Departamento de Matemática da Universidade de Coimbra

ESTATÍSTICA

LEI e LECD

2ª Frequência

Duração: 1h 30min 13-06-2023

Observação: Na resolução das questões deverá justificar o raciocínio utilizado e apresentar todos os cálculos efetuados.

---

1. Numa rede informática, o tempo de transmissão de um pacote de 10 MB de informação é representado por uma variável aleatória cujo valor médio era, há algum tempo, 7 segundos. O gestor da rede, considerando este tempo excessivo, decidiu efetuar algumas alterações na rede. Com o objetivo de avaliar se houve uma redução significativa do tempo de transmissão dos dados, foram realizados 50 ensaios independentes com pacotes de 10 MB e registados os correspondentes tempos de transmissão. Os valores observados, $x_1, x_2, ..., x_{50}$, conduziram aos seguintes resultados:

$$\sum_{i=1}^{50} x_i = 332.8 \quad \text{e} \quad \sum_{i=1}^{50} x_i^2 = 2231.4\,.$$

a) Calcule estimativas cêntricas da média e da variância do tempo de transmissão de um pacote de 10 MB.

b) Ao nível de significância 0.05, pode afirmar que as referidas alterações resultaram numa redução significativa do tempo médio de transmissão de um pacote de 10 MB?

2. Um fabricante de folhas de papel para impressoras possui um processo de produção que opera de modo contínuo. O comprimento das folhas, em milímetros, é bem modelado por uma variável aleatória com distribuição normal, cujo valor médio deverá ser $280\,mm$. O correspondente desvio padrão deverá ser inferior a $0.5\,mm$. Para verificar se as folhas são produzidas cumprindo estes requisitos, são recolhidas amostras de dimensão 25 ao longo do processo de produção.

a) Uma das amostras recolhidas apresenta desvio padrão igual a 0.437. A partir desta amostra, obtenha um intervalo de confiança para o desvio padrão do comprimento das folhas, com grau de confiança 0.95.

b) A partir da amostra referida na alínea anterior, foi obtido no software R o output seguinte.

```
        One Sample t-test
data: comprimento
t = -1.3257, df = 24, p-value = 0.1974
alternative hypothesis: true mean is not equal to 280
95 percent confidence interval:
   279.7034   280.0646
```

(i) Apresente a hipótese nula em teste neste output, indique a estatística de teste usada e a sua distribuição de probabilidade sob aquela hipótese.

(ii) Ao nível de significância 0.05, pode afirmar que as folhas estão a estão a ser fabricadas com o comprimento exigido?

(iii) Interprete os valores 279.7034 e 280.0646 que figuram no output.

v.p.f.

3. Pretende-se averiguar se determinada variável aleatória discreta, X, tem distribuição geométrica $\mathcal{G}(0.4)$, de suporte $\mathbb{N}_0$. No quadro abaixo encontra-se resumida uma amostra de X, de dimensão 100, bem como os valores $P(X = x_i)$, $i = 1, ..., 7$, para os diferentes valores observados de X, $x_1, ..., x_7$, admitindo que X tem a distribuição acima referida.

Observações (x_i)	0	1	2	3	4	5	6
Frequência absoluta	44	22	18	6	4	4	2
$P(X = x_i)$, $X \sim \mathcal{G}(0.4)$	0.4	0.24	0.144	0.0864	0.0518	0.0311	0.0187

a) Apresente as categorias adequadas para efetuar o teste do quiquadrado de ajustamento bem como as correspondentes probabilidades sob a hipótese "X tem distribuição $\mathcal{G}(0.4)$".

b) O p-valor do teste do quiquadrado efetuado no software R com base nas categorias e corrrespondentes probabilidades a que se refere a alínea a) é 0.7078. Aos níveis de significância usuais, o que pode concluir sobre a distribuição de probabilidade de X?

4. Seja X uma variável aleatória real seguindo uma lei uniforme num intervalo $[0, \theta]$, onde θ é um parâmetro real positivo, desconhecido. Nestas condições, tem-se $E(X) = \dfrac{\theta}{2}$ e $V(X) = \dfrac{\theta^2}{12}$. Seja $(X_1, ..., X_n)$ uma amostra aleatória de X, de dimensão n. Verifique que o estimador $T_n = \dfrac{1}{3}\overline{X}_n^2$ é assintoticamente cêntrico de $V(X)$ e deduza, a partir dele, um estimador cêntrico de $V(X)$.

FORMULÁRIO

Se $X \sim \mathcal{G}(p)$, de suporte $\mathbb{N}_0$, então $P(X = k) = p(1-p)^k,\ k \in \mathbb{N}_0,\ p \in]0, 1[$.

Cotação <u>prevista</u>: **1.a)** 1.25 **1.b)** 2.0 **2.a)** 2.25 **2.b)** 1.5 **3.** 2.0 **4.** 1.0

Proposta de resolução

1.a) Análogo ao Exercício 3.a) da Folha 7 (2023/2024).

Solução: X: "tempo de transmissão de um pacote de 10 MB de informação (em segundos)"

- Estimativa cêntrica para $E(X)$: $\overline{x} = 6.656$
- Estimativa cêntrica para $V(X)$: $\widehat{s}^2 \simeq 0.3323$

1.b) Análogo ao Exercício 2 da Folha 8 (2023/2024), mas com a estatística de teste

$$\frac{\overline{X} - 7}{\widehat{S}/\sqrt{50}} \overset{\bullet}{\sim} N(0,1), \text{ sob } H_0$$

porque X não é normal, o desvio padrão de X é desconhecido sob H_0 e $n = 50 > 30$.

A região crítica obtém-se como no Exercício 6 da Folha 8 (2023/2024):

$RC =]-\infty, -c]$, com $c > 0$ tal que

$$\begin{aligned} 0.05 &= P(\text{rejeitar } H_0/H_0 \text{ verdadeira}) \\ &= P(Z \leq -c), \quad Z \overset{\bullet}{\sim} N(0,1) \\ &\simeq P(U \leq -c), \quad U \sim N(0,1). \end{aligned}$$

Temos então

$$\begin{aligned} P(U \leq -c) \simeq 0.05 \quad &\Leftrightarrow \quad P(U \geq c) \simeq 0.05, \quad \text{pela simetria da lei } N(0,1) \\ &\Leftrightarrow \quad 1 - P(U < c) \simeq 0.05 \\ &\Leftrightarrow \quad 1 - F_U(c) \simeq 0.05, \quad \text{porque } U \text{ é contínua} \\ &\Leftrightarrow \quad F_U(c) \simeq 0.95 \Leftrightarrow c \simeq 1.645 \end{aligned}$$

Portanto, $RC =]-\infty, -1.645]$.

$Z_{obs} \simeq -4.2 \in RC$, pelo que rejeitamos H_0 ao n.s. 0.05.

Resposta: Sim.

2.a) X: "comprimento das folhas (em mm)", $X \sim N(m, \sigma)$.

NOTA: No enunciado afirma-se que m **deverá ser** 280 mm e que σ **deverá ser** inferior a 5 mm. **Não se afirma que** $m = 280\,mm$ nem que $\sigma < 5\,mm$. A amostra é recolhida com o **objetivo de averiguar se** isso acontece. Assim, em particular, não se pode admitir que $m = 280$.

A resolução é análoga à do Exercício 2.b) da Folha 7 e também à do exercício 3.b(ii) da mesma folha (2023/2024).

Solução: $\left]\sqrt{\frac{4.774225}{39.4}}, \sqrt{\frac{4.774225}{12.4}}\right[\simeq]0.3481, 0.6205[$

2.b(i) Hipótese nula: H_0: $m = 280$

Estatística de teste usada: $\frac{\overline{X} - 280}{\widehat{S}/\sqrt{25}}$, que, sob H_0, segue a distribuição de Student com 24 graus de liberdade, $T(24)$.

2.b(ii) p-valor$= 0.1974 > 0.05 \Rightarrow$ aceitar H_0 ao n.s. 0.05. Portanto, ao n.s. 0.05, podemos afirmar que as folhas estão a ser fabricadas com o comprimento exigido, em média.

2.b(iii) $]279.7034, 280.0646[$ é um intervalo de confiança para m com grau de confiança 0.95.

3.a) Corresponde ao Exercício 3 da Folha 9 (2023/2024).

Solução:

Categorias adequadas	$\{0\}$	$\{1\}$	$\{2\}$	$\{3\}$	$\{4\}$	$\{5, 6, 7, 8, ...\}$
$P(X = x_i)$, $X \sim \mathcal{G}(0.4)$	0.4	0.24	0.144	0.0864	0.0518	0.0778

3.b) p-valor$= 0.7078 > \alpha$, para os n.s. α usuais, $0.01 \leq \alpha \leq 0.1$. Assim, aos n.s. usuais, podemos concluir que $X \sim \mathcal{G}(0.4)$.

4. Corresponde ao Exercício 7.a), da Folha 6 (2023/2024).