ESTATÍSTICA DESCRITIVA

Média aritmética

A média é a medida mais utilizada pare resumir os dados e pode ser considerada uma medida justa. Uma definição, informal, para média é o valor pelo qual todos os demais podem ser substituídos de maneira que todos sejam iguais.

A média dos dados é calculada fazendo:

$$\mu = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$X = \{13, 18, 16, 12, 20, 12\}$$

Para os dados de exemplo:

$$\mu = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{13 + 18 + 16 + 12 + 20 + 12}{6} = \frac{91}{6} = 15.16667$$

ESTATÍSTICA DESCRITIVA

Média ponderada

Quando os dados são agrupados em k categorias o cálculo é feito por meio da média ponderada pela frequência (f_i repetições dos números):

$$\mu = \frac{\sum_{i=1}^{k} f_i \times x_i}{\sum_{i=1}^{k} f_i}$$

sendo $\sum_{i=1}^{k} f_i = n$

Considerando os dados $X = \{13, 18, 16, 12, 20, 12\}$, temos que:

x_i	f_i
12	2
13	1
16	1
18	1
20	1

$$\mu = \frac{12 \times 2 + 13 \times 1 + 16 \times 1 + 18 \times 1 + 20 \times 1}{2 + 1 + 1 + 1} = 15.16667$$

ESTATÍSTICA DESCRITIVA

Média geométrica

A media geométrica e utilizada para estudos sobre juros compostos, no contexto de dados financeiros e econômicos.

Considere $n = 6$ dados:

$$X = \{1.13, 1.18, 1.16, 1.12, 1.20, 1.12\}$$

A média geométrica dos dados é calculada fazendo:

$$\mu_{geo} = \sqrt[n]{\prod_{i=1}^{n} x_i}$$

O termo $\prod_{i=1}^{n} x_i$ indica o produtório $X_1 \cdot X_2 \cdot X_3 .... \cdot X_n$.

Para os dados de exemplo:

$$\mu_{geo} = \sqrt[n]{\prod_{i=1}^{n} X_i} = \sqrt[6]{1.13 \cdot 1.18 \cdot 1.16 \cdot 1.12 \cdot 1.20 \cdot 1.12} = \sqrt[6]{2.3282} = 1.1512$$

ESTATÍSTICA DESCRITIVA

Média geométrica

Lembrete: Para operacionalizar $\sqrt[n]{x}$ na calculadora lembre-se que:

$$\sqrt[a]{x^b} = x^{\frac{b}{a}}$$

Nesse caso, temos:

$$\sqrt[6]{x} = \sqrt[6]{x^1} = x^{\frac{1}{6}}$$

O exemplo mais simples e realizar a raiz quadrada de um número:

$$\sqrt[2]{10} = \sqrt[2]{10^1} = 10^{\frac{1}{2}} = 10^{0,5} = 3,1622$$

ESTATÍSTICA DESCRITIVA

Mediana

É a medida que indica o valor que está localizado no centro dos dados organizados de forma ordenada. Considere os valores $\{X_1, X_2, X_3, ..., X_n\}$. A representação desses valores de forma ordenada é feita com o uso de colchete no número de índice de cada valor $X_{[i]}$. Isto é, o valor $X_{[i]}$ indica o valor que está na i-ésima posição dos dados **ORDENADOS**.

Além disso, para determinar a mediana em dados desasgrupados é preciso saber se o número de valores n é ímpar ou par.

Caso seja ímpar: $\tilde{X} = X_{\left[\frac{n+1}{2}\right]}$ Caso seja par: $\tilde{X} = \frac{X_{\left[\frac{n}{2}\right]} + X_{\left[\frac{n}{2}+1\right]}}{2}$

ESTATÍSTICA DESCRITIVA

Mediana

Considere $n = 6$ dados coletados:

$$X = \{13, 18, 16, 12, 20, 12\}$$

Para calcular a mediana desses dados utilizaremos a fórmula para o número par de elementos. Além disso, os dados ordenados são:

$$X = \{12, 12, 13, 16, 18, 20\}$$

$$\tilde{X} = \frac{X_{\left[\frac{6}{2}\right]} + X_{\left[\frac{6}{2}+1\right]}}{2} = \frac{X_{[3]} + X_{[4]}}{2} = \frac{13+16}{2} = 14.5$$

ESTATÍSTICA DESCRITIVA

Moda

É o número com maior frequência no conjunto de dados, isto é, é o número que mais aparece. A moda pode não ser única, caso dois ou mais número tenham frequência igual. No caso de duas, o conjunto de dados é considerado bimodal. Caso todos os números apareçam o mesmo número de vezes, não há moda, sendo considerado amodal.

Considerando os dados $X = \{13, 18, 16, 12, 20, 12\}$ é fácil perceber que a moda dos dados é 12.

ESTATÍSTICA DESCRITIVA

Sensibilidade

A média, mediana e moda se comportam de maneira distinta à presença de valores extremos, isto é, valores muito grandes ou muito pequenos, que fogem do padrão observado dos dados.

Considere adicionar o valor de 1.000.000 aos valores utilizados anteriormente .Qual é a nova média? Qual a nova mediana? Qual a nova moda?

$$X = \{13,18,16,12,20,12,1000000\}.$$

ESTATÍSTICA DESCRITIVA

Sensibilidade

X = {13,18,16,12,20,12,1000000}.

$$\mu = \frac{\sum_{i=1}^{6} x_i}{6} = \frac{13 + 18 + 16 + 12 + 20 + 12 + 1.000.000}{7} = \frac{1.000.091}{7} = 142.870,1$$

A nova mediana é calculada utilizando a fórmula do número ímpar de valores:

$$\tilde{X} = X_{\left[\frac{n+1}{2}\right]} = X_{\left[\frac{7+1}{2}\right]} = X_{[4]} = 16$$

A moda não é afetada!!

ESTATÍSTICA DESCRITIVA

Variância

A variância σ^2 considera uma amplitude média que compara cada valor com um valor de referência, a média dos dados. Dessa forma cada valor é comparado com a média, e a diferença entre essa comparação é utilizada para determinar a variabilidade média.

A variância dos dados (populacionais) é calculada fazendo:

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

No entanto, com algum esforço matemático é possível obter uma **fórmula simplificada da variância**, que utiliza apenas somatórios.

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

Para os dados de exemplo:

$$\sigma^2 = \frac{\sum_{i=1}^{6} x_i^2}{6} - \left(\frac{\sum_{i=1}^{6} x_i}{6}\right)^2 = \frac{13^2 + 18^2 + 16^2 + 12^2 + 20^2 + 12^2}{6} - \left(\frac{91}{6}\right)^2 = \frac{1437}{6} - \frac{8281}{36} = 9.47222$$

ESTATÍSTICA DESCRITIVA

Desvio padrão

O desvio padrão (σ) é a raiz da variância e é mais adequado do que a variância, pois está na escala dos dados:

$$\sigma = \sqrt{\sigma^2} = \sqrt{9.4722} = 3.07$$

ESTATÍSTICA DESCRITIVA

Coeficiente de Variação (CV)

O coeficiente de variação (CV) é a medida correta para comparar variabilidade de dados diferentes. Essa medida é calculada utilizando a média (μ) e o desvio-padrão (σ) dos dados.

$$CV = \frac{\sigma}{\mu} \cdot 100\%$$

Dessa forma, o CV indica o percentual de variabilidade em relação a média dos dados.

A Gasto com cartão de crédito: O gasto média com cartão de crédito de uma certa família é de R$ 2.000,00 e o desvio padrão é de R$ 100,00.

$$CV_A = \frac{\sigma_A}{\mu_A} \cdot 100\% = \frac{100}{2000} \cdot 100\% = 5\%$$

B Gasto com energia elétrica: o gasto médio com energia elétrica em uma certa residência é de R$ 300,00 e o desvio padrão é de R$ 75,00

$$CV_B = \frac{\sigma_B}{\mu_B} \cdot 100\% = \frac{75}{300} \cdot 100\% = 25\%$$

ESTATÍSTICA DESCRITIVA

Medidas de posição dados agrupados

Fa ou Fi: Frequência absoluta

420

FR: Frequência relativa

420/8855 = 4,7%

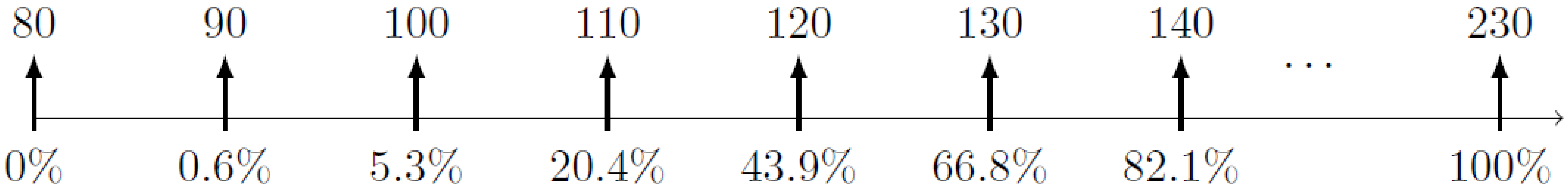
Frequência Relativa acumulada

$$\frac{52 + 420}{8855} = 0{,}6\% + 4{,}7\% = 5{,}3\%$$

Faixa	FA ou Fi	FR	FRA
[80-90)	52	0,6%	0,6%
[90-100)	420	4,7%	5,3%
[100-110)	1335	15,1%	20,4%
[110-120)	2083	23,5%	43,9%
[120-130)	2028	22,9%	66,8%
[130-140)	1354	15,3%	82,1%
[140-150)	748	8,4%	90,6%
[160-170)	240	2,7%	93,3%
[150-160)	318	3,6%	96,9%
[180-190)	72	0,8%	97,7%
[170-180)	138	1,6%	99,2%
[210-220)	13	0,1%	99,4%
[200-210)	21	0,2%	99,6%
[190-200)	28	0,3%	99,9%
[220-230)	5	0,1%	100,0%

ESTATÍSTICA DESCRITIVA

Quartis dados agrupados



Faixa	FA ou Fi	FR	FRA
[80-90)	52	0,6%	0,6%
[90-100)	420	4,7%	5,3%
[100-110)	1335	15,1%	20,4%
[110-120)	2083	23,5%	43,9%
[120-130)	2028	22,9%	66,8%
[130-140)	1354	15,3%	82,1%
[140-150)	748	8,4%	90,6%
[160-170)	240	2,7%	93,3%
[150-160)	318	3,6%	96,9%
[180-190)	72	0,8%	97,7%
[170-180)	138	1,6%	99,2%
[210-220)	13	0,1%	99,4%
[200-210)	21	0,2%	99,6%
[190-200)	28	0,3%	99,9%
[220-230)	5	0,1%	100,0%

Quartis:

Q1 = 1° QUARTIL ou posição 25%

Q2 = 2° QUARTIL ou posição 50%

Q3 = 3° QUARTIL ou posição 75%

Decis:

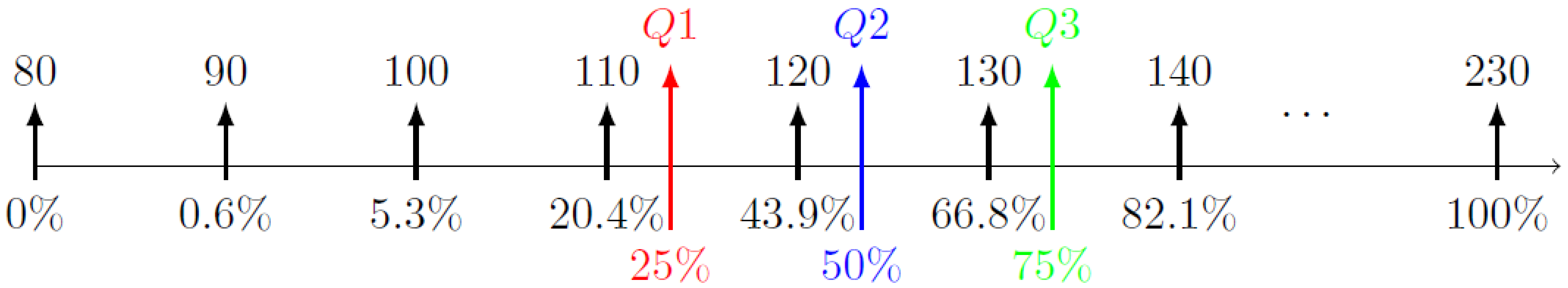
D1 = 1° DECIL ou posição 10%

D2 = 2° DECIL ou posição 20%

D9 = 9° DECIL ou posição 90%

ESTATÍSTICA DESCRITIVA

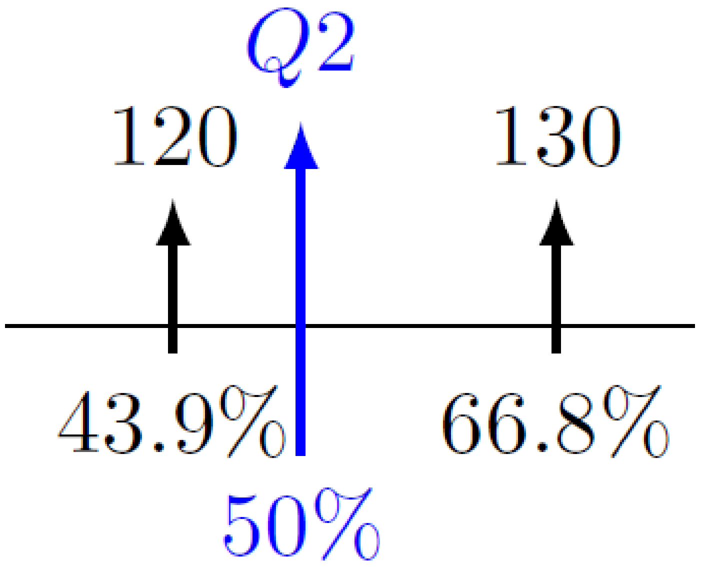
Quartis dados agrupados

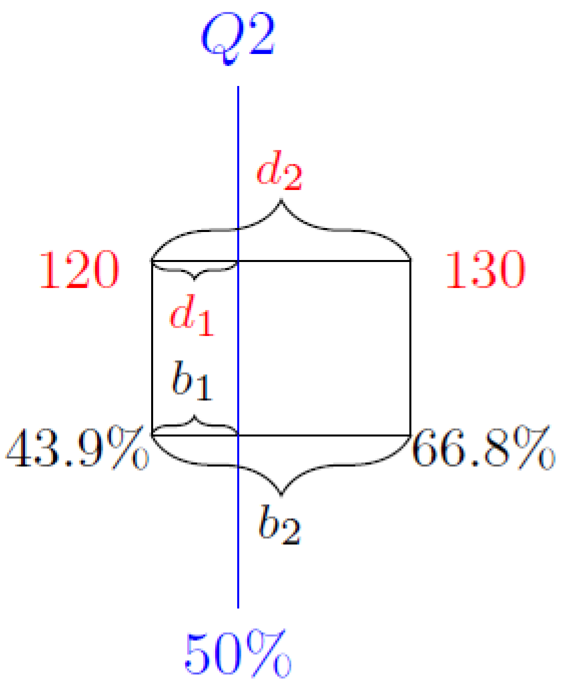


Faixa	FA ou Fi	FR	FRA
[80-90)	52	0,6%	0,6%
[90-100)	420	4,7%	5,3%
[100-110)	1335	15,1%	20,4%
[110-120)	2083	23,5%	43,9%
[120-130)	2028	22,9%	66,8%
[130-140)	1354	15,3%	82,1%
[140-150)	748	8,4%	90,6%
[160-170)	240	2,7%	93,3%
[150-160)	318	3,6%	96,9%
[180-190)	72	0,8%	97,7%
[170-180)	138	1,6%	99,2%
[210-220)	13	0,1%	99,4%
[200-210)	21	0,2%	99,6%
[190-200)	28	0,3%	99,9%
[220-230)	5	0,1%	100,0%

ESTATÍSTICA DESCRITIVA

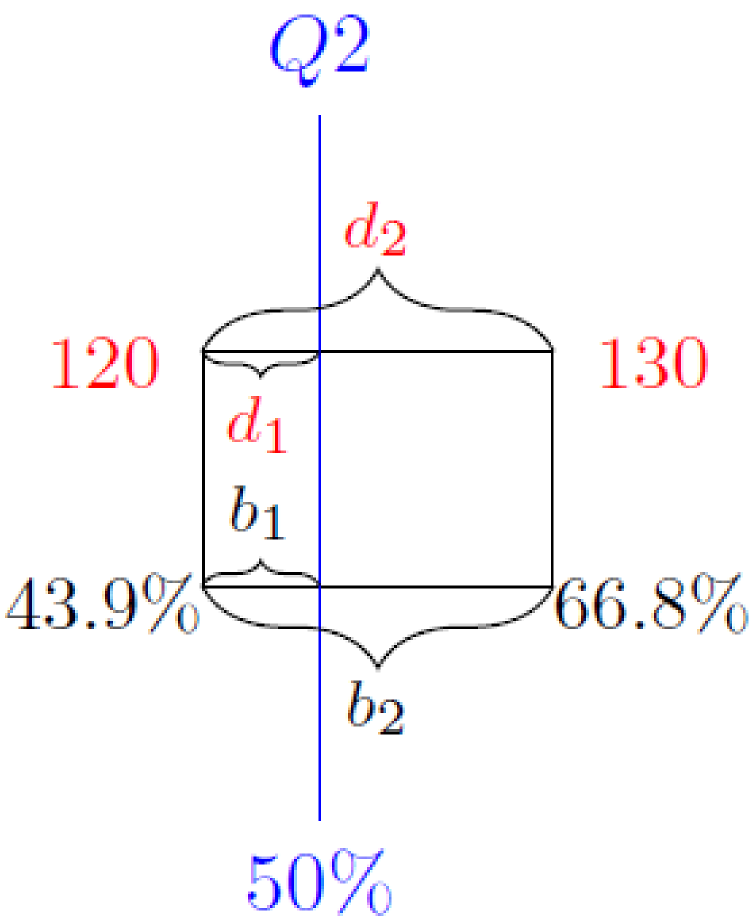
Quartis dados agrupados (Q2 = 50%)





ESTATÍSTICA DESCRITIVA

Quartis dados agrupados (Q2 = 50%)



$$\frac{d_1}{b_1} = \frac{d_2}{b_2}$$

$$\frac{Q_2 - 120}{50\% - 43.9\%} = \frac{130-120}{66.8\%-43.9\%}$$

$$\frac{Q_2 - 120}{6.1\%} = \frac{10}{22.9\%}$$

$$(Q_2 - 120) \times 22.9\% = 10 \times 6.1\%$$

$$Q_2 \times 22.9\% - 120 \times 22.9\% = 0.61$$

$$Q_2 \times 22.9\% - 27.5 = 0.61$$

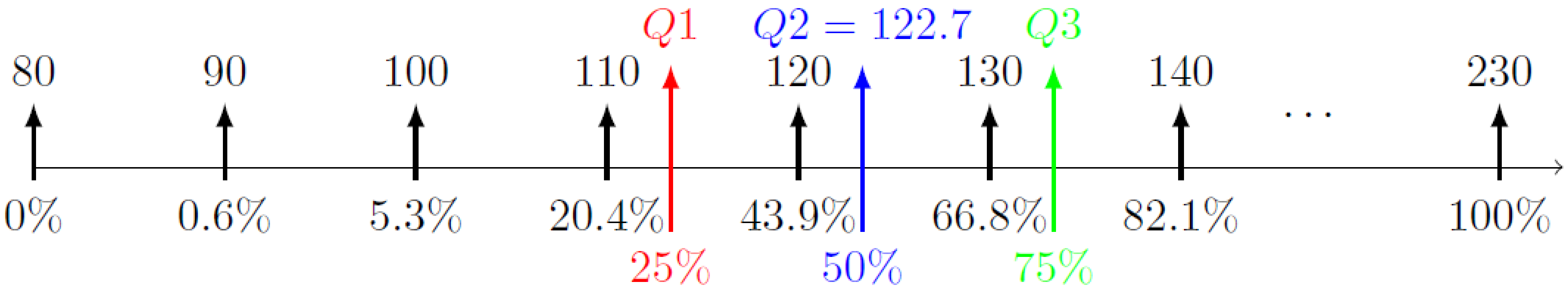
$$Q_2 \times 22.9\% = 28.11$$

$$Q_2 = \frac{28.11}{0.229}$$

$$Q_2 = 122.7$$

ESTATÍSTICA DESCRITIVA

Quartis dados agrupados (Q2 = 50%)



ESTATÍSTICA DESCRITIVA

Frequência relativa acumulada

PEREIRA, A. M.; SILVA, C. A. R.; QUEIROZ, D. C. A.; MORAES, M. J. B.; MELGES, J. L. P.; TASHIMA, M. M.; AKASAKI, J. L. revista Matéria, v.20, n.1, pp. 227– 238, 2015.

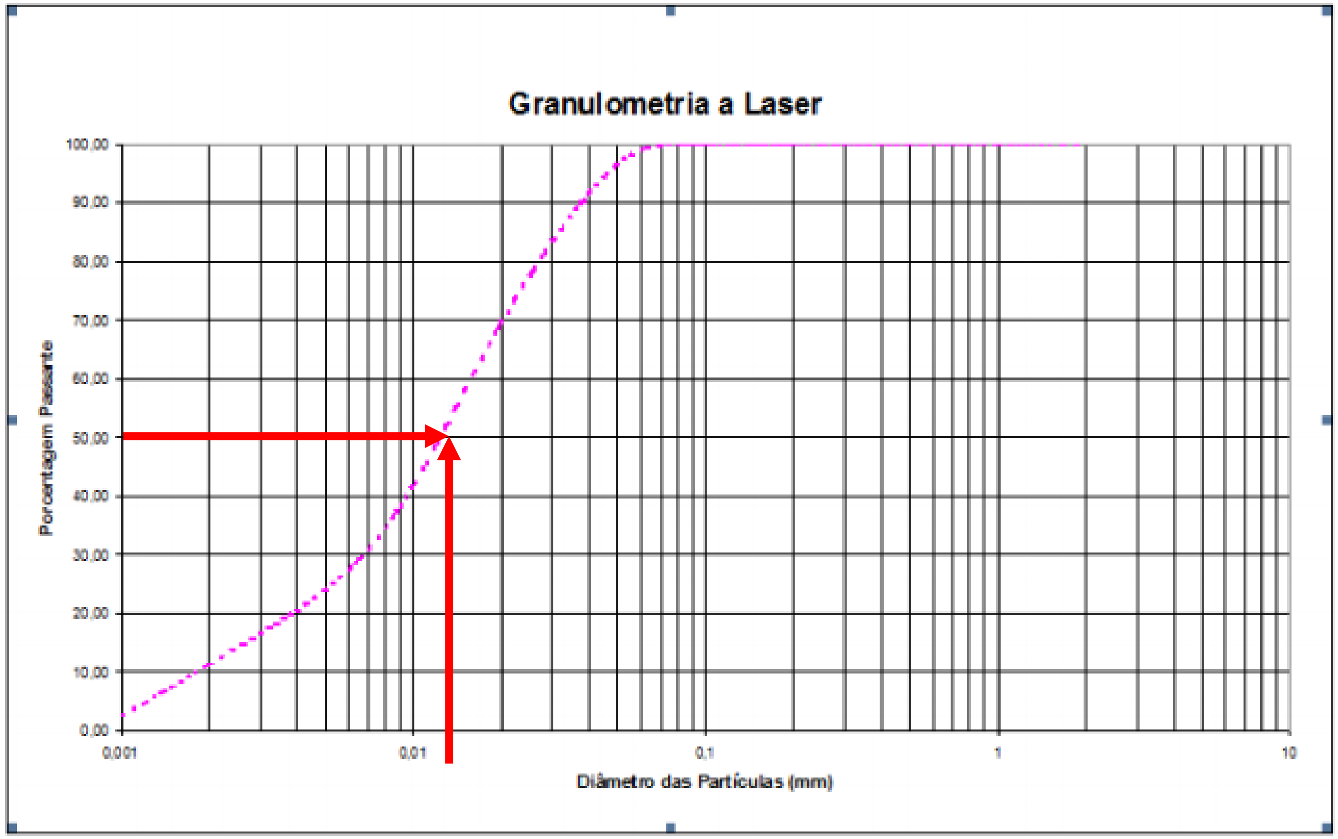


Figura 2: Distribuição granulométrica da CCA moída durante o período de 30 minutos.

ESTATÍSTICA DESCRITIVA

Histograma

Faixa	FA ou Fi	FR	FRA	Densidade
[80-90)	52	0,6%	0,6%	0,001
[90-100)	420	4,7%	5,3%	0,005
[100-110)	1335	15,1%	20,4%	0,015
[110-120)	2083	23,5%	43,9%	0,024
[120-130)	2028	22,9%	66,8%	0,023
[130-140)	1354	15,3%	82,1%	0,015
[140-150)	748	8,4%	90,6%	0,008
[160-170)	240	2,7%	93,3%	0,003
[150-160)	318	3,6%	96,9%	0,004
[180-190)	72	0,8%	97,7%	0,001
[170-180)	138	1,6%	99,2%	0,002
[210-220)	13	0,1%	99,4%	0,000
[200-210)	21	0,2%	99,6%	0,000
[190-200)	28	0,3%	99,9%	0,000
[220-230)	5	0,1%	100,0%	0,000
Total	8855	100,0%	-	-

Figura 1 – Dados agregados em intervalos

Fórmula de Sturges para o número de classes:

$$k = 3.322 \times log(n) + 1$$

k = 3.322 * log(8855) + 1
k = 31.19

Intervalo = (230-80)/31.19 = 4.80 ≈ 5.0

ESTATÍSTICA DESCRITIVA

Histograma

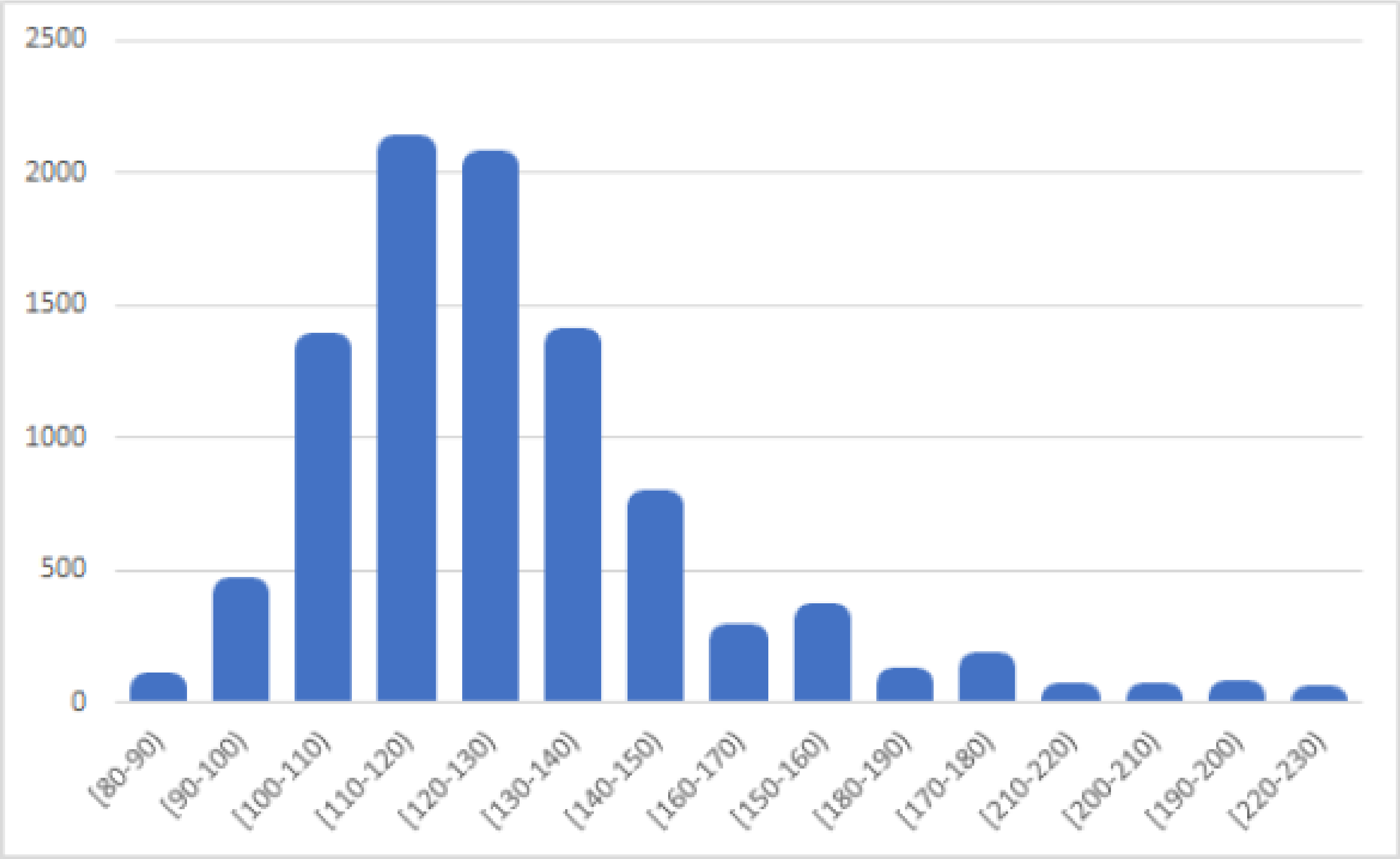


Figura 2 – Histograma com a frequência absoluta

ESTATÍSTICA DESCRITIVA

Boxplot

$$LS = Q3 + 1.5 \times (Q3 - Q1)$$

$$LI = Q1 - 1.5 \times (Q3 - Q1)$$

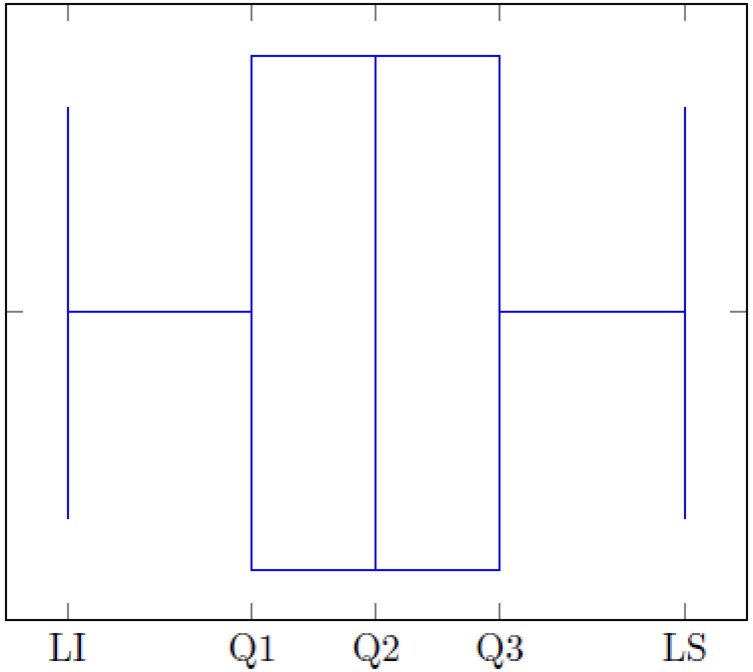


Figura 4 – Configuração do Boxplot

Q1: 1° quartil ou posição 25%
"25% dos valores estão abaixo do Q1"
"1 em cada 4 dos valores estão abaixo do Q1"

Q2: 2° quartil ou posição 50% (Mediana)
"50% dos valores estão abaixo do Q2"
"Metade dos valores estão abaixo do Q2"

Q5: 3° quartil ou posição 75%
"25% dos valores estão acima do Q3"
"apenas 1 em cada 4 dos valores estão acima do Q3"

ESTATÍSTICA DESCRITIVA

Assimetria

$$AS_1 = \frac{\mu - M_0}{\sigma}$$

$$AS_2 = \frac{3(\mu - Q2)}{\sigma}$$

$$AS_3 = \frac{(Q3 - Q2) - (Q2 - Q1)}{Q3 - Q1} = \frac{Q3 + Q1 - 2 \times Q2}{Q3 - Q1}$$

ESTATÍSTICA DESCRITIVA

Assimetria

Mediana

Moda

Média

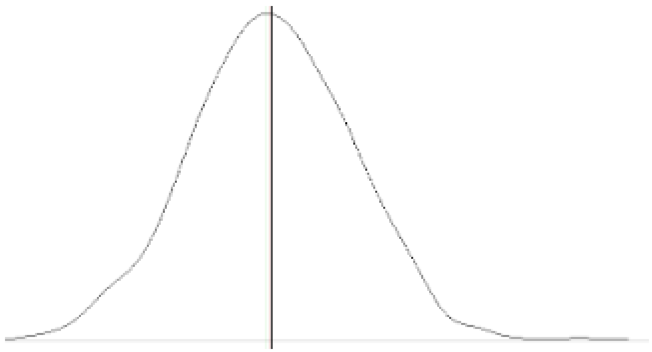


Figura 5 – Distribuição simétrica $AS_2 = 0$, Moda=Mediana=Média

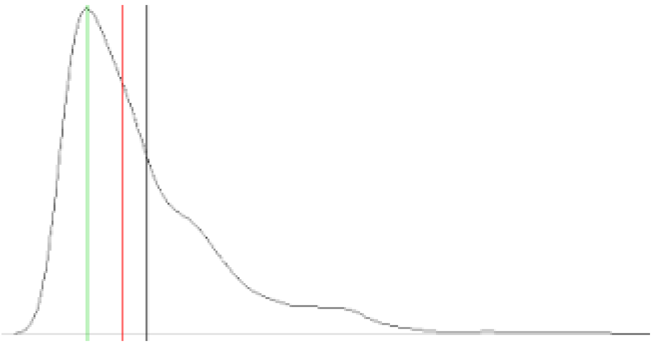


Figura 6 – Distribuição com assimetria positiva, Moda<Mediana<Média

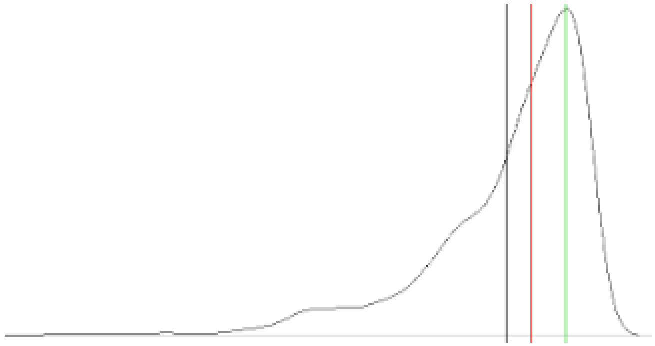


Figura 7 – Distribuição com assimetria negativa, Moda>Mediana>Média

ESTATÍSTICA DESCRITIVA

Assimetria

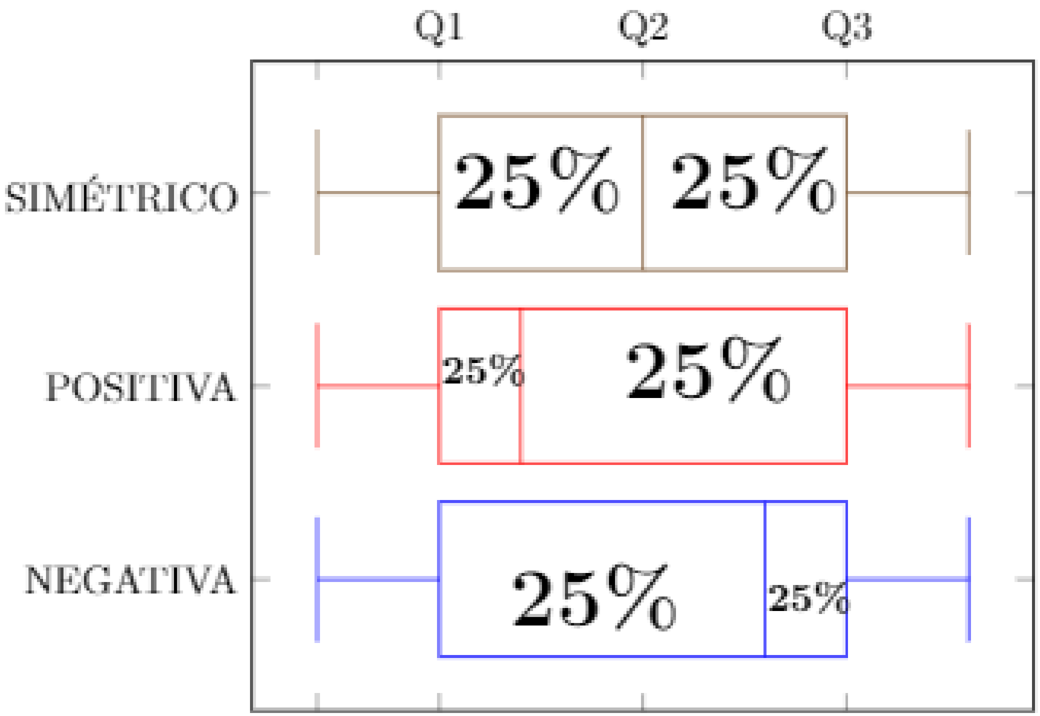


Figura 8 – Distribuição dos dados segundo tempo de cura.

$$AS_3 = \frac{(Q3 - Q2) - (Q2 - Q1)}{Q3 - Q1} = \frac{Q3 + Q1 - 2 \times Q2}{Q3 - Q1}$$

Referências

MORETTIN, P. A.; BUSSAB, W. O. *Estatística básica*. São Paulo: Editora Saraiva, 2017.