CAPÍTULO 3

MEDIDAS DE TENDÊNCIA CENTRAL E VARIABILIDADE

PPGEP
UFRGS

3.1. Medidas de Tendência Central

Há várias medidas de tendência central. Entre elas citamos a média aritmética, a mediana, a média harmônica, etc.

Cada uma dessas medidas apresenta vantagens e desvantagens, e a escolha depende dos objetivos desejados.

3.1.1. Média Aritmética

A média aritmética, ou simplesmente média, de um conjunto de *n* valores $x_1, \ldots, x_n$ é definida como:

$$\bar{X} = \frac{x_1 + \ldots + x_n}{n} = \frac{1}{n}\sum_{i=1}^{n} x_i$$

√ As letras gregas são usadas para representar parâmetros populacionais e as letras comuns parâmetros amostrais.

√ A média de uma amostra é representada por $\bar{X}$ e média de uma população é representada pela letra grega μ.

Exemplo: A média aritmética de 7,5 7,9 8,1 8,2 8,7 é

$$\bar{X} = \frac{7{,}5 + 7{,}9 + 8{,}1 + 8{,}2 + 8{,}7}{5} = 8{,}08$$

3.1.2. Média Aritmética para Dados Agrupados

É calculada quando a informação disponível é o valor médio do intervalo *i* (X_i) e a freqüência do intervalo *i* (f_i):

$$\bar{X} = \frac{f_1 X_1 + \ldots + f_K X_K}{f_1 + \ldots + f_K} = \frac{\sum_{i=1}^{K} f_i X_i}{\sum_{i=1}^{K} f_i}$$

Exemplo

Para os dados da Tabela 2.1. resulta:

12,58	12,97	13,45	13,53	13,59	13,61	13,62	13,78	13,97	14,21
14,47	14,51	14,53	14,58	14,65	14,78	14,83	14,97	15,06	15,13
15,17	15,23	15,29	15,37	15,40	15,45	15,51	15,62	15,67	15,73
15,83	15,98	16,01	16,11	16,17	16,23	16,35	16,43	16,49	16,52
16,67	16,83	16,97	17,05	17,13	17,22	17,3	17,48	17,8	18,47

Intervalos de classe	Freqüência absoluta
12,50 a 13,50	3
13,51 a 14,50	8
14,51 a 15,50	15
15,51 a 16,50	13
16,51 a 17,50	9
17,51 a 18,50	2

$$\overline{X} = \frac{3(13)+8(14)+15(15)+13(16)+9(17)+2(18)}{50} = 15{,}46$$

LOPP

3.1.3. Média Aritmética Ponderada

Algumas vezes associa-se a cada observação um peso W_i, onde esse peso representa a importância atribuída a cada observação. Nesse caso a média ponderada é calculada como:

$$\overline{X} = \frac{w_1 x_1 + \ldots + w_n x_n}{w_1 + \ldots + w_n} = \frac{\sum_{i=1}^{n} w_i \, x_i}{\sum_{i=1}^{n} w_i}$$

Exemplo: **O exame de seleção pode ser composto de três provas onde as duas primeiras tem peso 1 e a terceira tem peso 2. Um candidato com notas 70 75 e 90 terá média final:**

$$\overline{X} = \frac{1(70)+1(75)+2(90)}{4} = 81{,}25$$

LOPP

3.1.4. Mediana

Dado um conjunto de valores em ordem crescente, a mediana é definida como:

- √ **Se *n* é impar, o valor central;**
- √ **Se *n* é par, a média simples dos dois valores centrais.**

Exemplos

Exemplo 1: Na amostra 25 26 26 28 30 a mediana é

$$\tilde{x} = 26$$

Exemplo 2: Na amostra 71 73 74 75 77 79 a mediana é

$$\tilde{x} = \frac{(74+75)}{2} = 74{,}5$$

LOPP

3.1.5. Moda

A moda é o valor que ocorre com maior freqüência, ou seja, é o valor mais comum.

Exemplos

Exemplo 1: A amostra 23 25 25 26 26 26 27 29 tem moda 26.

Exemplo 2: A amostra 71 73 73 75 76 77 77 79 81 tem moda 73 e 77.

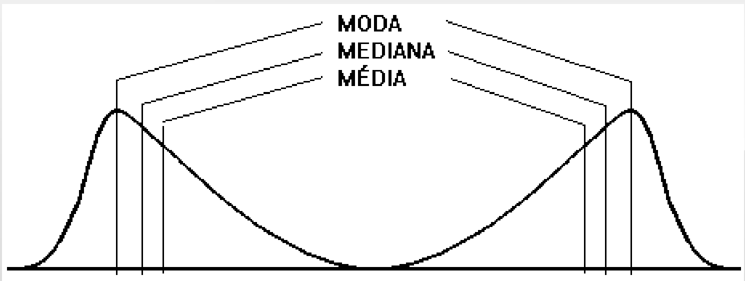
Nota: A moda pode ser múltipla ou pode não existir.

LOPP

3.1.6. Relações Empíricas entre Média, Moda e Mediana

Para distribuições simétricas a média, a mediana e a moda coincidem aproximadamente.

Para distribuições assimétricas observa-se o seguinte:



3.1.6. Relações Empíricas entre Média, Moda e Mediana

Exemplo

A relação entre média e mediana para as amostras a seguir é

A	Distribuição simétrica	10 12 14 16 18	$\bar{x} = 14 \; = \; \tilde{x} = 14$
B	Distribuição assimétrica à direita	10 12 14 16 23	$\bar{x} = 15 \; > \; \tilde{x} = 14$
C	Distribuição assimétrica à esquerda	05 12 14 16 18	$\bar{x} = 13 \; < \; \tilde{x} = 14$

3.1.7. Média Geométrica (G)

É a raiz de ordem *n* do produto dos valores da amostra:

$$G = \sqrt[n]{X_1 X_2 \ldots . X_n}$$

Exemplo

A média geométrica de 12 14 16 é:

$$G = \sqrt[3]{12 \times 14 \times 16} = 13{,}90$$

3.1.8. Média Harmônica (H)

É o inverso da média aritmética dos inversos das observações.

$$H = \frac{1}{\frac{1}{n}\sum \frac{1}{X_i}} = \frac{n}{\sum \frac{1}{X_i}}$$

Exemplo

A média harmônica de 12 14 16 é:

$$H = \frac{3}{\frac{1}{12} + \frac{1}{14} + \frac{1}{16}} = 13{,}81$$

Medidas de Tendência Central e Variabilidade

PPGEP

3.1.9. Relação entre Média Aritmética, Geométrica e Harmônica:

A média geométrica e a média harmônica são menores, ou no máximo igual, à média aritmética (ver eq. 3.11).

A igualdade só ocorre no caso em que todos os valores da amostra são idênticos.

Quanto maior a variabilidade, maior será a diferença entre as médias harmônica e geométrica e a média aritmética.

$$H \leq G \leq \overline{X}$$

Exemplo: **Para a amostra *12 14 16* tem-se**

$H = 13{,}81 < G = 13{,}90 < \overline{X} = 14{,}00$

LOPP

Medidas de Tendência Central e Variabilidade

PPGEP

3.1.10. Quartis

√ **Se um conjunto de dados é organizado em ordem crescente, o valor central é a mediana.**

√ **Valores que dividem o conjunto em quatro partes iguais são representados por Q_1, Q_2, Q_3 e denominam-se primeiro, segundo e terceiro quartis, respectivamente.**

√ **O segundo quartil (Q_2) é a mediana.**

√**O primeiro e o terceiro quartis são calculados usando-se o seguinte procedimento:**

LOPP

Medidas de Tendência Central e Variabilidade

PPGEP

3.1.10. Quartis

(1) partindo de uma amostra de tamanho *n*, colocar os valores em ordem crescente e identificar a ordem *i* (*1, 2, 3, …, n*) e o percentil *p(i) = (i-0,5)/n* associado a cada valor.

(2) identificar os valores associados aos percentis imediatamente acima e abaixo de 0,25; esses valores são chamados respectivamente de *x(inf)*, associado ao percentil *p(inf)*, e *x(sup)*, associado ao percentil *p(sup)*.

(3) e então calcular o primeiro quartil usando:

$$Q_1 = \frac{[p(\sup) - 0{,}25] \times x(\inf) + [0{,}25 - p(\inf)] \times x(\sup)}{p(\sup) - p(\inf)}$$

LOPP

Medidas de Tendência Central e Variabilidade

PPGEP

3.1.10. Quartis

(4) similarmente, para o terceiro quartil, identifica-se os valores associados aos percentis imediatamente acima e abaixo de 0,75; esses valores são chamados respectivamente de *x(inf)*, associado ao percentil *p(inf)*, e *x(sup)*, associado ao percentil *p(sup)*.

E então calcula-se o terceiro quartil usando:

$$Q_3 = \frac{[p(\sup) - 0{,}75] \times x(\inf) + [0{,}75 - p(\inf)] \times x(\sup)}{p(\sup) - p(\inf)}$$

LOPP

3.1.10. Quartis

Exemplo: Para a amostra a seguir calcular o primeiro e terceiro quartis: 13,3 13,5 17,2 13,8 12,3 12,7 13,0 14,5 14,9 15,8 13,1 13,3 14,1

x(i)	i	p(i) = (i-0,5)/n
12,3	1	0,038
12,7	2	0,115
13,0	3	0,192
13,1	4	0,269
13,3	5	0,346
13,3	6	0,423
13,5	7	0,500
13,8	8	0,577
14,1	9	0,654
14,5	10	0,731
14,9	11	0,808
15,8	12	0,885
17,2	13	0,962

LOPP

3.1.10. Quartis

(2) valores imediatamente acima e abaixo de 0,25: *x(inf)* = 13,0 e *x(sup)* = 13,1 associados com *p(inf)* = 0,192 e *p(sup)* = 0,269

$$Q_1 = \frac{[0{,}269 - 0{,}25]\times(13{,}0) + [0{,}25 - 0{,}192]\times 13{,}1}{0{,}269 - 0{,}192} = 13{,}08$$

(4) valores imediatamente acima e abaixo de 0,75: *x(inf)* = 14,5 e *x(sup)* = 14,9 associados com *p(inf)* = 0,731 e *p(sup)* = 0,808

$$Q_3 = \frac{[0{,}808 - 0{,}75]\times(14{,}5) + [0{,}75 - 0{,}731]\times 14{,}9}{0{,}808 - 0{,}731} = 14{,}60$$

LOPP

3.2. Medidas de Variabilidade

Invariavelmente as observações individuais irão apresentar alguma dispersão em torno do valor médio. Isso é chamado de variabilidade ou dispersão dos dados.

Há muitas medidas de variabilidade, como por exemplo, a amplitude total, o desvio padrão ou a distância inter-quartílica.

LOPP

3.2.1. Amplitude Total

- A amplitude total é definida como a diferença entre o maior e o menor valor das observações.

Exemplo : 8,5 8,7 8,9 10,1 10,5 10,7 11,5 11,9

A amplitude é: R = 11,9 - 8,5 = 3,4

- A amplitude é fácil de calcular e fornece uma idéia da magnitude da faixa de variação dos dados.
- Não informa a respeito da dispersão dos valores que caem entre os dois extremos.
- Quando n < 10 pode resultar em uma medida de variação bastante satisfatória.

LOPP

3.2.2. Desvio Padrão

Para uma amostra de *n* observações, $x_1, ..., x_n$, o desvio padrão *S* é definido como:

$$S = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}} \Rightarrow \text{amostra pequena } (n < 30)$$

√ A vantagem do desvio padrão é que trata-se de uma medida de variabilidade que leva em conta toda a informação contida na amostra.

√ A desvantagem é que seu cálculo é mais trabalhoso.

√ Quando a amostra é grande ($n > 30$) ou quando trata-se da população usa-se *n* no denominador.

√ O desvio-padrão de uma população é representado por σ.

Medidas de Tendência Central e Variabilidade

3.2.2. Desvio Padrão

Exemplo: para a amostra 10 12 14 16 18

A média é $\bar{x} = 14$ e o desvio-padrão é calculado :

Os desvios de cada valor em relação à média totalizam zero pois a média é o valor central:

$$10 - 14 = -4$$
$$12 - 14 = -2$$
$$14 - 14 = 0$$
$$16 - 14 = +2$$
$$18 - 14 = +4$$

$$S = \sqrt{\frac{(10-14)^2 + (12-14)^2 + (14-14)^2 + (16-14)^2 + (18-14)^2}{n-1}} = 3{,}16$$

Medidas de Tendência Central e Variabilidade

3.2.3. Variância

A variância S^2 é definida como o quadrado do desvio padrão.

$$S^2 = \frac{\sum (x_i - \bar{x})^2}{n-1} = \frac{\sum x_i^2 - [(\sum x_i)^2 / n]}{n-1}$$

A variância de uma população é representada pela letra grega σ^2.

A variância é o quadrado do desvio padrão, ou seja, σ^2

$\sigma^2 = 3{,}16^2 = 9{,}98$

Medidas de Tendência Central e Variabilidade

3.2.4. Amplitude Inter-quartílica

É definida como a amplitude do intervalo entre o primeiro e o terceiro quartis, ou seja:

$$Q = Q_3 - Q_1$$

√ As vezes também é usada a semi-amplitude inter-quartílica, que é a metade da anterior.

√ Trata-se de uma medida de variabilidade bastante robusta, que é pouco afetada pela presença de dados atípicos.

√ A amplitude inter-quartílica guarda a seguinte relação aproximada com o desvio padrão: *Q = (4/3)* x *desvio padrão*

Medidas de Tendência Central e Variabilidade

3.2.5. Coeficiente de Variação

É definido como o quociente entre o desvio padrão e a média e, em geral, é expresso em percentual.

$$CV = 100 \times \frac{S}{\overline{X}}$$

O coeficiente de variação é uma medida adimensional, útil para comparar resultados de amostras cujas unidades podem ser diferentes.

Uma desvantagem do coeficiente de variação é que ele deixa de ser útil quando a média é próxima de zero.

3.2.5. Coeficiente de Variação

EXEMPLO:

O resultado de dois grupos obtiveram os seguintes resultados:

Grupo 1:

- Média=2,49 mm, Desvio Padrão=0,12 mm

Grupo 2:

- Média=3,75 cm, Desvio Padrão=0,15 cm

Qual dos dois grupos é relativamente mais preciso?

CV1=0,12/2,49*100 = 4,8%

CV2=0,15/3,75*100 = 4,0% O segundo é mais preciso.

3.2.6. Variável Reduzida ou Padronizada

$$Z = \frac{X - \overline{X}}{S}$$

Ela mede a magnitude do desvio em relação à média, em unidades do desvio padrão.

Z = *1,5* significa uma observação desviada 1,5 desvios padrão para cima da média.

A variável reduzida é muito útil para comparar distribuições e detectar dados atípicos.

Dados são considerados atípicos quando $Z > 3$.

Exemplo:

O engenheiro está analisando as espessuras de peças fabricadas em duas máquinas de corte.

O operador mediu uma peça da máq. A com espessura de 90 mm e outra peça da máq. B com espessura de 100 mm.

O engenheiro deve considerar esses dados reais ou atípicos?

A máq. A possui média 51mm e desvio-padrão de 12mm.

Máq. A $Z = \frac{X - \overline{X}}{S} = \frac{90 - 51}{12} = 3{,}25$ Como $Z > 3$ é dado atípico

A máq. B possui média 72mm e desvio-padrão de 16mm.

Máq. B $Z = \frac{X - \overline{X}}{S} = \frac{100 - 72}{16} = 1{,}75$ Como $Z < 3$ não é dado atípico

PPGEP

Medidas de Tendência Central e Variabilidade

Exemplo:

Supondo que 51 fosse a média em uma prova de inglês, onde o desvio padrão é 12, para um candidato que obtivesse 90 acertos tem-se:

$$Z = \frac{X - \overline{X}}{S} = \frac{90 - 51}{12} = 3{,}25$$

Conclui-se que na prova de inglês este candidato está 3,25 desvios-padrão acima da média.

LOPP

PPGEP

Medidas de Tendência Central e Variabilidade

Exercícios

3.1. Para os dados do exercício 2.1, calcule a média aritmética e a mediana e verifique a relação *média* > *mediana*, válida para distribuições assimétricas à direita.

3.2. Ainda em relação aos dados do exercício 2.1, calcule a média aritmética usando a fórmula para dados agrupados e a tabela de freqüência que você construiu.

3.3. Para os dados do exercício 2.3., calcule a média e a mediana e verifique a relação média mediana para distribuições simétricas.

LOPP

PPGEP

Medidas de Tendência Central e Variabilidade

Exercícios

3.4. A partir dos dados do exercício 2.6., use a fórmula para o cálculo da média de dados agrupados e calcule a média para:

X1: Característica dimensional de uma peça;

X2: Tempo de uso (horas/semana) de um produto;

X3: Tempo até a falha de um produto.

LOPP

PPGEP

Medidas de Tendência Central e Variabilidade

Exercícios

3.5. As amostras a seguir representam valores de rugosidade medidos sobre superfície de peças produzidas por três máquinas diferentes.

M A	20,2	24,7	25,7	21,7	19,2	22,3	23,0	23,1	21,3	26,8	20,7	23,6	25,4	24,6	22,5
M B	21,3	22,7	22,5	23,8	20,4	23,3	23,7	23,4	25,5	22,4	23,1	21,7	24,3	24,7	22,2
M C	22,1	24,4	24,0	21,5	23,2	22,0	25,4	27,8	23,5	23,0	20,6	23,6	22,5	22,8	21,4

Para cada máquina, calcule a amplitude total, o desvio padrão e a amplitude inter-quartílica. Após, conclua a respeito de diferenças de variabilidade entre essas máquinas.

LOPP

Exercícios

3.6. Calcule o valor da variável reduzida *Z* para os pontos extremos das amostras que aparecem no exercício anterior. Após indique se há evidência de dados atípicos em alguma dessas amostras (obs: para *n=15*, um valor de *Z > 2,5* já seria evidência de dado atípico).

3.7. Caso haja indícios de dados atípicos, elimine esse resultado e refaça os cálculos da amplitude total, desvio padrão e amplitude inter-quartílica para a amostra correspondente. Se necessário, revise as conclusões do exercício 3.5.

Exercícios

3.8. Para a amostra a seguir, calcule a média, o desvio padrão e o coeficiente de variação. Calcule também os valores de *Z* para cada observação. Analise os valores de *Z* e indique se a amostra vem de uma população com distribuição simétrica, assimétrica à direita ou assimétrica à esquerda.

Tempos de uso (horas/semana) de um produto

4,5	7,0	9,0	11	13	14	16	16	17	18
18	20	21	22	22	23	24	24	24	24
25	25	26	26	26	27	28	28	29	30

Exercícios

3.9. Idem ao anterior

Tempos de uso (horas/semana) de um produto concorrente:

0,2	0,2	0,3	0,4	0,6	0,6	0,8	1,0	1,0	1,2
1,2	1,3	1,4	1,5	1,5	1,7	1,8	2,0	2,2	2,5
2,5	2,7	3,3	3,5	3,8	4,3	5,1	12,0	12,	15,0