Aula 04

ESTATÍSTICA APLICADA

Prof. Rita von Randow
rita.vonrandow@gmail.com

08/08/2024
a
15/08/2024

Agradecimento especial à professora Nanci de Oliveira por permitir que seu material seja utilizado no todo ou em partes durante este curso de Estatística.

MEDIDAS DE DISTRIBUIÇÃO DE FREQÜÊNCIAS

A- Medidas de Tendência Central

- Média
- Mediana
- Moda

B- Medidas de Dispersão ou Variabilidade

- Desvio Médio
- Variância
- Desvio Padrão
- Coeficiente de variação: medida de dispersão relativa

C- Medidas de Posição

- Quartil
- Decil
- Percentil

As medidas de posição central tem por objetivo resumir e representar um conjunto de dados. As medidas de posição central mais utilizadas são: **moda**, **mediana** e **média**.

A- Medidas de Tendência Central

MÉDIA ARITMÉTICA

- É a mais usual das medidas estatísticas.
- **É a soma dos valores de todos os dados dividido pelo número de dados.**

ATENÇÃO:
*Essas fórmulas somente serão usadas quando os dados **não estiverem agrupados em tabelas de distribuição de frequências**.*

μ = Média da população $\quad \bar{X} = média\ da\ amostra$

$$\mu = \frac{\sum_{i=1}^{n} x}{N} \quad \text{ou} \quad \bar{x} = \frac{\sum_{i=1}^{n} x}{n}$$

MEDIANA

- **É o dado que fica na posição central de um conjunto de dados que estão em ordem crescente ou decrescente.**

Qual valor está no centro dos dados ordenados?

Como encontrar a mediana?

Depois de ordenados os valores por ordem crescente ou decrescente, a **mediana** é:

- o valor que ocupa a posição central, se a quantidade desses valores for **ímpar**
- a média dos dois valores centrais, se a quantidade desses valores for **par**.

MODA

- É valor que mais se repete num conjunto de dados, ou seja, é o dado que ocorre com maior frequência.
- Se dois dados ocorrem com a mesma frequência elevada, cada dado é uma moda e os dados são chamados **bimodais**.
- Caso não haja dados repetidos, os dados são **amodais**.

Moda =
bola vermelha

Amodal

EXEMPLO 1:
Média, Moda e Mediana

Gastos em eletricidade:

Esta tabela não é de distribuição de frequências.

Meses	JAN.	FEV.	MAR.	ABR.	MAI.
Gasto (em €)	25€	22€	35€	28€	35€

- Média: 29 €

$$\bar{x} = \frac{25+22+35+28+35}{5} = \frac{145}{5} = 29 \Rightarrow \bar{x} = 29$$

Moda: 35 €

Mediana: 28 €

Rol: 22 25 **28** 35 35

Número impar de dados

Agora é com vocês! Calculem a média, mediana e moda nesse exemplo

EXEMPLO 2:
Média, Moda e Mediana

Gastos em eletricidade:

Meses	JAN.	FEV.	MAR.	ABR.	MAI.	JUN.
Gastos (em €)	25€	22€	35€	28€	35€	33€

- Média: 29,67 €

$$\bar{x} = \frac{25+22+35+28+35+33}{6} = \frac{178}{6} = 29,67 \Rightarrow \bar{x} = 29,67$$

Moda: 35 €

Mediana: 30,5 €

Rol: 22 25 **28** **33** 35 35

$$\frac{28+33}{2} = \frac{61}{2} = 30,5$$

Número par de dados

Média Ponderada

A média ponderada é utilizada quando cada valor de um conjunto de dados tem importâncias diferentes, isto é, aquelas com a maior importância (maior peso) devem influenciar mais na média. A média ponderada é:

$$\bar{x}_G = \sum_{i=1}^{k} x_i \times p_i$$

Onde p_i é o peso com as seguintes condições:

$$0 < p_i < 1 \quad e \quad \sum_{i=1}^{k} p_i = 1$$

MÉDIA DE DADOS AGRUPADOS

- A **média de uma distribuição de frequências** para uma amostra é dada por:

ATENÇÃO:
Esta fórmula é usada quando os dados estão agrupados em tabelas de distribuição de frequências.

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} \quad \text{onde} \quad n = \sum_{i=1}^{n} f_i$$

$$X_i = Pm_i = variável\ ou\ ponto\ médio\ da\ classe\ i$$

$$f_i = frequência\ da\ classe\ i$$

EXEMPLO 1
MÉDIA PARA DADOS AGRUPADOS

Use a distribuição de frequências abaixo para calcular a média do número de minutos que uma amostra de internautas gastou durante sua navegação mais recente na rede.

X_i	f_i
12,5	6
24,5	10
36,5	13
48,5	8
60,5	5
72,5	6
84,5	2
	Σf_i=50

Como os dados estão agrupados em uma distribuição de frequências, temos que usar a seguinte fórmula:

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} \quad \textit{onde} \quad n = \sum_{i=1}^{n} f_i$$

Solução

X_i	f_i	$X_i \cdot f_i$
12,5	6	75,0
24,5	10	245,0
36,5	13	474,5
48,5	8	388,0
60,5	5	302,5
72,5	6	435,0
84,5	2	169,0
	$\Sigma f_i = 50$	**$\Sigma X_i \cdot f_i = 2089,0$**

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{2089}{50} = 41,8$$

$$Logo,\quad \bar{x} = 41,8 \ minutos.$$

EXEMPLO 2
MÉDIA PARA DADOS AGRUPADOS

Use a distribuição de frequências abaixo para calcular a média dos estaturas, em cm, de 50 atletas.

Estaturas	f_i
150 ⊢ 155	2
155 ⊢ 160	10
160 ⊢ 165	12
165 ⊢ 170	15
170 ⊢ 175	5
175 ⊢ 180	6
	$\Sigma f_i = 50$

Como os dados estão agrupados em uma distribuição de frequências, temos que usar a seguinte fórmula, porém, inicialmente, temos que encontrar os ***pontos médios de cada classe X_i****:*

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} \quad \text{onde} \quad n = \sum_{i=1}^{n} f_i$$

Agora é com vocês!

EXEMPLO 2
MÉDIA PARA DADOS AGRUPADOS

Use a distribuição de frequências abaixo para calcular a média dos estaturas, em cm, de 50 atletas.

Estaturas	X_i	f_i
150 ├─ 155	152,5	2
155 ├─ 160	157,5	10
160 ├─ 165	162,5	12
165 ├─ 170	167,5	15
170 ├─ 175	172,5	5
175 ├─ 180	177,5	6
		$\Sigma f_i = 50$

Como os dados estão agrupados em uma distribuição de frequências, temos que usar a seguinte fórmula, porém, inicialmente, temos que encontrar os ***pontos médios de cada classe*** X_i *:*

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} \quad onde \quad n = \sum_{i=1}^{n} f_i$$

Solução

Estaturas	X_i	f_i	$X_i \cdot f_i$
150 ├─ 155	152,5	2	305
155 ├─ 160	157,5	10	1575
160 ├─ 165	162,5	12	1950
165 ├─ 170	167,5	15	2512,5
170 ├─ 175	172,5	5	862,5
175 ├─ 180	177,5	6	1065
		$\Sigma f_i = 50$	$\Sigma X_i \cdot f_i = 8270$

$$\bar{x} = \frac{\sum_{i=1}^{n} x_i f_i}{n} = \frac{8270}{50} = 165{,}4$$

Logo, $\bar{x} = 165{,}4 \; cm$.

MEDIDAS DE DISTRIBUIÇÃO DE FREQÜÊNCIAS

B- Medidas de Dispersão ou Variabilidade

- Amplitude Total
- Desvio Médio
- Variância
- Desvio Padrão
- Coeficiente de variação

MEDIDAS DE DISPERSÃO

- Numa pesquisa, os dados coletados podem:
  - assumir valores diferentes entre si
  - situar-se próximos ou não de uma medida de tendência central, como a média aritmética
  - situar-se próximos entre si ou não
  - ser discrepantes, em alguns casos

- Medidas de dispersão visam atribuir um valor numérico, da melhor forma possível, que expresse a homogeneidade ou não entre os diversos valores obtidos numa pesquisa.

AMPLITUDE TOTAL

- A **amplitude total** de um conjunto de dados é a diferença entre o valor máximo e o valor mínimo assumido pela variável.

$$\boxed{AT = (X_{máx} - X_{mín})\ do\ rol}$$

onde ***X*** é a variável.

DESVIO

- O **desvio** de um conjunto de dados é a diferença entre os valores da variável e a média do conjunto de dados, podendo ser para mais (+) ou para menos (-).

$$d_i = X_i - \bar{X}$$

EXEMPLO - DESVIO

A empresa A contratou dez pessoas com curso superior. O salário inicial nessa companhia é mostrado a seguir.

- Salário inicial na empresa A (em milhares de dólares)

Salário	37	38	39	41	41	41	42	44	45	47

- Determine o desvio de cada salário inicial para a empresa A.

Salário (milhares de dólares) X_i	Desvio (milhares de dólares) $X - \bar{X}_i$
37	-4,5
38	-3,5
39	-2,5
41	-0,5
41	-0,5
41	-0,5
42	0,5
44	2,5
45	3,5
47	5,5
$\Sigma X_i = 415$	$\Sigma(X_i - \bar{X}) = 0$

n=10

O salário médio inicial da empresa é

$$\bar{X} = \frac{\sum X_i}{n}$$

$$\bar{X} = \frac{415}{10} = 41,5$$

$$\bar{X} = 41,5$$

Observe que a soma dos desvios é zero. Isso ocorre para todo conjunto de dados. Se calcularmos a média dos desvios, obteremos sempre zero.

DESVIO MÉDIO

- Para resolver o problema anterior (média dos desvios igual a zero), calculamos o módulo dos desvios. Obtemos assim o **desvio médio**.

$$DM = \frac{\sum_{i=1}^{n} \left| x_i - \bar{x} \right|}{n}$$

EXEMPLO - DESVIO MÉDIO

Salário (milhares de dólares) X_i	Módulo do Desvio (milhares de dólares) $\lvert X - \bar{X}_i \rvert$
37	|-4,5| = 4,5
38	|-3,5| = 3,5
39	|-2,5| = 2,5
41	|-0,5| = 0,5
41	|-0,5| = 0,5
41	|-0,5| = 0,5
42	|0,5| = 0,5
44	|2,5| = 2,5
45	|3,5| = 3,5
47	|5,5| = 5,5
$\Sigma X_i = 415$	$\sum \lvert X_i - \bar{X} \rvert = 24$

O desvio médio do salário inicial da empresa A é:

$$DM = \frac{\sum_{i=1}^{n} \lvert x_i - \bar{x} \rvert}{n}$$

$$DM = \frac{24}{10} = 2{,}4 \text{ mil dólares}$$

VARIÂNCIA

- Dispensa o uso do MÓDULO.
- Usa os desvios ao quadrado.
- Em um conjunto de dados, a média dos quadrados dos desvios é chamada de **variância.**

VARIÂNCIA AMOSTRAL

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

- Para a **variância populacional** utilizamos:

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

DESVIO PADRÃO

- Resolve o problema dimensional da variância.
- **Desvio padrão** é a raiz quadrada da variância.

DESVIO PADRÃO AMOSTRAL

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

- Para o **desvio padrão populacional** utilizamos:

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

EXEMPLO
VARIÂNCIA E DESVIO PADRÃO

Obtenha a variância e o desvio padrão amostrais do salário inicial para a empresa A.

- Salário inicial na empresa A (em milhares de dólares)

Salário	37	38	39	41	41	41	42	44	45	47

Salário (milhares de dólares) X_i	Desvio ao quadrado (milhares de dólares) $(xi - \bar{x})^2$
37	$(37-41,5)^2 = (-4,5)^2 = 20,25$
38	$(38-41,5)^2 = (-3,5)^2 = 12,25$
39	$(39-41,5)^2 = (-2,5)^2 = 6,25$
41	$(41-41,5)^2 = (-0,5)^2 = 0,25$
41	$(41-41,5)^2 = (-0,5)^2 = 0,25$
41	$(41-41,5)^2 = (-0,5)^2 = 0,25$
42	$(42-41,5)^2 = (0,5)^2 = 0,25$
44	$(44-41,5)^2 = (2,5)^2 = 6,25$
45	$(45-41,5)^2 = (3,5)^2 = 12,25$
47	$(47-41,5)^2 = (5,5)^2 = 30,25$
$\Sigma X_i = 415$	$\Sigma(X_i - \bar{X})^2 = 88,5$

$$\bar{X} = \frac{\sum x_i}{n} = \frac{415}{10} = \mathbf{41,5}$$

Variância amostral:

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

$$s^2 = \frac{88,5}{10-1} = \frac{88,5}{9} = \mathbf{9,83}$$

s^2 = 9,83 mil dólares

Desvio padrão amostral:

$$s = \sqrt{9,83} \approx \mathbf{3,14}$$

s = 3,14 mil dólares

MEDIDA DE DISPERSÃO RELATIVA

- As medidas de dispersão que vimos anteriormente (amplitude total, desvio médio, variância e desvio padrão) somente são comparáveis quando se referem a mesma escala de medida e quando os grupos têm média não muito diferentes.
- **Para avaliar a variação dos dados de pesquisa através de um número índice (porcentagem), utilizamos o coeficiente de variação.**

COEFICIENTE DE VARIAÇÃO

- **Coeficiente de variação (C.V.)** é a razão entre o desvio padrão e a média, multiplicada por 100. Portanto, ele é dado **em porcentagem.**

$$C.V. = \frac{s}{\bar{\bar{X}}} \cdot 100\ (\%)$$

EXEMPLO
COEFICIENTE DE VARIAÇÃO

Compare a variação das idades dos 2 grupos seguintes.

1º grupo - Idade, em anos, de três crianças: 1, 3, 5

2º grupo- Idade, em anos, de três adultos: 53, 55, 57

Cálculo da média aritmética dos dois grupos:

1º grupo

$$\bar{X} = \frac{1+3+5}{3} = \frac{9}{3} = 3\ anos$$

2º grupo

$$\bar{X} = \frac{53+55+57}{3} = \frac{165}{3} = 55\ anos$$

Cálculo do desvio padrão (dispersão absoluta) dos dois grupos:

1º grupo

Xi	*$(X_i - \bar{X})^2$*
1	$(1 - 3)^2 = 4$
3	$(3 - 3)^2 = 0$
5	$(5 - 3)^2 = 4$
	$\Sigma(X - \bar{X})^2 = 8$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{8}{2} = 4$$

$$s = \sqrt{4} = 2\ \textbf{\textit{anos}}$$

2º grupo

Xi	*$(X_i - \bar{X})^2$*
53	$(53 - 55)^2 = 4$
55	$(55 - 55)^2 = 0$
57	$(57 - 55)^2 = 4$
	$\Sigma(X - \bar{X})^2 = 8$

$$s^2 = \frac{\sum_{i=1}^{n}(x_i - \bar{x})^2}{n-1} = \frac{8}{2} = 4$$

$$s = \sqrt{4} = 2\ \textbf{\textit{anos}}$$

A dispersão dos dados em torno da média é exatamente a mesma (s = 2 anos), porém, a diferença de 2 anos no 1º grupo é bastante significativo na mudança física, já no 2º grupo não o é.

Cálculo do coeficiente de variação (dispersão relativa) dos dois grupos:

1o grupo:

$$C.V. = \frac{s}{\overline{X}} \cdot 100\ (\%)$$

$$\boldsymbol{C.V. = \frac{2}{3} \cdot 100 = 66{,}7\%}$$

2o grupo:

$$C.V. = \frac{s}{\overline{X}} \cdot 100\ (\%)$$

$$\boldsymbol{C.V. = \frac{2}{55} \cdot 100 = 3{,}64\%}$$

Como o coeficiente de variação foi maior no 1º grupo, concluímos que o 1º grupo teve maior variação nos seus dados.

DESVIO PADRÃO PARA DADOS AGRUPADOS

- A fórmula para o desvio padrão da amostra de uma distribuição de frequência é:

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

onde ***n* = Σ*f*$_i$** é o número total de dados.

EXEMPLO 1- DESVIO PADRÃO PARA DADOS AGRUPADOS EM DISTRIBUIÇÃO DE FREQUÊNCIAS

Os resultados de uma amostra do número de crianças por família em uma região estão dispostos na tabela abaixo. Determine a média e o desvio padrão.

número de crianças por família

x_i	f_i
0	10
1	19
2	7
3	7
4	2
5	1
6	4
	$\Sigma f_i = 50$

número de famílias

X_i	f_i	$X_i \cdot f_i$	$(X_i - \bar{X})^2 \cdot f_i$
0	10	0	$(0 - 1{,}8)^2 \cdot 10 = (3{,}24) \cdot 10 = 32{,}4$
1	19	19	$(1 - 1{,}8)^2 \cdot 19 = (0{,}64) \cdot 19 = 12{,}16$
2	7	14	$(2 - 1{,}8)^2 \cdot 7 = (0{,}04) \cdot 7 = 0{,}28$
3	7	21	$(3 - 1{,}8)^2 \cdot 7 = (1{,}44) \cdot 7 = 10{,}08$
4	2	8	$(4 - 1{,}8)^2 \cdot 2 = (4{,}84) \cdot 2 = 9{,}68$
5	1	5	$(5 - 1{,}8)^2 \cdot 1 = (10{,}24) \cdot 1 = 10{,}24$
6	4	24	$(6 - 1{,}8)^2 \cdot 4 = (17{,}64) \cdot 4 = 70{,}56$
	$\Sigma f_i = 50$	$\Sigma X_i \cdot f_i = 91$	$\Sigma (X_i - \bar{X})^2 f_i = 145{,}40$

Média: $\bar{X} = \frac{\sum X_i \cdot f_i}{\sum f_i} = \frac{91}{50} = \mathbf{1{,}8}\ \boldsymbol{crianças}$

Desvio padrão: $s = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i}{n - 1}} = \sqrt{\dfrac{145{,}4}{49}} = \sqrt{2{,}967} = \boldsymbol{1{,}7\ crianças}$

Agora é com vocês! Calculem o desvio padrão para o exemplo abaixo

EXEMPLO 2- DESVIO PADRÃO PARA DADOS AGRUPADOS EM DISTRIBUIÇÃO DE FREQUÊNCIAS

O resultado de uma sondagem na qual mil adultos foram indagados sobre quanto gastavam anualmente na preparação de uma viagem de férias resultou na distribuição de frequência abaixo. Estime a média e o desvio padrão amostrais do conjunto de dados.

Gastos (US$)	f_i
0 ├─ 100	380
100 ├─ 200	230
200 ├─ 300	210
300 ├─ 400	50
400 ├─ 500	60
500 ├─ 600	70
	$\Sigma f_i = 1000$

Gastos (US$)	f_i	X_i	$X_i \cdot f_i$	$(xi - \bar{x})^2 f_i$
0 |— 100	380	**50**	19.000	$(50 - 189)^2 \cdot 380 = 7.341.980$
100 |— 200	230	**150**	34.500	$(150 - 189)^2 \cdot 230 = 349.830$
200 |— 300	210	**250**	52.500	$(250 - 189)^2 \cdot 210 = 781.410$
300 |— 400	50	**350**	17.500	$(350 - 189)^2 \cdot 50 = 1.296.050$
400 |— 500	60	**450**	27.000	$(450 - 189)^2 \cdot 60 = 4.087.260$
500 |— 600	70	**550**	38.500	$(550 - 189)^2 \cdot 70 = 9.122.470$
	$\Sigma f_i = 1000$		$\Sigma X_i \cdot f_i = 189.000$	$\Sigma(x_i - \bar{x})^2 fi = 22.979.000$

Média: $\bar{X} = \frac{\sum x_i \cdot f_i}{\sum f_i} = \frac{189000}{1000} = \mathbf{189}\ dólares$

Desvio padrão:

$$s = \sqrt{\frac{\sum_{i=1}^{n}(x_i - \bar{x})^2 \cdot f_i}{n - 1}} = \sqrt{\frac{22979000}{999}} = \sqrt{23002,002} = \mathbf{151,66\ dólares}$$

FÓRMULAS MEDIDAS DE DISPERSÃO

- As fórmulas seguintes são usadas para dados agrupados em distribuição de frequências.

Variância Amostral	$s^2 = \dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i}{n-1}$
Desvio Padrão Amostral	$s = \sqrt{\dfrac{\sum_{i=1}^{n} (x_i - \bar{x})^2 \cdot f_i}{n-1}}$

MEDIDAS DE POSIÇÃO

C- Medidas de Posição

- Quartil
- Decil
- Percentil

QUARTIS

- **Quartis** são números que dividem aproximadamente um conjunto ordenado de dados em **quatro partes iguais**.
- Portanto, há 3 quartis: $\mathbf{Q_1}$, $\mathbf{Q_2}$ e $\mathbf{Q_3}$.
- Cerca de ¼ dos dados é menor ou igual ao 1º quartil.
- Cerca de ½ (metade) dos dados é menor ou igual ao 2º quartil. O 2º quartil é igual à mediana do conjunto de dados.
- Cerca de ¾ dos dados é menor ou igual ao 3º quartil.

QUARTIS

Ou seja, cada parte deve ter 25% dos dados.

→→ Primeiro quartil é o valor onde 25% dos dados são menores ou iguais a ele, consequentemente 75% são maiores ou iguais a ele;

→→ Segundo quartil (mediana) é o valor onde 50% dos dados são menores ou iguais a ele e, 50% são maiores ou iguais a a ele;

→→ Terceiro quartil é é o valor onde 75% dos dados são menores ou iguais a ele, 25% são maiores ou iguais a ele.

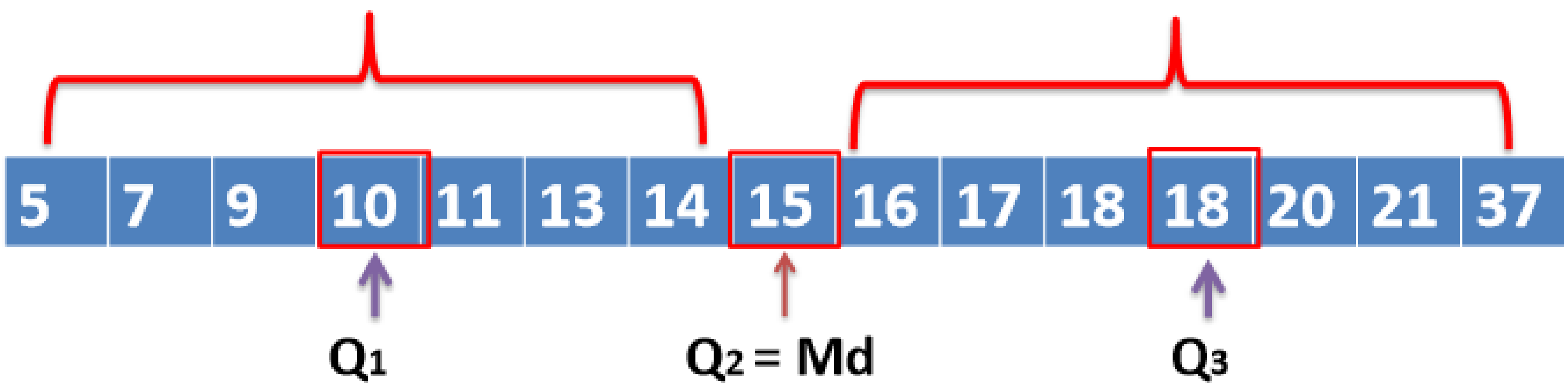
O segundo quartil coincide com a mediana, portanto, é pouco utilizado.

EXEMPLO - QUARTIS

A pontuação nos testes de 15 empregados envolvidos em um curso de treinamento está disposta a seguir. Obtenha os 3 quartis da pontuação dos testes.

13	9	18	15	14	21	7	10	11	20	5	18	37	16	17

Em 1° lugar, ordene o conjunto de dados e obtenha a mediana, que é igual ao 2° quartil.
Uma vez obtido o 2° quartil, pode-se dividir o conjunto de dados em 2 metades. O 1° e o 3° quartis são as medianas das metades inferior e superior do conjunto de dados.



Assim, cerca de ¼ dos empregados fez dez pontos ou menos, cerca da metade fez 15 pontos ou menos e cerca de ¾ conseguiu 18 pontos ou menos.

QUARTIS

No cálculo do quartil, deve-se primeiramente calcular a posição onde ele se encontra e, depois calculá-lo através de uma média ponderada de dois valores, caso seja necessário. O quartil pode ser obtida através de:

$$Q_i = x_{\left(\frac{i \times n}{4} + \frac{1}{2}\right)}$$

Com i = 1, 2 e 3 para se obter Q_1, Q_2 e Q_3 para um conjunto com n dados. O uso da média ponderada de dois valores é aplicado quando o resultado da posição $\left(\frac{i \times n}{4} + \frac{1}{2}\right)$ não é inteiro.

EXEMPLO 1

Para um conjunto com n = 10:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
9	10	12	14	16	17	17	19	20	22

O primeiro quartil e o terceiro quartil são:

$$Q_1 = x_{\left(\frac{1\times 10}{4}+\frac{1}{2}\right)} = x_{(3)} = 12$$

e

$$Q_3 = x_{\left(\frac{3\times 10}{4}+\frac{1}{2}\right)} = x_{(8)} = 19$$

Como as posições do primeiro quartil e terceiro quartil deram números inteiros, os quartis são associados à posição no conjunto ordenado de dados.

EXEMPLO 2

Para um conjunto com n = 11 tem-se:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}
9	10	12	14	16	17	17	19	20	22	24

O primeiro quartil é:

$$Q_1 = X_{\left(\frac{1 \times 11}{4} + \frac{1}{2}\right)} = X_{(3,25)}$$

Nesse caso, a “posição” 3,25 (número não inteiro) indica que o primeiro quartil está entre o terceiro e o quarto valor e, as casas decimais, 0,25, são interpretadas como proximidade maior ao terceiro do que o quarto elemento. As casas decimais também serão utilizadas como peso no cálculo do Q_1 e, quanto maior a proximidade maior será o peso para o cálculo de Q_1. A média ponderada entre x_3 e x_4 tem pesos iguais a (1-0,25) para x_3 e (0,25) para o x_4, respectivamente.

O procedimento para a obtenção dos quartis é inicialmente calcular:

$$Q_i = x_{\left(\frac{i \times n}{4} + \frac{1}{2}\right)} = x_{(j,d)}$$

Representando o resultado por $x_{(j,d)}$ com j o número inteiro e d a casa decimal da expressão, o quartil é dado por:

$$Q_i = x_{\left(\frac{i \times n}{4} + \frac{1}{2}\right)} = x_{(j,d)} = (1-d) \times x_{(j)} + (d) \times x_{(j+1)}$$

Para i =1, 2 e 3.

Assim, o primeiro quartil será:

$$Q_1 = x_{\left(\frac{1\times 11}{4}+\frac{1}{2}\right)} = x_{(3,25)}$$

Nesse caso, o primeiro quartil está entre o terceiro e o quarto valor e a casa decimal (0,25), indica que está mais próximo do terceiro elemento.
O primeiro quartil é:

$$Q_1 = x_{\left(\frac{1\times 11}{4}+\frac{1}{2}\right)} = x_{(3,25)} = (1 - 0{,}25) \times x_3 + (0{,}25) \times x_4$$
$$= (1 - 0{,}25) \times 12 + (0{,}25) \times 14 = 12{,}5$$

Assim, o terceiro quartil será:

$$Q_3 = x_{\left(\frac{3\times 11}{4}+\frac{1}{2}\right)} = x_{(8,75)}$$

Nesse caso, o terceiro quartil está entre o oitavo e o nono valor e a casa decimal (0,75), indica que está mais próximo do nono elemento.

O terceiro quartil é:

$$\begin{aligned} Q_3 = x_{\left(\frac{3\times 11}{4}+\frac{1}{2}\right)} = x_{(8,75)} &= (1-0,75)\times x_8 + (0,75)\times x_9 \\ &= (1-0,75)\times 19 + (0,75)\times 20 \\ &= 19,75 \end{aligned}$$

Exercício:

Calcule o primeiro, segundo e terceiro quartis do seguinte conjunto:

17
20
22
27
9
12
14
17
19
24
10
16

$$Q_i = x_{\left(\frac{i \times n}{4} + \frac{1}{2}\right)}$$

Resposta: Q1 = 13; Q2 = 17 e Q3 = 21

Exercício:

Calcule o primeiro, segundo e terceiro quartis do seguinte conjunto:

17
20
22
27
9
12
14
17
19
24
10
16

$$Q_i = X_{\left(\frac{i \times n}{4} + \frac{1}{2}\right)} = X_{(j,d)}$$

$$Q_i = X_{\left(\frac{i \times n}{4} + \frac{1}{2}\right)} = X_{(j,d)} = (1-d) \times X_{(j)} + (d) \times X_{(j+1)}$$

Para i =1, 2 e 3.

9 10 12 | 14 16 17 | 17 19 20 | 22 24 27

Q1=13 Q2=17 Q3=21

Resposta: Q1 = 13; Q2 = 17 e Q3 = 21

DECIS E PERCENTIS

- **DECIS** dividem o conjunto de dados em 10 partes iguais. São 9 decis: $\mathbf{D_1}$, $\mathbf{D_2}$, $\mathbf{D_3}$, ... $\mathbf{D_9}$
- **PERCENTIS** dividem o conjunto de dados em 100 partes iguais. São 99 percentis: $\mathbf{P_1}$, $\mathbf{P_2}$, $\mathbf{P_3}$, ... $\mathbf{P_{99}}$

- Os **percentis** são frequentemente usados na Educação e nos campos relacionados à Saúde para indicar como um indivíduo se compara com os outros em um determinado grupo. Pontuações em testes e medidas de crescimento infantil, por exemplo, são frequentemente expressos em percentis.

Desta forma, os **Decis** são calculados da seguinte forma:

$$D_i = x_{\left(\frac{i \times n}{10} + \frac{1}{2}\right)}$$

Com *i* variando de 1 até 9.

O cálculo é feito como o dos quartis e o resultado pode ser representado por $x_{(j,d)}$, com j a parte inteira e d a parte das casas decimais.

O cálculo dos decis é dado por:

$$D_i = x_{\left(\frac{i \times n}{10} + \frac{1}{2}\right)} = x_{(j,d)} = (1-d) \times x_{(j)} + (d) \times x_{(j+1)}$$

para i =1, 2, ... , 9.

Agora com vocês!

Calcule **o primeiro decil, o segundo decil e o nono decil** do conjunto ordenado de 20 números dado a seguir:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}	x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
9	10	12	14	16	17	17	19	20	22	24	27	30	31	32	34	35	37	38	39

$$D_i = x_{\left(\frac{i \times n}{10} + \frac{1}{2}\right)} = x_{(j,d)} = (1-d) \times x_{(j)} + (d) \times x_{(j+1)}$$

para i =1, 2, ... , 9.

$$D_1 = x_{\left(\frac{1\times 20}{10}+\frac{1}{2}\right)} = x_{(2,5)} = (1-0,5)\times x_{(2)} + (0,5)\times x_{(3)}$$

$$= (1-0,5)\times 10 + (0,5)\times 12$$

$$= 11$$

Interpretação: 10% dos dados são menores ou iguais a 11 ou 90% são maiores ou iguais 11.

$$D_2 = x_{\left(\frac{2\times 20}{10}+\frac{1}{2}\right)} = x_{(4,5)} = (1-0,5)\times x_{(4)} + (0,5)\times x_{(5)}$$

$$= (1-0,5)\times 14 + (0,5)\times 16$$

Interpretação: 20% dos dados são menores ou iguais a 15 ou 80% são maiores ou iguais a 15.

$$D_9 = X_{\left(\frac{9\times 20}{10}+\frac{1}{2}\right)} = X_{(18,5)} = (1-0,5)\times X_{(18)} + (0,5)\times X_{(19)}$$

$$= (1-0,5)\times 3\ 7 + (0,5)x38$$

$$= 37,5$$

Interpretação: 90% dos dados são menores ou iguais a 37,5 ou 10% são maiores ou iguais a 37,5.

Como são calculados os **Percentis?**

$$P_i = x_{\left(\frac{i \times n}{100} + \frac{1}{2}\right)}$$

Com *i* variando de 1 até 99 para obter P_1, P_2, ..., P_{99}.

Exemplo: Calcule o P_5, P_{95} da série ordenada a seguir:

x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8	x_9	x_{10}
9	10	12	14	16	17	17	19	20	22

x_{11}	x_{12}	x_{13}	x_{14}	x_{15}	x_{16}	x_{17}	x_{18}	x_{19}	x_{20}
24	27	30	31	32	34	35	37	38	39

Resposta: P5 = 9,5 e P95 = 38,5

APLICAÇÃO

Claudia é gerente de Marketing da empresa **Calça & Moda.** Ela gostaria de saber se a promoção de desconto na compra da segunda peça aumentou o gasto por cliente. Para isso, ela analisou no banco de dados o gasto de 20 clientes antes da promoção e, o gasto de 16 clientes durante a promoção. Os gastos observados são:

GASTOS ANTES DA PROMOÇÃO
(em reais)

141	117	129	115
113	69	133	96
86	168	63	151
104	104	59	115
90	138	131	84

GASTOS DURANTE A PROMOÇÃO
(em reais)

108	88	110	79
112	90	123	97
125	130	102	140
158	108	129	119

A gerente gostaria de saber o valor da compra dos 10% e 25% dos menores gastos e também dos 5% e 25% dos maiores gastos por compra.

Outra informação é a média dos gastos e a mediana dos gastos antes e durante a promoção.

Em termos de Quartis, Decis e Percentis. O que vocês deverão calcular?

As medidas de ordenamento desejadas pela Claudia são:
P10 = D1 : o valor dos 10% que menos gastam por compra;
P25 = Q1 : o valor dos 25% que menos gastam por compra;
P95 : o valor dos 5% que mais gastam por compra;
P75= Q3 : o valor dos 25% que mais gastam por compra.

Além da média e mediana de gastos antes e durante a promoção.

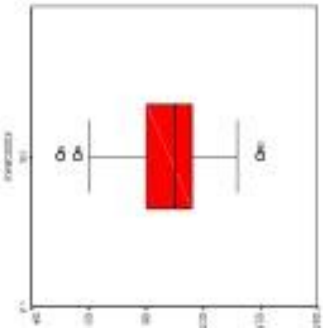
Os resultados obtidos foram:

	ANTES DA PROMOÇÃO	DURANTE A PROMOÇÃO
D_1	66	88,2
Q_1	88	99,5
Md	114	111
Q_3	132	127
P_{95}	159,5	152,6
média	110,3	113,63

APLICAÇÃO DOS QUARTIS: BOX PLOT

- Uma aplicação importante dos quartis é o **box plot**, também conhecidos como **plotes maria-chiquinha**, que é uma maneira de representar os dados.

- ***BOX PLOT*** é um **gráfico de dispersão** que relaciona os valores de uma variável com medidas estatísticas.



RESUMO CINCO-NÚMEROS DOS DADOS

- Para fazer um box plot é preciso conhecer os seguintes valores:

- Esses 5 valores são chamados de **resumo cinco-números** do conjunto de dados.

COMO FAZER UM BOX PLOT ?

1. Obtenha o resumo cinco-números do conjunto de dados

2. Construa uma escala horizontal que abranja a amplitude total dos dados.

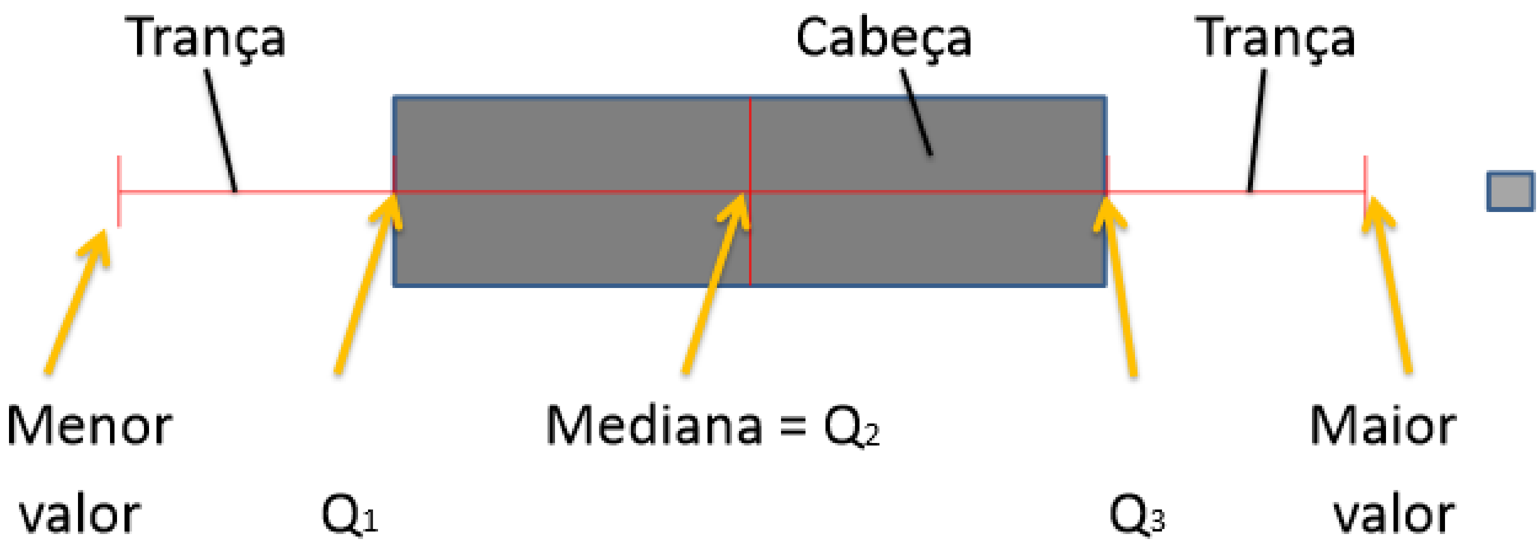
3. Plote os cinco números acima da escala horizontal.

4. Faça uma caixa acima da escala horizontal de Q1 a Q3 e trace uma reta vertical na caixa, passando por Q2.

5. Faça as “tranças” a partir da cabeça para o menor valor e o maior valor.

MODELO DE BOX PLOT

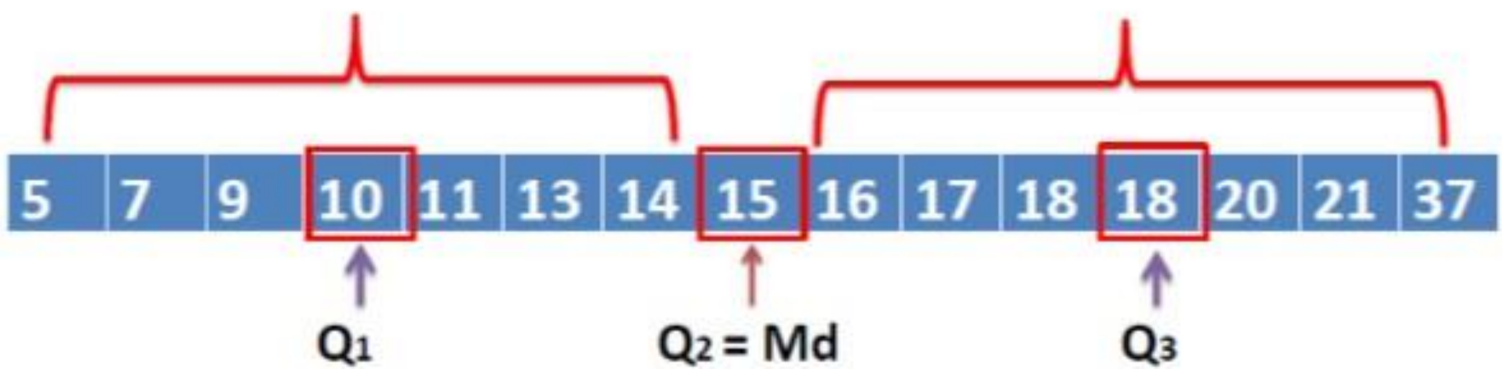
- Círculos, bem como quadradinhos indicam valores em uma pesquisa pelos quais diferem de grandeza dos demais, isto é, são dados distorcidos.

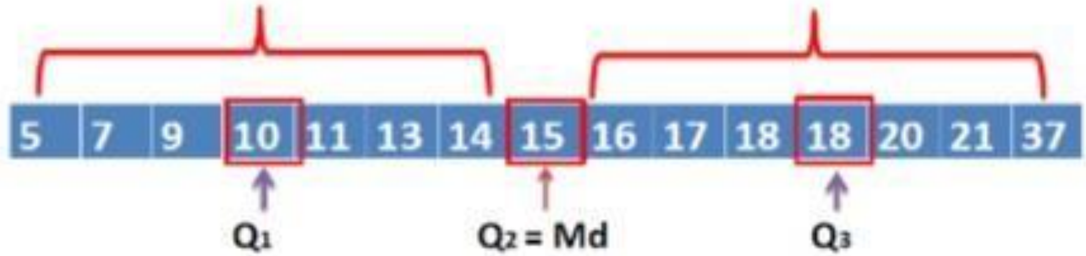


EXEMPLO - BOX PLOT

Faça um Box Plot que represente a pontuação dos 15 testes dados... (exemplo dos quartis).

O que você pode concluir a partir do gráfico?

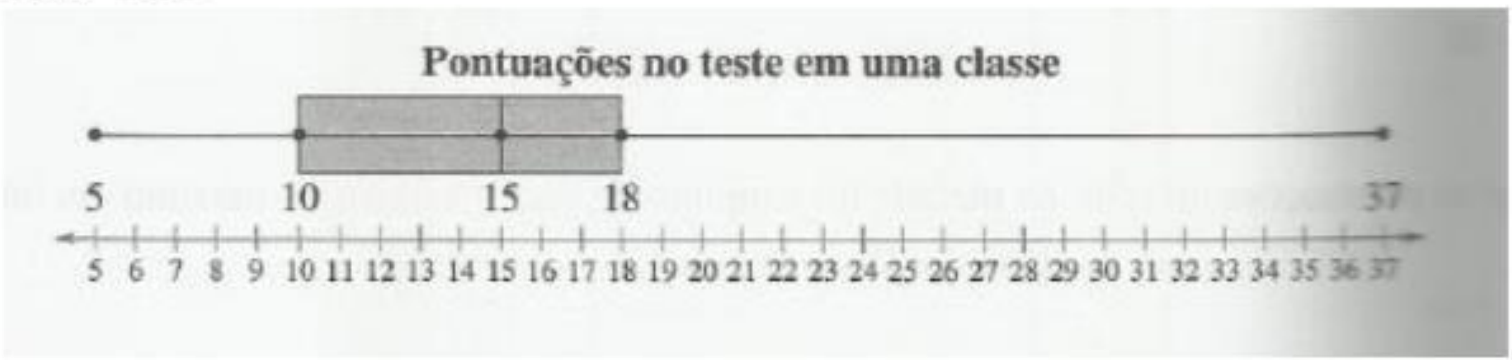




O **resumo cinco-números** das pontuações no teste são:

Menor valor = 5 $Q_1 = 10$ $Q_2 = 15$ $Q_3 = 18$ Maior valor = 37

BOX PLOT

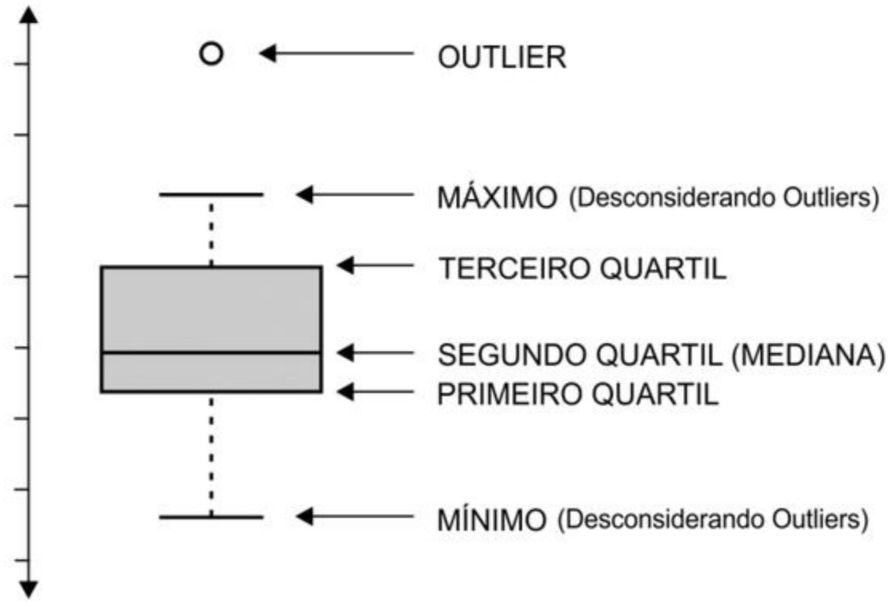


Uma das conclusões: cerca de metade das pontuações está entre 10 e 18.

Se eu tenho outliers no Boxplot

O limite de detecção de outliers é construído utilizando o intervalo interquartílico, dado pela distância entre o primeiro e o terceiro quartil. Sendo assim, os limites inferior e superior de detecção de outlier são dados por:

- Limite Inferior = Primeiro Quartil – 1,5 * (Terceiro Quartil – Primeiro Quartil)
- Limite Superior = Terceiro Quartil + 1,5 * (Terceiro Quartil – Primeiro Quartil)



Dados Agrupados em Intervalos

Quando os dados já estiverem agrupados em intervalos (na forma de tabela de frequência) e os dados brutos não são disponíveis para o cálculo exato das medidas de ordenamento, devemos obter essas medidas através de aproximações. Adota-se um procedimento para o cálculo do percentil que valerá para o cálculo do quartil e decil, pois essas medidas são casos particulares dos percentil.

Supondo o cálculo de Pi, com i = 1, 2, 3, ..., 99. O primeiro passo é calcular:

$$\left(\frac{i \times n}{100}\right)$$

que identifica a posição onde se encontra o percentil em questão e, depois calcular as frequências acumuladas.

O primeiro valor da frequência acumulada que ultrapassar ou igualar ao $\left(\frac{i \times n}{100}\right)$ é o intervalo que contém o Pi e, depois utilizar a fórmula de Pi dado a seguir. Essa expressão é obtida por relação de proporcionalidade (regra de três).

$$P_i = \mathit{LimInf} + h \times \frac{\frac{i \times n}{100} - F_{\text{acumulada anterior}}}{f_{\text{classe do i}}}$$

Onde:

- **LimInf:** é o limite inferior do intervalo que contém o percentil i;
- **h:** é o tamanho do intervalo que contém o percentil i;
- **n:** é o tamanho da amostra;
- $\mathbf{F_{acumulada\ anterior}}$**:** é a frequência acumulada anterior ao intervalo que contém o percentil i;
- $\mathbf{f_{classe\ do\ i}}$ **:** é a frequência do intervalo que contém o percentil i.

Exemplo:

Utilizando a tabela de frequência a seguir calcule o P_{10} como exemplo.

CLASSE				
LIMITE INFERIOR		LIMITE SUPERIOR	**FREQUÊNCIA**	**FREQUÊNCIA ACUMULADA**
10	⊢	15	5	5
15	⊢	20	12	17
20	⊢	25	8	25
25	⊢	30	3	28
		total (n)	28	

Para achar a posição do P_{10}, 1º calculo $\left(\frac{i \times n}{100}\right)$

$$i = 10 e\ n = 28 \Rightarrow \left(\frac{10 \times 28}{100}\right) = 2{,}8$$

$$\therefore P_{10}\ está\ na\ 1^{a} classe, de\ frequência\ acumulada\ de\ 0\ a\ 5$$

Exemplo:

Utilizando a tabela de frequência a seguir calcule o P_{10} como exemplo.

CLASSE			
LIMITE INFERIOR		LIMITE SUPERIOR	**FREQUÊNCIA**
10	⊢	15	5
15	⊢	20	12
20	⊢	25	8
25	⊢	30	3
		total (n)	28

LimInf = 10

CLASSE				**FREQUÊNCIA ACUMULADA**
LIMITE INFERIOR		LIMITE SUPERIOR	FREQUÊNCIA	
10	⊢	15	5	5
15	⊢	20	12	17
20	⊢	25	8	25
25	⊢	30	3	28
		total (n)	28	

$f_{\text{acumulada anterior}} = 0$

intervalo do P_{10}

h = 15 - 10 = 5

$f_{\text{10 percentil}} = 5$

Nesse caso, o P_{10} será:

$$P_i = LimInf + h \times \frac{\frac{i \times n}{100} - F_{\text{acumulada anterior}}}{f_{\text{classe do i}}}$$

$$= 10 + 5 \times \frac{\frac{10 \times 28}{100} - 0}{5}$$

$$= 12{,}8$$

Obs.: Se P_i está contido no primeiro intervalo a frequência acumulada anterior será igual a zero.

Exercício:

Em um escritório, o tempo por dia (em minutos), no horário de trabalho, que a Internet na área que João trabalha fica fora do ar em virtude de uma falha da prestadora de serviços, está apresentado na distribuição seguir.

Tempo parado em minutos	0 – 10	10 – 20	20 – 30	30 -- 40	40 – 50	Total
Fi	2	15	17	13	3	50

a) Calcule a média, moda e mediana.

b) Calcule o Q3, D3, P63.

Referências

SANTIAGO, M. S.; AKAMINE, C. T.; MORSELLI, N. V. Estatítica Aplicada à Gestão. UNIVESP. São Paulo, 2016.