

U.C. 21044
Estatística para as Ciências Sociais

8 de setembro de 2017

- INSTRUÇÕES -

- A prova é composta por **3** grupos de questões e respetivas alíneas, contém 4 página(s) e termina com a palavra **FIM**.
- Verifique o seu exemplar e, caso encontre alguma anomalia, dirija-se ao professor vigilante nos primeiros 15 minutos da prova, pois qualquer reclamação sobre defeito(s) de formatação e/ou de impressão não será aceite depois deste período.
- Todas as questões deverão ser **respondidas e justificadas** na folha de ponto. Todos os cabeçalhos e espaços reservados à identificação, deverão ser preenchidos com letra legível. Utilize unicamente tinta azul ou preta.
- É permitido o uso de máquina de calcular. É fornecido um formulário na segunda página deste enunciado.
- O aluno pode arbitrar um valor adequado da resposta a uma alínea que não tenha respondido, caso este seja necessário para a resolução de uma alínea posterior.
- Verifique no momento da entrega da(s) folha(s) de ponto se todas as páginas estão rubricadas pelo vigilante. Caso necessite de mais do que uma folha de ponto, deverá numerá-las no canto superior direito.
- Não serão aceites folhas de ponto dobradas ou danificadas. Exclui-se, para efeitos de classificação, toda e qualquer resposta apresentada em folhas de rascunho.
- Tenha em atenção que a prova tem a duração máxima de **2 horas e 30 minutos**.

COTAÇÃO E CRITÉRIOS DE CORREÇÃO:

- Clareza, correção, completude e justificação das respostas;
- Apresentação de fórmulas, tabelas e cálculos auxiliares;

COTAÇÕES																	
1. (7 valores)						2. (9 valores)						3. (4 valores)					
a)	b)	c)	d)	e)	f)	a)	b)	c)	d)	e)	f)	g)	h)	a)	b)	c)	d)
1.5	0.5	0.5	2	1.5	1	1	1	1	2	1	1	1	1	0.5	1	1.5	1

FORMULÁRIO

ESTATÍSTICA DESCRITIVA

População de dimensão N

$$\mu = \bar{X} = \frac{1}{N} \sum_{i=1}^N X_i = \frac{X_1 + X_2 + \dots + X_N}{N}; \quad \sigma = \sqrt{s^{*2}}; \quad CV = \frac{\sigma}{\mu}$$

$$s^{*2} = \sigma^2 = \frac{1}{N} \sum_{i=1}^N (X_i - \bar{X})^2 = \frac{(X_1 - \bar{X})^2 + (X_2 - \bar{X})^2 + \dots + (X_N - \bar{X})^2}{N};$$

Amostra de dimensão n

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i = \frac{x_1 + x_2 + \dots + x_n}{n}; \quad s = \sqrt{s^2}; \quad CV = \frac{s}{\bar{x}}$$

$$s^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{n-1};$$

Amostra de dim. n com dados agrupados em k classes de centro c_i com frequência n_i

$$\bar{x} \approx \frac{1}{n} \sum_{i=1}^k n_i c_i = \frac{n_1 c_1 + n_2 c_2 + \dots + n_k c_k}{n}; \quad s = \sqrt{s^2}; \quad CV = \frac{s}{\bar{x}}$$

$$s^2 \approx \frac{1}{n-1} \sum_{i=1}^k n_i (c_i - \bar{x})^2 = \frac{n_1 (c_1 - \bar{x})^2 + n_2 (c_2 - \bar{x})^2 + \dots + n_k (c_k - \bar{x})^2}{n-1};$$

$$\text{Mediana} = \begin{cases} x_{(\frac{n+1}{2})}, & n \text{ é ímpar}; \\ \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2}, & n \text{ é par}; \end{cases} \quad \begin{array}{l} Q_1 = \text{Mediana da 1ª sub-amostra}; \\ Q_3 = \text{Mediana da 2ª sub-amostra}; \end{array}$$

Amplitude = $\max(x_i) - \min(x_i)$; AIQ = $Q_3 - Q_1$;

Assimetria e Curtose

Grau de Assimetria: $G = \frac{3(\mu - Me)}{\sigma}$; Grau de curtose: $K = \frac{Q_3 - Q_1}{2(P_{90} - P_{10})}$

Coeficiente de Assimetria de Pearson: $G_1 = \frac{(\mu - Mo)}{\sigma}$; $G_2 = \frac{(Q_3 + Q_1 - 2Me)}{Q_3 - Q_1}$

REGRESSÃO E CORRELAÇÃO

$$R = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{(n-1)s_x s_y} = \frac{(x_1 - \bar{x})(y_1 - \bar{y}) + (x_2 - \bar{x})(y_2 - \bar{y}) + \dots + (x_n - \bar{x})(y_n - \bar{y})}{(n-1)s_x s_y}$$

Regressão linear: $\hat{y} = a + bx$; $R^2 \equiv$ coeficiente de determinação; $R_{Spearman} = 1 - \frac{6 \sum D_i^2}{n(n^2 - 1)}$

PROBABILIDADES

$$P(A) = \frac{\text{no. de casos de A}}{\text{no. de casos possíveis}}; \quad P(\bar{A}) = 1 - P(A); \quad P(A|B) = \frac{P(A \cap B)}{P(B)};$$

$$A \subset B \Rightarrow P(A) \leq P(B); \quad P(A - B) = P(A) - P(A \cap B);$$

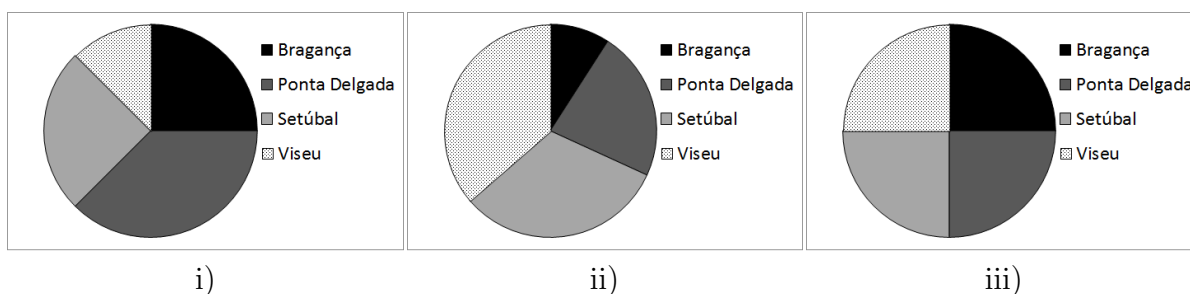
$$P(A \cup B) = P(A) + P(B) - P(A \cap B);$$

$$A \text{ e } B \text{ independentes} \Leftrightarrow P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A) \cdot P(B);$$

1. Para aferir a distribuição geográfica de famílias pobres em Portugal, recolheu-se o distrito de residência de 8 agregados familiares abaixo do limiar de pobreza, escolhidos aleatoriamente de uma base de dados da Santa Casa da Misericórdia, conforme os dados seguintes seguinte:

Ponta Delgada	Bragança	Ponta Delgada	Bragança
Ponta Delgada	Viseu	Setúbal	Setúbal

- (a) Indique qual a população, amostra e variável em estudo.
 (b) Indique se a amostra é adequada.
 (c) Classifique, justificando, a variável em estudo.
 (d) Determine, se possível, a tabela de frequências absolutas e relativas (simples e acumuladas) dessa variável.
 (e) Indique, justificando, qual dos 3 gráficos seguintes corresponde à distribuição da variável:



- (f) Determine, se possível, a média, moda e mediana dessa variável.
2. Pretende-se saber qual a relação entre a participação na plataforma e a nota final, para alunos de Estatística para as Ciências Sociais. Recolheram-se assim, o tempo médio semanal (em minutos) de participação na plataforma e a nota final para 8 alunos escolhidos ao acaso:

Tempo de participação (x_i)	10	80	150	20	100	178	240	60
Nota (y_i)	12	14	18	8	15	16	17	13

Sabe-se ainda que a média e a variância da variável tempo de participação e a variância da variável nota são dados por

$$\bar{x} = 104.75, \quad s_x^2 = 6357.64, \quad s_y^2 = 10.125,$$

que a reta de regressão linear que relaciona as duas variáveis é dada por

$$\hat{y} = 10.662 + 0.0331x$$

e que o coeficiente de correlação é dado por $R = 0.828$.

(V.s.s.f)

- (a) Classifique, justificando, as duas variáveis em estudo.
 - (b) Calcule, se possível, a média e mediana da variável nota.
 - (c) Calcule os quartis da variável nota.
 - (d) Indique qual das variáveis tem maior dispersão relativa.
 - (e) Indique, justificando, qual a variação prevista na nota pela reta de regressão linear quando o tempo de participação semanal decresce 30 minutos.
 - (f) Comente, justificando, a seguinte afirmação: "A nota média prevista para um estudante que participa 150 minutos semanais é 18."
 - (g) Comente a correlação entre as variáveis.
 - (h) Comente, justificando, a seguinte afirmação: "A variação da nota é explicada em cerca de 82.8% pela variação linear do tempo de participação."
3. Dos membros de uma associação, 80% têm pelo menos 65 anos e 60% são homens. Sabe-se que 40% dos membros são homens com pelo menos 65 anos.
- (a) Indique se os acontecimentos ser homem e ter pelo menos 65 anos são independentes.
 - (b) De entre os homens, qual a percentagem com pelo menos 65 anos?
 - (c) Esboce o diagrama de Venn respetivo, indicando a legenda da figura.
 - (d) Indique a proporção de mulheres com menos de 65 anos.

FIM