

Crise, Representação Política e Renovação Democrática

O caso Português no contexto da Europa do Sul

CIES-IUL

20 de janeiro de 2018
Criado por: Augusta Correia

Crise, Representação Política e Renovação Democrática

O caso Português no contexto da Europa do Sul

Conteúdo

CONTEXTUALIZAÇÃO DO ESTUDO	3
NORMAS DE CITAÇÃO	4
CONTACTOS	4
METODOLOGIA.....	6
UNIVERSO.....	6
RECOLHA DOS DADOS	7
AMOSTRA	8
MARGEM DE ERRO E INTERVALO DE CONFIANÇA.....	9
RESULTADOS.....	10
OS ESTUDOS ON-LINE E A APLICAÇÃO DE PESOS PARA CORRIGIR ENVIESAMENTOS	10
DISTRIBUIÇÃO DA AMOSTRA E CÁLCULO DE PESOS	17
SEXO	17
IDADE	18
REGIÃO	19
ESCOLARIDADE.....	20
CRUZAMENTO IDADE E EDUCAÇÃO	23
BIBLIOGRAFIA	24

Contextualização do estudo

O Centro de Investigação e Estudos de Sociologia do ISCTE - Instituto Universitário de Lisboa (CIES-IUL) em parceria com Instituto Português de Relações Internacionais da Universidade Nova de Lisboa o (IPRI-UNL) encontram-se neste momento a desenvolver um projeto de investigação denominado “Crise, Representação Política e Renovação da Democracia: O caso Português no Contexto do Sul da Europa”. O projeto coordenado pelos Professores Doutores André Freire, Emmanouil Tsatsanis e Marco Lisi é financiado pela Fundação para a Ciência e Tecnologia (Ref.^a PTDC/IVC-CPO/3098/2014) e conta com o apoio institucional da Comissão Nacional de Eleições (CNE) e da Assembleia da República Portuguesa. Este projeto insere-se em três redes internacionais de pesquisa que incluem vários países¹. O estudo conta ainda com a autorização da CNPD (Comissão Nacional de Proteção de Dados) e da CNE.

O projeto de investigação tem como objetivo analisar o caso de Portugal, um dos países da Europa do Sul mais severamente afetados pela crise, tanto numa perspetiva longitudinal como comparativa. De forma mais concreta pretende-se: em primeiro lugar, descrever como os padrões de representação política têm mudado ao longo do tempo, analisando como as ligações entre os cidadãos e os seus representantes têm evoluído desde a crise de 2008 e após a execução das políticas de austeridade que se seguiram, especialmente em Portugal e noutros países do Sul da Europa. Em segundo lugar, analisar os processos de desalinhamento (ou realinhamento) partidário de acordo com várias dimensões, sobretudo na arena eleitoral. Em terceiro lugar, compreender o impacto da crise e das políticas de austeridade no apoio (difuso e específico) dos cidadãos à democracia. Em quarto lugar, escrutinar o papel relativo dos partidos, dos sindicatos e dos movimentos sociais na canalização da participação e do protesto. Nesse sentido foram aplicados inquéritos a dois grandes grupos: os candidatos às últimas eleições legislativas (eleitos e não eleitos) e os eleitores. É de referir que, no caso dos eleitores, os investigadores recorreram ao inquérito online por painel onde os mesmos participantes serão inquiridos em duas vagas diferentes. O questionário foi aplicado pela empresa de pesquisa de mercado Netquest e as perguntas foram elaboradas em função das seguintes dimensões em análise:

- Democracia e representação
- Confiança nas instituições
- Ideologia e identidade partidária

¹ Comparative Candidate Survey (CCS)
Elites Parlamentarias de América Latina (PELA)
European Election Study (EES) & European Candidate Survey (ECS)
Individualized Representation and its Prerequisites

- Avaliação do desempenho do Governo
- Avaliação da situação económica do país e do próprio
- Opinião em relação às funções dos deputados
- Integração europeia
- Assuntos políticos e políticas públicas
- Participação política
- Sentido de voto
- Exposição aos meios de comunicação social
- A crise económica e as alternativas
- Dados pessoais e contexto sociodemográfico

O presente relatório tem como objetivo apresentar os aspetos metodológicos relativos à aplicação e resultados da primeira vaga de inquéritos aos eleitores.

Normas de citação

Freire, A., Lisi, M. e Tsatsanis, E. (2016), “Inquérito aos Cidadãos Portugueses 2016-2018”, Projeto de Investigação do CIES-IUL e do IPRI-NOVA, “Crise, Representação Política e Renovação da Democracia: O caso Português no Contexto do Sul da Europa”, FCT: PTDC/IVC-CPO/3098/2014, disponível em linha: <http://er.cies.iscte-iul.pt/>

Freire, A., Lisi, M. & Tsatsanis, E. (2017), “Portuguese Citizens Survey 2016-2018”, research project at CIES-IUL and IPRI-NOVA, “Crisis, Political Representation and Democratic Renewal: The Portuguese case in the Southern European context”, FCT: PTDC/IVC-CPO/3098/2014, available online at: <http://er.cies.iscte-iul.pt/>

Contactos

Para qualquer informação adicional por favor contacte:

André Freire, Professor Auxiliar com Agregação ISCTE-IUL (Instituto Universitário de Lisboa), Avenida das Forças Armadas, 1649-026, Lisboa, Portugal. Email: andre.freire@iscte.pt

Marco Lisi, Professor Auxiliar, Departamento de Estudos Políticos, Faculdade de Ciências Sociais e Humanas – Av. de Berna, 26 C, 1069-061 Lisboa | Portugal. Email: mlisi@fcsh.unl.pt

Qualquer questão relacionada com a base de dados, por favor, contate: Augusta Correia, CIES-IUL, Av. Das Forças Armadas, Edf.I, 1649-026 Lisboa, Portugal. Email: augusta_correia@iscte.pt ; ou Vera Ramalhete, IPRI-NOVA, Avenida de Berna, Avenida de Berna, 26-C, 1069-061 Lisboa. Email: vramalhete@fcsh.unl.pt

Metodologia

Universo

O **Universo** deste estudo corresponde à População residente em Portugal Continental com idades compreendida entre os 18 e os 70 anos. De acordo com os dados do INE (Censos 2011) esta população é constituída por 6 939 959 indivíduos. As tabelas que se seguem apresentam a **distribuição da população (casos e percentagem)** de acordo com os estratos pré-definidos à recolha de dados - sexo, idade e região NUTs II.

TABELA 1. DISTRIBUIÇÃO DA POPULAÇÃO PORTUGUESA RESIDENTE EM PORTUGAL CONTINENTAL COM IDADES COMPREENDIDAS ENTRE OS 18 E OS 70 ANOS POR SEXO E REGIÃO (NUTSII) (N)

Sexo	Idade	Frequências (N)					
		Região NUT's II (N)					
		Norte	Centro	Lisboa	Alentejo	Algarve	Total
Homem	[18-29]	267.008	152.718	194.686	49.090	30.039	693.541
	[30-39]	271.411	159.211	222.217	52.217	34.609	739.665
	[40-49]	273.139	160.439	189.442	51.837	32.193	707.050
	[50-59]	240.279	150.916	166.997	49.901	29.284	637.377
	[60-70]	197.569	141.600	162.978	45.511	27.055	574.713
Mulher	[18-29]	267.286	150.955	198.643	46.303	30.147	693.334
	[30-39]	286.438	166.255	235.847	51.664	35.561	775.765
	[40-49]	293.876	170.190	207.527	52.473	33.569	757.635
	[50-59]	262.511	161.153	194.239	50.589	30.274	698.766
	[60-70]	226.425	162.944	191.584	52.758	28.402	662.113
Total		2.585.942	1.576.381	1.964.160	502.343	311.133	6.939.959

Origem: Instituto Nacional de Estatística (Censos 2011)

TABELA 2. DISTRIBUIÇÃO DA POPULAÇÃO PORTUGUESA RESIDENTE EM PORTUGAL CONTINENTAL COM IDADES COMPREENDIDAS ENTRE OS 18 E OS 70 ANOS POR SEXO E REGIÃO (NUTSII) (%)

Sexo	Idade	Percentagem (%)					
		Região NUT's II					
		Norte	Centro	Lisboa	Alentejo	Algarve	Total
Homem	[18-29]	3,8	2,2	2,8	0,7	0,4	10,0
	[30-39]	3,9	2,3	3,2	0,8	0,5	10,7
	[40-49]	3,9	2,3	2,7	0,7	0,5	10,2
	[50-59]	3,5	2,2	2,4	0,7	0,4	9,2
	[60-70]	2,8	2,0	2,3	0,7	0,4	8,3
Mulher	[18-29]	3,9	2,2	2,9	0,7	0,4	10,0
	[30-39]	4,1	2,4	3,4	0,7	0,5	11,2
	[40-49]	4,2	2,5	3,0	0,8	0,5	10,9
	[50-59]	3,8	2,3	2,8	0,7	0,4	10,1
	[60-70]	3,3	2,3	2,8	0,8	0,4	9,5
Total		37,3	22,7	28,3	7,2	4,5	100,0

Origem: Instituto Nacional de Estatística (Censos 2011)

Recolha dos dados

A recolha de dados foi obtida através de inquérito on-line em painel. Foram enviados 7878 convites. A taxa de participação 51% (n=4044). Após o controlo de qualidade e a aplicação de filtros foram obtidos 2897 inquéritos validados (37% do total de convites e 72% dos participantes).

TABELA 3. RESUMO DAS ENTREVISTAS REALIZADAS

Entrevistas	Resultados
Filtrados	155
Filtrados por ISO	90
Quota full	177
Security question	113
Abandonos	612
Participantes	4044
Convites	7878
Completas	2897

Os dados foram recolhidos pela empresa Soluciones Netquest de Investigación SL. O processo de recolha de dados decorreu entre os dias 2 de maio de 2015 e 3 de julho de 2016.

Em média os inquiridos demoraram 37 minutos a responder ao questionário.

Dados da empresa que conduziu o inquérito:

Soluciones Netquest de Investigación SL

B-62470489

Calle Gran Capitán 2-4

Edificio Nexus I

08034 Barcelona

España

Tel: (34) 93 205 00 63

www.netquest.com

Amostra

Optou-se por uma amostra de 2500 inquéritos estratificada por sexo, região NUT's II e idade. No decorrer do processo de recolha de dados foi necessário aumentar o número de inquéritos no sentido de obter uma maior representatividade em todos os estratos. Assim, a amostra final é composta por 2897 inquéritos distribuídos de acordo com as tabelas que se seguem.

TABELA 4. DISTRIBUIÇÃO DA AMOSTRA POR ESTRATOS SEXO, IDADE E REGIÃO NUT'S II (n)

		Frequências (n)					
		Região NUT's II (n)					
Sexo	Idade	Norte	Centro	Lisboa	Alentejo	Algarve	Total
Homem	[18-29]	106	70	48	6	10	240
	[30-39]	100	64	85	21	10	280
	[40-49]	110	99	97	35	9	350
	[50-59]	148	88	78	38	12	364
	[60-70]	57	56	47	17	9	186
Mulher	[18-29]	171	94	76	20	8	369
	[30-39]	108	70	92	21	14	305
	[40-49]	106	95	112	23	12	348
	[50-59]	106	78	101	28	25	338
	[60-70]	37	25	42	7	6	117
Total		1049	739	778	216	115	2897

TABELA 5. DISTRIBUIÇÃO DA AMOSTRA POR ESTRATOS SEXO, IDADE E REGIÃO NUT'S II (%)

		Percentagem (%)					
		Região NUT's II					
Sexo	Idade	Norte	Centro	Lisboa	Alentejo	Algarve	Total
Homem	[18-29]	3,7	2,4	1,7	0,2	0,3	8,3
	[30-39]	3,5	2,2	2,9	0,7	0,3	9,7
	[40-49]	3,8	3,4	3,3	1,2	0,3	12,1
	[50-59]	5,1	3,0	2,7	1,3	0,4	12,6
	[60-70]	2,0	1,9	1,6	0,6	0,3	6,4
Mulher	[18-29]	5,9	3,2	2,6	0,7	0,3	12,7
	[30-39]	3,7	2,4	3,2	0,7	0,5	10,5
	[40-49]	3,7	3,3	3,9	0,8	0,4	12,0
	[50-59]	3,7	2,7	3,5	1,0	0,9	11,7
	[60-70]	1,3	0,9	1,4	0,2	0,2	4,0
Total		36,2	25,5	26,9	7,5	4,0	100

Margem de erro e Intervalo de Confiança

A margem de erro máxima da amostra é de $\pm 1,82\%$ para um nível de confiança de 95%. Ou seja, existe uma probabilidade de 95% do intervalo de confiança ($\pm 1,82\%$) em torno das estimativas emanadas do estudo incluírem a “real” percentagem presente na população.

Resultados

Os estudos on-line e a aplicação de pesos para corrigir enviesamentos

A aplicação de questionários on-line tem-se generalizado nas últimas décadas, com particular ênfase nos estudos ligados à pesquisa de mercados. De forma mais gradual os estudos on-line têm também ganho cada vez mais espaço nos meios académicos e governamental. O significativo aumento deste tipo de estudos levanta, no entanto, um conjunto de questões ligadas à metodologia e sobretudo à inferência estatística (Couper, 2011).

Como nas restantes técnicas de recolha de dados, aos estudos on-line estão associados erros de observação² e erros de não observação³. Quando os instrumentos de recolha de dados estão corretamente contruídos, a recolha de dados através de inquéritos on-line tem um impacto positivo na redução dos erros de observação, também chamados erros de medida, uma vez que permitem minimizar, por exemplo, os erros associados ao inquiridor, à digitalização dos dados ou mesmo à ordem pela qual as perguntas e/ou as categorias de resposta são apresentadas ao inquirido (Couper, 2011). Além disso, o facto de garantirem um maior anonimato facilita a recolha de dados que abordam questões sensíveis. No que toca aos estudos em painel ou longitudinais, como é o caso deste projeto, os estudos on-line também trazem vantagens, uma vez que facilitam a possibilidade de contactar novamente o inquirido e assim auscultar o mesmo em diferentes espaços temporais. São também inegáveis as vantagens dos inquéritos on-line em relação aos custos e ao tempo necessário para a recolha dos dados; já que são, geralmente, mais baratos e permitem a recolha de dados num período significativamente mais curto comparativamente às restantes metodologias.

² Também chamados erros de medida, os erros de observação dizem respeito a erros associados aos respondentes, aos instrumentos de recolha de informação, à forma como os dados foram recolhidos ou aos entrevistadores/inquiridores.

³ Erros de não observação referem-se a erros associados à cobertura da população alvo a estudar e posterior inferência dos resultados. Estão associados a este tipo de erros os problemas ligados ao desenho da amostra, às deficiências da base de sondagens (sampling frame) ou às não respostas.

Apesar dos benefícios, os estudos on-line colocam grandes desafios à extrapolação dos resultados para a população em estudo. São três os aspetos mais sensíveis associados aos estudos on-line: a cobertura da população, a amostragem e as taxas de não resposta (Couper, 2011).

A identificação de uma base de sondagens (*sampling frame*) é uma tarefa complexa, uma vez que a população de indivíduos que utiliza a internet é muito volátil e normalmente os estudos on-line pretendem inferir para toda a população e não só para os que recorrem à internet (Couper, 2011). Em Portugal, por exemplo, apenas 74% das habitações têm acesso à internet, quando comparado com a média europeia onde esse valor é de 85%. Por outro lado, a cobertura de internet pelo país não é homogénea. Enquanto nas cidades a taxa de cobertura de internet é de 81%, nas zonas sub-urbanas esse valor decai para 75% e no caso das zonas rurais a percentagem de casas com acesso à internet é de apenas 60%. É ainda importante referir que apenas 60% indicam utilizar a internet diariamente. E entre os que utilizam a internet (pelo menos uma vez por ano) 51% dos utilizadores portugueses diz não disponibilizar nenhuma informação pessoal na internet (um dos valores mais elevados observados entre os países da União Europeia, sendo que a média europeia é de 28%) (Eurobarómetro 2006). Ou seja, podemos concluir que nem todos os elementos da população têm uma probabilidade (conhecida) de serem seleccionados - condição necessária para uma amostra probabilística.

Paralelamente às questões ligadas à cobertura, também se levantam desafios no que toca a amostragem, uma vez que vários estudos indicam que são os mais novos e instruídos que recorrem à internet e também por isso os que mais participam nos inquéritos on-line (Couter 2011; Lynn 1996). As taxas de resposta (ou as não respostas) são também outras das questões que decorrem da utilização da internet como instrumento de recolha de dados. Embora esta questão seja hoje um problema de todas as técnicas (Couter, 2011).

Estas dificuldades associadas à cobertura da população, à amostragem e às taxas de resposta introduzem enviesamentos nos resultados. No caso dos estudos em painel, estas questões são particularmente

relevantes, uma vez que são os indivíduos que estão inscritos no painel que constituem a base de sondagens e muitos painéis falham na constituição de uma amostra representativa e ainda em conseguir que as pessoas participem numa segunda vaga do estudo, conduzindo, desta forma, a uma elevada taxa de não resposta parcial (Brick, 2013). O facto de alguns elementos da população terem uma probabilidade significativamente menor do que o desejável de participar no estudo é assim o grande problema das ciências sociais (Kelton & Flores-Cervantes, 2003) e, de forma acentuada, dos estudos on-line.

A correção dos enviesamentos introduzidos pelas taxas de não respostas, falhas na cobertura e ainda falhas no processo de amostragem é possível através da pós-estratificação também conhecidos por ponderadores (Couper 2011; Miratrix et al, 2017). Ainda que permaneça escassa a literatura sobre esta ferramenta e as suas implicações na análise e interpretação dos dados (Miratrix et al 2017; Lew, Himes-Cornell & Lee 2015; Hofler et al, 2005).

O recurso ao ponderador caracteriza-se pela aplicação de um fator multiplicador constante a cada caso presente na amostra (Dorofeev & Grant, 2006). O peso aplicado a cada respondente é tipicamente inverso à probabilidade de ser selecionado para o estudo. Existem várias metodologias para calcular o peso associado a cada indivíduo. As metodologias utilizadas são geralmente condicionadas pelas informações conhecidas sobre a população, o seu nível de detalhe e ainda o número de variáveis auxiliares consideradas na construção dos pesos. De uma maneira geral, destacam-se duas metodologias na construção dos pesos. Pesos por célula (cell weighting) – a mais utilizada - e pesos por RIM (RIM weighting - Random Iterative Method), que se caracterizam por serem construídos com base num processo iterativo com auxílio de softwares de análise de dados onde o peso é calculado individualmente considerando as variáveis auxiliares (Couter, 2011; Sharot 1986). Em ambos os casos, a criação dos pesos passa por identificar respondentes que sejam similares aos não respondentes em termos de informação auxiliar⁴ disponível tanto para os respondentes como para os não respondente e nesse caso aumentar o peso base dos respondentes de forma

⁴ Na maioria dos casos variáveis sociodemográficas uma vez que são as poucas informações que se sabe sobre os não respondentes.

a que estes passem também a representar os não respondentes (Kalton & Flores-Cervantes, 2003). Os ajustamentos introduzidos pelos pesos forçam a distribuição amostral de uma variável a aproximar-se da distribuição dessa mesma variável na população para a qual se quer extrapolar os dados. No caso do presente estudo, a criação dos pesos teve como objetivo aproximar os resultados da amostra aos dados observados na população portuguesa de acordo com as informações disponibilizados pela Instituto Nacional de Estatista – Censos 2011.

Embora a aplicação de ponderadores permita reduzir os enviesamentos conhecidos, é importante avaliar o impacto da sua utilização na interpretação dos dados, uma vez que ela não é isenta de riscos. Os pesos só deverão ser utilizados se as variáveis auxiliares estiverem relacionadas com a variável em estudo, caso contrário introduzem enviesamentos que não existem inicialmente na amostra (Lynn 1996; Pike 2007).

Outra das questões que se coloca à aplicação de pesos é o facto de estes implicarem um aumento da variância⁵ e conseqüentemente uma diminuição da precisão das estimativas. Quanto mais pequena for a amostra (menos de 100 casos) maior serão as conseqüências dos pesos na precisão dos resultados. Pelo que, o investigador terá de avaliar cuidadosamente a aplicação dos pesos. Ainda que, quando identificados os enviesamentos deverão ser corrigidos. Como referem Dorofeev & Grant (2006, p. 48): *“being aware of material biases and not making some attempt to improve the situation by applying weights requires a good explanation”*.

Face às implicações inerentes ao recurso aos pesos para reduzir os enviesamentos, a criação dos pesos e as implicações da sua utilização nas estimativas finais deverão ser cuidadosamente avaliadas. Como já foi referido, a maior conseqüência da aplicação do peso é o aumento da variância. No entanto, é possível calcular qual o Efeito do Peso (*WE*) na perda de precisão da amostra e desta forma permitir ao investigador tomar decisões sobre a utilização dos ponderadores. Neste caso, a identificação da dimensão da amostra calibrada (notação n_c) ou também denominada Dimensão da Amostra Efetiva (Baxter & Media, 2016) é

⁵ Em estatística existem dois tipos de variância. Ambos medem a dispersão dos resultados. O primeiro diz respeito à dispersão que se observa no interior de uma amostra. Exemplo o quanto a satisfação com o desempenho do governo varia em torno da média amostral. O segundo diz respeito a quanto as estimativas de um parâmetro variam de amostra para amostra. Quanto menor é esta variância maior é a precisão da estimativa. Raramente é possível calcular o valor de várias amostras para se calcular diretamente a variância, pelo que normalmente se recorre a teoria estatística para fazer esse cálculo. O valor obtido servirá de base para calcular o Erro Padrão e o Intervalo de Confiança de uma estimativa. É este último indicador que é afetado pela utilização de pesos.

uma importante ferramenta para avaliar o impacto dos ponderadores sobre a variância (Dorofeev & Grant, 2016).

A fórmula que se segue permite obter a Dimensão da Amostra Calibrada (notação n_c) (Dorofeev & Grant, 2016).

$$n_c = \frac{(\sum_{i=1}^n w_i)^2}{\sum_{i=1}^n w_i^2}$$

A grande vantagem obtida através do cálculo da Dimensão da Amostra Calibrada, além da sua simplicidade, é o facto de ela ser igual em todas as variáveis do estudo. A Dimensão da Amostra Calibrada é uma característica da amostra e não depende por isso das variáveis em estudo. Com base na n_c é possível calcular as margens de erro e o intervalo de confiança das estimativas obtidas através da amostra quando aplicado o peso (Dorofeev & Grant, 2006).

Ao dividirmos a dimensão amostra inicial (n) pela Dimensão da Amostra Calibrada (n_c) obtemos o efeito do peso - Weight Effect (WE) sobre a variância das estimativas. O WE é definido pela seguinte fórmula:

$$WE = \frac{n}{n_c}$$

A Dimensão da Amostra Calibrada é sempre inferior à dimensão original da amostra pelo que o WE assume sempre um valor superior a 1 (Dorofeev & Grant, 2006). O WE é um fator que aplicado à variância inicial (sem o efeito do peso) permite saber o valor da variância quando afetada pelo ponderador. O recurso aos ponderadores implica sempre um aumento da variância. Cabe ao investigador decidir qual a variância e o enviesamento que quer/pode assumir.

Kalton e Flores-Cervantes (2003) referem uma outra fórmula (apresentada por Kish, 1965) que permite obter os mesmos resultados de WE . Denominam-no *Medida F* ou *Fator de Inflação da Variância*, mas como já foi referido ambos permitem obter o mesmo resultado.

$$F = 1 + CV(w_i)^2$$

Por outro lado, quando invertido o valor de WE (ou de F) e o multiplicamos por 100 é possível obter a Eficiência do Peso (E) (Baxter e Media, 2016). Ou seja, a percentagem da precisão original que é mantida com a utilização do peso.

$$E = \frac{n_c}{n} \times 100\%$$

São vários os fatores que influenciam o impacto do peso na variância das estimativas. Entre eles está o valor dos pesos. Os pesos com valores extremos têm maior impacto na perda de precisão (Sharot, 1986; Biemer & Christ 2008, Dorofeev & Grant, 2006). Existem autores que defendem que pesos superiores a 2 devem ser “aparados” (Sharot, 1986). Já outros indicam que é necessário olhar cuidadosamente para pesos quando o *WE* é igual ou superior a 5 (Biemer e Christ 2008). Dado que quanto maior o valor do peso maior o seu impacto na precisão das estimativas. A amplitude entre o peso maior e menor também tem impacto na perda de precisão, pelo que a diferença entre peso maior e menor não deverá ser grande.

No sentido de minimizar a perda de precisão é necessário dar atenção às várias etapas da criação dos pesos, desde a escolha das variáveis auxiliares até à análise dos valores finais dos pesos, passando pela definição de classes de pesos. Em todas as etapas é necessário fazer uma análise e assumir a existência de um *trade off* entre o enviesamento da amostra e o aumento da variância.

A criação dos pesos começa com a identificação da ou das variáveis auxiliares e a definição das classes dessas variáveis. Só deverão ser consideradas as variáveis auxiliares que tenham uma relação com a variável em estudo. É preferível calcular os pesos recorrendo a um menor número de variáveis auxiliares uma vez que quanto maior o número de variáveis maior é o risco de o peso de uma variável interferir com outra (Tremblay, 1986 citado por Pike 2007) especialmente no caso dos pesos por célula. Neste sentido, privilegiámos a apresentação de pesos com base numa única variável auxiliar e apenas em casos excecionais são apresentados pesos com mais do que uma variável. No que respeita às classes, estas devem ser criadas de forma a serem mutuamente exclusivas e de forma a que não haja classes com poucos ou nenhuns casos. Kalton e Flores-Cervantes (2003) indicam que cada classe deve ter pelo menos 20 a 25 respondentes para serem estáveis de forma a minimizar o efeito em rede dos pesos. Também Tremblay (1986 citado por Pike 2007) indica três características que as classes devem ter para assegurar estimativas mais precisas: 1) cada classe deve ser o mais homogênea possível, i.é, uma baixa variância intra-classe; 2) deve existir variância inter classes e; 3) à semelhança do que refere Kalton e Flores-Cervantes (2003), as classes devem ter uma boa taxa de resposta.

Considerando especificamente este estudo, este espelha uma tendência generalizadas dos estudos on-line.

São essencialmente as pessoas mais novas e as mais escolarizadas que mais respondem aos inquéritos on-

line (Lynn, 1996) pelo que, neste estudo os pesos são particularmente relevantes quando consideradas as variáveis idade e escolaridade. Foram ainda criados pesos para corrigir enviesamentos verificados nas variáveis Sexo e Região, mas neste caso os enviesamentos não são expressivos.

Os pesos foram contruídos a partir de um rácio entre a proporção de elementos presente na população e a proporção de elementos presente na amostra dentro de cada sub-amostra. Os pesos superiores a 1 serão aplicados a sub-amostras que estão sub-representadas e os pesos inferiores a 1 serão aplicados a amostras sobre representadas. Há que ter em conta que os pesos maiores que 1 têm maior impacto na variância que pesos menores que 1 e que a amplitude entre o peso menor e o maior não deverá ser significativo. Assim, no sentido de diminuir o aumento da variância introduzida pelos pesos foi necessário fundir algumas classes de pesos.

Nas páginas que se seguem são apresentadas a distribuição da amostra e do universo nas diferentes variáveis auxiliares, os pesos associados a cada classe/categoria. Para cada ponderador é ainda calculada a Dimensão da Amostra Calibrada (n_c), o Efeito do Peso (WE), a Eficiência do Peso (E) e ainda a Margem de Erro da amostra quando aplicado o respetivo peso para um nível de confiança de 95%.

Distribuição da amostra e cálculo de pesos

Sexo

A distribuição da amostra pelas categorias da variável “sexo” é praticamente igual à verificada na população pelo que os pesos calculados não têm praticamente nenhum impacto na variância da amostra. Não havendo por esse motivo uma redução na precisão das estimativas. Neste caso, a margem de erro da amostra mantém-se igual ao valor inicial - $\pm 1,8\%$ para um nível de confiança de 95%. A Dimensão da Amostra Calibrada é de 2896 (menos 1 caso que a dimensão da amostra original).

TABELA 6. PESOS E DISTRIBUIÇÃO DA AMOSTRA E POPULAÇÃO POR SEXO

Sexo	Amostra		Proporção no Universo	Pesos
	n	%	%	
Homens	1420	49,02	48,30	0,98549
Mulheres	1477	50,98	51,70	1,01395
Total	2897	100,00	100,00	

Idade

Tal como já foi acima referido, são os mais jovens que tendem a participar de forma mais expressiva nos estudos online. O presente estudo reflete, ainda que de forma não excessivamente acentuada, essa realidade. O efeito dos pesos que permitem reduzir o enviesamento associado à idade (*WE*) sobre a variância é de 1,266. Ou seja, a aplicação dos pesos introduz no estudo uma perda de precisão de 21% (Eficiência do peso=79%). A Dimensão da Amostra Calibrada (n_c) quando aplicado o peso é de 2288. Neste caso a margem de erro máxima associada à amostra é de $\pm 2,05\%$ para um nível de confiança de 95%.

TABELA 7. PESOS E DISTRIBUIÇÃO DA AMOSTRA E POPULAÇÃO POR IDADE

Escalões etários	Amostra		Proporção no Universo	Pesos
	n	%	%	
[18-19]	58	2,0	3,2	1,579563
[20-24]	288	9,9	7,9	0,797367
[25-29]	263	9,1	8,9	0,981412
[30-39]	245	8,5	10,6	1,241499
[35-39]	340	11,7	11,3	0,964406
[40-44]	315	10,9	10,6	0,970208
[45-49]	383	13,2	10,5	0,797715
[50-54]	319	11	9,9	0,901009
[55-59]	383	13,2	9,3	0,707712
[60-64]	241	8,3	8,8	1,059284
[65-70]	62	2,1	9,0	4,299881
Total	2897	100	100	

Região

A distribuição da amostra pelas categorias da variável “Região” está muito próxima da distribuição verificada no universo, pelo que o enviesamento da amostra no que toca a esta variável não é significativo. Assim, os pesos calculados para corrigir esse enviesamento têm pouco impacto na variância das estimativas. A perda de precisão é de apenas 0,5%. Ou seja, a eficiência do peso é de 99,5%. A Dimensão da Amostra Calibrada (n_c) é, assim, de 2883 e o Efeito do peso (WE) é 1,005 (muito próximo de 1). Ao aplicar o peso a margem de erro máxima da amostra é de $\pm 1,85\%$ para um nível de confiança de 95%.

A tabela que se segue apresenta o valor dos pesos atribuídos a cada uma das categorias e a distribuição da amostra e da população pelas categorias da variável.

TABELA 8. PESOS E DISTRIBUIÇÃO DA AMOSTRA E POPULAÇÃO POR REGIÃO

Regiões	Amostra		Proporção no Universo	Pesos
	n	%	%	
Norte	1049	36,21	37,26	1,029046
Centro	739	25,51	22,71	0,890448
Lisboa	778	26,86	28,30	1,053874
Alentejo	216	7,46	7,24	0,970819
Algarve	115	3,97	4,48	1,129379
Total	2897	100,00	100,00	

Escolaridade

De forma a uniformizar os diferentes níveis de escolaridade e também permitir a comparação dos dados recolhidos no âmbito deste estudo com os publicados pelo Instituto Nacional de Estatística, os diferentes níveis de escolaridade foram convertidos de acordo com o ISCED 97. O ISCED é a classificação internacional de referência, desenvolvida pela UNESCO, que permite uniformizar os diferentes níveis de educação presentes nos vários países. Apesar de à data já existir uma nova classificação (ISCED 2011 e ISCED 2014) em 2011, data do último Censos, vigorava a classificação ISCED 97 por este motivo foi esta a classificação adotada para esta análise.

A tabela que se segue indica a correspondência entre os níveis de ensino presentes em Portugal e o ISCED 97.

TABELA 9. CONVERSÃO DOS NÍVEIS DE ENSINO DE ACORDO COM A CLASSIFICAÇÃO INTERNACIONAL - ISCED 97

Escolaridade	Classificação atual	ISCED 97	ISCED 97 REC
Sem estudos	Sem educação formal	ISCED 0	ISCED 0 e 1
Escola Primária	1º Ciclo	ISCED 1	
6ª Classe / 6º ano	2º Ciclo		
Ciclo preparatório			
5º ano antigo	3º Ciclo	ISCED 2	ISCED 2
7º,8º,9º anos unificados			
Escola Industrial/ Comercial			
7º ano antigo/ Propedêutico	Secundário	ISCED 3	ISCED 3
10º,11º,12º anos unificados			
Cursos de 2/3 anos com início a seguir ao 9º anos ou antigo 5º			
Cursos com início a seguir ao 11º/12º anos unificados ou 7º ano antigo	Pós-secundário não universitário	ISCED 4	ISCED 4, 5 e 6
Faculdade (completa ou incompleta)	Universitário	ISCED 5 e 6	

A distribuição da variável educação na amostra é a que apresenta a maior distorção face à distribuição verificada no universo. E seguindo a tendência generalizada verificada nos estudos on-line. A taxa de participação neste tipo de estudo é significativamente mais acentuada entre a população mais escolarizada. Na tabela seguinte é possível verificar uma inversão quando comparada a distribuição da amostra e a distribuição da população no que respeita aos diferentes níveis de escolaridade. O universo do estudo é caracterizado por um predomínio de indivíduos com baixa ou nenhuma escolaridade. Quase metade (43,4%) possuem no máximo o ensino preparatório, face aos menos de 10% da percentagem de indivíduos com o mesmo nível de escolaridade presente na amostra. Por outro lado, entre os indivíduos que participaram no estudo, 50% têm o ensino secundário completo, já entre a população esse grupo

representa apenas 18%. Essa disparidade conduz a que os pesos construídos para corrigir essa distorção tenham um maior impacto na variância. Para minimizar esse impacto e evitar a existência de pesos muito grandes e uma amplitude grande entre o peso menor e o mais elevado (fatores que potenciam ao aumento da variância) optou-se por fundir os níveis de educação situados nos extremos da escala de acordo com a tabela que se segue. A eficiência do peso aplicado a educação é de 41,3%, ou seja, a utilização do peso permitir reduzir o enviesamento, mas provoca uma perda de 58,7% na precisão das estimativas. A Dimensão da Amostra Calibrada, neste caso é 1196. O Efeito do Peso sobre a variância é de 2,423. Ao aplicar este ponderador a margem de erro máxima da amostra é $\pm 2,58\%$ para um nível de confiança de 95%.

TABELA 10. PESOS E DISTRIBUIÇÃO DA AMOSTRA E POPULAÇÃO POR EDUCAÇÃO

Escolaridade ISCED 97	Amostra		Proporção no Universo	Peso
	n	%	%	
ISCED 0 e 1	285	9,8	43,4	4,430920725
ISCED 2	376	13	19,8	1,526677164
ISCED 3	786	27,1	19,0	0,699668096
ISCED 4, 5 e 6	1450	50,1	17,8	0,354674019
Total	2897	100,00	100,00	

Na tentativa de reduzir a perda de precisão e simultaneamente minimizar o impacto dos enviesamentos que a amostra apresenta face à variável educação, optou-se por fundir os três primeiros níveis de educação e os três últimos. Neste caso a variável educação divide os indivíduos com escolaridade obrigatória ou menos, os indivíduos com o ensino secundário e os indivíduos com escolaridade superior ao ensino secundário em sub-grupos.

Ao aplicar o peso com base na variável educação recodificada em três categorias a precisão da amostra sofre uma perda de 48,7% face à precisão original. A eficiência do peso é de 51,3%. Neste caso a Dimensão da Amostra Calibrada é de 1485 ($WE=1,951$). Ao aplicar este ponderador à amostra a margem de erro das estimativas será de $\pm 2.55\%$ para um nível de confiança de 95%.

Neste caso cabe ao investigador ponderar qual o peso que mais se adequa à varável de estudo e as distorções que está disposto a aceitar.

TABELA 11. PESOS E DISTRIBUIÇÃO DA AMOSTRA E POPULAÇÃO POR EDUCAÇÃO

Escolaridade ISCED 97	Amostra		Proporção no Universo	Peso
	n	%	%	
ISCED 0, 1 e 2	661	22,8	63,3	2,774992379
ISCED 3	786	27,1	19,0	0,699668096
ISCED 4, 5 e 6	1450	50,1	17,8	0,354674019
Total	2897	100,00	100,00	

Cruzamento idade e educação

Uma vez que a idade e a educação são as variáveis demográficas com maiores enviesamentos, optou-se por analisar e criar um ponderador que permitisse minimizar os enviesamentos introduzidos pela ação combinada destas duas variáveis. Os pesos aqui apresentados (ver tabela abaixo) têm um efeito (*WE*) sobre a variância 2,202. Neste caso a Dimensão da Amostra Calibrada é 1316. A eficiência do peso é de 45,4%. Ao aplicar o ponderador a margem de erro máximo da amostra é de $\pm 2,7\%$ para um nível de confiança de 95%

TABELA 12. PESOS E DISTRIBUIÇÃO DA AMOSTRA E POPULAÇÃO POR IDADE E EDUCAÇÃO

Escolaridade	Idade	Amostra		Universo	Pesos
		n	%	%	
ISCED 0, 1 e 2	[18-29]	41	1,4	8,5	6,035558
	[30-39]	104	3,6	10,5	2,924881
	[40-49]	144	5,0	14,0	2,826586
	[50-59]	219	7,6	14,7	1,949012
	[60-70]	153	5,3	15,4	2,924309
ISCED 3	[18-29]	159	5,5	7,1	1,295761
	[30-39]	138	4,8	5,0	1,059808
	[40-49]	219	7,6	3,6	0,478567
	[50-59]	218	7,5	2,2	0,291475
	[60-70]	52	1,8	1,0	0,551411
ISCED 4, 5 e 6	[18-29]	409	14,1	4,3	0,306724
	[30-39]	343	11,8	6,3	0,531068
	[40-49]	335	11,6	3,4	0,297254
	[50-59]	265	9,1	2,3	0,254268
	[60-70]	98	3,4	1,4	0,410256
Total		2897	100	100	

Bibliografia

- Biemer, P. P., and Christ, S. L. (2008) "Weighting Survey Data", (chapter 17) in De Leeuw, E.D., Hox, J.J., and Dillman, D.A., *International Handbook of Survey Methodology*. New York: Lawrence Erlbaum
- Brick, JM. (2013) 'Unit Nonresponse and Weighting Adjustments: A Critical Review Discussion.' *Journal Of Official Statistics*, 29 (3). 329 - 353.
- Couper, M.P. (2011), "Web Survey Methodology: Interface Design, Sampling and Statistical Inference." *The International Statistical Seminar in the Basque Country, Vitoria-Gasteiz*, -15- November, http://www.eumat.es/productosServicios/vol53_c.html#axzz1v2GS94oO (visualizado a 10 de janeiro de 2018)
- Lew D.K., Himes-Cornell A., and Lee J. (2015) "Weighting and Imputation for Missing Data in a Cost and Earnings Fishery Survey," *Marine Resource Economics* 30, no. 2 (April 2015): 219-230.
- Dorofeev, Sergey, and Peter Grant. (2006). *Statistics for Real-Life Sample Surveys: non-simple-random samples and weighted data*. Cambridge, UK: Cambridge University Press.
- Eurobarometer (2006). Flash Eurobarometer 437 - Internet users' preferences for accessing content online. March 2016 (http://ec.europa.eu/eurostat/statistics-explained/index.php/Internet_access_and_use_statistics_-_households_and_individuals#Internet_access)
- Höfler, M., Pfister, H., Lieb, R., & Wittchen, H. (2005). The use of weights to account for non-response and drop-out. *Social Psychiatry and Psychiatric Epidemiology*, 40, 291-299.
- Kalton, G. and Flores-Cervantes, I. (2003) Weighting Methods. *Journal of Official Statistics*, 19, 81-97
- Lynn, P. (1996) "Weighting for Non-response", pp. 205-214 in Banks. R. et al. *Survey and statistical computing 1996*, ASC.
- Michael Baxter M. e Media K. (2016) A better rim weighting algorithm *International Journal of Market Research* 58(4):621-634 · August 2016
- Miratrix LW, Sekhon JS, Theodoridis AG, Campos LF. (2017). Worth Weighting? How to Think About and Use Sample Weights in Survey Experiments. Preprint on arXiv:1703.06808.
- Pike, G. R. (2007). Adjusting for nonresponse in surveys. In J. C. Smart (Ed.), *Higher education: Handbook of theory and research* (Vol. XXII, pp. 411-449). Dordrecht, The Netherlands: Springer.
- Sharot, T. (1986). Weighting survey results. *Journal of the Market Research Society*, 28, 3, pp. 269-284

