1.7.3 Medidas de dispersão(134).

Amplitude(h) é a diferença entre o valor máximo e o valor mínimo da variável,
ou seja: **$h = x_{máx.} - x_{mín.}$**

Exemplo 20(134)

Atividade7(135)

Nota: Quando os dados estão agrupados em classes, a amplitude é a diferença entre o limite superior da última classe e o limite inferior da primeira classe, como veremos na atividade 8, que se segue.

Atividade 8(135)

Nota: A **amplitude** é muito **sensíve**l a valores discrepantes.

Amplitude interquartil (136).

Amplitude interquartil (**Aq**) é a diferença entre o 3º quartil e o 1º quartil, ou seja:

$$Aq = Q_3 - Q_1$$

Exemplo 21(136)

Atividade 9(136)

Variância e Desvio padrão (137).

(Populacional/amostral)

Exemplo1(extra): Variância e desvio padrão **populacional(σ)**.
Consideremos os dados de uma população: 4; 5; 6.

1º calculemos a média populacional:

$$\mu = (4+5+6)/3 = 5$$

Variância:

$$V = \frac{(4-\mathbf{5})^2 + (5-\mathbf{5})^2 + (6-\mathbf{5})^2}{3} = \frac{1+0+1}{3} = \frac{2}{3}$$

O desvio padrão é a raiz quadrada da variância:

$$\sigma = \sqrt{V}$$

Neste caso, $\sigma = \sqrt{\frac{2}{3}} \approx 0.816$

Exemplo2(extra): Variância e desvio padrão **amostral(S)**.
Consideremos os dados de uma amostra: 4; 5; 6.

1º calculemos a média amostral.

$\bar{x}$= (4+5+6)/3 = 5

Variância amostral:

$$S^2 = \frac{(4-\mathbf{5})^2 + (5-\mathbf{5})^2 + (6-\mathbf{5})^2}{3-1} = \frac{1+0+1}{2} = \frac{2}{2} = 1$$

O desvio padrão é a raiz quadrada da variância:

$$S = \sqrt{S^2}$$

Neste caso, $S = \sqrt{1} = 1$

Variância populacional (V)

$$\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}$$

Desvio padrão populacional é a raiz quadrada da variância. $\sigma = \sqrt{V}$

Ou

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \mu)^2}{n}}$$

Nota: No caso de uma amostra, dividimos por **(n-1)** em vez de n.

Variância amostral (S^2)

$$\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}$$

Desvio padrão amostral (S)

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1}}$$

Exemplo (extra) Desvio padrão amostral(**S**).

Consideremos os dados da amostra: 2; 3; 7.

1º calculemos a média da amostra:

$\bar{x}$=(2+3+7)/3 = 4

Variância:

$$S^2 = \frac{(2-\mathbf{4})^2+(3-\mathbf{4})^2+(7-\mathbf{4})^2}{3-1} = \frac{4+1+9}{2} = \frac{14}{2} = 7$$

O desvio padrão é a raiz quadrada da variância:

$$S = \sqrt{S^2}$$

Neste caso, $S = \sqrt{7} \approx 2.646$

Nota: se os **dados** forem discretos e estiverem **agrupados**, usamos o mesmo raciocínio, agrupando os respetivos valores:

$$s = \sqrt{\frac{\sum_{i=1}^{M}(x_i-\bar{x})^2 \cdot f_i}{n-1}}$$

ou

$$\sigma = \sqrt{\frac{\sum_{i=1}^{M}(x_i-\mu)^2 \cdot f_i}{n}}$$

Exemplo 22 (137). Temos uma **populaçã**o-desvio padrão populacional(σ).

(Resolver, indicando todos os cálculos.)

1º Calculemos o valor médio populacional:

μ=(3×2+5×1+7×4+8×8+9×4+10×1)/20 = 149/20= 7.45

Variância:

$$V = \frac{(3-7.45)^2 \times 2 + (5-7.45)^2 \times 1 + (7-7.45)^2 \times 4 + (8-7.45)^2 \times 8 + (9-7.45)^2 \times 4 + (10-7.45)^2 \times}{20}$$

$$\frac{39.605+6.0025+0.81+2.42+9.61+6.5025}{20} = \frac{64.95}{20} = 3.2475$$

Desvio padrão:

$$\sigma = \sqrt{V} = \sqrt{3.2475} \approx 1.8$$

Nota: se preferir, pode usar os cálculos intermédios com uma tabela, tal como é feito no final da página 137 do livro.

Nota: como veremos a seguir, também podemos calcular o desvio padrão com a calculadora gráfica, no entanto, por vezes será pedido explicitamente para calcular apresentando todos os cálculos, como acabámos de fazer.

Calculadora gráfica:

Pode calcular do mesmo modo que calcula a média. O símbolo do desvio padrão populacional é σx e o amostral é Sx.

Exercício(extra): Resolva o exemplo anterior utilizando a calculadora gráfica. Comece por introduzir as duas listas. Na calculadora Casio, não esqueça de ir ao "SET" confirmar que a segunda lista contém as frequências. Na Calculadora Texas, Não esqueça de indicar "L1 e L2".
Os procedimentos são os mesmos que utilizámos anteriormente para a média.

Exemplo 23 (138)- Desvio padrão populacional. Com a calculadora gráfica.
Usamos as marcas das classes.

Elementos que pertencem ao intervalo] $\mu-\sigma, \mu+\sigma$[.

Exercício 22(163). O número de animais de estimação dos alunos de uma turma encontra-se na tabela.

Número de animais	Número de alunos
0	6
1	5
2	7
3	3
4	2

22.1) Calcule a média e o desvio padrão populacional para esta distribuição(2 c.d.). Pode usar a calculadora gráfica.

22.2) Quantos alunos têm um número de animais pertencente ao intervalo

] $\mu-\sigma, \mu+\sigma$[?

Resposta:

22.1) Usamos a calculadora gráfica e lançamos a primeira coluna como lista de x e a segunda como valores das frequências. Obtemos, (com 2c.d.) os valores:

$\mu \approx 1,57$; $\sigma \approx 1,25$

22.2) Basta agora fazer

] $\mu-\sigma, \mu+\sigma$[=]1.57-1.25; 1.57+1.25[=]0.32; 2.82[

neste intervalo temos os valores “1” e “2”, que correspondem às frequências 5 e 7 respetivamente. Assim, 5+7=12
R: 12 alunos têm um número de animais pertencente ao intervalo $]\mu-\sigma, \mu+\sigma[$.

Nota: Não vamos estudar os exemplos indicados na página 139 referentes a dados agrupados em classes com intervalos, por estar fora do programa do 10º ano.

Atividade 10- População(pág. 140) ~~10.5~~

Nota1:

1) Tal como a média, o desvio padrão também é uma medida **pouco resistente** pois é influenciado por valores discrepantes. No que diz respeito a **valores discrepantes** e comparando as três medidas de dispersão estudadas (amplitude, amplitude interquartil e desvio padrão), podemos afirmar que a amplitude interquartil é a mais robusta(ou resistente) das três.

2) O desvio padrão toma sempre valores maiores ou iguais a zero.
3)Se todos os dados forem iguais, o desvio padrão é zero.
Sugestão: experimente com a CG para a população 5; 5; 5; 5; 5.

4) Para amostras com a mesma média, quanto **mais dispersos** forem os dados, **maio**r é o valor do **desvio padrão**. ***Sugestão:*** experimente com a CG para as populações: 9; 10; 11 e 5; 10; 15. Obtenha a média e o desvio padrão para cada uma. Comente.

Propriedades do desvio padrão.(Extra)
-Muito importante!

Exemplo(Extra) Média-Resolva
Considere a população com os dados: 10; 12; 14; 17; 22.

1) Calcule o desvio padrão populacional. (R: $\sigma\approx4.195$)

2) Adicione 3 unidades a cada um dos elementos e calcule novamente o desvio padrão populacional . Comente os resultados obtidos. (R: $\sigma\approx4.195$)

3) Adicione 5 unidades a cada um dos elementos e calcule novamente o desvio padrão populacional. Comente os resultados obtidos. (R: $\sigma\approx4.195$)

1.4) Multiplique por 2 cada um dos dados e calcule o desvio padrão populacional. Comente os resultados obtidos.(R: $\sigma\approx8.39=2\times4.195$)

1.5) Multiplique por 3 cada um dos dados e calcule o desvio padrão populacional. Comente os resultados obtidos.(R: $\sigma\approx12.586\approx3\times4.195$)

Propriedades do desvio padrão:

Propriedade1: Se adicionarmos a cada um dos valores uma constante k, o novo desvio padrão não se altera.

Propriedade2: Se multiplicarmos cada um dos valores por uma constante k, o novo desvio padrão será igual ao desvio padrão original, multiplicado por k.

1.8 Introdução gráfica à análise de dados bivariados (142)

Exemplo 1(142)

(pág. 143)
A este tipo de gráfico chamamos **gráfico de correlação** ou **diagrama de dispersão**.

O conjunto dos pontos num gráfico de correlação designa-se por **nuvem de pontos**.

Nota: no exemplo acima, existe correlação entre as variáveis.

Calculadora gráfica: pág. 228 ou 230.

TI-Nspire:

https://pedronoia.net/nspire.pdf

Numworks:

https://www.numworks.com/pt

professores/ tutoriais/ regressão.

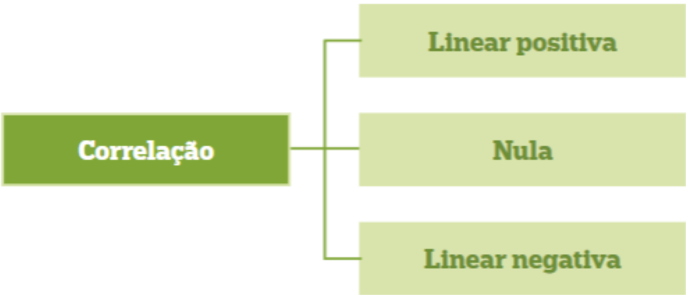
Exercício 24(164)

Atividade 1(143) (Solução: pág. 144)

A **correlação** diz-se **linear** se a nuvem de pontos se distribuir ao longo de uma linha reta, a reta de regressão.(145)

Correlação **linear positiva**- à medida que os valores de uma variável aumentam, os valores correspondentes da outra variável também aumentam.

Correlação **linear negativa**- à medida que os valores de uma variável aumentam, os valores correspondentes da outra variável diminuem.



Relação linear perfeita(145)

Exemplo 2(146)

Exemplo 3 (146)

1.9 Modelos de regressão linear
1.9.1 Reta de regressão (147).

Reta de regressão é a reta que melhor se ajusta aos pontos de um gráfico de correlação.(147)

Exemplo 1 (147)

Centro de gravidade da nuvem de pontos é o ponto cujas coordenadas são as médias das distribuições em análise; representa-se por $C\ (\mu_x, \mu_y)$

Calculadora gráfica: Casio pág. 228 texas 231

TI-Nspire:

https://pedronoia.net/nspire.pdf

Numworks:

https://www.numworks.com/pt

professores/ tutoriais/ regressão

1.9.2 Importância da reta de regressão(149).

Permite fazer previsões sobre uma das variáveis quando se conhecem os valores da outra.
Exemplo 2(149)

Atividade 1(149)

Nota importante: Na equação Y=ax+b, por vezes dão-nos o valor de y e pedem-nos o valor de x.

Exemplo extra(Muito importante):

Retomando a equação obtida na atividade anterior Y= 0.58x+8.08. Se nos disserem que um aluno obteve 16 valores e quisermos estimar o número de horas de estudo, fazemos y=16 e procuramos x:

16= 0.58x+8.08 ⇔ 16-8.08= 0.58x ⇔0.58x=7.92 ⇔ $x = \frac{7.92}{0.58}$ ⇔ $x \approx 13.655$

R: teria estudado aproximadamente 13.655 horas.

1.9.3 Limitações da reta de regressão(150).

Os pontos anómalos (outliers) podem influenciar o traçado da reta de regressão.

Exemplo 3(151).

1.10 Coeficiente de correlação (152).

O coeficiente de correlação linear mede o grau de associação linear entre duas variáveis. Representa-se por ***r*** e varia entre –1 e 1.

- $r = 1$, a correlação diz-se **total** (ou **perfeita**) **positiva**;
- $r = 0$, a correlação diz-se **nula**: não há correlação linear;
- $r = -1$, a correlação diz-se **total** (ou **perfeita**) **negativa**.

Para os valores intermédios, a correlação é tanto mais forte quanto mais próximo o valor de r se encontrar de 1 ou de –1, enfraquecendo à medida que se aproxima de zero.

Exemplo 1(152).

Exemplo 2(153)

<u>**Nota:**</u> o facto de o coeficiente de correlação ser elevado não significa que as duas variáveis tenham uma relação de causa-efeito.

Exemplo 3(153)

Calculadora gráfica: (228; 231)

TI-Nspire:

https://pedronoia.net/nspire.pdf

Numworks:

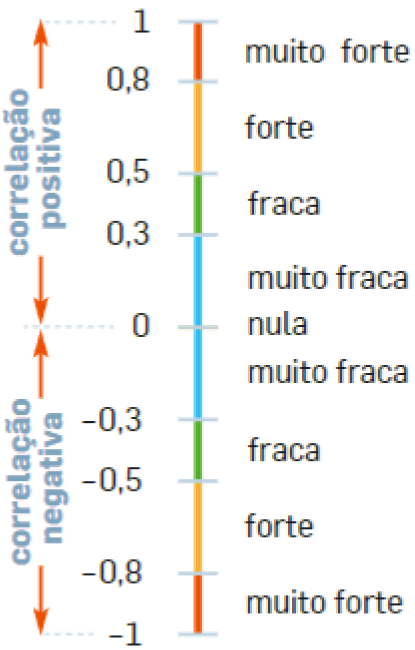
https://www.numworks.com/pt

professores/ tutoriais/ regressão.

Pág. 154

NOTA

Existem autores que definem uma escala para classificarem a correlação em função do valor do coeficiente de correlação, escala essa que não é única. Um exemplo pode ser o seguinte.



Atividade 1(1549

Atividade 2(154) ~~2.5~~

Erro no enunciado: em 2.1 é (x, y), (z, y) ~~(y,z)~~ e (z,x)

Exercício 27(164).

1.10.1 Limitações do coeficiente de correlação(155).

Nota: a existência de pontos outliers pode condicionar o valor do coeficiente de correlação.

Exemplo 4(155).

Nota: É importante uma investigação prévia dos pontos anómalos para um estudo mais rigoroso da correlação entre duas variáveis.

Atividade 3(156) ~~3.4~~

1.11 Tabelas de contingência(156)

Quando pelo menos uma das variáveis estatísticas em estudo é do tipo qualitativo, recorre-se à representação dos dados em **tabelas de contingência**.

Exemplo 1(156)

Exercício 28(164)

Atividade1(157)