UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Prova Final de Probabilidade e Estatística **26-07-2016**

Atenção: Não serão aceitas respostas sem justificativa. Resolver as questões nos espaços apropriados.

1. Suponha que a largura de um automóvel possa ser modelada por uma distribuição Normal com média 1,80m e desvio padrão 15cm. Admita também que para boa circulação, os carros devem ter uma distância de 50cm entre eles e de 30 cm entre eles e a parede. Essas mesmas distâncias devem ser preservadas quando os carros estiverem também estacionados.

   (a) Qual a probabilidade de que num corredor de 4,20 m possam passar carros indo e voltando? Você acha que a largura do corredor é suficiente?

   (b) Dimensionar a largura de uma garagem, para que, com probabilidade de pelo menos 0,7, cinco automóveis aleatoriamente escolhidos possam estacionar confortavelmente.

2. O tempo gasto pelo atendente de uma central para concluir o atendimento de um cliente, é independente dos demais atendimentos e pode ser descrito por uma variável aleatória com função densidade de probabilidade dada por:

$$f(x) = \begin{cases} 200x^{-3}, \text{para } x \geq a \\ 0, \text{ caso contrário.} \end{cases}$$

   (a) Calcule o valor da constante a para que $f(.)$ seja de fato uma função densidade.

   (b) Suponha que ao final de um atendimento, o atendente seja avaliado com uma pontuação de valor 0 (ruim), 5 (regular) ou 10 (bom). O atendimento é classificado como bom se ele demorar um tempo inferior a $2a$, como regular se ele demorar um tempo entre $2a$ e $3a$; e como ruim se ele demorar um tempo superior a $3a$. Definindo por Y a pontuação atribuida ao atendimento, forneça a função de distribuição acumulada de Y.

   (c) Suponha que no final do expediente, o atendente é pago com uma quantia dada por $Q = \dfrac{N^2}{5} - \dfrac{W^2}{6}$, em reais; onde N representa o número de clientes que ele atendeu ao longo do dia e W representa a quantidade desses atendimentos que foram classificados como ruins. Calcule o ganho médio desse atendente em um dia em que ele fizer 20 atendimentos.

3. Um fabricante de comprimidos de vitamina C concluiu, após sucessivas experiências, que para cada lote fabricado, o conteúdo das suas embalagens, tem distribuição normal. Além disso, constatou que o conteúdo médio das embalagens varia de um lote para outro embora o desvio padrão se mantenha igual a 20 comprimidos. Deseja-se estimar o número médio μ de comprimidos por embalagem, em um determinado lote.

   (a) Para uma amostra piloto com 15 embalagens extraídas aleatoriamente desse lote, qual é a probabilidade de que a estimativa da média não difira em mais de 5 comprimidos do parâmetro populacional.

   (b) Determine quantas embalagens adicionais (além das primeiras 15) deveriam ser sorteadas desse mesmo lote de modo a compor uma nova amostra, para a qual o módulo da diferença entre a estimativa e o parâmetro seja menor que 5 comprimidos com com 95% de confiança.

4. Dentro de alguns meses ocorrerão em todo Brasil eleições municipais para prefeitos e vereadores. Várias pesquisas deverão ser divulgadas periodicamente para acompanhar as intenções de votos relativas aos diversos candidatos.

   (a) O que significa quando um apresentador de telejornal afirma que, com confiança de 95%, um candidato a prefeito teve a percentagem das intenções de voto $(100 \times \hat{p})$ com uma margem de erro de dois pontos percentuais para mais ou para menos?

   (b) Considerando uma amostra de tamanho 1.500, se a proporção dos entrevistados que declararam sua intenção de votar no candidato foi $\hat{p} = 55\%$, determine o intervalo de confiança, não conservativo, de 95% para a proporção de votos do candidato.

   (c) Realize um teste de hipóteses, ao nível de significância de 5%, para verificar se o candidato com esta intenção de voto pode ficar tranquilo quanto a possibilidade de vencer já no primeiro turno (ter mais de 50% dos votos).

   Obs.: Embora na prática muitas vezes sejam adotados métodos de seleção mais elaborados, para simplificar, admita que nessas pesquisas a amostra é selecionada de forma totalmente aleatória.

Boa prova!

Soluções

1. Seja X_i = Largura de 1 automóvel. Sabemos que $X_i \sim N(180, 15^2)$

   (a) Seja a v.a. Y=ocupação do corredor com 2 carros e as folgas.
   $Y = X_1 + X_2 + 2 * 30 + 50 = X_1 + X_2 + 110$. Então: $E(Y) = E(X_1) + E(X_2) + 110 = 470$ e como X_1 e X_2 são v.a. independentes $Var(Y) = Var(X_1) + Var(X_2) = 450$. Assim:
   $Y \sim N(\mu_Y = 470; \quad \sigma_Y^2 = 450 = 21,21^2)$
   $P(Y < 420) = \Phi(\frac{420-470}{21,21}) = \Phi(-2,36) = 1 - \Phi(2,36) = 1 - 0,9909 = 0,0091$
   Como a probabilidade é muito baixa a largura do corredor não comporta, de forma confortável, a passagem dos 2 carros.

   (b) Deseja-se saber qual é o valor da largura, y_0, da garagem. Seja agora:
   $Y = \sum_{i=1}^{5} X_i + 2 \times 30 + 4 \times 50 = \sum_{i=1}^{5} X_i + 260$. Assim:
   $Y \sim N(\mu_Y = 1160; \quad \sigma_Y^2 = 5 \times 15^2 = 1125 = 33,54^2)$
   $P(Y < y_0) = 0,7 \Rightarrow \Phi(z_0 = \frac{y_0 - 1160}{33,54}) = 0,7$
   Mas, $\Phi(z_0) = 0,7 \Rightarrow z_0 = 0,52$. Assim,
   $y_0 = z_0 \times \sigma_Y + \mu_Y = 0,52 \times 33,54 + 1160 = 1177,44$,
   A largura da garagem deve ser de pelo menos 11m e 77 cm.

2. (a) Para que $f(x)$ seja de fato uma função densidade de probabilidade (f.d.p.), é preciso que:
   (I) $f(x) \geq 0, \ \forall x \in R$ (II) $\int_{-\infty}^{\infty} f(x)dx = 1$.
   Pela restrição (II), temos: $\int_a^\infty 200x^{-3}dx = 1 \quad \Rightarrow \quad \left[\frac{200x^{-2}}{-2}\right]_a^\infty = 1 \ \Rightarrow \ 0 - \left(-\frac{100}{a^2}\right) = 1 \ \Rightarrow \ a = \pm 10$.
   Porém, note que com $a = -10$, temos que o requisito (I) falha, pois $f(x) < 0$ para alguns valores de $x \in R$, como por exemplo $x = -5$. Por outro lado, se $a = 10$, temos também satisfeita a restrição (I) e consequentemente esse é o único valor de a que torna $f(x)$ uma f.d.p. de fato.

   (b) Temos que Y é uma variável aleatória discreta com suporte $R_y = \{0, 5, 10\}$ e as probabilidades de cada um desses valores dadas por:
   $P(Y = 10) = \int_{10}^{20} 200x^{-3}dx = 200\left[-\frac{1}{2x^2}\right]_{10}^{20} = -100\left(\frac{1}{400} - \frac{1}{100}\right) = \frac{3}{4} = \frac{27}{36}$.
   $P(Y = 5) = \int_{20}^{30} 200x^{-3}dx = 200\left[-\frac{1}{2x^2}\right]_{20}^{30} = -100\left(\frac{1}{900} - \frac{1}{400}\right) = \frac{1}{4} - \frac{1}{9} = \frac{5}{36}$.
   $P(Y = 0) = \int_{30}^{\infty} 200x^{-3}dx = 200\left[-\frac{1}{2x^2}\right]_{30}^{\infty} = -100\left(0 - \frac{1}{900}\right) = \frac{1}{9} = \frac{4}{36}$.
   Logo, temos a seguinte função distribuição acumulada para a variável Y:
   $$F(y) = \begin{cases} 0, \text{para } y < 0 \\ \frac{1}{9}, \text{para } 0 \leq y < 5 \\ \frac{1}{4}, \text{para } 5 \leq y < 10 \\ 1, \text{para } y \geq 10 \end{cases}.$$

   (c) Para $N = 20$ temos:
   $$E(Q) = E\left[\frac{20^2}{5} - \frac{W^2}{6}\right] = E\left[80 - \frac{W^2}{6}\right] = 80 - \frac{E(W^2)}{6}.$$
   Então, basta calcular o valor de $E(W^2)$ e para isso, vamos analisar qual a distribuição da variável W. Note que cada atendimento tem probabilidade $\frac{4}{36} = \frac{1}{9}$, independente dos demais, de receber uma avaliação ruim. Com isso, temos que a variável W pode ser representada como a soma de $N = 20$ variáveis independentes com distribuição Bernoulli de igual parâmetro $p = \frac{1}{9}$.
   Logo: $W \sim \text{Binom}\left(20, \frac{1}{9}\right)$
   $$E(W) = np = 20 \times \frac{1}{9} = \frac{20}{9} \qquad Var(W) = np(1-p) = 20 \times \frac{1}{9} \times \frac{8}{9} = \frac{160}{81}$$
   Agora, podemos escrever:
   $E(W^2) = Var(W) + E(W)^2 = \frac{160}{81} + \left(\frac{20}{9}\right)^2 = \frac{160+400}{81} = \frac{560}{81}$.
   Por fim, temos que:
   $$E(Q) = 80 - \frac{E(W^2)}{6} = 80 - \frac{1}{6} \times \frac{560}{81} = 80 - \frac{560}{486} \cong R\$ \ 78,85.$$

3. Seja $X =$ o número de comprimidos de vitamina C de uma embalagem.
Sabemos que $X \sim N(\mu,\ \sigma^2 = 20^2)$

   (a) $\hat{\mu} = \overline{X} = \frac{\sum X_i}{n} \sim N(\mu_{\overline{X}} = \mu,\ \sigma^2_{\overline{X}} = \frac{20^2}{15})$
   Então, $P(|\overline{X} - \mu| < 5) \ = \ P(|Z| < \frac{5\sqrt{15}}{20}) \ = \ \Phi(0,97) - \Phi(-0,97) \ = \ 2 \times \Phi(0,97) - 1 = 0,668$

   (b) Agora, $\overline{X} \sim N(\mu,\ \sigma^2_{\overline{X}} = \frac{20^2}{n})$
   Então, $P(|\overline{X} - \mu| < 5) \ = \ P(|Z| < \frac{5\sqrt{n}}{20}) \ = \ 0,95$
   Assim, $\frac{5\sqrt{n}}{20} = 1,96 \Rightarrow n = \left(\frac{1,96\times 20}{5}\right)^2 \ = \ 61,46 \quad \Longrightarrow \quad n = 62$. Assim, precisamos completar a amostra com mais 47 embalagens.

4. (a) O estudante deve falar de confiança e não de probabilidade. Pode citar que o intervalo de confiança é um estimador por intervalo para o parâmetro p (proporção do eleitorado disposto a votar no candidato).

   (b) O aluno deve responder usando o IC não conservador:

$$IC(p, 95\%) = \left(0,55 - 1,96\sqrt{\frac{0,55 \times 0,45}{1500}};\ 0,55 + 1,96\sqrt{\frac{0,55 \times 0,45}{1500}}\right) = (0,5248;\ 0,5751).$$

   (c) Hipóteses:

$$H_0 : p \leq p_0 = 50\% \text{ versus } H_0 : p > p_0 = 50\%.$$

   A estatística de teste sob H_0 :

$$Z = \frac{\hat{p} - p_0}{\sqrt{p_0(1-p_0)/n}} \sim N(0,1).$$

   Região Crítica:

$$RC = (p_0 + z_{1-\alpha}\sigma_0;\ +\infty) = (0,5 + 1,64\sqrt{0,5 \times 0,5/1500};\ +\infty) = (0,5212;\ +\infty).$$

   Tomada de decisão: Como $\hat{p} = 0,55 \in RC$, rejeitamos H_0, ou seja, temos evidências para acreditar que o candidato pode ficar tranquilo quanto às suas chances de vitória já no primeiro turno.

   A resposta pode ser obtida também via p-valor, da forma

$$\tilde{\alpha} = P(\hat{p} > 0,55) = P(Z > \frac{0,55 - 0,5}{\sqrt{0,5 \times 0,5/1500}}) = P(Z > 3,87) = 1 - \Phi(3,87).$$

   Porém o maior quantil disponível na tabela fornecida da Normal(0;1) é 3,59. Logo, concluímos que o p-valor é $\tilde{\alpha} \leq 1 - \Phi(3,59) = 1 - 0,9998 = 0,0002$. Portanto, como $0,0002 < 0,05 = \alpha$, isso implica que H_0 deve ser rejeitada.