

Data: novembro de 2023

Duração: 2 horas

Instruções:

- As questões devem ser respondidas em folha de teste da ESTSetúbal (não se aceitam questões escritas a lápis) e os cálculos devem ser entregues num script do R (ficheiro com extensão .R) que deve ser entregue via Moodle.
- Na folha de teste não pode responder a diferentes questões numa mesma folha de resposta. Coloque o seu nome e número em todas as folhas de teste.
- Não é permitido partilhar ficheiros durante a prova nem utilizar o email ou outras formas de comunicação.
- Só é permitido aceder ao Moodle no início do teste para obter os dados e no fim do teste para enviar o script (ficheiro com extensão .R).
- Durante o teste o seu computador deverá estar sempre no R ou RStudio, não é permitido aceder a outros programas.
- É permitida a consulta do caderno/apontamentos e dos slides das aulas em versão papel.
- O abandono da sala só poderá efetuar-se decorridos 60 minutos a partir do início da prova e implica a entrega da mesma.
- **Justifique convenientemente todas as respostas na folha de teste da ESTSetúbal** e recorra ao software R apenas para efetuar os cálculos necessários.

Questões:

1. Os seguintes dados referem-se ao peso das bagagens individuais numa amostra de 90 passageiros que embarcaram no aeroporto de Lisboa num dado voo.

peso da bagagem (kg)	[0, 5[[5, 10[[10, 15[[15, 20[[20, 25[[25, 30]
número de bagagens	5	5	A	30	B	10

- [0.5] (a) Indique e classifique a variável em estudo.
- [1.0] (b) Sabendo que 20% das bagagens desse voo tem peso inferior a 15 kg, calcule os valores de A e B.
- [1.0] (c) Considere $A = B = 20$, construa a tabela de frequências completa.
2. O RMS Titanic foi um navio britânico construído em Belfast, na Irlanda do Norte, que teve a sua viagem inaugural (e única) em 10 de Abril de 1912. No caminho entre a cidade inglesa de Southampton e a cidade de Nova York colidiu com um icebergue e naufragou nas águas geladas do Atlântico norte às 23h40 do dia 15 de abril de 1902. Estima-se que o navio levava 2224 pessoas a bordo, entre passageiros e tripulação, e mais de 1500 pessoas morreram em decorrência do naufrágio. No ficheiro titanic0.txt (Moodle) tem informação sobre alguns dos passageiros do Titanic. Nesse ficheiro encontram-se os seguintes campos:
- PassengerId = número de identificação do passageiro
 - Survived = se o passageiro sobreviveu (0 = Não; 1 = Sim)
 - Pclass = Classe do camarote do passageiro (1 = 1ª classe; 2 = 2ª classe; 3 = 3ª classe)
 - Sex = género (female = feminino, male = masculino)
 - Age: idade do passageiro (em anos)
 - Fare = Preço da passagem (em libras)

- [1.5] (a) Caso seja possível, identifique a População e a Amostra indicando as suas dimensões, a unidade estatística, as variáveis estatísticas e os dados estatísticos, classificando-os.
- [1.5] (b) Qual o género que sobreviveu mais e com que percentagem? Represente essa informação graficamente.
- [2.0] (c) Para fazer uma análise descritiva dos dados pretende-se classificar os passageiros de acordo com a sua idade. Os passageiros com idade máxima de 18 anos são classificados de "crianças", os passageiros com idade entre os 18 e 25 anos (inclusive) são classificados de "jovens", a classificação de "adultos" refere-se aos passageiros com idade entre os 25 anos e os 65 anos (inclusive), os restantes são classificados de "idosos". Construa a tabela de frequências completa usando a classificação definida para a variável idade e represente-a graficamente.
- [1.5] (d) Faça uma análise descritiva dos dados referentes ao preço da passagem calculando as medidas de localização central e dispersão adequadas e, caso considere adequado, construa o correspondente diagrama de extremos e quartis indicando se existem (e quantos) "outliers" moderados e severos.
3. Um posto de gasolina tem uma loja de conveniência que tem um pequeno bar onde se vende café. O número de cafés vendidos diariamente nesse bar é uma variável aleatória discreta X com a seguinte função de probabilidade:

x	50	100	150	200
$f(x)$	0.2	a	b	0.1

- [1.5] (a) Determine os valores a e b sabendo que, em média, são vendidos diariamente 115 cafés.
- [1.5] (b) Considere $a = 0.3$ e $b = 0.4$. Qual a probabilidade do número de cafés vendidos num dado dia ser superior a 150 sabendo que já foram vendidos pelo menos 100 cafés?
- [2.0] (c) O posto de gasolina é abastecido uma vez por semana e as vendas semanais de gasolina, em milhares de litros, é uma variável aleatória contínua Y com função densidade de probabilidade dada por:

$$f(y) = \begin{cases} y - 1 & , \quad 1 \leq y < 2 \\ 3 - y & , \quad 2 \leq y < 3 \\ 0 & , \quad \text{caso contrário} \end{cases}.$$

Calcule a variância do lucro semanal sabendo que o lucro semanal é dado por $2Y - 1$ e que, em média, semanalmente são vendidos 2000 litros de gasolina.

4. De 1000 declarações de IRS, sabe-se que 100 apresentam erros. Um fiscal das finanças selecionou 20 declarações, ao acaso, para analisar. Sabe-se que o número de declarações analisadas por hora tem distribuição de Poisson com variância 3.
- [1.5] (a) Qual a probabilidade do fiscal vir a encontrar mais do que uma declaração com erro?
- [1.5] (b) Sabendo que os funcionários das repartições de finanças trabalham 6 horas por dia, será razoável admitir que o funcionário conseguirá, num dia de trabalho, analisar pelo menos um quarto das declarações selecionadas?
- (c) O tempo que um contribuinte demora a preencher a declaração de IRS é uma variável aleatória que se admite ter distribuição normal com média 25.6 minutos e desvio padrão igual a 1.6 minutos.
- [1.5] i. Qual a probabilidade do tempo que um contribuinte demora a preencher a declaração de IRS estar entre 24.256 e 27.472 minutos?
- [1.5] ii. Qual o tempo máximo que 1.79% dos contribuintes demoram a preencher a declaração de IRS?

Fim do Teste