UFRJ - CCMN - IM - Departamento de Métodos Estatísticos
Probabilidade e Estatística - Estatística

Prova #2 22-05-2014

Atenção: Não serão aceitas respostas sem justificativa: as expressões que levaram a alguma resposta numérica devem ser indicadas nos espaços apropriados.

1. A tabela abaixo apresenta o número de gols sofridos pela seleção brasileira de futebol em cada uma das edições da copa do mundo.

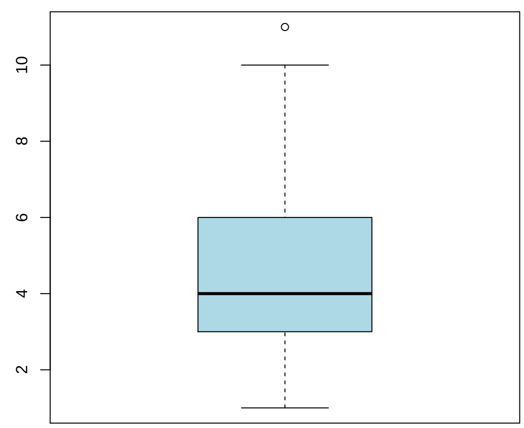
Ano	1930	1934	1938	1950	1954	1958	1962	1966	1970	1974
Gols sofridos	2	3	11	6	5	4	5	6	7	4
Ano	1978	1982	1986	1990	1994	1998	2002	2006	2010	
Gols sofridos	3	6	1	2	3	10	4	2	4	

(a) Qual é o número médio de gols sofridos pela seleção por edição da copa? Calcule o desvio padrão e o coeficiente de variação do número de gols sofridos.

(b) Calcule os quartis e construa um *box*-plot. Existe algum valor que pode ser considerado discrepante?

Resposta: **(a)** $\bar{x} = \frac{88}{19} = 4,631579$; $s^2 = \frac{\sum x_i^2 - n\bar{x}^2}{n-1} = \frac{532 - 19 \times 4,631579^2}{19-1} = 6,912281 \Rightarrow s = 2,629122$; $cv = \frac{s}{\bar{x}} \times 100 = \frac{2,629122}{4,631579} \times 100 = 56,7651\%$.

(b) $Q1 = 3$, $Q2 = 4$, $Q3 = 6$; $LI = 3 - 1,5 \times 3 = -1,5$, $LS = 6 + 1,5 \times 3 = 10,5$.



O valor 11 da copa de 1938 é discrepante.

2. Deseja-se estimar a proporção p das empresas localizadas em uma região que possuem um setor de treinamento de pessoal. Determine o tamanho mínimo de uma amostra aleatória simples a ser extraída do conjunto dessas empresas para que o erro absoluto de estimação seja inferior a $0,105$, com uma probabilidade de $0,98$, nas seguintes situações:

(a) Nada se sabe sobre o valor de p;

(b) Sabemos que pelo menos 80% das empresas dessa região possuem um setor de treinamento.

Resposta: Para garantir que $P[|\hat{p} - p| < 0,105] = 0,98$, devemos ter $n \geq (z_{0,99}/0,105)^2 p(1-p)$. Consultando a tabela da Normal padrão, obtemos $z_{0,99} = 2,33$.

(a) Fazendo $p = 0,5$, obtemos $n \geq (2,33/0,105)^2 \times 0,5 \times 0,5 \approx 124$.

(b) Fazendo $p = 0,8$, obtemos $n \geq (2,33/0,105)^2 \times 0,8 \times 0,2 \approx 79$. Obs.: Em ambos os casos foi feito um arredondamento para cima.

3. Suponhamos que se queira estimar, usando intervalo de confiança, o tempo médio que leva o princípio ativo de um remédio para chegar à corrente sanguínea de um paciente desde o momento que ele é ingerido.

(a) Suponha inicialmente que a distribuição de X é desconhecida e obtenha um intervalo de confiança de 95% para a média. Suponha que 80 medidas deste tempo X, em segundos, foram feitas e obteve-se:

$$\sum_{i=1}^{80} x_i = 1\,920 \text{ e } \sum_{i=1}^{80} x_i^2 = 48\,055.$$

Use como valor da variância, o obtido pelo calculo de s^2.

(b) Suponha agora que somente 25 medidas de X estão disponíveis e obteve-se:

$$\sum_{i=1}^{25} x_i = 600 \text{ e } \sum_{i=1}^{25} x_i^2 = 15\,018.$$

Se for possível dizer que X é normalmente distribuída, obtenha um intervalo de confiança de 95% para a média sem fazer nenhuma suposição sobre o valor da variância de X.

Resposta: **(a)** Temos $\bar{x} = 24$, $s^2 = 25$ e $z_{1-\alpha/2} = 1,96$, onde $1 - \alpha/2 = 0,975$. Logo, os limites do intervalo são:

$$l_1 = 24 - 1,96\frac{5}{\sqrt{80}} = 24 - 1,0957 \text{ e } l_2 = 24 + 1,96\frac{5}{\sqrt{80}} = 24 + 1,0957.$$

(b) Já que a variável segue uma distribuição Normal, temos $\bar{x} = 24$, $s^2 = 25,75$ e $t_{1-\alpha/2;24} = 2,06$, onde $1 - \alpha/2 = 0,975$ e usamos 24 graus de liberdade. Logo, os limites do intervalo são:

$$l_1 = 24 - 2,06\frac{5,07}{\sqrt{25}} = 24 - 2,089 \text{ e } l_2 = 24 + 2,06\frac{5,07}{\sqrt{25}} = 24 + 2,089.$$

4. Pretende-se realizar um teste de hipótese de H_0: $\mu \geq 10$ versus H_1: $\mu < 10$ com n=25, $\sigma_X = 3$ e $\alpha = 0,1$.

(a) Aponte a estatística de teste e o critério de decisão (ou seja, as regiões de rejeição e de aceitação) a serem usados.

(b) Diga, no contexto de teste de hipóteses, o que são: o erro I e o erro II. Determine a P(erro de tipo II) para $\mu = 8,5$.

(c) Que decisão deve ser tomada se $\overline{x}$=8,2 nesta amostra com 25 elementos.

Resposta: **(a)** H_0: $\mu = 10$ versus H_1: $\mu < 10$, sendo a hipóteses H_0 considerada como simples a região de rejeição deveria ser $\overline{X} \leq \overline{x}_c$. Assim, $\Phi(z_c) = 0,10 \Longrightarrow \Phi(-z_c) = 0,90 \Longrightarrow -z_c = 1,28 \Longrightarrow z_c = -1,28$. Como $z_c = \frac{\overline{x}_c - \mu_0}{\sigma/\sqrt{n}} \Longrightarrow \overline{x}_c = -1,28 \times 3/5 + 10 = 9,23$. Assim, rejeita-se H_0 se $\overline{x}_{obs} < 9,23$, caso contrário não há evidência para a rejeição de H_0 .

(b) Erro I: Rejeitar H_0 quando H_0 é verdadeira. Erro II: Aceitar H_0 quando H_0 é falsa. A $P(erro\,II)$ é obtida supondo $\mu = 8,5$ e calculando $P(\overline{X}_c < 9,23) = \Phi(\frac{9,23-8,5}{3/5}) \approx \Phi(1,22) = 0,8888$.

(c) Como $\overline{x}_{obs} = 8,2 < 9,23$ há evidência para a rejeição da hipótese H_0.

Boa prova!