USP – ICMC – SME0122 - Introdução à Inferência Estatística
1ª prova – 2º/2013 – 15/10/2013

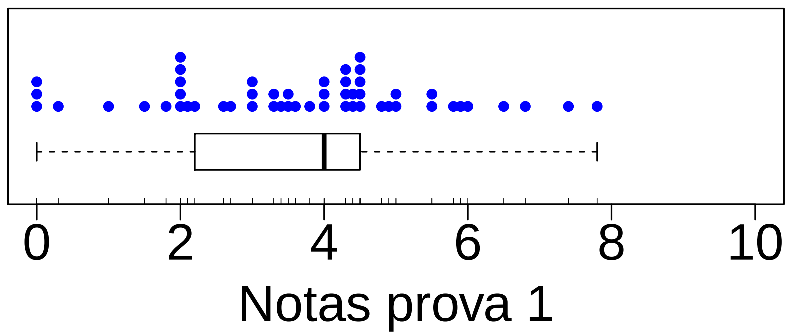

Figura 1: Gráficos de pontos e de caixa das notas da 1ª prova.

S O L U Ç Ã O

1. Antes do tratamento com um certo medicamento, 11 pessoas com problema de insônia dormiam em média 2 h por noite. Como efeito do tratamento, espera-se que ocorra um acréscimo no tempo de sono. Entretanto, alguns especialistas acreditam que o tratamento não apresenta efeito nenhum. Com base nos dados abaixo, referentes ao número de horas dormidas, você concorda com estes especialistas?

   3,1 1,8 2,7 2,4 2,9 0,2 3,7 5,1 8,3 2,1 2,4

   Solução. Supomos que o tempo de sono, em h, é uma variável aleatória X com distribuição $\mathrm{N}(\mu, \sigma^2)$. A existência de algum efeito do medicamento significa redução ou acréscimo no tempo de sono. Assim, as hipóteses a testar são $\mathrm{H}_0 : \mu = 2$ contra $\mathrm{H}_1 : \mu \neq 2$ (ou seja, $\mu_0 = 2$). O nível de significância adotado é de 5%. Consultando a Tábua III com $p = 5\%$ e $n - 1 = 10$ graus de liberdade obtemos $t_c = 2,228$. As estimativas da média e do desvio padrão amostral são $\overline{x} = 3,16$ h e $s = 2,09$ h. Com estes valores obtemos a estatística de teste

$$T = \frac{\sqrt{n}(\overline{x} - \mu_0)}{s} = 1,833.$$

   Portanto, $|T| < t_c$ e H_0 não é rejeitada. Com base nos dados e com um nível de significância de 5%, podemos concordar com os especialistas. A diferença do tempo de sono, igual a $3,16 - 2,0 =$ 1,16 h, não é significativa.

2. A utilização da capacidade de um processador durante um certo intervalo de tempo é uma variável aleatória (X) com função densidade $f(x;\theta) = \theta x^{\theta - 1}$, se $0 < x < 1$ e $f(x;\theta) = 0$, caso contrário, com $\theta > 0$. Em um experimento foram coletados os dados abaixo.

   0,81 0,93 0,74 0,99 0,84 0,92 0,80 0,97 0,41

   (a) Apresente uma estimativa para o parâmetro θ.

   (b) Sabendo que $[-4,12/\sum_{i=1}^{n}\log(X_i); -15,76/\sum_{i=1}^{n}\log(X_i)]$ é um intervalo de confiança de 95% para θ, você afirmaria que a utilização da capacidade do processador tem distribuição uniforme?

   Solução. (a) Será apresentada a estimativa de máxima verossimilhança de θ. A função verossimilhança é

$$L(\theta) = \prod_{i=1}^{n} f(x;\theta) = \prod_{i=1}^{n} \theta x_i^{\theta - 1} = \theta^n \left(\prod_{i=1}^{n} x_i \right)^{\theta - 1}.$$

Para facilitar a maximização tomamos o logaritmo (natural), dado por

$$\ell(\theta) = n\log(\theta) + (\theta - 1)\sum_{i=1}^{n}\log(x_i),$$

cuja derivada primeira em relação a θ é

$$\frac{\partial}{\partial\theta}\ell(\theta) = \frac{n}{\theta} - \sum_{i=1}^{n}\log(x_i).$$

Igualando a 0 obtemos $\theta = -n/\sum_{i=1}^{n}\log(x_i)$. Como

$$\frac{\partial^2}{\partial\theta^2}\ell(\theta) = -\frac{n}{\theta^2} < 0,$$

temos que $\widehat{\theta} = -n/\sum_{i=1}^{n}\log(X_i)$ é o estimador de máxima verossimilhança de θ. Utilizando os dados coletados obtemos a estimativa 4,51.

(b) A partir do intervalo de confiança de 95% do enunciado e com os dados coletados obtemos a estimativa intervalar $[2,06;7,89]$. A distribuição uniforme corresponde a $\theta = 1$. Como $1 \notin [2,06;7,89]$, os dados não indicam que a distribuição da utilização da capacidade do processador é uniforme.

3. Em uma amostra aleatória de 75 itens, 12 tiveram um acabamento mais áspero do que permitem as especificações.

   (a) Apresente um intervalo de confiança de 95% para a proporção de itens que não atendem às especificações.

   (b) Quantos itens devem compor uma amostra se pretendemos estar 95% confiantes de que o erro de estimação da proporção de itens que não atendem às especificações não ultrapassa 0,05?

   Solução. (a) A proporção de itens na população com acabamento mais áspero do que permitem as especificações é denotada por p. A partir da amostra de $n = 75$ itens, a proporção amostral tem estimativa $\overline{\mathrm{p}} = 12/75 = 0,16$. Consultando a Tábua III com $p = 5\%$ obtemos $z_{\alpha/2} = 1,960$. Pela abordagem otimista o erro máximo é

$$\mathrm{E} = z_{\alpha/2}\sqrt{\frac{\overline{\mathrm{p}}(1-\overline{\mathrm{p}})}{n}} = 0,083,$$

   de modo que uma estimativa intervalar é dada por $[\overline{\mathrm{p}} - \mathrm{E}, \overline{\mathrm{p}} + \mathrm{E}] = [0,077;0,243]$.

   Pela abordagem conservativa o erro máximo é

$$\mathrm{E} = z_{\alpha/2}\sqrt{\frac{1}{4n}} = 0,113,$$

   de modo que uma estimativa intervalar é dada por $[\overline{\mathrm{p}} - \mathrm{E}, \overline{\mathrm{p}} + \mathrm{E}] = [0,047;0,273]$.

   (b) O erro máximo é $\mathrm{E} = 0,05$. O tamanho da amostra deve satisfazer

$$n \geqq \frac{z_{\alpha/2}^2\,\mathrm{p}(1-\mathrm{p})}{\mathrm{E}^2}.$$

   Com a estimativa de p do item (a) obtemos $n \geqq 206,5$. Tomamos $n = 207$.

   Pela abordagem conservativa, $n \geqq \{z_{\alpha/2}/(2\mathrm{E})\}^2 = 384,2$. Tomamos $n = 385$.

4. Um trabalhador que morava na cidade A mudou-se para a cidade B, onde continuou a receber o mesmo salário. Após a mudança, ele afirmou que o salário médio dos trabalhadores das duas cidades aumentou. Esta afirmação pode ser verdadeira?
   Solução. Sim. Provamos com os seguintes argumentos. Pelo enunciado percebe-se que não ocorreram mudanças nos salários em si. O número de trabalhadores em A e B é m e n, respectivamente, com $n > 1$ e $m > 1$. Os salários nas duas cidades são $x_1, x_2, \ldots, x_m$ e $y_1, y_2, \ldots, y_n$, com médias $\overline{x} = \sum_{i=1}^{m} x_i/m$ e $\overline{y} = \sum_{i=1}^{n} y_i/n$. Supomos que $\overline{x} > \overline{y}$. Sem perda de generalidade, digamos que x_1 é o salário do

trabalhador que se mudou de A para B. Supomos que $\overline{y} < x_1 < \overline{x}$. Após a mudança os salários médios passam a ser

$$\overline{x}^* = \frac{x_2 + \cdots + x_m}{m-1} \quad \text{e} \quad \overline{y}^* = \frac{x_1 + y_1 + \cdots + y_n}{n+1},$$

que podem ser escritos como

$$\overline{x}^* = \frac{m\overline{x} - x_1}{m-1} = \frac{m}{m-1}\overline{x} - \frac{x_1}{m-1}$$

e

$$\overline{y}^* = \frac{x_1 + n\overline{y}}{n+1} = \frac{x_1}{n+1} + \frac{n}{n+1}\overline{y},$$

de modo que

$$\overline{x}^* - \overline{x} = \frac{\overline{x} - x_1}{m-1} \quad \text{e} \quad \overline{y}^* - \overline{y} = \frac{x_1 - \overline{y}}{n+1}.$$

Levando em conta que $\overline{y} < x_1 < \overline{x}$, concluímos que $\overline{x}^* - \overline{x} > 0$ e $\overline{y}^* - \overline{y} > 0$.

5. Uma variável aleatória X com distribuição Poisson(θ) é tal que $E(X) = Var(X) = \theta$. Portanto, para uma amostra aleatória de tamanho n, pelo teorema central do limite obtemos que

$$\frac{\overline{X} - \theta}{\sqrt{\theta/n}}$$

tem distribuição N(0,1), aproximadamente, em que $\overline{X} = \sum_{i=1}^{n} X_i/n$.

A partir da observação de dados históricos sabe-se que o número de solicitações de um serviço durante um certo intervalo de tempo é uma variável aleatória com distribuição Poisson. Os dados abaixo foram coletados. Apresente um intervalo de confiança de 95% aproximado para o número médio de solicitações.

13 10 16 9 16 10 8 9 10 16 11 7

Solução. O número de solicitações de um serviço durante um certo intervalo de tempo é denotado por X. Como $E(X) = \theta$, por substituição, um estimador para θ é $\widehat{\theta} = \overline{X}$. Consultando a Tábua III com $p = 5\%$ obtemos $z_{\alpha/2} = 1,960$. Utilizando a distribuição do enunciado temos que

$$P\left(-z_{\alpha/2} \leqq \frac{\overline{X} - \theta}{\sqrt{\theta/n}} \leqq z_{\alpha/2}\right) \cong 0,95,$$

de modo que

$$P\left(-z_{\alpha/2}\sqrt{\frac{\theta}{n}} \leqq \overline{X} - \theta \leqq z_{\alpha/2}\sqrt{\frac{\theta}{n}}\right) \cong 0,95$$

e

$$P\left(\overline{X} - z_{\alpha/2}\sqrt{\frac{\theta}{n}} \leqq \theta \leqq \overline{X} + z_{\alpha/2}\sqrt{\frac{\theta}{n}}\right) \cong 0,95.$$

Assim, obtivemos um intervalo para θ com probabilidade aproximadamente igual a 0,95. Os limites do intervalo acima dependem de θ. Substituindo pelo estimador $\widehat{\theta} = \overline{X}$ resulta em

$$\left[\overline{X} - z_{\alpha/2}\sqrt{\frac{\overline{X}}{n}},\ \overline{X} + z_{\alpha/2}\sqrt{\frac{\overline{X}}{n}}\right].$$

Com os dados coletados em uma amostra de tamanho $n = 12$, a estimativa de θ é $\overline{x} = 11,25$. Substituindo $\overline{x}$, $z_{\alpha/2}$ e n, uma estimativa intervalar para o número de médio de solicitações de um serviço durante um certo intervalo de tempo é $[9,35; 13,15]$.