



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Elementos de Estatística

Segundo Semestre/2018

Resumo dos Dados

- ▶ Já vimos como resumir conjuntos de dados provenientes de variáveis qualitativas e quantitativas utilizando tabelas e gráficos;
- ▶ Para variáveis aleatórias quantitativas pode-se utilizar, além das tabelas e gráficos, medidas que resumem o conjunto de dados;



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Medidas Resumo

- ▶ **Medidas de Tendência Central:**

- ▶ Média;
- ▶ Mediana;
- ▶ Moda;

- ▶ **Medidas de Dispersão:**

- ▶ Amplitude;
- ▶ Quartis;
- ▶ Variância;
- ▶ Desvio Padrão;
- ▶ Coeficiente de variação.



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Medidas de Tendência Central

- ▶ Medidas em torno das quais as observações se distribuem;
- ▶ As medidas de tendência central, ou medidas de posição, mais estudadas são:
 - ▶ Média;
 - ▶ Mediana;
 - ▶ Moda.



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Média

- ▶ Média (\bar{x}_{obs}):
- ▶ Considere a amostra $(x_1, x_2, x_3, \dots, x_n)$, a média observada é dada por:
- ▶
$$\bar{x}_{obs} = \frac{x_1 + x_2 + \dots + x_n}{n} = \frac{1}{n} \sum_{i=1}^n x_i.$$



Exemplo 1 – média

- Suponha que parafusos a serem utilizados em tomadas elétricas são embalados em caixas rotuladas como contendo 100 unidades. Em uma construção, 10 caixas de um lote tiveram o número de parafusos contados, fornecendo os valores 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, o número médio de parafusos será dado por:

$$\bar{x}_{obs} = \frac{98 + 102 + 100 + \dots + 95 + 99 + 100}{10}$$

$$\bar{x}_{obs} = \frac{986}{10} = 98,6 \text{ parafusos}$$



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Mediana

- ▶ É o valor que ocupa a posição central dos dados ordenados.
- ▶ Para encontrar a mediana deve-se ordenar os dados do menor para o maior;
- ▶ A mediana relativa a um conjunto de dados pode ser definida como:

$$\text{▶ } md_{obs} = \begin{cases} x_{\left(\frac{n+1}{2}\right)}, & \text{se } n \text{ for ímpar} \\ \frac{x_{\left(\frac{n}{2}\right)} + x_{\left(\frac{n}{2}+1\right)}}{2}, & \text{se } n \text{ for par} \end{cases}$$



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – mediana

- ▶ Considere o mesmo conjunto de dados do último exemplo: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, a mediana será dada por:
- ▶ Primeiro colocamos os dados em ordem:

Ordem	1	2	3	4	5	6	7	8	9	10
Número de Parafusos	95	96	97	98	99	99	100	100	100	102

$$\begin{aligned}n = 10 \Rightarrow md_{obs} &= \frac{x_{(\frac{n}{2})} + x_{(\frac{n}{2}+1)}}{2} = \frac{x_{(\frac{10}{2})} + x_{(\frac{10}{2}+1)}}{2} = \\&= \frac{x_{(5)} + x_{(5+1)}}{2} = \frac{x_{(5)} + x_{(6)}}{2} = \frac{99+99}{2} = 99 \text{ parafusos}\end{aligned}$$



Moda

- ▶ A moda de um conjunto de dados é a observação que aparece com maior frequência no conjunto;
- ▶ Um conjunto pode ser unimodal, bimodal ou multimodal;
- ▶ Caso todos os valores tenham a mesma frequência, não é possível determinar a moda do conjunto de valores, conjunto amodal.



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – moda

- ▶ Considere o mesmo conjunto de dados do último exemplo: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, a moda será dada por:
- ▶ Utilizando os dados já ordenados, temos:

Ordem	1	2	3	4	5	6	7	8	9	10
Número de Parafusos	95	96	97	98	99	99	100	100	100	102

- ▶ O valor 100, é o mais frequente, aparecendo 3 vezes, logo:
- ▶ $mo_{obs} = 100$ parafusos



Medidas de Tendência Central

- ▶ Podem ser utilizadas conjuntamente para auxiliar a análise dos dados;
- ▶ Ou, em determinadas situações uma delas pode ser mais conveniente do que a outra:
 - ▶ No caso de haver um ou mais dados que se afastam do geral das observações (valores discrepantes ou outliers) a média passa a ser uma medida de tendência central inadequada, sendo a mediana uma medida mais indicada.
 - ▶ No caso de conjuntos multimodais ou amodais, a média ou a mediana são mais indicadas para representar a tendência central.



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 2 – valor atípico

- ▶ Considere o mesmo conjunto de dados do último exemplo, porém considerando que uma das caixas com 1 parafusos na realidade tivesse 45 parafusos: 98, 102, 100, 45, 99, 97, 96, 95, 99, 100. Para essas caixas, a moda será dado por:
- ▶ Utilizando os dados já ordenados, temos:

Ordem	1	2	3	4	5	6	7	8	9	10
Número de Parafusos	45	95	96	97	98	99	99	100	100	102

- ▶ $\bar{x}_{obs} = \frac{98+102+\dots+99+100}{10} = 93,1$ parafusos
- ▶ $md_{obs} = \frac{x_5+x_6}{2} = \frac{98+99}{2} = 98,5$ parafusos
- ▶ $mo_{obs_1} = 99; mo_{obs_2} = 100$ parafusos



Exemplo 2 – valor atípico

- ▶ $\bar{x}_{obs} = \frac{98+102+\dots+99+100}{10} = 93,1$ parafusos
- ▶ $md_{obs} = \frac{x_5+x_6}{2} = \frac{98+99}{2} = 98,5$ parafusos
- ▶ $mo_{obs_1} = 95; mo_{obs_2} = 99; mo_{obs_3} = 100$ parafusos
- ▶ Ao inserir um valor atípico no conjunto de dados utilizado, pode perceber que:
- ▶ A média foi bastante influenciada, passou de 98,6 para 93,1 parafusos, se tornando inadequada;
- ▶ O conjunto passou a ser multimodal, o que também torna a utilização da moda inapropriada;
- ▶ A mediana foi a medida que menos sofreu influência do valor atípico, sendo a mais adequada nesse caso.



Medidas de Tendência Central

- ▶ Nem sempre se trabalha, ou se tem interesse, no conjunto de dados originais, obtidos por medidas diretas;
- ▶ Comumente, o interesse está em uma função dessas medidas diretas, sendo necessário considerar os valores originais multiplicados ou acrescidos de constantes para obter um novo conjunto de valores (medidas indiretas);
- ▶ Veremos, a seguir, como as medidas de tendência central se alteram e como podem ser obtidas a partir do conjunto original (medidas diretas).



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 3 – medidas indiretas

- ▶ Nas caixas de parafusos do exemplo 1, vamos admitir um custo de c por parafuso e de e pela embalagem da caixa. Desejamos calcular as medidas de posição do *custo total* (T), definido como a soma dos custos dos parafusos e da embalagem. Iniciamos, calculando as novas medidas de tendência central apenas para o *custo líquido por caixa* (L), isto é, o custo dos parafusos contidos na caixa sem a embalagem. Temos:
- ▶ $Média\ de\ L = \bar{l}_{obs} = \frac{98c+102c+\dots+100c}{10} = \frac{986c}{10} = 98,6c$
- ▶ $mediana = md_{obs}(L) = \frac{99c+99c}{2} = 99c$
- ▶ $moda = mo_{obs}(L) = 100c$



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 3 – medidas indiretas

- ▶ Vamos incluir agora o custo da embalagem. As caixas custarão:
- ▶ $98c + e, 102c + e, 100c + e, \dots, 99c + e, 100c + e$
- ▶ $Média\ de\ T = \bar{t}_{obs} = \frac{(98c+e)+(102c+e)+\dots+(100c+e)}{10} =$
 $= \frac{986c + 10e}{10} = 98,6c + e$
- ▶ $mediana = md_{obs}(T) = \frac{(99c+e)+(99c+e)}{2} = 99c + e$
- ▶ $moda = mo_{obs}(L) = 100c + e$



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Medidas de Tendência Central

- ▶ Considerando o exemplo 3, pode-se perceber que:
 - ▶ A multiplicação de uma constante c resultou em que as novas medidas de tendência central são as antigas multiplicadas por c ;
 - ▶ O acréscimo por uma quantidade e teve o efeito de somar essa mesma constante às medidas de tendência central.



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exercício 1 – Parte 1

- ▶ Foram coletadas 150 observações de uma variável, representando o *número de vestibulares FUVEST* (um por ano) que um mesmo estudante prestou. Assim, foi observado que 75 estudantes prestaram vestibular FUVEST, uma única vez, e assim por diante. Os dados estão na tabela abaixo:

Nº de vestibulares FUVEST	1	2	3	4
Nº de estudantes	75	47	21	7

- ▶ Calcule as medidas de tendência central da variável número de vestibulares.



Exercício 1 – Parte 2

- ▶ Pode ser de interesse estudar o gasto dos alunos associado com as despesas do vestibular. Para simplificar um pouco a situação, vamos supor que se atribui, para cada aluno, uma despesa fixa de R\$1300, relativa à preparação e mais R\$50 para cada vestibular prestado. De posse dessas informações, vamos calcular as medidas de tendência central da variável D : despesa com vestibular.

Nº de vestibulares FUVEST	1	2	3	4
Nº de estudantes	75	47	21	7



Exercício 2

- ▶ Um estudante está procurando um estágio para o próximo ano. As companhias A e B têm programas de estágios e oferecem uma remuneração por 20 horas semanais com as seguintes características (em salários mínimos):

Companhia	A	B
Média	2,5	2,0
Mediana	1,7	1,9
Moda	1,5	1,9

- ▶ Qual companhia o aluno deverá escolher? Justifique.



Simetria

- ▶ Um conjunto de dados é dito simétrico se os dados se distribuem de igualmente ao redor da média;
- ▶ Pode-se dizer que um conjunto de dados é simétrico quando a média, mediana e moda são dadas pelo mesmo valor;
- ▶ O conhecimento da simetria de um conjunto auxilia a interpretação do mesmo.



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Dados

- ▶ Suponha, por exemplo, que um questionário foi aplicado aos alunos do primeiro ano de uma escola fornecendo as seguintes informações:
 - ▶ Id: identificação do aluno
 - ▶ Turma: turma a que o aluno foi alocado (A ou B)
 - ▶ Sexo: F se feminino, M se masculino
 - ▶ Idade: idade em anos completos
 - ▶ Alt: altura em metros
 - ▶ Peso: peso em quilogramas
 - ▶ Filhos: número de filhos na família
 - ▶ Fuma: hábito de fumas, sim ou não
 - ▶ Toler: tolerância ao cigarro (I – indiferente; P – incomoda pouco; M – incomoda muito)
 - ▶ Exerc: horas de atividade física, por semana
 - ▶ Cine: número de vezes que vai ao cinema por semana
 - ▶ OpCine: opinião a respeito das salas de cinema na cidade (B – regular a boa; M – muito boa)
 - ▶ TV: horas gastas assistindo TV, por semana
 - ▶ OpTV: opinião a respeito da qualidade da programação na TV (R – ruim; M – média; B – boa; e N – não sabe)



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Informações de Questionário Infantil – dados brutos – Tabela 1

Id	Turma	Sexo	Idade	Alt	Peso	Filh	Fuma	Toler	Exer	Cine	OpCne	TV	OpTV
1	A	F	17	1,6	50,5	2	NÃO	P	0	1	B	16	R
2	A	F	18	1,69	55	1	NÃO	M	0	1	B	7	R
3	A	M	18	1,85	72,8	2	NÃO	P	5	2	M	15	R
4	A	M	25	1,85	80,9	2	NÃO	P	5	2	B	20	R
5	A	F	19	1,58	55	1	NÃO	M	2	2	B	5	R
6	A	M	19	1,76	60	3	NÃO	M	2	1	B	2	R
7	A	F	20	1,6	58	1	NÃO	P	3	1	B	7	R
8	A	F	18	1,64	47	1	SIM	I	2	2	M	10	R
9	A	F	18	1,62	57,8	3	NÃO	M	3	3	M	12	R
10	A	F	17	1,64	58	2	NÃO	M	2	2	M	10	R
11	A	F	18	1,72	70	1	SIM	I	10	2	B	8	N
12	A	F	18	1,66	54	3	NÃO	M	0	2	B	0	R
13	A	F	21	1,7	58	2	NÃO	M	6	1	M	30	R
14	A	M	19	1,78	68,5	1	SIM	I	5	1	M	2	N
15	A	F	18	1,65	63,5	1	NÃO	I	4	1	B	10	R
16	A	F	19	1,63	47,4	3	NÃO	P	0	1	B	18	R
17	A	F	17	1,82	66	1	NÃO	P	3	1	B	10	N
18	A	M	18	1,8	58,2	2	NÃO	P	3	4	B	10	R
19	A	F	20	1,6	54,5	1	NÃO	P	3	2	B	5	R
20	A	F	18	1,68	52,5	3	NÃO	M	7	2	B	14	M
21	A	F	21	1,7	60	2	NÃO	P	8	2	B	5	R
22	A	F	18	1,65	58,5	1	NÃO	M	0	3	B	5	R
23	A	F	18	1,57	49,2	1	SIM	I	5	4	B	10	R
24	A	F	20	1,55	48	1	SIM	I	0	1	M	28	R
25	A	F	20	1,69	51,6	2	NÃO	P	8	5	M	4	N
26	A	F	19	1,54	57	2	NÃO	I	6	2	B	5	R
27	B	F	23	1,62	63	2	NÃO	M	8	2	M	5	R
28	B	F	18	1,62	52	1	NÃO	P	1	1	M	10	R
29	B	F	18	1,57	49	2	NÃO	P	3	1	B	12	R
30	B	F	25	1,65	59	4	NÃO	M	1	2	M	2	R
31	B	F	18	1,61	52	1	NÃO	P	2	2	M	6	N
32	B	M	17	1,71	73	1	NÃO	P	1	1	B	20	R
33	B	F	17	1,65	56	3	NÃO	M	2	1	B	14	R
34	B	F	17	1,67	58	1	NÃO	M	4	2	B	10	R
35	B	M	18	1,73	87	1	NÃO	M	7	1	B	25	B
36	B	F	18	1,6	47	1	NÃO	P	5	1	M	14	R
37	B	M	17	1,7	95	1	NÃO	P	10	2	M	12	N
38	B	M	21	1,85	84	1	SIM	I	6	4	B	10	R
39	B	F	18	1,7	60	1	NÃO	P	5	2	B	12	R
40	B	M	18	1,73	73	1	NÃO	M	4	1	B	2	R
41	B	F	17	1,7	55	1	NÃO	I	5	4	B	10	B
42	B	F	23	1,45	44	2	NÃO	M	2	2	B	25	R
43	B	M	24	1,76	75	2	NÃO	I	7	0	M	14	N
44	B	F	18	1,68	55	1	NÃO	P	5	1	B	8	R
45	B	F	18	1,55	49	1	NÃO	M	0	1	M	10	R
46	B	F	19	1,7	50	7	NÃO	M	0	1	B	8	R
47	B	F	19	1,55	54,5	2	NÃO	M	4	3	B	3	R
48	B	F	18	1,6	50	1	NÃO	P	2	1	B	5	R
49	B	M	17	1,8	71	1	NÃO	P	7	0	M	14	R
50	B	M	18	1,83	86	1	NÃO	P	7	0	M	20	B

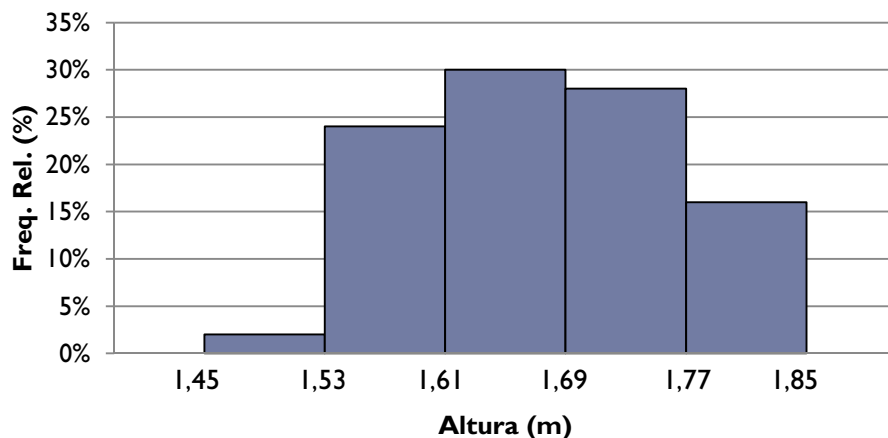


Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Simetria

Altura dos alunos entrevistados (X)

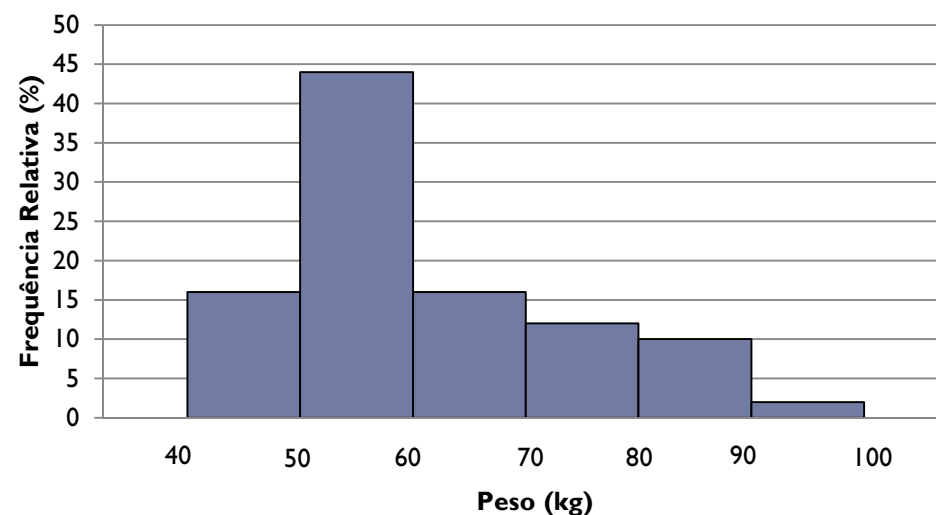


$$\bar{x}_{obs} = 1,67 \text{ metros}$$

$$md_{obs}(X) = 1,67 \text{ metros}$$

$$mo_{obs}(X) = 1,7 \text{ metros}$$

Peso dos alunos entrevistados (Y)



$$\bar{y}_{obs} = 60,2 \text{ kg}$$

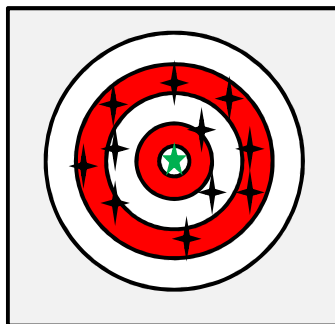
$$md_{obs}(Y) = 57,9 \text{ kg}$$

$$mo_{obs}(Y) = 55 \text{ kg}$$



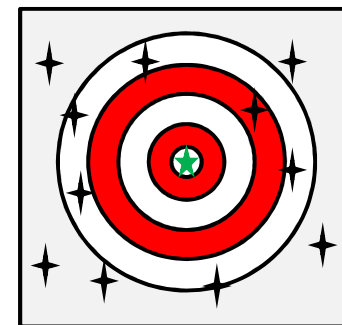
Medidas de Dispersão

- ▶ As medidas de tendência central indicam em torno de qual valor os dados se distribuem;
- ▶ Para dados com pequena dispersão, ou variabilidade, as medidas de tendência central fornecem uma descrição apropriada dos dados;
- ▶ Porém, para dados com uma grande dispersão as medidas de tendência central podem não ser tão apropriadas na descrição dos mesmos.



★ Medida de tendência central

✦ Observações



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Medidas de Dispersão

- ▶ Algumas das medidas de dispersão, ou de variabilidade, mais utilizadas são:
 - ▶ Amplitude;
 - ▶ Quartis;
 - ▶ Variância;
 - ▶ Desvio Padrão;
 - ▶ Coeficiente de variação.



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Amplitude

- ▶ Fornece uma descrição da variabilidade de um conjunto de dados;
- ▶ A amplitude é dada pela diferença entre os valores máximo e mínimo de um conjunto de dados;
- ▶ Assim como a média, a amplitude apresenta uma grande sensibilidade à valores atípicos.



Exemplo 1 – amplitude

- ▶ Considere o mesmo conjunto de dados (caixas de parafusos) utilizado anteriormente: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, a amplitude será dada por:

	mínimo									máximo
Ordem	1	2	3	4	5	6	7	8	9	10
Número de Parafusos	95	96	97	98	99	99	100	100	100	102

$$amplitude_{obs} = 102 - 95 = 7 \text{ parafusos}$$



Quartis

- ▶ Valores que dividem o conjunto de dados em quatro partes contendo 25% dos dados cada:
 - ▶ Primeiro Quartil (Q_1): 25% das observações estão abaixo e os outros 75% das observações estão acima do primeiro quartil.
 - ▶ Segundo Quartil (Q_2) ou mediana (md): 50% das observações estão abaixo e os outros 50% das observações estão acima do primeiro quartil.
 - ▶ Terceiro Quartil (Q_3): 75% das observações estão abaixo e os outros 25% das observações estão acima do terceiro quartil.
- ▶ Assim como para a mediana, o primeiro passo para encontrar o quartil desejado é ordenar os dados;
- ▶ Existem técnicas distintas para encontrar os quartis desejados;
- ▶ A técnica mais simples é calcular a mediana e em seguida utilizar o mesmo pensamento pra calcular os quartis.



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – quartis

- Considere o mesmo conjunto de dados (caixas de parafusos) utilizado anteriormente: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, os quartis serão dados por:

Ordem	1	2	3	4	5	6	7	8	9	10
Número de Parafusos	95	96	97	98	99	99	100	100	100	102

$$md_{obs} = 99$$

Ordem	1	2	3	4	5	md
Número de Parafusos	95	96	97	98	99	99

$$Q_1 = \frac{97 + 98}{2} = 97,5$$

Ordem	md	6	7	8	9	10
Número de Parafusos	99	99	100	100	100	102

$$Q_3 = \frac{100 + 100}{2} = 100$$



Resumo dos 5 números

- ▶ Resumo dado por 5 valores que ajuda a entender a variabilidade e simetria dos dados:
 - ▶ Mínimo;
 - ▶ Primeiro Quartil;
 - ▶ Mediana;
 - ▶ Terceiro Quartil;
 - ▶ Máximo.



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – resumo dos 5 números

- Considere o mesmo conjunto de dados (caixas de parafusos) utilizado anteriormente: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, o resumo de 5 números será dado por:

	$md_{obs} = 99$									
	mínimo									
Ordem	1	2	3	4	5	6	7	8	9	10
Número de Parafusos	95	96	97	98	99	99	100	100	100	102
	máximo									
	$Q_1 = 97,5$									
	$Q_3 = 100$									



Box-Plot

- ▶ Exibe um resumo dos dados de maneira simplificada;
- ▶ O Box-Plot possui informação sobre o resumo dos 5 números;
- ▶ De maneira simplificada informa sobre, entre outras coisas, a variabilidade e simetria dos dados.



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – resumo dos 5 números

- Considere o mesmo conjunto de dados (caixas de parafusos) utilizado anteriormente: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, o resumo de 5 números será dado por:

$$\min = 95$$

$$Q_1 = 97,5$$

$$md_{obs} = 99$$

$$Q_3 = 100$$

$$\max = 102$$



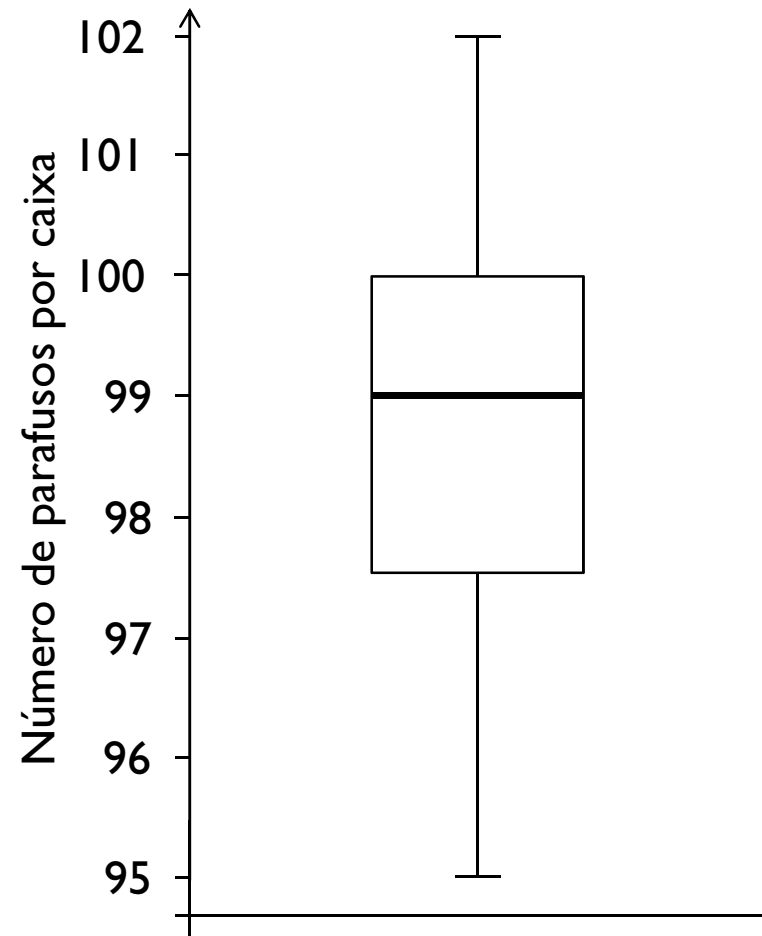
Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – Box-Plot

$min = 95$
 $Q_1 = 97,5$
 $md_{obs} = 99$
 $Q_3 = 100$
 $max = 102$



Exercício 4

- ▶ Considere os dados dos pesos dos 50 alunos entrevistados (dados brutos apresentados no slide 23) divididos por sexo:
- ▶ Calcule o resumo dos 5 números para os pesos das alunas e dos alunos separadamente;
- ▶ Construa dois box-plots (dispostos paralelamente um ao outro) um para o peso das alunas e um para o peso dos alunos;
- ▶ O peso dos alunos entrevistados se comporta de maneira igual para ambos os sexos? Justifique.



Exercício 4

Id	Sexo	Peso	Ordem
18M		58,2	1
6M		60	2
14M		68,5	3
49M		71	4
3M		72,8	5
32M		73	6
40M		73	7
43M		75	8
4M		80,9	9
38M		84	10
50M		86	11
35M		87	12
37M		95	13

Id	Sexo	Peso	Ordem
42F		44	1
8F		47	2
36F		47	3
16F		47,4	4
24F		48	5
29F		49	6
45F		49	7
23F		49,2	8
46F		50	9
48F		50	10
1F		50,5	11
25F		51,6	12
28F		52	13
31F		52	14
20F		52,5	15
12F		54	16
19F		54,5	17
47F		54,5	18
2F		55	19

Id	Sexo	Peso	Ordem
5F		55	20
41F		55	21
44F		55	22
33F		56	23
26F		57	24
9F		57,8	25
7F		58	26
10F		58	27
13F		58	28
34F		58	29
22F		58,5	30
30F		59	31
21F		60	32
39F		60	33
27F		63	34
15F		63,5	35
17F		66	36
11F		70	37



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Variância

- ▶ A variância de um conjunto de dados objetiva quantificar a variabilidade ao redor da média aritmética das observações.
- ▶ A variância de um conjunto de dados é dada por:
- ▶ $var_{obs} = \frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2$
- ▶ A variância possui um inconveniente: se as observações forem medidas em *cm* a variância será dada em cm^2 . A unidade de medida da variância será sempre a unidade de medida das observações elevada ao quadrado.



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 1 – variância

- Considere o mesmo conjunto de dados (caixas de parafusos) utilizado anteriormente: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, a variância será dada por:

$$var_{obs} = \frac{(98 - 98,6)^2 + (102 - 98,6)^2 + \dots + (99 - 98,6)^2 + (100 - 98,6)^2}{10}$$

$$var_{obs} = \frac{(-0,6)^2 + (3,4)^2 + \dots + (0,4)^2 + (1,4)^2}{10} = \frac{40,4}{10} = 4,04$$



Exemplo 1 – variância

i	x_i	$x_i - \bar{x}_{obs}$	$(x_i - \bar{x}_{obs})^2$
1	98	-0,6	0,36
2	102	3,4	11,56
3	100	1,4	1,96
4	100	1,4	1,96
5	99	0,4	0,16
6	97	-1,6	2,56
7	96	-2,6	6,76
8	95	-3,6	12,96
9	99	0,4	0,16
10	100	1,4	1,96
Total			40,40

$$\bar{x}_{obs} = 98,6$$

$$var_{obs} = \frac{40,40}{10} = 4,04$$



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 3 – medidas indiretas

- ▶ Nas caixas de parafusos do exemplo 1, vamos admitir um custo de c por parafuso e de e pela embalagem da caixa. Desejamos calcular a variância do *custo total* (T), definido como a soma dos custos dos parafusos e da embalagem. Iniciamos, calculando a nova variância apenas para o *custo líquido por caixa* (L), isto é, o custo dos parafusos contidos na caixa sem a embalagem.



Exemplo 3 – medidas indiretas

i	x_i	l_i	t_i
1	98	$98c$	$98c + e$
2	102	$102c$	$102c + e$
3	100	$100c$	$100c + e$
4	100	$100c$	$100c + e$
5	99	$99c$	$99c + e$
6	97	$97c$	$97c + e$
7	96	$96c$	$96c + e$
8	95	$95c$	$95c + e$
9	99	$99c$	$99c + e$
10	100	$100c$	$100c + e$
média	98,6	$98,6c$	$98,6c + e$



Exemplo 3 – medidas indiretas

i	l_i	$x_i - \bar{x}_{obs}$	$(x_i - \bar{x}_{obs})^2$
1	$98c$	$-0,6c$	$0,36c^2$
2	$102c$	$3,4c$	$11,56c^2$
3	$100c$	$1,4c$	$1,96c^2$
4	$100c$	$1,4c$	$1,96c^2$
5	$99c$	$0,4c$	$0,16c^2$
6	$97c$	$-1,6c$	$2,56c^2$
7	$96c$	$-2,6c$	$6,76c^2$
8	$95c$	$-3,6c$	$12,96c^2$
9	$99c$	$0,4c$	$0,16c^2$
10	$100c$	$1,4c$	$1,96c^2$
Total			$40,40c^2$

$$\bar{l}_{i_{obs}} = 98,6c$$

$$var_{obs} = \frac{40,40c^2}{10} = 4,04c^2$$



Departamento de Estatística

UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exemplo 3 – medidas indiretas

i	t_i	$x_i - \bar{x}_{obs}$	$(x_i - \bar{x}_{obs})^2$
1	$98c + e$	$-0,6c$	$0,36c^2$
2	$102c + e$	$3,4c$	$11,56c^2$
3	$100c + e$	$1,4c$	$1,96c^2$
4	$100c + e$	$1,4c$	$1,96c^2$
5	$99c + e$	$0,4c$	$0,16c^2$
6	$97c + e$	$-1,6c$	$2,56c^2$
7	$96c + e$	$-2,6c$	$6,76c^2$
8	$95c + e$	$-3,6c$	$12,96c^2$
9	$99c + e$	$0,4c$	$0,16c^2$
10	$100c + e$	$1,4c$	$1,96c^2$
Total			$40,40c^2$

$$\bar{t}_{i_{obs}} = 98,6c + e$$

$$var_{obs} = \frac{40,40c^2}{10} = 4,04c^2$$



Medidas de Dispersão

- ▶ Considerando o exemplo 3, pode-se perceber que:
 - ▶ A multiplicação de uma constante c resultou em que a nova variância é a antiga multiplicada por c^2 ;
 - ▶ O acréscimo por uma quantidade e não causou alteração no novo cálculo da variância.



Departamento de Estatística
UNIVERSIDADE FEDERAL DE JUIZ DE FORA



Exercício 1 – Parte 3

- ▶ No Exercício 1 (parte 2), definimos a quantidade D , *despesa no vestibular*, obtida a partir de X (*número de vestibulares prestados*) pela expressão $D = 50X + 1300$. Calcule a variância de D

Nº de vestibulares FUVEST	1	2	3	4
Nº de estudantes	75	47	21	7



Desvio Padrão

- ▶ O desvio padrão é dado pela raiz quadrada da variância:
- ▶ $dp_{obs} = \sqrt{var_{obs}} = \sqrt{\frac{1}{n} \sum_{i=1}^n (x_i - \bar{x})^2}$
- ▶ A grande vantagem do desvio padrão é o fato dele ter a mesma unidade de medida das observações;
- ▶ Assim como a variância o desvio padrão fornece uma medida de variabilidade ao redor da média do conjunto observado;
- ▶ No entanto o valor dado pelo desvio padrão costuma ser mais direto para a compreensão do quanto os dados se distanciam da sua média aritmética.



Exemplo 1 – desvio padrão

- ▶ Considere o mesmo conjunto de dados (caixas de parafusos) utilizado anteriormente: 98, 102, 100, 100, 99, 97, 96, 95, 99, 100. Para essas caixas, o desvio padrão será dado por:

$$dp_{obs} = \sqrt{var_{obs}} = \sqrt{4,04} = 2,01$$



Coeficiente de Variação

- ▶ Em alguns casos é interessante relacionar a média aritmética com o desvio padrão.
- ▶ O coeficiente de variação fornece uma medida livre de dimensão e representada como uma percentagem, indicando a importância da variação dos dados:
- ▶ $CV = \frac{dp}{\bar{x}} \times 100\%$.
- ▶ Como o coeficiente de variação não possui dimensão, ele pode ser utilizado para comparar a variabilidade entre dois conjuntos de dados distintos;
- ▶ Quanto menor o CV, maior a homogeneidade entre os dados.



Exercício 5

- ▶ Considere os dados dos pesos e alturas dos 50 alunos entrevistados (dados brutos apresentados no slide 23):
- ▶ Calcule o coeficiente de variação para as variáveis peso (Y) e altura (X), sabendo que:
 - ▶ $\bar{x}_{obs} = 1,672m$; $dp(X)_{obs} = 0,0897m$
 - ▶ $\bar{y}_{obs} = 60,188kg$; $dp(Y)_{obs} = 11,634kg$;
- ▶ Qual variável é a mais homogênea?
- ▶ Como podemos interpretar os resultados obtidos?

