UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Probabilidade e Estatística - Prova # 02

Atenção: Só serão aceitas respostas com justificativa. - Rio, 13/06/2017

1. Um caminhão pode transportar uma carga máxima de três toneladas. Sua carga é composta por caixas transportando dois tipos de produtos, tipo A e tipo B. Suponha que o peso de caixas com produto do tipo A tenha distribuição normal com média 50 kg e variância 25 kg^2 e que o peso de caixas com produto do tipo B tenha distribuição normal com média 38 kg e variância 16 kg^2. Suponha ainda que pesos de caixas distintas sejam independentes.

   (a) Calcule a média e variância da carga tranportada quando o caminhão é carregado com 25 caixas com produto do tipo A e 45 com caixas com produto do tipo B.

   (b) Com a mesma carga descrita no item (a), calcule a probabilidade do caminhão ter sido carregado com uma carga superior a máxima permitida.

   (c) Se o caminhão só for carregado com produto do tipo A, qual o maior número de caixas que ele pode transportar de modo que a probabilidade de exceder a carga máxima seja inferior a 2%.

Solução:

(a) Usamos a propriedade de que para soma de variáveis aleatórias independentes a média é igual a soma das médias e a variância é igual a soma das variâncias. Assim, a carga transportada tem distribuição normal com média

$$\mu = 25 \times 50 + 45 \times 38 = 2960\text{kg}$$

e

$$\sigma^2 = 25 \times 25 + 45 \times 16 = 1345\mathsf{kg}^2\,.$$

(b) Usamos a propriedade de que soma de variáveis aleatórias independentes com distribuição normal também é normal. Assim, pelo item (a) temos que a carga transportada X é normal com média 2960 e variância 1345. A probabilidade do caminhão ter sido carregado com uma carga superior a máxima é

$$P(X > 3000) = P\Big(\frac{X - 2960}{\sqrt{1345}} > \frac{3000 - 2960}{\sqrt{1345}}\Big) = 1 - \Phi(1.09) = 1 - 0.8621 = 0.1379\,.$$

(c) Seja k o número máximo de caixas que podem ser transportadas de forma que a probabilidade de exceder a carga máxima seja inferior a 2% e Y a carga transportada com este número de caixas. Temos que Y tem distribuição normal com média $50k$ e variância $25k$. Assim

$$0.02 \geq P(Y > 3000) = 1 - \Phi\Big(\frac{3000 - 50k}{5\sqrt{k}}\Big)$$

Observe que a função

$$\frac{3000 - 50x}{5\sqrt{x}}$$

é decrescente em x, assim k está bem definido como o maior valor tal que

$$\frac{3000 - 50k}{5\sqrt{k}} \geq z_{0.98} = 2.05\,.$$

Ou seja $k = \lfloor w^2 \rfloor$ onde w é a maior raiz de

$$50w^2 + 10.25w - 3000 = 0\,.$$

Daí

$$k = \left\lfloor \Big(\frac{-10.25 + \sqrt{10.25^2 + 4 \times 50 \times 3000}}{2 \times 50}\Big)^2 \right\rfloor = \lfloor (7.64)^2 \rfloor = 58\,.$$

2. Os dados a seguir referem-se a 9 observações sobre o tempo gasto de acondicionamento de um produto em minutos (Y) e seu volume em toneladas (X).

Obs.	1	2	3	4	5	6	7	8	9		
X	48	72	63	82	88	109	112	123	140	$\sum x = 837$	$\sum x^2 = 85.079$
Y	84	108	110	133	144	152	180	196	231	$\sum y = 1.338$	$\sum y^2 = 216.526$

$\sum xy = 135.492$

(a) Calcule a correlação entre tempo e volume.

(b) Observe que nesse caso, uma reta que descreva o comportamento do tempo em função do volume deveria começar pela origem uma vez que, se não há volume, o tempo de acondicionamento deve ser nulo. Assim, um modelo apropriado para explicar esses dados é $Y_i = b \times X_i + \mathsf{erro}_i$.

Determine o estimador de b usando o critério de mínimos quadrados, ou seja, minimizando a soma dos quadrados dos erros.

(c) Com base no modelo proposto, estime o tempo de acondicionamento para um volume de 100 toneladas.

Solução:

$$8s_{xy} = 135492 - \frac{837 \times 1338}{9} = 11058 \text{ ton.min}$$

$$8s_y^2 = 216526 - \frac{1338^2}{9} = 17610 \text{ min}^2$$

$$8s_x^2 = 85079 - \frac{837^2}{9} = 7238 \text{ ton}^2$$

(a)

$$r = \frac{11058}{\sqrt{17610 \times 7238}} \approx 0,979$$

(b)

$$y_i = bx_i + e_i$$

tal que $e_i = y_i - bx_i$ e $\sum e_i^2 = \sum (y_i - bx_i)^2 = f(b)$

$$f'(b) = -2\sum x_i(y_i - bx_i) = 0 \quad \leftrightarrow \quad b = \frac{\sum xy}{\sum xx}$$

$f''(b) = 2\sum x_i^2 = 2\sum xx \geq 0$

Também é possível, observando que $f(b)$ é uma parábola com a concavidade voltada para cima, obter b como o vértice da parábola, que fornece o valor mínimo de $f(b)$.

Logo, de fato, $b = \frac{\sum xy}{\sum xx}$ é ponto de mínimo absoluto.

(c) Temos $b = \frac{135492}{85079} \approx 1,593$ tal que

$$\text{tempo estimado} \approx 1,593 \times 100 = 159,3 \text{ minutos.}$$

3. Uma amostra aleatória de tamanho 16 de uma população normal com média μ e variância σ^2 resultou em $\sum_{i=1}^{16} x_i = 24$ e $\sum_{i=1}^{16} x_i^2 = 36,5822$. Construa um intervalo de confiança de 95% para média populacional μ se a variância populacional

(a) $\sigma^2 = 0,04$;

(b) não é conhecida.

Solução:

Dos dados temos $\bar{x} = 1,5$, $s^2 \approx 0,0388$ e $s \approx 0,197$.

(a) Como a variância é conhecida o intervalo de 95% de confiança para μ é dado por

$$\bar{x} \pm z_{(0,975)} \frac{\sigma}{\sqrt{n}} : 1,5 \pm 1,96 \frac{0,2}{4} : 1,5 \pm 0,098 : (1,402; \quad 1,598)$$

(b) Nesse caso, como a população é normal, temos que a distribuição de $T = \frac{\bar{X}-\mu}{S/\sqrt{n}}$ é uma t com $n-1$ graus de liberdade tal que o intervalo de 95% de confiança para μ é dado por

$$\bar{x} \pm t_{(0,975,15)} \frac{s}{\sqrt{n}} : 1,5 \pm 2,131 \frac{0,197}{4} : 1,5 \pm 0,105 : (1,395; \quad 1,605)$$

4. Queremos testar $\mathsf{H}_0 : p \geq 0,85$ contra $\mathsf{H}_1 : p < 0,85$ em que p é a proporção de peças que atendem as especificações de fábrica. Para isso foi selecionada uma amostra aleatória de 40 peças dessa população, verificando-se que trinta e duas (32) peças na amostra atendem as especificações de fábrica.

   (a) Qual a decisão a ser tomada ao nível $\alpha = 5\%$? Por que?

   (b) Qual o p-valor (nível crítico) desse teste?

   (c) Qual a probabilidade de se cometer o Erro Tipo II, no teste obtido em (a), se a proporção populacional de peças que atendem as especificações de fábrica for p = 0,75?

Solução:

(a) Como $np_0(1-p_0) = 40 \times 0,85 \times 0,15 = 5,1 > 3$ a aproximação normal da binomial é adequada.

Temos $n = 40$, $\hat{p} = 0,80$. Ao nível de significância de 5% a região crítica é dada por $Z_0 < -z_{(0,95)} = -1,645$ ou, equivalentemente, $\hat{p} < 0,757$ em que $Z_0 = \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}$.

O valor amostral obtido foi $z_0 \approx -0,8856$ e não está na região crítica ou, equivalentemente, $\hat{p} = 0,8$ não está na região crítica.

Logo, ao nível de significância de 5%, não rejeitamos H_0.

(b) $P - \mathsf{valor} = P(\hat{p} \leq 0,8 | p = 0,85) \approx \phi\left(\frac{0,8-0,85}{\sqrt{\frac{0,85\times0,15}{40}}}\right) \approx \phi(-0,89) = 0,1867$

(c) $\beta(0,75) = P(\hat{p} > 0,757 | p = 0,75) \approx 1 - \phi\left(\frac{0,757-0,75}{\sqrt{\frac{0,75\times0,25}{40}}}\right) \approx 1 - \phi(0,1) \approx 0,4602$