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

Prova #2 25-11-2014

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

1. Um grupo de engenheiros deseja estudar a relação entre o número médio de itens defeituosos produzidos pelas máquinas a cada lote de 1.000 peças e o tempo de funcionamento dessas máquinas, em anos. Para isso, coletaram esses dados para 10 máquinas:

X = Tempo de funcionamento	1	1	2	4	5	6	9	9	11	x_{10}
Y = Número médio de itens defeituosos	6	8	15	12	13	18	28	21	26	y_{10}

Sabe-se que: $\sum_{i=1}^{10} x_i = 63$ $\quad \sum_{i=1}^{10} y_i = 182$ $\quad \sum_{i=1}^{10} x_i^2 = 591$ $\quad \sum_{i=1}^{10} y_i^2 = 4088$ $\quad \sum_{i=1}^{10} x_i y_i = 1517$

(a) Quais os valores da observação faltante (x_{10}, y_{10})? Construa o gráfico de dispersão das variáveis X e Y.

(b) Calcule o coeficiente de correlação amostral entre X e Y, interpretando o valor obtido.

(c) Calcule os coeficientes da reta de regressão $Y = a + bX$ e utilize a equação desta reta para estimar a proporção esperada de itens defeituosos entre os que forem produzidos por uma máquina já em funcionamento por 10 anos.

2. Sejam $X_1, X_2, \ldots X_n$ uma amostra aleatória da variável aleatória X com distribuição Bernoulli(p).

(a) Qual a distribuição de $Y = X_1 + X_2 + \ldots + X_n$?

(b) Considere os dois seguintes estimadores para p: $T_1 = \dfrac{Y}{n}$ e $T_2 = \dfrac{Y + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}$. Esses estimadores são não tendenciosos(não viciados) para p?

(c) Expresse o erro quadrático médio (EQM) desses dois estimadores em função de n e p.

(d) Suponha n=4 e faça um único gráfico mostrando o comportamento do EQM em função de p para os dois estimadores.

3. Um procedimento químico utilizado por uma indústria gera em média μ mg de cloreto de sódio como produto final de uma reação. A quantidade de cloreto obtida a cada vez apresenta um comportamento aleatório $N(\mu, \sigma^2)$, onde $\sigma^2 = 2500$. A partir de 100 repetições do mesmo procedimento, obteve-se em média 200 mg de cloreto.

(a) Calcule o intervalo de confiança de 99% para μ. Interprete o resultado.

(b) Considere que a indústria modificou o procedimento anterior e que agora a variância da quantidade de cloreto produzida não é mais conhecida. Com uma amostra de 51 repetições desse novo procedimento, obteve-se média e desvio-padrão amostrais respectivamente iguais a 250 mg e 60 mg. Obtenha o intervalo de confiança de 99% para o μ que o novo procedimento é capaz de produzir.

(c) Com base nos itens anteriores, você acredita que o novo procedimento seja capaz de produzir, em média, uma quantidade de cloreto de sódio diferente da que era obtida anteriormente? Qual é o nível de confiança associado à sua conclusão? (Suponha que as seleções de amostra relativas aos dois procedimentos químicos tenham sido realizadas independentemente uma da outra).

4. O teor da substância S no minério M_0 se comporta como uma Normal de média 25% e desvio padrão 10%. Já o teor da substância S no minério M_1 também se comporta como uma Normal, porém esta com média 30% e desvio padrão 10%. A Companhia C dispõe de um volume apreciável de um minério cuja natureza (M_0 ou M_1) é desconhecida. Como S é uma impureza indesejável, só interessa à Cia. C utilizar esse material no caso de ele ser M_0. Serão coletadas n unidades amostrais desse material a serem analisadas quimicamente e, com base nos resultados, se pretende testar H_0: o minério é M_0 contra H_1: o minério é M_1.

(a) Qual deve ser o tamanho n da amostra para que as probabilidades do Erro I e do Erro II sejam respectivamente de 0,01 e 0,10?

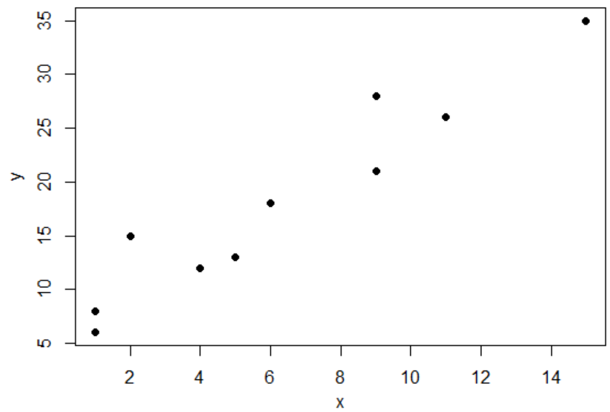
(b) Usando o n obtido em (a), qual a decisão a ser tomada se $\bar{X}_{\text{obs}} = 28\%$? Justifique a sua resposta.

Sugestão: Note que quanto maior for a média amostral, maiores serão os motivos para se rejeitar H_0.

Boa prova!

SOLUÇÕES

1. **(a)** $X_{10} = \sum_{i=1}^{10} X_i - \sum_{i=1}^{9} X_i = 63 - (1 + \ldots + 11) = 15$ $\qquad Y_{10} = \sum_{i=1}^{10} Y_i - \sum_{i=1}^{9} Y_i = 182 - (6 + \ldots + 26) = 35$



(b)

$$r_{xy} = \frac{\sum_{i=1}^{10} X_i Y_i - n\bar{X}\bar{Y}}{\sqrt{\sum_{i=1}^{10} X_i^2 - n\bar{X}^2}\sqrt{\sum_{i=1}^{10} Y_i^2 - n\bar{Y}^2}} = \frac{1517 - 10 \times 6,3 \times 18,2}{\sqrt{591 - 10 \times (6,3)^2}\sqrt{4088 - 10 \times (18,2)^2}} = 0,95464$$

Como esse valor está próximo de 1, as variáveis número médio de itens defeituosos produzidos e tempo de funcionamento da máquina tem forte relação linear positiva.

(c)

$$\hat{b} = \frac{\sum_{i=1}^{10} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{10} X_i^2 - n\bar{X}^2} = \frac{1517 - 10 \times 6,3 \times 18,2}{591 - 10 \times (6,3)^2} = 1,908 \qquad \hat{a} = \bar{Y} - \hat{b}\bar{X} = 18,2 - 1,908 \times 6,3 = 6,178$$

Quando X=10 temos que o Y estimado será: $\hat{Y} = \hat{a} + \hat{b} \times 10 = 6,178 + 1,908 \times 10 = 25,258$. Então a proporção estimada será: $\frac{25,258}{1000} = 2,526\%$

2. **(a)** $X_1, \ldots, X_n$ são iid's Bernoulli(p) $\Rightarrow Y = \sum_{i=1}^{n} X_i \sim$ binomial(n,p).

(b) $E(T_1) = E\left(\frac{Y}{n}\right) = \frac{np}{n} = p \quad \Rightarrow \quad T_1$ é não tendencioso(não viciado) para p.

$E(T_2) = E\left(\frac{Y + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}\right) = \frac{np + \frac{\sqrt{n}}{2}}{n + \sqrt{n}} \quad \Rightarrow \quad T_2$ é tendencioso(viciado) para p.

(c) Viés$(T_1) = 0$

Viés$(T_2) = \frac{np + \frac{\sqrt{n}}{2}}{n + \sqrt{n}} - p = \frac{np + \frac{\sqrt{n}}{2} - np - \sqrt{n}p}{n + \sqrt{n}} = \frac{\sqrt{n}\left(\frac{1}{2} - p\right)}{\sqrt{n}(\sqrt{n} + 1)} = \frac{\frac{1}{2} - p}{\sqrt{n} + 1}$

$$\text{Var}(T_1) = \text{Var}\left(\frac{Y}{n}\right) = \frac{np(1-p)}{n^2} = \frac{p(1-p)}{n}$$

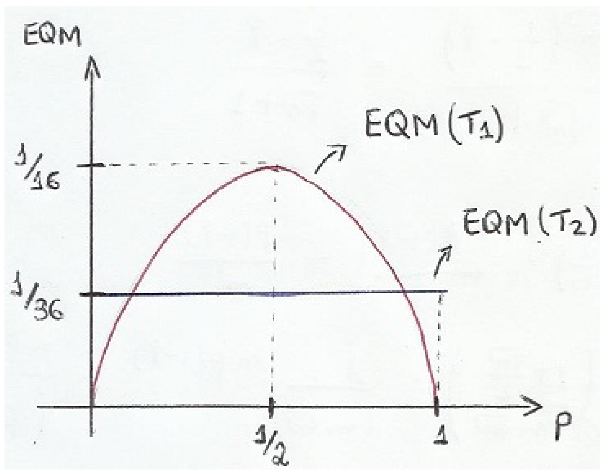
$$\text{Var}(T_2) = \text{Var}\left(\frac{Y + \frac{\sqrt{n}}{2}}{n + \sqrt{n}}\right) = \frac{np(1-p)}{(n+\sqrt{n})^2} = \frac{np(1-p)}{(\sqrt{n}(\sqrt{n}+1))^2} = \frac{p(1-p)}{(\sqrt{n}+1)^2}$$

Como $\text{EQM}(T_i) = \text{Viés}(T_i)^2 + \text{Var}(T_i)$, temos que:

$$\text{EQM}(T_1) = \frac{p(1-p)}{n}$$

$$\text{EQM}(T_2) = \left(\frac{\frac{1}{2} - p}{\sqrt{n}+1}\right)^2 + \frac{p(1-p)}{(\sqrt{n}+1)^2} = \frac{\frac{1}{4} - p + p^2 + p - p^2}{(\sqrt{n}+1)^2} = \frac{1}{4(\sqrt{n}+1)^2}$$

(d) Com n=4, temos: $\text{EQM}(T_1) = \dfrac{p(1-p)}{4}$ e $\text{EQM}(T_2) = \dfrac{1}{36}$.



3. **(a)**

$$\bar{x} = 200;\ \sigma = 50;\ n = 100;\ q_{0,995} = 2,573$$

$$\bar{x} \pm q_{0,995} \times \frac{\sigma}{\sqrt{n}} = \left(200 - 2,573 \times \frac{50}{10};\ 200 + 2,573 \times \frac{50}{10}\right) = (187,135;\ 212,865)$$

(b)

$$\bar{x} = 250;\ s = 60;\ n = 51;\ t_{0,995} = 2,678; q_{0,995} = 2,573$$

Resolvendo utilizando a t-student:

$$\bar{x} \pm t_{0,995} \times \frac{s}{\sqrt{n}} = \left(250 - 2,678 \times \frac{60}{\sqrt{51}};\ 250 + 2,678 \times \frac{60}{\sqrt{51}}\right) = (227,5;\ 272,5)$$

Resolvendo utilizando a aproximação para normal:

$$\bar{x} \pm q_{0,995} \times \frac{s}{\sqrt{n}} = \left(250 - 2,573 \times \frac{60}{\sqrt{51}};\ 250 + 2,573 \times \frac{60}{\sqrt{51}}\right) = (228,4;\ 271,6)$$

(c) Ambos os intervalos conterão simultaneamente a respectiva média populacional com probabilidade $0,99^2 = 0,9801$. Logo, como a interseção dois dois intervalos é vazia, as médias devem ser diferentes ao nível de confiança de $98,01\%$

4. **(a)** Sejam $\mu_0 = 25, \mu_1 = 30, \alpha = 0,01, \beta = 0,10$, e $\sigma = 10$.

Se H_0 é verdadeira, $X_1, X_2, \ldots X_n$ são v.a's iid's com distribuição $N(\mu_0; \sigma^2)$, e se H_0 é falsa, $X_1, X_2, \ldots X_n$ são v.a's iid's com distribuição $N(\mu_1; \sigma^2)$

Já que se trata de um teste de média, a estatística de teste será a média amostral $\bar{X}$.

Como $\mu_0 = 25 < 30 = \mu_1$, o critério de decisão será do tipo: Rejeitar H_0 se $\bar{X}_{\text{obs}} > \bar{x_c}$, onde o ponto de corte $\bar{x_c}$ é uma constante a determinar. Caso contrário, aceitar H_0.

Então $\alpha = \text{P(Erro I)} = \text{P}(\bar{X} > \bar{x_c})$, se H_0 é verdadeira. Neste caso $\bar{X} \sim N\left(\mu_0; \frac{\sigma^2}{n}\right)$.

Logo, padronizando temos: $\alpha = P\left(\frac{\bar{X} - \mu_0}{\frac{\sigma}{\sqrt{n}}} > \frac{\bar{x_c} - \mu_0}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z > \frac{(\bar{x_c} - \mu_0)\sqrt{n}}{\sigma}\right)$, o que implica em $\text{Z}_{(1-\alpha)} = \frac{(\bar{x_c} - \mu_0)\sqrt{n}}{\sigma} \quad \Rightarrow \quad \bar{x_c} = \frac{\sigma Z_{(1-\alpha)}}{\sqrt{n}} + \mu_0 \qquad (I)$

Por outro lado, $\beta = \text{P(Erro II)} = \text{P}(\bar{X} \leq \bar{x_c})$, se H_0 é falsa. Neste caso $\bar{X} \sim N\left(\mu_1; \frac{\sigma^2}{n}\right)$.

Logo, padronizando temos: $\beta = P\left(\frac{\bar{X} - \mu_1}{\frac{\sigma}{\sqrt{n}}} \leq \frac{\bar{x_c} - \mu_1}{\frac{\sigma}{\sqrt{n}}}\right) = P\left(Z \leq \frac{(\bar{x_c} - \mu_1)\sqrt{n}}{\sigma}\right)$, o que implica em $\text{Z}_\beta = \frac{(\bar{x_c} - \mu_1)\sqrt{n}}{\sigma} \quad \Rightarrow \quad \bar{x_c} = \frac{\sigma Z_\beta}{\sqrt{n}} + \mu_1 \qquad (II)$

De (I) e (II) decorre que $\frac{\sigma Z_{(1-\alpha)}}{\sqrt{n}} + \mu_0 = \frac{\sigma Z_\beta}{\sqrt{n}} + \mu_1$, o que implica em:

$$n = \left(\frac{\sigma(Z_{(1-\alpha)} - Z_\beta)}{\mu_1 - \mu_0}\right)^2 = \left(\frac{10(2,33 - (-1,28))}{30 - 25}\right)^2 \cong 52.$$

(b) De (I) temos que $\bar{x_c} = \dfrac{10 \times 2,33}{\sqrt{52}} + 25 = 28,23$.

Como $\bar{X}_{\text{obs}} = 28 < 28,23, H_0$ deve ser aceita.