Estatística II
Época de Recurso - 29 de junho de 2023
Tópicos de Resolução

1

a) $E(X) = \int_0^\infty \lambda e^{-\lambda x} dx = \frac{1}{\lambda}$ (distribuição exponencial)
Método dos Momentos: $\bar{X} = \sum_{i=1}^n X_i$
$E(X) = \bar{X} \implies \frac{1}{\lambda} = \bar{X} \implies \tilde{\lambda} = \frac{1}{\bar{X}}$

b) Função verosimilhança: $L(\lambda|x_1, x_2..x_n) = \prod_{i=1}^n (\lambda e^{-\lambda x_i}) = \lambda^n e^{-\lambda \sum_i x_i}$
$l(\lambda|\mathbf{x}) = \ln(\lambda^n e^{-\sum_{i=1}^n x_i}) = nln(\lambda) - \lambda \sum_{i=1}^n x_i$
$\frac{dl}{d\lambda} = n\frac{1}{\lambda} - \sum_{i=1}^n x_i = 0 \implies \lambda = \frac{n}{\sum_{i=1}^n x_i} = \frac{1}{\bar{x}}$
$\frac{d^2 l}{d\lambda^2} = -n\frac{1}{\lambda^2} < 0, \implies \hat{\lambda} = (\bar{X})^{-1}$ é um maximizante de $L(\lambda)$.

c) $\hat{\lambda} = (\bar{x})^{-1} = (450/30)^{-1} = 1/15$
$\text{Prob}(X \leq 15|\lambda = 15) = 1 - e^{-\lambda x} = 1 - e^{-15/15} = 1 - e^{-1} = 0.6321$

2

a) IC para μ: VF: $\frac{\bar{X}-\mu}{S'/\sqrt{n}} \sim t_{(n-1)}$
$\implies IC(95\%) = (\bar{x} \pm t_{\alpha/2}\frac{s'}{\sqrt{n}})$
Média amostral: $\bar{x} = 956/10 = 95.6$,
Variância amostral: $s^2 = 91607/10 - (95.6)^2 = 21.34$
Variância amostral corrigida: $s'^2 = \frac{10}{9}s^2 = 1.11 \times 21.34 = 23.71$
IC para μ: $(95.6 \pm 2.26\frac{\sqrt{23.71}}{\sqrt{10}}) = (92.12, 99.08)$

IC para σ^2: VF: $\frac{(n-1)S'^2}{\sigma^2} \sim \chi^2_{(n-1)}$
$\implies IC(95\%) = \left((n-1)s'^2/q_{\alpha/2}, (n-1)s'^2/q_{1-\alpha/2}\right)$
$n - 1 = 9$ graus de liberdade

$q_{1-\alpha/2} = 2.70$ $q_{\alpha/2} = 19.023$
IC para σ^2: $(23.71 \times 9/19.023, 23.71 \times 9/2.70) = (11.22, 79.03)$

b) Falso.
Repetindo a construção do intervalo de confiança (com base em amostras independentes) espera-se que 95% dos intervalos contenham o verdadeiro valor de θ.
A probabilidade de o verdadeiro valor de θ estar contido dentro do intervalo de confiança é 0 ou 1.

c) Estimativa para μ: $\boxed{\bar{x} = 95.6}$
Estimativa para a variância corrigida: $\boxed{s'^2 = 23.71}$
$X \sim N(\hat{\mu}, \hat{\sigma}^2) \implies Z = (X - \hat{\mu})/\hat{\sigma}) \sim N(0,1) \implies$
$Prob(X \leq 100) = Prob(Z \leq z)$
onde $z = (100 - 95.6)/\sqrt{23.71} = 0.90$
Da tabela 4: $\Phi(0.90) = 0.8159$

d) Teste de hipóteses para a igualdade de médias
$H_0 : \mu_x = \mu_y$ vs. $H_1 : \mu_x \neq \mu_y$
E.T. : $T = \dfrac{\bar{X} - \bar{Y} - (\mu_X - \mu_Y)}{\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)}\sqrt{\frac{(m-1)S_X'^2 + (n-1)S_Y'^2}{m+n-2}}} \sim t_{(18)}$
$t_{obs} = \frac{\bar{x}-\bar{y}}{s^*}$ onde $s^* = \sqrt{\frac{1}{m} + \frac{1}{n}}\sqrt{\frac{(m-1)s_x'^2 + (n-1)s_y'^2}{m+n-2}}$
$W_\alpha = \{t_{obs} : |t_{obs}| > t_{\alpha/2}\}$
$t_{\alpha/2} = 2.101$
$\bar{y} = 986/10 = 98.6$
Variância amostral: $s_y^2 = 99600/10 - (98.6)^2 = 238.04$
Variância amostral corrigida: $s_y'^2 = \frac{10}{9} s_y^2 = 1.11 \times 238.04 = 264.22$
$t_{obs} = \frac{95.6 - 98.6}{5.368} = -0.056$
$t_{obs} \notin W \Rightarrow$ Não se rejeita H_0.
Não há evidência estatística para a diferença de médias.

3

a) $X \sim B(1, \theta)$, $n = 1000$
H_0: $\theta \leq 0.25$ H_1: $\theta > 0.25$
$\bar{x} = 270/1000 = 0.27$
$Z = \dfrac{\bar{X} - \theta}{\sqrt{\frac{\theta(1-\theta)}{n}}} \sim N(0,1)$
$z_{obs} = \frac{0.27 - 0.25}{0.014} = 1.43$
$W = \{z_{obs} : z_{obs} > z_\alpha\}$

Para $\alpha = 0.05, z_\alpha = 1.645$ $z_{obs} \notin W \implies$ Não rejeitar H_0.
Para $\alpha = 0.10, z_\alpha = 1.28$ $z_{obs} \in W \implies$ Rejeitar H_0 .
Rejeita-se a suposição inicial do grupo contra a candidata ao nível de 10% mas não ao nível de 5%.

b) valor-p $= Prob(z > z_{obs}) = 1 - \Phi(z_{obs}) = 1 - 0.9279 = 0.07$, significância observada do teste, correspondente à probabilidade de cometer um erro de 1ª espécie (ou seja, rejeitar H_0 sendo esta verdadeira). O valor-p dá a probabilidade de o valor observado para a estatística na amostra quando a hipótese nula é verdadeira e, como tal, permite não só rejeitar ou não a hipótese nula mas também situar a estatística observada dentro da região de rejeição permitindo saber quão longe está da fronteira da região de rejeição para uma dada dimensão do teste.
Tem-se valor-p $= 0.07 > 0.05 = \alpha$ mas valor-p $= 0.07 < 0.10 = \alpha$, sendo a conclusão equivalente à da alínea anterior.

4

Teste de ajustamento (não paramétrico).
H_0: X segue a distribuição definida por f $(N_j = np_{0j})$ vs.
H_1: X não segue a distribuição definida por f $(N_j \neq np_{0j})$

Frequências observadas: $f_0 = 30/50 = 0.6, f_1 = 14/50 = 0.28, f_{2+} = 6/50 = 0.12$
E.T.: $Q = \sum_{j=1}^{m} \dfrac{(N_j - np_{0j})^2}{np_{0j}} \sim \chi^2_{(m-1)}$, com $m - 1 = 2$
$Q_{obs} = \frac{(30-0.6\times50)^2}{0.6\times50} + \frac{(14-0.25\times50)^2}{0.25\times50} + \frac{(6-0.15\times50)^2}{0.15\times50} = 0 + 0.18 + 0.3 = 0.48$
$p_{obs} = P(Q > Q_{obs}) \approx P(Q > 0.51) = 0.975 > 0.05 \implies$ Não rejeitar H_0 ao nível de 5%, concluindo que os dados são consistentes com a hipótese definida.
Alternativamente, pela região de rejeição, tem-se:
$Q_{2,0.05} = 5.991$ e $W = \{Q : Q > 5.991\}, Q_{obs} \notin W$.

5

a) Teste de significância global:
H_0: $\beta_1 = \beta_2 = \beta_3 = \beta_4 = 0$ vs. H_1: $\exists \beta_j \neq 0,\ j = 1, 2, 3, 4$.
$F_{\text{obs}} = 86.87$ e valor-p $< 0.001 < 0.05$
Rejeitamos H_0. A regressão é globalmente significativa.

Teste de significância individual para β_1:
H_0: $\beta_1 = 0$ vs. H_1: $\beta_1 \neq 0$.

A variável "years" corresponde a x_1 com β_1: $\hat{\beta}_1 = 64455$ e $se(\hat{\beta}_1) = 17771$.
E.T.: $t_1 = \frac{\hat{\beta}_1}{se(\hat{\beta}_1)} \sim t_{(349)}$
$W = \{t : |t| > t_{\alpha/2}\} = \{t : |t| > 1.96\}$
$t_{\text{obs}} = 3.627 \in W \implies$ Rejeita-se H_0, concluindo que a variável "years" é estatisticamente significativa ao nível de 5%.
Alternativamente, pelo valor-p: $P(t > t_{obs}) = 2.95 \times 10^{-5} < \alpha$.

b) Ambas as variáveis são estatisticamente significativas, como se pode concluir pelos valor-p observados, e têm um efeito positivo no salário do jogador.
$\hat{\beta}_2 = 19633 \implies$ por cada ponto adicional no índice de performance no jogo, o salário do jogador aumenta, em média, 19633\$ ceteris paribus.
$\hat{\beta}_3 = 3149 \implies$ por cada jogo adicional por ano na liga, o salário do jogador aumenta, em média, 3149\$ ceteris paribus.

c) E.T.: $t_3 = \frac{\hat{\beta}_3}{se(\hat{\beta}_3)} \sim t_{(349)}$
IC a 90%: $\alpha = 0.1$, $t_{\alpha/2} = 2.576$
IC$=\hat{\beta}_3 \pm t_{\alpha/2} \times se(\hat{\beta}_3) = (3149 - 1.645 \times 1566, 3149 + 1.645 \times 1566) = (572.93, 5725.07)$.
Conclui-se que a variável é significativa pois $\{0\} \notin$ IC.

d) Intervalo de previsão
Começando de $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \hat{\beta}_3 x_3$, queremos estimar $\theta_0 = \beta_0 + \beta_1 c_1 + \beta_2 c_2 + \beta_3 c_3$ onde $c_1 = 5, c_2 = 110, c_3 = 215$.
O estimador de θ_0 é $\hat{\theta}_0 = \hat{\beta}_0 + \hat{\beta}_1 c_1 + \hat{\beta}_2 c_2 + \hat{\beta}_3 c_3$.
Para obter um intervalo de confiança para θ_0 precisamos do erro padrão para θ_0. Escrevemos

$$\beta_0 = \theta_0 - \beta_1 c_1 - \beta_2 c_2 - \beta_3 c_3$$

e substituímos na equação da regressão linear original:

$$y = \theta_0 + \beta_1(x_1 - c_1) + \beta_2(x_2 - c_2) + \beta_3(x_3 - c_3).$$

Realizando uma regressão sobre as novas variáveis $x_j - c_j$ obtemos uma estimativa e erro padrão para θ_0. Do Anexo II, retiramos $\hat{\theta}_0 = 1512888$ e $se(\hat{\theta}_0)$=111352.
IC de previsão a 99% : $(1512888 \pm 2.576 \times 111352) = (1.22, 1.80)$ milhões de dólares.

e) $\hat{\beta}_1 = 0.07 \implies$ por cada ano adicional de jogo estima-se que o salário do jogador aumente, em média, $\beta_1 \times 100 = 7\%$, ceteris paribus.

Passos para prever o salário dados valores para x_1, x_2, x_3 através de um modelo ajustado de log(salary) ($\ln(y)$):

1. Obter $\hat{m}_i = \exp(\widehat{\ln(y_i)})$ para as n observações da amostra.
2. Estimar o declive da regressão $y_i = a\hat{m}_i$ sem termo independente $\hat{a}$.
3. As previsões pontuais para y, obtêm-se através de $\hat{y} = \hat{a}\exp(\widehat{ln(y_i)})$