UFRJ - CCMN - IM - Departamento de Métodos Estatísticos
Probabilidade e Estatística - Estatística

Prova Final 14-07-2015

Atenção: Não serão aceitas respostas sem justificativa: as expressões que levaram a alguma resposta numérica devem ser indicadas nos espaços apropriados.

1. Em um bairro existem 20mil residências que podem ser atendidas por 3 empresas de TV a cabo. A empresa A tem 2100 assinantes, a empresa B tem 1850 e a empresa C tem 2600 assinantes, sendo que algumas residências em condomínios subscrevem aos serviços de mais de uma empresa. Assim, temos 420 residências que são assinantes de A e B, 120 de A e C, 180 de B e C e 30 que são assinantes das três empresas. Se uma residência desse bairro é sorteada ao acaso, qual é a probabilidade de

   (a) Ser assinante apenas da empresa A?

   (b) Assinar pelo menos uma delas?

   (c) Não ter TV a cabo?

2. O tempo de travessia T de uma balsa é uma variável aleatória que assume os valores 10, 15 e 30 minutos com probabilidades iguais. Carlos pega esta balsa diariamente para ir ao trabalho. Entretando, ele não consegue pegar a conexão com o ônibus da firma se o tempo gasto com a travessia for $T = 30$. Não pegando o ônibus ele consegue um taxi com probabilidade 1/2, e, neste caso, chega a tempo no serviço. Não conseguindo um taxi ele vai a pé e chega atrasado. Sua firma tem a política de tolerar no máximo 20% de atrasos por ano.

   (a) Calcule a probabilidade de que, num dia normal de trabalho, Carlos chegue atrasado no serviço.

   (b) Calcule a probabilidade de que, num ano de trabalho, Carlos não consiga cumprir com a política de tolerância da empresa com respeito a atrasos, admitindo que haja 300 dias de trabalho por ano. Você acha que Carlos conseguirá se manter no emprego?

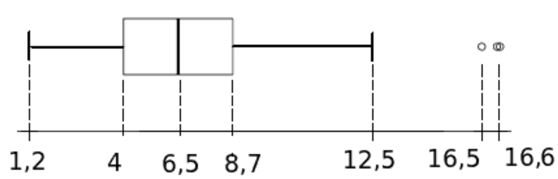
3. Deseja-se estimar a probabilidade de acerto (de encaçapar a bola com uma jogada) do campeão nacional de sinuca. Um intervalo de 95% de confiança foi obtido para estimar esta probabilidade: $(0,66;\ 0,94)$.

   (a) Forneça uma estimativa pontual para a probabilidade de acerto.

   (b) Se o intervalo é conservador (ou conservativo), qual é o tamanho da amostra?

   (c) Calcule, para este mesmo tamanho de amostra, o intervalo não conservador (ou não conservativo). Há diferença? Por quê?

4. Considere uma amostra de 100 tempos de falha de um componente elétrico (meses) que tenha gerado o seguinte boxplot:



   (a) Qual conjunto contém a maior quantidade de elementos da amostra referente aos tempos de falha: $(-\infty; 6,5]$ ou $[6,5; \infty)$? Justifique.
   (b) Pressupondo que os quartis não serão afetados pela inclusão da nova observação, um tempo de falha de 12,7 meses seria considerado um outlier de acordo com o boxplot? Justifique.
   (c) Encarando agora esse boxplot como se ele fosse uma distribuição de probabilidade, ou seja, encarando os quartis amostrais como se fossem quartis populacionais, suponha que 50 componentes sejam sorteados (com reposição e de modo independente) e que se anote o tempo que levou até o componente falhar. Usando o TCL, obtenha uma aproximação para a probabilidade de observarmos mais de 35 componentes cujos tempos de falha tenham sido menores que 8,7 meses.

Boa prova!

Gabarito

1. **(a)** Sejam $n(\Omega) = 20000$, $n(A) = 2100$, $n(B) = 1850$, $n(C) = 2600$, $n(A \cap B) = 420$, $n(A \cap C) = 120$, $n(B \cap C) = 180$ e $n(A \cap B \cap C) = 30$.

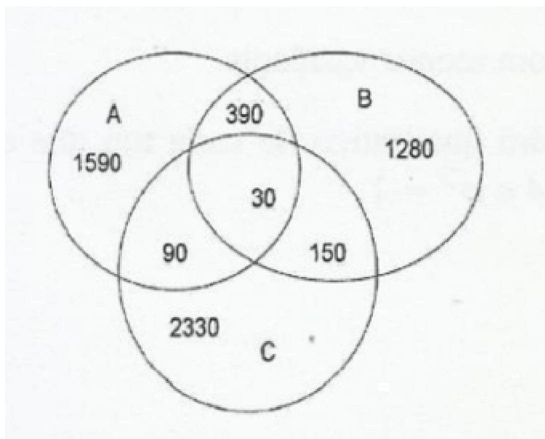
   Então,

$$P(A) = \frac{2100}{20000},\ P(B) = \frac{1850}{20000},\ P(C) = \frac{2600}{20000}$$
$$P(A \cap B) = \frac{420}{20000},\ P(A \cap C) = \frac{120}{20000},\ P(B \cap C) = \frac{180}{20000}$$
$$P(A \cap B \cap C) = \frac{30}{20000}.$$

   A probabilidade pedida é $P(A \cap \overline{(B \cup C)})$.

   O conjunto A pode ser representado como união disjunta dos eventos (vide figura):

$$A = [A \cap \overline{(B \cup C)}] \cup [A \cap C] \cup [(A \cap B) - (A \cap B \cap C)]$$



   Assim, a probabilidade $P(A)$ pode ser obtida como a soma das probabilidade que compõem a representação anterior, isto é:

$$\begin{aligned} P(A) &= P[A \cap \overline{(B \cup C)}] + P(A \cap C) + P[(A \cap B) - (A \cap B \cap C)] \\ &= P[A \cap \overline{(B \cup C)}] + P(A \cap C) + P(A \cap B) - P(A \cap B \cap C). \end{aligned}$$

   Logo, a probabilidade requerida é

$$\begin{aligned} P[A \cap \overline{(B \cup C)}] &= P(A) - P(A \cap C) - P(A \cap B) + P(A \cap B \cap C) \\ &= \frac{2100}{20000} - \frac{120}{20000} - \frac{420}{20000} + \frac{30}{20000} \\ &= 0,0795. \end{aligned}$$

   (b) A probabilidade pedida é $P(A \cup B \cup C)$

$$\begin{aligned} P(A \cup B \cup C) &= P(A) + P(B) + P(C) - P(A \cap B) - P(A \cap C) - P(B \cap C) + P(A \cap B \cap C) \\ &= \frac{2100}{20000} + \frac{1850}{20000} + \frac{2600}{20000} - \frac{120}{20000} - \frac{420}{20000} - \frac{180}{20000} + \frac{30}{20000} \\ &= 0,293. \end{aligned}$$

(c) A probabilidade pedida é $P(A \cup B \cup C) = 1 - P(A \cup B \cup C) = 1 - 0,293 = 0,707$.

2. **(a)** Através da árvore temos 4 resultados possíveis para o experimento, sendo que em 3 deles o Carlos chega no horário e apenas quando perde o taxi ele chega atrasado, com probabilidade 1/6.

   (b)

   Seja $Y \sim Binomial(n, p)$, $n = 300$ e $p = 1/6$. Temos $E[Y] = 50$ e $var(Y) = 250/6$. Calculando

$$P(Y > 60) = P\left(Z > \frac{60 - 50 + 0,5}{\sqrt{250/6}}\right) = P(Z > 1,63) = 0,052.$$

   Assim, a probabilidade de Carlos ter pelo menos 60 atrasos (20%) durante 300 dias é pequena, 5.2%. Ele deve conseguir se manter no emprego.

3. **(a)** Como o intervalo é simétrico em torno da estimativa pontual, esta estimativa é então o ponto médio do intervalo, isto é, $\hat{p} = (0,66 + 0,94)/2 = 0,80$.

   (b) Como $\hat{p} = 0,80$, então

$$\begin{aligned} 0,14 &= z_{0,975}\frac{1}{2\sqrt{n}} = 1,96\frac{1}{2\sqrt{n}} \\ \Rightarrow n &= \left(\frac{1,96}{2 \times 0,14}\right)^2 = 49. \end{aligned}$$

   (c) O intervalo não conservativo é da forma

$$\begin{aligned} IC(0,95) &= \left(\hat{p} - z_{0,975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}};\ \hat{p} + z_{0,975}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right) \\ &= \left(0,8 - 1,96\sqrt{\frac{0,8 \times 0,2}{49}};\ 0,8 + 1,96\sqrt{\frac{0,8 \times 0,2}{49}}\right) \\ &= (0,688;\ 0,912). \end{aligned}$$

   Sim! O intervalo de confiança conservativo apresenta maior amplitude, pois considera o cenário mais conservador (maior amplitude possível), isto é, quando $p(1-p) = 1/4$.

4. **(a)** Como 6,5 é a mediana, esse valor deixa 50% dos dados a sua esquerda e 50% à direita, portanto ambos os intervalos contêm a mesma quantidade de dados.

   (b) A cerca superior vale $Q_3 + 1,5DIQ = 8,7 + 1,5 \times 4,7 = 15,75$. Como a observação 12,7 é menor que 15,75 e é claramente maior que a cerca inferior, não é considerada um outlier.

   (c) Defina as variáveis aleatórias X_i que assumem o valor 1 se o tempo de falha da i-ésima componente sorteada é menor que 8,7 meses e 0 caso contrário. X_i são iid's com distribuição Bernoulli(0,75) pois 8,7 é o terceiro quartil populacional.

   Seja $Y = \sum_{i=1}^{50} X_i$. Logo, $Y \sim$ Binomial$(50; 0,75)$.
   Pelo TCL, temos que $Y \approx N(np, np(1-p))$, onde $p = 0,75$ e $n = 50$, ou seja,

$$Y \approx N(37,5;\ \ 9,375).$$

   Assim, temos (aplicando a correção de continuidade)

$$\begin{aligned} P(Y > 35) &\approx P(Y \geq 35,5) = P\left(\frac{Y - 37,5}{\sqrt{9,375}} > \frac{35,5 - 37,5}{\sqrt{9,375}}\right) \approx \\ &\approx 1 - \phi\left(\frac{-2}{\sqrt{9,375}}\right) = 1 - \phi(-0,6532) = \\ &= \phi(0,6532) = 0,7432. \end{aligned}$$