Universidade Federal do Rio de Janeiro - UFRJ
$2^{\underline{a}}$ Prova de Estatística Unificada

Turma: Engenharia — Data: 08/12/2011

---

1. Um levantamento obtido, junto aos funcionários de um pequeno escritório, busca relacionar as variáveis: *tempo de carreira (X)* e *número de diferentes empregos nos últimos 5 anos (Y)*.

X	8	9	10	11	12
Y	3	2	2	2	1

   a) calcule a covariância entre X e Y;

   b) estime os coeficientes α e β da reta de regressão simples ($y_i = \alpha + \beta x_i + \varepsilon_i$).

2. O tempo de vida X (em horas) de um tipo de componente eletrônico fabricado por certa empresa segue distribuição conforme dada abaixo: ($\beta > 0$ constante)

$$f(x\,;\,\beta) = \begin{cases} \beta^2\, x\, e^{-\beta x} & , \text{ se } \ x \geq 0, \\ 0 & , \text{ se } \ x < 0, \end{cases}$$

   Tomada uma amostra aleatória de 100 componentes deste tipo, e denotando por X_i o tempo de vida do i-ésimo componente na amostra:

   a) obtenha o estimador de máxima verossimilhança do parâmetro β;

   b) obtenha o estimador de máxima verossimilhança de $\mu = 2/\beta$.

3. Pesquisadores desejam estudar o tempo gasto por engenheiros para executarem determinada tarefa. Para isto, foram selecionados aleatoriamente 64 engenheiros. Observou-se, nessa amostra, que a soma do tempo gasto por eles foi de 192 horas enquanto a soma dos quadrados dos tempos foi de 828 horas^2.

   a) Determine o intervalo de 95% de confiança para a média populacional do tempo gasto pelos engenheiros para resolverem a tarefa. Justifique.

   b) Sabendo-se que, dentre os 64 engenheiros selecionados, 16 eram recém formados, determine o intervalo de 99% de confiança (conservativo) para a proporção populacional de engenheiros recém-formados.

4. Em uma linha de produção, discos devem ser fabricados com no máximo 48 mm de diâmetro. Sabemos que a distribuição dos diâmetros dos discos segue um modelo Normal. Uma amostra aleatória de 16 discos é analisada, e para esta obtemos uma média de 49,301 mm de diâmetro e variância de 4(mm^2). Com base nesta amostra:

   a) teste a hipótese da média dos diâmetros dos discos produzidos ser menor ou igual a 48mm contra a média dos diâmetros dos discos produzidos ser maior que 48mm, ao nível de 5% de significância.

   b) obtenha o p-valor deste teste.

5. Uma montadora de computadores foi informada de que uma nova bateria com tempo de vida superior as atuais foi lançada no mercado. Com o objetivo de testar a veracidade da informação, um engenheiro responsável pela linha de produção realiza 25 ensaios com a bateria atual obtendo uma vida média $\bar{x}$ (horas), e 16 ensaios com a nova bateria, obtendo uma vida média $\bar{y}$ (horas).

   a) Supondo que em ambos os casos o tempo de vida seja uma variável aleatória com distribuição Normal com mesmo desvio-padrão $\sigma = 0,3$, elabore um teste de hipótese para a veracidade da informação, explicitando a estatística de teste e determinando a região de rejeição ao nível de $\alpha = 0,005$.

   b) Se $\bar{x} = 8,2$ e $\bar{y} = 8,5$ a que conclusão se chegaria?
   Obs. use a aproximação $\sqrt{41 \times (0,3)^2/(25 \times 16)} \approx 0,1$

Respostas

1. (a) $\overline{x} = 10$; $\overline{y} = 2$; $s_x^2 = 2,5$; $s_y^2 = 0,5$ e $\text{cov}(x, y) = -1,0$.

   (b) $\hat{\alpha} = 8$ e $\hat{\beta} = -0,6$.

2. (a) $L(\beta) = \prod_{i=1}^{100} \beta^2\, x_i\, \exp\{-\beta x_i\} = \beta^{200} \left(\prod_{i=1}^{100} x_i\right) \exp\left\{-\beta \sum_{i=1}^{100} x_i\right\}$.

   $\ln L(\beta) = 200 \ln \beta + \sum_{i=1}^{100} \ln x_i - \beta \sum_{i=1}^{100} x_i$. $\frac{d}{d\beta} \ln L(\beta) = \frac{200}{\beta} - \sum_{i=1}^{100} x_i = 0 \Leftrightarrow \beta = \frac{200}{\sum_{i=1}^{100} x_i} = \frac{2}{\bar{x}_{100}}$

   $\Rightarrow EMV(\beta) = \frac{200}{\sum_{i=1}^{100} X_i} = \frac{2}{\overline{X}_{100}}$.

   (b) $\mu = \frac{2}{\beta}$

   $\Rightarrow \text{EMV }(\mu) = \frac{2}{EMV(\beta)} = \frac{2}{(2 \times 100)/\sum_{i=1}^{100} X_i} = \frac{\sum_{i=1}^{100} X_i}{100} = \overline{X}_{100}$.

3. (a) Temos que $n = 64$, $\sum_{i=1}^{64} x_i = 192$ e $\sum_{i=1}^{64} x_i^2 = 828$.
   O intervalo de 95% de confiança para a média populacional μ é da forma

   $$\left(\overline{x} - t_{0,975}\frac{S}{\sqrt{n}}\, ; \overline{x} + t_{0,975}\frac{S}{\sqrt{n}}\right),$$

   em que $t_{0,975}$ é o quantil $0,975$ da distribuição t de Student com 63 graus de liberdade,

   $$\overline{x} = \frac{\sum_{i=1}^{64}}{64} = \frac{192}{64} = 3$$

   e

   $$S = \sqrt{\frac{\sum_{i=1}^{64} x_i^2 - n\overline{x}^2}{n-1}} = \sqrt{\frac{828 - 64 \times 3^2}{63}} = \sqrt{4} = 2.$$

   Como o tamanho da amostra, n, é grande, a distribuição t de Student se aproxima da distribuição normal padrão. Assim, o intervalo de 95% de confiança para a média pode ser também

   $$\left(\overline{x} - z_{0,975}\frac{S}{\sqrt{n}}; \overline{x} + z_{0,975}\frac{S}{\sqrt{n}}\right) = \left(3 - 1,96 \times \frac{2}{8}; 3 + 1,96 \times \frac{2}{8}\right) = (2,51;\, 3,49)$$

   em que $z_{0,975} = 1,96$ é o quantil $0,975$ da distribuição normal padrão.
   Podemos dizer que, com 95% de confiança, o intervalo acima compreende o verdadeiro valor da média populacional, μ.

   (b) Temos que $\widehat{p} = 16/64 = 1/4 = 0,25$.
   Como o tamanho da amostra, n, é grande, podemos obter um intervalo de 99% de confiança para a proporção populacional, p, de duas maneiras.

   Intervalo conservartivo:

   $$\begin{aligned}\left(\widehat{p} - z_{0,995}\frac{1}{2\sqrt{n}}; \widehat{p} + z_{0,995}\frac{1}{2\sqrt{n}}\right) &= \left(0,25 - 2,57 \times \frac{1}{2 \times 8}; 0,25 + 2,57 \times \frac{1}{2 \times 8}\right) \\ &= (0,0894;\, 0,4106)\,.\end{aligned}$$

   Intervalo não-conservartivo:

   $$\begin{aligned}\left(\widehat{p} - z_{0,995}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}; \widehat{p} + z_{0,995}\sqrt{\frac{\widehat{p}(1-\widehat{p})}{n}}\right) &= \left(0,25 - 2,57\sqrt{\frac{0,25 \times 0,75}{64}}; 0,25 + 2,57\sqrt{\frac{0,25 \times 0,75}{64}}\right) \\ &= (0,1109;\, 0,3891)\,,\end{aligned}$$

   sabendo que $\sqrt{0,25 \times 0,75} = 0,4330$.
   Aqui, $z_{0,995}$ é o quantil $0,995$ da distribuição normal padrão.

4. $n = 16$; $\bar{x} = 49,301$; $s^2 = 4$.

   (a) (1) $H_0 : \mu \leq 48$ (ou $H_0 : \mu = 48$)
   $H_1 : \mu > 48$.

   (2) $\alpha = 0,05$.

   (3) Estatística de teste: $T = \dfrac{\overline{X} - 48}{\sqrt{4/16}} = \dfrac{\overline{X} - 48}{1/2}$.

   (4) Sob H_0, T tem distribuição t de Student com 15 graus de liberdade.

   (5) Região crítica: $R = [t_{0,95;15}\,;\,\infty) = [1,753\,;\,\infty)$.

   (6) $t = \dfrac{49,301 - 48}{1/2} = 2,602 \in R \Rightarrow$ Rejeitamos H_0.

   OBS.: Os passos (3) - (6) também podem ser feitos da seguinte maneira:

   (3') Estatística de teste: $\overline{X}$.

   (4') Sob H_0, $\overline{X} \sim N(48\,;\,\sigma^2/16)$.

   (5') Região crítica: $R = [48 + t_{0,95;15}\sqrt{\frac{s^2}{n}}\,;\,\infty) = [48,8765\,;\,\infty)$.

   (6') $\bar{x} = 49,301 \in R \Rightarrow$ Rejeitamos H_0.

   (b) $\tilde{\alpha} = P(T \in [2,602\,;\,\infty)) = P(T > 2,602) = 1 - P(T \leq 2,602)$, onde T é distribuído conforme H_0. Logo, pela tabela da distribuição t de Student, concluímos que
   $\tilde{\alpha} = 1 - 0,99 = 0,01$.

5. Informações do exercício:

$$\begin{array}{lll} n_X = 25 & X_i \overset{iid}{\sim} N(\mu_X;\,(0,3)^2), & i = 1,\ldots,25 \\ n_Y = 16 & Y_j \overset{iid}{\sim} N(\mu_Y;\,(0,3)^2), & j = 1,\ldots,16 \end{array}$$

   (a) Hipóteses: $H_0 :\ \mu_X \geq \mu_Y$ versus $H_A :\ \mu_X < \mu_Y$.
   Estatística de tese: $\dfrac{\bar{X} - \bar{Y}}{\sqrt{(0,3)^2/25 + (0,3)^2/16}}$
   Cálculo da região crítica: $P(Z < z_c) = 0,005 \ \Rightarrow\ z_c = -2,578$.
   Logo, $RC = \{z \in \mathbb{R} |\ z < -2,578\}$.

   (b) Para $\bar{x} = 8,2$ e $\bar{y} = 8,5$ temos

$$z = \frac{8,2 - 8,5}{\sqrt{(0,3)^2/25 + (0,3)^2/16}} \approx -3.$$

   Como $z \in RC$, rejeitamos H_0.