Departamento de Matemática

Probabilidades e Estatística

LEMat, LMAC, MEAmbi, MEBiol, MEBiom, MEEC, MEFT, MEQ LEAN, MEAer, MEMec

1º semestre – 2011/2012 — 2º Teste (Época Normal)

7/01/2012 – **8:30** — Duração: 1 hora e 30 minutos

Grupo I — **1.0** + **1.0** + **3.0** + **3.0** + **2.0** = **10.0** valores

Exercício 1

(a)

- **V.a. de interesse**

  $X \sim \text{Binomial}(n, p)$, $p \in [0,1]$

- **A.a.**

  X_1 a.a. de dimensão unitária proveniente da população X

  $X_1 \sim X$

- **Parâmetro** DESCONHECIDO

  p

- **Estimador de** p

  $T = \frac{X_1 + \frac{\sqrt{n}}{2}}{n+\sqrt{n}}$

- **Estimador centrado de** p ?

  T é um estimador centrado de p sse $E(T) = p, \forall p \in [0,1]$. Ora, $E(X_1) = E(X) = np$ donde

$$
\begin{aligned}
E(T) &= E\left(\frac{X_1 + \frac{\sqrt{n}}{2}}{n+\sqrt{n}}\right) \\
&= \frac{E(X_1) + \frac{\sqrt{n}}{2}}{n+\sqrt{n}} \\
&= \frac{np + \frac{\sqrt{n}}{2}}{n+\sqrt{n}} \\
&\neq p,
\end{aligned}
$$

  para algum $p \in [0,1]$. Assim, conclui-se que T não um estimador centrado de p.

- **Obs.**

  É curioso notar que, para $p = \frac{1}{2}$, $E(T) = \frac{\frac{1}{2}+\frac{\sqrt{n}}{2}}{n+\sqrt{n}} = \frac{1}{2}$. Pode então afirmar-se que $E(T) \neq p, \forall p \in]0,1[\backslash\{\frac{1}{2}\}$.

(b)

- **Outro resultado**

  Dado que $V(X_1) = V(X) = np(1-p)$ tem-se

$$
\begin{aligned}
\lim_{n\to+\infty} V(T) &= \lim_{n\to+\infty} V\left(\frac{X_1 + \frac{\sqrt{n}}{2}}{n+\sqrt{n}}\right) \\
&= \lim_{n\to+\infty} \frac{V(X_1)}{(n+\sqrt{n})^2} \\
&= \lim_{n\to+\infty} \frac{np(1-p)}{(n+\sqrt{n})^2} \\
&= \lim_{n\to+\infty} \frac{p(1-p)}{n+2\sqrt{n}+1} \\
&= 0.
\end{aligned}
$$

Exercício 2

(a)

- **V.a. de interesse**
  X = pressão de trabalho de peça de aço comum
  Y = pressão de trabalho de peça de aço inoxidável

- **Situação**
  $X \sim \text{Normal}(\mu_X, \sigma_X^2) \perp\!\!\!\perp Y \sim \text{Normal}(\mu_Y, \sigma_Y^2)$
  $(\mu_X - \mu_Y)$ DESCONHECIDO
  $\sigma_X = 4.0$, $\sigma_Y = 5.0$
  $n_X = 20$, $n_Y = 25$

- **Passo 1 — Selecção da v.a. fulcral para** $(\mu_X - \mu_Y)$

$$Z = \frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim \text{Normal}(0, 1)$$

  uma vez que é suposto determinar um IC para a diferença de valores esperados de duas populações normais independentes, com variâncias conhecidas.

- **Passo 2 — Obtenção dos quantis de probabilidade**
  Ao ter em conta que, neste caso, $(1 - \alpha) \times 100\% = 90\%$ i.e., $\alpha = 0.1$, usaremos os quantis de probabilidade:

$$\begin{cases} a_\alpha = -\Phi^{-1}(1 - \alpha/2) = -\Phi^{-1}(0.95) \stackrel{tabela}{=} -1.6449 \\ b_\alpha = \Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.95) = 1.6449. \end{cases}$$

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

$$P\left(a_\alpha \leq Z \leq b_\alpha\right) = 1 - \alpha$$
$$\ldots$$
$$P\left[(\bar{X} - \bar{Y}) - \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}} \leq \mu_X - \mu_Y \leq (\bar{X} - \bar{Y}) + \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right] = 1 - \alpha.$$

- **Passo 4 — Concretização**
  Atendendo a que $\Phi^{-1}(1 - \alpha/2) = \Phi^{-1}(0.95) \stackrel{tabela}{=} 1.6449$

$$n_X = 20$$
$$\sum_{i=1}^{n_X} x_i = 596$$
$$\bar{x} = \frac{1}{n_X} \sum_{i=1}^{n_X} x_i = \frac{596}{20} = 29.8$$

$$n_Y = 25$$
$$\sum_{i=1}^{n_Y} y_i = 867.5$$
$$\bar{y} = \frac{1}{n_Y} \sum_{i=1}^{n_Y} y_i = \frac{867.5}{25} = 34.7$$

  e

$$IC_{(1-\alpha)\times 100\%}(\mu_X - \mu_Y) = \left[(\bar{x} - \bar{y}) \pm \Phi^{-1}(1 - \alpha/2) \times \sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}\right],$$

  segue-se:

$$\begin{aligned} IC_{90\%}(\mu_X - \mu_Y) &= \left[(29.8 - 34.7) \pm 1.6449 \times \sqrt{\frac{4.0^2}{20} + \frac{5.0^2}{25}}\right] \\ &= [-4.9 \pm 1.6449 \times 1.341641] \\ &= [-7.106865,\ -2.693135]. \end{aligned}$$

(b)
- **Hipóteses**
  $H_0 : \mu_X - \mu_Y = \mu_0 = 0$
  $H_1 : \mu_X - \mu_Y \neq \mu_0$

- **Nível de significância**
  $\alpha = 5\%$

- **Estatística de teste**

  $$T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \sim_{H_0} \text{Normal}(0,1)$$

  dado que se pretende efectuar um teste sobre a diferença de valores esperados de duas populações normais, com variâncias conhecidas.

- **Região de rejeição de** H_0 (para valores da estatística de teste)
  Estamos a lidar com um teste bilateral ($H_1 : \mu_X - \mu_Y \neq \mu_0$) e com uma estatística de teste com f.d.p. simétrica em relação à origem, logo a região de rejeição de H_0 (para valores da estatística de teste) é uma reunião de intervalos do tipo

  $$W = (-\infty, -c) \cup (c, +\infty),$$

  onde $c : P(\text{Rejeitar } H_0 | H_0) = \alpha_0$, i.e.,

  $$c = \Phi^{-1}(1 - \alpha_0/2) = \Phi^{-1}(0.975) \overset{tabela}{=} 1.96.$$

- **Decisão**
  O valor observado da estatística de teste é igual a

  $$\begin{aligned} t &= \frac{(\bar{x} - \bar{y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \\ &= \frac{(29.8 - 34.7) - 0}{\sqrt{\frac{4.0^2}{20} + \frac{5.0^2}{25}}} \\ &\simeq -3.652244. \end{aligned}$$

  Como $t = -3.652244 \in W = (-\infty, -1.96) \cup (1.96, +\infty)$, devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 5\%$.

(c)
- **Probabilidade pedida**
  Pretende calcular-se o valor da função potência no ponto $\mu_X - \mu_Y = 1$. Ora, é sabido que $\frac{(\bar{X}-\bar{Y})-(\mu_X-\mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}} \sim \text{Normal}(0,1)$, donde se segue:

  $$\begin{aligned} &P\left[\text{Rejeitar } H_0 | \mu_X - \mu_Y = 1\right] \\ &= P\left[T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} < -1.96 \text{ ou } T = \frac{(\bar{X} - \bar{Y}) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} > 1.96 \;\middle|\; \mu_X - \mu_Y = 1\right] \\ &= P\left[\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y) + [(\mu_X - \mu_Y) - \mu_0]}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} < -1.96 \text{ ou}\right. \\ &\qquad \left.\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y) + [(\mu_X - \mu_Y) - \mu_0]}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} > 1.96 \;\middle|\; \mu_X - \mu_Y = 1\right] \\ &= P\left[\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} < -1.96 - \frac{(\mu_X - \mu_Y) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \text{ ou}\right. \\ &\qquad \left.\frac{(\bar{X} - \bar{Y}) - (\mu_X - \mu_Y)}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} > 1.96 - \frac{(\mu_X - \mu_Y) - \mu_0}{\sqrt{\frac{\sigma_X^2}{n_X} + \frac{\sigma_Y^2}{n_Y}}} \;\middle|\; \mu_X - \mu_Y = 1\right] \end{aligned}$$

$$
\begin{aligned}
&= \quad \Phi\left[-1.96-\frac{(\mu_X-\mu_Y)-\mu_0}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}\right]+1-\Phi\left[1.96-\frac{(\mu_X-\mu_Y)-\mu_0}{\sqrt{\frac{\sigma_X^2}{n_X}+\frac{\sigma_Y^2}{n_Y}}}\right] \\
&= \quad \Phi\left(-1.96-\frac{1-0}{\sqrt{\frac{4.0^2}{20}+\frac{5.0^2}{25}}}\right)+1-\Phi\left(1.96-\frac{(1-0)}{\sqrt{\frac{4.0^2}{n_X}+\frac{5.0^2}{25}}}\right) \\
&\simeq \quad \Phi(-2.71)+[1-\Phi(1.21)] \\
&= \quad [1-\Phi(2.71)]+[1-\Phi(1.21)] \\
&\overset{tabela}{=} \quad (1-0.9966)+(1-0.8869) \\
&= \quad 0.1165.
\end{aligned}
$$

Grupo II	**4.0 + 3.0 + 3.0 = 10.0** valores

Exercício 1

(a)

- **V.a. de interesse**
  $X =$ Indicador da cor do telemóvel

- **Hipóteses**
  Seja $p_i = P(X = i)$, $i = 1, \ldots, 5$. Se não houver preferência por qualquer cor $p_i = p_i^0 = \frac{1}{5}$, $i = 1, \ldots, 5$. Assim sendo, confrontar-se-ão as seguintes hipóteses:
  $H_0 : p_i = p_i^0 = \frac{1}{5}$, $i = 1, \ldots, 5$
  $H_1 : \exists i : p_i \neq p_i^0$
  ou
  $H_0 : X \sim$ Uniforme Discreta$(\{1, 2, 3, 4, 5\})$
  $H_0 : X \not\sim$ Uniforme Discreta$(\{1, 2, 3, 4, 5\})$

- **Estatística de Teste**
  $$T = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i} \ \overset{a}{\sim}_{H_0} \ \chi^2_{(k-\beta-1)},$$
  onde:
  $k =$ No. de classes $= 5$
  $O_i =$ Frequência absoluta observável da classe i
  $E_i =$ Frequência absoluta esperada, sob H_0, da classe i
  $\beta =$ No. de parâmetros a estimar $= 0.$[1]

- **Região de rejeição de H_0** (para valores de T)
  A região de rejeição de H_0 escrita para valores de T é um intervalo à direita $W = (c, +\infty)$.

- **Frequências absolutas esperadas sob H_0**
  Atendendo a que, sob H_0, $p_i = p_i^0 = \frac{1}{5}$, $i = 1, \ldots, 5$, tem-se $E_i = n \times p_i^0 = 300 \times \frac{1}{5} = 60$, $i = 1, \ldots, 5.$[2]

- **Decisão**
  No cálculo do valor observado da estatística de teste convém adiantar a seguinte tabela auxiliar:

[1]Dado que a distribuição conjecturada em H_0 está completamente especificada, i.e., H_0 é uma hipótese simples.
[2] Notar que não é necessário qualquer agrupamento de classes uma vez que $E_i \geq 1$,$\forall i$ e $E_i \geq 5$ em pelo menos 80% das classes.

i	Classe i	Freq. abs. obs. o_i	Freq. abs. esper. sob H_0 $E_i = n \times p_i^0$	Parcelas valor obs. estat. teste $\frac{(o_i - E_i)^2}{E_i}$
1	{1}	40	60	$\frac{(40-60)^2}{60} \simeq 6.667$
2	{2}	55	60	0.417
3	{3}	65	60	0.417
4	{4}	88	60	13.067
5	{5}	52	60	1.067
		$\sum_{i=1}^k o_i = n = 300$	$\sum_{i=1}^k E_i = n = 300$	$t = \sum_{i=1}^k \frac{(o_i - E_i)^2}{E_i} \simeq 21.635$

- **Decisão (cont.) e intervalo para o p-value**

  Dado que a região de rejeição deste teste é um intervalo à direita, tem-se:

$$\begin{aligned} p - value &= P(T > t \mid H_0) \\ &= P(T > 21.635 \mid H_0) \\ &\simeq 1 - F_{\chi^2_{(5-1-0)}}(21.635) \end{aligned}$$

  Recorrendo às tabelas de quantis da distribuição do qui-quadrado podemos adiantar um intervalo para o *p-value*. Com efeito,

$$\begin{aligned} F^{-1}_{\chi^2_{(4)}}(0.9995) = 20.00 &< 21.635 \\ 0.9995 &< F_{\chi^2_{(4)}}(21.635) \\ 0.0005 = 1 - 0.9995 &> 1 - F_{\chi^2_{(4)}}(21.635). \end{aligned}$$

  Deste modo, o intervalo aproximado para o *p-value* é $(0, 0.0005)$, pelo que devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \geq 0.05\%$, nomeadamente aos níveis usuais de significância de 1%, 5% e 10%.

- **Decisão (cont.) e obtenção do p-value usando máquina de calcular**

  Uma vez que a região de rejeição deste teste é um intervalo à direita, tem-se:

$$\begin{aligned} p - value &= P(T > t \mid H_0) \\ &= P(T > 21.635 \mid H_0) \\ &\simeq 1 - F_{\chi^2_{(4)}}(21.635) \\ &\simeq 0.000237, \end{aligned}$$

  pelo que

  - não devemos rejeitar H_0 a qualquer n.s. $\alpha_0 \leq 0.0237\%$;
  - rejeitar H_0 a qualquer n.s. $\alpha_0 > 0.0237\%$, nomeadamente aos níveis usuais de significância de 1%, 5% e 10%.

Exercício 2

(a)

- **[Modelo de RLS**

  $Y_i = \beta_0 + \beta_1 x_i + \epsilon_i$

  $Y_i =$ resultado da medição da $i-$ésima concentração conhecida

  $x_i =$ valor da $i-$ésima concentração conhecida

  $\epsilon_i =$ erro aleatório associado à medição da $i-$ésima concentração conhecida

- **Hipóteses de trabalho**

  $\epsilon_i \sim_{i.i.d.}$ Normal$(0, \sigma^2)$, $i = 1, \ldots, n$ (hipótese de trabalho)

  $\beta_0, \beta_1, \sigma^2$ DESCONHECIDOS]

- **Estimativas de β_0 e β_1**

  Dado que

  $n = 12$

$\bar{x} = 5.5$
$\sum_{i=1}^{n} x_i^2 = 498$
$\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2 = 498 - 12 \times 5.5^2 = 135$

$\bar{y} = 5.492$
$\sum_{i=1}^{n} y_i^2 = 501.846$
$\sum_{i=1}^{n} y_i^2 - n\,(\bar{y})^2 = 501.846 - 12 \times 5.492^2 = 139.901232$

$\sum_{i=1}^{n} x_i y_i = 499.767$
$\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = 499.767 - 12 \times 5.5 \times 5.492 = 137.295,$

as estimativas de β_1 e β_0 são, para este modelo, iguais a:

$$\begin{aligned}
\hat{\beta}_1 &= \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2} \\
&= \frac{137.295}{135} \\
&= 1.017 \\
\hat{\beta}_0 &= \bar{y} - \hat{\beta}_1 \times \bar{x} \\
&= 5.492 - 1.017 \times 5.5 \\
&= -0.1015
\end{aligned}$$

- **Estimativa de σ^2**

  Atendendo ao facto de $V(\epsilon) = \sigma^2$ e ao formulário, tem-se:

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n-2}\left[\left(\sum_{i=1}^{n} y_i^2 - n\,\bar{y}^2\right) - (\hat{\beta}_1)^2\left(\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2\right)\right] \\
&= \frac{1}{12-2}\left(139.901232 - 1.017^2 \times 135\right) \\
&= 0.027221.
\end{aligned}$$

(b)

- **Hipóteses**

  $H_0 : \beta_1 = \beta_{1,0} = 1$ vs. $H_1 : \beta_1 \neq \beta_{1,0}$

- **Nível de significância**

  $\alpha_0 = 5\%$

- **Estatística de teste**

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

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

  Estamos a lidar com um teste bilateral ($H_1 : \beta_1 \neq \beta_{1,0}$), pelo que a região de rejeição de H_0 é $W = (-\infty, -c) \cup (c, +\infty)$, onde

$$\begin{aligned}
c \;:\;\; & P(\text{Rejeitar } H_0 | H_0) = \alpha_0 \\
& c = F^{-1}_{t_{(n-2)}}\left(1 - \frac{\alpha_0}{2}\right) \\
& c = F^{-1}_{t_{(10)}}(0.975) \\
& c = 2.228
\end{aligned}$$

- **Decisão**

Tendo em conta resultados da alínea a), o valor observado da estatística de teste é dado por

$$
\begin{aligned}
t &= \frac{\hat{\beta}_1 - \beta_{1,0}}{\sqrt{\frac{\hat{\sigma}^2}{\sum_{i=1}^n x_i^2 - n\,\bar{x}^2}}} \\
&= \frac{1.017 - 1}{\sqrt{\frac{0.027221}{135}}} \\
&= 1.197177
\end{aligned}
$$

Como $t = 1.197177 \not\in W = (-\infty, -2.228) \cup (2.228, +\infty)$, não devemos rejeitar H_0 ao n.s. $\alpha_0 = 5\%$, assim como para n.s. inferiores a 5%.