Probabilidades e Estatística E - versão A

2º Teste: 28-05-2014

Duração: 1h30

**Nome completo:**________________________________________ **Nº de aluno:**__________

- **Nos grupos 1 a 5, assinale com uma cruz sobre V para verdadeiro ou sobre F para falso, o valor lógico de cada uma das afirmações.**
- **Uma resposta correcta vale 1 valor, uma resposta incorrecta desconta 0.4 valores e uma não resposta nada desconta.**
- **O grupo 6 deve ser respondido no enunciado.**

1. O armazenamento de faturas em formato digital, feito com recurso a um conhecido software, resulta em ficheiros com dimensão média de 100KB e com um desvio padrão de 10KB. Considere um conjunto de 36 ficheiros escolhidos ao acaso. (3.0)

[X] [F] A probabilidade aproximada da média das dimensões dos 36 ficheiros ser no máximo 104KB é 0.9918.

**R:** Dada a amostra aleatória $X_1, ..., X_{36}$ dos 36 ficheiros, sabe-se que $\mu = E(X_i) = 100KB, i = 1, ..., 36$ e $\sigma = \sigma(X_i) = 10, i = 1, ..., 36$. Pelo Teorema Limite Central,

$$Z = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} \overset{a}{\sim} N(0,1).$$

Queremos,

$$P(\bar{X} \leq 104) = P\left(\frac{\bar{X} - 100}{10/6} \leq \frac{104 - 100}{10/6}\right) = \phi(2.4) = 0.9918$$

[V] [X] A probabilidade aproximada da dimensão total dos 36 ficheiros ser superior a 3540KB é 0.1587.

**R:** Dada a amostra aleatória $X_1, ..., X_{36}$ dos 36 ficheiros, sabe-se que $\mu = \mathbb{E}(X_i) = 100KB, i = 1, ..., 36$ e $\sigma = \sigma(X_i) = 10, i = 1, ..., 36$. Pelo Teorema Limite Central,

$$Z = \frac{S_n - n\mu}{\sqrt{n}\sigma} \overset{a}{\sim} N(0,1).$$

Queremos,

$$P(S_{36} > 3540) = 1 - P\left(\frac{S_{36} - 36 \times 100}{6 \times 10} \leq \frac{3540 - 36 \times 100}{6 \times 10}\right) = 1 - \phi(-1) = \phi(1) = 0.8413$$

[V] [X] A média de variáveis aleatórias independentes e identicamente distribuidas tem sempre distribuição normal.

**R:** Apenas a média de v.a. normais terá distribuição normal, para as outras distribuições (para as quais exista valor médio e variância), a normalidade é apenas assimptótica como evidenciada pelo T.L.C.

2. Seja $X_1, ..., X_n$ uma amostra aleatória de uma população com distribuição $U(0, b)$ e $\hat{b}$ um estimador de $b$. (3.0)

$\boxed{\text{V}}$ $\boxed{\text{X}}$ Se $E[\hat{b}] = \frac{n}{n+1}b$ então $\hat{b}$ é estimador centrado para $b$.

**R:** O estimador $\hat{b}$ será estimador centrado para $b$ se $E[\hat{b}] = b$.

$\boxed{\text{X}}$ $\boxed{\text{F}}$ O estimador dos momentos de $b$ é $b^* = 2\bar{X}$.

**R:** O estimador dos momentos será obtido resolvendo a equação $E(X) = \bar{X}$ como $X \sim U(0, b)$ teremos $\frac{0+b}{2} = \bar{X} \Leftrightarrow b = 2\bar{X}$ pelo que o estimador dos momentos será $b^* = 2\bar{X}$.

$\boxed{\text{X}}$ $\boxed{\text{F}}$ Se $EQM(b^*) = \frac{b^2}{3n}$ então $b^*$ é um estimador consistente de $b$.

**R:** O estimador $b^*$ será estimador consistente para $b$ se $\lim_{n\to\infty} EQM(b^*) = 0$, o que se verifica.

3. Recolheu-se uma amostra aleatória $x_1, ..., x_{25}$, correspondente a velocidades de transmissão de dados (em MB/s) através de uma certa ligação à internet. Admitindo que a velocidade de transmissão de dados dessa ligação segue uma distribuição normal e tendo-se observado $\bar{x} = 20$ e $s^2 = 4$. (3.0)

$\boxed{\text{X}}$ $\boxed{\text{F}}$ $IC_{95\%}(\mu) \equiv [19.176; 20.824]$.

**R:** Como estamos no caso de população normal com $\sigma$ desconhecido,

$$IC_{95\%}(\mu) \equiv \left[\bar{X} - t_{n-1;0.025}\frac{S}{\sqrt{n}}; \bar{X} + t_{n-1;0.025}\frac{S}{\sqrt{n}}\right] \equiv [19.176; 20.824]$$

$\boxed{\text{X}}$ $\boxed{\text{F}}$ $IC_{95\%}(\mu) \subseteq IC_{99\%}(\mu)$.

**R:** Como $t_{n-1:0.025} < t_{n-1:0.005}$ teremos $IC_{95\%}(\mu) \subseteq IC_{99\%}(\mu)$.

$\boxed{\text{V}}$ $\boxed{\text{X}}$ $IC_{95\%}(\sigma^2) \equiv [2.637; 6.957]$ (arredondado a 3 casas décimais).

**R:** Temos que,

$$IC_{95\%}(\sigma^2) \equiv \left[\frac{(n-1)S^2}{\chi^2_{n-1:0.025}}; \frac{(n-1)S^2}{\chi^2_{n-1:0.975}}\right] \equiv [2.437; 7.742]$$

4. A velocidade real de processamento (em GHz) do novo processador $i_{2014}$ é uma variável aleatória $X$: (3.0)

$\boxed{\text{X}}\ \boxed{\text{F}}$ O fabricante garante que a velocidade média de processamento é superior a 3.5GHz. As hipóteses do teste a realizar para se testar essa garantia, são:

$$H_0 : \mu \leq 3.5 \text{ vs } H_1 : \mu > 3.5.$$

**R:** A afirmação que é feita é no sentido de $\mu > 3.5$ pelo que na realização do teste as hipóteses serão definidas por:

$$H_0 : \mu \leq 3.5 \text{ vs } H_1 : \mu > 3.5.$$

$\boxed{\text{V}}\ \boxed{\text{X}}$ Para testar se a proporção dos processadores $i_{2014}$ com velocidade real superior a 3.5GHz é de $p = 10\%$, a estatística do teste a realizar, quando se utiliza uma amostra de dimensão $n$, é:

$$Z = \frac{\hat{P} - 0.1}{\sqrt{n \times 0.1 \times 0.9}} \overset{a}{\sim} N(0, 1).$$

**R:** A estatística de teste relativamente a um teste sobre uma proporção $p$ é:

$$Z = \frac{\hat{P} - 0.1}{\sqrt{0.1 \times 0.9/n}} \overset{a}{\sim} N(0, 1).$$

$\boxed{\text{V}}\ \boxed{\text{X}}$ Num teste de hipóteses bilateral para um nível de significância de 5%, em que a estatística do teste tem distribuição $N(0, 1)$, a região de rejeição é:

$$R_{5\%} = ]-\infty; -z_{0.05}[\cup]z_{0.05}, +\infty[.$$

**R:** Se a estatística de teste tem distribuição $N(0, 1)$ a região de rejeição para $\alpha = 5\%$ será:

$$R_{5\%} = ]-\infty; -z_{0.025}[\cup]z_{0.025}, +\infty[.$$

5. Considerere a seguinte amostra de dimensão $n = 30$:

19.1, 17.2, 21.3, 19.4, 18.8, 22.0, 20.7, 21.7, 16.6, 17.8, 19.7, 21.5, 19.7, 21.8, 20.6,
20.3, 20.3, 19.5, 17.4, 16.3, 20.9, 22.8, 19.2, 18.7, 23.8, 20.7, 17.4, 22.2, 23.7, 18.6

Relativamente à aleatoriedade da amostra: (3.0)

$\boxed{\text{X}}\ \boxed{\text{F}}$ O número de sequências observadas na correspondente amostra de sinais é $v_0 = 17$.

**R:** A correspondente amostra de sinais é:

$$. - + - - + - + - +++ - + -- = - - - ++ - - + - - ++ -$$

pelo que $v_0 = 17$.

$\boxed{\text{X}}\ \boxed{\text{F}}$ Na estatística, $Z = \frac{V - \frac{2n-1}{3}}{\sqrt{\frac{16n-29}{90}}}$ o valor de $n$ correspondente à amostra considerada é $n = 29$.

**R:** Como na amostra de sinais há um empate (=) teremos que o valor de $n$ na estatística do teste vale $n = 29$.

$\boxed{\text{V}}\ \boxed{\text{X}}$ Para um valor observado da estatística de teste de $z_{obs} = 2.91$ deve-se aceitar a aleatoriedade da amostra, para $\alpha = 5\%$.

**R:** Para $\alpha = 5\%$ a região de rejeição será:

$$R_{5\%} = ]-\infty; -1.96[\cup]1.96, +\infty[$$

pelo que se deve rejeitar a aleatoriedade da amostra.

**Nome completo:**____________________________________________ **Nº de aluno:**__________

# Justifique detalhadamente as suas respostas

6. Para testar a conjetura sobre se uma população tem distribuição $N(20, 2^2)$ recolheu-se uma amostra de 30 observações e construiu-se a seguinte tabela de frequências:

| Classes | $]-\infty, 17]$ | $]17, 18.5]$ | $]18.5, 20]$ | $]20, 21.5]$ | $]21.5, 23]$ | $]23, +\infty[$ |
|---|---|---|---|---|---|---|
| Frequência Observada ($O_i$) | 2 | 4 | 9 | 8 | 5 | 2 |
| Frequência Esperada ($E_i$) | 2 | 4.8 | 8.2 | 8.2 | 4.8 | 2 |

(5.0)

(a) Complete a tabela de frequências, apresentando os **cálculos detalhados**.

**R:** Para se completar a tabela de frequências teremos de determinar $E_1$ e $E_2$, ou seja, os valores esperados de observações nas duas primeiras classes, se for verdade que a população ($X$) tem distribuição $N(20, 2^2)$.
Uma vez que a distribuição $N(20, 2^2))$ é simétrica em relação a $\mu = 20$, teremos:

$$p_1 = P(X \leq 17 | X \sim N(20, 2^2)) = P(X > 23 | X \sim N(20, 2^2)) = p_6,$$

e portanto

$$E_1 = n \times p_1 = n \times p_6 = E_6 = 2.$$

De forma análoga teremos que $p_2 = p_5$ e portanto $E_2 = E_5 = 4.8$.
Em alternativa, pode-se calcular

$$p_1 = P(X \leq 17 | X \sim N(20, 2^2)) = P\left(\frac{X-20}{2} \leq \frac{17-20}{2} \middle| X \sim N(20, 2^2)\right) = \phi(-1.5) = 0.0668$$
$$\Rightarrow E_1 = 30 \times 0.0668 = 2$$

e calcula-se $E_2$ de forma análoga.
Devemos ter em atenção que na tabela de frequências existem $E_i$'s que são inferiores a 5, pelo que teremos de agrupar algumas classes. Assim procedendo, obteremos a nova tabela de frequências:

| Classes | $]-\infty, 18.5]$ | $]18.5, 20]$ | $]20, 21.5]$ | $]21.5, +\infty[$ |
|---|---|---|---|---|
| Frequência Observada ($O_i$) | 6 | 9 | 8 | 7 |
| Frequência Esperada ($E_i$) | 6.8 | 8.2 | 8.2 | 6.8 |

(b) Realize de **forma detalhada** o teste de ajustamento do Qui-quadrado para testar a conjetura apresentada. Utilize um nível de significância de $\alpha = 5\%$.

**R:** Para realizar o teste de ajustamento do qui-quadrado teremos de seguir os seguintes passos:

- Hipóteses do teste:
$$H_0 : X \sim N(20, 2^2) \quad vs \quad H_1 : X \nsim N(20, 2^2)$$
- Estatística do teste:
$$X^2 = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i} \overset{sob\ H_0}{\sim} \chi^2_{k-p-1=3}$$
uma vez que o número de classes é $k = 4$ e como não se estimaram parâmetros, $p = 0$.
- Região de rejeição para $\alpha = 5\%$:
$$R_{0.05} \equiv ]\chi^2_{3:0.05}; +\infty[ \equiv ]7.81; +\infty[$$
- Regra de decisão e decisão: Devemos rejeitar $H_0$ se $x^2_{obs} \in R_{0.05}$.
Como
$$x^2_{obs} = \frac{(6-6.8)^2}{6.8} + \frac{(9-8.2)^2}{8.2} + \frac{(8-8.2)^2}{8.2} + \frac{(7-6.8)^2}{6.8} = 0.18 \notin R_{0.05}$$
não se rejeita $H_0$, podendo-se concluir que os dados dão suporte à conjetura de a população ter distribuição $N(20, 2^2)$.