Prova #2 23-02-2016

Atenção: Não serão aceitas respostas sem justificativa.

1. O funcionamento de uma bateria pressupõe uma alternância permanente entre tempos de recarga e de utilização (descarga). O tempo de recarga pode ser modelado como uma variável aleatória X, que tem distribuição exponencial de média 12 horas. O tempo de utilização, até a bateria descarregar, pode ser modelado como uma variável aleatória Y, que tem distribuição Uniforme$(20, 32)$, em horas. Suponha ainda que todos os tempos envolvidos nesse processo sejam independentes entre si.

   (a) Defina por T a variável aleatória que indica o tempo decorrido entre a bateria começar uma recarga e iniciar a recarga seguinte. Forneça a esperança e a variância da variável T.

   (b) Quando temos uma sequência infinita de variáveis aleatórias $X_1, X_2, \dots X_n, \dots$, o Teorema Central do Limite nos permite aproximar a distribuição da variável $Y = \sum_{i=1}^{n} X_i$ por uma distribuição normal. Que requisitos o número n de parcelas do somatório e as variáveis X_i precisam satisfazer para que esse Teorema possa ser utilizado?

   (c) Suponha que a bateria tenha uma vida útil de 100 recargas, ou seja, após começar com a recarga inicial, ela deixa de funcionar quando for iniciar a 101ª recarga. Com auxílio do Teorema mencionado no item anterior, calcule aproximadamente a probabilidade de que a bateria tenha uma vida útil de menos de 160 dias.

2. Os dados da tabela abaixo representam o número de queixas a respeito das condições de trabalho por parte dos empregados em dez filiais de uma determinada empresa durante os anos de 2009 e 2010.

Ano	Número de queixas									
2009	18	10	17	10	17	9	14	12	9	18
2010	9	10	8	11	0	9	5	11	8	10

   (a) Calcule o número médio de queixas e o desvio padrão do número de queixas para cada ano.

   (b) Calcule os quartis do número de queixas e construa um *box-plot* dos dados para cada ano. Os dois *box-plots* devem aparecer juntos na mesma figura, ou seja, utilize a mesma escala de forma que os gráficos fiquem comparáveis.

   (c) Existe diferença aparente de um ano para outro em termos de valor central? E em termos de dispersão? Explique. Existe(m) valor(es) discrepante(s)? Explique.

3. Deseja-se estimar a proporção p dos usuários de determinado serviço público, que se consideram satisfeitos com a sua qualidade. Com base em levantamentos anteriores, sabe-se que $0,1 \leq p \leq 0,7$.

   (a) Em uma amostra piloto de tamanho 40, extraída ao acaso dessa população, 17 usuários revelaram estar satisfeitos. Obtenha um intervalo não conservativo ao nível de confiança de $0,99$ para p com base nessa amostra piloto.

   (b) Deseja-se agora dimensionar uma segunda amostra que permita estimar p de forma que o erro absoluto, $| \hat{p} - p |$, seja menor que $0,05$ com $0,98$ de probabilidade. Quantos usuários adicionais (aos da piloto) precisariam ser ouvidos para se garantir esse nível de precisão?

   (c) Admita agora que o objetivo fosse estimar p com um erro relativo, $\frac{|\hat{p}-p|}{p}$, menor que 0,2 com 0,95 de probabilidade. Quantos usuários adicionais (aos da piloto) precisariam ser ouvidos para se garantir esse novo nível de precisão?

4. A vida média de uma amostra de 100 lâmpadas de certa marca é de 1615 horas. Por similaridade com outros processos de fabricação, suporemos que o desvio padrão populacional é igual a 120 horas. Utilizando um nível de significância de 5%, desejamos testar se a duração média de todas as lâmpadas dessa marca é igual ou é diferente de 1600 horas. Obs.: Admita que o tamanho $n = 100$ da amostra é suficientemente grande para que seja aplicável o Teorema central do limite.

   (a) Qual é a conclusão do teste?

   (b) Determine a probabilidade do erro tipo II, se a média populacional fosse 1620 horas.

Boa prova!

Gabarito

1. **(a)** $X \sim Exp(1/12)$, $Y \sim U(20,32)$. Temos que $T = X + Y$. Assim,

$$E(T) = E(X+Y) = E(X) + E(Y) = \overbrace{E(Exp(1/12))}^{1/\lambda} + \overbrace{E(U(20,32))}^{(a+b)/2} = 12 + \frac{20+32}{2} = 12 + 26 = 38.$$

$$Var(T) = Var(X+Y) = \overbrace{Var(X) + Var(Y)}^{X,Y ind.} = \overbrace{Var(Exp(1/12))}^{1/\lambda^2} + \overbrace{Var(U(20,32))}^{(b-a)^2/12}$$
$$= \frac{1}{(1/12)^2} + \frac{(32-20)^2}{12} = 144 + 12 = 156.$$

(b) Premissas: X_i's devem ser independentes e identicamente distribuídas (i.i.d.) com média e variância ambas finitas e n deve ser suficientemente grande para que a distribuição de Y já esteja bem próxima de uma normal, pois essa convergência pressupõe que $n \to \infty$.

(c) Seja T_i o tempo entre a i-ésima recarga e a $(i+1)$-ésima recarga serem iniciadas. Pelo item (a), temos que $E(T_i) = 38$ e $Var(T_i) = 156$. Queremos calcular

$$P\left(\sum_{t=1}^{100} T_i < 160 \times 24\right) = P\left(\sum_{t=1}^{100} T_i < 3840\right).$$

Note que $T_1, \ldots, T_{100}$ são i.i.d's. Então, pelo TCL:

$$\sum_{t=1}^{100} T_i \sim N(100E(T_i), 100Var(T_i)) \equiv N(3800, 15600).$$

Portanto,

$$P\left(\sum_{t=1}^{100} T_i < 3840\right) = P\left(Z < \frac{3840-3800}{\sqrt{15600}}\right) = P(Z < 0,32) = 0,6255.$$

2. **(a)** Sejam $(x_1, \ldots, x_{10})$ e $(y_1, \ldots, y_{10})$ as reclamações em 2009 e 2010, respectivamente. Assim, as médias são

$$\bar{x} = \sum_{i=1}^{10} x_i/10 = 13,4 \quad \text{e} \quad \bar{y} = \sum_{i=1}^{10} y_i/10 = 8,1.$$

As variâncias são

$$s_x^2 = \frac{\sum_{i=1}^{10} x_i^2 - \left(\sum_{i=1}^{10} x_i\right)^2/10}{9} = \frac{1928 - 134^2/10}{9} = 14,711$$

$$s_y^2 = \frac{\sum_{i=1}^{10} y_i^2 - \left(\sum_{i=1}^{10} y_i\right)^2/10}{9} = \frac{757 - 81^2/10}{9} = 11,211.$$

Portanto, os desvios padrões são $s_x = \sqrt{14,71} = 3,835$ e $s_y = \sqrt{11,289} = 3,348$.

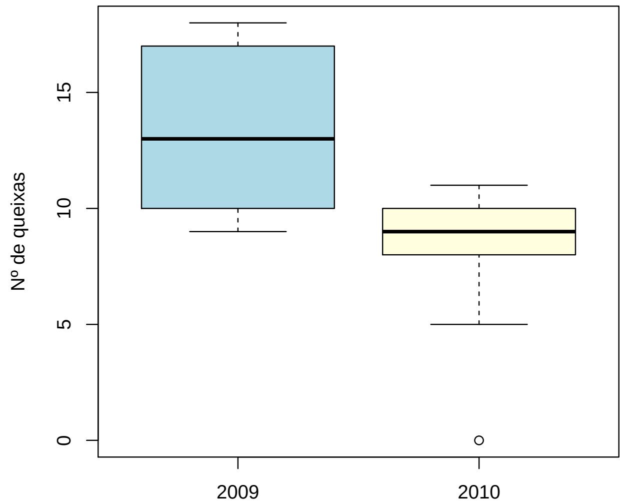
(b) Os quartis são

$$Q1_x = x_{(3,25)} = 10 \quad \text{e} \quad Q1_y = y_{(3,25)} = 8$$
$$Q2_x = x_{(5,5)} = 13 \quad \text{e} \quad Q2_y = y_{(5,5)} = 9$$
$$Q3_x = x_{(7,25)} = 17 \quad \text{e} \quad Q3_y = y_{(7,25)} = 10.$$

As cercas são

$$LI_x = 10 - 1,5 \times (17-10) = -0,5 \quad \text{e} \quad LI_y = 8 - 1,5 \times (10-8) = 5,0$$

$$LS_x = 17 + 1,5 \times (17-10) = 27,5 \quad \text{e} \quad LS_y = 10 + 1,5 \times (10-8) = 13,0.$$



(c)

- Existe diferença aparente entre os dados de 2009 e 2010.
- Com relação à posição, o número de queixas em 2010 é menor que em 2009.
- Há maior dispersão no número de queixas em 2009 comparado com 2010.
- O valor 0, observado em 2010, é discrepante.

3. **(a)** O intervalo não conservativo ao nível de confiança de $0,99$ para p é dado por: $\hat{p} \pm z_{1-\frac{\alpha}{2}}\sqrt{\frac{\hat{p}(1-\hat{p})}{n}}$. Neste caso, temos $\hat{p} = 17/40 = 0,425$. Por outro lado, $1-\alpha = 0,99$ implica que $1-\alpha/2 = 0,995$. Consultando a tabela da N(0,1), temos $z_{0,995} = 2,58$. O intervalo de confiança é, portanto, $0,425 \pm 2,58\sqrt{\frac{0,425\times 0,575}{40}}$. Fazendo os cálculos: $(0,223; 0,627)$.

   (b) Para que tenhamos $P(\mid \hat{p}-p \mid < 0,05) = 0,98$, o tamanho da amostra deve ser dado pela expressão: $n = \left(\frac{z_{0,99}}{0,05}\right)^2 p(1-p)$ Como $0,1 \leq p \leq 0,7$, o valor máximo do produto $p(1-p)$ dentro desse intervalo é $0,5\times(1-0,5) = 0,25$. Por outro lado, $z_{0,99} = 2,33$. Então devemos usar uma amostra de tamanho $n = \left(\frac{2,33}{0,05}\right)^2 \times 0,25 = 542,89 \approx 543$. Como a amostra piloto foi de tamanho 40, serão necessários 503 usuários adicionais aos da piloto.

   (c) Para que tenhamos $P\left[\frac{|\hat{p}-p|}{p} < 0,20\right] = P[\mid \hat{p}-p \mid < 0,20p] = 0,95$, o tamanho da amostra deve ser dado pela expressão: $n = \left(\frac{z_{0,975}}{0,20p}\right)^2 p(1-p) = \left(\frac{1,96}{0,20}\right)^2 \frac{1-p}{p}$, porque $z_{0,975} = 1,96$. Ora, esta é uma função decrescente de p. Então, como $0,1 \leq p \leq 0,7$, o valor máximo do quociente $\frac{1-p}{p}$ nesse intervalo é $\frac{1-0,1}{0,1} = 9$. Então devemos usar uma amostra de tamanho $n = \left(\frac{1,96}{0,20}\right)^2 \times 9 = 864,36 \approx 865$. Como a amostra piloto foi de tamanho 40, serão necessários 825 usuários adicionais aos da piloto.

4. (a) Temos o teste bilateral dado por $\left\{\begin{array}{l} H_0: \ \mu = 1600 \\ H_1: \ \mu \neq 1600 \end{array}\right.$, cuja região de rejeição é dada por

$$
\begin{aligned}
\text{RC} &= \{\overline{x} : (\overline{x} < \mu_0 - z_{1-\alpha/2}\sigma/\sqrt{n}) \cup (\overline{x} > \mu_0 + z_{1-\alpha/2}\sigma/\sqrt{n})\} \\
&= \{\overline{x} : (\overline{x} < 1600 - 1,96 \times 120/10) \cup (\overline{x} > 1600 + 1,96 \times 120/10)\} \\
&= \{\overline{x} : (\overline{x} < 1576,48) \cup (\overline{x} > 1623,52)\} \\
&= \{z_{obs} : (z_{obs} < -z_{1-\alpha/2}) \cup (z_{obs} > z_{1-\alpha/2}) \\
&= \{z_{obs} : (z_{obs} < -1,96) \cup (z_{obs} > 1,96)
\end{aligned}
$$

com

$$z_{obs} = \frac{\overline{x} - \mu_0}{\sigma/\sqrt{n}} = \frac{1615 - 1600}{120/10} = 1,25.$$

Utilizamos o desvio padrão populacional conhecido e consequentemente o quantil de 0,975 da distribuição normal padrão. **Conclusão:** Temos que $\overline{x} = 1615$ não pertence a região crítica, portanto não rejeitamos a hipótese nula de que a média populacional seja igual a 1600 horas.

(b)

$$
\begin{aligned}
\beta &= \Pr(\text{“Erro Tipo II”}) = \Pr(\text{“Não rejeitar } H_0\text{”}), \text{ se “}H_1 \text{ é verdadeira”} \\
&= \Pr(1576,48 \leqslant \overline{X} \leqslant 1623,52), \text{ se } \mu = 1620 \\
&= \Pr\left(\frac{1576,48 - 1620}{120/10} \leqslant Z \leqslant \frac{1623,52 - 1620}{120/10}\right) \\
&= \Pr(-3,6267 \leqslant Z \leqslant 0,2933) \\
&= \Phi(0,2933) - \Phi(-3,6267) = 0,6152.
\end{aligned}
$$