UFRJ - CCMN - IM - Departamento de Métodos Estatísticos

Segunda Avaliação de Probabilidade e Estatística **19-06-2018**
Atenção: Não serão aceitas respostas sem justificativa.
Resolver as questões nos espaços apropriados.

1. Os lucros diários de determinado comerciante com a venda de um de seus produtos podem ser considerados variáveis aleatórias independentes. Suponha que o valor médio e o desvio padrão dos lucros diários sejam de respectivamente $\mu = R\$60,00$ e $\sigma = 6,40$. Usando a aproximação fornecida pelo Teorema Central do Limite, calcule:
   (a) A probabilidade do lucro **trimestral** (90 dias) ser superior a $R\$5500,00$.
   (b) O lucro **semestral** (180 dias) máximo a ser garantido com 98% de chance.
   (c) Qual o número mínimo de dias necessários para garantir um lucro de pelo menos $R\$5000,00$ com 95% de chance.

2. Dois especialistas X e Y estimaram o tempo em horas a ser gasto no desenvolvimento de 11 projetos:

Projeto	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**	**11**
X	8,5	6,7	4,4	4,8	6,1	5,4	5,7	5,9	6,7	5,6	6,1
Y	7,6	7,4	6,2	4,6	7,0	6,5	6,5	7,8	6,6	7,8	7,9

   (a) Fazer um gráfico ramo-folha (ramo = parte inteira e folha = parte decimal) para X.
   (b) Calcule os três quartis para as duas variáveis X e Y. Construa um box-plot dos dados para cada especialista. Os dois box-plots devem aparecer juntos na mesma figura, utilize a mesma escala de forma que os gráficos fiquem comparáveis. Comente os gráficos
   (c) Construa um diagrama de dispersão.
   (d) Se, num conjunto de n pontos no plano, existir um deles que está muito destoante do comportamento geral, qual será o efeito sobre o coeficiente de correlação de se excluir este ponto?

3. Em uma eleição de 2º turno um instituto de opinião pretende estimar, numa pesquisa de boca de urna, a proporção p de eleitores que votaram no candidato do partido liberal. Responda às seguintes questões.
   (a) Determine o número de eleitores que devem ser consultados de modo que a proporção p possa ser estimada com margem de erro de 0,01 e coeficiente de confiança de 95%?
   (b) Se as pesquisas de opinião do dia anterior indicam claramente que o candidato deverá ter entre 25% e 40% dos votos, você conseguiria reduzir o tamanho amostral calculado em (a) com essa informação? Justifique.
   (c) Suponha que o instituto tenha consultado 1200 eleitores, dos quais 564 afirmam terem votado no candidato do partido liberal. Obtenha um intervalo de confiança não conservativo com coeficiente de confiança de 95% para a proporção p.

4. O fabricante de determinado modelo de automóvel afirma que seu desempenho médio é de 12 km/ℓ de gasolina. Testes foram feitos em 36 desses veículos, escolhidos ao acaso, e apurou-se um desempenho médio de 10,8 km/ℓ. Admita que o desempenho siga o modelo Normal com variância igual a 16 (km/ℓ)^2. Através do p-valor execute os dois testes, itens (a) e (b) a seguir, com relação a afirmação do fabricante:
   (a) $H_0 : \mu \geq 12km/\ell$ contra a alternativa $H_1 : \mu < 12km/\ell$. Que decisão deve ser tomada ao nível de 5%?
   (b) $H_0 : \mu = 12km/\ell$ contra a alternativa $H_1 : \mu \neq 12km/\ell$. Que decisão deve ser tomada ao nível de 5%?
   (c) No item (a) obtenha a região de rejeição com $\alpha = 5\%$. Se, de fato, $\mu = 10km/\ell$, isso corresponderia a que valor da probabilidade do erro tipo II?

Boa Prova!

Solução

1. (a) Seja $X = \sum_{j=1}^{90} X_j$ o lucro trimestral, onde X_j é o lucro do j-ésimo dia. Seja $Z \sim \mathcal{N}(0,1)$, pelo TCL $P(X \geq 5500) = P\Big(\frac{X-90\times 60}{\sqrt{90}\times 6{,}4} \geq \frac{5500-90\times 60}{\sqrt{90}\times 6{,}4}\Big) \approx P(Z \geq 1,65) = 0,0495\,.$

   (b) Seja $X = \sum_{j=1}^{180} X_j$ o lucro semestral. Queremos encontrar o quantil a 98% de X. Ou seja, queremos x tal que

$$0,98 = P(X \leq x) \approx P\Big(Z \leq \frac{x - 180 \times 60}{\sqrt{180} \times 6,4}\Big) = P\Big(Z \leq \frac{x - 10800}{85,865}\Big)\,.$$

   Como $P(Z \leq 2,05) = 0,98$ temos: $\quad 2,05 = \frac{x-10800}{85{,}865} \quad \Longrightarrow \quad x = 10800 + 2,05 \times 85,865 = 10976,02\,.$

   (c) Seja $X_n = \sum_{j=1}^{n} X_j$ o lucro obtido em n dias. Queremos encontrar o menor n tal que

$$0,95 \geq P(X_n \geq 5000) \approx P\Big(Z \leq \frac{5000 - 60n}{6,4\sqrt{n}}\Big)\,.$$

   Como $P(Z \geq -1,64) = 0,95$, temos

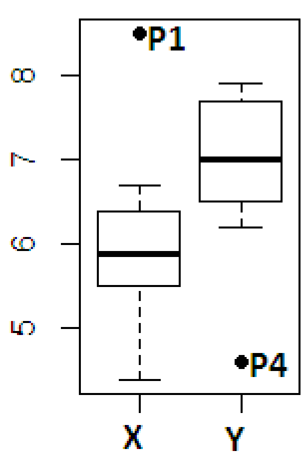
$$5000 - 60n = -1,64 \times 6,4 \times \sqrt{n} \quad \text{ou} \quad 60n - 10,496\sqrt{n} - 5000 = 0$$

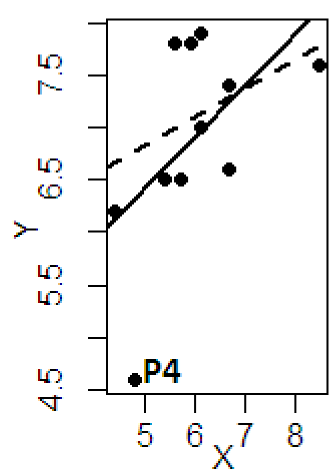
   Então n será quadrado de uma raiz positiva da equação acima. A única raiz positiva é aproximadamente $9,217$ cujo quadrado é $84,95$, logo devemos escolher $n = 85$. Ou seja, são necessários pelo menos 85 dias para garantir um lucro de pelo menos R$5000,00 com 95% de chance.

2. (a) O gráfico ramo-folha para X:

```
4 | 48
5 | 4679
6 | 1177
7 |
8 | 5
```

   (b) Pos(Q1)= (3+n)/4=14/4=3,5; Pos (Q2)= (11+1)/2 = 6; Pos(Q3)=(3n+1)/4=8,5.
   Os valores de X em ordem crescente são: 4,4 4,8 5,4 5,6 5,7 5,9 6,1 6,1 6,7 6,7 8,5. Assim,
   Q1=$(X_3 + X_4)/2 = (5,4 + 5,6)/2 = 5,5$; Q2=$X_6 = 5,9$; Q3=$(X_8 + X_9)/2 = (6,1 + 6,7)/2 = 6,4$
   DIQ=Q3-Q1=0,9; CI=5,5-1,5×0,9=4,15; CS=6,4+1,5×0,9=7,75
   Os valores de Y em ordem crescente são: 4,6 6,2 6,5 6,5 6,6 7,0 7,4 7,6 7,8 7,8 7,9. Assim,
   Q1=$(Y_3 + Y_4)/2 = (6,5 + 6,5)/2 = 6,5$; Q2=$Y_6 = 7,0$; Q3=$(Y_8 + Y_9)/2 = (7,6 + 7,8)/2 = 7,7$
   DIQ=Q3-Q1=1,2; CI=6,5-1,5×01,2=4,7; CS=7,7+1,5×1,2=9,5





   Os dois box-plots têm valores discrepantes, por excesso em X e por falta em Y. Tirando estes pontos a variável Y se mostra simétrica, mas a X ainda apresenta uma leve assimetria.

   (c) ver o gráfico

   (d) O resultado esperado dessa exclusão deve ser um aumento do coeficiente de correlação em módulo.

3. (a) Como $d = 0,01$, $\alpha = 0,05$ e não há informação a priori sobre a proporção p, o tamanho amostral n deve satisfazer

$$n = (1/4)(1,96/0,01)^2 = 9604.$$

(b) Sabemos a priori que $0,25 \leq p \leq 0,4$. Como a função $p \mapsto p(1-p)$ é crescente no intervalo $[0,25; 0,4]$ podemos reduzir o tamanho amostral n para

$$(0,4)(0,6)(1,96/0,01)^2 \approx 9220.$$

(c) Como $\hat{p} = 0,47$, 0 intervalo de confiança não conservativo com com coeficiente de confiança de 95% para p é dado por:

$$\left(0,47 - 1,96\sqrt{\frac{(0,47)(0,53)}{1200}}; 0,47 + 1,96\sqrt{\frac{(0,47)(0,53)}{1200}}\right) = (0,4418; 0,4982).$$

4. (a) X = desempenho do automóvel, X $\sim N(\mu, \sigma^2)$

O teste é unilateral. Vamos usar o caso mais desfavorável a H_0, o mais dificil de decidir entre H_0 e H_1. Assim, H_0:μ=12 versus H_1:μ <12. Supondo H_0 verdadeira a estatística de teste é:

$\overline{X} \sim N(\mu_0, \sigma^2/n)$ ou no eixo da Normal padronizada $Z = \frac{\overline{X}-\mu_0}{\sigma/\sqrt{n}}$. Como, na amostra, $\overline{X}$=10,8

p-valor=$P(Z < \frac{10,8-12}{4/6} = P(Z < -1.80) = 1 - \Phi(1.80) = 0,0359$.

Como o p-valor, uma medida de quanto os dados concordam com H_0, é baixo, menor que 5%, devemos rejeitar a hipótese H_0.

(b) O teste é bilateral. p-valor=$2 \times P(Z < \frac{10,8-12}{4/6} = 2 \times 0,0359 = 0,0718$.

Para este teste os dados não apresentam uma evidência para rejeção da hipótese H_0 ao nível de 5%.

(c) RC=$\{\overline{x} < -1,64 \times 4/6 + 12\} = \{\overline{x} < 10,91\}$. Assim, se $\mu = 10$,

$\beta = P(\overline{X} > 10,91) = 1 - \Phi(\frac{10,91-10}{4/6}) = 1 - \Phi(1,37) = 1 - 0,9147 = 0,0853$.