Prof. MSc. Marcos Alexandruk

E-mail: alexandruk@uni9.pro.br

https://github.com/alexandruk/analisededados

Estatística

Estatística

A palavra estatística tem origem no latim "**status**" e relaciona-se com "**estado**".

No início, a palavra era usada para se referir ao "**cidadão político**".

Posteriormente, passou a ser utilizada em alemão com o sentido de "**conjunto de dados do Estado**", de onde decorre o seu significado desde o século XIX.

BATISTA, Carolina. Estatística. Toda Matéria, 2021. Disponível em: https://www.todamateria.com.br/estatistica-conceito-fases-metodo/. Acesso em: 23/02/2021.

Estatística

“Estatística é uma ciência exata que estuda a coleta, a organização, a análise e registro de dados por amostras.

Utilizada desde a Antiguidade, quando se registravam os nascimentos e as mortes das pessoas, é um método de pesquisa fundamental para tomar decisões. Isso porque fundamenta suas conclusões nos estudos realizados.”

BATISTA, Carolina. Estatística. Toda Matéria, 2021. Disponível em: https://www.todamateria.com.br/estatistica-conceito-fases-metodo/. Acesso em: 23/02/2021.

Estatística

Estatística é ciência que tem por fim a **pesquisa e a comparação dos fatos gerais e particulares** verificados no movimento das sociedades.

Objetivo geral da estatística

O objetivo da estatística é a análise e interpretação dos fenômenos sociais de qualquer natureza, para planejamento de ações.

Fases do Método Estatístico

Definição do problema
Determinar como a coleta de dados pode solucionar um problema

Planejamento
Elaborar como fazer o levantamento dos dados

Coleta de Dados
Reunir os dados e definir da periodicidade da coleta (contínua, ocasional, etc.)

Correção dos dados
Conferir os dados para afastar eventuais erros

Apuração dos dados
Organização e contagem dos dados

Apresentação dos dados
Demonstrar o resultado da coleta dos dados (gráficos e tabelas)

Análise dos dados
Exame detalhado e interpretação dos dados

BATISTA, Carolina. Estatística. Toda Matéria, 2021. Disponível em: https://www.todamateria.com.br/estatistica-conceito-fases-metodo/. Acesso em: 23/02/2021. (adaptado)

Importância da Estatística na Engenharia

A probabilidade e estatística pode contribuir para mitigação dos erros e favorecer a análise de um projeto em construção, considerando as mais diversas situações, de forma que dados estatísticos podem auxiliar os testes de desempenho e o controle de qualidade.

Análise Descritiva

A análise descritiva dos dados se limita a calcular algumas medidas de posição e variabilidade, como a média e variância, por exemplo.

Inferência

Inferência estatística é um ramo da Estatística cujo objetivo é fazer afirmações a partir de um conjunto de valores representativo (amostra) sobre um universo.

Em geral, a inferência estatística está associada à coleta, à redução, à análise e à modelagem dos dados.

Tipos de Variáveis

Variáveis qualitativas: apresentam algum tipo de atributo do elemento pesquisado. (educação, estado civil, sexo, etc.)

Variáveis quantitativas: apontam para um impacto no elemento pesquisado e contribuem na análise.

Variáveis quantitativas discretas: Quando podemos expressar as variáveis por um número inteiro em certa contagem, chamamos de variável quantitativa discreta. (número de filhos, quantidade de veículos, etc.)

Variáveis quantitativas contínuas: Quando destacamos uma variável por intermédio de uma medida, chamamos de variável quantitativa contínua. (tempo, temperatura, pressão, etc.)

Distribuições de frequências

No estudo de uma variável, devemos dispor um maior interesse em conhecer a distribuição dessa variável por meio das possíveis realizações dela e dispor seus valores, de modo que se tenha uma boa ideia global dessa distribuição.

Distribuições de frequências

Frequência de porcentagens de 20 empregados segundo o grau de instrução:

Grau de instrução	Contagem	Frequência	Proporção	Porcentagem
1º grau	8	8	0,4	40%
2º grau	7	7	0,35	35%
Superior	5	5	0,25	25%
Total	**20**	**20**	**1,00**	**100%**

Distribuições de frequências

Frequência absoluta acumulada e frequência relativa acumulada:

Grau de instrução	Frequência Absoluta	Frequência Relativa	Frequência Absoluta Acumulada	Frequência Relativa Acumulada
1º grau	8	40%	8	40%
2º grau	7	35%	8 + 7 = 15	40% + 35% = 75%
Superior	5	25%	15 + 5 = 20	75% + 25% = 100%
Total	**20**	**100%**	**20**	**100%**

As frequências acumuladas são extremamente úteis quando o objetivo é saber a quantidade ou a porcentagem até determinada característica.

Amplitude total

Alturas de 32 crianças de 1 a 4 anos:

73,93	71,51	66,83	64,17	66,16	65,7	64,78	65,81
63,15	62,56	61,88	60,94	60,3	60,15	56,57	55,86
71,47	70,09	64,44	63,27	66,06	65,09	64,73	64,16
62,69	61,91	61,49	60,73	60,24	59,37	56,03	**55,77**

Amplitude total = Valor Máximo – Valor Mínimo

Amplitude total = 73,93 – 55,77

Amplitude total = 18,16

Números de classes

Alturas de 32 crianças de 1 a 4 anos:

73,93	71,51	66,83	64,17	66,16	65,7	64,78	65,81
63,15	62,56	61,88	60,94	60,3	60,15	56,57	55,86
71,47	70,09	64,44	63,27	66,06	65,09	64,73	64,16
62,69	61,91	61,49	60,73	60,24	59,37	56,03	**55,77**

Número de classes = SQRT (n)

Número de classes = SQRT (32)

Número de classes = 5,65 (aproximado para 6)

SQRT => Raiz Quadrada

Amplitude do intervalo

Amplitude do intervalo = Amplitude total / número de classes

Amplitude do intervalo = 18,16 / 6

Amplitude do intervalo = 3,02

Classes	fi	Fi	fr	Fr
55 ⊢	4	4	12,50%	12,50%
58 ⊢	6	10	18,75%	31,25%
61 ⊢	7	17	21,88%	53,13%
64 ⊢	11	28	34,38%	87,50%
67 ⊢	1	29	3,13%	90,63%
70 ⊢ [ERRO]	3	32	9,38%	100%
Total	**32**		**100%**	

Amplitude do intervalo

Amplitude do intervalo = Amplitude total / número de classes

Amplitude do intervalo = 18,16 / 6

Amplitude do intervalo = 3,02

Classes	fi	Fi	fr	Fr
55 ⊢ 58	4	4	12,50%	12,50%
58 ⊢ 61	6	10	18,75%	31,25%
61 ⊢ 64	7	17	21,88%	53,13%
64 ⊢ 67	11	28	34,38%	87,50%
67 ⊢ 70	1	29	3,13%	90,63%
70 ⊢ 73 **[ERRO]**	3	32	9,38%	100%
Total	**32**		**100%**	

[ERRO] O maior valor é 73,93 (está acima de 73)

Amplitude do intervalo

Amplitude do intervalo = Amplitude total / número de classes

Amplitude do intervalo = 18,16 / 6

Amplitude do intervalo = 3,02 (arredondar para 4)

Classes	fi	Fi	fr	Fr
55 ⊢ 59				
59 ⊢ 63				
63 ⊢ 67				
67 ⊢ 71				
71 ⊢ 75				
75 ⊢ 79				
Total	**32**		**100%**	

Regra de Sturges

$k = 1 + 3,3 * LOG(n)$

k = Número de classes

n = Total de dados

$k = 1+3,3*LOG(20)$

$k = 5,293399$ $\boxed{k \approx 5}$

$A_{Total} = Valor_{Max} - Valor_{Min}$

$A_{Total} = 42 - 15$ $\boxed{A_T = 27}$

$h = A_{Total}/k$

$h = 27/5$ $h = 5,4$ $\boxed{h \approx 6}$

h = Amplitude do Intervalo

(Arredondar para cima)

Pesquisa: Idade				
17	18	16	24	23
42	40	36	15	18
26	23	23	24	28
41	16	18	20	27

IDADE	fi
15 ⊢ 21	8
15 ⊢ 27	6
27 ⊢ 33	2
33 ⊢ 39	1
39 ⊢ 45	3

Exercício

SALÁRIOS					
20,50	9,50	15,30	17,20	24,10	19,90
15,40	12,70	7,40	16,50	15,30	26,20
14,90	7,80	23,30	15,90	11,80	18,40
13,40	14,30	16,20	16,70	9,20	16,80
9,80	20,10	17,80	17,10	12,60	15,90

Classes	fi
7,40 ⊢ 10,40	
10,40 ⊢ 13,60	
13,60 ⊢ 16,80	
16,80 ⊢ 20,00	
20,00 ⊢ 23,20	
23,20 ⊢ 26,40	

Amplitude Total (A_{Total}) =	
Total de dados (n) =	
Número de classes (k) =	
Amplitude do intervalo (h) =	

Exercício

SALÁRIOS					
20,50	9,50	15,30	17,20	24,10	19,90
15,40	12,70	7,40	16,50	15,30	26,20
14,90	7,80	23,30	15,90	11,80	18,40
13,40	14,30	16,20	16,70	9,20	16,80
9,80	20,10	17,80	17,10	12,60	15,90

Classes	fi
7,40 ⊢ 10,60	
10,60 ⊢ 13,80	
13,80 ⊢ 17,00	
17,00 ⊢ 20,20	
20,20 ⊢ 23,40	
23,40 ⊢ 26,60	

$k = 1 + 3{,}3 * LOG(n)$

k = Número de classes — $k = 1+3{,}3*LOG(30)$

n = Total de dados — $k = 6$ — $\boxed{k \approx 6}$

$A_{Total} = Valor_{Max} - Valor_{Min}$ — $A_{Total} = 26{,}20 - 7{,}40$ — $\boxed{A_T = 18{,}80}$

$h = A_{Total}/k$ — $h = 18{,}80/6$ — $h = 3{,}13$ — $\boxed{h \approx 3{,}20}$

h = Amplitude do Intervalo — *(Arredondar para cima)*

Exercício

SALÁRIOS					
20,50	9,50	15,30	17,20	24,10	19,90
15,40	12,70	7,40	16,50	15,30	26,20
14,90	7,80	23,30	15,90	11,80	18,40
13,40	14,30	16,20	16,70	9,20	16,80
9,80	20,10	17,80	17,10	12,60	15,90

Classes	fi
7,40 ⊢ 10,60	5
10,60 ⊢ 13,80	4
13,80 ⊢ 17,00	11
17,00 ⊢ 20,20	6
20,20 ⊢ 23,40	2
23,40 ⊢ 26,60	2
TOTAL VALORES	30

Amplitude Total (A_{Total}) =	18,80
Total de dados (n) =	30
Número de classes (k) =	6
Amplitude do intervalo (h) =	3,20

Exercício

Classes	fi	Fi	fr	Fr
7,40 ⊢ 10,60	5	5	16,66%	16,66%
10,60 ⊢ 13,80	4	9	13,33%	30,00%
13,80 ⊢ 17,00	11	20	36,66%	66,66%
17,00 ⊢ 20,20	6	26	20,00%	86,66%
20,20 ⊢ 23,40	2	28	6,66%	93,33%
23,40 ⊢ 26,60	2	30	6,66%	100,00%
Total	30		100%%	

fi = frequência absoluta

Fi = frequência absoluta acumulada

fr = frequência relativa

Fr = frequência relativa acumulada

Referências

DACHS J. N. W. Análise de dados e regressão. São Paulo: IME USP, 1978.

LEVIN J. Estatística aplicada a Ciências Humanas. São Paulo: Harper e Row do Brasil, 1978.

MORETTIN P. A. Introdução a estatística para ciências exatas. São Paulo: Atual Editora, 1981.

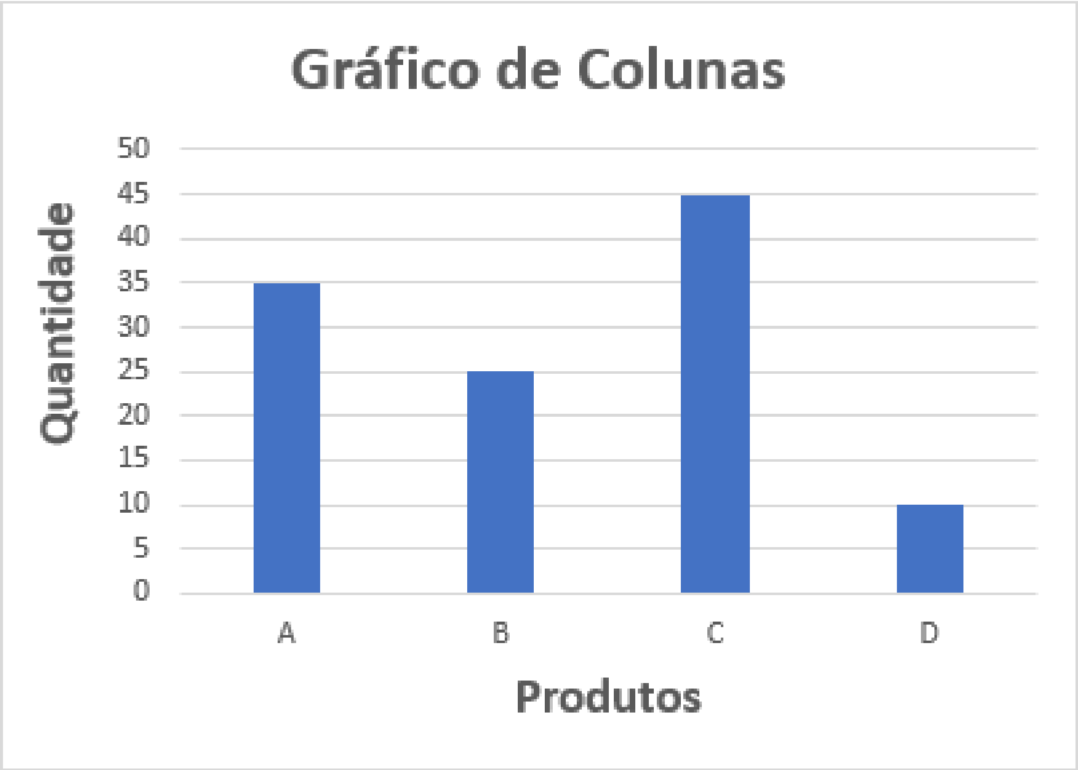
Prof. MSc. Marcos Alexandruk

E-mail: alexandruk@uni9.pro.br

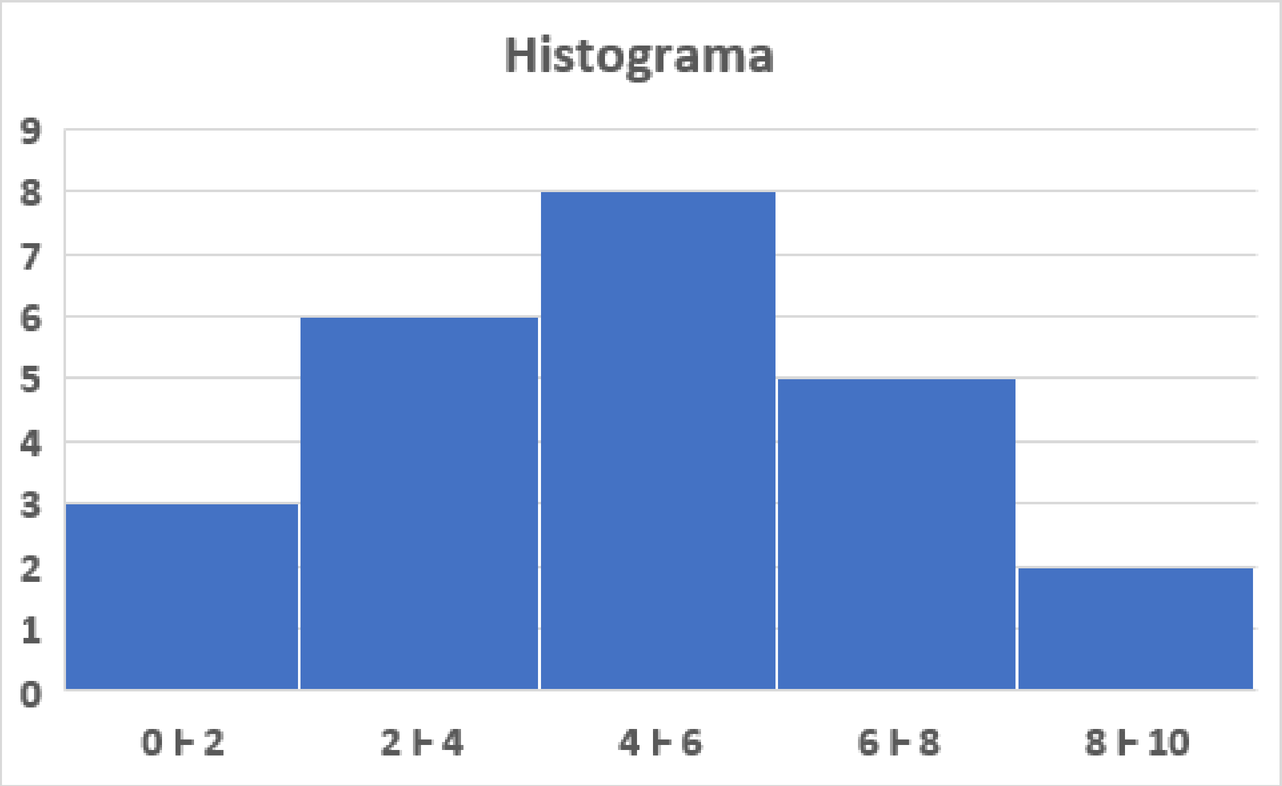
https://github.com/alexandruk/analisededados

Representações gráficas

Produtos	Quantidade
A	35
B	25
C	45
D	10

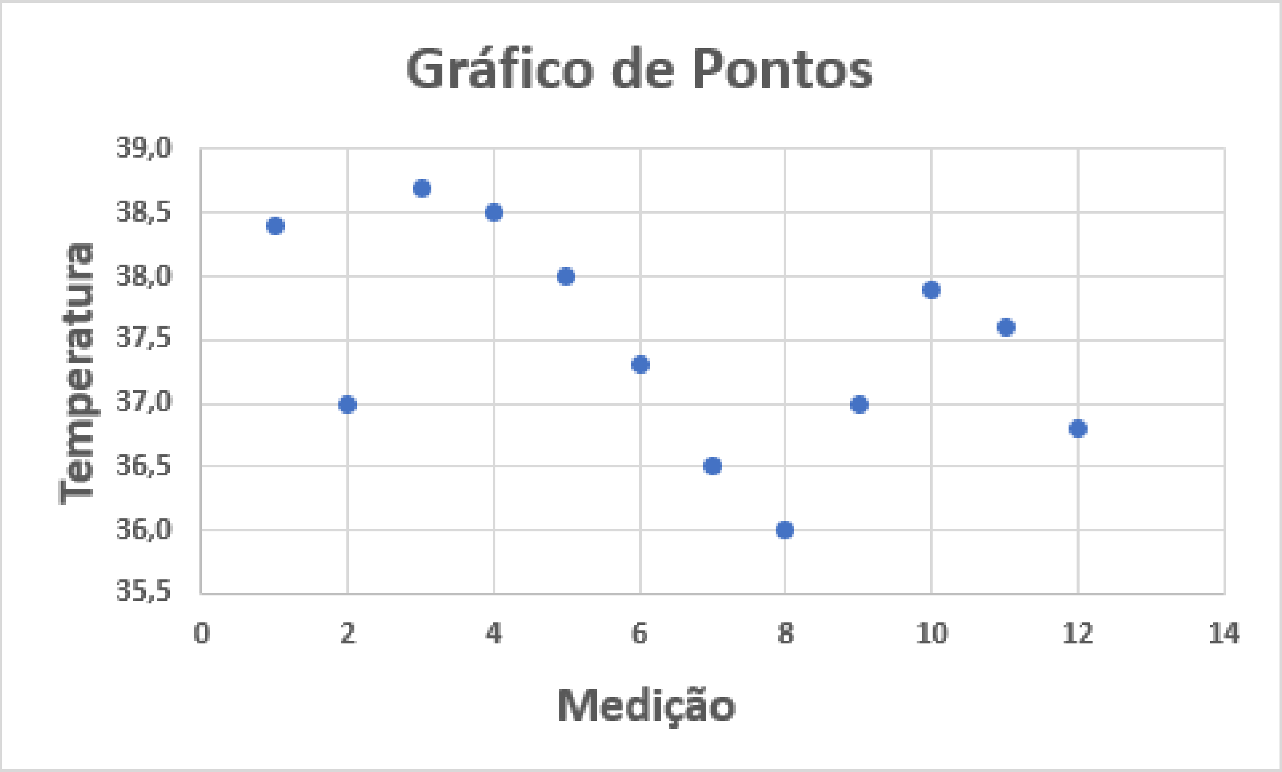


Classes	Frequências
0 ⊢ 2	3
2 ⊢ 4	6
4 ⊢ 6	8
6 ⊢ 8	5
8 ⊢ 10	2

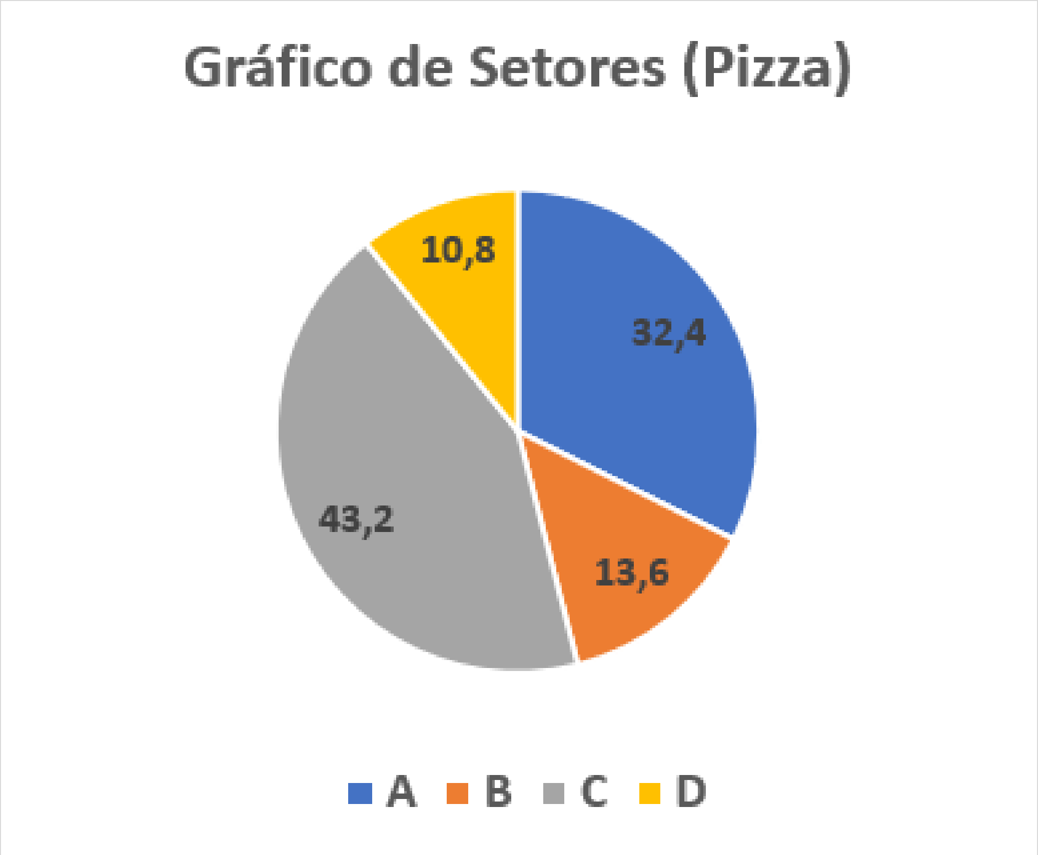


Medição	Temperatura
1	38,4
2	37,0
3	38,7
4	38,5
5	38,0
6	37,3
7	36,5
8	36,0
9	37,0
10	37,9
11	37,6
12	36,8

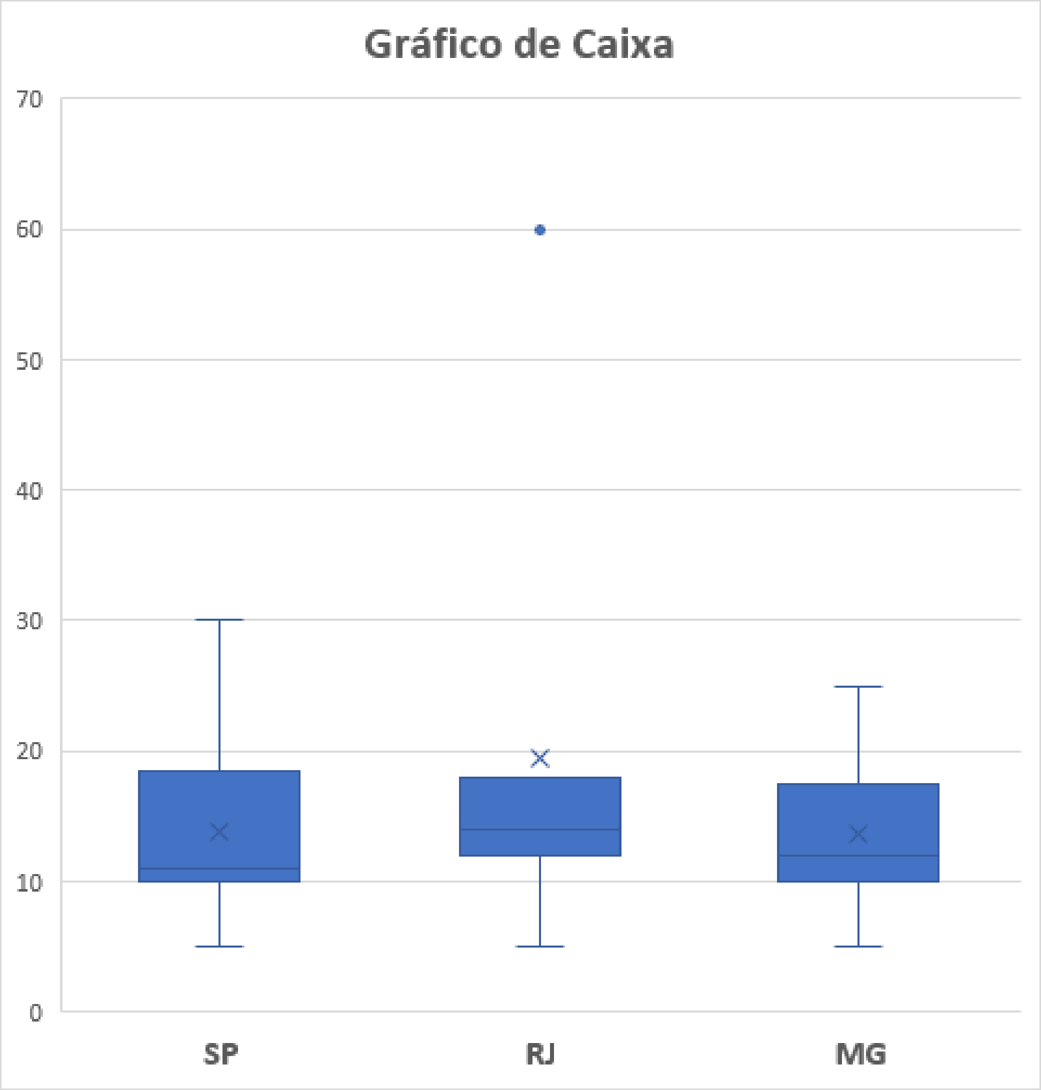
Medição	Temperatura
1	38,4
2	37,0
3	38,7
4	38,5
5	38,0
6	37,3
7	36,5
8	36,0
9	37,0
10	37,9
11	37,6
12	36,8



Produtos	Quantidade (%)
A	32,4
B	13,6
C	43,2
D	10,8



SP	10
SP	5
SP	30
SP	12
SP	10
SP	20
SP	14
SP	10
RJ	12
RJ	60
RJ	5
RJ	15
RJ	18
RJ	12
RJ	14
MG	10
MG	10
MG	12
MG	5
MG	14
MG	25
MG	12
MG	20
MG	15



Prof. MSc. Marcos Alexandruk

E-mail: alexandruk@uni9.pro.br

https://github.com/alexandruk/analisededados

Medidas de tendência central:
média aritmética; média geométrica; média harmônica

Média aritmética

1º caso: dados não agrupados

A média aritmética dos valores x_1 , x_2 , x_3 , ... , x_n é o quociente entre a soma desses valores e o seu número total n.

$$\bar{x} = \frac{x_1 + x_2 + x_3 + \cdots + x_n}{n} \quad \text{ou} \quad \bar{x} = \frac{\Sigma x_i}{n}$$

Exemplo: Determinar a média aritmética dos valores: 3, 7, 8, 10 e 11.

$$\bar{x} = \frac{3+7+8+10+11}{5} = 7{,}8$$

Média aritmética

2º caso: dados agrupados sem intervalos

Se os elementos x_1 , x_2 , x_3 , ... , x_n apresentam, respectivamente, frequências f_1 , f_2 ,f_3 , ... , f_n , então:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n}{n} \text{ ou } \bar{x} = \frac{\Sigma x_i f_i}{n}$$

Exemplo: dada a amostra: 2, 5, 5, 5, 5, 6, 6, 6, 8, 8, a média será:

$$\bar{x} = \frac{2\,.\,1 + 5\,.\,4 + 6\,.\,3 + 8\,.\,2}{10} = \frac{56}{10} = 5{,}6$$

x_i	f_i	$x_i f_i$
2	1	2
5	4	20
6	3	18
8	2	16
Total	10	56

Média aritmética

3º caso: dados agrupados com intervalos

Quando os dados estão agrupados, aceita-se, por convenção, que as frequências se distribuam uniformemente ao longo da classe e que, portanto, o seu ponto médio (x) é o valor representativo do conjunto. Então:

$$\bar{x} = \frac{x_1 f_1 + x_2 f_2 + x_3 f_3 + \cdots + x_n f_n}{n} \quad \text{ou} \quad \bar{x} = \frac{\Sigma x_i f_i}{n}$$

Exemplo: dada a amostra conforme a tabela, a média será:

$$\bar{x} = \frac{3{,}5\,.\,1 + 6{,}5\,.\,10 + 9{,}5\,.\,8 + 12{,}5\,.\,1}{20} = \frac{157}{20} = 7{,}85$$

Classe	x_i	f_i	$x_i f_i$
2 ⊢ 5	3,5	1	3,5
5 ⊢ 8	6,5	10	65
8 ⊢ 11	9,5	8	76
11 ⊢ 14	12,5	1	12,5
Total		20	157

Média geométrica

A média geométrica de um conjunto de números positivos é definida como o **produto de todos os membros do conjunto elevado ao inverso do número de membros**. Indica a tendência central ou o valor típico de um conjunto de números usando o produto dos seus valores.

A média geométrica é frequentemente utilizada quando comparamos diferentes itens – encontrando uma única "figura representativa" para esses itens – quando cada um desses itens possuem múltiplas propriedades que possuem diferentes escalas numéricas. Por exemplo, a média geométrica pode nos dar uma "média" significativa para comparar duas companhias que estão sendo classificadas numa escala de 0 a 5 para suas sustentabilidades ambientais e sendo classificadas de 0 a 100 para suas viabilidades financeiras. Se a média aritmética fosse usada em vez da média geométrica, a viabilidade financeira pesaria mais pois seu alcance numérico é grande, logo uma pequena mudança percentual na classificação financeira (por exemplo: uma mudança de 80 para 90) faria uma grande diferença na média aritmética do que uma grande diferença percentual na classificação da sustentabilidade ambiental (por exemplo uma mudança de 2 para 5 na escala).

Média geométrica

Sejam x1 , x2 , x3 , ... , xn valores da variável X, associadas, respectivamente, às frequências f1 , f2 , f3 , ... , fn. Então, a média geométrica de x é definida por:

$$M_g = \sqrt[n]{{x_1}^{f_1} \,.\, {x_2}^{f_2} \,.\, {x_3}^{f_3} \,.\, \ldots \,.\, {x_n}^{f_n}}$$

Em particular, se f_1 , f_2 , f_3 , ... , f_n = 1. temos:

$$M_g = \sqrt[n]{x_1 \,.\, x_2 \,.\, x_3 \,.\, \ldots \,.\, x_n}$$

Exemplo: Dada a tabela de distribuição de frequências, temos:

x_i	f_i
1	8
2	6
3	5
5	3
Total	22

$$M_g = \sqrt[22]{1^8 \,.\, 2^6 \,.\, 3^5 \,.\, 5^3} = \sqrt[22]{1944000} = 1{,}93$$

Média harmônica

A média harmônica é definida como a quantidade de elementos no conjunto, divida pela soma do inverso dos elementos do conjunto.

A média aritmética é muitas vezes utilizada erroneamente em casos que exigem a média harmônica. Um exemplo é o cálculo da velocidade média em um percurso de ida e volta em uma mesma via, em que a ida é percorrida a 60 km/h e a volta a 40 km/h a média aritmética de 50 está incorreta. A velocidade média no percurso total é a média harmônica de 40 e 60, ou seja 48km/h.

Exemplo:

Distância de A a B = 120 Km | Velocidade média = 40 Km/h | Duração da viagem = 120 /40 = 3 horas

Distância de B a A = 120 Km | Velocidade média = 60 Km/h | Duração da viagem = 120 /60 = 2 horas

Distância total de A a B + de B a A = 120 Km + 120 Km = 240 Km

Duração total da viagem = 3 horas + 2 horas = 5 horas

Velocidade média da viagem (de A a B + de B a A) = 240 / 5 = 48 Km/h

Média harmônica

Se os elementos x_1 , x_2 , x_3 , ... , x_n apresentam, respectivamente, frequências f_1 , f_2 , f_3 , ... , f_n , então, a média harmônica é definida como o inverso da média aritmética do inverso dos valores:

$$M_h = \frac{n}{\frac{f_1}{x_1} + \frac{f_2}{x_2} + \frac{f_3}{x_3} + \cdots + \frac{f_n}{x_n}} = \frac{n}{\sum \frac{f_i}{x_i}}$$

Em particular, se f_1 , f_2 , f_3 , ... , f_n = 1. temos:

$$M_h = \frac{n}{\frac{1}{x_1} + \frac{1}{x_2} + \frac{1}{x_3} + \cdots + \frac{1}{x_n}} = \frac{n}{\sum \frac{1}{x_i}}$$

x_i	f_i
1	8
2	6
3	5
5	3
Total	22

Exemplo: Dada a tabela de distribuição de frequências, temos:

$$M_h = \frac{22}{\frac{8}{1} + \frac{6}{2} + \frac{5}{3} + \frac{3}{5}} = \frac{22}{\frac{398}{30}} = 22.\frac{30}{398} = \frac{330}{199} = 1{,}66$$

Prof. MSc. Marcos Alexandruk

E-mail: alexandruk@uni9.pro.br

https://github.com/alexandruk/analisededados

Medidas de tendência central: moda e mediana

Objetivo

Calcular e interpretar as medidas de tendência central: a **moda** e a **mediana** de uma distribuição, destacando as suas diferenças e usos.

Média, moda e mediana

Apesar de ser bastante utilizada a média aritmética, nem sempre é a medida mais adequada para se analisar um agrupamento de dados.

Veja o exemplo. Numa certa empresa com 200 empregados, os salários são os seguintes:

Salários (em salários mínimos)	Número de empregados
1	100
2	30
3	30
4	5
5	25
10	5
25	3
40	2

Calculando o salário médio desses empregados, obtemos 3 salários mínimos.

Este número está correto do ponto de vista aritmético, mas não é representativo da condição salarial da maioria dos empregados. Afinal, 130 (65% do total) deles, ganham menos do que este valor. Por outro lado, de acordo com a tabela, 5 empregados (2,5%) ganham mais do que 20 salários mínimos, o que "puxa" a média para cima.

Neste caso, é mais conveniente usarmos outro tipo de medida como valor representativo do salário dos empregados, conforme veremos nesta aula.

Moda

Dada uma coleção de números, a moda é o valor que ocorre com **maior frequência**.

Assim, no exemplo citado, o salário mais frequente é o salário mínimo, que é recebido por 100 empregados, isto é, 1 salário mínimo.

Observações:

- Existem casos em que a moda não existe — os valores não se repetem ou todos os valores têm a mesma frequência (distribuição amodal).
- Em alguns casos, pode haver mais de uma moda, ou seja, a distribuição dos valores pode ser bimodal, trimodal, etc.

Moda (M_o)

1º Caso: dados não agrupados

É o valor de maior frequência ou que aparece mais vezes em um conjunto de dados.

Exemplo: **7, 8, 8, 9, 10, 10, 10, 12, 15**

O elemento de maior frequência é o 10, que aparece três vezes.

Portanto $\mathbf{M_o = 10}$ (distribuição unimodal).

Exemplo: **3, 5, 8, 10, 12 e 13**

Todos os elementos da série apresentam a mesma frequência, logo, a série é **amodal**.

Exemplo: **2, 2, 5, 5, 8, 9**

Os elementos 2 e 5 têm frequência 2. Logo, temos $\mathbf{M_o = 2}$ **e** $\mathbf{M_o = 5}$ (distribuição **bimodal**).

Moda (M_o)

2º Caso: dados agrupados sem intervalos

Basta identificar o elemento de maior frequência.

x_i	f_i
0	2
2	4
3	**5**
4	3
6	1

Portanto, $\mathbf{M_o = 3}$

Moda (M_o)

3º Caso: dados agrupados com intervalos

Neste caso, consideramos como moda o valor compreendido entre os limites da **classe modal**, ou seja, **aquela que apresenta a maior frequência**. Tal valor é dado por:

$$M_o = l_i + \frac{\Delta_1}{\Delta_1 + \Delta_2} . h$$

Em que:

$l_i =$ limite inferior da classe modal

$\Delta_1 =$ diferença entre a frequência (f_i) da classe modal e a imediatamente anterior

$\Delta_2 =$ diferença entre a frequência (f_i) da classe modal e a imediatamente posterior

$h =$ amplitude da classe modal

Moda (M_o)

3º Caso: dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i
0 ⊢ 10	1
10 ⊢ 20	3
20 ⊢ 30	6
30 ⊢ 40	2

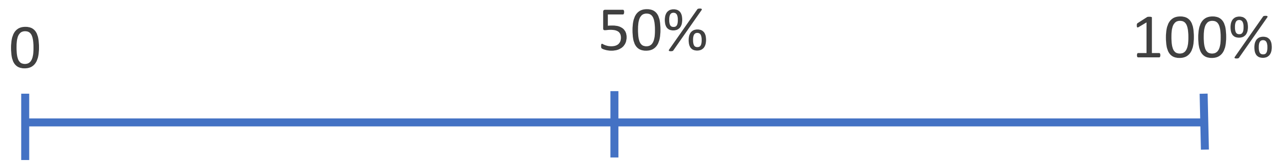
1º passo: identifica-se a classe modal (aquela que apresenta maior frequência)
No caso, trata-se da 3ª classe 20 ⊢ 30 (fi=6)

2º passo: aplica-se a fórmula. No caso temos:

$$l_i = 20$$
$$\Delta_1 = 6 - 3 = 3$$
$$\Delta_2 = 6 - 2 = 4$$
$$h = 30 - 20 = 10$$
$$M_o = 20 + \frac{3}{3+4} \,.\, 10$$
$$M_o = 20 + \frac{30}{7}$$
$$M_o = \frac{140 + 30}{7}$$
$$M_o = \frac{170}{7}$$
$$M_o = 24{,}29$$

Mediana ($\tilde{x}$)

Dada uma coleção de números colocados em ordem crescente, a mediana ($\tilde{x}$) é o valor que divide a amostra em duas partes iguais.



50% dos valores da série são valores menores ou iguais a $\tilde{x}$ e 50% dos valores da série são maiores ou iguais a $\tilde{x}$.

Mediana ($\tilde{x}$)

1º Caso: dados não agrupados

Quando temos um número ímpar de elementos, dispostos em ordem crescente, a mediana é definida como sendo o elemento central, de ordem $\frac{n+1}{2}$

Exemplo: **1, 2, 3, 4, 5**

$\tilde{x} = 3$

Se a coleção tiver um número par de elementos, também dispostos em ordem crescente, a mediana é definida como a média aritmética dos dois valores centrais, de ordem $\frac{n}{2}$ e $\frac{n}{2} + 1$

Exemplo: **2, 4, 6, 8, 10, 12**

$$\tilde{x} = \frac{6+8}{2} = \frac{14}{2} = 7$$

Mediana ($\bar{x}$)

2º Caso: dados agrupados sem intervalos

Basta considerar a frequência acumulada e localizar a mediana, procedendo da mesma forma que no caso anterior.

Exemplo 1: Dada a distribuição:

x_i	f_i
12	1
14	2
15	1
16	2
17	1
20	2
Total	9

Como n = 9 é ímpar, logo será o elemento de ordem $\frac{n+1}{2}$ ou seja:

$$\bar{x} = \frac{n+1}{2} = \frac{9+1}{2} = \frac{10}{2} = 5 \text{ (5º elemento)}$$

Mediana ($\bar{x}$)

2º Caso: dados agrupados sem intervalos

Exemplo 2: Dada a distribuição:

x_i	f_i
12	2
14	2
15	2
16	2
17	2
20	2
Total	12

Como n = 12 é par, logo será o elemento de ordem $\frac{n+1}{2}$ e $\frac{n+1}{2}$ + 1 ou seja:

$$\bar{x} = \frac{n}{2} e \frac{n}{2} + 1 = \frac{12}{2} \text{ e } \frac{12}{2} + 1 = 6 \text{ e } 7 \text{ (6º e 7º elemento)}$$

Mediana ($\bar{x}$)

3º Caso: dados agrupados com intervalos

Neste caso, devemos inicialmente localizar a classe mediana. Para isso seguimos os seguintes passos:

1º passo: calculamos a ordem $\frac{1}{2}$. Independente se n é par ou ímpar.

2º passo: pela F_i (Frequência acumulada) identificamos a classe que contém a mediana.

3º passo: utilizamos a fórmula:

$$\bar{x} = l_i + \frac{\left(\frac{n}{2} - \Sigma_f\right)}{F_{Md}} \cdot h$$

Em que:

l_i = limite inferior da classe modal

n = tamanho total da amostra ou número de elementos

$\sum_f$ = soma das frequências anteriores à classe mediana

h = amplitude da classe mediana

F_{Md} = Frequência da classe mediana

Mediana ($\bar{x}$)

3º Caso: dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i	F_i
3 ⊢ 6	2	2
6 ⊢ 9	5	7
9 ⊢ 12	8	15
12 ⊢ 15	3	18
15 ⊢ 18	1	19
Total	19	

1º passo: calcula-se $\frac{n}{2}$. Como n = 19, temos:

$\frac{19}{2} = 9{,}5$ (elemento)

2º passo: Identifica-se a classe mediana pela F_i.

Neste caso a classe mediana é a 3ª: 9 ⊢ 12

3º passo: Aplica-se a fórmula:

$l_i = 9$
$\Sigma_f = 7$
$h = 12 - 9 = 3$
$F_{Md} = 8$

Portanto:

$$\bar{x} = l_i + \frac{\left(\frac{n}{2} - \Sigma_f\right)}{F_{Md}} \cdot h$$

$$\bar{x} = 9 + \frac{9{,}5 - 7}{8} \cdot 3 = 9 + \frac{2{,}5}{8} \cdot 3 = 9 + \frac{7{,}5}{8} = 9{,}9375$$

Conclusão 50% dos valores são menores ou iguais a 9,94 e 50% são maiores ou iguais a 9,94

Prof. MSc. Marcos Alexandruk

E-mail: alexandruk@uni9.pro.br

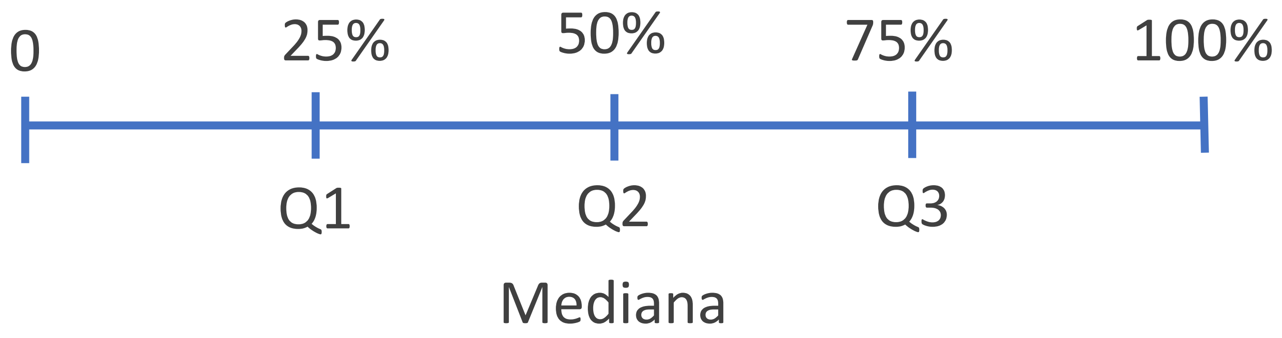
https://github.com/alexandruk/analisededados

Separatrizes

Quartil, Decil, Percentil

Quartis (Q_1, Q_2 (Md) e Q_3)

Os **quartis** dividem um conjunto ordenado de dados em quatro partes iguais, com cada parte representando 25%.



Quartis (Q_1, Q_2 (Md) e Q_3)

As notas de nove alunos em uma determinada prova estão apresentadas a seguir:

73, 74, 77, 52, 85, 59, 73, 84, 92

Determine a mediana, o 1º e o 3º quartil.

1º passo: Ordenar os elementos em ordem crescente:

52, 59, 73, 73, 74, 77, 84, 85, 92

2º passo: Determinar a mediana (o elemento central):

Q2 (Mediana) = 52, 59, 73, 73, 74, 77, 84, 85, 92

3º passo: Determinar Q1 (1º Quartil):

52, 59, 73, 73, 74, 77, 84, 85, 92

Q1 = (59+73)/2 = 122/2 = 66

4º passo: Determinar Q3 (3º Quartil):

52, 59, 73, 73, 74, 77, 84, 85, 92

Q3 = (84+85)/2 = 169/2 = 84,5

Quartis (Q_1)

Dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i	F_i
160 ⊢ 164	7	7
164 ⊢ 168	4	11
168 ⊢ 172	5	16
172 ⊢ 176	8	24
176 ⊢ 180	16	40
	$\sum f_i = 40$	

$$* = l_i + \frac{(k.\sum fi - F_{i\,anterior})}{f_i} \cdot h$$

Q_1: calcular $\frac{1\,.\,n}{4}$. Como n = 40, temos:

$\frac{40}{4} = 10º$ (elemento)

2º passo: Identifica-se a classe do Q_1 pela F_i.

Neste caso a classe Q_1 é a 2ª: 164 ⊢ 168

3º passo: Aplica-se a fórmula:

$l_i = 164$
$k = \frac{1}{4}$
$\Sigma_{fi} = 40$
$F_{i\,anterior} = 7$
$f_i = 4$
$h = 168 - 164 = 4$

Portanto:

$$Q_1 = 164 + \frac{\left(\frac{1}{4}.40 - 7\right)}{4} \cdot 4$$

$$Q_1 = 164 + \frac{(10 - 7)}{\cancel{4}} \cdot \cancel{4}$$

$$Q_1 = 164 + 3 = 167$$

Quartis (Q_2)

Dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i	F_i
160 ⊢ 164	7	7
164 ⊢ 168	4	11
168 ⊢ 172	5	16
172 ⊢ 176	8	24
176 ⊢ 180	16	40
	$\sum f_i = 40$	

$$* = l_i + \frac{(k.\sum fi - F_{i\,anterior})}{f_i} \cdot h$$

Q_2: calcular $\frac{2\,.\,n}{4}$. Como n = 40, temos:

$\frac{80}{4} = 20º$ (elemento)

2º passo: Identifica-se a classe do Q_2 pela F_i.

Neste caso a classe Q_2 é a 4ª: 172 ⊢ 176

3º passo: Aplica-se a fórmula:

$l_i = 172$
$k = \frac{1}{2}$
$\Sigma_{fi} = 40$
$F_{i\,anterior} = 16$
$f_i = 8$
$h = 176 - 172 = 4$

Portanto:

$$Q_2 = 172 + \frac{\left(\frac{1}{2}.40 - 16\right)}{8} \cdot 4$$

$$Q_2 = 172 + \frac{(20 - 16)}{\cancel{8}\ 2} \cdot \cancel{4}$$

$$Q_2 = 172 + 2 = 174$$

Quartis (Q_3)

Dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i	F_i
160 ⊢ 164	7	7
164 ⊢ 168	4	11
168 ⊢ 172	5	16
172 ⊢ 176	8	24
176 ⊢ 180	16	40
	$\sum f_i = 40$	

$$* = l_i + \frac{(k.\sum fi - F_{i\,anterior})}{f_i} \cdot h$$

Q_3: calcular $\frac{3\,.\,n}{4}$. Como n = 40, temos:

$\frac{120}{4} = 30º$ (elemento)

2º passo: Identifica-se a classe do Q_3 pela F_i.

Neste caso a classe Q_3 é a 5ª: 176 ⊢ 180

3º passo: Aplica-se a fórmula:

$l_i = 176$

$k = \frac{3}{4}$

$\Sigma_{fi} = 40$

$F_{i\,anterior} = 24$

$f_i = 16$

$h = 180 - 176 = 4$

Portanto:

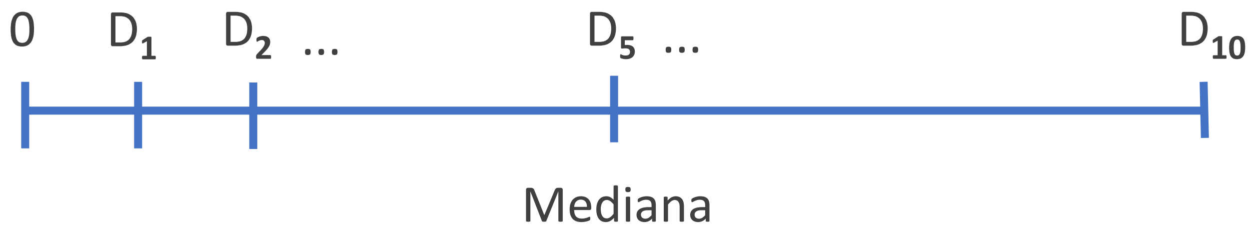
$$Q_3 = 176 + \frac{\left(\frac{3}{4}.40 - 24\right)}{16} \cdot 4$$

$$Q_3 = 176 + \frac{(30 - 24)}{\cancel{16}\ 4} \cdot \cancel{4}$$

$$Q_3 = 176 + \frac{6}{4} = 176 + 1{,}5 = 177{,}5$$

Decil (D_1, D_2 , D_3 ... D_{10})

Os **decis** dividem um conjunto ordenado de dados em 10 partes iguais, com cada parte representando 10%.



Decil

Dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i	F_i
160 ⊢ 162	7	7
162 ⊢ 164	4	11
164 ⊢ 166	8	19
166 ⊢ 168	9	28
168 ⊢ 170	12	40
	$\sum f_i = 40$	

$$* = l_i + \frac{(k.\sum fi - F_{i\,anterior})}{f_i} \cdot h$$

D_2: calcular $\frac{2\,.\,n}{10}$. Como n = 40, temos:

$\frac{80}{10} = 8^{\underline{o}}$ (elemento)

2º passo: Identifica-se a classe D_2 pela F_i.

Neste caso a classe D_2 é a 2ª: 162 ⊢ 164

3º passo: Aplica-se a fórmula:

$l_i = 162$

$k = \frac{2}{10}$

$\Sigma_{fi} = 40$

$F_{i\,anterior} = 7$

$f_i = 4$

$h = 164 - 162 = 2$

Portanto:

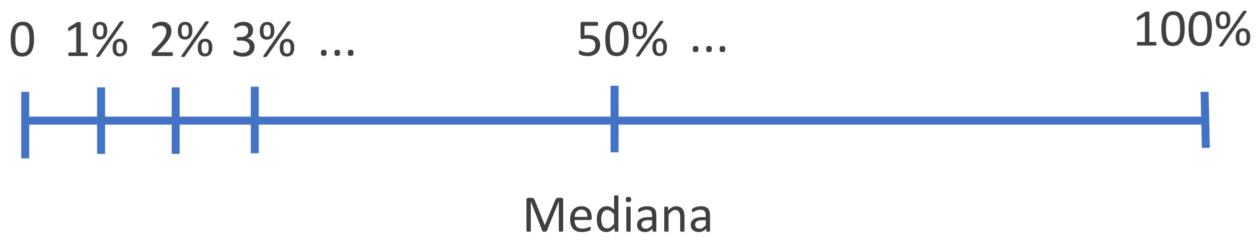
$$D_2 = 162 + \frac{\left(\frac{2}{10}.40 - 7\right)}{4} \cdot 2$$

$$D_2 = 162 + \frac{(8-7)}{\cancel{4}\,2} \cdot \cancel{2}$$

$$D_2 = 162 + \frac{1}{2} = 162 + 0{,}5 = 162{,}5$$

Percentil (P_1, P_2 , P_3 ... P_{100})

Os **percentis** dividem um conjunto ordenado de dados em 100 partes iguais, com cada parte representando 1%.



Percentil

Dados agrupados com intervalos

Exemplo: Dada a tabela:

classe	f_i	F_i
160 ⊢ 162	7	7
162 ⊢ 164	4	11
164 ⊢ 166	8	19
166 ⊢ 168	9	28
168 ⊢ 170	12	40
	$\sum f_i = 40$	

$$* = l_i + \frac{(k.\sum fi - F_{i\,anterior})}{f_i} \cdot h$$

P_{20}: calcular $\frac{20\,.\,n}{100}$. Como n = 40, temos: $\frac{800}{100} = 8º$ (elemento)

2º passo: Identifica-se a classe P_{20} pela F_i.

Neste caso a classe P_{20} é a 2ª: 162 ⊢ 164

3º passo: Aplica-se a fórmula:

$l_i = 162$
$k = \frac{20}{100}$
$\Sigma_{fi} = 40$
$F_{i\,anterior} = 7$
$f_i = 4$
$h = 164 - 162 = 2$

Portanto:

$$P_{20} = 162 + \frac{\left(\frac{20}{100}.40 - 7\right)}{4} \cdot 2$$

$$P_{20} = 162 + \frac{(8-7)}{\cancel{4}\,2} \cdot \cancel{2}$$

$$P_{20} = 162 + \frac{1}{2} = 162 + 0{,}5 = 162{,}5$$

R (quartil, decil e percentil)

```
x <- c(69, 70, 75, 66, 83, 88, 66, 63, 61, 68, 73, 57, 52, 58, 77)
```

quartis

```
quantile(x)
```

decis

```
quantile(x, prob = seq(0, 1, length = 11))
```

percentis

```
quantile(x, prob = seq(0, 1, length = 101))
```

resumo

```
summary(x)
```

R (entrada de dados externos – arquivo .csv)

Criar o arquivo teste.txt:

```
nome,idade
Fulano,20
Beltrano,30
Sicrano,40
```

Importar os dados:

```
teste<-read.table("c:/Aulas/teste.txt",header=T,sep=",")
```

Medidas de Dispersão

desvio médio, variância, desvio padrão e coeficiente de variação

Média, moda e mediana

A **média**, apesar de ser uma medida muito utilizada em Estatística, é muitas vezes insuficiente para caracterizar aceitavelmente uma distribuição.

A **moda** e a **mediana** também são medidas que nem sempre são suficientes para caracterizar um conjunto de dados.

Em alguns casos, temos que recorrer a outros parâmetros, chamados de medidas de dispersão.

As medidas de dispersão são medidas estatísticas utilizadas para avaliar o grau de variabilidade ou dispersão dos valores em torno da média. Servem para medir a representatividade da média.

Média, moda e mediana

Observe as séries:

a. 10, 1, 18, 20, 35, 3, 7, 15, 11, 10
b. 12, 13, 13, 14, 12, 14, 12, 14, 13, 13
c. 13, 13, 13, 13, 13, 13, 13, 13, 13, 13

Estes dados possuem a mesma média 13. No entanto, são sequências completamente distintas do ponto de vista da variabilidade de dados.

Na série "c" não há dispersão.

Comparando-se as séries "a" e "b", percebe-se que "a" apresenta maior dispersão em torno da média do que "b".

Isso indica que necessitamos de outro tipo de medida para distinguir e comparar os três conjuntos de dados.

O critério frequentemente usado para tal fim é aquele que mede a maior ou menor dispersão dos dados em torno da média, e as medidas mais usadas são:

- **desvio médio**
- **variância**
- **desvio padrão**
- **coeficiente de variação**

Desvio médio (Dm)

É a análise dos desvios em torno da média. Calculamos inicialmente a média da amostra ($\bar{x}$):

Em seguida, identificamos a distância de cada elemento da amostra para sua média:

$$|d_i| = |x_i - \bar{x}|$$

Finalmente, calculamos o desvio médio:

$$\frac{\sum |d_i| F_i}{n} \quad \text{ou} \quad \frac{\sum |x_i - \bar{x}| F_i}{n}$$

Onde x_i é a variável, $\bar{x}$ a média e n o número de dados da amostra.

Dessa forma, o desvio médio é a média aritmética dos valores absolutos dos desvios.

x_i	F_i	$x_i F_i$	$\lvert d_i\rvert = \lvert x_i - \bar{x}\rvert$	$\lvert d_i\rvert F$
2	5	10	|2 - 4,17| = 2,17	2,17 x 5 = 10,85
3	4	12	|3 - 4,17| = 1,17	1,17 x 4 = 4,68
5	4	20	|5 - 4,17| = 0,83	0,83 x 4 = 3,32
6	2	12	|6 - 4,17| = 1,83	1,83 x 2 = 3,66
7	3	21	|7 - 4,17| = 2,83	2,83 x 3 = 8,49
Total	18	75		31

$$\bar{x} = \frac{\sum x_i F_i}{n} = \frac{75}{18} = 4{,}17$$

$$Dm = \frac{\sum |\mathrm{di}| F_i}{n} = \frac{31}{18} = 1{,}72$$

Variância (Var)

É a média aritmética dos quadrados dos desvios. Logo:

$$Var = \frac{\sum \mathrm{d}_i^2 F_i}{n}$$

x_i	F_i	$x_i F_i$	$\lvert d_i \rvert = \lvert x_i - \bar{x} \rvert$	di^2	$d_i^2 F_i$
2	5	10	|2 - 4,17| = 2,17	4,71	23,55
3	4	12	|3 - 4,17| = 1,17	1,37	5,48
5	4	20	|5 - 4,17| = 0,83	0,69	2,76
6	2	12	|6 - 4,17| = 1,83	3,35	6,7
7	3	21	|7 - 4,17| = 2,83	8,01	24,03
Total	18	75			62,52

$$Var = \frac{\sum \mathrm{d}_i^2 F_i}{n} = \frac{62{,}52}{18} = 3{,}47$$

Desvio padrão (Dp)

Como para calcular a variância trabalhamos com os quadrados dos desvios, podemos ter uma incompatibilidade em relação às unidades dos valores da variável considerada.

Para contornar esse problema, temos o desvio padrão, que é a raiz quadrada da variância:

$$Dp = \sqrt{Var}$$

x_i	F_i	$x_i F_i$	$\lvert d_i \rvert = \lvert x_i - \bar{x} \rvert$	di^2	$d_i^2 F_i$
2	5	10	|2 - 4,17| = 2,17	4,71	23,55
3	4	12	|3 - 4,17| = 1,17	1,37	5,48
5	4	20	|5 - 4,17| = 0,83	0,69	2,76
6	2	12	|6 - 4,17| = 1,83	3,35	6,7
7	3	21	|7 - 4,17| = 2,83	8,01	24,03
Total	18	75			62,52

$$Var = \frac{\sum d_i^2 F_i}{n} = \frac{62,52}{18} = 3,47$$

$$Dp = \sqrt{Var} = \sqrt{3,47} = 1,86$$

Resumindo: a distribuição possui média **4,17**. Isto é, seus valores estão em torno de **4,17** e seu grau de concentração é de **1,72**, medido pelo desvio médio e de **1,86**, medido pelo desvio padrão.

Coeficiente de Variação (CV)

O desvio padrão por si só não nos diz muita coisa; para contornar esta dificuldade, usamos o coeficiente de variação.

Trata-se de uma medida relativa de dispersão útil para a comparação em termos relativos do grau de concentração em torno da média de séries distintas.

É expresso em porcentagens e dado por:

$$CV = \frac{Dp}{\bar{x}} \cdot 100$$

Onde Dp é o desvio padrão e $\bar{x}$, a média da distribuição.

Diz-se que a distribuição possui pequena variabilidade (dispersão) quando o CV apresentar valor até 15%; média dispersão quando estiver acima de 15% até 30% e grande dispersão quando superar 30%.

Coeficiente de Variação (CV)

Considere a tabela abaixo:

x_i	F_i	x_iF_i	$\|d_i\| = \|x_i - \bar{x}\|$	di^2	$d_i^2 F_i$
2	5	10	|2 - 4,17| = 2,17	4,71	23,55
3	4	12	|3 - 4,17| = 1,17	1,37	5,48
5	4	20	|5 - 4,17| = 0,83	0,69	2,76
6	2	12	|6 - 4,17| = 1,83	3,35	6,7
7	3	21	|7 - 4,17| = 2,83	8,01	24,03
Total	18	75			62,52

$$Var = \frac{\sum d_i^2 F_i}{n} = \frac{62{,}52}{18} = 3{,}47$$

$$Dp = \sqrt{Var} = \sqrt{3{,}47} = 1{,}86$$

$$\bar{x} = 4{,}17$$

$$CV = \frac{Dp}{\bar{x}} \cdot 100$$

$$CV = \frac{1{,}86}{4{,}17} .100 = 44{,}60\%$$

$$\boldsymbol{alta\ dipers\tilde{a}o}$$

Baixa dispersão CV ≤ 15%

Média dispersão: 15% < CV < 30%

Alta dispesão: CV ≥ 30%

Coeficiente de Variação (CV)

Exemplo: Numa empresa, o salário médio dos homens é de R$ 4.000,00, com desvio padrão de R$ 1.500,00 e, o das mulheres, é em média de R$ 3.000,00, com desvio padrão de R$ 1.200,00. Então:

$$CV_H = \frac{1500}{4000} . 100 = 37{,}5\%$$

$$CV_M = \frac{1200}{3000} . 100 = 40\%$$

Logo, podemos concluir que os salários das mulheres apresentam maior dispersão que os dos homens.

De modo geral, quanto menor o CV, menos dispersos estão os dados em torno da média, que passa a ser mais representativa do conjunto de dados.

Medidas de Dispersão (Exemplo)

Encontre o desvio médio, o desvio padrão e o coeficiente de variação da distribuição:

Classes	x_i	F_i	x_iF_i	$\lvert d_i \rvert$	$\lvert d_i \rvert F_i$	di^2	$d_i^2F_i$
2 ⊢ 4	3	2	6	4,2	8,4	17,64	35,28
4 ⊢ 6	5	4	20	2,2	8,8	4,48	19,36
6 ⊢ 8	7	7	49	0,2	1,4	0,04	0,28
8 ⊢ 10	9	4	36	1,8	7,2	3,24	12,96
10 ⊢ 12	11	3	33	3,8	11,4	14,44	43,32
		20	144		37,2		111,20

Introdução à teoria da amostragem

Introdução à teoria da amostragem

Objetivo:

Determinar o espaço amostral, os eventos desse espaço e calcular o número de elementos destes conjuntos.

Experimento aleatório (E)

Experimentos aleatórios são aqueles que, mesmo repetidos várias vezes sob condições semelhantes, **apresentam resultados imprevisíveis**.

Exemplo:

Em uma afirmação do tipo: "é provável que meu time ganhe a partida de hoje" pode resultar:

- Que o time perca.
- Que o time ganhe.
- Que o time empate.

O resultado é imprevisível e depende do acaso. Fenômenos como esses são chamados **fenômenos aleatórios** ou **experimentos aleatórios**.

Espaço amostral (S)

É o conjunto de todos os possíveis resultados de um experimento aleatório (E). Indicamos o espaço amostral por S e o número de elementos de S por n(S).

Exemplo:

E: jogar um dado cúbico e observar o número da face de cima:

S = {1, 2, 3, 4, 5, 6}

n(S) = 6

E: jogar uma moeda e observar o resultado:

S = {cara, coroa}

n(S) = 2

E: lançar duas moedas e observar o resultado na face de cada uma:

S = {(cara, cara), (cara, coroa), (coroa, cara), (coroa, coroa)}

n(S) = 4

Evento

É qualquer subconjunto do espaço amostral S de um experimento aleatório (E).

Exemplo:

E: lançar um dado cúbico e observar o número da face de cima:

S = {1, 2, 3, 4, 5, 6}

n(S) = 6

Exemplos de eventos de S:

A: sair número par: {2, 4, 6}

n(A) = 3

B: sair número primo: {2, 3, 5}

n(B) = 3

Combinações de eventos

Uma urna contém 3 bolas pretas e 3 bolas vermelhas. Dessa urna são retiradas, sucessivamente, 3 bolas.

Determine o espaço amostral.

Determine os eventos.

A: as 3 bolas têm a mesma cor

B: exatamente 2 das bolas são pretas

C: as 3 bolas são vermelhas

D: o número de bolas pretas é igual ao número de bolas vermelhas

Solução:

S = {(P, P, P), (P, P, V), (P, V, P), (P, V, V), (V, P, P), (V, P, V), (V, V, P), (V, V, V)}

n(S) = 8

A = {(P, P, P), (V, V, V)}

n(A) = 2

B = {(P, P, V), (P, V, P), (V, P, P)}

n(B) = 3

C = {(V, V, V)}

n(C) = 1

D = { }

n(D) = 0

O conjunto vazio é chamado evento impossível.
Quando o evento coincide com o espaço amostral, ele é chamado evento certo.

Combinações de eventos

União de dois eventos

Sejam **A** e **B** dois eventos, então **A U B** é um evento que ocorrerá se, e somente se, **A** ou **B** ocorrem.

Considere o exemplo:

E: lançamento um dado cúbico e observação do número voltado para cima

S = {1, 2, 3, 4, 5, 6}

A: ocorrência de um número ímpar: **{1, 3, 5}**

B: ocorrência de um número par primo: **{2}**

Logo, A U B: ocorrência de um número ímpar ou um número par primo:

A U B = {1, 2, 3, 5}

Combinações de eventos

União de dois eventos

Sejam **A** e **B** dois eventos, então **A ∩ B** é um evento que ocorrerá se, e somente se, **A** e **B** ocorrem simultaneamente.

Observação: em particular, se **A ∩ B = Ø** , **A** e **B** são chamados mutuamente exclusivos.

Considere o exemplo:

E: lançamento um dado cúbico e observação do número voltado para cima

S = {1, 2, 3, 4, 5, 6}

A: ocorrência de um número par: **{2, 4, 6}**

B: ocorrência de um número múltiplo de 4: **{4}**

Logo, A ∩ B: ocorrência de um número par e múltiplo de 4:

A ∩ B = {4}

Combinações de eventos

Complementar de um evento

Dado um evento **A**, o conjunto formado pelos elementos de S que não pertencem a **A** se chama evento complementar de **A** em relação a S e indica-se por $\bar{A}$.

Considere o exemplo:

E: lançamento um dado cúbico e observação do número voltado para cima

S = {1, 2, 3, 4, 5, 6}

A: ocorrência de um número par: **{2, 4, 6}**

$\bar{A} = \{1, 3, 5\}$

$A \cup \bar{A} = \{1, 2, 3, 4, 5, 6\} = S$

$A \cap \bar{A} = \emptyset$

Fatorial, permutação, arranjo e combinação

Fatorial

O fatorial de um número natural n é representado por **n!** (lê-se n fatorial), em que:

n! = n . (n – 1) . (n – 2) . ... . 3 . 2 . 1, para n ≥ 2

1! = 1

0! = 1

Exemplos:

4! = 4 . 3 . 2 . 1 = 24

7! = 7 . 6 . 5 . 4 . 3 . 2 . 1 = 5040

n! = n . (n – 1)! ou n . (n – 1) . (n – 2)! ...

Exemplo: simplificar as expressões:

$$\frac{8!}{5!.3!} = \frac{8.7.6.\cancel{5!}}{\cancel{5!}.3!} = \frac{8.7.6}{3.2.1} = 56$$

$$\frac{(n+1)!}{n!} = \frac{(n+1).\cancel{n!}}{\cancel{n!}} = n+1$$

Permutação

Dado um conjunto de n elementos, chama-se permutação simples dos n elementos, qualquer sequência – agrupamento ordenado – desses n elementos.

$\mathbf{P_n = n!}$

Exemplos:

Dados três cartões coloridos: Amarelo (A), Laranja (L) e Vemelho (V). Quantas sequências diferentes podemos obter ao enfileirar os três cartões?

$\mathbf{P_n = 3! = 3 . 2 . 1 = 6}$

(A,L,V) (A,V,L) (L,A,V) (L,V,A) (V,A,L) (V,L,A)

Um restaurante funciona cinco dias por semana. Elabrou, portanto, cinco menus diferentes, designados como M1, M2, M3, M4 e M5. De quantas maneiras é possível escolher esses menus de forma que não haja repetição na semana?

$\mathbf{P_n = 5! = 5 . 4 . 3 . 2 . 1 = 120}$

Arranjo

Chama-se arranjo simples de n elementos distintos tomados p a p ($p \leq n$), todo agrupamento **ordenado** formado por p elementos escolhidos entre os n elementos dados. *(muda a ordem, muda o grupo: AB ≠ BA)*

Arranjo Simples

Considere um conjunto com **n** elementos distintos. Qualquer sequência de **p** desses elementos (todos distintos) é chamada de Arranjo Simples ($0 \leq p \leq n$, com n e p naturais). Dizemos arranjo simples de n elementos tomados p a p, e simbolizamos por $A_{n,p}$

Esse arranjo simples pode ser calculado da seguinte forma:

$$A_{n,p} = \frac{n!}{(n-p)!}$$

Arranjos com a totalidade dos elementos, ou seja, $A_{n,n}$ são as permutações de n elementos.

$$A_{n,n} = P_n = n!$$

Exemplo:

Em uma urna de sorteio de prêmios existem dez bolas enumeradas de 0 a 9. Determine o número de possibilidades existentes num sorteio cujo prêmio é formado por uma sequência de 6 algarismos.

$$A_{10,6} = \frac{n!}{(n-p)!} = \frac{10!}{(10-6)!} = \frac{10!}{4!} = \frac{10.9.8.7.6.5.\cancel{4!}}{\cancel{4!}} = 10.9.8.7.6.5 = 151200$$

Combinação

Denominam-se combinações simples de n elementos tomados p a p ($p \leq n$), aos diferentes subconjuntos que contêm p elementos, **sem referência à ordem**. *(muda a ordem, o grupo permanece o mesmo: AB = BA)*

Uma combinação simples é representada da seguinte forma:

$$C_{n,p} = \frac{n!}{p!\,(n-p)!}$$

Exemplo:

Dado o conjunto de 4 elementos {a,b,c,d}, quantas combinações formadas com 2 elementos são possíveis?

$$C_{4,2} = \frac{n!}{p!\,(n-p)!} = \frac{4!}{2!\,(4-2)!} = \frac{4\,.\,3\,.\,\cancel{2!}}{2!\,.\,\cancel{2!}} = \frac{12}{2} = 6$$

{a,b} {a,c} {a,d} {b,c} {b,d} {c,d}

$C_{n,p}$ ou $\binom{n}{p}$ são formas de indicar o número de combinações de n elementos p a p, com $n \geq p$.

$\binom{n}{p}$ também é chamado **coeficiente binomial** porque é utilizado em uma fórmula matemática chamada **teorema binomial**.

Exercícios

Fatorial: Calcule os seguintes fatoriais:

$4!$

$$\frac{10!}{8!}$$

Permutação: Dados cinco cartões coloridos, quantas sequências diferentes podemos obter ao enfileirar os cinco cartões?

Arranjo: Suponha que, em uma corrida de oito cavalos, você esteja tentando acertar a ordem de chegada dos três primeiros finalistas, sem nada saber sobre os cavalos. Quantas finais são possíveis?

Combinação: Sabe-se que um júri foi formado por 7 pessoas, selecionadas de um grupo de 21 pessoas. Neste caso, temos um agrupamento de ordem 7 (as 7 pessoas que formam o júri). A ordem de escolha dessas 7 pessoas não muda o grupo. Portanto, quantas combinações são possíveis?

Quantos jogos diferentes (com seis números em cada jogo) uma pessoa precisaria fazer para ter 100% de certeza de ganhar o prêmio em um determinado sorteio da Mega Sena? (São sorteados 6 números dentre 60 em cada sorteio.)

Probabilidade de um evento

Probabilidade de ocorrência de um evento

Dado um experimento aleatório, sendo S o seu espaço amostral, vamos admitir que todos os elementos de S tenham a mesma chance de acontecer, chamamos de probabilidade de um evento A o número real P(A), tal que:

$$P(A) = \frac{n(A)}{n(S)}$$

Em que:

n(A) = número de elementos do evento A

n(S) = número de elementos do espaço amostral S

Probabilidade de ocorrência de um evento

Considerando o lançamento de um dado, determine a probabilidade de ocorrer na face superior:

Um número par

Solução:

Temos que: S = {1, 2, 3, 4, 5, 6}, logo n(S) = 6

A = {2, 4, 6}, logo n(A) = 3

Então: P(A) = $\frac{3}{6} = \frac{1}{2}$ ou 50%

Probabilidade de ocorrência de um evento

Considerando o lançamento de um dado, determine a probabilidade de ocorrer na face superior:

O número 2

Solução:

Temos que: S = {1, 2, 3, 4, 5, 6}, logo n(S) = 6

B = {2}, logo n(B) = 1

Então: P(B) = $\frac{1}{6}$, ou 16,67%

Probabilidade de ocorrência de um evento

Considerando o lançamento de um dado, determine a probabilidade de ocorrer na face superior:

Um número menor ou igual a 6

Solução:

Temos que: S = {1, 2, 3, 4, 5, 6}, logo n(S) = 6

C = {1, 2, 3, 4, 5, 6}, logo n(C) = 6

Então: P(C) = $\frac{6}{6}$ = 1, ou 100%

Probabilidade de ocorrência de um evento

Considerando o lançamento de um dado, determine a probabilidade de ocorrer na face superior:

Um número maior que 6

Solução:

Temos que: S = {1, 2, 3, 4, 5, 6}, logo n(S) = 6

D = {...}, logo n(D) = 0

Então: P(D) = $\frac{0}{6}$ = 0, ou 0%

Eventos complementares

Sabemos que um evento pode ocorrer ou não. Sendo **p** a probabilidade de que ele ocorra (**sucesso**) e **q** a probabilidade de que ele não ocorra (**insucesso**), então: **p + q = 1** ou **q = 1 – p**:

Exemplo: vimos que, no lançamento de um dado, a probabilidade de ocorrer o número 2 na face superior é $\frac{1}{6}$ ou 16,67%.

Logo, a probabilidade de não tirar 2 no lançamento é $1-\frac{1}{6}=\frac{5}{6}$ ou 83,33%.

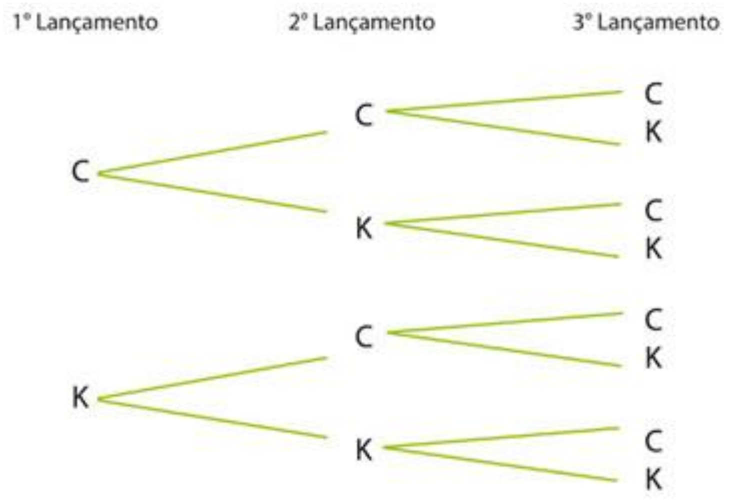
Exercício resolvido 1

Uma moeda é lançada 3 vezes sucessivamente. Qual a probabilidade de obtermos:

- Resultados iguais: 3 vezes cara ou 3 vezes coroa
- Pelo menos uma cara
- Exatamente uma cara
- Número de coroas maior que o número de caras

Exercício resolvido 1

- **Resultados iguais: 3 vezes cara (C) ou 3 vezes coroa (K)**



Solução:

$S = \{(CCC), (CCK), (CKC), (CKK), (KCC), (KCK), (KKC), (KKK)\}$

$n(S) = 8$

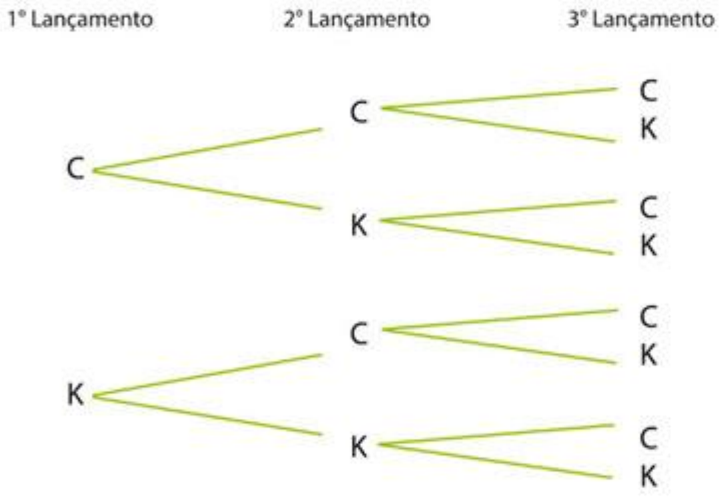
$A = \{(CCC), (KKK)\}$

$n(A) = 2$

$\text{Logo}, P(A) = \frac{2}{8} = \frac{1}{4} = 0{,}25 = 25\%$

Exercício resolvido 1

- **Pelo menos uma cara (C)**



Solução:

$S = \{(CCC), (CCK), (CKC), (CKK), (KCC), (KCK), (KKC), (KKK)\}$

$n(S) = 8$

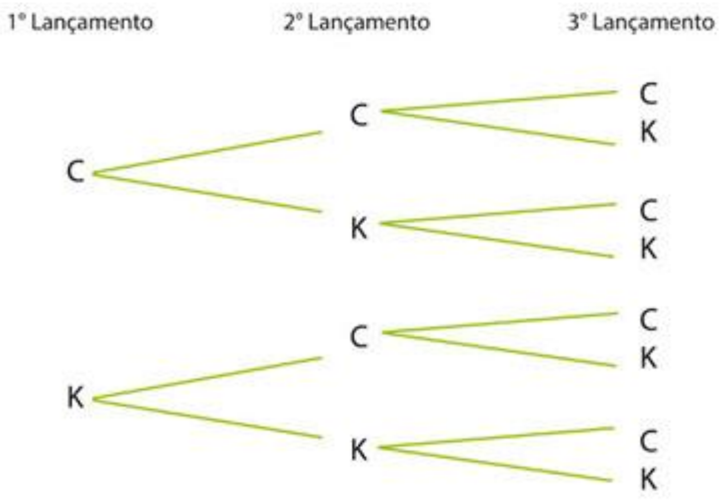
$B = \{(CCC), (CCK), (CKC), (CKK), (KCC), (KCK), (KKC)\}$

$n(B) = 7$

$\text{Logo, } P(B) = \frac{7}{8} = 0{,}875 = 87{,}5\%$

Exercício resolvido 1

- **Exatamente uma cara (C)**



Solução:

$S = \{(CCC), (CCK), (CKC), (CKK), (KCC), (KCK), (KKC), (KKK)\}$

$n(S) = 8$

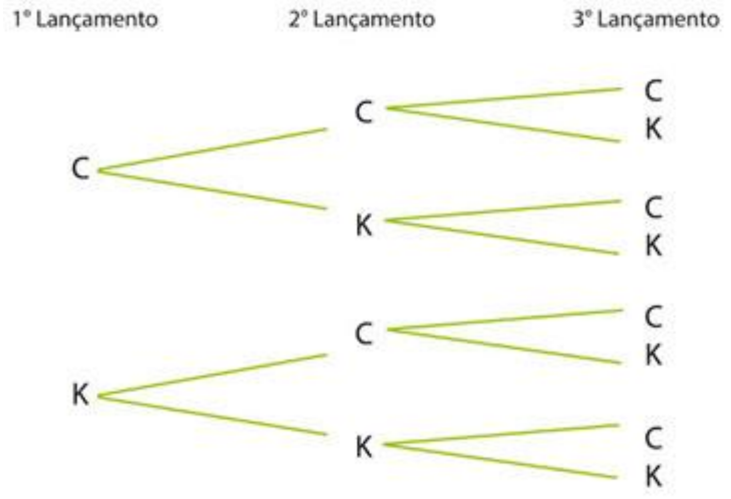
$C = \{(CKK), (KCK), (KKC)\}$

$n(C) = 3$

Logo, $P(C) = \dfrac{3}{8} = 0{,}375 = 37{,}5\%$

Exercício resolvido 1

- **Número de coroas (K) maior que o número de caras (C)**



Solução:

$S = \{(CCC), (CCK), (CKC), (CKK), (KCC), (KCK), (KKC), (KKK)\}$

$n(S) = 8$

$D = \{(CKK), (KCK), (KKC), (KKK)\}$

$n(D) = 4$

$\text{Logo, } P(D) = \frac{4}{8} = \frac{1}{2} = 0{,}50 = 50\%$

Exercício resolvido 2

Uma equipe de **doze pessoa**s é formada por **nove homens** e **três mulheres**. Dessas pessoas, duas serão sorteadas para compor uma comissão. Qual é a probabilidade de a comissão ser formada por:

- Duas mulheres
- Dois homens
- Um homem e uma mulher

Solução: Vamos, primeiramente, calcular o número de elementos do espaço amostral S; para isso, devemos considerar um grupo de 12 pessoas, do qual serão retirados 2 elementos, não importando a ordem, o que corresponde ao número de combinações de 12, tomados 2 a 2:

$$C_{n,p} = \frac{n!}{p!\,(n-p)!}$$

$$n(S) = C_{12,2} = \frac{12!}{(12-2)!\;2!} = \frac{12!}{10!\;2!} = \frac{10\cdot 11\cdot 10!}{10!\;2!} = \frac{12\cdot 11}{2\cdot 1} = \frac{132}{2} = 66$$

Exercício resolvido 2

- **Comissões formadas por 2 mulheres, de um total de 3 mulheres:**

$$n(A) = C_{3,2} = \frac{3!}{(3-2)!\,2!} = \frac{3!}{1!\,2!} = \frac{6}{2} = 3$$

$$Logo, P(A) = \frac{n(A)}{n(S)} = \frac{3}{66} = \frac{1}{22} = 0{,}045 = 4{,}5\%$$

Exercício resolvido 2

- **Comissões formadas por 2 homens, de um total de 9 homens:**

$$n(B) = C_{9,2} = \frac{9!}{(9-2)!\,2!} = \frac{9\,.\,8\,.\,7!}{7!\,2!} = \frac{72}{2} = 36$$

$$Logo\ P(B) = \frac{n(B)}{n(S)} = \frac{36}{66} = \frac{6}{11} = 0{,}545 = 54{,}5\%$$

Exercício resolvido 2

- **Comissões formadas por 1 homem (de um total de 9) e 1 mulher (de um total de 3):**

$$n(C) = C_{9,1} \,.\, C_{3,1} = \frac{9!}{(9-1)!\,1!} \cdot \frac{3!}{(3-1)!\,1!} = \frac{9!}{8!} \cdot \frac{3!}{2!} = \frac{9\,.\,8!}{8!} \cdot \frac{3\,.\,2!}{2!} = 9\,.\,3 = 27$$

$$Logo\ P(C) = \frac{n(C)}{n(S)} = \frac{27}{66} = \frac{9}{22} = 0{,}409 = 40{,}9\%$$