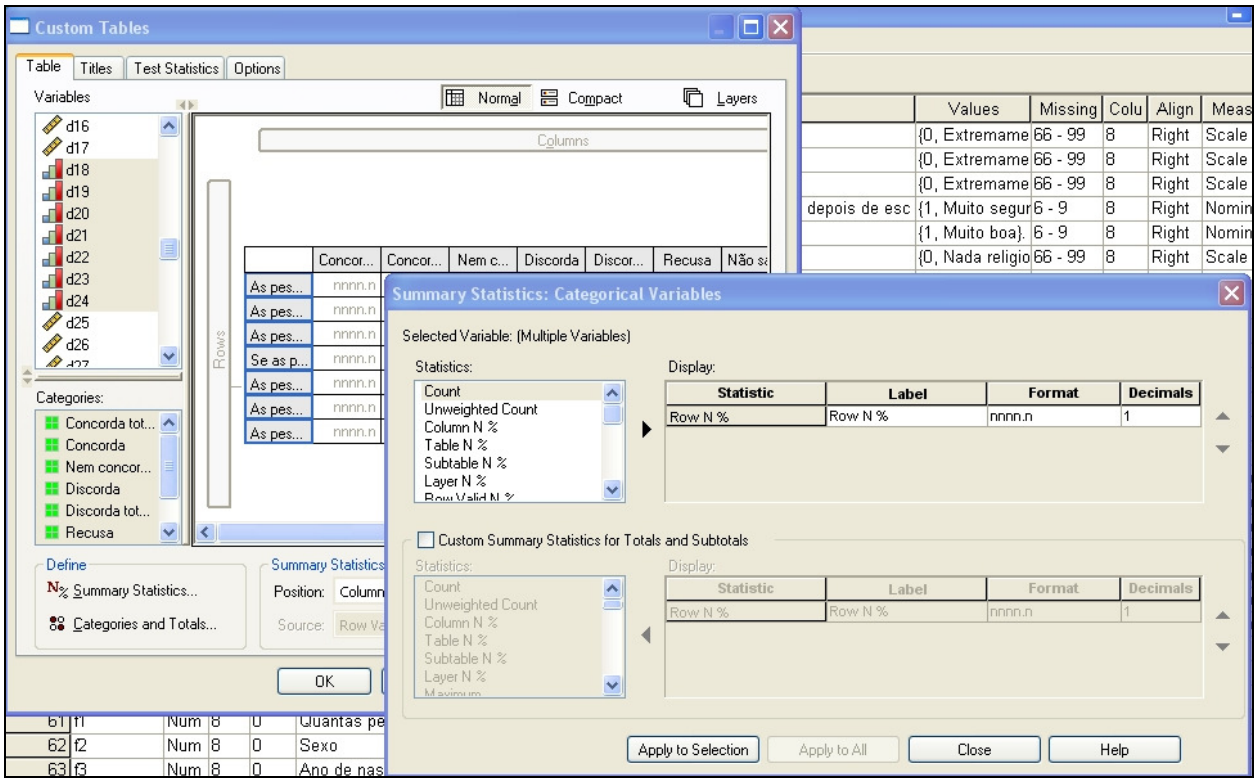


Os resultados são os seguintes[3]:

		N	%
Sexo	Masculino	630	41.7
	Feminino	881	58.3
	Total	**1511**	**100.0**
Idade	Até 30 anos	341	22.6
	31 - 50 anos	505	33.4
	51 - 65 anos	315	20.8
	> 65 anos	350	23.2
	Total	**1511**	**100.0**
Anos de escolaridade concluídos	Até 9 anos	1046	69.2
	10 - 12 anos	251	16.6
	> 12 anos	212	14.0
	NR	2	.1
	Total	**1511**	**100.0**

[3] Nota: o quadro foi modificado no respectivo editor, a que se acede "clicando" duas vezes sobre o mesmo.

1.1.3. Quadro sínteses de frequências dos indicadores das questões *d18* a *d24*



O resultado é o seguinte:

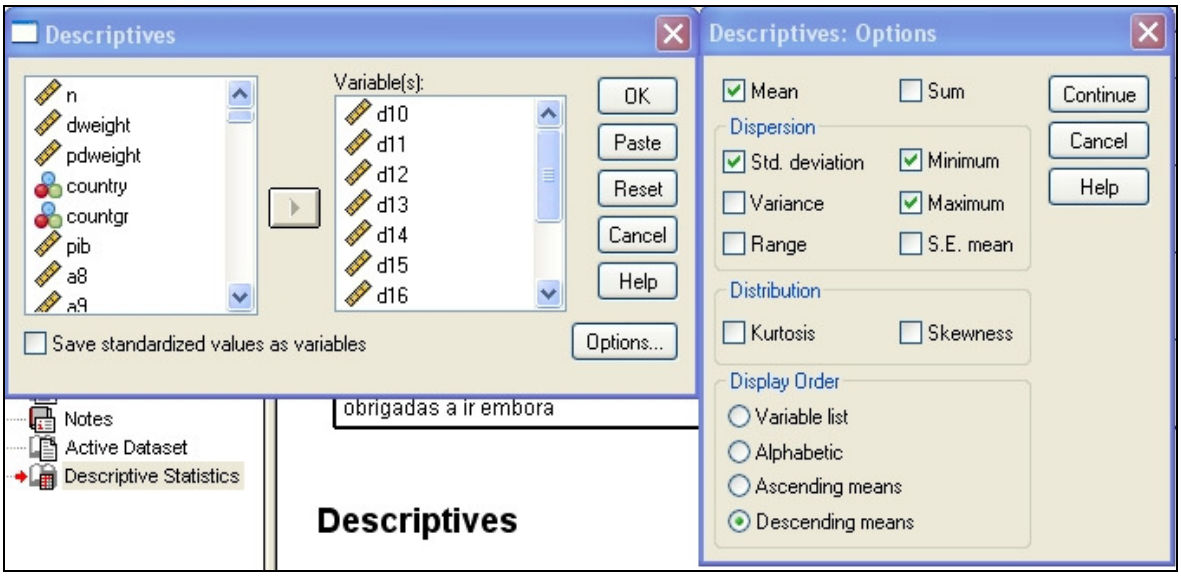
	Concorda totalmente	Concorda	Nem concorda	Discorda	Discorda totalmente	Total
As pessoas que vêm viver e trabalhar para cá fazem com que os salários baixem	11.7	41.3	15.1	25.4	6.5	100.0
As pessoas que vêm viver e trabalhar para cá, em regra, prejudicam mais as expectativas económicas dos pobres do que dos ricos	14.1	46.1	19.4	16.2	4.3	100.0
As pessoas que vêm viver e trabalhar para cá ajudam a preencher lugares em que há falta de trabalhadores	13.7	53.7	16.7	12.3	3.6	100.0
Se as pessoas que vieram viver e trabalhar para cá estiverem desempregadas por muito tempo deviam ser obrigadas a ir embora	14.3	42.3	22.1	16.2	5.2	100.0
As pessoas que vieram viver para cá devem ter os mesmos direitos do que todas as outras pessoas	25.9	54.0	12.5	6.0	1.6	100.0
As pessoas que vieram viver para cá cometerem um crime grave, devem ser obrigadas a ir embora	43.3	40.1	9.3	6.2	1.0	100.0
As pessoas que vieram viver para cá cometerem qualquer crime, devem ser obrigadas a ir embora	26.6	38.5	19.0	13.2	2.7	100.0

1.2. Estatísticas descritivas e gráfico de perfil

Nota muito importante: A análise estatística, excepto nos quadros de frequências, deve incidir apenas nas respostas válidas. Assim, antes de efectuar qualquer análise estatística, torna-se necessário definir e activar os respectivos ***missing values*** (não responde/não sabe/não se aplica) na coluna ***missing*** da base de dados.

1.2.1. Utilizando o procedimento *Descriptives*[4]

Indicadores das questões *d10* a *d17*:



Os resultados são os seguintes:

Qual a importância dos seguintes aspectos para deixar vir para cá alguém que nasceu, foi educado e viveu fora de Portugal:

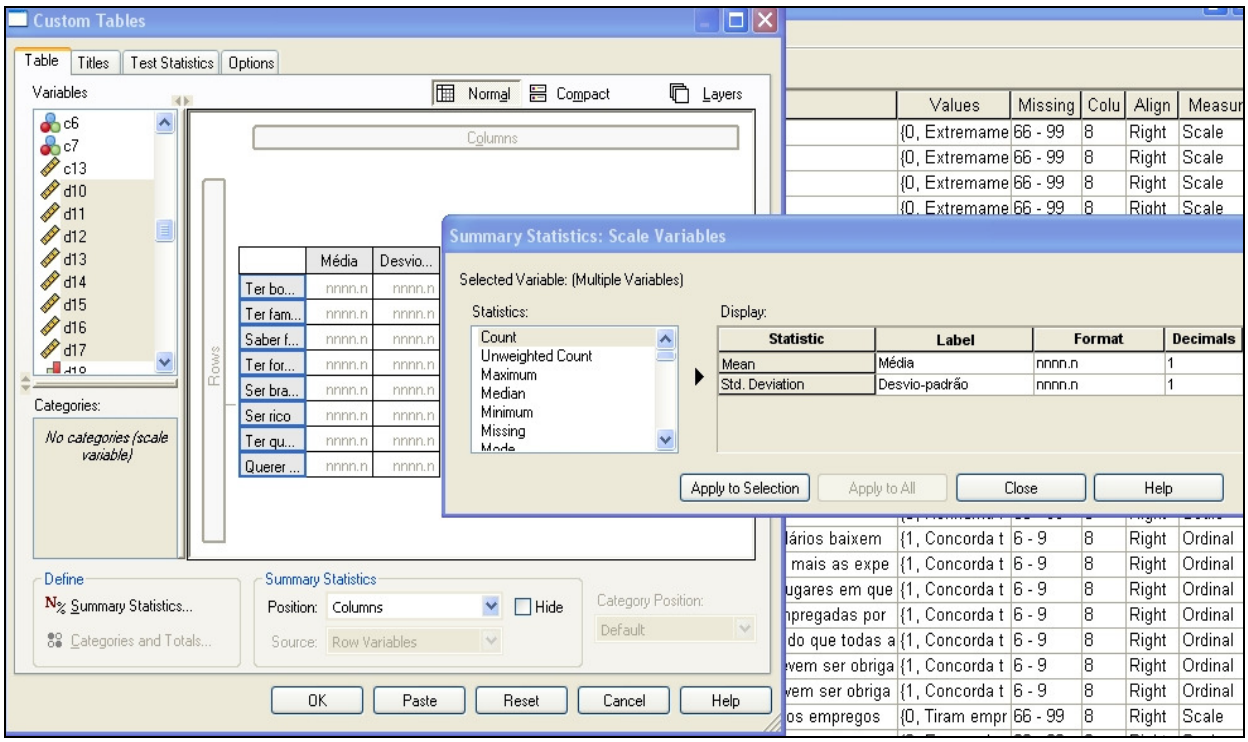
Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation
Ter qualificações profissionais de que o país precis	1448	0	10	7.46	2.241
Querer adaptar-se ao mesmo modo de vida do país	1449	0	10	7.11	2.442
Ser rico	1447	0	10	6.81	2.807
Ter familiares próximos a viver cá	1454	0	10	6.48	2.425
Ter boas qualificações académicas	1416	0	10	6.06	2.536
Saber falar a língua oficial do país	1466	0	10	6.05	2.633
Ter formação cristã	1440	0	10	3.79	2.971
Ser branco	1451	0	10	2.85	2.874
Valid N (listwise)	1330				

[4] Nota: as tabelas geradas pelo procedimento *Descriptives* são em formato "rascunho" destinando-se apenas ao controlo e validação da base de dados, com o objectivo de eliminar erros de introdução de dados.

1.2.2. Utilizando o procedimento *Tables*

Indicadores das questões *d10* a *d17*:



Os resultados são os seguintes:

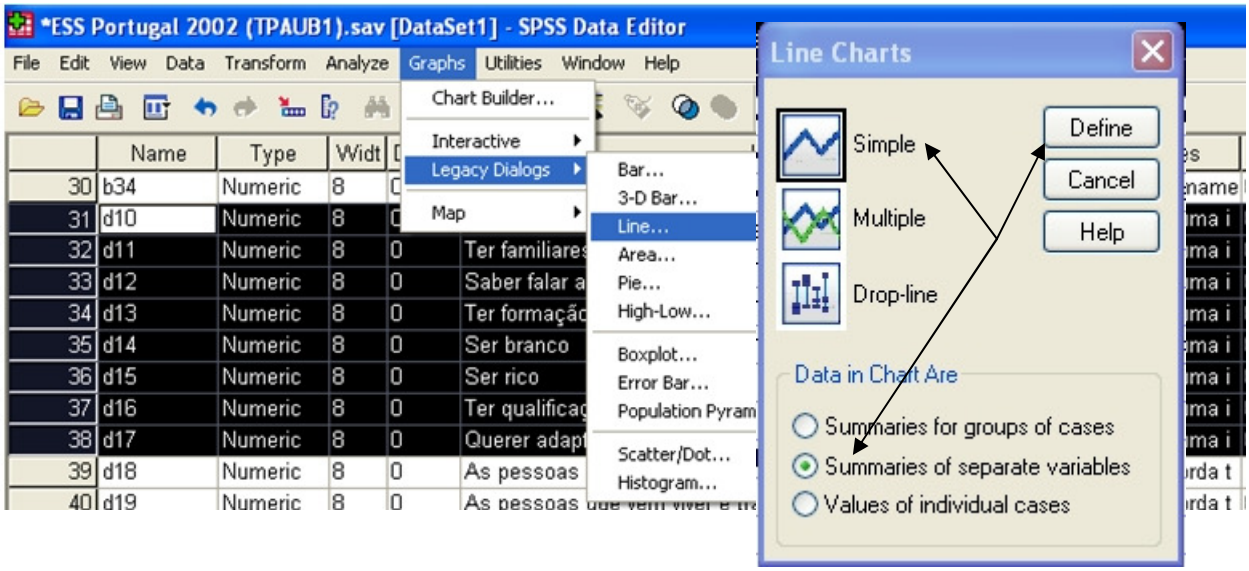
Qual a importância dos seguintes aspectos para deixar vir para cá alguém que nasceu, foi educado e viveu fora de Portugal:

	Média	Desvio-padrão
Ter boas qualificações académicas	6.1	2.5
Ter familiares próximos a viver cá	6.5	2.4
Saber falar a língua oficial do país	6.1	2.6
Ter formação cristã	3.8	3.0
Ser branco	2.8	2.9
Ser rico	6.8	2.8
Ter qualificações profissionais de que o país precisa	7.5	2.2
Querer adaptar-se ao mesmo modo de vida do país	7.1	2.4

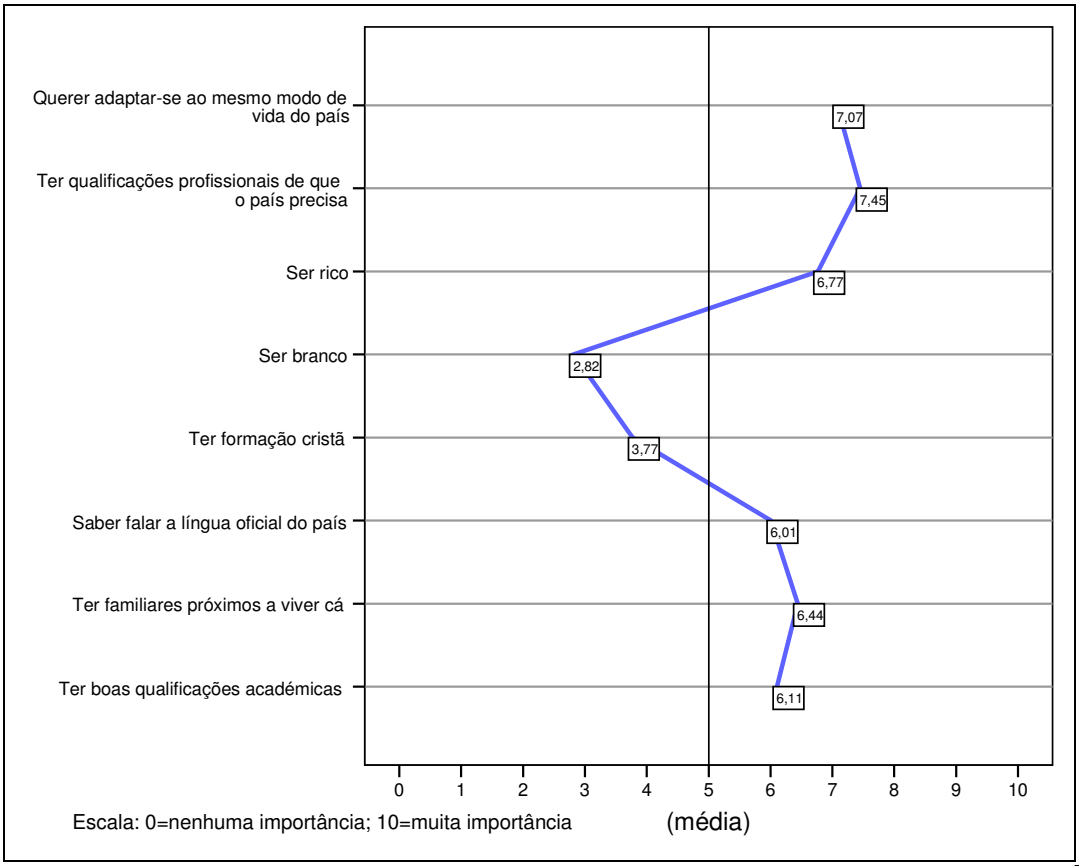
Escala: 0=nenhuma importância; 10=muita importância

1.2.3. Gráfico de perfil

a) Feito em SPSS[5]:



Qual a importância dos seguintes aspectos para deixar vir para cá alguém que nasceu, foi educado e viveu fora de Portugal:

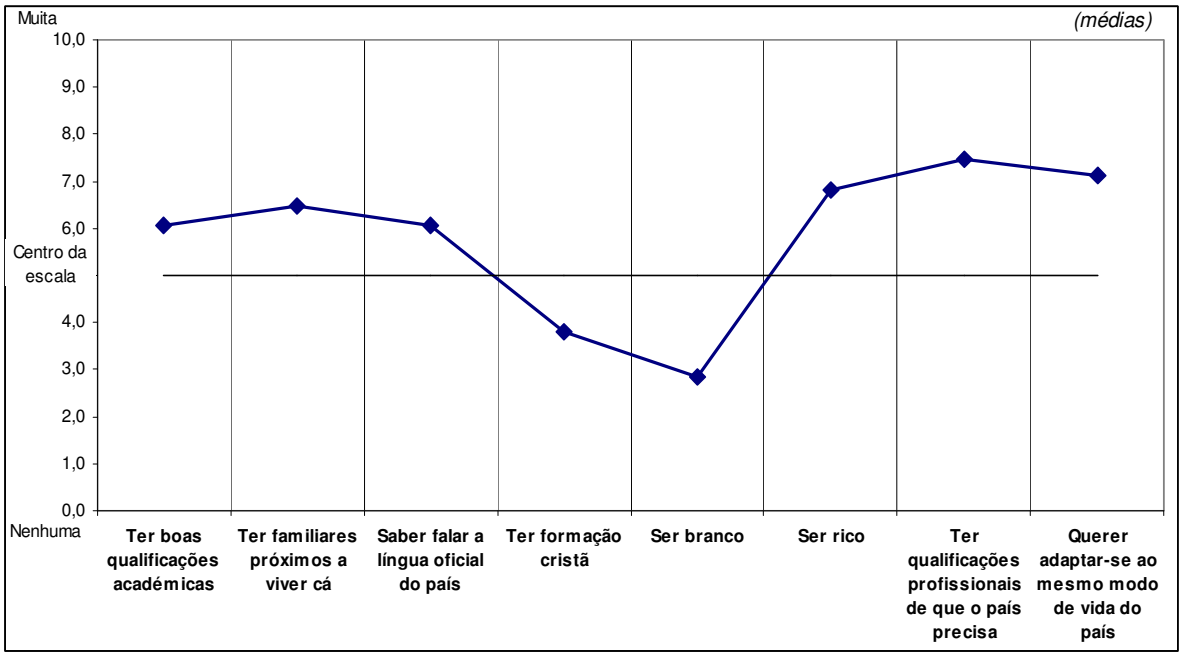


[5] A forma de fazer gráficos em SPSS tem variado nas várias versões do software. O gráfico aqui reproduzido foi feito na versão 15.0.

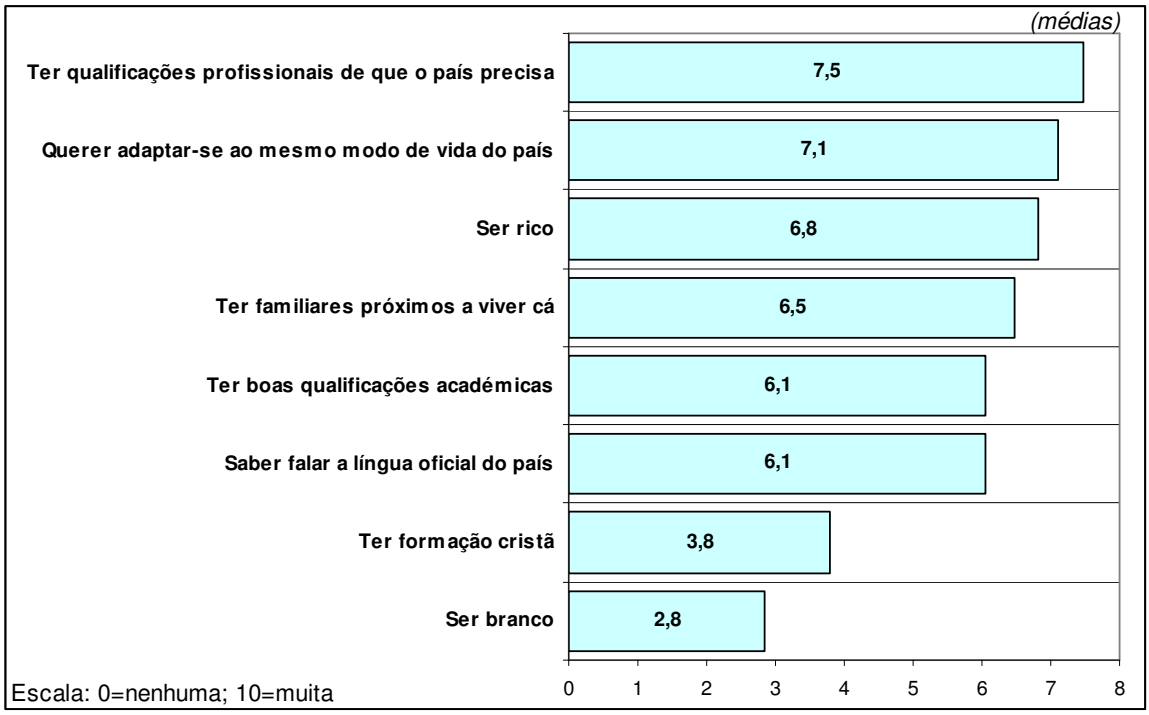
b) Feito em Excel

Alternativamente aos gráficos produzidos em SPSS, podem os mesmos serem produzidos em Excel, utilizando os recursos gráficos do software. Para o efeito, basta "copiar" o quadro de resultados no *output* através do comando "copy" e "colá-lo numa folha em Excel. Os gráficos seguintes foram feitos dessa forma:

Qual a importância dos seguintes aspectos para deixar vir para cá alguém que nasceu, foi educado e viveu fora de Portugal:



Qual a importância dos seguintes aspectos para deixar vir para cá alguém que nasceu, foi educado e viveu fora de Portugal:



1.3. Testes de aderência não paramétricos

Os testes de aderência não paramétricos permitem verificar se os dados amostrais se ajustam bem ou não a uma determinada distribuição teórica. Baseiam-se na Lei dos Grandes Números e consistem na comparação das frequências amostrais (observadas) com as frequências teóricas (esperadas).

1.3.1. χ^2 (Qui-quadrado)

É útil para verificar se a distribuição das frequências observadas dos dados se ajusta a um modelo teórico pré-determinado. Compara as frequências observadas na amostra e as frequências esperadas para a distribuição conhecida na população e testa se as diferenças são estatisticamente significativas ou não. Pode aplicar-se a todos os níveis de medida das variáveis, uma vez que admite como nível mínimo de medida, o nível nominal.

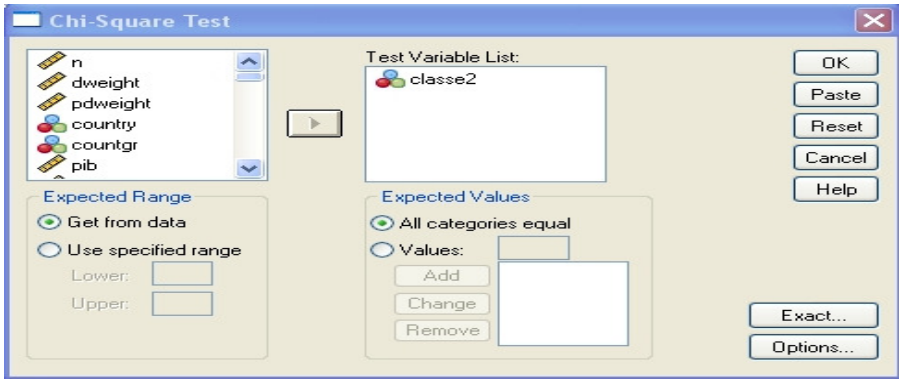
O SPSS testa, por defeito, se a distribuição é uniforme (*All categories equal*), mas pode-se especificar outra distribuição (*Values*)[6].

Exemplo: Pretende-se testar se a classe social (***classe2***) tem uma distribuição uniforme no universo.
Como a variável é nominal, o procedimento consiste em realizar o teste de aderência do χ^2[7]:

Hipóteses do teste (bilateral):

H_o: A classe social ***(classe2)*** tem uma distribuição uniforme na população

H_1: A classe social ***(classe2)*** não tem uma distribuição uniforme na população



O resultado é o seguinte:

Classes sociais (ACM) próprio

	Observed N	Expected N	Residual
Empresários, dirigentes e profissionais liberais	175	256.6	-81.6
Profissionais técnicos e de enquadramento	191	256.6	-65.6
Trabalhadores independentes	81	256.6	-175.6
Empregados executantes	413	256.6	156.4
Operários	423	256.6	166.4
Total	1283		

Test Statistics

	Classes sociais (ACM) próprio
Chi-Square [a]	366.123
df	4
Asymp. Sig.	.000

a. 0 cells (.0%) have expected frequencies less than 5. The minimum expected cell frequency is 256.6.

Interpretação: Rejeita-se a hipótese nula. A classe social não tem uma distribuição uniforme no universo ($\chi^2(4)=366{,}123$; $p=0{,}000$).

[6] Ver, por exemplo, o Anexo 1: Construção de um ponderador, no manual *Breve Introdução ao SPSS e Tratamento Preliminar de Dados*.
[7] Testa a probabilidade de as categorias da variável terem frequência igual.

1.3.2. *Kolmogorov-Smirnov*

"O teste de Kolmogorov-Srnirnov (K-S) é usado pata decidir se a distribuição da variável sob estudo (F (X)) numa determinada amostra provém de uma população com uma distribuição específica Fo (X). Neste caso porém estamos apenas interessados em testar se a distribuição da variável é ou não normal"[8].

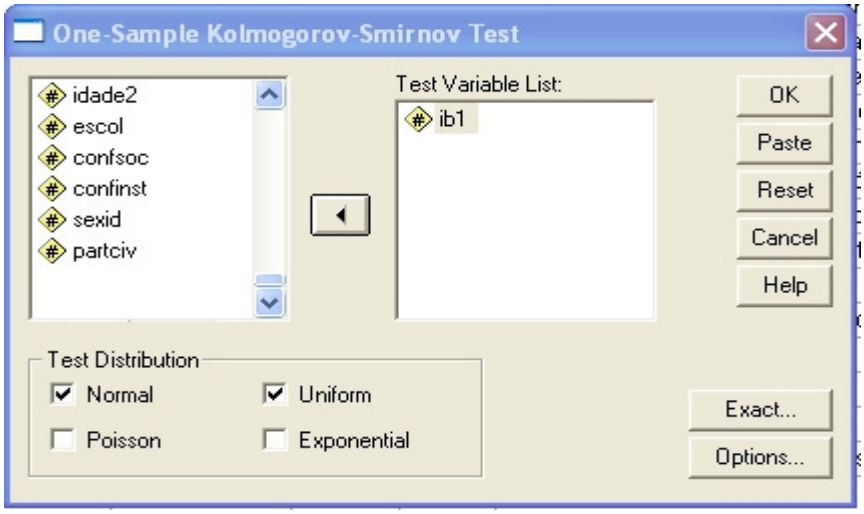
Exemplo: Pretende-se testar se o interesse pela política (***ib1***) segue uma distribuição normal ou uniforme no universo.

Como a variável é ordinal, o procedimento consiste em realizar o teste de aderência de *Kolmogorov-Smirnov* para as duas distribuições.

Hipóteses dos testes (bilaterais):

H_o: O interesse pela política (***ib1***) tem uma distribuição normal/uniforme na população

H_1: O interesse pela política (***ib1***) não tem uma distribuição normal/uniforme na população



O resultado é o seguinte:

One-Sample Kolmogorov-Smirnov Test

		Qual o seu interesse pela política
N		1503
Normal Parameters[a,b]	Mean	2.13
	Std. Deviation	.961
Most Extreme Differences	Absolute	.206
	Positive	.206
	Negative	-.198
Kolmogorov-Smirnov Z		8.001
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Normal.

b. Calculated from data.

One-Sample Kolmogorov-Smirnov Test 2

		Qual o seu interesse pela política
N		1503
Uniform Parameters[a,b]	Minimum	1
	Maximum	4
Most Extreme Differences	Absolute	.325
	Positive	.325
	Negative	-.078
Kolmogorov-Smirnov Z		12.613
Asymp. Sig. (2-tailed)		.000

a. Test distribution is Uniform.

b. Calculated from data.

Interpretação: Rejeita-se a hipótese nula nos dois casos. A variável interesse pela política não segue uma distribuição normal (*K-S*=8,001; *p*=0,000), nem uniforme (*K-S*=12,613; *p*=0,000) na população.

[8] Maroco, J (2007), *Análise Estatística - Com Utilização do SPSS*, Lisboa, Sílabo: 134.

1.3.3. Teste de aderência paramétrico: *t-Student* para uma amostra

O teste *t-Student* para uma amostra é um teste paramétrico que testa se "a média populacional é ou não igual a um determinado valor"[9]. Admite como nível de medida mínimo o nível intervalar (quantitativo).

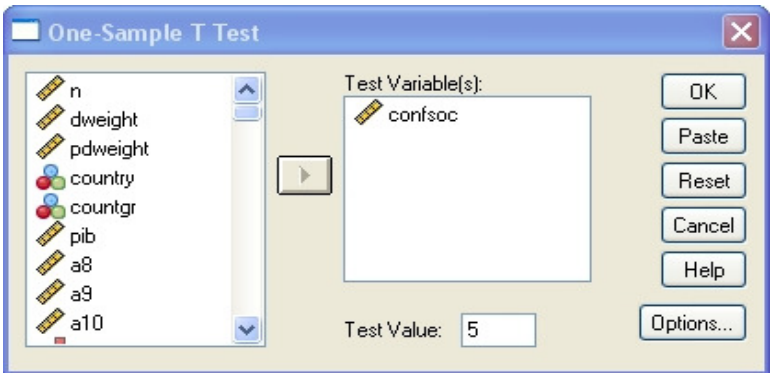
Exemplo: Pretende-se testar se a média populacional da confiança social difere da média que corresponde ao centro da escala (5)[10] dos indicadores que compõem a variável (***confsoc***).

Como a variável é quantitativa, o procedimento consiste em realizar o teste *t* de aderência (*One-Sample T-Test*), comparando se a média observada difere significativamente de 5.

Hipóteses do teste (bilateral):

H_o: A confiança social tem média igual a 5 na população

H_1: A confiança social tem média diferente de 5 na população



O resultado é o seguinte:

One-Sample Statistics

	N	Mean	Std. Deviation	Std. Error Mean
Indice sintético de Confiança social	1480	4.316	1.7469	.0454

One-Sample Test

	Test Value = 5					
					95% Confidence Interval of the Difference	
	t	df	Sig. (2-tailed)	Mean Difference	Lower	Upper
Indice sintético de Confiança social	-15.073	1479	.000	-.684	-.774	-.595

Interpretação: Rejeita-se a hipótese nula. A média observada, 4,3, difere significativamente da média de referência (5) ($t(1479)= -15{,}073$; $p=0{,}000$).

[9] Maroco, idem: 144.
[10] A escala de medida dos indicadores de confiança social varia entre 0=nenhuma e 10=toda.

Análise de Questões de Resposta Múltipla

2. Análise de questões de resposta múltipla

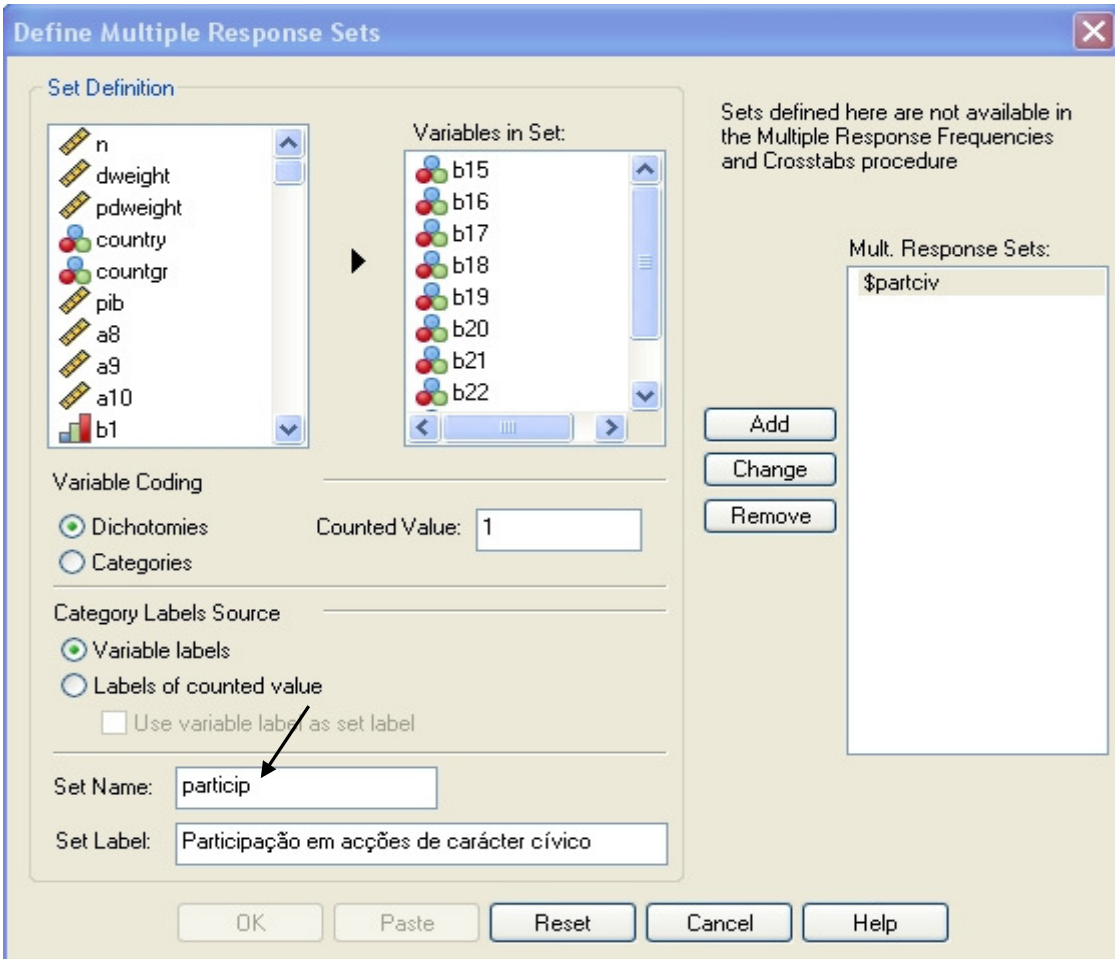
As perguntas de **Escolha Múltipla** são muito utilizadas em inquéritos por questionário. Caracterizam-se pela resposta a um conjunto de items, podendo consistir na simples "marcação" do item ou, quando as respostas são dicotómicas, na selecção de uma das duas opções (e.g. sim/não, certo/errado, concordo/discordo, gosto/não gosto, etc.).

2.1. Criar uma variável agregada com a resposta simultânea aos diversos items

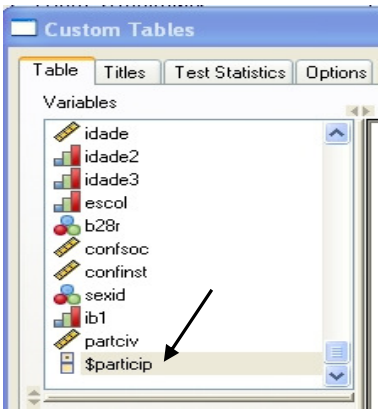
Exemplo: Os items ***b15*** a ***b24*** do questionário *European Social Survey* são o exemplo de uma questão de resposta múltipla:

> *Há várias acções que se podem desenvolver para melhorar as coisas ou para evitar que corram mal. Durante os últimos 12 meses, fez alguma das seguintes coisas?*

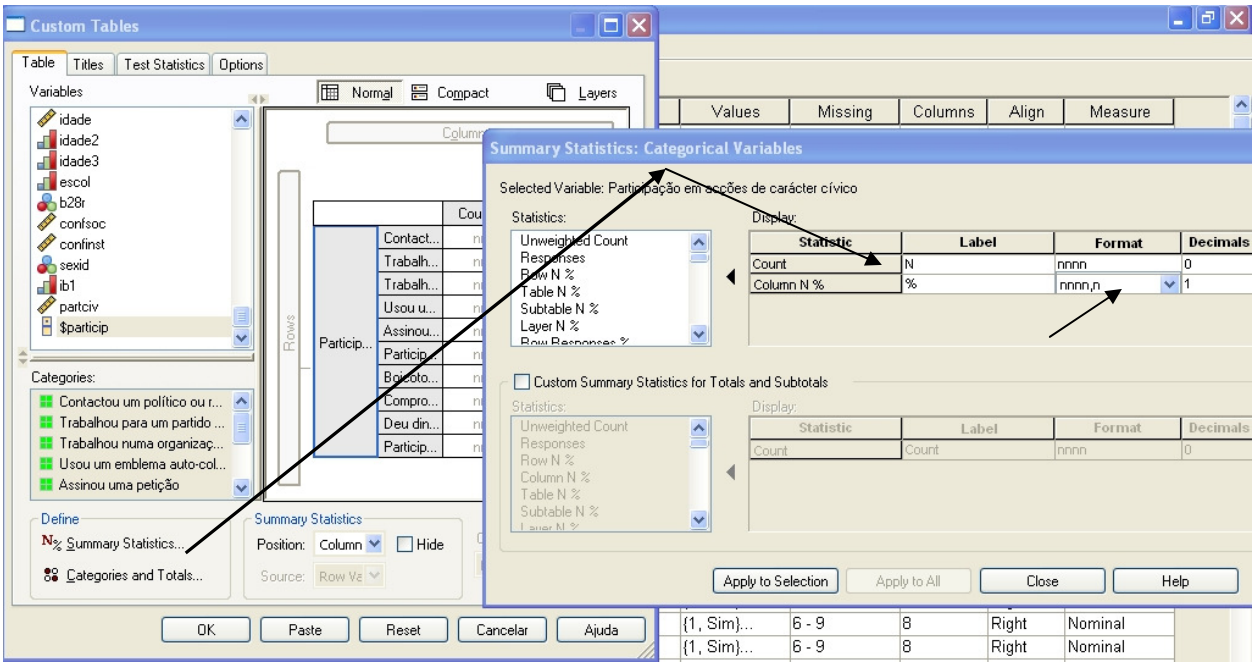
Em primeiro lugar, vamos criar a variável que traduz a participação cívica dos inquiridos - "***partcip***" - que agrega todas as respostas ***sim***:



A variável "***partcip***" é uma variável virtual que, embora não faça parte da base de dados, está visível no procedimento *Tables* e podemos, por conseguinte, solicitar as respectivas frequências e cruzamentos



2.2. Frequências de uma questão de resposta múltipla

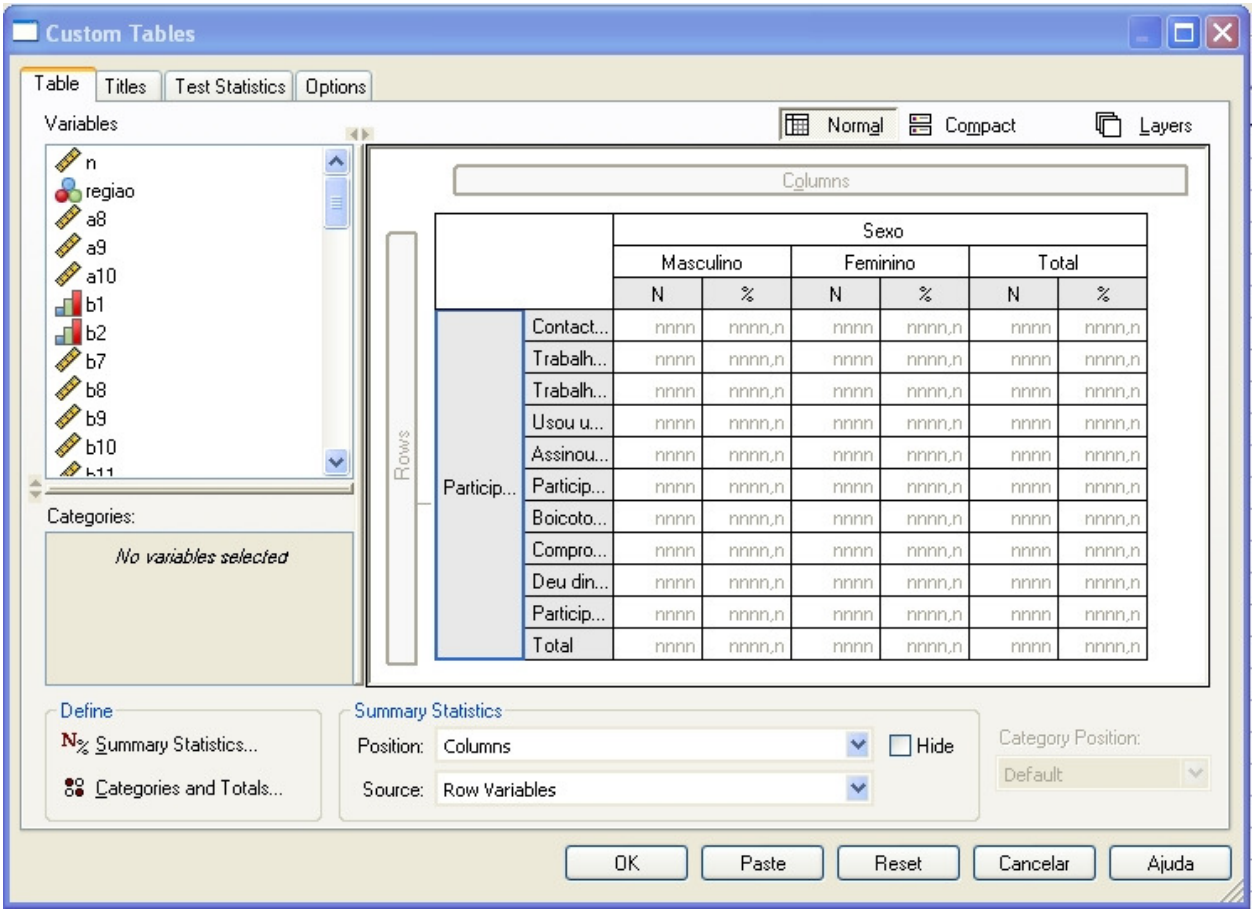


O resultado é o seguinte:

Participação em acções de carácter cívico	N	%
Contactou um político ou representante do governo/poder local	168	49,9
Comprou produtos por razões de ordem política, ética ou ambiental	112	33,2
Assinou uma petição	101	30,0
Usou um emblema auto-colante de campanha/movimento	94	27,9
Trabalhou numa organização ou associação de outro tipo	79	23,4
Participou numa manifestação	63	18,7
Trabalhou para um partido político/movimento cívico	58	17,2
Deu dinheiro a um grupo ou organização política	54	16,0
Boicotou determinados produtos	48	14,2
Participou em acções de protesto ilegais	5	1,5
Nº de inquiridos que responderam	**337**	**100,0**

Nota: apenas 337 inquiridos responderam "sim" a, pelo menos, um item. Os restantes, ou não responderam ou responderam "não".

2.3. Cruzamento de uma questão de resposta múltipla



O resultado é o seguinte:

Participação em acções de carácter cívico	Sexo					
	Masculino		Feminino		**Total**	
	N	%	N	%	N	%
Contactou um político ou representante do governo/poder local	94	57,7	74	42,5	**168**	**49,9**
Comprou produtos por razões de ordem política, ética ou ambiental	45	27,6	67	38,5	**112**	**33,2**
Assinou uma petição	49	30,1	52	29,9	**101**	**30,0**
Usou um emblema auto-colante de campanha/movimento	43	26,4	51	29,3	**94**	**27,9**
Trabalhou numa organização ou associação de outro tipo	48	29,4	31	17,8	**79**	**23,4**
Participou numa manifestação	33	20,2	30	17,2	**63**	**18,7**
Trabalhou para um partido político/movimento cívico	43	26,4	15	8,6	**58**	**17,2**
Deu dinheiro a um grupo ou organização política	36	22,1	18	10,3	**54**	**16,0**
Boicotou determinados produtos	22	13,5	26	14,9	**48**	**14,2**
Participou em acções de protesto ilegais	4	2,5	1	,6	**5**	**1,5**
Total	**163**	**100,0**	**174**	**100,0**	**337**	**100,0**

Análise de Dados Bivariada

3. Análise de dados bivariada

Nota: A base de dados que vamos utilizar neste capítulo é um extracto da base de dados do *European Social Survey* (*round* 1) com os resultados da aplicação do questionário em Portugal, modificada com os exercícios do capítulo 1 (ficheiro **ESS Portugal 2002 (TPAUB2).sav**)

3.1. Cruzamentos e teste de independência χ^2 (Qui-quadrado)

O teste do χ^2 de independência "serve para testar se duas ou mais populações (ou grupos) independentes diferem relativamente a uma determinada característica, i.e. se a frequência com que os elementos da amostra se repartem pelas classes de uma variável nominal categorizada é ou não idêntica"[11]. Admite como nível mínimo de mediada o nível nominal.

Exemplo: Pretende-se saber se há relação entre o sexo e o facto de ter comprado produtos por razões de ordem política, ética ou ambiental.

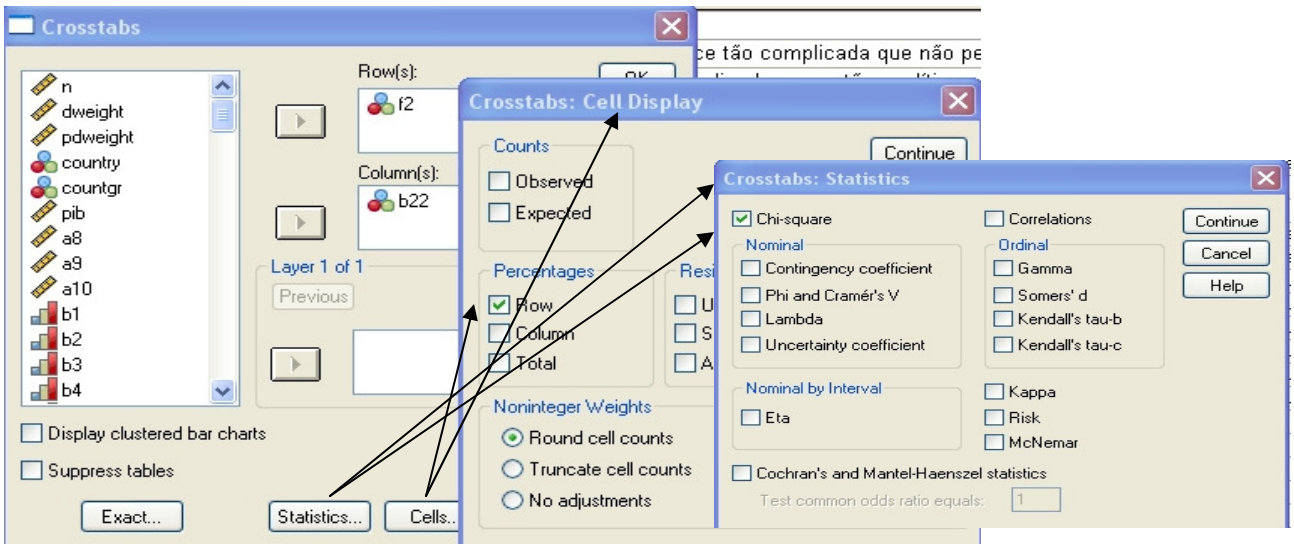
O procedimento consiste em cruzar as variáveis sexo (***f2***) e (***b22***) e solicitar o teste de independência do χ^2*.

Hipótese do teste (bilateral):

H_o: A compra de produtos por razões de ordem política, ética ou ambiental, é igual entre homens e mulheres

H_1: A compra de produtos por razões de ordem política, ética ou ambiental, é diferente entre homens e mulheres

3.1.1. Utilizando o comando *Crosstabs*



[11] Maroco, idem: 103.

* Para a selecção dos testes estatísticos, ver o Anexo 1.

O resultado é o seguinte:

Sexo * Comprou produtos por razões de ordem política, ética ou ambiental Crosstabulation

			Sim	Não	Total
Sexo	Masculino	Count	45	579	624
		Row %	7,2%	92,8%	100,0%
	Feminino	Count	67	804	871
		Row %	7,7%	92,3%	100,0%
Total		Count	112	1383	1495
		Row %	7,5%	92,5%	100,0%

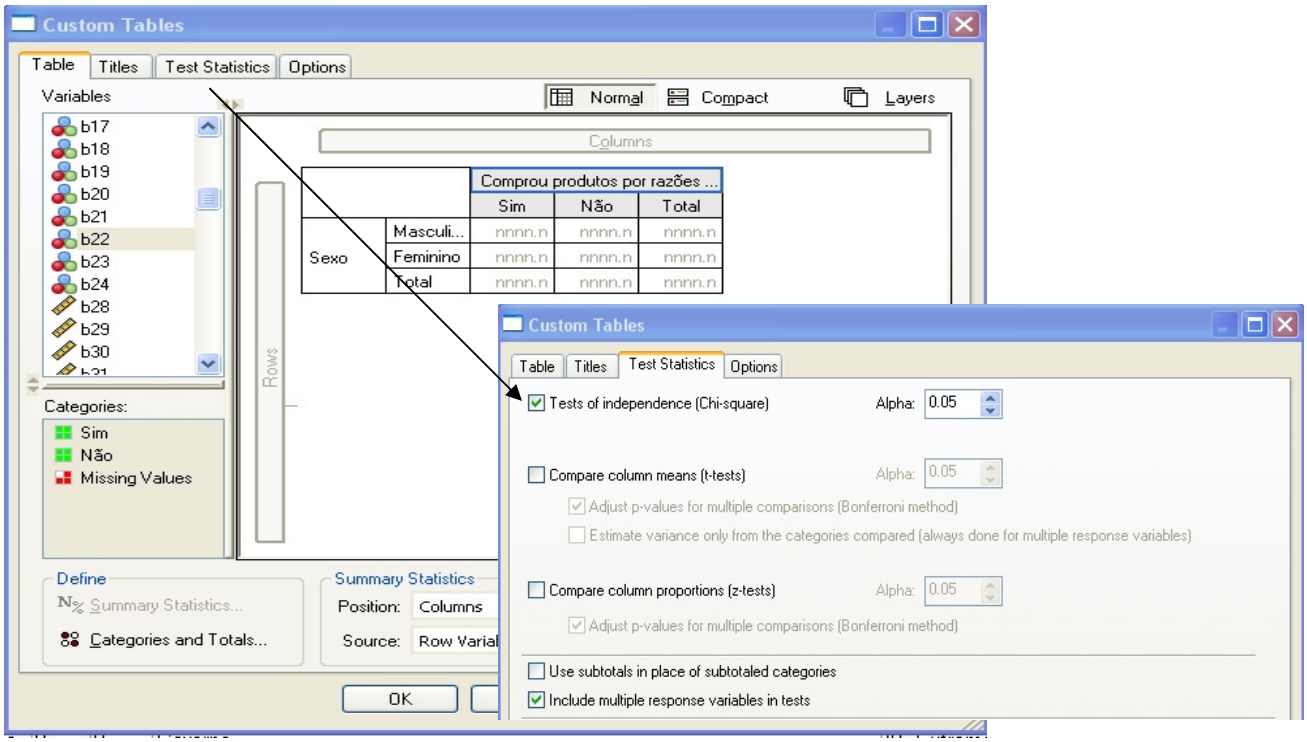
Chi-Square Tests

	Value	df	Asymp. Sig. (2-sided)	Exact Sig. (2-sided)	Exact Sig. (1-sided)
Pearson Chi-Square	,121[b]	1	,728		
Continuity Correction[a]	,062	1	,804		
Likelihood Ratio	,122	1	,727		
Fisher's Exact Test				,766	,404
Linear-by-Linear Association	,121	1	,728		
N of Valid Cases	1495				

a. Computed only for a 2x2 table

b. 0 cells (,0%) have expected count less than 5. The minimum expect count is 46,75.

3.1.2. Utilizando o comando *Tables*



O resultado é o seguinte:

Sexo	Comprou produtos por razões de ordem política, ética ou ambiental			
	Sim		Não	
	N	%	N	%
Masculino	45	7,2	579	92,8
Feminino	67	7,7	804	92,3
Total	**112**	**7,5**	**1383**	**92,5**

Pearson Chi-Square Tests

		Comprou produtos por razões de ordem política, ética ou ambiental
Sexo	Chi-square	,121
	df	1
	Sig.	,728

Results are based on nonempty rows and columns in each innermost subtable.

Interpretação: Embora as mulheres comprem ligeiramente mais que os homens, rejeita-se a hipótese nula (χ^2 (1)=0,121; p>0,05). Ou seja, as diferenças entre homens e mulheres não são estatisticamente significativas.

3.2. Testes não paramétricos (procedimento *Nonparametric Tests*)

Os testes não paramétricos não estão condicionados por qualquer distribuição de probabilidades dos dados em análise, ou seja, não estão sujeitos aos condicionalismos da verificação dos pressupostos, como acontece nos testes paramétricos, e constituem uma alternativa à sua utilização, quando aqueles não se verificam e a sua violação é "grave".

3.2.1. Duas amostras independentes: Teste de *Mann-Whitney*

"O teste de Wilcoxon-Mann-Whitney ou simplesmente teste de Mann-Whitney, é o teste não-paramétrico adequado para comparar as funções de distribuição de uma variável pelo menos ordinal medida em duas amostras independentes"[12].

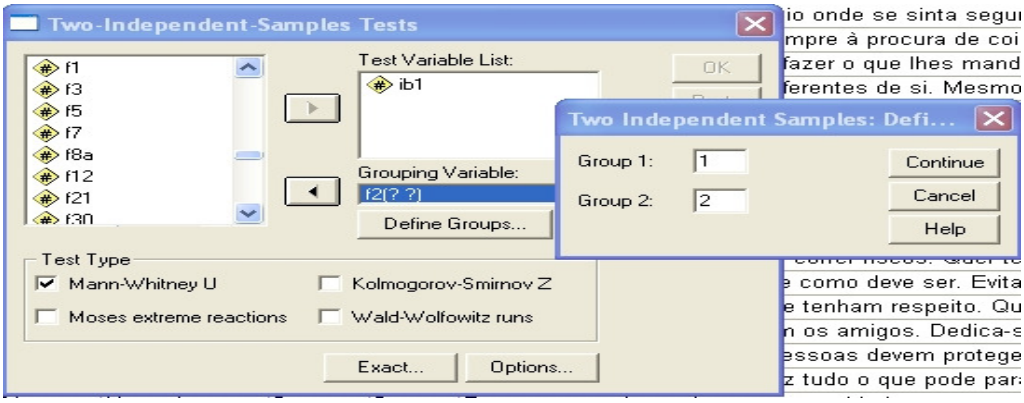
Exemplo: Pretende-se testar se há relação entre o sexo (***f2***) e o interesse pela política (***ib1***).

O procedimento consiste na realização do teste não paramétrico para 2 amostras independentes (***Mann-Whitney***)[13].

Hipótese do teste (bilateral):

H_o: Homens e mulheres têm o mesmo interesse pela política

H_1: Homens e mulheres não têm o mesmo interesse pela política



O resultado é o seguinte:

Ranks

	Sexo	N	Mean Rank	Sum of Ranks
Qual o seu interesse pela política	Masculino	628	815.82	512336.00
	Feminino	875	706.19	617920.00
	Total	1503		

Test Statistics [a]

	Qual o seu interesse pela política
Mann-Whitney U	234670.000
Wilcoxon W	617920.000
Z	-5.058
Asymp. Sig. (2-tailed)	.000

a. Grouping Variable: Sexo

Interpretação: Rejeita-se a hipótese nula. Os homens têm mais interesse pela política do que as mulheres. A média das ordenações (*Mean Rank*) é superior nos homens[14] e as diferenças são estatisticamente significativas (*M-W*=234670; *p*=0,000).

[12] Maroco, idem: 219. Este teste pode também ser utilizado como alternativa ao teste *t-Student* para amostras independentes, nomeadamente quando os pressupostos deste teste não são válidos.
[13] Consultar o Anexo 1.
[14] A escala é crescente.

3.2.2. K amostras independentes (*Kruskal-Wallis*)

"O teste de Kruskal-Walis pode ser considerado como a alternativa não-paramétrica à ANOVA *one-way* (Kruskal & Wallis, 1952). Este teste pode ser então usado para testar se duas ou mais amostras provém de uma mesma população ou se de populações diferentes ou se, de igual modo, as amostras provém de populações com a mesma distribuição"[15].

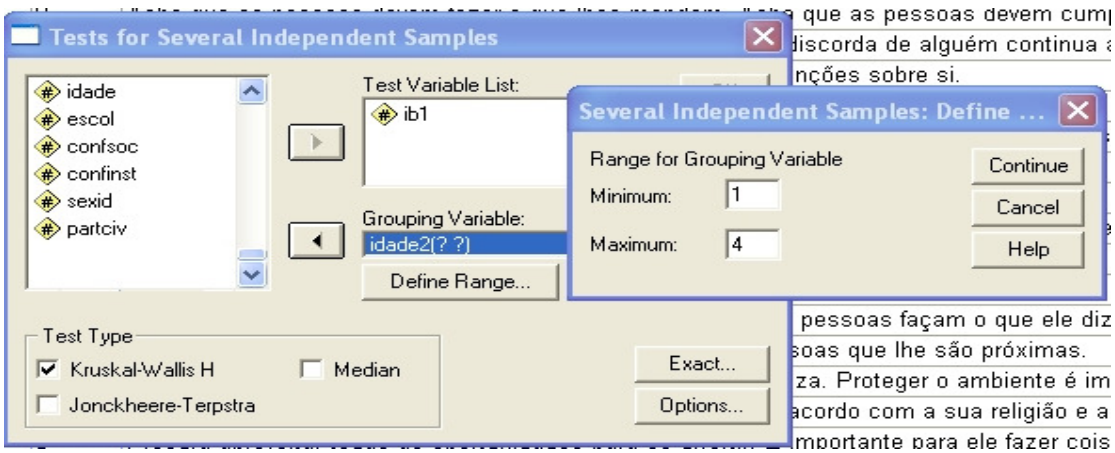
Exemplo: Pretende-se testar se há relação entre a idade (***idade2***) e o interesse pela política (***ib1***).

O procedimento consiste na realização do teste não paramétrico para k amostras independentes (***Kruskal-Wallis***)[16].

Hipótese do teste (bilateral):

H_o : O interesse pela política tem igual distribuição nos diversos escalões etários

H_1: O interesse pela política não tem igual distribuição nos diversos escalões etários



O resultado é o seguinte:

		Qual o seu interesse pela política				
		Nenhum interesse	Pouco interesse	Algum interesse	Muito interesse	Total
Idade	Até 30 anos	27.4	33.6	31.9	7.1	100.0
	31 - 50 anos	28.0	28.8	32.9	10.3	100.0
	51 - 65 anos	30.7	30.4	29.7	9.3	100.0
	> 65 anos	45.8	25.1	25.6	3.5	100.0
	Total	32.5	29.3	30.3	7.8	100.0

Ranks

	Idade	N	Mean Rank
Qual o seu interesse pela política	Até 30 anos	339	777.35
	31 - 50 anos	504	803.46
	51 - 65 anos	313	768.74
	> 65 anos	347	637.38
	Total	1503	

Test Statistics [a,b]

	Qual o seu interesse pela política
Chi-Square	36.088
df	3
Asymp. Sig.	.000

a. Kruskal Wallis Test

b. Grouping Variable: Idade

Interpretação: Rejeita-se a hipótese nula. Os indivíduos entre 31 a 50 anos são os que revela maior interesse pela política (*mean rank*=803,046) e os mais velhos são os que revelam menos interesse (*mean rank*=637,38). As diferenças são estatisticamente significativas (*K-W* (3)=36,088; *p*=0,000).

[15] Maroco, idem: 227.
[16] Consultar o Anexo 1.

3.3. Testes paramétricos

Os Testes paramétricos são baseados em parâmetros da amostra, por exemplo média e desvio padrão, assumindo a normalidade e a homogeneidade das distribuições.

3.3.1. Duas amostras emparelhadas (*t Student* para amostras emparelhadas)

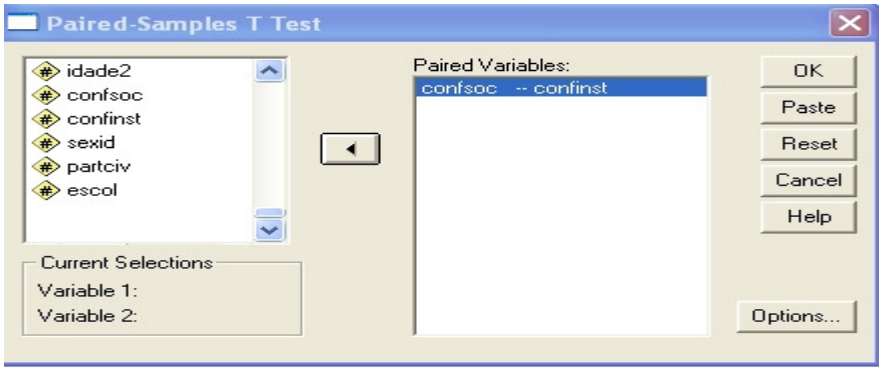
"O teste *t-Student* para amostras emparelhadas é o teste a utilizar quando se pretende comparar duas populações de onde foram extraídas duas amostras emparelhadas relativamente a uma variável dependente quantitativa"[17].

Exemplo: Pretende-se testar se a média da confiança social (***confsoc***) é idêntica, ou não, à média da confiança institucional (***confinst***).
O procedimento consiste na realização do teste paramétrico para duas amostras emparelhadas (***Paired-Samples T-Test***)[18].

Hipótese do teste (bilateral):

H_o: A média da confiança social é igual à média da confiança institucional

H_1: A média da confiança social não é igual à média da confiança institucional



O resultado é o seguinte:

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Índice sintético de Confiança social	4,313	1319	1,7165	,0473
	Índice sintético de Confiança institucional	4,087	1319	1,7606	,0485

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Índice sintético de Confiança social & Índice sintético de Confiança institucional	1319	,310	,000

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Índice sintético de Confiança social - Índice sintético de Confiança institucional	,2255	2,0428	,0562	,1151	,3358	4,009	1318	,000

Interpretação: Rejeita-se a hipótese nula (t (1318)=4,009; p=0,000). (A média da confiança social (4,313) é ligeiramente superior à média da confiança institucional (4,087). A correlação entre as duas variáveis é média fraca (0,310) e estatisticamente significativa (p=0,000).

[17] Maroco, idem: 270. Ao contrário do teste *t-Student* para amostras independentes, no caso de apenas duas amostras emparelhadas, não é necessário assegurar a homocedasticidade das variâncias.
[18] Consultar o Anexo 1.

3.3.2. Duas amostras independentes (*t Student* de independência)

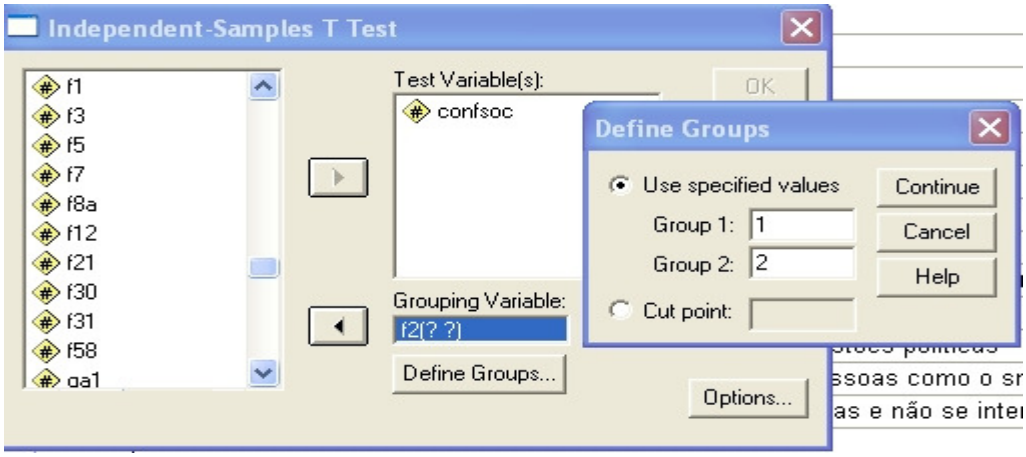
"O teste *t-Student* serve também para testar se as médias de duas populações são ou não significativamente diferentes. Este teste requer que as duas amostras tenham sido obtidas aleatoriamente de duas populações e que as variáveis dependentes possuam distribuição normal com variâncias homogéneas"[19].

Exemplo: Pretende-se testar se há relação entre o sexo (***f2***) e a confiança social (***confsoc***).
O procedimento consiste na realização do teste paramétrico para duas amostras independentes (***Independent-Samples T-Test***)[20].

Hipótese do teste (bilateral):

H_o: A média da confiança social é igual entre homens e mulheres

H_1: A média da confiança social é diferente entre homens e mulheres



O resultado é o seguinte:

Group Statistics

	Sexo	N	Mean	Std. Deviation	Std. Error Mean
Indice sintético de Confiança social	Masculino	612	4.412	1.7476	.0706
	Feminino	868	4.248	1.7442	.0592

Independent Samples Test

		Levene's Test for Equality of Variances		t-test for Equality of Means						
									95% Confidence Interval of the Difference	
		F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Indice sintético de Confiança social	Equal variances assumed	.182	.669	1.781	1478	.075	.164	.0921	-.0167	.3448
	Equal variances not assumed			1.780	1313.926	.075	.164	.0922	-.0167	.3449

Interpretação: Não se rejeita a hipótese nula. Os homens (4,412) revelam mais confiança social que as mulheres (4,248)[21], mas a diferença não é estatisticamente significativa ($t\,(1478)=1{,}781$; $p>0{,}05$).

[19] Maroco, idem: 147-148.
[20] Consultar o Anexo 1.
[21] O índice de confiança social varia entre 0=nenhuma confiança e 10=toda a confiança.

3.3.3. k amostras independentes (Análise de Variância simples paramétrica - ANOVA)

"A comparação de médias de duas ou mais populações de onde foram extraídas amostras aleatórias e independentes pode fazer-se através de uma metodologia proposta por Sir Ronald Fisher e genericamente designada por Análise de Variância (abreviadamence ANOVA do inglês *Analysis of Variance*) (Fisher, 1935) se a distribuição da variável em estudo for Normal e se as variâncias populacionais forem homogéneas"[22].

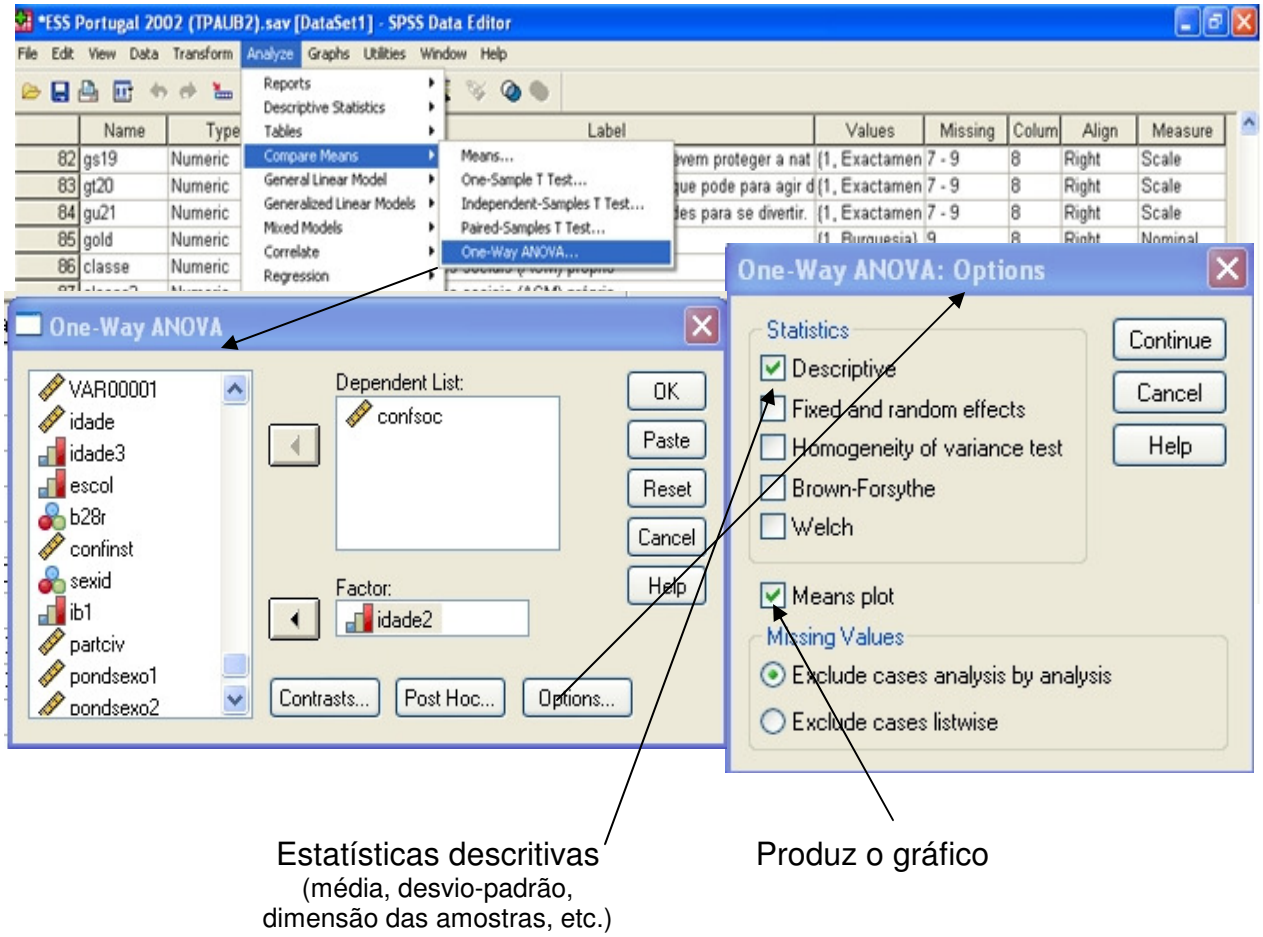
Exemplo: Pretende-se testar se há relação entre a idade (***idade2***) e a confiança social (***confsoc***).

O procedimento consiste na realização da Análise de Variância Simples Paramétrica (***One-way Anova***)[23].

Hipótese do teste (bilateral):

H_o: A média da confiança social é igual em todos os escalões etários

H_1: A média da confiança social é diferente em pelo menos um escalão etário



[22] Maroco, idem: 154.
[23] Consultar o Anexo 1.

O resultado é o seguinte:

Descriptives

Índice sintético de Confiança social

	N	Mean	Std. Deviation	Std. Error	95% Confidence Interval for Mean		Minimum	Maximum
					Lower Bound	Upper Bound		
Até 30 anos	338	4.722	1.5585	.0848	4.555	4.889	.0	9.3
31 - 50 anos	497	4.058	1.6747	.0751	3.911	4.206	.0	9.3
51 - 65 anos	312	4.068	1.8340	.1038	3.864	4.273	.0	10.0
> 65 anos	333	4.519	1.8500	.1014	4.319	4.718	.0	10.0
Total	1480	4.316	1.7469	.0454	4.226	4.405	.0	10.0

Nota: como as amostras são superiores a 30 (N), não se torna necessário testar a "normalidade"

Test of Homogeneity of Variances

Índice sintético de Confiança social

Levene Statistic	df1	df2	Sig.
4.050	3	1476	.007

Hipótese para a "Homogeneidade de variâncias":

H_o: as variâncias são homogéneas

H_1:as variâncias não são homogéneas

Tomada de decisão: Rejeita-se a hipótese nula, não se verifica o pressuposto.

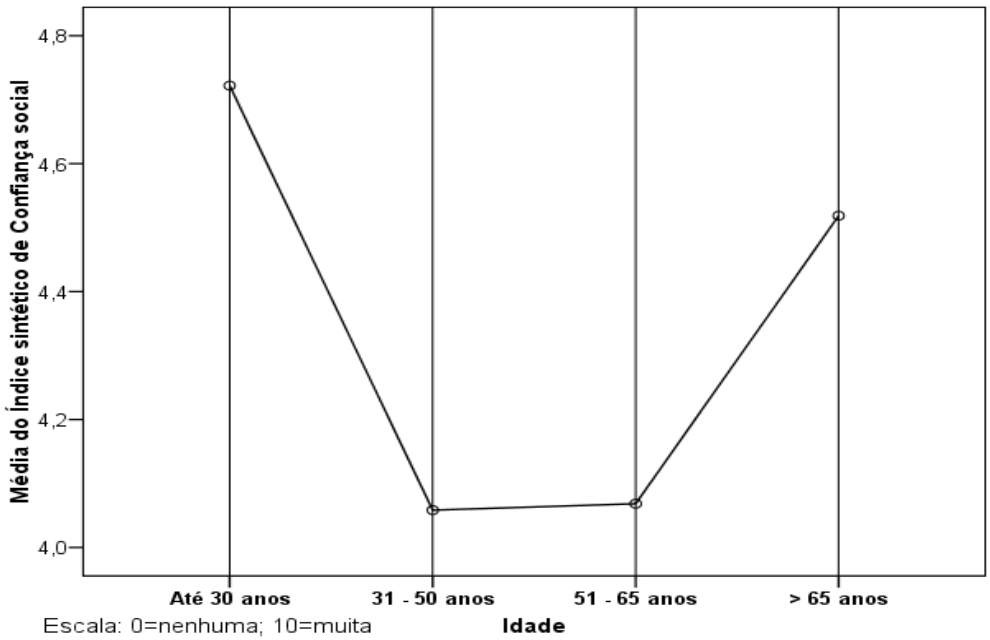
No entanto, a "violação deste pressuposto não tem consequências sérias se os grupos ou amostras tiverem dimensões razoáveis"[24]. Na prática, a Anova é robusta à violação do pressuposto se a maior das amostras não for superior ao dobro da menor.

ANOVA

Índice sintético de Confiança social

	Sum of Squares	df	Mean Square	F	Sig.
Between Groups	121.466	3	40.489	13.607	.000
Within Groups	4391.954	1476	2.976		
Total	4513.420	1479			

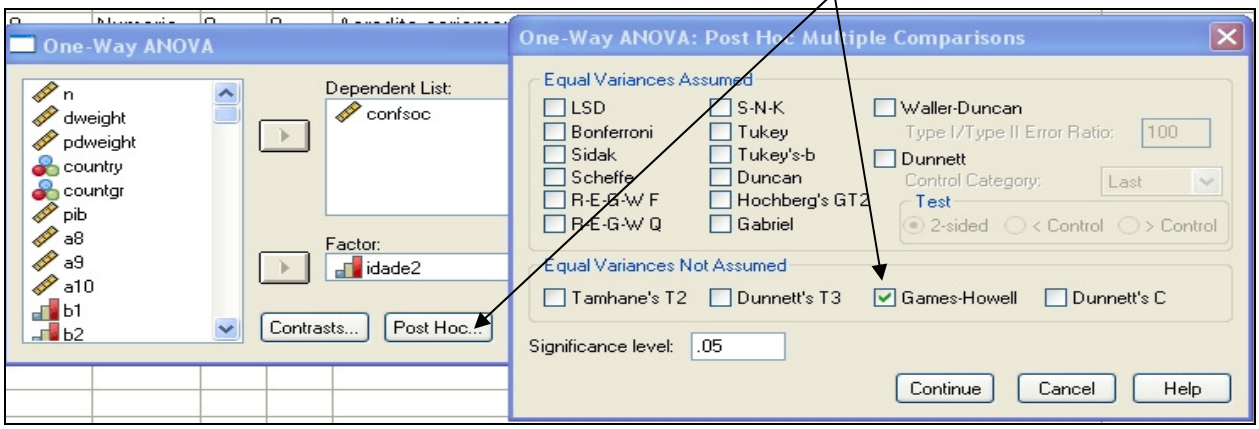
Tomada de decisão:
Rejeita-se a hipótese nula, há pelo menos um escalão etário em que a média da confiança social é significativamente diferente.



Nota: Tendo-se rejeitado a hipótese nula, conclui-se que há pelo menos um escalão etário onde a confiança social apresenta valores médios significativamente diferentes dos restantes escalões. Nesta situação importa, por conseguinte, saber quais são os escalões que diferem uns dos outros. Para o efeito realiza-se um teste de comparações múltiplas à posteriori (***Post Hoc***).

[24] Murteira, B. (1990), Probabilidades e Estatística, Lisboa McGraw-Hill, vol.II: :349.

O SPSS disponibiliza vários testes para este fim, sendo os mais utilizados, o teste de *Scheffe*[25], no caso de as variâncias serem iguais, e o teste *Games-Howell* no caso de serem diferentes. Neste caso, uma vez que se rejeita a hipótese de as variâncias serem iguais (*p*=0,007), vamos solicitar o teste *Games-Howell*. Par o efeito, repete-se o procedimento e selecciona-se o teste no separador "PostHoc":



O resultado é o seguinte:

Multiple Comparisons

Dependent Variable: Índice sintético de Confiança social
Games-Howell

(I) Idade	(J) Idade	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	95% Confidence Interval Upper Bound
Até 30 anos	Até 30 anos					
	31 - 50 anos	,6635*	,1133	,000	,372	,955
	51 - 65 anos	,6535*	,1340	,000	,308	,999
	> 65 anos	,2034	,1321	,415	-,137	,544
31 - 50 anos	Até 30 anos	-,6635*	,1133	,000	-,955	-,372
	31 - 50 anos					
	51 - 65 anos	-,0100	,1282	1,000	-,340	,320
	> 65 anos	-,4602*	,1262	,002	-,785	-,135
51 - 65 anos	Até 30 anos	-,6535*	,1340	,000	-,999	-,308
	31 - 50 anos	,0100	,1282	1,000	-,320	,340
	51 - 65 anos					
	> 65 anos	-,4501*	,1451	,011	-,824	-,076
> 65 anos	Até 30 anos	-,2034	,1321	,415	-,544	,137
	31 - 50 anos	,4602*	,1262	,002	,135	,785
	51 - 65 anos	,4501*	,1451	,011	,076	,824
	> 65 anos					

*. The mean difference is significant at the .05 level.

Estamos agora em condições de interpretar o resultado da Análise de Variância.

Interpretação: Os mais novos (4,722), seguidos dos mais velhos (4,519) são os que mais confiam. Os escalões intermédios 31-50 anos (4,058) e 51-65 anos (4,068) confiam um pouco menos. As diferenças são estatisticamente significativas (*F* (3)=13,787; *p*=0,000). O quadro seguinte sintetiza as diferenças de médias significativas entre os quatro escalões etários:

Confiança Social: diferenças de médias entre escalões etários

	Até 30 anos	31-50 anos	51-65 anos
Até 30 anos			
31-50 anos	0,6635**		
51-65 anos	0,6535**	*n.s.*	
> 65 anos	*n.s.*	0,4602*	0,4501*

* *p< 0,05;* ** *p<0,001;*

[25] Que é também o mais conservador,

4. Correlação

Em estatística, a *correlação*, também chamada de *coeficiente de correlação*, indica a força e a direcção do relacionamento entre duas variáveis. Refere-se à medida da relação entre duas variáveis, embora a correlação não implique causalidade. Neste sentido geral, existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados.

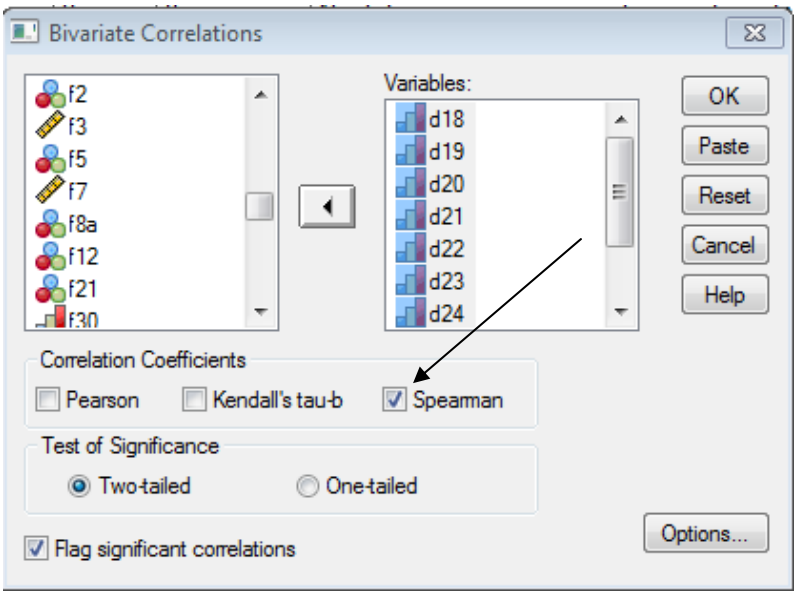
O coeficiente de correlação varia entre –1 e 1[26] e deve ser interpretado da seguinte forma:

0:	ausência de correlação
+/-]0 – 0,25]:	muito fraca
+/-]0,25 – 0,40]:	fraca
+/-]0,40 – 0,60]:	moderada
+/-]0,60 – 0,75]:	moderada forte
+/-]0,75 – 0,90]:	forte
+/-]0,90 – 1[:	muito forte
+/- 1:	correlação perfeita

4.1. Correlação não linear de Spearman (*rho* de Spearman)

"O coeficiente de correlação de Spearman mede a intensidade da relação entre variáveis ordinais. Utiliza os valores de ordem das observações em vez do seu valor observado. Deste modo, este coeficiente não é sensível a assimetrias na distribuição, nem à presença de outliers, não exigindo que os dados provenham de duas populações normais. Aplica-se igualmente em variáveis intervalo/rácio como alternativa ao R de Pearson, quando neste último se viola a normalidade"[27].

Exemplo: Correlação entre as variáveis atitudinais sobre os imigrantes (***d18*** a ***d24***). Trata-se de variáveis ordinais pelo que o coeficiente de correlação mais adequado a este tipo de variáveis é o *rho de Spearman*:



[26] O sinal – significa uma relação negativa e a ausência de sinal uma relação positiva.
[27] Pestana, M. H. e J. N. Gageiro (2000), *Análise de dados para Ciências Sociais – A Complementaridade do SPSS*, Lisboa, Sílabo. 2ª edição.

O resultado é o seguinte[28]:

Correlações (*Spearman's rho*)

		d18	d19	d20	d21	d22	d23
d18 As pessoas que vêm viver e trabalhar para cá fazem com que os salários baixem	Rho						
	Sig.	.					
	N	1406					
d19 As pessoas que vêm viver e trabalhar para cá, em regra, prejudicam mais as expectativas económicas dos pobres do que dos ricos	Rho	,569(**)					
	Sig.	,000	.				
	N	1366	1420				
d20 As pessoas que vêm viver e trabalhar para cá ajudam a preencher lugares em que há falta de trabalhadores	Rho	-,188(**)	-,204(**)				
	Sig.	,000	,000	.			
	N	1378	1391	1453			
d21 Se as pessoas que vieram viver e trabalhar para cá estiverem desempregadas por muito tempo deviam ser obrigadas a ir embora	Rho	,250(**)	,327(**)	-,134(**)			
	Sig.	,000	,000	,000	.		
	N	1353	1371	1391	1422		
d22 As pessoas que vieram viver para cá devem ter os mesmos direitos do que todas as outras pessoas	Rho	-,223(**)	-,187(**)	,318(**)	-,111(**)		
	Sig.	,000	,000	,000	,000	.	
	N	1382	1398	1428	1405	1462	
d23 As pessoas que vieram viver para cá cometerem um crime grave, devem ser obrigadas a ir embora	Rho	,226(**)	,271(**)	-,051	,307(**)	,037	
	Sig.	,000	,000	,057	,000	,161	.
	N	1378	1391	1418	1396	1429	1455
d24 As pessoas que vieram viver para cá cometerem qualquer crime, devem ser obrigadas a ir embora	Rho	,250(**)	,219(**)	-,110(**)	,341(**)	-,073(**)	,594(**)
	Sig.	,000	,000	,000	,000	,006	,000
	N	1376	1388	1413	1392	1425	1444

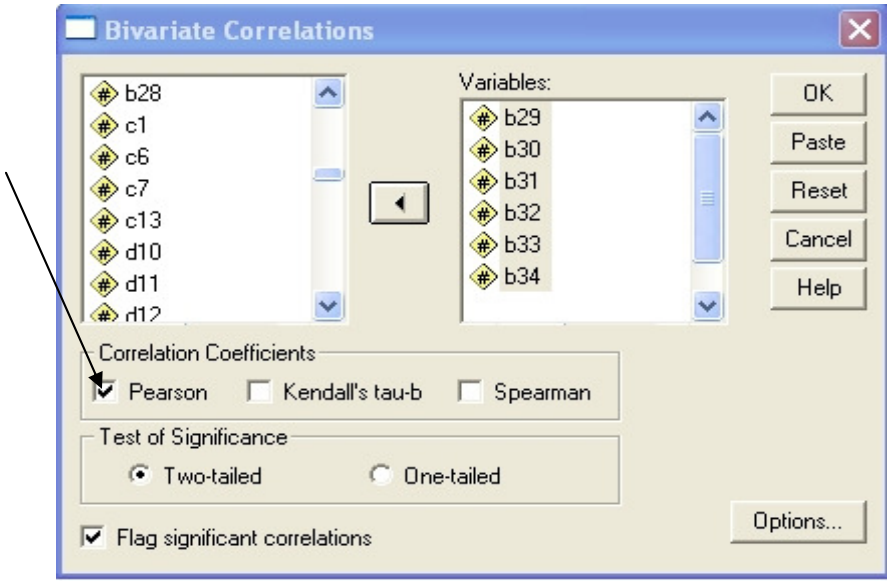
** Correlation is significant at the 0.01 level (2-tailed).

[28] O output foi melhorado no Word.

4.2. Correlação linear simples (*r* de Pearson)

A correlação linear simples permite obter uma medida (*coeficiente de correlação r de* Pearson) através da qual se determina a força ou intensidade de uma associação linear entre duas ou mais variáveis quantitativas ou tratadas como tal (escalas tipo *Likert*).

Exemplo: Correlação entre as variáveis satisfação com a vida (***b29***), com a economia (***b30***), com o Governo (***b31***), com a democracia (***b32***), com a educação (***b33***) e com os serviços de saúde (***b34***):



O resultado é o seguinte:

Correlations

		Satisfação com a vida em geral	Economia	Governo	Democracia	Educação
Satisfação com a vida em geral	Pearson Correlation					
	Sig. (2-tailed)					
	N					
Economia	Pearson Correlation	,339**				
	Sig. (2-tailed)	,000				
	N	1441				
Governo	Pearson Correlation	,280**	,578**			
	Sig. (2-tailed)	,000	,000			
	N	1413	1392			
Democracia	Pearson Correlation	,348**	,403**	,507**		
	Sig. (2-tailed)	,000	,000	,000		
	N	1371	1353	1339		
Educação	Pearson Correlation	,205**	,361**	,289**	,300**	
	Sig. (2-tailed)	,000	,000	,000	,000	
	N	1429	1389	1367	1338	
Serviços de Saúde	Pearson Correlation	,195**	,396**	,340**	,294**	,537**
	Sig. (2-tailed)	,000	,000	,000	,000	,000
	N	1489	1440	1412	1370	1433

**. Correlation is significant at the 0.01 level (2-tailed).

Interpretação: as correlações são positivas e significativas entre todas as variáveis; (*p*=0,000), sendo a menor entre a satisfação com a vida e com a educação (0,205) e a maior entre a satisfação com o Governo e com a economia (0,578

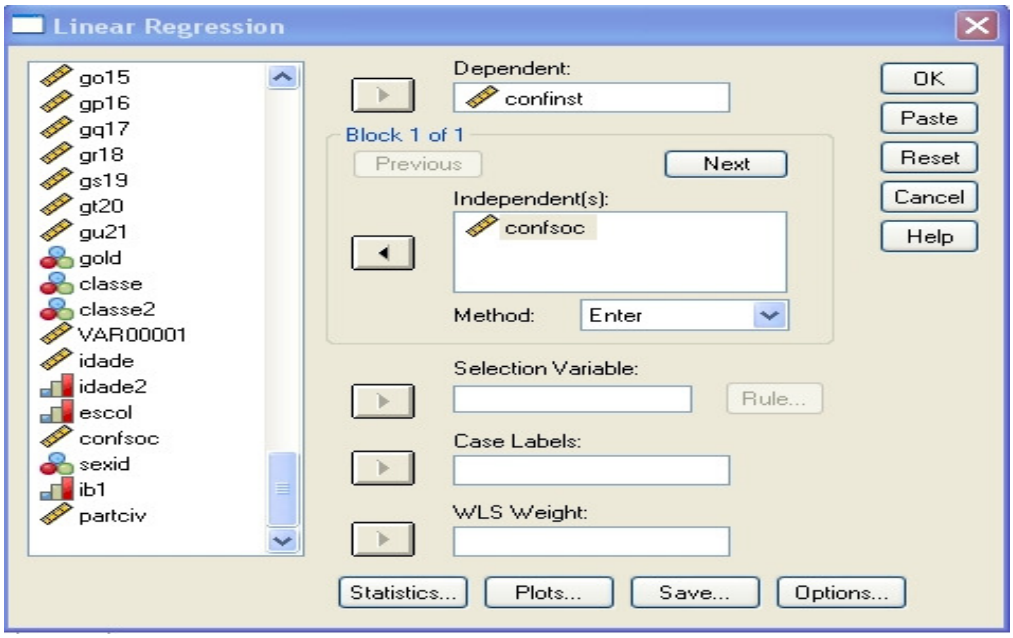
5. Regressão linear simples

A regressão linear, como referem Bryman e Cramer, "*é um poderoso instrumento para resumir a natureza da associação entre variáveis e para fazer previsões acerca dos valores da variável dependente*".[29]

Na regressão linear simples, o objectivo é sintetizar a associação entre duas variáveis (independente e dependente), produzindo uma linha (recta de regressão) que se aproxime dos dados recolhidos. Ou seja, prever *Y* (variável dependente) a partir de *X* (variável independente).

Exemplo: pretende-se saber em que medida é que a confiança social (***confsoc***) – variável independente – explica a confiança institucional (***confinst***) – variável dependente.

O procedimento consiste na realização da Análise de regressão linear simples entre as duas variáveis:



O resultado é o seguinte:

Model Summary

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate
1	.310[a]	.096	.095	1.6746

a. Predictors: (Constant), Índice sintético de Confiança social

Coeficiente de correlação
Quantifica a força da associação entre as duas variáveis

Coeficiente de determinação
Quantifica a percentagem de variação da variável dependente que é explicada pela variação da variável independente

R2 ajustado
Valor do R^2 ajustado às variáveis independentes. É o que deve ser interpretado na Regressão linear Múltipla

Erro estandardizado da estimativa
Dispersão dos resíduos relativamente ao modelo estimado
Interpretação: os níveis de confiança institucional previstos desviam-se, em média, **1,6746** dos níveis de confiança institucional observados

A correlação entre as variáveis é fraca (*R*=0,310) e o coeficiente de determinação é muito fraco (R^2=0,096). Ou seja, apenas 9,6% da variação da *confiança institucional* – variável dependente – é explicada pela variação da *confiança social* – variável independente.

[29] Alan Bryman e Duncan Cramer, *op.cit.*: 212

O resultado é o seguinte:

ANOVA [b]

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	392.354	1	392.354	139.912	.000[a]
	Residual	3693.263	1317	2.804		
	Total	4085.617	1318			

a. Predictors: (Constant), Índice sintético de Confiança social

b. Dependent Variable: Índice sintético de Confiança institucional

O teste *F* (quadro Anova) dá-nos informação sobre a adequabilidade do modelo. Tem as seguintes hipóteses estatísticas:

H_0: $\rho^2=0$ ou H_0: $\beta_1 = 0$ (R^2 na população é igual a 0 ou β_1 na população é igual a 0)

H_0: $\rho^2\neq 0$ ou H_0: $\beta_1 \neq 0$ (R^2 na população é diferente de 0 ou β_1 na população é diferente de 0)

Neste caso (F (1)=139,912; p=0,000), rejeita-se a hipótese de isso acontecer. O modelo de regressão linear é adequado.

Coefficients [a]

Model		Unstandardized Coefficients		Standardized Coefficients	t	Sig.
		B	Std. Error	Beta		
1	(Constant)	2.716	.125		21.778	.000
	Índice sintético de Confiança social	.318	.027	.310	11.828	.000

a. Dependent Variable: Índice sintético de Confiança institucional

Estimativas para a ordenada na origem (constant) e para o coeficiente de regressão (declive da recta), que permitem escrever a equação da recta estimada:

Confiança institucional* = 2,176 + 0,318 *Confiança social

Estimativa estandardizada para o coeficiente de regressão (Beta = R na regressão linear simples).

Nas estimativas estandardizadas a ordenada na origem é zero, simplificando-se a equação da recta estimada:

Confiança institucional* = 0,310 *confiança social

Na Regressão linear os testes *t-Student* permitem inferir sobre os parâmetros (ordenada na origem e coeficiente de regressão) a partir das suas estimativas.

Habitualmente testa-se a nulidade desses parâmetros:

H_0: $\beta_0 = 0$;

H_0: $\beta_1 = 0$;

Neste caso, rejeita-se a hipótese nula nos dois parâmetros: a recta não passa na origem (0) e o β_1 é diferente de 0.

Bibliografia

Aranaz, M. F. (2003) *SPSS para Windows – Análisis estadístico*, Madrid, McGraw-Hill.

Bravo, R. S. (1989) *Técnicas de Investigación Social – Teoria Y Ejercícios*, Madrid, Paraninfo, &ª edição.

Bryman, A. e D. Cramer (2003) *Análise de Dados em Ciências Sociais.* Oeiras, Celta.

Coelho, J. P., L. M. Cunha, I. G. Martins (2008) *Inferência Estatística - com utilização do SPSS e G*power*, Lisboa, Edições Sílabo.

Ferrando, M. G. (1989) *Socioestadística - Introdución a la estadística en Sociologia*, Madrid, Alinza Editorial, 3ª edição.

Guimarães, R. e J. A. S. Cabral (1997), *Estatística*, Lisboa, McGraw-Hill.

Maroco, J. (2007) *Análise Estatística Com o PASW Statistics (ex-SPSS)*, Lisboa, Wook.

Murteira, B. (1990) *Probabilidades e Estatística*, Lisboa, McGraw-Hill, vol.II.

Pestana, M. H., J. N. Gageiro (2003) *Análise de Dados para as Ciências Sociais – A complementaridade do SPSS*, Lisboa, Sílabo, 3ª edição.

Rada, V. D. (2002) *Técnicas de Análisis Multivariante para Investigación Social y Comercial (Ejemplos práticos utilizando SPSS versión 11)*, Madrid, Ra-Ma.

Vinacua, B. V. (2002) *Análisis Estadístico com SPSS para Windows, Volumen I – Estadística Básica*, Madrid, McGraw-Hill, 2ª edição.

Links úteis relacionados com a matéria deste manual

http://support.spss.com/ProductsExt/SPSS/ESD/17/Download/User%20Manuals/English/SPSS%20Statistics%20Brief%20Guide%2017.0.pdf

http://support.spss.com/ProductsExt/SPSS/Documentation/Manuals/15.0/SPSS%20Base%20User's%20Guide%2015.0.pdf

http://www2.uca.es/serv/ai/formacion/spss/Inicio.pdf

http://www.stat.tamu.edu/spss.php/

http://ritco.biz/files/Statistical%20Analyses%20using%20SPSS.pdf

http://calcnet.mth.cmich.edu/org/spss/

http://support.spss.com/ProductsExt/SPSS/Documentation/Manuals/16.0/SPSS%20Tables%2016.0%20.pdf

Anexo 1

Testes e procedimentos estatísticos mais utilizados em análise de dados bivariada

ANÁLISE DE DADOS BIVARIADA INFERENCIAL: TESTES E PROCEDIMENTOS MAIS UTILIZADOS

Variável a testar e nº de amostras		Teste	Comandos SPSS	Observações
2 variáveis na escala **nominal**[1]		χ^2 (Qui-quadrado) independência	Analyze / Descriptive Statistics / Crosstabs / Statistics / Chi-square	Em tabelas 2 X 2, este comando produz também o teste de Fisher, que por vezes é mais adequado do que o χ^2[2]
Nível de medida da variável a testar	**Amostras independentes**			
Pelo menos na escala **Ordinal**	2	**Kolmogorov-Smirnov Independência**	Analyze Nonparametric tests / 2 Independent Samples	
		Mann-Whitney		Utilizado como alternativa não paramétrica ao teste *t-Student.*
	K	**Kruskal-Wallis**	Analyze Nonparametric tests / K Independent Samples	Utilizado como alternativa não paramétrica à análise de variância simples paramétrica (One-way ANOVA).
Quantitativa ou tratada como tal (Escalas tipo *Likert*)	2	**Teste *t Student***	Analyze Compare Means Independent-Samples T-Test	**Pressupostos** • A variável a testar deve ter distribuição Normal. Nas pequenas amostras (n's ≤ 30) há que testar a normalidade dessa variável nas k amostras.[3] • Homocedasticidade: as variâncias da variável a testar devem ser iguais nas k amostras.[4]
	K	**Análise de variância simples paramétrica (One-way ANOVA)**	Analyze Compare Means One-Way Anova	

[1] Verifica-se na bibliografia a execução de testes de Independência do χ^2 com uma, ou mesmo duas variáveis ordinais, no entanto, é necessário ter em atenção que este teste não tem em conta a ordem das categorias.

[2] Quando n<20 ou quando 20≤ n ≤40 e alguma frequência esperada for inferior a 5, deve interpretar-se o teste de Fisher.

[3] Recorde-se que a "violação deste pressuposto não tem consequências sérias se os grupos ou amostras tiverem dimensões razoáveis", em Murteira, B. (1990), *Probabilidades e Estatística*, Lisboa McGraw-Hill, vol.II: :349. Se alguma das amostras for de pequena dimensão, ter-se-á de averiguar a normalidade da variável dependente em cada uma das *K* amostras, através do *Teste de aderência de Kolmogorov-Smirnov (K-S)***.**
Está provado que o Teste de Aderência de Kolmogorov-Smirnov é um teste conservador (i.e. tende a não rejeitar a H_o quando esta é falsa) nos casos em que é testada a normalidade da distribuição sem especificar *a priori* a média e a variância da normal a ajustar, sendo assim necessário recorrer a estimativas desses parâmetros. Na correcção de Lilliefors são usadas tabelas que permitem fazer o ajustamento à Normal na ausência desses parâmetros. Para n's ≤ 50 o SPSS disponibiliza também o teste de Shapiro-Wilk. Assim, quando se testa a normalidade em pequenas amostras (n's ≤ 30) deve interpretar-se o teste de Shapiro-Wilk. Ambos os testes estão disponíveis em: **[Analyze ➔ Descriptive Statistics ➔ Explore ➔ Plots ➔ Normality plots with tests]**

[4] Uma nota relativa às implicações da rejeição da hipótese da homocedasticidade: "na prática, a homogeneidade da variância só se torna importante quando as dimensões das amostras (grupos ou células) forem muito diferentes, isto é, quando a maior amostra tiver uma dimensão pelo menos dupla da dimensão da menor amostra. Quando as amostras não são fortemente desequilibradas, o efeito da heterogeneidade da variância, mesmo se acentuada, é pouco significativo." in Guimarães, Rui e Cabral, José A. Sarsfield (1997), *Estatística*, Lisboa, McGraw-Hill, pág.465.

ANÁLISE DE DADOS BIVARIADA INFERENCIAL: TESTES E PROCEDIMENTOS MAIS UTILIZADOS

(continuação)

Amostras emparelhadas		**Teste**	**Comandos SPSS**	**Observações**
Nível de medida das variáveis a testar	**Amostras emparelhadas**			
Pelo menos na escala **Ordinal**	2	**Wilcoxon**	Analyze Nonparametric tests 2 Related Samples	Pode ser utilizado como alternativa não paramétrica ao teste *t-Student* para diferença de médias em 2 amostras emparelhadas.
	K	**Friedman**	Analyze Nonparametric tests K Related Samples	
Comparar a média de 2 variáveis **quantitativas** ou tratadas como tal (Escalas tipo *Likert*)		**Teste *t-Student* para amostras emparelhadas**	Analyze Compare Means Paired-Samples T-Test	**Pressuposto** A variável que resulta das diferenças deve ter uma distribuição normal. Se o n ≤ 30 há que testar a normalidade.

2 variáveis quantitativas (ou tratadas como tal, escalas tipo *Likert*)

Quantificar a magnitude da relação linear entre 2 variáveis (*X* e *Y*) quantitativas ou tratadas como tal.	**Correlação linear simples**	Coeficiente de correlação linear (***r*** de Pearson) (Varia entre –1 e 1)	Analyze Correlate Bivariate
Quantificar a percentagem de variação de *Y* que é explicada pela variação de *X*.	**Regressão linear simples**	Coeficiente de determinação (r^2) (Varia entre 0 e 1)	Analyze Regression Linear

ANEXO 2

European Social Survey

(*round* 1 - 2002)

Extracto do Questionário

Versão completa do questionário em:

http://www.europeansocialsurvey.org/

European Social Survey

(*round* 1 - 2002)

Extracto do Questionário adaptado

Questionário integral disponível em: http://www.europeansocialsurvey.org/

O INQUÉRITO DE 2002

O objectivo central do ESS é o de desenvolver e conduzir um estudo sistemático centrado na mudança de valores, atitudes, atributos e padrões de comportamento entre os europeus, de modo a compreender a sua distribuição e variação, bem como o sentido e a intensidade da mudança, dentro de cada país e entre países.

O questionário, cuja aplicação é feita de dois em dois anos através de entrevista pessoal com cerca de uma hora de duração, consiste num módulo permanente, e em dois módulos rotativos, que serão repetidos em intervalos, cada um deles dedicado a um tema ou tópico específico. Assim, enquanto os módulos rotativos têm como objectivo permitir um conhecimento detalhado das respectivas temáticas, com interesse académico ou para o estabelecimento de medidas políticas, o módulo permanente permitirá monitorar mudanças e continuidades numa vasta gama de variáveis sócio-económicas, sócio-políticas, sócio-psicológicas e sócio-demográficas.

A8: De uma forma geral, acha que todo o cuidado é pouco quando se lida com as pessoas ou acha que se pode confiar na maioria das pessoas?

Todo o cuidado é pouco										**A maioria das pessoas é de confiança**	(Recusa)	(NS)	(NR)
0	1	2	3	4	5	6	7	8	9	10	77	88	99

A9 : Acha que a maior parte das pessoas tentam aproveitar-se de si sempre que podem, ou pensa que a maior parte das pessoas são honestas?

Tentam aproveitar-se de mim										**São honestas**	(Recusa)	(NS)	(NR)
0	1	2	3	4	5	6	7	8	9	10	77	88	99

A10: Acha que, na maior parte das vezes, as pessoas estão preocupadas com elas próprias ou acha que tentam ajudar os outros?

As pessoas estão preocupadas com elas próprias										**As pessoas tentam ajudar os outros**	(Recusa)	(NS)	(NR)
0	1	2	3	4	5	6	7	8	9	10	77	88	99

VOU AGORA FAZER-LHE ALGUMAS PERGUNTAS SOBRE POLÍTICA

B1: De um modo geral, qual o seu interesse pela política?

Muito interesse,	1
Algum interesse,	2
Pouco interesse,	3
ou, Nenhum interesse?	4
(Recusa)	7
(Não sabe)	8
(Não responde)	9

Diga-me, por favor, qual a confiança pessoal que tem em cada uma das instituições que lhe vou dizer:

		Nenhuma confiança										**Toda a confiança**	(Recusa)	(NS)	(NR)
B7	Parlamento Nacional?	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B8	Sistema Jurídico?	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B9	Polícia?	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B10	Políticos?	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B11	Parlamento Europeu?	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B12	Nações Unidas?	0	1	2	3	4	5	6	7	8	9	10	77	88	99

Há várias acções que se podem desenvolver para melhorar as coisas em Portugal ou para evitar que corram mal. Durante os últimos 12 meses, fez alguma das seguintes coisas?

		Sim	**Não**	(Recusa)	(NS)	(NR)
B15	Contactou um político, um representante do governo central ou um representante do poder local	1	2	7	8	9
B16	Trabalhou para um partido político ou movimento cívico.	1	2	7	8	9
B17	Trabalhou numa organização ou associação de outro tipo	1	2	7	8	9
B18	Usou um emblema auto-colante de campanha/movimento	1	2	7	8	9
B19	Assinou uma petição	1	2	7	8	9
B20	Participou numa manifestação.	1	2	7	8	9
B21	Boicotou determinados produtos	1	2	7	8	9
B22	Comprou de propósito determinados produtos por razões de ordem política, ética ou ambiental.	1	2	7	8	9
B23	Deu dinheiro a um grupo ou organização política	1	2	7	8	9
B24	Participou em acções de protesto ilegais (corte de estrada, boicote eleitoral).	1	2	7	8	9

B28: Em política é costume falar-se de esquerda e direita. Como é que se posicionaria nesta escala, em que 0 representa a posição mais à esquerda e 10 a posição mais à direita?

Esquerda										**Direita**	(Recusa)	(NS)	(NR)
0	1	2	3	4	5	6	7	8	9	10	77	88	99

B29: Tudo somado, qual é o seu grau de satisfação com a vida em geral?
B30: De um modo geral qual o seu grau de satisfação com o estado actual da economia portuguesa?
B31: Qual é o seu grau de satisfação com a forma como o Governo está a actuar?
B32: Qual o seu grau de satisfação com o funcionamento da democracia em Portugal?

	Extrema-mente insatisfeito										**Extrema-mente satisfeito**	(Recusa)	(NS)	(NR)
B29	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B30	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B31	0	1	2	3	4	5	6	7	8	9	10	77	88	99
B32	0	1	2	3	4	5	6	7	8	9	10	77	88	99

B33: Utilizando a seguinte escala, diga, por favor, como avalia, no geral, o estado da Educação em Portugal, hoje em dia?

Extremamente mau										**Extremamente bom**	(Recusa)	(NS)	(NR)
0	1	2	3	4	5	6	7	8	9	10	77	88	99

B34: E relativamente aos serviços de saúde em geral? Qual o seu grau de satisfação com os Serviços de Saúde em Portugal hoje em dia?

Extremamente mau										**Extremamente bom**	(Recusa)	(NS)	(NR)
0	1	2	3	4	5	6	7	8	9	10	77	88	99

VOU AGORA FAZER-LHE ALGUMAS PERGUNTAS SOBRE MIGRAÇÕES

Da seguinte lista de aspectos diga qual a importância que deve ser dada a cada um deles para deixar vir para cá alguém que nasceu, foi educado e viveu fora de Portugal.

		Não deve ser dada importância nenhuma										**Deve ser dada muita importância**	(Recusa)	(NS)	(NR)
D10	ter boas qualificações académicas	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D11	ter familiares próximos a viver cá	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D12	saber falar Português	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D13	ter formação cristã	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D14	ser branco	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D15	ser rico	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D16	ter qualificações profissionais de que Portugal precisa	0	1	2	3	4	5	6	7	8	9	10	77	88	99
D17	Querer adaptar-se ao mesmo modo de vida dos Portugueses	0	1	2	3	4	5	6	7	8	9	10	77	88	99

Diga, por favor, em que medida concorda ou discorda com cada uma das seguintes afirmações.

		Concorda totalmente	**Concorda**	**Nem concorda nem discorda**	**Discorda**	**Discorda totalmente**	(Recusa)	(NS)	(NR)
D18	As pessoas que vêm viver e trabalhar para cá fazem com que os salários baixem.	1	2	3	4	5	7	8	9
D19	As pessoas que vêm viver e trabalhar para cá, em regra, prejudicam mais as expectativas económicas dos pobres do que dos ricos.	1	2	3	4	5	7	8	9
D20	As pessoas que vêm viver e trabalhar para cá ajudam a preencher lugares em que há falta de trabalhadores.	1	2	3	4	5	7	8	9
D21	Se as pessoas que vieram viver e trabalhar para cá estiverem desempregadas por muito tempo deviam ser obrigadas a ir embora.	1	2	3	4	5	7	8	9
D22	As pessoas que vieram viver para cá devem ter os mesmos direitos do que todas as outras pessoas.	1	2	3	4	5	7	8	9
D23	Se as pessoas que vieram viver para cá cometerem **um crime grave**, devem ser obrigadas a ir embora.	1	2	3	4	5	7	8	9
D24	Se as pessoas que vieram viver para cá cometerem **qualquer crime**, devem ser obrigadas a ir embora.	1	2	3	4	5	7	8	9

FINALMENTE GOSTARIA DE LHE FAZER ALGUMAS PERGUNTAS SOBRE SI E AS OUTRAS PESSOAS DO SEU AGREGADO FAMILIAR

F2: CODIFIQUE O SEXO [ ]

F3: E em que ano nasceu?

Recusa 77 // Não sabe 88 // Não Responde 99

F7: Quantos anos completos de escolaridade terminou?

ESCREVER O NÚMERO:

Recusa 77
Não sabe 88
Não responde 99