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

Prova # 02 26-11-2013

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

---

1. Dado um conjunto de pares (x_i, y_i), i variando de 1 até n.

   (a) Que relação existe entre as médias das variáveis X e Y=a+X, onde "a" é uma constante (isto é, $y_i = a + x_i$)? E entre as variâncias de X e Y? E entre as distâncias interquartis (DIQ) de X e Y? Qual o coeficiente de correlação de X e Y?

   (b) Responda as mesmas quatro perguntas do item (a) se Y=bX, onde "b" é uma constante positiva.

   Solução

   (a) $\bar{y} = \frac{\sum(a+x_i)}{n} = a + \bar{x}$;
   $s_y^2 = \frac{\sum(y_i-\bar{y})^2}{n-1} = \frac{\sum(a+x_i-(a+\bar{x}))^2}{n-1} = s_x^2$;
   $Q_1(y) = y_{(\frac{n+3}{4})} = a + x_{(\frac{n+3}{4})} = a + Q_1(x)$ e $Q_3(y) = y_{(\frac{3n+1}{4})} = a + x_{(\frac{3n+1}{4})} = a + Q_3(x)$
   $DIQ(y) = Q_3(y) - Q_1(y) = DIQ(x)$;
   $r_{xy} = \frac{\sum x_iy_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} = \frac{\sum x_i(a+x_i) - n\bar{x}(a+\bar{x})}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum(a+x_i)^2 - n(a+\bar{x})^2)}} = 1$

   (b) $\bar{y} = \frac{\sum(bx_i)}{n} = b\ \bar{x}$;
   $s_y^2 = \frac{\sum(y_i-\bar{y})^2}{n-1} = \frac{\sum(bx_i-(b\bar{x}))^2}{n-1} = b^2 s_x^2$;
   $Q_1(y) = y_{(\frac{n+3}{4})} = b\ x_{(\frac{n+3}{4})} = b\ Q_1(x)$ e $Q_3(y) = y_{(\frac{3n+1}{4})} = b\ x_{(\frac{3n+1}{4})} = b\ Q_3(x)$, porque b>0. Assim, $DIQ(y) = Q_3(y) - Q_1(y) = b\ DIQ(x)$;
   $r_{xy} = \frac{\sum x_iy_i - n\bar{x}\bar{y}}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum y_i^2 - n\bar{y}^2)}} = \frac{\sum x_i(bx_i) - n\bar{x}(b\bar{x})}{\sqrt{(\sum x_i^2 - n\bar{x}^2)(\sum(bx_i)^2 - n(b\bar{x})^2)}} = \frac{b}{\sqrt{b^2}} = 1$, porque b>0

2. Os dados abaixo correspondem a medições do QI (Quociente de Inteligência) para uma amostra de 30 pessoas.

```
117 106 106 100 90 91 130  96 98 99 112 130 102 101 94
100 126 83 106 121 79 103 126 78 79  97 100  93 103 107
```

   Deseja-se estimar a média populacional μ dessa variável, supondo distribuição Normal.

   (a) Qual a probabilidade de que o erro absoluto na estimação com base nessa amostra seja inferior a 3 unidades?

   (b) Qual seria o tamanho de uma nova amostra para que, com 95% de probabilidade, o erro absoluto de estimação seja inferior a 3 unidades?

   Obs.: Para facilitar os cálculos, informa-se que: $\sum x_i = 3.073$ e $\sum x_i^2 = 320.757$.

   Solução
   Sabemos que $(X_1, ..., X_{30})$ é uma amostra aleatória de uma distribuição $N(\mu; \sigma^2)$.
   Daí $\bar{X} \sim N(\mu, \sigma^2/30)$.

   (a) $P[|\bar{X} - \mu| < 3] = P[\frac{|\bar{X}-\mu|}{S/\sqrt{30}} < \frac{3\sqrt{30}}{S}]$
   Como $T = \frac{\bar{X}-\mu}{S/\sqrt{30}} \sim t$ de Student com 29 g.l. e $S_{obs} = \sqrt{\frac{320757 - 3073^2/30}{29}} = 14,36$,

$P[|\bar{X}-\mu|<3] = P[|T| < \frac{3\sqrt{30}}{14,36}] = P[|T| < 1,144]$
Da tabela da t, vemos que $P[T < 0,854] = 0,8$, o que implica que $P[|T| < 0,854] = 0,6$, e que $P[T < 1,311] = 0,9$, o que implica que $P[|T| < 1,311] = 0,8$.
Daí, $0,6 < P[|\bar{X}-\mu| < 3] < 0,8$.
Por outro lado, para 29 g.l. a distribuição t de Student já se aproxima muito da Normal Padrão. Então:
$P[|\bar{X}-\mu| < 3] = P[|Z| < 1,144] = 2\Phi(1,144) - 1 = 2 \times 0,8729 - 1 = 0,7458.$

(b) Para que tenhamos $P[|\bar{X}-\mu| < 3] = 0,95$, o tamanho n da amostra terá de ser maior que 30. Então, com maior razão, podemos aproximar a t de Student com n-1 g.l. pela N(0;1). Daí ,

$$n = \left(\frac{z_{0,975} \times S_{obs}}{3}\right)^2 = \left(\frac{1,96 \times 14,36}{3}\right)^2 \approx 88.$$

3. Sabe-se que a altura dos indivíduos de uma certa população é Normalmente distribuída. Em uma amostra aleatória de tamanho 16, obteve-se o seguinte intervalo de 95% de confiança para a altura média populacional $(177,66;\ 185,5)$ (em cm).

(a) Forneça estimativas pontuais para a altura média e para o desvio padrão da altura.

(b) Qual seria o novo intervalo se a confiança exigida fosse 99%?

(c) Qual deveria ser o tamanho de uma nova amostra para que o novo intervalo tivesse a mesma amplitude do intervalo anterior, porém agora com 99% de confiança, supondo que o desvio padrão populacional seja igual a 9 cm?

Solução

(a) Estimativa pontual para a média: $\overline{x} = \frac{177,66+185,50}{2} = 181,58$.
Estimativa pontual para o desvio padrão: temos que $185,50 - 181,58 = 3,92$ e, portanto,

$$3,92 = t_{1-\alpha/2}\frac{s}{\sqrt{n}} = \overbrace{t_{0,975}}^{2,13}\frac{s}{\sqrt{16}} \Rightarrow s = 7,3615.$$

(b)

$$\begin{aligned} IC(0,99) &= \left(\overline{x} - t_{0,995}\frac{s}{\sqrt{n}};\ \overline{x} + t_{0,995}\frac{s}{\sqrt{n}}\right) \\ &= \left(181,58 - 2,95 \times \frac{7,3615}{4};\ 181,58 + 2,95 \times \frac{7,3615}{4}\right) = (176,12;\ 186,98). \end{aligned}$$

(c) Se $\sigma = 9$ cm e o novo Int. Confiança para μ ao nível 0,99 tem a mesma amplitude do anterior), temos:
185,5 - 177,66 $= 2\ z_{0,995} \times \frac{9}{\sqrt{n}}$, o que implica que $n = \left(\frac{2\times 2,58\times 9}{7,84}\right)^2 \approx 35,087, \quad n = 36.$

4. Suponha que os tempos de vida, em horas, de certos semicondutores a laser sejam Normalmente distribuidos com média μ e variância σ^2.
Para testar $H_0 : \mu \leq 7.000$ versus $H_1 : \mu > 7.000$, uma amostra aleatória de tamanho 25 foi observada e apresentou os seguintes resultados:
$\sum x_i = 179.500, \quad \sum(x_i - \bar{x})^2 = 8.640.000$

(a) Qual a sua conclusão, ao nível de significância de 1%?

(b) Determine o menor intervalo no qual você sabe que o p-valor associado aos dados obtidos se encontra.

(c) Com base no intervalo obtido no item (b), diga qual seria sua decisão se o nível de significância fosse de 5%. E se fosse 10%?

Solução

Vamos testar $H_0 : \mu \leq 7000$ horas contra $H_1 : \mu > 7000$ horas.

(a) Usaremos $\alpha = 0,01$ e uma amostra de tamanho n = 25. Já que σ é desconhecido, a estatística de teste é $T = \frac{\bar{x}-7000}{S/\sqrt{25}}$, cuja distribuição é uma t de Student com 24 g.l., se $\mu = 7000$.
A região de rejeição é definida por Tobs $> t_{0,99;24} = 2,492$. (Pela tabela da t)
Como $\sum x_i = 179500, \bar{x} = \frac{179500}{25} = 7180$.
Como $\sum(x_i - \bar{x})^2 = 8640000, s = \sqrt{\frac{8640000}{24}} = 600$.
Então, $T_{obs} = \frac{7180-7000}{600/\sqrt{25}} = 1,5$.
Uma vez que $T_{obs} = 1,5 < 2,492, H_0$ deve ser aceita.

(b) p-valor $=P[T > 1,5]$, onde $T \sim t$ de Student com 24 g.l.
Pela tabela da t, vemos que $P[T < 1,318] = 0,90$ e $P[T < 1,711] = 0,95$.
Daí, $P[T > 1,711] < p - valor < P[T > 1,318]$, ou seja, $0,05 < p - valor < 0,10$.

(c) A partir da conclusão do item (b), vemos que se $\alpha = 0,05, H_0$ seria aceita; e se $\alpha = 0,10$, H_0 seria rejeitada. Isto porque o p-valor é o menor valor de α para o qual H_0 ainda seria rejeitada, com os dados disponíveis.

Boa prova!