Probabilidades e Estatística

— TODOS OS CURSOS —

2ºsemestre – 2019/2020
27/06/2020 – **13:00**

Duração: 60 minutos **2ºteste**

Pergunta 1

A variável aleatória X representa o número de acessos horários a um servidor e possui função de probabilidade

$$P(X=x)=\begin{cases}\frac{(x+2)(x+1)}{2}p^3(1-p)^x & x=0,1,2,\ldots\\ 0 & \text{outros casos,}\end{cases}$$

onde p é uma probabilidade desconhecida. Considere que $(X_1,\ldots,X_5)$ é uma amostra aleatória de dimensão 5 de X. Obtenha a estimativa de máxima verosimilhança de $P(X=x)$ baseada na seguinte concretização de $(X_1,\ldots,X_5)$: (x_1,x_2,x_3,x_4,x_5).

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

- **V.a. de interesse**

  X = número de acessos horários a um servidor

- **F.p. de X**

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

- **Parâmetro desconhecido**

  $p,\quad 0<p<1$

- **Obtenção da estimativa de MV de θ**

  Passo 1 — Função de verosimilhança

$$\begin{aligned}
L(p\mid \underline{x}) &= P(\underline{X}=\underline{x})\\
&\overset{X_i\, indep}{=} \prod_{i=1}^n P(X_i=x_i)\\
&\overset{X_i\sim X}{=} \prod_{i=1}^n P(X=x_i)\\
&= \prod_{i=1}^n \left[\frac{(x_i+1)\,(x_i+2)}{2}\,(1-p)^{x_i}\,p^3\right]\\
&= 2^{-n}\left[\prod_{i=1}^n (x_i+1)\,(x_i+2)\right](1-p)^{\sum_{i=1}^n x_i}\,p^{3n},\quad 0<p<1
\end{aligned}$$

  Passo 2 — Função de log-verosimilhança (caso $\bar{x}>0$)

$$\ln L(p\mid \underline{x}) = -n\ln(2)+\sum_{i=1}^n \ln[(x_i+1)\,(x_i+2)]+\ln(1-p)\sum_{i=1}^n x_i+3n\ln(p)$$

  Passo 3 — Maximização

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

$$\begin{cases}-\frac{\sum_{i=1}^n x_i}{1-\hat{p}}+\frac{3n}{\hat{p}}=0\\ -\frac{\sum_{i=1}^n x_i}{(1-\hat{p})^2}-\frac{3n}{\hat{p}^2}<0 & \text{(prop. verdadeira porque } \bar{x}\geq 0 \text{ e } n>0)\end{cases}$$

$$\begin{cases}-\hat{p}\,n\bar{x}+3n-3n\hat{p}=0\\ -\end{cases}$$

Passo 4 — Estimativa de MV de λ

$$\hat{p} = \frac{3}{\bar{x}+3}, \quad \text{onde } \bar{x} = \tfrac{1}{5}\textstyle\sum_{i=1}^{5} x_i$$

- **Outro parâmetro desconhecido**

  $h(p) = P(X = x) = \frac{(x+2)(x+1)}{2}\, p^3 (1-p)^x$

- **Estimativa de MV de $h(p)$**

$$\begin{aligned} \widehat{h(p)} &\overset{prop.inv.EMV}{=} h(\hat{p}) \\ &= \frac{(x+2)(x+1)}{2}\,\hat{p}^3 (1-\hat{p})^x. \end{aligned}$$

Pergunta 2

Considere que o tempo de vida de certo tipo de lâmpadas (na unidade milhar de hora) tem distribuição normal com valor esperado e variância desconhecidos. Depois de obtidas as durações de n lâmpadas selecionadas casualmente, verificou-se que a variância amostral corrigida é igual a s^2. Determine um intervalo de confiança a $(1-\alpha)\times 100\%$ para o verdadeiro valor do desvio padrão.

- **V.a. de interesse**

  $X =$ tempo de vida de certo tipo de lâmpada

- **V.a. fulcral para σ**

  $Z = \frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{(n-1)}$

- **Situação**

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

  μ desconhecido

  σ^2 DESCONHECIDO

- **Obtenção do IC para σ**

  Passo 1 — Seleção da v.a. fulcral para σ

$$Z = \frac{(n-1)\,S^2}{\sigma^2} \sim \chi^2_{(n-1)}$$

  Passo 2 — Obtenção dos quantis de probabilidade

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

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

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

$$P\left[a_\alpha \le \frac{(n-1)\,S^2}{\sigma^2} \le b_\alpha\right] = 1-\alpha$$

$$P\left[\frac{1}{b_\alpha} \le \frac{\sigma^2}{(n-1)\,S^2} \le \frac{1}{a_\alpha}\right] = 1-\alpha$$

$$P\left[\sqrt{\frac{(n-1)\,S^2}{b_\alpha}} \le \sigma \le \sqrt{\frac{(n-1)\,S^2}{a_\alpha}}\right] = 1-\alpha$$

Passo 4 — Concretização

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

Pergunta 3

Numa sondagem à opinião pública, em dado círculo eleitoral, foram inquiridas n pessoas e $n\bar{x}$ delas manifestaram-se favoráveis a determinado partido político. Rejeita-se ou não a hipótese de este partido ter 60% das preferências naquele círculo eleitoral? Considere uma hipótese alternativa bilateral e decida com base no valor-p aproximado.

- **Hipóteses**

  $H_0 : p = p_0 = 0.60$

  $H_1 : p \neq p_0$

- **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 bilateral, r.r. de H_0 é reunião de dois intervalos simétricos, $W = (-\infty, -c) \cup (c, +\infty)$.

- **Decisão (com base no valor-p)**

$$\begin{aligned} valor-p &= P\left[|T| > t = \frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}} \mid H_0\right] \\ &\simeq 2 \times \left[1 - \Phi\left(\left|\frac{\bar{x}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right|\right)\right] \end{aligned}$$

  Devemos escolher a resposta, das quatro apresentadas em nota de rodapé,[1] 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. $\alpha_0 > valor-p$.

Pergunta 4

Seja X a variável aleatória que representa o número semanal de avarias de um sistema eletrónico. Uma engenheira eletrotécnica defende a hipótese H_0 de que X possui função de probabilidade $P(X = x) = \frac{(x+2)(x+1)}{2} p^3 (1-p)^x,\ x = 0, 1, 2, \ldots$

A contabilização do número semanal de tais avarias, em n semanas selecionadas casualmente, conduziu à seguinte tabela de frequências:

Número semanal de avarias	0	1	2	3	superior a 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 E_1 e E_5 (aproximando-as às décimas), averigúe se H_0 é consistente com este conjunto de dados. Decida com base no valor-p.

---

[1] 1. Rejeita-se para 1%, 5% e 10%. 2. Rejeita-se para 5% e 10% e não se rejeita para 1%. 3. Rejeita-se para 10% e não se rejeita para 1% e 5%. 4. Não se rejeita para 1%, 5% e 10%.

- **V.a. de interesse**

  $X =$ número semanal de avarias de um sistema eletrónico

- **Hipóteses**

  $H_0 : P(X = x) = \frac{(x+2)(x+1)}{2} p^3 (1-p)^x, \ x = 0,1,2,\ldots$

  $H_1 : \neg H_0$

- **Estatística de teste**

  $T = \sum_{i=1}^{k} \frac{(O_i - E_i}{E_i} \overset{a}{\sim}_{H_0} \chi_{(k-\beta-1)},$

  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**

$$
\begin{aligned}
E_1 &= n \times P(X = 0 \mid H_0) \\
&= n \times \frac{(0+2)(0+1)}{2} p^3 (1-p)^x \\
E_5 &= n - \sum_{i=1}^{4} E_i
\end{aligned}
$$

- **Decisão (com base no valor-p)**

$$
\begin{aligned}
valor - p &= P\left(T > t = \frac{(o_i - E_i)^2}{E_i} \mid H_0\right) \\
&\simeq 1 - F_{\chi^2_{(k-1)}}\left(\sum_{i=1}^{k} \frac{(o_i - E_i)^2}{E_i}\right)
\end{aligned}
$$

  Devemos escolher a resposta, das quatro apresentadas em nota de rodapé,[2] que se coaduna com:

  - a não rejeição de H_0 a qualquer n.s. $\alpha_0 \le valor - p$;
  - a rejeição de H_0 a qualquer n.s. $\alpha_0 > valor - p$.

Pergunta 5

Para avaliar a relação entre a densidade populacional (Y, em centenas de habitantes por quilómetro quadrado) das áreas residenciais de determinada cidade e a distância (x, em quilómetro) ao centro da mesma, recolheu-se uma amostra casual de dimensão n que conduziu aos seguintes valores:

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

onde $x_i \in [x_{(1)}, x_{(n)}]$ para $i = 1,2,\ldots,n$.

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

- **Hipóteses de trabalho**

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

[2] 1. Rejeita-se para 1%, 5% e 10%. 2. Rejeita-se para 5% e 10% e não se rejeita para 1%. 3. Rejeita-se para 10% e não se rejeita para 1% e 5%. 4. Não se rejeita para 1%, 5% e 10%.

- **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 $a_\alpha \le Z \le b_\alpha$

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

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

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

Passo 4 — Concretização

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

onde:

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

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

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