CE-003: Estatística II - Turma: O, 1a Prova (25/04/2012)

GRR: ___________________________ **Nome:** ____

1. Foram feitas medições dos tempos de atendimento e solução de solicitações feitas por cliente de dois provedores de serviços (A e B). Os valores obtidos estão representados nos gráficos da figura a seguir.
   a. Indique qual *boxplot* da figura à direita correspondente cada curva da figura à esquerda. Justifique sua resposta.
   b. Em um dos provedores a média foi de 44.6 e a mediana 40.6, enquanto que no outro a média foi 49.5 e a mediana 49.2. Quais valores correspondem a cada provedor? Justifique sua resposta.
   c. Interprete e discuta cada um dos gráficos, comparando os provedores do serviço.

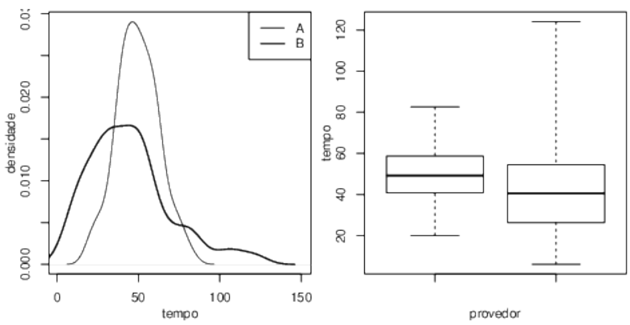


Figura 1: Tempo de atendimento de solicitações de dois provedores de serviços.

Solução:
Pontos para notar/comentar: assimetria, amplitude dos valores, variabilidade, diferença entre medianas

2. A tabela a seguir apresenta as notas de matemática no vestibular e na disciplina de cálculo de alguns alunos selecionados ao acaso. Pretende-se examinar os desempenhos nestas provas e se há relação entre os desempenhos.

Aluno	Vestibular	Cálculo	Aluno	Vestibular	Cálculo
1	37	65	7	35	50
2	57	92	8	80	90
3	34	56	9	65	88
4	40	70	10	47	71
5	21	52	11	28	52
6	28	73	12	67	88

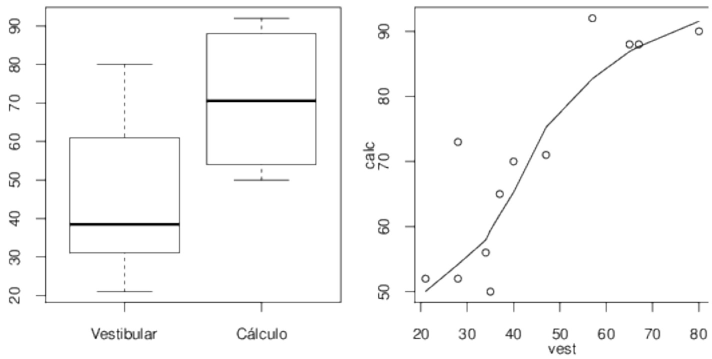
   a. Calcule a mediana, quartis e amplitude interquartílica das notas de cálculo.
   b. Calcule o coeficiente de variação das notas do vestibular e de cálculo.
   c. Construa um diagrama "ramo-e-folhas"com todas as notas (vestibular e cálculo) e marque (sublinhe) nas "folhas"os dados da prova de cálculo.
   d. Faça um gráfico com os diagramas "box-plot"das duas notas (um "boxplot"para cada).
   e. Construa um gráfico adequado para representar os dados das duas provas conjuntamente. Calcule medida(s) de associação adequada(s).
   f. Compare, interprete e discuta os resultados.

Solução:

a.
```
medianaV      q1V      q3V     AIQV
    38.5     31.0     61.0     30.0

medianaC      q1C      q3C     AIQC
    70.5     54.0     88.0     34.0
```
b.
```
      mediaV varianciaV        sdV
       44.92     338.27      18.39

      mediaC varianciaC        sdV
       70.58     255.17      15.97

CVvestibular    CVcalculo
       40.95        22.63
```
c.
```
  The decimal point is 1 digit(s) to the right of the |

  2 | 188457
  4 | 0702267
  6 | 557013
  8 | 08802

  The decimal point is 1 digit(s) to the right of the |

  2 | 188
```

```
3 | 457
4 | 07
5 | 02267
6 | 557
7 | 013
8 | 088
9 | 02
```

---

d. 

---

e.
```
pearson  kendall spearman
 0.8675   0.6357   0.7750
```

f. Comentários:

O CV permite comparar a variabilidade de grupos de diferentes médias, que é o caso neste exemplo. A medida Mostra que as notas de cálculo são mais homogêneas do que as do vestibular, em relação às suas médias, embora as variabilidade absolutas sejam semelhantes.

Os gráficos box-plot e ramo-e-folhas mostram valores nitidamente mais elevados para notas de cálculo, com variabilidades absolutas semelhantes, uma leve assimetria nas notas do vestibular com maior concentração de valores baixos e sem presença de observações discrepantes.

O diagrama de dispersão mostra uma relação ligeiramente não linear, positiva e sem presença de dados Discrepantes, embora com os dados dispostos em dois grupos separados de valores baixos e altos. Desta forma os diferentes coeficientes de correlação apresentam valores um pouco diferentes como de Pearson mais elevado devido à posição dos grupos distintos e moderada associação.

---

3. Em um grupo de estudantes 45% são do curso A, 25% do curso B o restante do curso C. A proporção de mulheres em cada curso um dos cursos é de 20, 50 e 75%, respectivamente. Se um estudante é sorteado qual a probabilidade de:
   a. seja homem;
   b. seja do curso A, sabendo que foi sorteada uma mulher;
   c. seja do curso C sabendo que foi sorteado um homem.

Solução:

$$P[A] = 0{,}45 \quad ; \quad P[B] = 0{,}25 \quad ; \quad P[C] = 0{,}30$$
$$P[M|A] = 0{,}20 \quad ; \quad P[M|B] = 0{,}50 \quad ; \quad P[M|C] = 0{,}75$$
$$P[H|A] = 0{,}80 \quad ; \quad P[H|B] = 0{,}50 \quad ; \quad P[H|C] = 0{,}25$$

a. $P[H] = 1 - P[M] = 1 - (P[M \cap A] + P[M \cap B] + P[M \cap C]) = 1 - (P[M|A] \cdot P[A] + P[M|B] \cdot P[B] + P[M|C] \cdot P[C]) = 1 - (0{,}20 \cdot 0{,}45 + 0{,}50 \cdot 0{,}25 + 0{,}75 \cdot 0{,}30) = 1 - 0.44 = 0.56$

b. $P[A|M] = \frac{P[A \cap M]}{P[M]} = \frac{P[M|A] \cdot P[A]}{P[M]} = \frac{0.09}{0.44} = 0.205$

c. $P[C|H] = \frac{P[C \cap H]}{1 - P[M]} = \frac{P[H|C] \cdot P[C]}{P[H]} = \frac{0.075}{0.56} = 0.134$

---

4. A probabilidade de ocorrer falha/corrupção na cópia de um arquivo é de 0,03. Em 10.000 cópias feitas em um sistema mostre como calcular as probabilidades a seguir de ao menos duas maneiras diferentes.
   a. De que não ocorram falhas;
   b. de que ocorram o máximo três falhas;
   c. de que ocorram no máximo quatro falhas, sabendo que ao menos uma falha ocorreu.

Solução:

$$P[F] = 0{,}03 \quad ; \quad P[\overline{F}] = 0{,}97$$
$$X : \text{número de falhas}$$
$$\text{Solução 1: } X_B \sim B(n = 10.000, p = P[F] = 0{,}03)$$
$$\text{Solução 2: } X_P \approx P(\lambda = n \cdot p = 10.000 \cdot 0{,}03 = 300)$$
$$\text{Solução 3: } X_N \approx N(\mu = n \cdot p = 10.000 \cdot 0{,}03 = 300 \quad ; \quad \sigma^2 = n \cdot p \cdot (1 - p) = 291)$$

a.

$$P[\overline{F}_1 \ldots \overline{F}_{10.000}] \overset{ind}{=} 0{,}97^{10.000} = 5.216e - 133$$
$$P[X = 0] = P[X_B = 0] = \binom{10000}{0}(0{,}03)^0(1 - 0{,}03)^{10000} = 5.216e - 133$$
$$P[X = 0] \approx P[X_P = 0] = \frac{e^{-300}300^0}{0!} = 5.148e - 131$$

b.

$$P[X \leq 3] = P[X_B \leq 3] = \sum_{i=0}^{3} \binom{10000}{i} (0{,}03)^i (1 - 0{,}03)^{10000-i} = 2.596e - 126$$

$$P[X \leq 3] = P[X_P \leq 3] = \sum_{i=0}^{3} = \frac{e^{-300} 300^i}{i!} = 2.34e - 124$$

c.

$$P[X \leq 4 | X \geq 1] =$$

$$\frac{P[1 \leq X \leq 4]}{P[X \geq 1]} = \frac{P[X = 1] + P[X = 2] + P[X = 3] + P[X = 4]}{1 - P[X = 0]} =$$

$$= \frac{P[X_B = 1] + P[X_B = 2] + P[X_B = 3] + P[X_B = 4]}{1 - P[X_B = 0]} = \frac{\sum_{i=1}^{4} \binom{10000}{i} (0{,}03)^i (1 - 0{,}03)^{10000-i}}{1 - P[X_B = 0]} = 2.013e - 124$$

$$\approx \frac{P[X_P = 1] + P[X_P = 2] + P[X_P = 3] + P[X_P = 4]}{1 - P[X_P = 0]} = \frac{\sum_{i=1}^{4} \frac{e^{-300} 300^i}{i!}}{1 - P[X_B = 0]} = 1.761e - 122 =$$

5. Considere a função $f_X(x) = k(1 + 2x)\ \ I_{(0,2)}(x)$
   - mostre que $f(x)$ é função de densidade de probabilidade (f.d.p.) para algum valor de k e determine o valor de k.
   - obtenha $F(x)$
   - obtenha $E[X]$
   - obtenha $P[X > 1]$
   - obtenha a tal que $P[X < a] = 0{,}6$

Solução:

- $$\int_0^2 f(x) \mathrm{d}x = 1 \longrightarrow k(2 + 4) = 1 \longrightarrow k = 1/6$$
  $$0 \leq f(x) \leq 1 \quad \forall \quad 0 < x < 2$$

  ```
  > fx <- function(x, k) ifelse(x > 0 & x<2, k*(1+2*x), 0)
  > integrate(fx, 0, 2, k=1/6)
  ```

  ```
  1 with absolute error < 1.1e-14
  ```

- $$F(x) = \int_0^x f(x) \mathrm{d}x = \frac{x(1 + x)}{6}$$
- $$E(x) = \int_0^2 x f(x) \mathrm{d}x = 1.222$$
- $$P[X > 1] = \int_1^2 f(x) \mathrm{d}x = \frac{1}{6}(1 + 3) = 0.667$$
- $$P[X < a] = 0{,}6 \longrightarrow \int_0^a f(x) \mathrm{d}x = 0{,}6 \longrightarrow \frac{a(1 + a)}{6} = 0{,}6 \longrightarrow a = 1.462$$

---

6. O tempo de montagem de um determinado mecanismo é uma v.a. uniforme mo intervalo de 30 a 40 segundos. Determine:
   - as expressões de $f(x)$ e $F(x)$;
   - a probabilidade da montagem ser feita em menos que 33 segundos;
   - a probabilidade da montagem ser feita em menos que 38 segundos, sabendo que foi maior que 35 segundos;
   - o tempo abaixo do qual 80% das montagens são feitas;
   - o tempo esperado para montagem de 5.000 peças;
   - o custo esperado para montagem de 5.000 peças sabendo que montagens abaixo de 33 segundos tem custo de R$1,00, entre 33 e 38 segundos o custo é de R$1,50 e acima de 38 segundos o custo é de R$3,00.

Solução:

$$X: \text{ tempo de montagem } X \sim U[30, 40]$$

- $$f(x) = \frac{1}{40 - 30}\ I_{[30,40]}(x) = \frac{1}{10}\ I_{[30,40]}(x)$$
  $$F(x) = \frac{x - 30}{40 - 30}\ I_{[30,40]}(x) = \frac{x - 30}{10}\ I_{[30,40]}(x)$$
- $P[X < 33] = F(33) = \frac{33-30}{10} = 0{,}30$
- $P[X < 38 | X > 35] = \frac{P[35 < X < 38]}{P[X > 35]} = \frac{F(38) - F(35)}{1 - F(35)} = 0{,}60$
- $P[X < t_1] = 0{,}80 \longrightarrow t_1 = 38$
- $5.000 \cdot E[X] = 5.000 \cdot \frac{30+40}{2} = 175000$ segundos = 48.61 horas
- Y : custo por peça

y	1,00	1,50	3,00
P[Y=y]	$P[X \leq 33] = 0.3$	$P[33 < X < 38] = 0.5$	$P[X \geq 38] = 0.2$

Custo = $5.000 \cdot E[Y] = 5.000 \cdot (1{,}00 \cdot P[Y = 1{,}00] + 1{,}50 \cdot P[Y = 1{,}50]) + 3{,}00 \cdot P[Y = 3{,}00]) = 8250.$
