

COMPLEMENTOS DE ESTATÍSTICA

LICENCIATURA EM GESTÃO DA QUALIDADE

2024/25

ficha de exercícios

introdução à amostragem

1. Indique, para cada caso, qual é a população e, desta, qual a amostra selecionada:
 - (a) Para avaliar a eficácia de uma campanha de promoção da reciclagem nos lares de uma cidade, 87 lares foram analisados sobre se fazem a separação do lixo;
 - (b) Para estudar o *share* de um programa de televisão em Portugal, alguns telespetadores foram entrevistados sobre a sintonia no respetivo canal no horário do programa;
 - (c) Foram realizadas entrevistas a 2300 eleitores com o objetivo de avaliar a intenção de voto nas próximas eleições legislativas em Portugal.
2. Dois investigadores *A* e *B* recolheram duas amostras separadamente de dimensões 100 e 200, respetivamente.

Pode dizer-se que a amostra recolhida pelo investigador *B* é mais representativa da população? Justifique a sua resposta.
3. Uma empresa de telecomunicações dispõe de uma base de dados com 250 dos seus clientes (*cellular.sav*) que foram classificados segundo a propensão para mudarem de operadora.
 - (a) Na impossibilidade de inquirir todos os clientes, a empresa pretende conduzir um estudo baseado numa amostra com 20% do universo.
 - i. Obtenha, a partir do spss, amostras aleatórias: simples e sistemática, utilizando o valor personalizado igual a 5.
 - ii. Para cada uma das amostras anteriores obtenha intervalos de confiança a 95% para a faturação mensal média e para a percentagem dos clientes com propensão ≥ 50 de mudar de operadora, e verifique se os IC contêm os parâmetros populacionais.
 - (b) A empresa pretende contactar os 10 clientes com maior propensão para mudar de operadora para evitar que este troquem a empresa. Esta opção traduz-se em que tipo de amostragem?

4. Considere os números inteiros positivos de 0 até 100 (AmostragemEx4.sav).
 - (a) Obtenha uma amostra de dimensão 40 através de uma amostragem (utilizando o valor personalizado igual a 20):
 - i. aleatória simples (sem reposição)
 - ii. aleatória simples com reposição
 - iii. sistemática
 - iv. estratificada, estratificando pela variável par/ímpar, com afetação proporcional $\frac{n}{N}$
 - v. por grupos, considerando os 10 grupos: $1 \rightarrow 10, \dots, 91 \rightarrow 100$
 - (b) Para cada uma das amostras anteriores calcule intervalos de confiança a 95% aproximados para:
 - i. a média
 - ii. a proporção de números pares
 - (c) Verifique se os intervalos de confiança obtidos contêm os parâmetros populacionais e analise criticamente os resultados.
5. A base de dados tree_car.sav contém o preço da compra do primeiro veículo preço de compra do veículo e algumas variáveis demográficas dos 3110 compradores de um determinado stand de venda de automóveis.
 - (a) Calcule a média e o desvio padrão amostrais do *preço de compra do veículo* para cada uma das categorias das variáveis *género*, *salário anual* (em m€), *habilitação académica* e *estado civil*.
 - (b) Qual ou quais das variáveis anteriores considera mais adequadas para serem consideradas para estratificar uma amostra aleatória?
 - (c) Determine a estratificação de uma amostra com 5% da população com as variáveis relevantes identificadas na alínea anterior.
6. Considere o ficheiro casas.sav com os dados de 40 registos de casas vendidas no último mês pela imobiliária ÁguedaCasa.
 - (a)
 - i. Obtenha uma amostra aleatória simples, sem reposição, com 15 imóveis de entre os 40 registados, através da posição 00 – 30, de cima para baixo, da tabela de números pseudo-aleatórios.
 - ii. Com a amostra obtida na alínea anterior, calcule um intervalo de confiança a 80% adequado para a área média, μ , de um imóvel registado na base de dados.
 - (b) A imobiliária ÁguedaCasa pretende entrevistar os novos proprietários com o objetivo de identificar potenciais fatores que levaram à aquisição da casa.

- i. Na impossibilidade de entrevistar todos os novos proprietários, a imobiliária pretende obter uma amostra aleatória, sem reposição, com 75% das 40 casas vendidas, para entrevistar os seus novos proprietários. Obtenha a referida amostra utilizando como valor personalizado no SPSS o número 31.
 - ii. Considerando a amostra da alínea anterior, calcule um intervalo de confiança a 90% para a percentagem, π , das casas vendidas nesse mês que não têm garagem e verifique se este intervalo contém o verdadeiro parâmetro populacional.
7. Uma rede de hotéis quer construir um intervalo de confiança a 99% para o número médio de quartos ocupados por noite em todas as unidades espalhadas pelo país. Qual deve ser a dimensão da amostra se pretendemos uma margem de erro máxima de 50 quartos. Tenha em atenção que numa amostra piloto obteve-se $s = 156$.
8. Qual a dimensão da amostra que um gestor deve recolher se desejar estimar o tempo médio de espera dos clientes ao balcão, sabendo que, de estudos anteriores, $s = 3$ minutos e que se pretende estimar o tempo médio com uma margem de erro não superior a meio minuto para uma confiança de 90%.
9. Pretende-se construir um intervalo de confiança a 99% para a proporção do clientes de uma loja que pagam as suas compras com cartão de crédito. Sabe-se que atualmente a loja tem registados 4600 clientes.
Quantos clientes devem ser analisados quanto ao meio de pagamento das compras de modo que esse intervalo não tenha uma amplitude superior a 4%?

Soluções

1. a) população: todos os lares da cidade, amostra: os 87 lares analisados; b) população: todas as pessoas que estavam a assistir TV na hora em que o programa foi transmitido, amostra: os telespetadores entrevistados no estudo; c) população: todos os eleitores em Portugal, amostra: os 2300 eleitores entrevistados
2. Não. Uma amostra de 100 elementos que representam significativamente a população é melhor que uma amostra de 200 elementos selecionada sem o emprego de procedimentos adequados. Para decidir qual amostra é mais representativa, é necessário conhecer na íntegra os procedimentos utilizados para a recolha das duas amostras, pois estes influenciam na qualidade da amostra.
3. a) os resultados apresentados foram obtidos com
coletar amostra: opções de seleção/valor personalizado igual a 5
 $N = 250, n \geq 30$, intervalos aproximados pelo TLC

amostragem aleatória simples

$$\mu = 64,08 \pm 1,96 \frac{20,961}{\sqrt{50}} \sqrt{\frac{250-50}{250-1}} = 64,08 \pm 5,21 \rightarrow I_{95\%}(\mu) = [58,87\text{€}, 69,29\text{€}]$$

$$p = 0,18 \pm 1,96 \sqrt{\frac{0,18(1-0,18)}{50}} \sqrt{\frac{250-50}{250-1}} = 0,1800 \pm 0,0367 \rightarrow I_{95\%}(p) = [8,46\%, 27,54\%]$$

$$\mu = 63,40\text{€} \in I_{95\%}(\mu) \text{ e } p = \frac{50}{250} = 20\% \in I_{95\%}(p)$$

amostragem sistemática

$$\mu = 61,94 \pm 1,96 \frac{18,72}{\sqrt{50}} \sqrt{\frac{250-50}{250-1}} = 61,94 \pm 4,65 \rightarrow I_{95\%}(\mu) = [57,29\text{€}, 66,59\text{€}]$$

$$p = 0,20 \pm 1,96 \frac{0,20(1-0,20)}{\sqrt{50}} \sqrt{\frac{250-50}{250-1}} = 0,2000 \pm 0,0397 \rightarrow I_{95\%}(p) = [10,06\%, 29,94\%]$$

$$\mu = 63,40\text{€} \in I_{95\%}(\mu) \text{ e } p = \frac{50}{250} = 20\% \in I_{95\%}(p)$$

b) amostragem intencional

4. b) i. *coletar amostra: opções de seleção/valor personalizado igual a 20*

amostragem aleatória simples, sem reposição

$$\mu = 56,58 \pm 1,96 \frac{24,866}{\sqrt{40}} \sqrt{\frac{100-40}{100-1}} = 56,58 \pm 6,00 \rightarrow I_{95\%}(\mu) = [50,58; 62,58]$$

amostragem aleatória simples, com reposição

$$\mu = 58,68 \pm 1,96 \frac{30,000}{\sqrt{40}} = 58,68 \pm 9,30 \rightarrow I_{95\%}(\mu) = [49,38; 67,98]$$

amostragem aleatória sistemática (IC aprox. sem reposição)

$$\mu = 50,50 \pm 1,96 \frac{29,216}{\sqrt{40}} = 50,50 \pm 9,05 \rightarrow I_{95\%}(\mu) = [41,45; 59,55]$$

amostragem aleatória estratificada, com afetação proporcional

– cálculo da variância da média amostral considerando a estratificação, sem reposição

$$\hat{\sigma}_{\bar{X}_{est}}^2 = s_{\bar{X}_{est}}^2 = \left(\frac{50}{100}\right)^2 \frac{576,042}{20} \frac{50-20}{50-1} + \left(\frac{50}{100}\right)^2 \frac{644,042}{20} \frac{50-20}{50-1} = 4,408 + 4,929$$

$$= 9,337$$

$$\mu = 59,00 \pm 1,96\sqrt{9,337} = 59,00 \pm 5,989 \rightarrow I_{95\%}(\mu) = [53,011; 64,989]$$

– cálculo da variância da média amostral considerando a uma a.a.s. (sem reposição)

$$s^2 = 624,308, \hat{\sigma}_{\bar{x}}^2 = \frac{624,308}{40} \frac{100 - 40}{100 - 1} = 9,459,$$

$$\mu = 59,00 \pm 1,96\sqrt{9,459} = 59,00 \pm 6,028 \rightarrow I_{95\%}(\mu) = [52,972; 65,028]$$

amostragem aleatória por clusters (IC aprox. através a.a.s., sem reposição)

clusters selecionados; 7,8,9,10

$$\mu = 80,50 \pm 1,96 \frac{11,690}{\sqrt{40}} \sqrt{\frac{100 - 40}{100 - 1}} = 80,50 \pm 2,820 \rightarrow I_{95\%}(\mu) = [77,68; 83,32]$$

5. a) obter no SPSS (*analisar/explorar*)

b) as variáveis que apresentam preços de compra dos veículos mais desiguais dentro das categorias são o salário médio anual e as habilitações académicas. Não há diferenças significativas entre masculino/feminino e casado/não casado (se necessário pode confirmar-se pela aplicação do teste t para a comparação de 2 médias de amostras independentes).

c) No SPSS *analisar/estatísticas descritivas/tabela de referência cruzada* obtém-se o número de elementos da população na estratificação:

			Income category in thousands				Total
			Under \$25	\$25 – \$49	\$50 – \$74	\$75+	
Level of education	Did not complete high school	Contagem	173	270	106	133	682
		% do Total	5,6%	8,7%	3,4%	4,3%	21,9%
	High school degree	Contagem	189	353	177	254	973
		% do Total	6,1%	11,4%	5,7%	8,2%	31,3%
	Some college	Contagem	113	238	118	174	643
		% do Total	3,6%	7,7%	3,8%	5,6%	20,7%
	College degree	Contagem	99	213	131	193	636
		% do Total	3,2%	6,8%	4,2%	6,2%	20,5%
	Post-undergraduate degree	Contagem	15	56	35	70	176
		% do Total	0,5%	1,8%	1,1%	2,3%	5,7%
Total	Contagem	589	1130	567	824	3110	
	% do Total	18,9%	36,3%	18,2%	26,5%	100,0%	

Como se pretende uma fração de amostragem de 5%, a amostra estratificada tem de ser constituída da seguinte forma

		salário anual			
		$\leq 25m\text{€}$	$25m\text{€} - 49m\text{€}$	$50m\text{€} - 74m\text{€}$	$\geq 75m\text{€}$
hab.	básico	$173 \times 5\% \approx 9$	$270 \times 5\% \approx 14$	$106 \times 5\% \approx 5$	$133 \times 5\% \approx 7$
acad.	secund.	$189 \times 5\% \approx 9$	$353 \times 5\% \approx 18$	$177 \times 5\% \approx 9$	$254 \times 5\% \approx 13$
	freq. sup.	$113 \times 5\% \approx 6$	$238 \times 5\% \approx 12$	$118 \times 5\% \approx 6$	$174 \times 5\% \approx 9$
	licenc.	$99 \times 5\% \approx 5$	$213 \times 5\% \approx 11$	$131 \times 5\% \approx 7$	$193 \times 5\% \approx 10$
	pós-grad.	$15 \times 5\% \approx 1$	$56 \times 5\% \approx 3$	$35 \times 5\% \approx 2$	$70 \times 5\% \approx 4$

Note que pretendia-se uma amostra com $n = 3110 \times 5\% \approx 156$, mas após a estratificação e arredondamentos temos $n = 160$, isto é, uma fração amostral real de 5,14%.

6. a)

i. registos {14, 18, 01, 21, 40, 30, 16, 05, ~~05~~, ~~05~~, 36, 25, ~~18~~, 27, 08, 29, 07, 26}

ii. $N = 40$, $n = 15$, população finita e amostragem sem reposição \rightarrow considerar fator de correção
Como $n < 30$, temos de admitir que a variável área é normalmente distribuída e usamos a t de Student.

O IC obtido no SPSS considera o quantil da distribuição t de Student, logo temos que a margem de erro do IC é $\frac{104,4679-93,8441}{2} = 5,312$. O fator de correção é $FC = \sqrt{\frac{40-15}{40-1}} = 0,8006$, logo

$$\begin{aligned}\mu_{\text{área}} &= 99,156 \pm 5,312 \times 0,8006 \\ &= 99,156 \pm 4,253 \\ I_{80\%}(\mu_{\text{área}}) &= [94,90\text{m}^2; 103,41\text{m}^2]\end{aligned}$$

Descriptives

		Statistic	Std. Error
Area	Mean	99,1560	3,94930
	80% Confidence Interval for Mean	Lower Bound	93,8441
		Upper Bound	104,4679
	5% Trimmed Mean	98,2128	
	Median	96,8100	
	Variance	233,955	
	Std. Deviation	15,29557	
	Minimum	76,59	
	Maximum	138,70	
	Range	62,11	
	Interquartile Range	8,47	
	Skewness	1,204	,580
	Kurtosis	2,375	1,121

Sendo $t_{0,90;14} = IDF.T(0.90, 14) = 1,345$, podemos fazer o cálculo

$$\mu_{\text{área}} = 99,156 \pm 1,345 \times \frac{15,296}{\sqrt{15}} \times \sqrt{\frac{40-15}{40-1}}$$

b)

i. registos {1-2; 4; 6; 8-12; 15; 17; 19-31; 33-37; 40}

Método

Tipo:
Amostragem aleatória simples

☒ Sem reposição (SS)
☐ Com reposição (SS)
☐ Usar estimação CS para análise

Unidades:
Proporções

☒ Valor:

0,75

Que tipo de valor semente deseja usar?

☐ Um número escolhido aleatoriamente
☒ Valor customizado:

31

Insira um valor semente customizada se desejar reproduzir a amostra mais tarde.

ii. $\hat{\pi}_{\text{sem garagem}} = \frac{21}{30} = 0,70$

$N = 40, n = 30$, população finita e amostragem sem reposição \rightarrow usar fator de correção

$$\begin{aligned}\pi_{\text{sem garagem}} &= 0,70 \pm 1,645 \times \sqrt{\frac{0,70 \cdot (1 - 0,70)}{30}} \times \sqrt{\frac{40 - 30}{40 - 1}} \\ &= 0,70 \pm 0,07 \\ I_{90\%}(\pi_{\text{sem garagem}}) &= [63,0\%, 77,0\%]\end{aligned}$$

7. $n \geq 65$

8. $n \geq 98$

9. $n \geq 2182$