Introdução à Probabilidade e Estatística

MA

2º Trimestre – 2024/2025
26/11/2024 — 12:30, VA4
MAP45 #1 — 45 minutos

- Justifique convenientemente as suas respostas e escreva os resultados com casas decimais.
- Só pode sair da sala vinte e cinco minutos após o início do MAP, devendo nesse caso entregar a sua prova ou desistir da mesma. Escreva o seu número e nome completo abaixo.

Número: **Nome completo**:

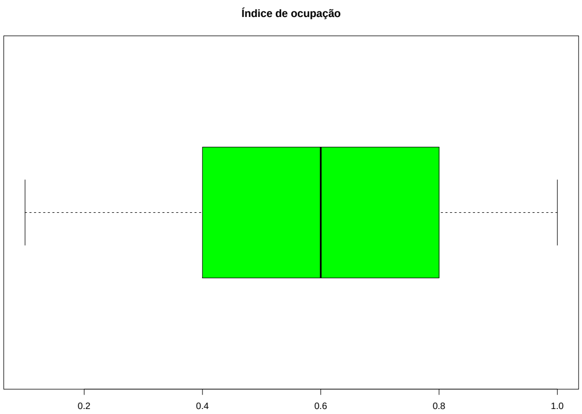
1. Admita que a amostra ordenada seguinte diz respeito ao índice de ocupação (quociente entre a área total de implantação e a área de solo a que o índice diz respeito) arredondado às décimas de $n = 25$ lotes de terreno:

$0.1, 0.1, 0.2, 0.3, 0.3, \quad 0.3, 0.4, 0.5, 0.5, 0.5, \quad 0.5, 0.5, 0.6, 0.6, 0.6, \quad 0.7, 0.7, 0.8, 0.8, 0.8, \quad 0.8, 0.9, 0.9, 0.9, 1.0.$

(a) Obtenha a média, a variância e o segundo coeficiente de assimetria de Pearson, sabendo que $\sum_{i=1}^{n} x_i = 14.3$ e $\sum_{i=1}^{n} x_i^2 = 9.79$. Comente este último valor. **(1.5)**

- **Média**
  $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{14.3}{25} = 0.572$
- **Variância**
  $s^2 = \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 - n\,(\bar{x})^2\right] = \frac{1}{25-1}\left(9.79 - 25 \times 0.572^2\right) = \frac{1.6104}{24} = 0.0671.$
- **2o. coeficiente de assimetria de Pearson**
  Como $n = 25$ é ímpar, a mediana amostral é igual a $me = q_{1/2} = x_{((n+1)/2)} = x_{(26/2)} = x_{(13)} = 0.6$ e
  $$scp_2 = 3 \times \frac{\bar{x} - me}{s} = 3 \times \frac{0.572 - 0.6}{\sqrt{0.0671}} \simeq -0.324278.$$
- **Comentário**
  Este valor negativo de scp_2 próximo de zero sugere uma ligeira assimetria negativa [como sugere o diagrama de extremos-e-quartis abaixo].

(b) Obtenha os cinco valores-chave usados para traçar o seguinte diagrama de extremos-e-quartis associado ao conjunto de dados acima. **(2.5)**



Por que razão não foram assinaladas quaisquer observações discordantes (*outliers*) neste diagrama?

- **1o. quartil**
  Como $n/4 = 25/4$ não é um inteiro, temos $q_{1/4} = x_{([n/4]+1)} = x_{([6.25]+1)} = x_{(7)} = 0.4.$
- **2o. quartil ou mediana,** $q_{1/2}$
  $me = q_{1/2} \stackrel{(a)}{=} 0.6.$
- **3o. quartil**
  Como $3n/4 = 75/4$ não é um inteiro, temos $q_{3/4} = x_{([3n/4]+1)} = x_{([18.75]+1)} = x_{(19)} = 0.8.$

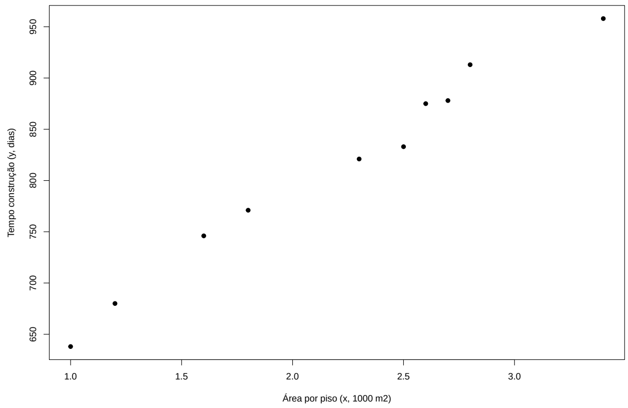
- **Restantes dois pontos**
  Uma vez que a amplitude inter-quartil é igual a $IQR = q_{3/4} - q_{1/4} = 0.8 - 0.4 = 0.4$, temos:
  $$\text{“mínimo”} = \max\{x_{(1)}, q_{1/4} - 1.5 \times IQR\} = \max\{0.1, 0.4 - 1.5 \times 0.4\} = \max\{0.1, -0.2\} = 0.1;$$
  $$\text{“máximo”} = \min\{x_{(n)}, q_{3/4} + 1.5 \times IQR\} = \min\{1.0, 0.8 + 1.5 \times 0.4\} = \min\{1.0, 1.4\} = 1.0.$$

- **Comentário**
  Não foram assinaladas quaisquer observações discordantes neste diagrama, pois
  $x_i \in [\text{“mínimo”}, \text{“máximo”}], \forall i = 1, \ldots, n$.

(c) Determine a percentagem de observações que não excedem o mais frequente dos valores da amostra. **(0.5)**

- **Percentagem pedida**
  Atendendo a que a dimensão da amostra é igual a $n = 25$, o valor mais frequente entre os registados acima é $mo = 0.5$ e há 12 observações inferiores ou iguais a $mo = 0.5$ na amostra acima, a percentagem pedida é dada por
  $\frac{\text{no. observações}_{\leq mo}}{n} \times 100\% = \frac{12}{25} \times 100\% = 48\%.$

2. O diagrama de dispersão abaixo diz respeito a um conjunto de $n = 10$ observações, (x_i, y_i), $i = 1, \ldots, n$, das variáveis área por piso (x, em $10^3 m^2$) e tempo de construção (y, em dias) de edifícios de 10 pisos. **(2.1)**



Adiante-se que: $\bar{x} = 2.19$; $\sum_{i=1}^n x_i^2 = 53.23$; $\bar{y} = 811.3$; $\sum_{i=1}^n y_i^2 = 6676173$; $\sum_{i=1}^n x_i y_i = 18465.4$.

Calcule o coeficiente de correlação amostral entre estas duas variáveis e comente brevemente o valor obtido.

- **Coeficiente de correlação amostral (de Pearson)**
  $$r = \frac{\sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}}{\sqrt{(\sum_{i=1}^n x_i^2 - n\bar{x}^2) \times (\sum_{i=1}^n y_i^2 - n\bar{y}^2)}} = \frac{18465.4 - 10 \times 2.19 \times 811.3}{\sqrt{(53.23 - 10 \times 2.19^2) \times (6676173 - 10 \times 811.3^2)}} \simeq \frac{697.93}{\sqrt{5.269 \times 94096.1}} \simeq 0.991202.$$

- **Comentário**
  As variáveis área por piso (x) e tempo de construção (y) tendem a variar no mesmo sentido relativamente às respectivas médias (pois $r > 0$) e estão linear e fortemente correlacionadas (já que $r \simeq 0.991202$ está muito próximo de 1)[, facto este corroborado também pelo diagrama de dispersão].