

Gabarito - Lista 4

Mensuração e Estatística Descritiva

Gabarito

Exercício 1 [3 pontos]. A tabela abaixo mostra o número de habitantes por km² em alguns países do Leste Europeu. Calcule as seguintes estatísticas e mostre todas as etapas dos cálculos, nos casos em que eles forem necessários (essa é a última vez que vamos pedir para você fazer isso na mão).

- a Média
- b Mediana
- c Moda
- d Variância
- e Desvio-padrão

Country	Density
Poland	123
Czech Rep.	132
Croatia	85
Bosnia	87
Romania	96
Russia	8.8
Estonia	36
Latvia	40
Lithuania	58
Ukraine	87
Belarus	50
Georgia	81

a

$$\bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{123 + 132 + 85 + 87 + 96 + 8.8 + 36 + 40 + 58 + 87 + 50 + 81}{12} = 73,65$$

b

$$8.8 \ 36 \ 40 \ 50 \ 58 \ 81 \ || \ 85 \ 87 \ 87 \ 96 \ 123 \ 132 \Rightarrow Med(y) = \frac{81 + 85}{2} = 83$$

c 87

d Para ajudar a fazer o cálculo da variância, sugerimos que se construa a seguinte tabela:

country	density	$y_i - \bar{y}$	$(y_i - \bar{y})^2$
Poland	123	49,35	2435,42
CzechRep.	132	58,35	3404,72
Croatia	85	11,35	128,82
Bosnia	87	13,35	178,22
Romania	96	22,35	499,52
Russia	8,8	-64,85	4205,52
Estonia	36	-37,65	1417,52
Latvia	40	-33,65	1132,32
Lithuania	58	-15,65	244,92
Ukraine	87	13,35	178,22
Belarus	50	-23,65	559,32
Georgia	81	7,35	54,02
Soma	883,8	-	14438,57
n	12	-	-
Média	73,65	-	-

Em seguida, podemos utilizar a soma do quadrado dos desvios para completar a fórmula da variância:

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^n (y_i - \bar{y})^2 = \frac{14438,57}{11} = 1312,6$$

e $\sigma = \sqrt{1312,6} = 36,23$

Exercício 2 [2 pontos]. Referindo-se aos conceitos de média e desvio-padrão, Agresti e Finlay ora utilizam caracteres romanos (\bar{y} e s , respectivamente), ora utilizam caracteres gregos (μ e σ , respectivamente). Por que existe essa diferença? Explique em que casos são utilizadas cada uma das notações.

Agresti e Finlay utilizam uma notação comum nos livros-texto de introdução à Estatística: usar caracteres romanos para se referir a amostras e caracteres gregos para se referir a parâmetros populacionais. Assim, \bar{y} se refere à média amostral, enquanto μ se refere à média populacional.

Vale notar que, de modo geral, os livros abrem uma exceção a essa regra e utilizam p para a proporção populacional. Isso acontece porque o equivalente grego seria π , e ficaria muito esquisito calcular uma proporção populacional e concluir, por exemplo, que $\pi = 0,73$.

Exercício 3 [2 pontos]. Um colega de departamento conseguiu um financiamento para fazer uma *survey* com amostra aleatória simples, de maneira que todos os indivíduos da população de interesse têm a mesma chance de participar da pesquisa. Assim, diz ele, as estimativas de sua pesquisa não sofrerão problemas erro amostral. Você concorda com ele? Discuta.

O pesquisador está com uma concepção equivocada do que significa erro amostral. Erro amostral é a variação esperada da média (ou de qualquer outra estatística) quando repetimos a pesquisa diversas vezes. O erro amostral não está associado, a princípio, a falhas de amostragem ou de desenho de pesquisa, embora possa ser amplificado por esses fatores. Em outras palavras: mesmo uma pesquisa com desenho e amostragem perfeitos terá erro amostral.

Exercício 4 [3 pontos]. A imagem abaixo mostra as estatísticas descritivas básicas para a variável *gdp_cap*, disponível no banco de dados *world95.dta*. A variável é medida em dólares, de maneira que o PIB *per capita* médio dos países do banco é de \$5833.33.

. sum gdp_cap					
Variable	Obs	Mean	Std. Dev.	Min	Max
gdp_cap	102	5883.333	6603.27	122	23474

- Para convertermos o PIB *per capita* dos países para Reais, teríamos que multiplicar todos os valores pela taxa de câmbio. Se fizéssemos isso, a média da variável seria a mesma? E o desvio-padrão? Explique.
- O que acontece com a média e o desvio-padrão da variável PIB per capita se adicionarmos 100 dólares em todos os países?

Dica: Existem várias maneiras de responder esse exercício. Você pode apresentar uma resposta analítica, um exemplo com os cálculos feitos à mão ou uma simulação no Stata. Qualquer estratégia será aceita, desde que produza os resultados corretos.

- Ao multiplicar uma variável de média \bar{y} e desvio-padrão s por um valor v qualquer, a média passa a ser $v\bar{y}$ e o desvio-padrão passa a ser $|v|s$. Assim, supondo uma taxa de câmbio de R\$3,59, a nova média seria de 21.121,15, e o novo desvio-padrão seria de 23.705,74.
- Ao adicionar uma constante v a uma variável de média \bar{y} e desvio-padrão s , a média passa a ser $\bar{y} + v$, e o desvio-padrão não se altera (pense que você apenas está deslocando a distribuição à direita no eixo x). Assim, a média passaria a ser 5.983,33, e o desvio-padrão permaneceria 6603,27.