Bioestatística — CE001
Prof. Fernando de Pol Mayer — Departamento de Estatística — DEST
Exercícios: medidas resumo
Nome: GABARITO GRR:

---

1. Estime as medidas de centro (média, mediana, moda) para amostras de altura de uma espécie de árvore (metros), coletadas em quatro áreas diferentes:
   a) Área A: 9,2 10,8 10,6 11,1 12,1 9,6 11,2 8,4 12,9 12,1 14,4 11,1 11,1 9,7 8,4 12,3 10,7 12,9 9,1 12,8;
   b) Área B: 12,5 18,5 21,3 14,3 18,5 19,0 10,8 23,1 17,4 10,7 14,3 16,3 18,0 7,1 12,8 14,7 11,3 8,2 13,8;
   c) Área C: 21,3 28,7 15,8 24,0 13,7 18,1 12,6 14,6 6,1 19,8 22,3 15,7 16,3 18,2 15,7 6,6 9,3 1,3 19,0;
   d) Área D: 13,7 8,6 14,9 10,2 14,0 10,5 15,0 5,2 10,0 11,7 18,7 9,3 7,9 6,5 11,5 12,0 8,3 8,3 9,8 4,7.

```
$a
  media mediana    moda
 11.025  11.100  11.100

$b
   media  mediana    moda1    moda2
14.87368 14.30000 14.30000 18.50000

$c
   media  mediana     moda
15.74211 15.80000 15.70000

$d
  media mediana    moda
  10.54   10.10    8.30
```

---

2. Calcule a amplitude, a variância, o desvio-padrão, e o coeficiente de variação para as quatro amostras do exercício anterior.

```
              a         b         c         d
media 11.025000 14.873684 15.742105 10.540000
AMP    6.000000 16.000000 27.400000 14.000000
VAR    2.657763 18.556491 43.942573 12.318316
DP     1.630265  4.307725  6.628919  3.509746
CV    14.786982 28.962055 42.109485 33.299296
```

---

3. Descreva comparativamente as quatro áreas quanto à altura das árvores, utilizando as estatísticas que você calculou. (*Exemplo de resposta*) Em média, a área C possui as árvores mais altas ($\bar{x}_C = 14,874$), enquanto que a área D possui as árvores mais baixas ($\bar{x}_D = 10,54$). Em todas as áreas, o valor da mediana está muito próximo do valor da média e da moda, o que indica que a distribuição das alturas em todas as áreas é aproximadamente simétrica. A maior amplitude de variação de alturas foi observada na área C, que também apresentou a maior variabilidade das observações em relação à média, como pode ser observado pelos valores da variância ($s_C^2 = 43,943$), e do desvio-padrão ($s_C = 6,629$). A área com árvores de alturas mais homogêneas foi a A, pois a amplitude foi de 6 m, e a varibilidade das alturas em torno da média foi a menor quando comparada com as demais áreas ($s_A^2 = 2,658$ e $s_A = 1,63$). Estas diferenças de variabilidade podem ser observadas através do coeficiente de variação, que foi de 42,1% para a área C, e de 14,8% para a área A. A área D apresentou um CV de 33,3%, enquanto que o CV da área B foi de 28,9%.

---

4. Um exame vestibular para uma faculdade tem 80 questões, sendo 40 de português e 40 de matemática. Para os 20 melhores classificados, apresentamos o número de acertos em cada disciplina.
   - Português: 35, 35, 34, 32, 31, 30, 26, 26, 24, 23, 23, 12, 11, 20, 17, 12, 14, 20, 8, 10
   - Matemática: 31, 29, 27, 28, 28, 26, 30, 28, 25, 23, 21, 32, 31, 20, 21, 25, 20, 13, 23, 20

   (a) Calcule as medidas de centro: média, mediana e moda para cada grupo

```
$port
  media mediana   moda1   moda2   moda3   moda4   moda5
  22.15   23.00   12.00   20.00   23.00   26.00   35.00
```

```
$mat
  media mediana   moda1   moda2
  25.05   25.50   20.00   28.00
```

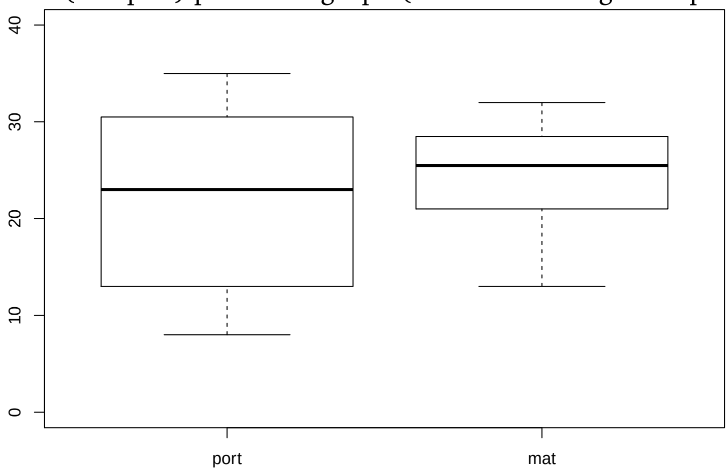
(b) Calcule as medidas de variabilidade: variância, desvio-padrão, e coeficiente de variação para cada grupo

```
         port       mat
VAR 80.134211 23.839474
DP   8.951771  4.882568
CV  40.414318 19.491291
```

(c) Calcule o resumo dos cinco números para cada grupo

```
    port  mat
Min  8.0 13.0
Q1  13.0 21.0
Q2  23.0 25.5
Q3  30.5 28.5
Max 35.0 32.0
```

(d) Construa um gráfico de caixa (box plot) para cada grupo (em um mesmo gráfico para comparação)

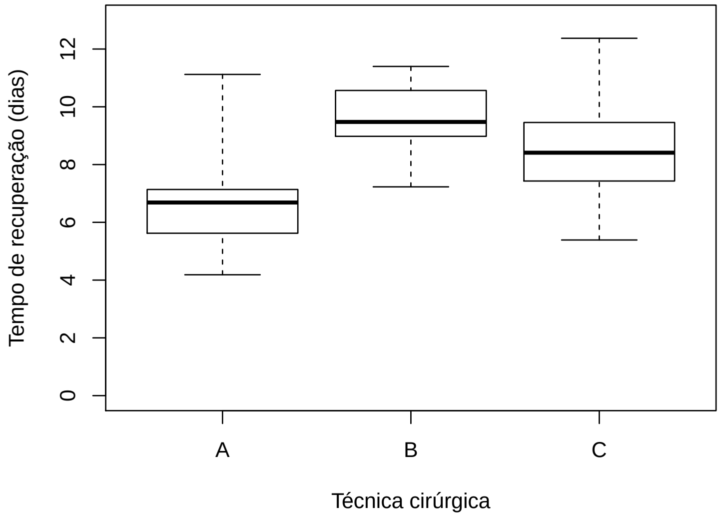


(e) Com todos os resultados obtidos, descreve comparativamente estes dois grupos em termos de medidas de tendência central, variabilidade, amplitude e distribuição (simetria) dos dados.
(*Exemplo de resposta*) Em média, o número de acertos em matemática ($\bar{x}_{mat} = 25,05$) foi maior do que o número de acertos em português ($\bar{x}_{port} = 22,15$). A diferença entre os valores médios e a mediana mostra que existe uma leve assimetria negativa (ou à esquerda) para os dois casos ($\bar{x} <$ Me), embora esta diferença seja mais pronunciada nas notas de português. A amplitude do número de acertos em português foi de $\text{AMP}_{port} = 35 - 8 = 27$, maior do que a amplitude obervada para o número de acertos em matemática, que foi de $\text{AMP}_{port} = 32 - 13 = 19$. A variabilidade dos acertos em torno da média também foi maior para as notas de português, com variância de $s^2_{port} = 80,134$ e desvio-padrão de $s_{port} = 8,951$. Já para a matemátiva, a variabilidade em torno da média foi menor, com $s^2_{mat} = 23,839$ e desvio-padrão $s_{mat} = 4,883$. Resumindo estas informações sobre variabilidade, nota-se que o coeficiente de variação para português foi de 40,4%, enquanto que para a matemática foi menor, com aproximadamente 19,5%. Através do resumo dos cinco números e do gráfico de caixa, percebe-se que 50% dos acertos foram entre 13 e 30,5 em português (diferença entre Q1 e Q3), e entre 21 e 28,5 em matemática, mostrando novamente a menor variabilidade observada para a matemática.

(f) Você acha que os aprovados são melhores em português ou matemática?

---

5. Deseja-se comparar três técnicas cirúrgicas para a extração do dente siso. Cada uma das técnicas foi aplicada a 30 pacientes, e os resultados são apresentados em diagramas de caixa abaixo.



(a) Encontre os valores (aproximados) para a mediana, os quartis, máximo e mínimo.

```
## Observação: estes são os valores exatos!
            A         B         C
Min  4.184541  7.228075  5.388393
Q1   5.623099  8.977717  7.430463
Q2   6.686161  9.476333  8.410870
Q3   7.137260 10.565402  9.456991
Max 11.120594 11.397466 12.372938
```

(b) Discuta a variabilidade do tempo de recuperação em cada técnica.
(*Exemplo de resposta*) A amplitude de dias de recuperação para as técnicas A, B, e C, são, respectivamente: $\text{AMP}_A = 11,121 - 4,185 = 6,936$, $\text{AMP}_B = 11,398 - 7,228 = 4,17$, $\text{AMP}_C = 12,373 - 5,388 = 6,985$. Com isso, percebe-se que a técnica A apresenta um tempo de recuperação (aproximado) entre 4 e 11 dias, a técnica B entre 7 e 11 dias, e a técnica C entre 5 e 12 dias. A maior variação de tempo de recuperação foi então da técnica C (maior amplitude), enquanto que a técnica B apresentou a menor variação. Nota-se, em termos de mediana, que a técnica A apresentou o menor tempo de recuperação (Me = 6,69), embora com uma grande variabilidade. Além disso, percebe-se através do gráfico de caixa, que a técnica A possui uma distribuição com assimetria positiva, como pode ser observado pela proximidade da mediana com o terceiro quuartil (Q3), além de uma extensão maior da caixa até o valor máximo. A técnica B apresentou um tempo de recuperação mediano de 9,47 dias, o maior observado entre as 3 técnicas. Percebe-se também que a distribuição do tempo de recuperação para esta técnica é levemente assimétrica à esquerda, pois a mediana está deslocada para baixo dentro da caixa. O tempo de recuperação mediano para a técnica C foi de 8,41 dias. Para esta técnica, 50% das pessoas tiveram um tempo de recuperação entre 7,43 (Q1) e 9,46 (Q3) dias.

(c) Se você é otimista, qual técnica escolheria?

---

6. A distribuição das estaturas, em centímetros, de alunos de um curso colegial está representada na tabela de frequência abaixo. Calcule a média, a variância, e o desvio-padrão das estaturas.

Classes	Frequência	PM = x_i	$x_i f_i$	x_i^2	$x_i^2 f_i$
$135 \vdash 145$	15	140	2100	19600	294000
$145 \vdash 155$	150	150	22500	22500	3375000
$155 \vdash 165$	250	160	40000	25600	6400000
$165 \vdash 175$	70	170	11900	28900	2023000
$175 \vdash 185$	10	180	1800	32400	324000
$185 \vdash 195$	5	190	950	36100	180500
Total	500		79250		12596500

$$
\begin{aligned}
\bar{x} &= \frac{1}{n}\sum_{i=1}^{n} x_i f_i \\
&= \frac{1}{500}\cdot 79250 \\
&= 158,5
\end{aligned}
$$

$$
\begin{aligned}
s^2 &= \frac{1}{n-1}\left[\sum_{i=1}^{n} x_i^2 f_i - \frac{(\sum_{i=1}^{n} x_i f_i)^2}{n}\right] \\
&= \frac{1}{500-1}\left[12596500 - \frac{(79250)^2}{500}\right] \\
&= 70,892
\end{aligned}
$$

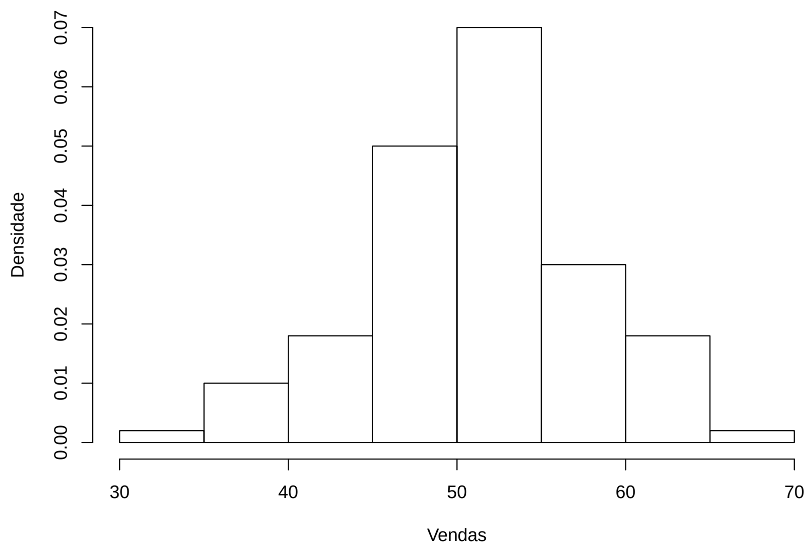
$$
\begin{aligned}
s &= \sqrt{70,892} \\
&= 8,419
\end{aligned}
$$

---

7. Os dados abaixo representam as vendas semanais, em classes de salários mínimos, de vendedores de gêneros alimentícios:

Vendas	Vendedores
$30 \vdash 35$	2
$35 \vdash 40$	10
$40 \vdash 45$	18
$45 \vdash 50$	50
$50 \vdash 55$	70
$55 \vdash 60$	30
$60 \vdash 65$	18
$65 \vdash 70$	2

(a) Faça o histograma das observações



(b) Calcule a média da amostra, $\bar{x}$

```
[1] 51.2
```

(c) Calcule o desvio-padrão da amostra, s

```
[1] 6.618912
```

(d) Qual a porcentagem das observações compreendidas entre $\bar{x} - 2s$ e $\bar{x} + 2s$, *aproximadamente*?

```
## xbarra - 2s
[1] 37.96218
## xbarra + 2s
[1] 64.43782
## Percentual aproximado de observações nesse intervalo (use as
## frequências acumuladas relativas e o ponto médio das classes)
[1] 0.93
```

(e) Calcule a mediana

```
[1] 52.5
```

---

8. O departamento pessoal de uma certa firmafez um levantamento dos salários dos 120 funcionários do setor administrativo, obtendo os resultados (em salários mínimos) da tabela abaixo:

Salários	Freq.relativa
$0 \vdash 2$	0.25
$2 \vdash 4$	0.40
$4 \vdash 6$	0.20
$6 \vdash 10$	0.15

(a) Calcule a média, a mediana, a variância e o desvio-padrão

```
## Média
[1] 3.65
## Variância
[1] 5.1275
## Desvio-padrão
[1] 2.264398
## Mediana (pela frequência acumulada relativa)
[1] 3
```

(b) Se for concedido um aumento de 100% para todos os 120 funcionários, haverá alteração na média? E na variância? Justifique sua resposta.

```
## Média -----------------------------------------------------------
[1] 7.3
## Sim, a média é multiplicada por 2
## Variância -------------------------------------------------------
[1] 20.51
## Sim, a variância é multiplicada por 4
```

(c) Se for concedido um abono de dois salários mínimos para todos os 120 funcionários, haverá alteração na média? E na mediana? E na variância? Justifique sua resposta.

```
## Média -----------------------------------------------------------
[1] 5.65
## Sim, a média aumenta em duas unidades
## Mediana ---------------------------------------------------------
[1] 5
## Sim, a mediana aumenta em duas unidades
## Variância -------------------------------------------------------
[1] 5.1275
## Não, a variância não se altera
```

---

9. O resultado de uma prova de estatística aplicada à 25 alunos foi o seguinte:

```
[1]  9  9  8  8  9 10  8  8  9  8 10  7  7  9  9  7  8  9  4  7  7  8 10  9  9
```

Como os alunos possuiam diferentes níveis educacionais, decidiu-se calcular o desempenho relativo de cada candidato, para facilitar a interpretação dos resultados. Esse medida de desempenho relativo será obtida da seguinte forma:

1. Calcula-se a média $\bar{x}$ e o desvio-padrão s da amostra
2. A nota x_i de cada aluno será padronizada da seguinte forma:

$$z_i = \frac{x_i - \bar{x}}{s}$$

criando assim uma nova variável Z que corresponde ao conjunto de notas padronizadas.

Com isso:

(a) Calcule as notas padronizadas de todos os funcionários.

```
## Média
[1] 8.24
## Desvio-padrão
[1] 1.273735
## Valores padronizados
 [1]  0.5966706  0.5966706 -0.1884223 -0.1884223  0.5966706  1.3817634
 [7] -0.1884223 -0.1884223  0.5966706 -0.1884223  1.3817634 -0.9735152
[13] -0.9735152  0.5966706  0.5966706 -0.9735152 -0.1884223  0.5966706
[19] -3.3287938 -0.9735152 -0.9735152 -0.1884223  1.3817634  0.5966706
[25]  0.5966706
```

(b) Com os resultados obtidos acima, calcule a média ($\bar{z}$) e o desvio-padrão (s_z) de Z.

```
## Média de Z
[1] 0
## Desvio-padrão de Z
[1] 1
```

(c) Se alguma das notas padronizadas estiver acima de $2s_z$ ou abaixo de $-2s_z$, esse aluno deve ser considerado "atípico". Existe algum nessa situação?

```
## Sim, é o valor:
[1] -3.328794
## Correspondente à nota
[1] 4
```

(d) Interprete o significado de Z.

```
## Z é uma variável padronizada, que mede o número de desvios-padrões
## que cada observação se afasta da média.
```