DM DEPARTAMENTO DE MATEMÁTICA TÉCNICO LISBOA	**Probabilidades e Estatística** **LEAN, LEE, LEGI, LERC, MEAmbi, MEEC, MEM, MEMec, MEQ**	2º semestre – 2020/2021 18/06/2021 – **9:00**

Duração: 60+15 minutos **Teste 2A**

Justifique convenientemente todas as respostas

1. Um engenheiro mecânico propõe a utilização do seguinte estimador da variância populacional σ^2: $T = \frac{n-1}{n+1} S^2$, onde $S^2 = \frac{1}{n-1}\left(\sum_{i=1}^n X_i^2 - n\bar{X}^2\right)$ é a variância corrigida da amostra aleatória de dimensão n, $(X_1, \ldots, X_n)$. (4.0)

Após ter deduzido $E\left(S^2\right)$, mostre que o enviesamento de T é igual a $c\sigma^2$, onde c é uma constante real. Obtenha o valor de c quando $n = 178$.

- **Estimador de σ^2**

  $S^2 = \frac{1}{n-1}[(\sum_{i=1}^n X_i^2) - n(\bar{X})^2]$, onde $X_i \overset{i.i.d.}{\sim} X$, $i = 1, \ldots, n$

- **Valor esperado de S^2**

  [Ao notarmos que $E(Z^2) = V(Z) + E^2(Z)$, $E(\bar{X}) = E(X) = \mu$ e $V(\bar{X}) = \frac{V(X)}{n} = \frac{\sigma^2}{n}$, segue-se, para qualquer valor positivo de σ^2,]

$$
\begin{aligned}
E(S^2) &= \frac{1}{n-1} E\left[\left(\sum_{i=1}^n X_i^2\right) - n\left(\bar{X}\right)^2\right] \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^n E\left(X_i^2\right) - n E\left[\left(\bar{X}\right)^2\right]\right\} \\
&= \frac{1}{n-1}\left\{\sum_{i=1}^n \left[V(X_i) + E^2(X_i)\right] - n \times \left[V(\bar{X}) + E^2(\bar{X})\right]\right\} \\
&= \frac{1}{n-1}\left[\sum_{i=1}^n (\sigma^2 + \mu^2) - n \times \left(\frac{\sigma^2}{n} + \mu^2\right)\right] \\
&= \frac{1}{n-1}\left(n\sigma^2 + n\mu^2 - \sigma^2 - n\mu^2\right) \\
&= \frac{1}{n-1}(n-1)\sigma^2 \\
&= \sigma^2.
\end{aligned}
$$

  [Logo S^2 é um estimador centrado de σ^2.]

- **Outro estimador de σ^2**

  $T = \frac{n-1}{n+1} S^2$

- **Enviesamento de T e obtenção de c**

$$
\begin{aligned}
E\left[\frac{(n-1)S^2}{n+1}\right] - \sigma^2 &= \frac{n-1}{n+1} E(S^2) - \sigma^2 \\
&= \frac{n-1}{n+1}\sigma^2 - \sigma^2 \\
&= \frac{(n-1)-(n+1)}{n+1}\sigma^2 \\
&= -\frac{2}{n+1}\sigma^2,
\end{aligned}
$$

  i.e., $c = -\frac{2}{179} \simeq -0.0112$.

2. Pretende-se comparar a potência esperada (em cavalos-vapor) de duas marcas de carros, A e B. Para tal foram observados $n_A = 100$ e $n_B = 100$ carros, selecionados ao acaso de entre as marcas A e B respectivamente, tendo-se obtido as seguintes médias e variâncias corrigidas amostrais: $\bar{x}_A = 362$ e $s_A^2 = 31.2$, para os carros da marca A; e $\bar{x}_B = 361$ e $s_B^2 = 21.2$, para os carros da marca B. (4.0)

Considere que as medições obtidas são concretizações de duas amostras aleatórias independentes provenientes de duas populações com distribuições arbitrárias, valores esperados μ_A e μ_B desconhecidos e variâncias σ_A^2 e σ_B^2 desconhecidas.

Obtenha um intervalo de confiança aproximado a 90% para $\mu_A - \mu_B$.

- **V.a. de interesse**

  X_A = potência de carro da marca A

  X_B = potência de carro da marca B

- **Situação**

  X_A e X_B v.a. independentes com distribuições arbitrárias

  $(\mu_A - \mu_B)$ DESCONHECIDO

  σ_A^2 e σ_B^2 desconhecidas

- **Obtenção do IC para $\mu_A - \mu_B$**

  Passo 1 — Seleção da v.a. fulcral para $(\mu_A - \mu_B)$

$$Z = \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \sim^a \text{normal}(0,1)$$

  Passo 2 — Obtenção dos quantis de probabilidade

$$\begin{cases} a_\alpha = \Phi(\alpha/2) = -\Phi(1-\alpha/2) \overset{\alpha=0.1}{=} -\Phi(0.95) \overset{tabelas, calc}{=} -1.6449 \\ b_\alpha = \Phi(1-\alpha/2) = \Phi(0.95) = 1.6449 \end{cases}$$

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

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

$$P\left[a_\alpha \leq \frac{(\bar{X}_A - \bar{X}_B) - (\mu_A - \mu_B)}{\sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}} \leq b_\alpha\right] \simeq 1-\alpha$$

$$P\left[(\bar{X}_A - \bar{X}_B) - \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}} \leq \mu_A - \mu_B \leq (\bar{X}_A - \bar{X}_B) + \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{S_A^2}{n_A} + \frac{S_B^2}{n_B}}\right] \simeq 1-\alpha$$

  Passo 4 — Concretização

  Tendo em conta os quantis acima, as dimensões amostrais n_A e n_B, bem como as concretizações de $\bar{X}_A$, $\bar{X}_B$, S_A^2 e S_B^2 e a expressão geral do IC,

$$IC_{(1-\alpha)\times 100\%}(\mu_A - \mu_B) \simeq \left[(\bar{x}_A - \bar{x}_B) \pm \Phi^{-1}(1-\alpha/2) \times \sqrt{\frac{s_A^2}{n_A} + \frac{s_B^2}{n_B}}\right],$$

  temos

$$
\begin{aligned}
IC_{90\%}(\mu_A-\mu_B) &\simeq \left[(362-361)\pm 1.6449\times\sqrt{\frac{31.2}{100}+\frac{21.2}{100}}\right]\\
&\simeq [1\pm 1.6449\times, 0.723878]\\
&\simeq [1\pm 1.1907]\\
&\simeq [-0.1907, 2.1907].
\end{aligned}
$$

3. As normas de segurança especificam uma probabilidade máxima $p_0 = 0.01$ para a existência de falhas nas asas de uma aeronave após a realização de determinada missão. Em $n = 75$ missões realizadas, foram observadas falhas em zero dessas missões. (4.0)

Obtenha o valor-p aproximado do teste quando são confrontadas as hipóteses $H_0 : p = p_0$ e $H_1 : p > p_0$, onde p representa a probabilidade desconhecida de existência de falhas nas asas de uma aeronave selecionada ao acaso após a realização de uma missão. Decida com base no valor-p aproximado que obteve.

- **V.a. de interesse**

  $$X = \begin{cases} 1, & \text{se existem falhas nas asas da aeronave após a realização de uma missão} \\ 0, & \text{caso contrário} \end{cases}$$

- **Situação**

  $X \sim \text{Bernoulli}(p)$

  p DESCONHECIDO

- **Hipóteses**

  $H_0 : p = p_0 = 0.01$

  $H_1 : p > p_0$

- **Estatística de teste**

  $$T = \frac{\bar{X}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \overset{a}{\sim}_{H_0} \text{normal}(0,1)$$

- **Região de rejeição de H_0**

  Teste unilateral superior, logo a região de rejeição de H_0 é do tipo $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**

  Atendendo a que o valor observado da estatística de teste é igual a

  $$
  \begin{aligned}
  t &= \frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\\
  &= \frac{\frac{a}{n}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\\
  &= \frac{\frac{0}{75}-0.01}{\sqrt{\frac{0.01\times(1-0.01)}{75}}}\\
  &= -0.870388
  \end{aligned}
  $$

  e ao tipo de região de rejeição de H_0, temos

$$
\begin{aligned}
valor-p &= P(T > t \mid H_0) \\
&\simeq 1-\Phi(t) \\
&\simeq 1-\Phi(-0.87) \\
&= \Phi(0.87) \\
&\overset{tabelas,calc.}{=} 0.8078.
\end{aligned}
$$

Por consequência, devemos:

- não rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 80.78\%$, designadamente a qualquer dos n.u.s. $(1\%, 5\%, 10\%)$;
- rejeitar H_0 a qualquer n.s. $\alpha_0 > 80.78\%$.

4. Admita que a variável aleatória X representa a distância (em milhares de km) percorrida por um veículo até à falha de determinada componente usada nesse mesmo veículo. Ao avaliar a fiabilidade dessa componente, uma engenheira conjeturou a hipótese H_0 de que X possui função de distribuição $F_0(x) = 1 - \exp\left[-\left(\frac{x}{254}\right)^2\right]$, para $x \geq 0$. (4.0)

A concretização de uma amostra aleatória de dimensão $n = 300$ proveniente da população X conduziu à seguinte tabela de frequências:

Classe	$[0,100]$	$]100,200]$	$]200,300]$	$]300,400]$	$]400,\infty[$
Frequência absoluta observada	50	115	84	42	o_5
Frequência absoluta esperada sob H_0	E_1	95.5	87.0	49.2	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 décimas), averigue se H_0 é consistente com este conjunto de dados. Decida com base no valor-p aproximado.

- **V.a. de interesse**

  X = distância (em milhares de km) percorrida por um veículo até à falha da componente

- **Hipóteses**

  $H_0 : F_X(x) = F_0(x), \quad x \in \mathbb{R}$

  $H_1 : \neg H_0$

- **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;
- O_i = freq. abs. observável da classe i;
- E_i = freq. abs. esperada sob H_0 da classe i;
- $\beta = 0$.

- **Frequência absolutas observada e esperadas sob H_0 omissas**

$$
\begin{aligned}
o_5 &= n - \sum_{i=1}^{4} o_i \\
&= 300 - (50 + 115 + 84 + 42) \\
&= 9 \\
E_1 &= n \times P(X \leq 100 \mid H_0) \\
&= n \times F_0(100) \\
&= 300 \times \left\{ 1 - \exp\left[-\left(\frac{100}{254}\right)^2 \right] \right\} \\
&\simeq 43.1 \\
E_5 &= n - \sum_{i=1}^{4} E_i \\
&\simeq 300 - (43.1 + 95.5 + 87.0 + 49.2) \\
&= 25.2.
\end{aligned}
$$

- **Região de rejeição de H_0** (para valores de T)

  Tratando-se de um teste de ajustamento do qui-quadrado, a região de rejeição de H_0 escrita para valores observados de T é o intervalo à direita $W = (c, +\infty)$.

- **Decisão (com base no valor-p)**

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esp. sob H_0 E_i	Parcelas valor obs. estat. teste $\frac{(o_i - e_i)^2}{e_i}$
1	$[0, 100]$	50	43.1	$\frac{(50-43.1)^2}{43.1} = 1.1046$
2	$]100, 200]$	115	95.5	3.9817
3	$]200, 300]$	84	87.0	0.1034
4	$]300, 400]$	42	49.2	1.0537
5	$]400, \infty[$	9	25.2	10.4143
		$\sum_{i=1}^{k} o_i = n$ $= 300$	$\sum_{i=1}^{k} e_i = n$ $= 300$	$t = \sum_{i=1}^{k} \frac{(o_i - e_i)^2}{e_i}$ $= 16.6577$

Dado que o valor observado da estatística de teste é $t = 16.6577$ e $W = (c, +\infty)$, obtemos

$$
\begin{aligned}
valor - p &= P(T > t \mid H_0) \\
&\simeq 1 - F_{\chi^2_{(k-1)}}(t) \\
&= 1 - F_{\chi^2_{(4)}}(16.6577) \\
&\simeq 0.002252
\end{aligned}
$$

e devemos

- não rejeitar de H_0 a qualquer n.s. $\alpha_0 \leq valor - p = 0.2252\%$;
- rejeitar de H_0 a qualquer n.s. $\alpha_0 > valor - p = 0.2252\%$, nomeadamente aos n.u.s. $(1\%, 5\%, 10\%)$.

Alternativamente e recorrendo às tabelas de quantis da distribuição do qui-quadrado, podemos obter um intervalo para o valor-p deste teste:

$$
\begin{aligned}
F^{-1}_{\chi^2_{(4)}}(0.995) = 14.86 &< t = 16.6577 < 18.47 = F^{-1}_{\chi^2_{(4)}}(0.999) \\
0.995 &< F_{\chi^2_{(4)}}(16.6577) < 0.999
\end{aligned}
$$

$$0.001 = 1 - 0.999 \quad < \quad valor - p \simeq 1 - F_{\chi^2_{(4)}}(16.6577) < 1 - 0.995 = 0.005.$$

Assim, podemos adiantar que:

- não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.1\%$;
- devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 0.5\%$, em particular aos n.u.s. $(1\%, 5\%, 10\%)$.

5. Por forma a estudar a relação entre a temperatura da superfície das estradas (x, em graus Fahrenheit) e a deflexão dos pavimentos (Y) em determinada região, foi obtido o seguinte conjunto de resultados referentes a $n = 10$ observações casuais: (4.0)

$$\sum_{i=1}^n x_i = 743, \quad \sum_{i=1}^n x_i^2 = 57\,125, \quad \sum_{i=1}^n y_i = 6.15, \quad \sum_{i=1}^n y_i^2 = 3.8945, \quad \sum_{i=1}^n x_i y_i = 465.69$$

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

Após ter enunciado as hipóteses de trabalho que entender convenientes, obtenha o intervalo de confiança a 99% para β_0 e a amplitude deste intervalo de confiança, tirando partido dos resultados obtidos.

- **Modelo de RLS**

  Y = deflexão do pavimento (v.a. resposta)

  x = temperatura da superfície da estrada (variável explicativa)

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

- **Hipóteses de trabalho**

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

- **Estimativas de MV de β_0 e β_1; estimativa de σ^2**

  Importa notar que

  - $n = 10$
  - $\sum_{i=1}^n x_i = 743$

    $\bar{x} = \frac{1}{n}\sum_{i=1}^n x_i = \frac{743}{10} = 74.3$

    $\sum_{i=1}^n x_i^2 = 57\,125$

    $\sum_{i=1}^n x_i^2 - n\,\bar{x}^2 = 57\,125 - 10 \times 74.3^2 = 1920.1$
  - $\sum_{i=1}^n y_i = 6.15$

    $\bar{y} = \frac{1}{n}\sum_{i=1}^n y_i = \frac{6.15}{10} = 0.615$

    $\sum_{i=1}^n y_i^2 = 3.8945$

    $\sum_{i=1}^n y_i^2 - n\,\bar{y}^2 = 3.8945 - 10 \times 0.615^2 = 0.112250$
  - $\sum_{i=1}^n x_i y_i = 465.69$

    $\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y} = 465.69 - 10 \times 74.3 \times 0.615 = 8.745.$

  Logo,

$$\hat{\beta}_1 \quad = \quad \frac{\sum_{i=1}^n x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2} = \frac{8.745}{1920.1} \simeq 0.004554$$

$$
\begin{aligned}
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \bar{x} \\
&= 0.615 - 0.004554 \times 74.3 \\
&\simeq 0.276638 \\
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right) - \left(\hat{\beta}_1\right)^2 \left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right] \\
&\simeq \frac{1}{10-2}\left(0.112250 - 0.004554^2 \times 1920.1\right) \\
&\simeq 0.009054
\end{aligned}
$$

- **Obtenção do IC para β_0**

Passo 1 — Seleção da v.a. fulcral

$$
Z = \frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\right)}} \sim t_{(n-2)}
$$

Passo 2 — Obtenção dos quantis de probabilidade

$$
\begin{cases}
a_\alpha = F_{t_{(n-2)}}(\alpha/2) = -F_{t_{(10-2)}}(1-0.01/2) = -F_{t_{(8)}}(0.995) \stackrel{tabelas,calc.}{=} -3.355 \\
b_\alpha = F_{t_{(n-2)}}(1-\alpha/2) = F_{t_{(8)}}(0.995) \stackrel{tabelas,calc.}{=} 3.355
\end{cases}
$$

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

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

$$
P\left[a_\alpha \le (\hat{\beta}_0 - \beta_0)/\sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\right)} \le b_\alpha\right] = 1-\alpha
$$

$$
P\left[\hat{\beta}_0 - b_\alpha \times \sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\right)} \le \beta_0 \le \hat{\beta}_0 - a_\alpha \times \sqrt{\hat{\sigma}^2 \times \left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}\right)}\right] = 1-\alpha
$$

Passo 4 — Concretização

Tendo em conta a expressão geral do IC para β_0,

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

e os resultados anteriores, o IC pretendido e a sua amplitude são iguais a

$$
\begin{aligned}
IC_{99\%}(\beta_0) &\simeq \left[0.276638 \pm 3.355 \times \sqrt{0.009054 \times \left(\frac{1}{10} + \frac{74.3^2}{1920.1}\right)}\right] \\
&\simeq [0.276638 \pm 3.355 \times 0.164124] \\
&\simeq [-0.273997, 0.827337] \\
0.827337 - (-0.273997) &= 1.101334.
\end{aligned}
$$