UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Prova Final de Probabilidade e Estatística **20-12-2016**

Atenção: Não serão aceitas respostas sem justificativa. Usar os espaços apropriados.

1. Segundo o IBGE, em 2012, o número de pessoas na Região Norte representa 8% da população; na Região Nordeste: 28%; na Região Sudeste: 42%; na Região Sul: 14%; e na Região Centro-Oeste: 8%. Uma pesquisa de 2012 revelou a proporção de leitores em cada região. Na Região Norte era de 47%; 51% na Região Nosdeste; 50% na Região Sudeste; 43% na Região Sul; e 53% na Região Centro-Oeste.
   (a) Qual é a proporção de leitores no Brasil segundo os dados de 2012?
   (b) Se uma pessoa da população brasileira sorteada aleatoriamente é leitora, qual é a probabilidade dela ser da Região Sudeste?

2. Em uma determinada localidade, a renda de um habitante em milhares de u.m. é uma v.a. X com função densidade de probabilidade:

$$f_X(x) \quad = \quad \begin{cases} \frac{x+1}{10} & 0 \leq x < 2 \\ \frac{18-3x}{40} & 2 \leq x < 6 \\ 0 & \text{outro caso} \end{cases}$$

   (a) Determine a função de distribuição acumulada de X
   (b) Escolhida uma pessoa ao acaso, qual é a probabilidade de sua renda exceder 3000 u.m.?
   (c) Determine a renda mediana nessa localidade.

3. Seja X o tempo gasto para se encontrar uma vaga no estacionamento de um Shopping Center. Foram coletadas n = 10 observações independentes desse tempo, em minutos:

   4,47 5,47 5,79 3,50 3,17 6,33 5,88 1,37 3,97 8,17

   (a) Calcule a mediana Q2 da amostra, que, como se sabe, é uma estimativa natural da mediana populacional q_2(X).
   (b) Suponha agora que X $\sim N(\mu, \sigma^2)$. Nessas condições, a média amostral $\bar{X}$ é um estimador não tendencioso de q_2(X)? Por quê?
   (c) Obtenha uma outra estimativa pontual para a mediana populacional q_2(X), usando para isso o estimador$\bar{X}$, proposto no item (b).
   (d) Admitindo agora que o tempo para se encontrar a vaga no estacionamento segue uma distribuição $N(\mu; \sigma^2)$ e que o tamanho n da amostra é bastante grande, qual desses dois estimadores de q_2(X) é o mais preciso: a mediana amostral Q2 ou a média amostral $\bar{X}$? Por quê? (Para responder esta pergunta você pode usar a informação de que, para n grande, a distribuição de probabilidade de Q2 é aproximadamente uma Normal com média q_2(X) e variância $\frac{1}{4n(f(q_2(X)))^2}$, onde f(.) é a função de densidade da $N(\mu; \sigma^2)$.)

4. Para mostrar que seu produto é econômico, um fabricante de automóveis fez um experimento com n=16 carros, do mesmo tipo e no mesmo percurso. Com os dados, que podem ser modelados por uma distribuição Normal, obteve um intervalo de confiança (IC) com 95% para μ, a média de consumo em Km/litro. O IC de 95% foi (8,85; 14,15)
   (a) Para decidir se efetua a compra, um consumidor interessado solicitou um teste de hipótese (TH) com α=0,01, $H_0 : \mu \geq 13$ versus $H_1 : \mu < 13$. Tomando como base o IC acima, exiba a região de rejeição do TH.
   (b) Calcule o p-valor do TH. Qual decisão deve ser tomada?
   (c) O que você acha da afirmação: “Se em um TH H_0 for rejeitada com α=0,01 este mesmo TH terá H_0 rejeitada, também, ao nível α=0,05"? Está ou não correta? Justifique sua resposta no contexto de TH.

Boa prova!

Soluções

1. Sejam os eventos
   N: pessoa é da Região Norte;
   O: pessoa é da Região Nordeste;
   U: a pessoa é da Região Sudeste;
   S: a pessoa é da Região Sul;
   C: a pessoa é da Região Centro-Oeste; e
   L: a pessoa é leitora. Temos

$$P(N) = 0,08;\ P(O) = 0,28;\ P(U) = 0,42;\ P(S) = 0,14;\ P(C) = 0,08$$

   e

$$P(L \mid N) = 0,47;\ P(L \mid O) = 0,51;\ P(L \mid U) = 0,50;\ P(L \mid S) = 0,43;\ P(L \mid C) = 0,53.$$

   (a) Pelo teorema da probabilidade total, temos

$$\begin{aligned} P(L) &= P(L \mid N)P(N) + P(L \mid O)P(O) \\ &\quad + P(L \mid U)P(U) + P(L \mid S)P(S) + P(L \mid C)P(C) \\ &= 0,493. \end{aligned}$$

   (b) Pelo teorema de Bayes, temos

$$P(U \mid L) = \frac{P(L \mid U)P(U)}{P(L)} = \frac{0,5 \times 0,42}{0,493} = 0,426.$$

2. (a) A função de distribuição de X é dada por

$$F_X(x) = \begin{cases} 0 & x < 0 \\ \int_0^x \frac{u+1}{10} du = \frac{x^2}{20} + \frac{x}{10} & 0 \le x < 2 \\ \int_0^2 \frac{u+1}{10} + \int_2^x \frac{18-3u}{40} du = \frac{18}{40}x - \frac{3}{80}x^2 - 0,35 & 2 \le x < 6 \\ 1 & x \ge 6 \end{cases}$$

   (b) $P(X > 3) = 1 - P(X \le 3) = 1 - F_X(3) = 1 - [\frac{18}{40}(3) - \frac{3}{80}(3)^2 - 0,35] = 0,3375$

   (c) Encontrar a renda mediana equivale a resolver $P(X \le m) = 0,5$. Assim, tem-se que $\frac{18}{40}m - \frac{3}{80}m^2 - 0,35 = 0,5$ ou equivalentemente $\frac{3}{80}m^2 - \frac{18}{40}m + 0,85 = 0.0$. O que leva as soluções $m \approx 9,65$ ou $m \approx 2,35$ u.m.. A única solução possível é $m = 2,35$.

3. (a) Colocando os valores em ordem crescente, temos:
   1,37 3,17 3,50 3,97 4,47 5,47 5,79 5,88 6,33 8,17
   Logo Q2 $= \frac{4,47+5,47}{2} = 4,97$ minutos.

   (b) Sabemos que $\mathrm{E}(\bar{X}) = \mu$. Por outro lado, por ser a $\sim N(\mu,\ \sigma^2)$ uma distribuição simétrica em torno de $\mathrm{x} = \mu$, $\mathrm{q_2(X)} = \mu$. Logo, a média amostral $\bar{X}$ é um estimador não tendencioso para $\mathrm{q_2(X)} = \mu$.

   (c) Com base nos dados, calcula-se $\bar{x} = \frac{4,47+5,47+\cdots+8,17}{10} = 4,81$. A nova estimativa é então 4,81 minutos.

   (d) Vamos comparar os dois estimadores em termos de seus EQM's. Como ambos são assintoticamente não tendenciosos, basta comparar suas variâncias.

$$Var(\bar{X}) = \frac{\sigma^2}{n}$$

   Por outro lado, sabemos que $Var(Q2) = \frac{1}{4n(f(q_2(X)))^2}$, onde:
   $f(q_2(X)) = f(\mu) = \frac{1}{\sqrt{2\pi}\sigma} exp(-\frac{(\mu-\mu)^2}{2\sigma^2}) = \frac{1}{\sqrt{2\pi}\sigma}$. Substituindo na expressão acima, obtemos:

$$Var(Q2) = \frac{1}{4n(f(q_2(X)))^2} = \frac{2\pi\sigma^2}{4n} = \frac{\pi\sigma^2}{2n}.$$

   Como $\frac{\pi}{2} > 1$, $Var(\bar{X}) < Var(Q2)$, o que significa que $\bar{X}$ é o mais preciso dos 2 estimadores.

4. (a) $\overline{x} = \frac{l_{inf}+l_{sup}}{2} = \frac{14{,}15+8{,}85}{2} = 11,5$
$d = \frac{14{,}15-8{,}85}{2} = 2,65 = t_{1-\frac{\alpha}{2};n-1} \times \frac{s}{\sqrt{n}} = t_{0{,}975;15}\frac{s}{\sqrt{n}}$. Assim, $s = \frac{2{,}65\times 4}{2{,}131} = 4,97$
$H_0 \geq 13$ versus $H_1 < 13$, com α=0,01, então $t_{tab} = t_{0{,}99;15} = 2,602$
$t_{obs} = \frac{\overline{x}-\mu_0}{\frac{s}{\sqrt{n}}} = \frac{11{,}5-13}{\frac{4{,}97}{4}} = -1,2$
RC $= \{t|t < t_{tab} = -2,602\}$, onde RC, região Crítica ou Rejeição.
Como $-2,602 = t_{tab} < t_{obs} = -1,2$ não há evidência para a rejeição de H_0.

(b) Como t_{obs}= -1,2 e olhando a linha 15=n-1 g.l. na tabela da t-student observamos: $0,8 < P(T < 1,2) < 0,9$. Assim, $0,1 <$ p-valor $< 0,2$ que corresponde a um valor não raro de acontecer. Como o p-valor é uma medida de quanto os dados observados concordam com a hipótese H_0, então sendo p-valor grande, não devemos rejeitar H_0.

(c) Sim! Por que se a rejeição corresponde um nível menor, $\alpha = 0,01$, o valor observado da estatística de teste estará no interior da região de rejeição $R_{0{,}01} \subset R_{0{,}05}$. Assim, rejeita-se, também ao nível de $\alpha = 0,05$. Ou seja, a região de rejeição $R_{0{,}01}$, correspondente a α = 0,01, está necessariamente contida na região de rejeição $R_{0{,}05}$, correspondente a α = 0,05.