Elementos de Probabilidades e Estatística (21037)
Ano letivo 2016-17
E-Fólio A – 17 a 26 de abril 2017
Critérios de correção e orientações de resposta

No presente relatório apresentam-se os critérios de correção e um exemplo de resolução das questões colocadas no e-fólio A. Também são salientados alguns pontos que os estudantes devem ter em conta e apresentar na resolução e esclarecimentos sobre os erros mais comuns.
Foram aceites outras resoluções entregues pelos alunos, desde que equivalentes (com raciocínio e cálculos corretos).

Chama-se a atenção de que o e-fólio **é um teste exclusivamente individual**, feito sob compromisso de honra do estudante ser o **seu único autor**. Os casos que tenham evidenciado existência de fraude podem ser penalizados ou ficar sob escrutínio mais apertado nas próximas avaliações.

Critérios de Avaliação e Cotação

- Na avaliação das respostas será tido em conta: a clareza e objetividade, a correção científica das respostas, o rigor da linguagem matemática/estatística que é utilizada, a evidência de pesquisa, as justificações dos cálculos e a interpretação dos resultados no contexto do problema.
- A cotação do e-Fólio A é de **4 valores**, assim distribuídos:

Questão 1 – 1.5 valores Questão 2 – 1.5 valores Questão 3 – 1 valor

Questão 1 (1.5 val)

Os dados do quadro que se segue representam o tempo (em segundos) que uma aplicação informática direcionada para os Sistemas de Informação Geográfica (SIG) demora a responder a um conjunto de instruções mais complexas. Os dados são relativos uma amostra 40 computadores portáteis.

5,2	6,4	5,7	8,3	7,0	5,4	4,8	9,1	5,5	6,2
5,9	5,7	6,3	5,1	7,4	6,2	8,9	7,3	5,4	8,8
5,2	6,8	5,1	6,7	8,2	7,1	4,3	5,0	8,2	5,5
7,1	5,8	6,3	10,2	7,5	9,6	5,2	7,8	5,6	8,1

Com base neste dados, indique, justificando:

a) **(0.3 v)** a escala de medida dos dados, o máximo, o mínimo e a amplitude dos tempos de resposta do sistema.

Para uma resolução completa é esperado:

- ✓ indicar a variável considerada e classificação do tipo de variável, no âmbito da classificação Matemática/Estatística.

  Sendo os dados relativos a medições de tempo, a variável subjacente é contínua (toma valores num intervalo de números reais), numa escala de razão (rácios comparáveis, existência de zero absoluto).

- ✓ Indicar o conceito e cálculos das estatísticas solicitadas (o máximo, o mínimo e a amplitude).

✓ Apresentar os valores finais corretos.

b) **(0.4 v)** a média, a moda e a mediana dos tempos de resposta. Como classifica os dados, em relação à assimetria?

Para uma resolução completa é esperado:

✓ indicar o conceito e expressões das estatísticas amostrais pedidas.

✓ Fazer a substituição com os dados do enunciado, apresentar alguns cálculos parciais (mesmo aquando da utilização do Excel ou de calculadora).

✓ Concluir sobre a assimetria dos dados, indicando o critério utilizado.

Nota: a situação em que o aluno indica a classe modal depois de agrupar os dados em classes também foi considerada embora não responda exatamente ao solicitado.

.**Média:** $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{\sum_{i=1}^{n} x_i}{n}$, ou seja, $\bar{x} = \frac{\sum_{i=1}^{40} x_i}{40} = \frac{5{,}2 + 6{,}4 + ... + 8{,}1}{40}$.

Moda: Valor de X (tempos de resposta) mais frequente (definição), quando este valor existe e é único (a distribuição pode ser amodal – sem moda – p.ex. dist uniforme -; bimodal – duas modas, etcpor exemplo)

Mediana: Indicar a definição

Quartil de 50%. Se ***n*** é par, a mediana m_e é a média aritmética das observações se encontram nas duas posições centrais, $\frac{n}{2}$ $\frac{(n+1)}{2}$, estando a amostra ordenada por ordem crescente.

Se n é ímpar a mediana m_e é a observação x se encontra na posição $\frac{(n+1)}{2}$, estando a amostra ordenada por ordem crescente.

Sendo n par, ordenando a amostra ordenada, a mediana é a média aritmética das observações das posições 20 e 21 $m_e = \frac{6{,}3 + 6{,}3}{2} = 6{,}3$ segundos.

Moda < Mediana < Média → assimetria positiva – prolongamento da distribuição na cauda direita.

Resumo dos resultados

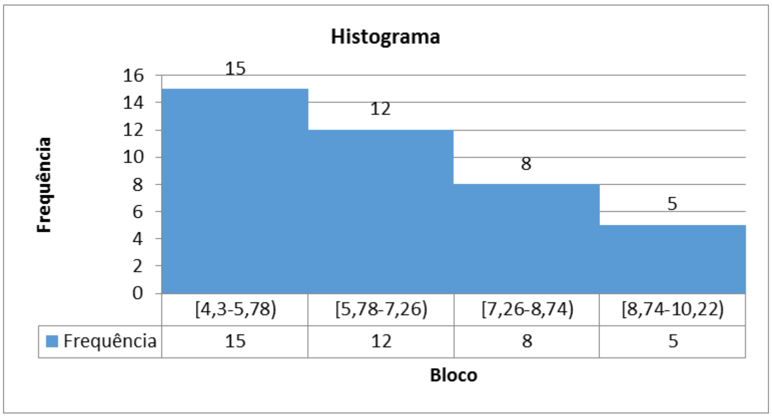
Máximo	10,2
Mínimo	4,3
Amplitude	5,9
Moda	5,2
Mediana	6,3
Media	6,64
variancia	2,091
desvio-padrão	1,446

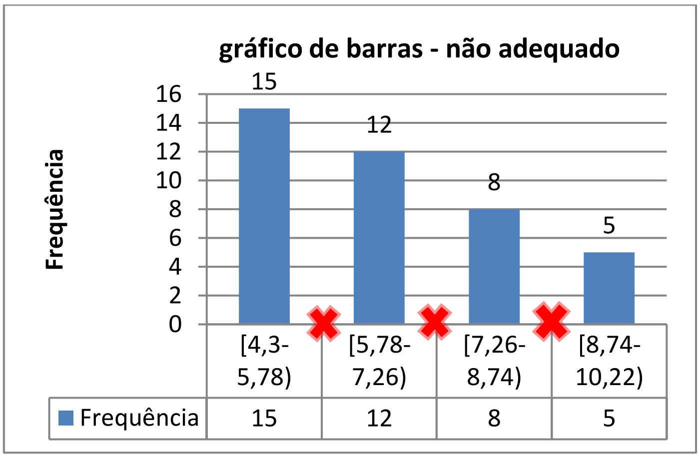
c) **(0.8 v)** Elabore um gráfico que entenda ser o mais adequado para representar as frequências dos tempos de resposta em 4 categorias fixadas (para as quais deve inclusive definir os limites mínimo e máximo, justificando).

Para uma resolução completa é esperado:

- ✓ Fixando o número de classes pedido (4), indicar a decisão sobre a definição das mesmas atendendo às condições para melhor representação: mesma amplitude (caso contrário exigiria uma uniformização das frequências), os limites devem garantir que todos os valores observados pertencem a uma classe.
- ✓ As classes devem ser abertas num limite (inferior ou superior) e fechadas no outro (mutuamente exclusivas e exaustivas).
- ✓ O gráfico mais adequado a variáveis numa escala contínua (escala subjacente é o tempo) é o histograma (classes contíguas).

No exemplo de resolução a seguir, a amplitude de classe é o quociente Amplitude/nº de classes 5,9/4. A primeira classe inicia no mínimo da amostra (inclusive). Considerou-se um arredondamento a 2 casas decimais.
Outros critérios foram aceites desde que válidos, p. ex: Verificar que o máximo e mínimo estão contidos nos intervalos. Para tal o aluno pode subtrair uma quantidade para o mínimo ficar no interior da classe; colocar o mínimo no centro da classe e a partir daí somar a amplitude. O gráfico adequado é o histograma (preferencialmente de frequências relativas), pois o tempo é uma variável contínua!
As frequências (absolutas e relativas) podem variar ligeiramente com a decisão sobre os limites e amplitude das classes, contudo a assimetria positiva deve ser evidente no histograma (e ir de encontro aos resultados da alínea anterior).





Questão 2 (1.5 val)

No quadro que se segue encontram-se registadas as vendas de manuais de Matemática para 12º ano de escolaridade, registadas numa semana em 10 livrarias de rua (junto a escolas, etc).

Nº da livraria	1	2	3	4	5	6	7	8	9	10
Vendas (unidades)	9	12	6	11	a	6	7	13	8	2a

a) **(0.75 v)** Sabe-se que a livraria nº 5 atingiu metade das vendas da livraria nº 10, e que a média do número de livros vendidos é de 9 livros por livraria. Com estes dados, complete o quadro com as frequências em falta calcule ainda o desvio-padrão das vendas dessa semana. Justifique (se for necessário, pode efetuar pequenos arredondamentos).

Para uma resolução completa é esperado:

- ✓ Formular o problema em termos estatísticos/matemáticos com os dados do enunciado, resolver a equação e completar o quadro.
- ✓ Apresentar a expressão usada para o desvio-padrão, substituir com os dados e indicar o resultado.
- ✓ Interpretar/concluir no contexto.

$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i \Rightarrow 9 = \frac{1}{10}\sum_{i=1}^{10} x_i = \frac{1}{10}(9+12+\ldots+a+\ldots.+8+2a)$, donde a=6.

A livraria nº5 vendeu 6 livros e a livraria nº10 vendeu 12 livros, na mesma semana.

$s = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(x_i-\bar{x})^2} \cong 2{,}65$ o desvio padrão das vendas nas 10 livrarias é de 2,65 livros.

b) **(0.75 v)** As livrarias nº 3, 7 e 9 vendem livros na modalidade online, através do seu site. Sabe-se que a probabilidade destas livrarias realizarem diariamente **apenas uma** venda é $\frac{1}{5}$, $\frac{1}{7}$ e $\frac{1}{8}$ respetivamente. Suponha que as vendas das 3 livrarias são independentes. Calcule a probabilidade de, num dia escolhido ao acaso, pelo menos uma das três livrarias ter realizado apenas uma venda. Explicite o problema em causa na notação dos acontecimentos e justifique os resultados.

Para uma resolução completa é esperado:

- ✓ Definir os acontecimentos elementares (individuais)
- ✓ Definir o acontecimento de interesse e formular a probabilidade pedida na linguagem dos acontecimentos e das probabilidades.
- ✓ Indicar a propriedade de independência entre acontecimentos, necessária para a resolução do problema.
- ✓ Substituir valores, calcular a probabilidade pedida e apresentar os resultados.

Foram aceites resoluções equivalentes que utilizaram notação diferente.

Definir os acontecimentos

A_3 – A livraria 3 realiza apenas uma venda online no dia selecionado.
A_7 – A livraria 7 realiza apenas uma venda online no dia selecionado.
A_9 – A livraria 9 realiza apenas uma venda online no dia selecionado.

Sabe-se que as probabilidades de ocorrência destes acontecimentos são, respetivamente, $P(A_3) = \frac{1}{5}$; $P(A_7) = \frac{1}{7}$; $P(A_9) = \frac{1}{8}$

Sabe-se ainda que as vendas das 3 livrarias são independentes, ou seja, verificam a igualdade: $P(A_i \cap A_j) = P(A_i) \times P(A_j)$ e $P(A_i \cap A_j \cap A_k) = P(A_i) \times P(A_j) \times P(A_k)$

O acontecimento "*num dia escolhido ao acaso, pelo menos uma das três livrarias ter realizado apenas uma venda*" traduz-se, utilizando as operações entre acontecimentos, na União dos três acontecimentos - $(A_i \cup A_j \cup A_k)$. O acontecimento "*pelo menos uma das três livrarias*.... " (*isto é, no mínimo uma livraria*...), inclui as seguintes possibilidades:

- Realiza-se uma venda em apenas uma das livrarias (A_3, ou A_7 ou A_9), OU ainda
- Realiza-se apenas uma venda em duas das livrarias (3 possibilidades de combinação), OU ainda
- Realiza-se apenas uma venda nas três livrarias.

Basta verificar-se uma das possibilidades, para que a União ocorra (seja verdadeira).
Por definição

$$P(A_3 \cup A_7 \cup A_9) = P(A_3) + P(A_7) + P(A_9) - P(A_3 \cap A_7) - P(A_3 \cap A_9) - P(A_7 \cap A_9) + P(A_3 \cap A_7 \cap A_9)$$

Aplicando a independência, tem-se

$$P(A_3 \cup A_7 \cup A_9) = \frac{1}{5} + \frac{1}{7} + \frac{1}{8} + \frac{1}{5} \times \frac{1}{7} + \frac{1}{5} \times \frac{1}{8} + \frac{1}{7} \times \frac{1}{8} - \frac{1}{5} \times \frac{1}{7} \times \frac{1}{8} = \frac{112}{280} = 0{,}4$$

A probabilidade de pelo menos uma livraria realizar apenas uma venda no dia escolhido é 0.4 (40%).

Questão 3 (1 val)

Uma Companhia de Seguros especializada no ramo de acidentes de trabalho divide as suas clientes (também empresas) por três grandes categorias, **A**, **B**, **C**, de acordo com o número de trabalhadores. A carteira de clientes da Seguradora está distribuída da seguinte forma: 30% das empresas seguradas pertencem à categoria **A**, 43% pertencem à categoria **B** e as restantes estão na categoria **C**.
A seguradora possui estimativas para a probabilidade das empresas comunicarem ocorrências de acidente de trabalho em cada período de seis meses, que são, respetivamente, 15% (para a categoria A), 8% (categoria B) e 10% (categoria C).
Nos dois semestres mais recentes foi comunicada apenas uma ocorrência à Companhia Seguradora. Determine a probabilidade de ambas terem origem na mesma categoria de empresas. Assuma os pressupostos que ache necessários e justifique os resultados.

Para uma resolução completa é esperado:

- ✓ Definir os acontecimentos elementares (individuais) e os acontecimentos condicionais
- ✓ Recorrer à probabilidade total (analiticamente ou diagrame de árvore) e à fórmula de Bayes.
- ✓ Formular o acontecimento final que é pedido e utilizar os conhecimentos/pressupostos necessários (independência, acontecimentos mutuamente exclusivos).
- ✓ Substituir valores, calcular a probabilidade pedida e apresentar os resultados.

Dados do enunciado:
Considere-se os seguintes acontecimentos identificados no enunciado:

A – Empresa segurada pertence à categoria A
B – Empresa segurada pertence à categoria B
C – Empresa segurada pertence à categoria C
O – Ocorrência de acidente de trabalho num semestre.

P(A)=0,3 (30%); P(B)=0,43 (43%)

Os acontecimentosformam uma partição do espaço dos clientes da Companhia de seguros (são não vazios, são disjuntos 2 a 2 e a sua união é o universo das categorias). Então, P(A)+P(B)+P(C) = 1, e daqui se calcula P(C)=1-0,3-0,43=0,27.

No enunciado é indicada a probabilidade de ser comunicada ocorrência de acidente de trabalho para cada uma das categorias de empresa, ou seja, a probabilidade de ser comunicada ocorrência de acidente de trabalho, **dada cada uma das categorias** (a categoria condiciona a probabilidade de comunicar a ocorrência). Assim, tem-se:

$P(O|A) = 0{,}15 \quad P(O|B) = 0{,}08 \quad P(O|C) = 0{,}10$

(nota: alguns alunos interpretaram incorretamente estas probabilidades condicionadas como probabilidade de intersecção $P(A \cap B)$, etc.).

Utilizando o Teorema da Probabilidade Total podemos determinar a probabilidade total de ser comunicada ocorrência (ou 1 ocorrência), fazendo as respetivas substituições:

$$P(O) = P(O|A)P(A) + P(O|B)P(B) + P(O|C)P(C) = 0{,}15 \times 0{,}3 + 0{,}08 \times 0{,}43 + 0{,}10 \times 0{,}27$$

$$= 0{,}045 + 0{,}0344 + 0{,}027 = 0.1064$$

Com a presente informação, e tendo sido comunicada uma ocorrência (ou, ***sabendo que*** foi comunicada uma ocorrência), podemos determinar a probabilidade de ter tido origem em cada uma das categorias de empresa, A, B e C. Para tal, recorremos ao teorema/fórmula de Bayes:

$$P(A|O) = \frac{P(A \cap O)}{P(O)} = \frac{P(O|A)P(A)}{P(O)} = \frac{0{,}045}{0.1064} = 0.423 \ .$$

E pelo mesmo raciocínio obtém-se $P(B|O) = \frac{0{,}0344}{0.1064} = 0.323$ e $P(C|O) = \frac{0{,}027}{0.1064} = 0.254$.

Assumindo que a comunicação de ocorrências por uma empresa é independente da comunicação por parte de outra empresa, e independente de semestre para semestre podemos determinar a probabilidade final que é pedida

Formulando (foram tidas em conta outras interpretações apresentadas, quando fez sentido):
Foi comunicada uma ocorrência em cada um dos semestres recentes (2 ocorrências). Qual a probabilidade de terem origem na mesma categoria de empresa?

Temos aqui 3 situações possíveis: Foram ocorrências comunicadas por empresas da categoria A **OU** ocorrências comunicadas por empresas da categoria B **OU** ocorrências comunicadas por empresas da categoria C.
Sendo os acontecimentos independentes entre semestres e disjuntos (se ocorreu AA, não ocorreu BB, etc), a probabilidade total é a soma das probabilidades das 3 possibilidades:

$$P(Ambas\ da\ mesma\ categoria, dado\ que\ houve\ 2\ ocorrências)$$
$$= P(A|O) \times P(A|O) + P(B|O) \times P(B|O) + P(C|O) \times P(C|O)$$
$$= (0{,}423)^2 + (0{,}323)^2 + (0{,}254)^2 = 0{,}3477.$$

FIM