UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Avaliação Final de Probabilidade e Estatística **05-07-2018**
Atenção: Não serão aceitas respostas sem justificativa.
Resolver as questões nos espaços apropriados.

1. Todas as receitas de analgésico que chegam a um ambulatório médico prescrevem aspirina (com 60% de probabilidade) ou dipirona sódica (com 40% de probabilidade). Num determinado dia há em estoque 70 comprimidos de aspirina e 50 comprimidos de dipirona sódica. Se nesse dia são prescritas 100 receitas de analgésicos, calcule:
   (a) A média e o desvio padrão do número de prescrições de aspirina. Indique a expressão matemática para o modelo probabilístico desta variável;
   (b) A probabilidade aproximada de todas as 100 receitas serem atendidas.

2. No processo de modelagem de um problema, sabe-se que determinada variável tem uma distribuição de probabilidade simétrica. Entretanto há controvérsias sobre se ela deve ser uma Uniforme ou uma Normal. Considere as variáveis aleatórias X $\sim$ U[a; b] e Y $\sim N(\mu;\sigma^2)$, tais que E(X) $=\mu$ e Var(X) $=\sigma^2$. Para comparar as distribuições de X e de Y, calcule:
   (a) $\frac{P(\mu-\sigma<Y<\mu+\sigma)}{P(\mu-\sigma<X<\mu+\sigma)}$
   (b) $\frac{P(\mu-\sqrt{3}\sigma<Y<\mu+\sqrt{3}\sigma)}{P(\mu-\sqrt{3}\sigma<X<\mu+\sqrt{3}\sigma)}$
   (c) $\frac{diq(Y)}{diq(X)}$, (diq, significa distância interquartil).

3. Determinado estudo visa relacionar a renda mensal familiar com o consumo mensal em despesas fundamentais (saúde, alimentação, educação). Foram coletados dados de renda e consumo de 36 famílias residentes em determinada região da cidade e, a partir deles, foram obtidos os seguintes valores:

$$\sum x_i = 800,8\,; \quad \sum x_i^2 = 20759,62\,; \quad \sum y_i = 545,8\,; \quad \sum y_i^2 = 9568,44\,; \quad \sum x_i\, y_i = 13543,25\,,$$

   onde x representa a variável renda e y representa a variável consumo (ambos em salários mínimos).
   (a) Calcule média e desvio padrão para ambas x e y.
   (b) Forneça a reta de regressão associada ao par (renda,consumo).
   (c) Calcule a correlação. Com base no valor obtido, diga se a reta de regressão parece fornecer uma boa descrição da relação entre renda e consumo.

4. Numa linha de produção um engenheiro precisa estimar a proporção de peças defeituosas. Ele deseja que sua estimativa não se desvie da proporção verdadeira por mais de 0,06 com probabilidade de 95%.
   (a) Quantas peças se precisa examinar para satisfazer essa exigência?
   (b) Foi verificado que esse tipo de defeito afeta no máximo 15% das peças. Utilizando essa informação adicional, para quanto seria possível diminuir o tamanho da amostra?
   (c) Com esse último tamanho amostral (item b) foi extraída uma amostra para a qual se obteve $\hat{p} = 0,115$. Quanto seria o p-valor do teste, $H_0 : p \leq 0,10$ versus $H_1 : p > 0,10$? Que decisão deveria ser tomada ao nível de significância de 5%?

Boa Prova!

Solução

1. Vamos considerar a variável aleatória X: número de prescrições de aspirina. Note que sabendo-se o número de prescrições de aspirina sabe-se também o número de prescrições de dipirona sódica. Suposição X $\sim$ Bin(100; 0, 60). Então,

   (a) E(X) = np = 100 $\times$ 0,60 = 60
   Var(X) = np(1 - p) = 100 $\times$ 0,60 $\times$ 0,40 = 24
   DP(X) = $\sqrt{24} \approx$ 4,90
   p(x) = $\begin{pmatrix} 100 \\ x \end{pmatrix} \times 0,6^x \times 0,4^{100-x}$; x=0,...,100

   (b) Assim, a distribuição de X pode ser aproximada por Y $\sim$ N(60; 4,90^2).
   Do enunciado pode-se deduzir que P(todas as prescrições serem atendidas) = P(50 $\leq X \leq$ 70).
   Então, Fazendo a correção de continuidade:
   P(50 - 0,5 $\leq X \leq$ 70 + 0,5) $\approx$ P(49,5 $\leq Y \leq$ 70,5)
   = P ((49,5 - 60)/4,90 $\leq Z \leq$ (70,5 - 60)/ 4,90) = P(-2,14 $\leq Z \leq$2,14)
   = P(Z $\leq$ 2,14) - P(Z $\leq$ -2,14)
   = Φ(2,14) - [1 - Φ(2,14)] = 0,9838 - [1 - 0,9838] = 0,9676 ou (96,76%).
   Portanto, a probabilidade de todas as prescrições serem atendidas é aproximadamente 0,9676 (ou 96,76%)

2. $E(X) = E(Y) \Longrightarrow \frac{a+b}{2} = \mu$ e $DP(X) = DP(Y) \Longrightarrow \frac{(b-a)}{\sqrt{12}} = \sigma \Longrightarrow \frac{\sigma}{b-a} = \frac{1}{2\sqrt{3}}$ (*)
   Esta identidade (*) será usada nos 3 itens a seguir.

   (a) $P(\mu - \sigma < Y < \mu + \sigma) = P(-1 < \frac{Y-\mu}{\sigma} < 1)) = 2\Phi(1) - 1$
   $P(\mu - \sigma < X < \mu + \sigma) = \frac{(\mu+\sigma)-(\mu-\sigma)}{b-a} = \frac{2\sigma}{b-a} = \frac{1}{\sqrt{3}}$
   Então, $\frac{P(\mu-\sigma<Y<\mu+\sigma)}{P(\mu-\sigma<X<\mu+\sigma)} = (2\Phi(1) - 1)\sqrt{3} = 1,182$

   (b) $P(\mu - \sqrt{3}\sigma < Y < \mu + \sqrt{3}\sigma) = P(-\sqrt{3} < \frac{Y-\mu}{\sigma} < \sqrt{3}) = 2\Phi(\sqrt{3}) - 1$
   $P(\mu - \sqrt{3}\sigma < X < \mu + \sqrt{3}\sigma) = \frac{(\mu+\sqrt{3}\sigma)-(\mu-\sqrt{3}\sigma)}{b-a} = \frac{2\sqrt{3}\sigma}{b-a} = 1$
   Então, $\frac{P(\mu-\sqrt{3}\sigma<Y<\mu+\sqrt{3}\sigma)}{P(\mu-\sqrt{3}\sigma<X<\mu+\sqrt{3}\sigma)} = 2\Phi(\sqrt{3}) - 1 = 0,917$

   (c) diq(Y) = q_3(Y) - q_1(Y) = $(\mu + z_{0,75}\ \sigma) - (\mu + z_{0,25}\ \sigma)$,
   onde $z_{0,75} = 0,675$ e $z_{0,25} = -0,675$
   Portanto, diq(Y) = 2 x 0,675 σ = 1,35 σ.
   diq(X) = q_3(X) - q_1(X) = ((1/4) a+(3/4) b) - ((3/4) a+(1/4) b) = (b - a)/2
   Então, (diq(Y))/(diq(X)) = (1,35 σ)/((b - a)/2) = (1,35$\times$2)/(2$\sqrt{3}$) = 0,779.

3. (a)

$$\bar{x} = \frac{800,8}{36} = 22,24 \qquad \bar{y} = \frac{545,8}{36} = 15,16$$

   e

$$s_x = \sqrt{\frac{20759,62 - 36 \times 22,24^2}{35}} = 9,186 \qquad s_y = \sqrt{\frac{9568,44 - 36 \times 15,16^2}{35}} = 6,082\,.$$

   (b) A covariância amostral é

$$s_{xy} = \frac{13543,25 - 36 \times 22,24 \times 15,16}{35} = 40,158\,.$$

   Daí obtemos o coeficiente angular da reta de regressão

$$b = \frac{s_{xy}}{s_x^2} = \frac{40,158}{9,186^2} = 0,4759$$

   e o intercepto

$$a = \bar{y} - b\bar{x} = 15,16 - 0,4759 \times 22,24 = 4,576\,.$$

   Portanto a reta de regressão é $y = 0,4759x + 4,576$.

   (c) A correlação é

$$r_{xy} = \frac{s_{xy}}{s_x s_y} = \frac{40,158}{9,186 \times 6,082} = 0,7188\,.$$

   A correlação é suficientemente próxima de um para indicar que a reta de regressão possivelmente terá um bom

4. (a) O erro absoluto da estimativa é d=0,06. $P(|\hat{p} - p| < d) = 0,95$, com $\Phi(z) = 0,975$, temos z=1,96. Como não temos informação alguma sobre p, utilizaremos p = 0,5 que maximiza p(1-p) e calculamos o tamanho amostral:

$$n = \left(\frac{z}{d}\right)^2 p(1-p) = \left(\frac{1,96}{0,06}\right)^2 0,25 = 266,7 \approx 267$$

(b) Se p for no máximo 0,15, o novo tamanho diminuido seria:

$$n = \left(\frac{z}{d}\right)^2 p(1-p) = \left(\frac{1,96}{0,06}\right)^2 0,15 \times 0,85 = 136,06 \approx 137$$

(c) p-valor $= P\left(Z > \frac{\hat{p}-p_0}{\sqrt{\frac{p_0(1-p_0)}{n}}}\right) = 1 - \Phi\left(\frac{0,115-0,10}{\sqrt{\frac{0,10\times 0,90}{137}}}\right) = 1 - \Phi(0,59) = 1 - 0,7224 = 0,2776$

Como o p-valor é razoavelmente grande não se tem evidência para a rejeição da hipótese H_0. Esse experimento, com essa amostra, confirma que p deve ser menor que 10%.