UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Prova Final de Probabilidade e Estatística

Atenção: Só serão aceitas respostas com justificativa. - 27/06/2017

---

1. Considere um reino encantado bem, bem distante. Há uma chance de 50% de que uma rainha carregue o gene da magia. Se ela for portadora do gene, então cada príncipe terá, um independentemente do outro, 50% de chance de ter poderes mágicos. Se a rainha não for portadora desses poderes, os príncipes não herdarão o dom da magia. A rainha teve três filhos.

   (a) Se a rainha for portadora do gene da magia, qual a probabilidade de pelo menos um de seus três filhos possuir o gene da magia?

   (b) Supondo que todos os filhos da rainha não têm poderes mágicos, qual a probabilidade de que ela seja portadora do gene da magia?

2. Admita que o intervalo de tempo T entre duas chegadas sucessivas de mensagens de Whatsapp ao celular de determinada pessoa segue uma distribuição Exponencial, e que a probabilidade de que esse intervalo de tempo seja maior que 20 minutos é igual a 0,135.

   (a) Em média, quanto tempo após a chegada de uma mensagem deve chegar a próxima mensagem?

   (b) Qual a probabilidade de que T seja no máximo 40 minutos, dado que T é no mínimo 25 minutos?

   (c) Qual o desvio padrão do número de mensagens que a pessoa recebe em meia hora?

   (d) Qual a probabilidade de que, ao longo de 1 hora, ela receba pelo menos 4 mensagens?

3. Sejam T_1 e T_2 dois estimadores não tendenciosos do parâmetro θ tais que

$$\mathsf{Var}(T_1) = 1, \qquad \mathsf{Var}(T_2) = 4 \qquad \text{e} \qquad \mathsf{COV}(T_1, T_2) = -1.$$

   Considere o estimador de θ definido por

$$T = \alpha T_1 + (1-\alpha)T_2, \quad \alpha \in \mathrm{R}.$$

   (a) Para que valores de α, T é um estimador não tendencioso de θ?

   (b) Expresse a variância de T em função de α.

   (c) Para que valor de α, a variância de T é mínima?

4. Uma companhia produz peças para um aparelho. A medição do tamanho, em mm, para 30 espécimes de peças desse tipo resultou nas observações abaixo:

76	79	88	88	90	90	90	92	93	93
94	94	95	95	96	98	98	98	98	99
100	100	101	103	103	107	110	112	117	119

   (a) Calcule o p-valor do teste de hipótese para a média populacional dessa variável ser no máximo 95 contra a alternativa de ser maior do que 95. Baseado no p-valor, que conclusão se tem ao nível α=5%;

   (b) Teste a hipótese de que a proporção de peças com comprimento estritamente superior a 100 mm seja maior ou igual ($\geq$) que 30%, ao nível α de 8% ;

   (c) Utilize os dados como uma amostra piloto para dimensionar uma nova amostra com base na qual se pode obter uma estimativa da média populacional do comprimento das peças, de modo a que o coeficiente de variação do estimador, $\overline{X}$ seja da ordem de 0,01.

   Para esta amostra temos: $\sum_{i=1}^{30} x_i = 2916$ e $\sum_{i=1}^{30} x_i^2 = 286084$.

Boa Prova!

Solução

1. Considere os eventos: R: "rainha possui o gene da magia" e A_i: "príncipe i possui o gene da magia", com $i = 1, 2, 3$.

   (a) A probabilidade desejada, usando independência condicional, é

$$\begin{aligned} P(\cup_{i=1}^{3} A_i|R) &= \sum_{i=1}^{3} P(A_i|R) - \sum_{i>j} P(A_i \cap A_j|R) + P(\cap_{i=1}^{3} A_i|R) \\ &= (1/2) + (1/2) + (1/2) - (1/2)^2 - (1/2)^2 - (1/2)^2 + (1/2)^3 \\ &= 7/8. \end{aligned}$$

   (b)

$$P(R|\cap_{i=1}^{3} A_i^c) = \frac{P(R \cap (\cap_{i=1}^{3} A_i^c))}{P(\cap_{i=1}^{3} A_i^c)} = \frac{P(R)[\prod_{i=1}^{3} P(A_i^c|R)]}{P(R)[\prod_{i=1}^{3} P(A_i^c|R)] + P(R^c)[\prod_{i=1}^{3} P(A_i^c|R^c)]}.$$

   Calculando obtém-se

$$P(R|\cap_{i=1}^{3} A_i^c) = \frac{(1/2)(1/2)(1/2)(1/2)}{(1/2)(1/2)(1/2)(1/2) + (1/2)(1)(1)(1)} = \frac{1}{9}$$

2. (a) $T \sim Exp(\lambda)$. Sabemos que $0{,}135 = P(T > 20) = e^{-20\lambda}$.
   Logo, $ln(0,135) = -2 = -20\lambda, \quad \lambda = 0,1$. Assim, $E(T) = 1/\lambda = 10$ minutos.

   (b) Pela Propriedade da Perda de memória da Exponencial: $P(T > 40|T > 25) = P(T > 40 - 25) = P(T > 15) = e^{-0,1x15} = e^{-1,5} = 0,223$.
   Portanto, $P(T < 40|T > 25) = 1 - 0,223 = 0,777$.

   (c) Seja X o número de chegadas de mensagens em meia hora, ou seja, 30 minutos. Então, pela relação existente entre os modelos Exponencial e de Poisson,$X \sim Poisson(\mu X)$, onde $\mu X = 30\lambda = 30x0,1 = 3$. Logo $DP(X) = \sqrt{3} = 1,73$.

   (d) Seja Y o número de chegadas de mensagens em 1 hora = 60 minutos. Então, $Y \sim Poisson(\mu_Y)$, onde $\mu_Y = 60\lambda = 60x0,1 = 6$.
   Daí, $P(Y \geq 4) = 1\breve{} P(Y \leq 3) = 1\breve{}(P(X = 0) + P(X = 1) + P(X = 2) + P(X = 3) =$
   $= 1\breve{} e^{-6}(1 + 6 + 6^2/2 + 6^3/6) = 0,849$.

3. (a) T é estimador não tendencioso de θ para todo $\alpha \in R$, pois $\mathsf{E}[T] = \mathsf{E}[\alpha T_1 + (1-\alpha)T_2] = \alpha \underbrace{\mathsf{E}[T_1]}_{=\theta} + (1-\alpha)\underbrace{\mathsf{E}[T_2]}_{=\theta} =$
   $\alpha\theta + (1-\alpha)\theta = \theta$.

   (b) $\mathsf{Var}(T) = \alpha^2 \underbrace{\mathsf{Var}(T_1)}_{=1} + (1-\alpha)^2 \underbrace{\mathsf{Var}(T_2)}_{=4} + 2\alpha(1-\alpha)\underbrace{\mathsf{COV}(T_1, T_2)}_{=-1} =$

$$= \alpha^2 + 4(1 - 2\alpha + \alpha^2) - 2\alpha + 2\alpha^2 = 7\alpha^2 - 10\alpha + 4$$

   (c) Observe que a expressão da variância em função de α é uma parábola com a concavidade voltada para cima de tal modo que o valor mínimo da variância é obtido no vértice da parábola, ou seja, para $\alpha = \dfrac{-(-10)}{2 \times 7} = \dfrac{5}{7}$.

4. Com base na amostra: $\overline{x} = \frac{\sum_{i=1}^{30} x_i}{n_1} = \frac{2916}{30} = 97,2mm$;
   $s^2 = \frac{\sum x_i^2 - n\overline{x}^2}{n-1} = \frac{286084 - 30\times(97,2)^2}{29} = 91,34mm^2$; e $s = \sqrt{91,34} = 9,6mm$

   (a) A estatística de teste a ser usada aqui é a média amostral, $\overline{x}$. Como o tamanho amostral é razoavelmente grande, sob H0, $\overline{X} \sim N(95, s/\sqrt{n})$. Devido à forma como foram definidas as hipóteses H0: $\mu \leq 95$ versus H1: $\mu > 95$, é natural que valores baixos de $\overline{x}$ nos levem a aceitar H0, enquanto valores altos de $\overline{x}$ nos levem a rejeitar H0. Então,
   p-valor=P($\overline{x} > 97,2$)=1-$\Phi(\frac{97,2-95}{9,6/\sqrt{30}}) = 1 - \Phi(1,25) = 0,1056$. Como o p-valor é uma probabilidade que representa uma medida de quanto os dados concordam com a hipótese H0, 11% é um valor alto comparado com 5%. Assim, não temos evidência para a rejeição de H0.

   (b) A estatística de teste a ser usada aqui é a proporção amostral, $\hat{p}$. Como o tamanho amostral é razoavelmente grande, sob H0, $\hat{p} \sim N(p, p(1-p)/n)$. Devido à forma como foram definidas as hipóteses H0: $\hat{p} \geq 0,3$ versus H1: $\hat{p} < 0,3$, é natural que valores altos de $\hat{p}$ nos levem a aceitar H0, enquanto valores baixos de $\hat{p}$ nos levem a rejeitar H0. Na amostra temos 8 valores estritamente maiores que 100, $\hat{p} = 8/30 = 0,27$ Então,
   p-valor=P($\hat{p} < 0,27$)=$\Phi(\frac{0,27-0,3}{\sqrt{0,3\times0,7/30}}) = \Phi(-0,40) = 1 - \Phi(0,40) = 0,35$. Como o p-valor é aproximadamente 35% sendo um valor alto comparado com 8%, não temos evidência para a rejeição de H0.

(c) Com a nova amostra, de tamanho n, o coeficiente de variação do estimador de μ será

$$cv(\overline{X}) = (DP(\overline{X}))/(E(\overline{X}) = \frac{\sigma/\sqrt{n}}{\mu}.$$

Como μ e σ são ambos desconhecidos, vamos substituí-los por suas estimativas baseadas na piloto, de tamanho $\mathrm{n}_1 = 30$, ou seja, $\overline{x}_1 = 97,2mm$ e $s_1 = 9,6mm$, respectivamente. Então, para que $cv(\overline{X})$ seja aproximadamente igual a 0,01, devemos ter: $\frac{\frac{9,6}{\sqrt{n}}}{97,2} = 0,01$, o que implica que $n \approx \left(\frac{9,6}{97,2 \times 0,01}\right)^2 \approx 98$ espécimes.