

Análise exploratória de dados

Parte 5

Prof.: Eduardo Vargas Ferreira

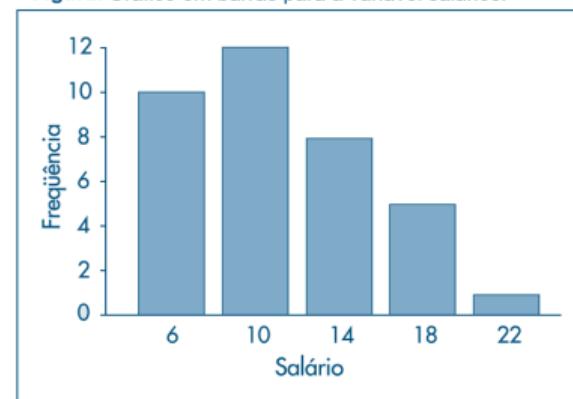


Tabelas e gráficos

- Vimos que as tabelas e gráficos, muitas vezes, fornecem mais informações sobre o comportamento dos dados do que a própria tabela original.

Classe de salários	Ponto médio	Frequência
4,00 ← 8,00	6	10
8,00 ← 12,00	10	12
12,00 ← 16,00	14	8
16,00 ← 20,00	18	5
20,00 ← 24,00	22	1
Total	-	36

Figura: Gráfico em barras para a variável salários.

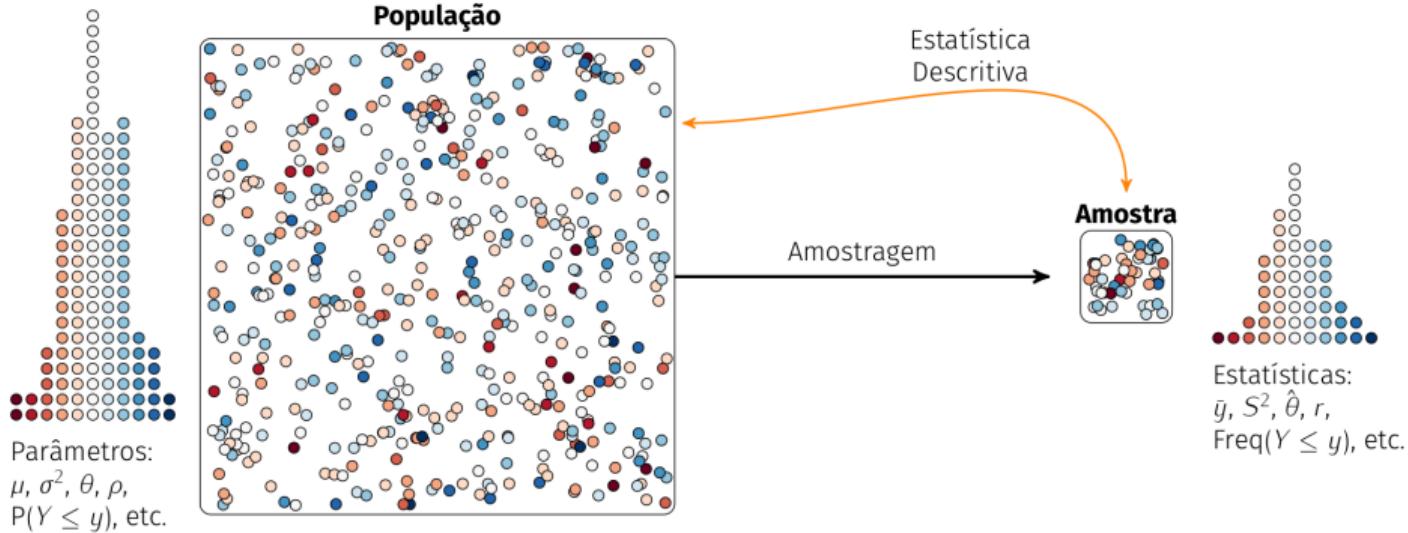


Fonte: Estatística Básica (Bussab e Morettin, 2017)

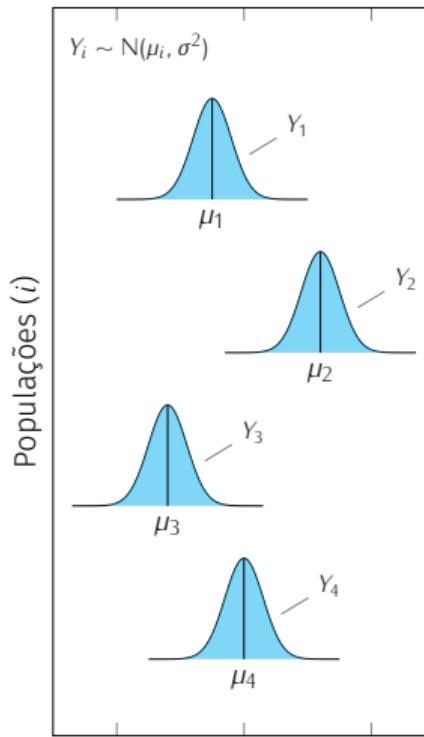
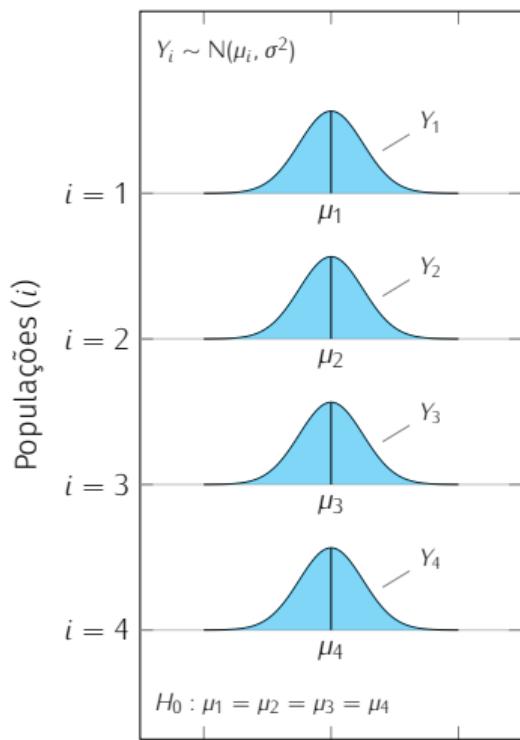
- Entretanto, em certos casos, queremos resumir ainda mais estes dados, fornecendo um ou alguns valores que sejam representativos da série toda.

Medidas-Resumo

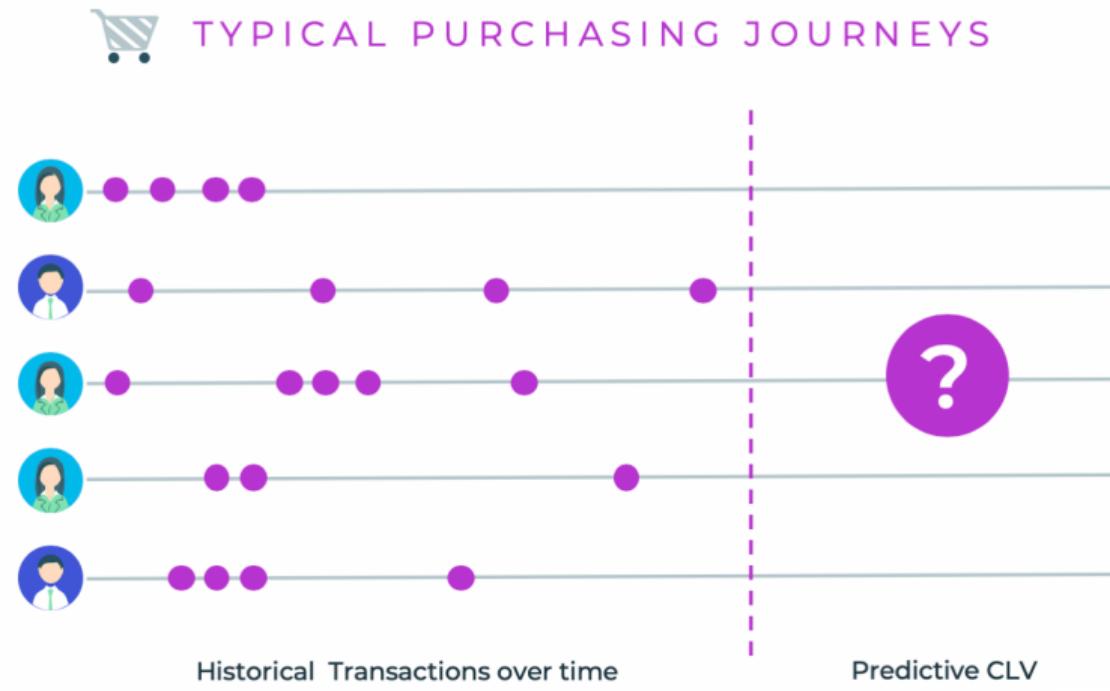
Medidas-Resumo



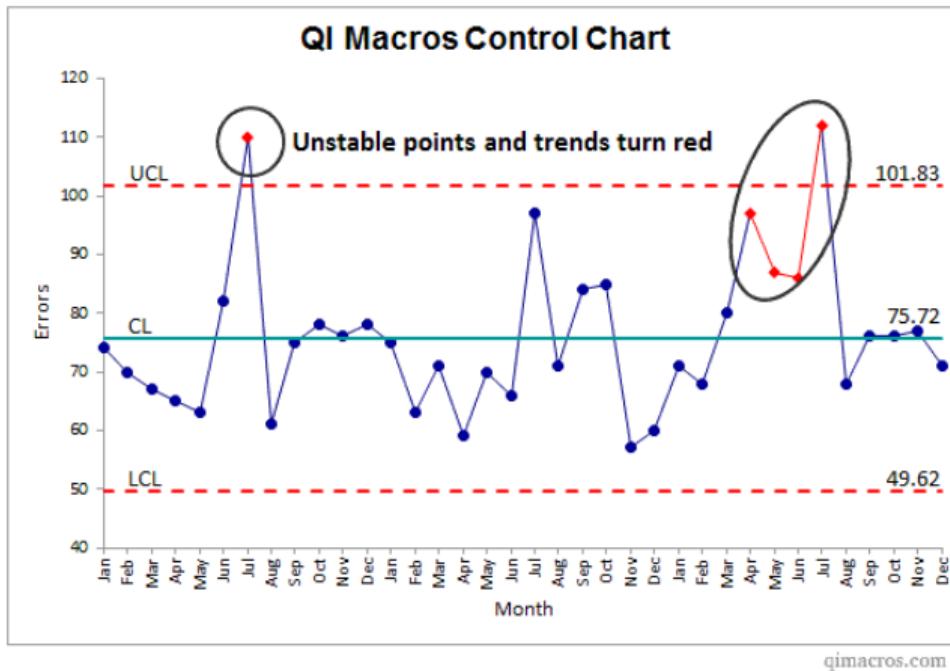
Exemplo: testes de hipóteses



Exemplo: customer lifetime value

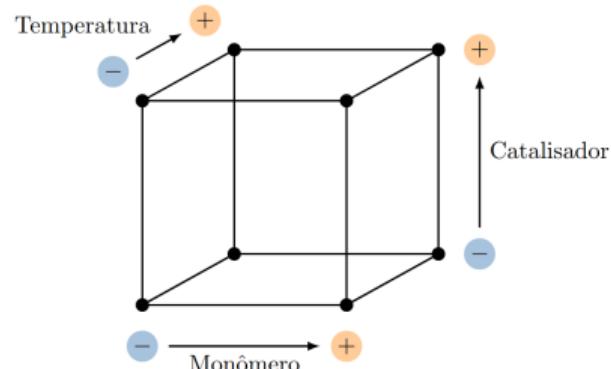


Exemplo: control charts



Exemplo: repetição do mesmo experimento

Temperatura (°C)	Monômero (g)	Catalisador (g)	Rendimento (%)	Massa molar (g/mol)
80	25	2	82	11000
100	25	2	79	13000
80	50	2	86	9000
80	25	4	82	11000
100	50	2	92	9000
100	25	4	97	5000
80	50	2	96	7000
100	50	4	81	10000



Exemplo: idade média dos clientes

Pergunta 1

Idade média dos clientes?

Pergunta 2

Duração média das refeições?

Pergunta 3

Haverá pratos veganos?



Exemplo: duração das refeições

Pergunta 1

Idade média dos clientes?

Pergunta 2

Duração média das refeições?

Pergunta 3

Haverá pratos veganos?



Exemplo: pratos veganos

Pergunta 1

Idade média dos clientes?

Pergunta 2

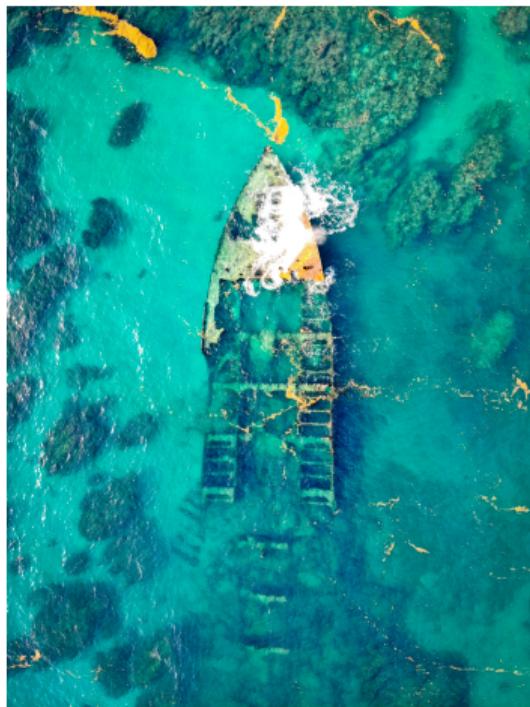
Duração média das refeições?

Pergunta 3

Haverá pratos veganos?



Exemplo: investigando o naufrágio do Titanic

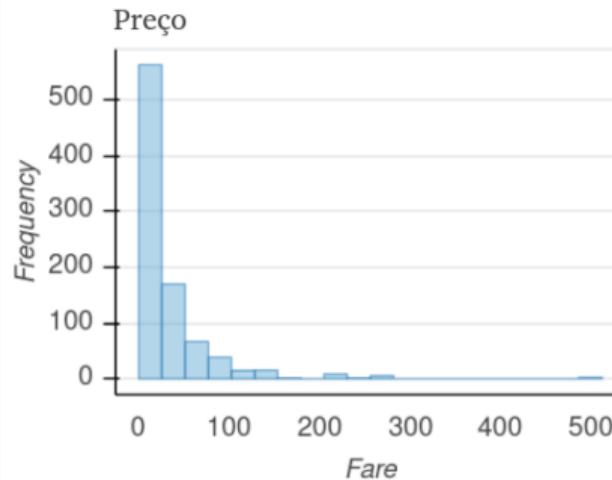


Dataset Statistics

Number of Variables	12
Number of Rows	891
Missing Cells	866
Missing Cells (%)	8.1%
Duplicate Rows	0
Duplicate Rows (%)	0.0%
Total Size in Memory	315.0 KB
Average Row Size in Memory	362.1 B
Numerical Variables	7
Categorical Variables	5

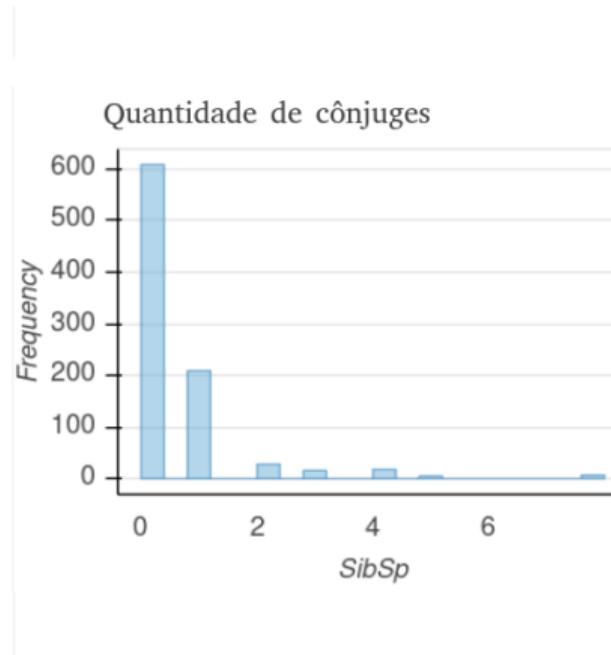
Exemplo: investigando o naufrágio do Titanic

Distinct Count	248
Unique (%)	27.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	13.9 KB
Mean	32.2042
Minimum	0
Maximum	512.3292
Zeros	15
Zeros (%)	1.7%



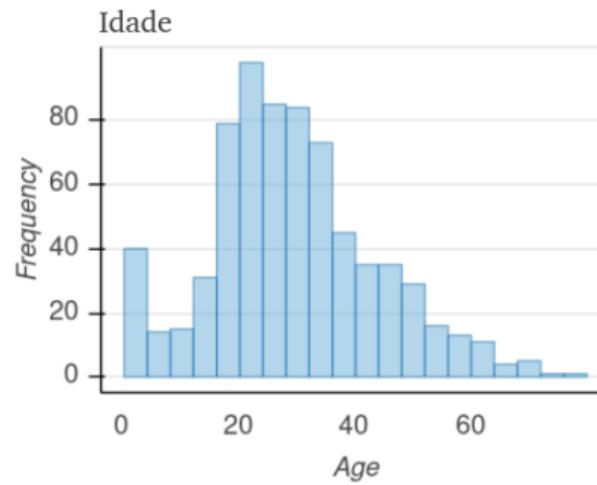
Exemplo: investigando o naufrágio do Titanic

Distinct Count	7
Unique (%)	0.8%
Missing	0
Missing (%)	0.0%
Infinite	0
Infinite (%)	0.0%
Memory Size	13.9 KB
Mean	0.523
Minimum	0
Maximum	8
Zeros	608
Zeros (%)	68.2%



Exemplo: investigando o naufrágio do Titanic

Distinct Count	88
Unique (%)	12.3%
Missing	177
Missing (%)	19.9%
Infinite	0
Infinite (%)	0.0%
Memory Size	11.2 KB
Mean	29.6991
Minimum	0.42
Maximum	80
Zeros	0
Zeros (%)	0.0%



Tabelas e gráficos

Tabela: informações socioeconómicas de 20 funcionários da empresa.

ID	Estado civil	Grau de instrução	Nº de filhos	Salário*	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	-	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	interior
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	-	5,73	20	10	capital
5	solteiro	superior	3	6,26	40	07	interior
6	casado	ensino fundamental	-	6,66	28	00	outra
7	solteiro	ensino fundamental	-	6,86	41	00	interior
8	solteiro	ensino fundamental	2	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	outra
10	solteiro	ensino médio	1	7,44	23	06	interior
11	casado	ensino médio	2	8,12	33	10	interior
12	solteiro	ensino fundamental	5	8,46	27	06	interior
13	solteiro	ensino médio	3	8,74	37	11	interior
14	casado	ensino médio	4	8,95	44	10	interior
15	casado	ensino médio	-	9,13	30	11	capital
16	solteiro	superior	4	9,35	38	06	capital
17	casado	superior	4	9,77	31	10	interior
18	casado	superior	-	9,80	39	08	outra
19	solteiro	ensino fundamental	3	10,53	25	10	interior
20	solteiro	ensino médio	2	10,76	37	08	capital

* número de salários mínimos

Fonte: dados hipotéticos

É mais comum se ter quantos filhos?

Tabela: informações socioeconómicas de 20 funcionários da empresa.

ID	Estado civil	Grau de instrução	Nº de filhos	Salário*	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	-	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	interior
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	-	5,73	20	10	capital
5	solteiro	superior	3	6,26	40	07	interior
6	casado	ensino fundamental	-	6,66	28	00	outra
7	solteiro	ensino fundamental	-	6,86	41	00	interior
8	solteiro	ensino fundamental	2	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	outra
10	solteiro	ensino médio	1	7,44	23	06	interior
11	casado	ensino médio	2	8,12	33	10	interior
12	solteiro	ensino fundamental	5	8,46	27	06	interior
13	solteiro	ensino médio	3	8,74	37	11	interior
14	casado	ensino médio	4	8,95	44	10	interior
15	casado	ensino médio	-	9,13	30	11	capital
16	solteiro	superior	4	9,35	38	06	capital
17	casado	superior	4	9,77	31	10	interior
18	casado	superior	-	9,80	39	08	outra
19	solteiro	ensino fundamental	3	10,53	25	10	interior
20	solteiro	ensino médio	2	10,76	37	08	capital

* número de salários mínimos

Fonte: dados hipotéticos

Qual o salário médio dos empregados?

Tabela: informações socioeconómicas de 20 funcionários da empresa.

ID	Estado civil	Grau de instrução	Nº de filhos	Salário*	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	-	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	interior
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	-	5,73	20	10	capital
5	solteiro	superior	3	6,26	40	07	interior
6	casado	ensino fundamental	-	6,66	28	00	outra
7	solteiro	ensino fundamental	-	6,86	41	00	interior
8	solteiro	ensino fundamental	2	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	outra
10	solteiro	ensino médio	1	7,44	23	06	interior
11	casado	ensino médio	2	8,12	33	10	interior
12	solteiro	ensino fundamental	5	8,46	27	06	interior
13	solteiro	ensino médio	3	8,74	37	11	interior
14	casado	ensino médio	4	8,95	44	10	interior
15	casado	ensino médio	-	9,13	30	11	capital
16	solteiro	superior	4	9,35	38	06	capital
17	casado	superior	4	9,77	31	10	interior
18	casado	superior	-	9,80	39	08	outra
19	solteiro	ensino fundamental	3	10,53	25	10	interior
20	solteiro	ensino médio	2	10,76	37	08	capital

* número de salários mínimos

Fonte: dados hipotéticos

Existe discrepância no salário, eles variam muito?

Tabela: informações socioeconómicas de 20 funcionários da empresa.

ID	Estado civil	Grau de instrução	Nº de filhos	Salário*	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	-	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	interior
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	-	5,73	20	10	capital
5	solteiro	superior	3	6,26	40	07	interior
6	casado	ensino fundamental	-	6,66	28	00	outra
7	solteiro	ensino fundamental	-	6,86	41	00	interior
8	solteiro	ensino fundamental	2	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	outra
10	solteiro	ensino médio	1	7,44	23	06	interior
11	casado	ensino médio	2	8,12	33	10	interior
12	solteiro	ensino fundamental	5	8,46	27	06	interior
13	solteiro	ensino médio	3	8,74	37	11	interior
14	casado	ensino médio	4	8,95	44	10	interior
15	casado	ensino médio	-	9,13	30	11	capital
16	solteiro	superior	4	9,35	38	06	capital
17	casado	superior	4	9,77	31	10	interior
18	casado	superior	-	9,80	39	08	outra
19	solteiro	ensino fundamental	3	10,53	25	10	interior
20	solteiro	ensino médio	2	10,76	37	08	capital

* número de salários mínimos

Fonte: dados hipotéticos

Qual a amplitude de idade dos empregados?

Tabela: informações socioeconómicas de 20 funcionários da empresa.

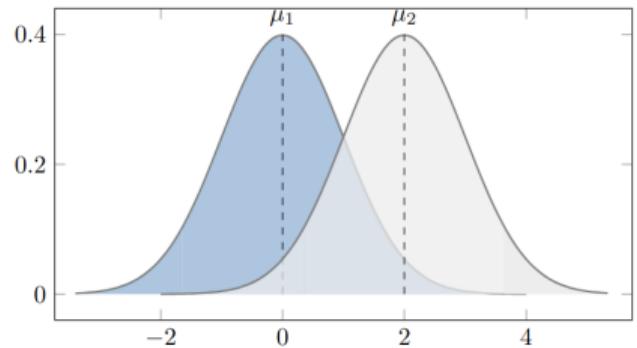
ID	Estado civil	Grau de instrução	Nº de filhos	Salário*	Idade		Região de procedência
					anos	meses	
1	solteiro	ensino fundamental	-	4,00	26	03	interior
2	casado	ensino fundamental	1	4,56	32	10	interior
3	casado	ensino fundamental	2	5,25	36	05	capital
4	solteiro	ensino médio	-	5,73	20	10	capital
5	solteiro	superior	3	6,26	40	07	interior
6	casado	ensino fundamental	-	6,66	28	00	outra
7	solteiro	ensino fundamental	-	6,86	41	00	interior
8	solteiro	ensino fundamental	2	7,39	43	04	capital
9	casado	ensino médio	1	7,59	34	10	outra
10	solteiro	ensino médio	1	7,44	23	06	interior
11	casado	ensino médio	2	8,12	33	10	interior
12	solteiro	ensino fundamental	5	8,46	27	06	interior
13	solteiro	ensino médio	3	8,74	37	11	interior
14	casado	ensino médio	4	8,95	44	10	interior
15	casado	ensino médio	-	9,13	30	11	capital
16	solteiro	superior	4	9,35	38	06	capital
17	casado	superior	4	9,77	31	10	interior
18	casado	superior	-	9,80	39	08	outra
19	solteiro	ensino fundamental	3	10,53	25	10	interior
20	solteiro	ensino médio	2	10,76	37	08	capital

* número de salários mínimos

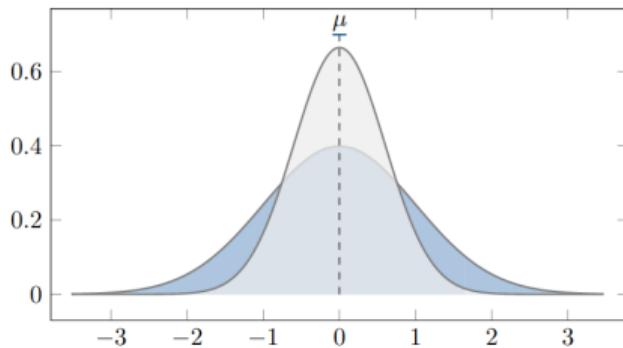
Fonte: dados hipotéticos

Medidas de posição e dispersão

Medidas de posição



Medidas de dispersão



Referências

- Bussab, WO; Morettin, PA. Estatística Básica. São Paulo: Editora Saraiva, 2006 (5^a Edição).
- Magalhães, MN; Lima, ACP. Noções de Probabilidade e Estatística. São Paulo: EDUSP, 2008.

