DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA

Probabilidades e Estatística

LEIC-A, MEBiol, MEBiom, MEM, MEEC, MEFT, MEMec, MEQ; MEAer; LERC, LEIC-T

1º semestre – 2020/2021
15/01/2021 – **11:00**

Duração: 60+15 minutos

Teste 2B

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

1. Admita que o número de programas examinados de modo independente até que se observe o primeiro programa que não compile é representado pela variável aleatória X com distribuição geométrica com parâmetro p, onde p é uma probabilidade desconhecida. Determine a estimativa de máxima verosimilhança de p, atendendo à amostra $(x_1, \ldots, x_5)$ proveniente da população X.

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

- **V.a. de interesse**
  X = no. de programas examinados ... até que se observe o 1o. programa que não compile
  $X \sim$ geométrica(p)

- **F.p. de X**
  $P(X = x) = (1-p)^{x-1}\, p, \quad x = 1, 2, \ldots$

- **Parâmetro desconhecido**
  $p, \quad p \in (0,1)$

- **Amostra**
  $\underline{x} = (x_1, \ldots, x_n)$ amostra de dimensão n proveniente da população X

- **Obtenção da estimativa de MV de p**

  Passo 1 — Função de verosimilhança

$$\begin{aligned} L(p \mid \underline{x}) &= P(\underline{X} = \underline{x}) \\ &\overset{X_i \overset{i.i.d}{\sim} X}{=} \prod_{i=1}^{n} P(X = x_i) \\ &= \prod_{i=1}^{n} \left[(1-p)^{x_i - 1}\, p\right] \\ &= (1-p)^{\sum_{i=1}^{n} x_i - n}\, p^n, \quad 0 < p < 1 \end{aligned}$$

  Passo 2 — Função de log-verosimilhança

$$\ln L(p|\underline{x}) = \left(\sum_{i=1}^{n} x_i - n\right) \times \ln(1-p) + n \times \ln(p)$$

  Passo 3 — Maximização

  A estimativa de MV de p é doravante representada por $\hat{p}$ e

$$\hat{p} \ : \ \begin{cases} \left.\dfrac{d \ln L(p|\underline{x})}{dp}\right|_{p=\hat{p}} = 0 & \text{(ponto de estacionaridade)} \\ \left.\dfrac{d^2 \ln L(p|\underline{x})}{dp^2}\right|_{p=\hat{p}} < 0 & \text{(ponto de máximo)} \end{cases}$$

$$\hat{p} \quad : \quad \begin{cases} -\frac{\sum_{i=1}^{n} x_i - n}{1-\hat{p}} + \frac{n}{\hat{p}} = 0 \\ -\frac{\sum_{i=1}^{n} x_i - n}{(1-\hat{p})^2} - \frac{n}{\hat{p}^2} < 0 \end{cases} \qquad \begin{cases} -\hat{p}\sum_{i=1}^{n} x_i + n\hat{p} + n - n\hat{p} = 0 \quad (\hat{p} \neq 0,1) \\ \text{prop. verdadeira já que} \quad \sum_{i=1}^{n} x_i \geq n, \quad n \in \mathbb{N} \end{cases}$$

Passo 4 — Estimativa de MV de p

$$\hat{p} = \frac{1}{\bar{x}}$$

2. Admita que a resistência mecânica de certo material cerâmico possui distribuição normal com valor esperado e variância desconhecidos. Uma vez medidas as resistências mecânicas de n especímenes selecionados casualmente, verificou-se que a média amostral e a variância amostral corrigida são iguais a $\bar{x}$ e s^2, respetivamente.

Obtenha um intervalo de confiança a $(1-\alpha) \times 100\%$ para o valor esperado da resistência mecânica.

Assinale a sua resposta com uma cruz.

- ⊙ [,]
- ⊙ [,]
- ⊙ [,]
- ⊙ [,]

- **V.a. de interesse**

  X = resistência mecânica de certo material cerâmico

- **Situação**

  $X \sim \text{normal}(\mu, \sigma^2)$

  μ DESCONHECIDO

  σ^2 desconhecido

- **Obtenção do IC para μ**

  Passo 1 — Selecção da v.a. fulcral para μ

  $$Z \quad = \quad \frac{\bar{X} - \mu}{S/\sqrt{n}} \sim t_{(n-1)}$$

  [dado que é suposto determinar um IC para o valor esperado de uma população normal, com variância desconhecida.]

  Passo 2 — Obtenção dos quantis de probabilidade

  Far-se-á uso dos quantis

  $$(a_\alpha, b_\alpha) \quad : \quad \begin{cases} P(a_\alpha \leq Z \leq b_\alpha) = 1 - \alpha \\ P(Z < a_\alpha) = P(Z > b_\alpha) = \alpha/2. \end{cases} \qquad \begin{cases} a_\alpha = F^{-1}_{t_{(n-1)}}(\alpha/2) \\ b_\alpha = F^{-1}_{t_{(n-1)}}(1-\alpha/2). \end{cases}$$

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

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

  $$P\left(a_\alpha \leq \frac{\bar{X}-\mu}{S/\sqrt{n}} \leq b_\alpha\right) = 1 - \alpha$$

  $$P\left(\bar{X} - b_\alpha \times \frac{S}{\sqrt{n}} \leq \mu \leq \bar{X} - a_\alpha \times \frac{S}{\sqrt{n}}\right) = 1 - \alpha$$

Passo 4 — Concretização

Tendo em conta os quantis acima e o facto de as concretizações de $\bar{X}$ e S^2 serem iguais a $\bar{x}$ e s^2 (respetivamente), temos

$$IC_{(1-\alpha)\times 100\%}(\mu) = \left[\bar{x} - F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}, \quad \bar{x} + F^{-1}_{t_{(n-1)}}(1-\alpha/2) \times \frac{s}{\sqrt{n}}\right].$$

3. Considere que a variável aleatória X_1 (respetivamente X_2) representa a espessura de arruela proveniente da unidade fabril 1 (respetivamente da unidade fabril 2). Ao selecionar-se casualmente $n_1 = n_1$ e $n_2 = n_2$ arruelas da produção diária das unidades fabris 1 e 2, obtiveram-se os seguintes resultados: $\bar{x}_1 = \bar{x}_1$, $s_1^2 = s_1^2$; $\bar{x}_2 = \bar{x}_2$, $s_2^2 = s_2^2$.

Admita que as variáveis aleatórias X_1 e X_2 são independentes e confronte as hipóteses $H_0 : E(X_1) = E(X_2)$ e $H_1 : E(X_1) \neq E(X_2)$. 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_i = espessura de arruela proveniente da unidade fabril i, $i = 1,2$

- **Situação**

  X_i v.a. com dist. arbitrária, com valor esperado μ_i e variância σ_i^2, $i = 1,2$

  $X_1 \perp\!\!\!\perp X_2$

  $(\mu_1 - \mu_2)$ DESCONHECIDO

  σ_1^2 e σ_2^2 desconhecidos

- **Hipóteses**

  $H_0 : \mu_1 - \mu_2 = \mu_0 = 0$

  $H_1 : \mu_1 - \mu_2 \neq \mu_0$

- **Estatística de teste**

$$T = \frac{(\bar{X}_1 - \bar{X}_2) - \mu_0}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$$

  [uma vez que se pretende efectuar um teste de igualdade dos valores esperados de duas populações com distribuições arbitrárias independentes e com variâncias desconhecidas, dispondo de duas amostras suficientemente grandes.]

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

  Estamos a lidar com um teste bilateral ($H_1 : \mu_1 - \mu_2 \neq \mu_0$), logo a região de rejeição de H_0 (para valores da estatística de teste) é do tipo $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base no valor-p)**

$$valor-p = P(T > |t| \mid H_0) \simeq 2 \times \left[1 - \Phi\left(\left|\frac{(\bar{x}_1 - \bar{x}_2) - \mu_0}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}\right|\right)\right]$$

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. Seja X a variável aleatória que representa o número de inspeções automóvel solicitadas diariamente a uma oficina mecânica. Uma engenheira sustenta a hipótese H_0 de que X possui função de probabilidade

$$P(X = x) = \frac{1}{6}(x+3)(x+2)(x+1)\, p^4\, (1-p)^x, \quad x = 0, 1, 2, \ldots$$

A contabilização do número de tais solicitações, em n dias selecionados casualmente, conduziu à seguinte tabela de frequências:

Número diário de inspeções automóvel	0	1	2	3	>3
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 as frequências absolutas esperadas sob H_0 omissas, E_2 e E_4 (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 inspeções automóvel solicitadas diariamente à oficina mecânica

- **Hipóteses**
  $H_0 : P(X = x) = \frac{1}{6}(x+3)(x+2)(x+1)\, p^4\, (1-p)^x, \quad x = 0, 1, 2, \ldots$
  $H_1 : \neg H_0$

- **Estatística de teste**
  $T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \overset{a}{\sim}_{H_0} \chi_{(k-\beta-1)}$,
  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 $W = (c, +\infty)$.

- **Frequências absolutas esperadas sob H_0 omissas**

$$\begin{aligned} E_2 &= n \times P(X = 1 \mid H_0) \\ &= n \times \frac{(1+3)\times(1+2)\times(1+1)}{6} p^4 (1-p)^2 \\ E_4 &= n - (E_1 + E_2 + E_3 + E_5) \end{aligned}$$

- **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. Por forma a avaliar a relação entre a percentagem de hidrocarbonetos presentes no condensador principal de uma unidade de destilação (x) e a pureza do oxigénio produzido (Y), recolheu-se uma amostra casual de dimensão $n = 10$ que conduziu aos seguintes resultados:

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

onde $[\min_{i=1,\dots,n} x_i, \max_{i=1,\dots,n} x_i] = [x_{(1)}, x_{(n)}]$.

Com base no modelo de regressão linear simples, determine um intervalo de confiança a $(1-\alpha) \times 100\%$ para o valor esperado de Y quando $x = x_0$.

Assinale a sua resposta com uma cruz.

- ⊙ [,]
- ⊙ [,]
- ⊙ [,]
- ⊙ [,]

- **Hipóteses de trabalho**
  $\epsilon_i \stackrel{i.i.d.}{\sim} \text{normal}(0, \sigma^2)$, $i = 1, \dots, n$

- **Obtenção do IC a** $(1-\alpha) \times 100\%$ **para** $E(Y \mid x_0)$

  Passo 1 — V.a. fulcral para $E(Y \mid x_0) = \beta_0 + \beta_1 x_0$

$$Z \quad = \quad \frac{(\hat\beta_0 + \hat\beta_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{\left(x_0 - \frac{\sum_{i=1}^n x_i}{n}\right)^2}{\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}\right]}} \sim t_{(n-2)}$$

  Passo 2 — Quantis de probabilidade

$$\begin{aligned} a_\alpha &= -F^{-1}_{t_{(n-2)}}(1-\alpha/2) \\ b_\alpha &= F^{-1}_{t_{(n-2)}}(1-\alpha/2) \end{aligned}$$

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

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

$$P\left\{ a_\alpha \leq \frac{(\hat\beta_0 + \hat\beta_1 x_0) - (\beta_0 + \beta_1 x_0)}{\sqrt{\hat\sigma^2 \times \left[\frac{1}{n} + \frac{\left(x_0 - \frac{\sum_{i=1}^n x_i}{n}\right)^2}{\sum_{i=1}^n x_i^2 - n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}\right]}} \leq b_\alpha \right\} = 1 - \alpha$$

$$P\left\{(\hat{\beta}_0+\hat{\beta}_1 x_0)-F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat{\sigma}^2\times\left[\frac{1}{n}+\frac{\left(x_0-\frac{\sum_{i=1}^n x_i}{n}\right)^2}{\sum_{i=1}^n x_i^2-n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}\right]}\le\beta_0+\beta_1 x_0\le\right.$$

$$\left.(\hat{\beta}_0+\hat{\beta}_1 x_0)+F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat{\sigma}^2\times\left[\frac{1}{n}+\frac{\left(x_0-\frac{\sum_{i=1}^n x_i}{n}\right)^2}{\sum_{i=1}^n x_i^2-n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}\right]}\right\}=1-\alpha$$

Passo 4 — Concretização

$$IC_{(1-\alpha)\times 100\%}(\beta_0+\beta_1 x_0) = \left[(\hat{\beta}_0+\hat{\beta}_1 x_0)\pm F^{-1}_{t_{(n-2)}}(1-\alpha/2)\times\sqrt{\hat{\sigma}^2\times\left[\frac{1}{n}+\frac{\left(x_0-\frac{\sum_{i=1}^n x_i}{n}\right)^2}{\sum_{i=1}^n x_i^2-n\left(\frac{\sum_{i=1}^n x_i}{n}\right)^2}\right]}\right]$$

onde:

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

$\hat{\beta}_0=\bar{y}-\hat{\beta}_1\times\frac{\sum_{i=1}^n x_i}{n}$;

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