AD2 da disciplina Probabilidade e Estatística

GABARITO REVISADO

Professores: Otton Teixeira da Silveira Filho e Regina Célia P. Leal Toledo
02.2008

i) (2,0 pontos com cada item valendo 0,5 ponto)
Verifique se as expressões a seguir são funções densidade de probabilidade (as funções se anulam fora dos intervalos especificados).

Resposta: Integraremos cada função dentro do intervalo onde ela é diferente de zero.

$a.\ f(x)=3x\ ,\ se\ 0\leq\ x\ \leq 1.$

Resposta

A função é não negativa no domínio mas

$$\int_0^1 3x\,dx=3\int_0^1 x\,dx=3\frac{x^2}{2}|_0^1=\frac{3}{2}\neq 1$$

e, portanto, não é densidade de probabilidade.

$b.\ f(x)=\cos(x)\ ,0\leq x\leq \pi/2.$

A função é não negativa no domínio e

$$\int_0^{\pi/2}\cos(x)dx=sen(x)|_0^{\pi/2}=1-0=1$$

portanto é densidade de probabilidade.

$c.\ f(x)=(x-3)/2\ ,\ se\ 3\leq\ x\ \leq 5.$

A função é não negativa no domínio e

$$\int_3^5\frac{(x-3)}{2}dx=\frac{1}{2}\left(\int_3^5 x\,dx-3\int_3^5 dx\right)=\frac{1}{2}\left(\frac{x^2}{2}|_3^5-3x|_3^5\right)=\frac{1}{2}(12{,}5-4{,}5-15+9)=1$$

portanto é densidade de probabilidade.

Esta questão poderia ser feita notando que a distribuição tem a forma de um triângulo de altura 1 e base 2. Assim a área é dada por (1 . 2)/2 = 1.

$d.$
$(2+x)/4, \ se \ -2 \leq x \leq 0;$
$(2-x)/4, \ se \ 0 \leq x \leq 2.$

Esta função é não negativa no domínio. Há duas maneiras de verificar se a integral é igual a 1. Uma maneira é fazer o cálculo direto e outra é obtida se for observado que a função define um triângulo de base 4 e altura ½. Assim, a área do triângulo será

(4 . ½)/ 2 = 1

Também é uma distribuição.

ii) (1,5 pontos)
Foi sorteada uma amostra de 18 postos de saúde da rede pública em uma determinada cidade e anotado o número de casos de dengue em cada uma delas no mês de setembro. Os resultados foram: 10, 8, 5, 4, 3, 7, 1, 11, 3, 6, 6, 7, 3, 4, 8, 9, 5, 5. Deseja-se estimar o número médio de casos e sua variância para apoio à população devido ao início da época chuvosa. Obtenha seguintes estimadores:

$\mu_1 = \frac{(valormínimo + valormáximo)}{2}$ e $\mu_2 = \bar{X}$

Resposta: $\mu_1 = \frac{11+1}{2} = 6$ e

$$\mu_2 = \frac{10+8+5+4+3+7+1+11+3+6+6+7+3+4+8+9+5+5}{18} = 5{,}8333$$

(0,5 ponto no total)

Calcule também os estimadores da variância (use as duas estimativas de média)

$$\sigma^2 = \frac{(valormáximo)^2 - (valormínimo)^2}{2}$$

Resposta: $\sigma^2 = \frac{10^2 - 3^2}{2} = 45{,}5$

Atenção! Houve um erro de digitação e a fórmula para o estimador abaixo está incorreta!

$$\sigma^2 = \frac{1}{n-1} \sum_{i=1}^{n} \left(X_i^2 - n\bar{X}^2 \right)$$

A expressão correta é a que se segue:

$$\sigma^2 = \frac{1}{n-1}\left(\sum_{i=1}^{n} X_i^2 - n\,\bar{X}^2\right)$$

Calculemos o somatório do quadrado dos valores

$$\sum_{i=1}^{n} X_i^2 = 100+64+25+16+9+49+1+121+9+36+36+49+9+16+64+81$$
$$+25+25=735$$

Com a primeira estimativa da média teremos

$$\sigma^2 = \frac{1}{17}\left[735-18(6,0)^2\right] = 5,117$$

com a segunda estimativa para a média

$$\sigma^2 = \frac{1}{17}\left[735-18(5,833)^2\right] = 7,210$$

(0,5 ponto no total)

Diga qual o estimador mais adequado para a média e para a variância. (0,5 ponto)

Resposta: Os segundos estimadores da média e da variância são os mais adequados pois levam em consideração a totalidade dos dados. O fato dos valores das médias diferirem de pouco apenas indicam uma situação particular e que não pode ser generalizada.

iii) (1,5 pontos)

Se investiga numa empresa de informática a possibilidade de mudança da linguagem na qual são elaboradas as aplicações desenvolvidas. O fator avaliado é a velocidade de implementação nas linguagens. Uma equipe de programadores, experientes na linguagem usada atualmente (linguagem 1) e numa proposta para novo uso (linguagem 2), desenvolveram programas a partir do mesmo algorítmo. Abaixo vai uma tabela com os resultados de tempos obtidos para o desenvolvimento de programas baseados no algorítmo.

linguagem 1	linguagem 2
17	18
16	14
21	19
14	11
18	23
24	21
16	10
14	13

21	19
23	24
13	15
18	20
22	18
14	16
20	21

a) Calcule um intervalo de confiança de 95% para a diferença das médias no tempo de programação (1,0 ponto);

Partimos da suposição de que o número de dados é suficiente para usarmos a distribuição Normal. Tomaremos a primeira linguagem como referência e verificaremos se os dados levantados sobre a segunda linguagem estão fora ou não do intervalo de confiança da primeira.

Resposta: Examinando as amostras, temos que a média aritmética de tempo de elaboração do algoritmo na linguagem 1 é aproximadamente $\mu_1 = 18{,}07$ **.**
Para obtermos uma estimativa para variância, usaremos a variância amostral dada por

$$S^2 = \frac{1}{n-1} \sum_1^n X_i^2 - n\bar{X}^2 = \frac{1}{14} \sum_1^{15} X_i^2 - 15\ \bar{X}^2 ¿$$

que para a linguagem 1 nos dá:

$$S^2 = 12{,}92 \Rightarrow S \approx 3{,}59$$

Vamos supor que o tamanho da amostra é grande o suficiente para que possamos usar o Teorema do Limite Central. Daí podemos usar para a distribuição

$$N(18{,}07\,;3{,}59)$$

Sendo o intervalo de confiança dado por

$$IC(\mu,\gamma) = \left[\bar{X} - z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\,;\bar{X} + z_{\gamma/2}\frac{\sigma}{\sqrt{n}}\right]$$

para o valor de confiança igual a $\gamma = 0{,}95$ **retiramos da tabela da distribuição Normal o valor de** $z_{\gamma/2} = 1{,}96$ **. Daí teremos**

$$IC(\mu, 0{,}95) = \left[18{,}07 - 1{,}96\frac{3{,}59}{\sqrt{15}}\,;18{,}07 + 1{,}96\frac{3{,}59}{\sqrt{15}}\right] = [16{,}25\,;19{,}88]$$

b) Por estes dados podemos concluir que a linguagem deve ser mudada (0,5 ponto)?

Resposta: Observe que a média para a linguagem 2 é $\mu_1 = 17{,}466$ que está dentro do intervalo de confiança. Isto indica que dentro do valor de confiança requerido, a mudança de linguagem não tem vantagens.

iv) (1,5 pontos)
Uma amostra de 20 observações de uma variável aleatória Normal forneceu média de 5,5 e variância amostral 4. Deseja-se testar, ao nível de significância de 5%, se a média na população é 5,7 (1,0 ponto). Qual é a conclusão? (0,5 ponto).

Resposta: Supondo a distribuição Normal como válida, calculemos a região crítica por

$$RC = \{x \in \mathbb{R} : x < p_{c1}\, ou\, x > p_{c2}\}$$

e para o nível de significância exigido teremos

$$P(\hat{p} < p_{c1} : H_o) = \frac{0{,}05}{2} \text{ e } P(\hat{p} > p_{c2} : H_o) = \frac{0{,}05}{2}$$

Como supomos valer a distribuição Normal, podemos calcular a probabilidade acima como

$$\hat{P} \sim N(5{,}5 ; 4/20) = N(5{,}5 ; 0{,}2)$$

então

$$P(\hat{p} < p_{c1} : H_0) = P\left(\frac{\hat{p} - 5{,}5}{\sqrt{(0{,}2)}} < \frac{p_{c1} - 5{,}5}{\sqrt{(0{,}2)}}\right) = \frac{0{,}05}{2}$$

pela tabela de distribuição Normal temos que

$$\frac{p_{c1} - 5{,}5}{\sqrt{(0{,}2)}} = -1{,}65 \text{ **ou** } p_{c1} = 4{,}762$$

Da mesma forma

$$\frac{p_{c2} - 5{,}5}{\sqrt{(0{,}2)}} = 1{,}65 \text{ **ou** } p_{c2} = 6{,}237$$

e então temos

$$RC = \{x \in \mathbb{R} : x < 6{,}237\, ou\, x > 4{,}762\}$$

v) (1,5 pontos)
Numa eleição concorriam dois candidatos (A e B). Um deles, o candidato A, obteve 60% dos votos. Numa seção eleitoral, julgada representativa do eleitorado, qual a probabilidade de do candidato A ter conseguido entre 550 e 730 votos de 1000 votos? (1,0 ponto) Qual a probabilidade do candidato B ter conseguido o mesmo número de votos?

(0,5 ponto)

Resposta: Vamos supor válida a distribuição Normal. Sendo assim, a probabilidade será dada por

$$P(a \le X \le b) = P\left(\frac{a-\mu}{\sigma} \le \frac{X-\mu}{\sigma} \le \frac{b-\mu}{\sigma}\right) = P\left(\frac{a-\mu}{\sigma} \le Z \le \frac{b-\mu}{\sigma}\right)$$

Neste caso particular a proporção amostral é de 0,6. Para este valor a estimativa amostral da variância é dada por

$$\sigma^2 = p\frac{(1-p)}{n} = 0{,}6\frac{(1-0{,}6)}{1000} = 0{,}00024$$

Assim, teremos para a probabilidade pedida a expressão

$$P(550 \le X \le 730) = P\left(\frac{550-600}{0{,}0155} \le \frac{X-\mu}{\sigma} \le \frac{730-600}{0{,}0155}\right) = P\left(\frac{550-600}{0{,}0155} \le Z \le \frac{730-600}{0{,}0155}\right)$$

ou

$$P(550 \le X \le 730) = P(-3225{,}8 \le Z \le 8387{,}1) = 1$$

Por complemento a resposta para o candidato B será virtualmente nula a probabilidade.

vi) (2,0 pontos)
Uma indústria metalúrgica afirma que o nível de um determinado sal metálico de suas águas servidas ultrapassou o nível de 28,0 mg/m^3 em 12 dias no último mês (de 30 dias). Um estudo independente, suspeitando dos dados da empresa, fez medidas independentes durante 21 dias e observando que o nível de 28,0 mg/m^3 foi ultrapassado 9 vezes. Qual seria a probabilidade dos dados da metalúrgica estarem errados a um nível de 5%?

Resposta: Façamos alguns cálculos preliminares.

Pelos dados da indústria a proporção amostral de ocorrência de ultrapassagem do nível de referência do sal foi de

$$P = \frac{12}{30} = 0{,}4$$

Devemos testar a hipótese deste valor ter sido ultrapassado a partir do relatório independente. Como estamos trabalhando com proporção amostral, usemos o estimador de variância dado por

$$\sigma^2 = p\frac{(1-p)}{n}$$

Suporemos podemos usar a distribuição Normal, ou seja,

$\hat{P} \sim N(p, p(1-p)/n)$

A Região Critica é dada por

$RC = \{x \in \mathbb{R} : x < p_{c1}\, ou\, x > p_{c2}\}$

Para o valor do nível exigido temos que

$P(\hat{p} < p_{c1} : H_o) = \frac{0{,}05}{2}$ **e** $P(\hat{p} > p_{c2} : H_o) = \frac{0{,}05}{2}$

Pela hipótese p = 0,4 e a distribuição será

$\hat{P} \sim N(0{,}4,\, 0{,}24/21) = N(0{,}4,\, 0{,}0114)$

Logo

$$P(\hat{p} < p_{c1} : H_0) = P\left(\frac{\hat{p} - 0{,}4}{\sqrt{(0{,}0114)}} < \frac{p_{c1} - 0{,}4}{\sqrt{(0{,}0114)}}\right) = \frac{0{,}05}{2}$$

Pela tabela da distribuição Normal temos que

$\frac{p_{c1} - 0{,}4}{\sqrt{(0{,}0114)}} = -1{,}65$ **o que nos dá** $p_{c1} = 0{,}223$

por sua vez teremos $\frac{p_{c2} - 0{,}4}{\sqrt{(0{,}0114)}} = 1{,}65$ **o que nos dá** $p_{c2} = 0{,}455$.

Portanto

$RC = \{x \in \mathbb{R} : x < 0{,}223\, ou\, x > 0{,}455\}$

ou seja, não existe base para discordar do relatório da indústria metalúrgica.