Probabilidades e Estatística

LEGM, LEAN, MEC, MEAmbi; LMAC; LEE, LEGI

DM
DEPARTAMENTO DE MATEMÁTICA
TÉCNICO LISBOA

1º semestre – 2020/2021
15/01/2021 – **09:00**

Duração: 60+15 minutos **Teste 2A**

Nº : ________ **Nome:** ______________________________ **Curso:** ________ **Sala:** ________

1. Em certa experiência, uma fonte radioativa emite pelo menos uma partícula alfa com probabilidade desconhecida p e não emite qualquer partícula alfa com probabilidade $1-p$. Admita que X é uma variável aleatória que representa o número total de experiências em que ocorre a emissão de pelo menos uma partícula alfa, em duas experiências independentes.

Seja (X_1, X_2) uma amostra aleatória de dimensão dois extraída da população X. Considere os estimadores

$$T_1 = (a\,X_1 + b\,X_2)/(a+b) \quad \text{e} \quad T_2 = (c\,X_1 + d\,X_2)/(c+d),$$

na estimação do parâmetro $E(X)$. Obtenha a razão entre os erros quadráticos médios de T_1 e T_2.

Preencha a caixa abaixo com o resultado obtido com, pelo menos, quatro casas decimais.

$\boxed{}$

- **V.a. de interesse distribuição, valor esperado e variância**

  $X=$ número total de experiências em que ocorre a emissão de pelo menos uma partícula alfa, em duas experiências independentes

  $X \sim \text{binomial}(2, p)$

  $E(X) = 2p$

  $V(X) = 2p(1-p)$

- **Estimadores de** $E(X) = 2p$

  $T_1 = \frac{a\,X_1 + b\,X_2}{a+b}$

  $T_2 = \frac{c\,X_1 + d\,X_2}{c+d}$

- **Erros quadráticos médios**

$$\begin{aligned}
EQM_{2p}(T_1) &= V(T_1) + [E(T_1) - 2p]^2 \\
&= \frac{a^2+b^2}{(a+b)^2} V(X) + \left[\frac{a+b}{a+b} E(X) - 2p\right]^2 \\
&= \frac{a^2+b^2}{(a+b)^2} 2p(1-p) + (2p-2p)^2 \\
&= \frac{a^2+b^2}{(a+b)^2} 2p(1-p)
\end{aligned}$$

$$\begin{aligned}
EQM_{2p}(T_2) &= V(T_2) + [E(T_2) - 2p]^2 \\
&= \frac{c^2+d^2}{(c+d)^2} V(X) + \left[\frac{c+d}{c+d} E(X) - 2p\right]^2 \\
&= \frac{c^2+d^2}{(c+d)^2} 2p(1-p)
\end{aligned}$$

- **Razão pedida**

$$\frac{EQM_{2p}(T_1)}{EQM_{2p}(T_2)} = \frac{\frac{a^2+b^2}{(a+b)^2}\,2p(1-p)}{\frac{c^2+d^2}{(c+d)^2}\,2p(1-p)} = \frac{(a^2+b^2)\,(c+d)^2}{(a+b)^2\,(c^2+d^2)}$$

2. Para avaliar a diferença entre os valores esperados dos ritmos cardíacos (em batimentos por minuto) de homens e de mulheres, foram recolhidas ao acaso medições de n_H homens e n_M mulheres, tendo-se obtido as médias e variâncias corrigidas amostrais seguintes: $\bar{x}_H = \bar{x}_H$ e $s^2_H = s^2_H$ para os homens; $\bar{x}_M = \bar{x}_M$ e $s^2_M = s^2_M$ para as mulheres.

Considere que os dois grupos de medições são concretizações de duas amostras aleatórias independentes, provenientes de distribuições normais com valores esperados desconhecidos μ_H e μ_M e variâncias σ^2_H e σ^2_M desconhecidas mas iguais.

Obtenha um intervalo de confiança a $(1-\alpha)\times 100\%$ para a diferença de valores esperados $\mu_H - \mu_M$.

Assinale a sua resposta com uma cruz.

⦿ [,]

⦿ [,]

⦿ [,]

⦿ [,]

- **V.a. de interesse**

  X_H = ritmo cardíaco de homem

  X_M = ritmo cardíaco de mulher

- **Distribuição de X_H e X_M**

  $X_H \sim \text{normal}(\mu_H, \sigma^2_H) \perp\!\!\!\perp X_2 \sim \text{normal}(\mu_M, \sigma^2_M)$

  $(\mu_H - \mu_M)$ DESCONHECIDO

  σ^2_H e σ^2_M desconhecidas mas iguais

  $n_H < 30$, $n_M < 30$

- **Obtenção do IC para $\mu_H - \mu_M$**

  Passo 1 — Seleção da v.a. fulcral para $(\mu_H - \mu_M)$

$$Z = \frac{(\bar{X}_H - \bar{X}_M) - (\mu_H - \mu_M)}{\sqrt{\frac{(n_H-1)\,S^2_H + (n_M-1)\,S^2_M}{n_H+n_M-2} \times \left(\frac{1}{n_H} + \frac{1}{n_M}\right)}} = \frac{(\bar{X}_H - \bar{X}_M) - (\mu_H - \mu_M)}{S} \sim t_{(n_H+n_M-2)}$$

  Passo 2 — Obtenção dos quantis de probabilidade

$$\begin{cases} a_\alpha = F^{-1}_{t_{(n_H+n_M-2)}}(\alpha/2) = -F^{-1}_{t_{(n_H+n_M-2)}}(1-\alpha/2) \\ b_\alpha = F_{t_{(n_H+n_M-2)}}(1-\alpha/2) \end{cases}$$

  Passo 3 — Inversão da desigualdade $a_\alpha \le Z \le b_\alpha$

$$P(a_\alpha \le Z \le b_\alpha) = 1-\alpha$$

$$P\left[a_\alpha \le \frac{(\bar{X}_H - \bar{X}_M) - (\mu_H - \mu_M)}{S} \le b_\alpha\right] = 1-\alpha$$

$$P\left[(\bar{X}_H - \bar{X}_M) - b_\alpha \times S \le \mu \le (\bar{X}_H - \bar{X}_M) - a_\alpha \times S\right] = 1-\alpha$$

Passo 4 — Concretização

$$IC_{(1-\alpha)\times 100\%}(\mu_H-\mu_M) = \left[(\bar{x}_H - \bar{x}_M) \pm F^{-1}_{t_{(n_H+n_M-2)}}(1-\frac{\alpha}{2}) \times \sqrt{\frac{(n_H-1)\, s_H^2 + (n_M-1)\, s_M^2}{n_H+n_M-2}\left(\frac{1}{n_H}+\frac{1}{n_M}\right)} \right].$$

3. Um fabricante de cadeira de rodas elétricas afirma que o diâmetro (em cm) das rodas traseiras das cadeiras por ele fabricadas possui valor esperado igual a 61 cm. Foram selecionadas aleatoriamente n cadeiras, tendo-se observado uma média e uma variância amostral corrigida dos diâmetros das rodas traseiras iguais a $\bar{x} = \bar{x}$ e $s^2 = s^2$, respetivamente.

Apoiarão os dados a conjetura do fabricante? Decida com base no valor-p aproximado.

Assinale a sua resposta com uma cruz.

- Rejeita-se H_0 a 1%, 5% e 10%.
- Rejeita-se H_0 a 5% e 10% e não se rejeita H_0 a 1%.
- Rejeita-se H_0 a 10% e não se rejeita H_0 a 1% e 5%.
- Não se rejeita H_0 a 1%, 5% e 10%.

- **V.a. de interesse**
  X = diâmetro das rodas traseiras (em cm)

- **Situação**
  X v.a. com dist. arbitrária, com valor esperado μ e variância σ^2

  μ DESCONHECIDO

  σ^2 desconhecido

- **Hipóteses**
  $H_0 : \mu = \mu_0 = 61$
  $H_1 : \mu \neq \mu_0$

- **Estatística de teste**

$$T \quad = \quad \frac{\bar{X}-\mu_0}{\frac{S}{\sqrt{n}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$$

  [uma vez que se pretende efectuar um teste para o valor esperado de uma população com distribuição arbitrária e variância desconhecida, dispondo de uma amostra suficientemente grande.]

- **Região de rejeição de H_0**
  O teste é bilateral, logo a região de rejeição de H_0 é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base no valor-p)**

$$valor-p \quad = \quad 2 \times P(T > |t| \mid H_0) \simeq 2 \times \left[1 - \Phi\left(T > \left|\frac{\bar{x}-\mu_0}{\frac{s}{\sqrt{n}}}\right|\right)\right]$$

  Devemos escolher a resposta, das quatro apresentadas acima, que se coaduna com:
  - a não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq valor-p$;
  - a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor-p$.

4. Os testes PCR (*polymerase chain reaction*) são frequentemente usados para detectar a presença do vírus SARS-CoV-2 em amostras de DNA. Considere que a variável aleatória X representa o número de testes PCR efetuados até se encontrar uma primeira amostra contaminada. Uma especialista em saúde pública defende a hipótese H_0 de que X possui distribuição geométrica com parâmetro $p = 0.10$. A concretização de uma amostra aleatória de dimensão n proveniente da população X conduziu à seguinte tabela de frequências

Classe	$\{1,2,3,4\}$	$\{5,6,7\}$	$\{8,9,10\}$	$\{11,12\}$	$\{13,14,\dots\}$
Frequência absoluta observada	o_1	o_2	o_3	o_4	o_5
Frequência absoluta esperada sob H_0	E_1	E_2	E_3	E_4	E_5

Após ter calculado a frequência absoluta observada omissa o_5, bem como as frequências absolutas esperadas sob H_0 omissas E_1 e E_5 (aproximando-as às centésimas), averigue se H_0 é consistente com este conjunto de dados. Decida com base no valor-p aproximado.

Assinale a sua resposta com uma cruz.

- Rejeita-se H_0 a 1%, 5% e 10%.
- Rejeita-se H_0 a 5% e 10% e não se rejeita H_0 a 1%.
- Rejeita-se H_0 a 10% e não se rejeita H_0 a 1% e 5%.
- Não se rejeita H_0 a 1%, 5% e 10%.

- **V.a. de interesse**

  X = número de testes PCR efetuados até se encontrar uma primeira amostra contaminada

- **Hipóteses**

  $H_0 : X \sim \text{geométrica}(p = 0.10)$

  $H_1 : \neg H_0$

- **Estatística de teste**

  $T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \stackrel{a}{\sim}_{H_0} \chi_{(k-\beta-1)}$,

  onde: k = no. de classes = 5; $\beta = 0$; o_i = freq. abs. observada da classe i; E_i = freq. abs. esperada sob H_0 da classe i.

- **Região de rejeição de H_0**

  Ao lidarmos com um teste de ajustamento do qui-quadrado, a região de rejeição de H_0 é um intervalo à direita $W = (c, +\infty)$.

- **Frequência absoluta observada omissa**

$$o_5 = n - \sum_{i=1}^{k-1} o_i$$

- **Frequências absolutas esperadas sob H_0 omissas**

  Uma vez que

$$\begin{aligned}
P(X = x \mid H_0) &= (1-p)^{x-1} \times p, \quad x \in \mathbb{N} \\
F_{x|H_0}(x) &= P(X \leq x \mid H_0) \\
&= \sum_{i=1}^{x} P(X = x \mid H_0) \\
&= 1 - (1-p)^x, \quad x \in \mathbb{N},
\end{aligned}$$

  temos

$$\begin{aligned}
E_1 &= n \times F_{X|H_0}(4) \\
&= n \times 0.3439
\end{aligned}$$

$$E_5 \quad = \quad n - \sum_{i=1}^{k-1} E_i.$$

- **Decisão (com base no valor-p)**

$$valor-p \quad = \quad P\left(T > t = \sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i} \mid H_0\right) \simeq 1 - F_{\chi^2_{(k-1)}}\left(\sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i}\right)$$

Devemos escolher a resposta, das quatro apresentadas acima, que se coaduna com:

- a não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq valor-p$;
- a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor-p$.

5. Uma empresa, dedicada à comercialização e reparação de portáteis, decidiu analisar a relação entre a duração do tempo de reparação (Y) e o número de componentes eletrónicas a serem reparadas (x). A empresa tem vindo a registar tempos de reparação (em minutos) dos portáteis e respetivo número de componentes a serem reparadas. Os resultados obtidos são os seguintes:

Número de componentes (x)	x_1	x_2	x_3	x_4	x_5	x_6
Tempo de reparação (Y)	31	32	35	37	41	59

$\sum_{i=1}^{n} x_i = \sum_{i=1}^{n} x_i$, $\sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i^2$, $\sum_{i=1}^{n} y_i = \sum_{i=1}^{n} y_i$, $\sum_{i=1}^{n} y_i^2 = \sum_{i=1}^{n} y_i^2$, $\sum_{i=1}^{n} x_i y_i = \sum_{i=1}^{n} x_i y_i$,

Admita que as variáveis x e Y estão relacionadas de acordo com o modelo de regressão linear simples: $Y = \beta_0 + \beta_1 x + \epsilon$.

O que pode concluir sobre a significância do modelo de regressão? Decida com base no valor-p.

Assinale a sua resposta com uma cruz.

- ⦿ Rejeita-se H_0 a 1%, 5% e 10%.
- ⦿ Rejeita-se H_0 a 5% e 10% e não se rejeita H_0 a 1%.
- ⦿ Rejeita-se H_0 a 10% e não se rejeita H_0 a 1% e 5%.
- ⦿ Não se rejeita H_0 a 1%, 5% e 10%.

- **Modelo de RLS**

  Y = tempo de reparação (v.a. resposta)

  x = número de componentes (variável explicativa)

  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i, \quad i = 1, \ldots, n$

- **Hipóteses de trabalho**

  $\epsilon_i \overset{i.i.d.}{\sim} \text{normal}(0, \sigma^2), \quad i = 1, \ldots, n$

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 0$

  $H_1 : \beta_1 \neq \beta_{1,0}$

- **Estatística de teste**

$$T = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_i x_i^2 - n\bar{x}^2}}} \overset{H_0}{\sim} t_{(n-2)}$$

- **Região de rejeição de H_0** (para valores da estatística de teste)

  Dado que o teste é bilateral, a região de rejeição de H_0 é a reunião de intervalos $W = (-\infty, -c) \cup (c, +\infty)$.

- **Valor observado da estatística de teste**

$$t = \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_i x_i^2 - n\bar{x}^2}}}$$

  onde

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n x_i y_i - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)\left(\frac{\sum_{i=1}^n y_i}{n}\right)}{\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}$$

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^n y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^n x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{n-2}\left\{\left[\sum_{i=1}^n y_i^2 - n\left(\frac{\sum_{i=1}^n y_i}{n}\right)^2\right] - (\hat{\beta}_1)^2\left[\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2\right]\right\}.
\end{aligned}$$

  [Convinha que t fosse calculado com base nas fórmulas acima que tiram partido do valor de n e das somas que se encontram no enunciado do problema.]

- **Decisão (com base no valor-p)**

  $valor - p = 2 \times \left[1 - F_{t_{(n-2)}}(|t|)\right]$

  Devemos escolher a resposta, das quatro apresentadas acima, que se coaduna com:

  - a não rejeição de H_0 a qualquer n.s. $\alpha_0 \leq valor - p$;
  - a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor - p$.