

ACH3553 - Estatística I

Curso de Gestão de Políticas Públicas

Aula 6

Alexandre Ribeiro Leichsenring
alexandre.leichsenring@usp.br

EACH

1 Medidas de dispersão

- Amplitude
- Amplitude do intervalo interquartil
- Variância e Desvio Padrão
- Coeficiente de Variação

1 Medidas de dispersão

- Amplitude
- Amplitude do intervalo interquartil
- Variância e Desvio Padrão
- Coeficiente de Variação

Medidas de dispersão

Segundo o dicionário Houaiss:

dispersão

substantivo feminino (a1789)

ato ou efeito de dispersar(-se)

1 separação (de pessoas ou coisas) por diferentes lugares

2 p.ext. debandada, correria, desbarato

3 bio eco processo de separação de organismos anteriormente agrupados, como, p.ex., o afastamento da prole de seus pais ou o deslocamento de indivíduos de uma região de alta densidade populacional para outra, de baixa densidade

4 est grau de flutuação mostrado por uma variável aleatória

5 fisquím disseminação de uma substância em um fluido no qual não é solúvel

5.1 fisquím o resultante dessa disseminação (aplicável esp. a emulsões e suspensões estáveis, como leite, *fog*, *smog*)

6 ópt variação do índice de refração num meio de acordo com o comprimento de onda da radiação

7 psic falta de concentração; desatenção

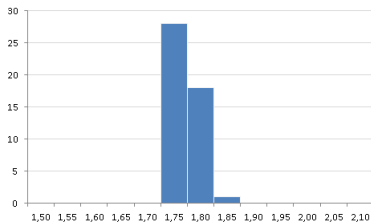
Medidas de dispersão

- Medidas de tendência central representam um *valor típico* de um conjunto de dados
 - ▶ o valor ao redor do qual os valores do conjunto estão distribuídos
- Como estão distribuídos esses valores ao redor da medida de tendência central?
 - ▶ **Medidas de dispersão**
- Também chamadas de medidas de variabilidade
- Principais medidas de variabilidade:
 - ▶ Amplitude
 - ▶ Variação interquartil
 - ▶ Variância
 - ▶ Desvio Padrão

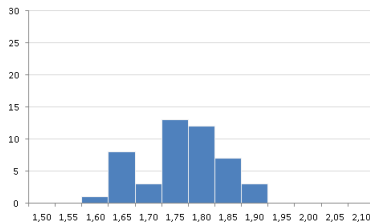
Dispersão

Vejamos a distribuição da altura dos indivíduos de 3 grupos diferentes.

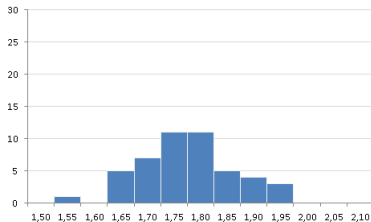
Grupo 1



Grupo 2



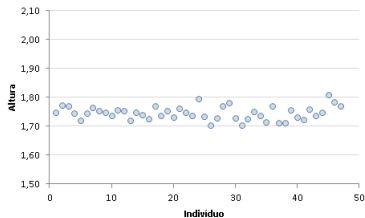
Grupo 3



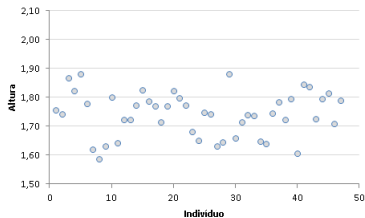
Dispersão

Outra forma de examinar.

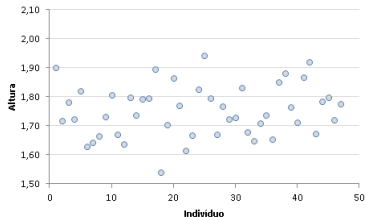
Grupo 1



Grupo 2



Grupo 3



Amplitude

Definição

Amplitude é a diferença entre o maior e o menor valor de uma distribuição.

$$A = x_{\text{máx}} - x_{\text{mín}}$$

- Medida simples
 - Cálculo rápido e fácil
 - Depende apenas de dois valores ($x_{\text{máx}}$ e $x_{\text{mín}}$)
 - Medida influenciada por apenas um valor
- ⇒ Medida de variabilidade apenas aproximada

Amplitude

Exemplo

Com relação às alturas dos 3 grupos apresentados anteriormente:

	Grupo 1	Grupo 2	Grupo 3
Máximo ($X_{\max}$)	1,79	1,98	1,97
Mínimo ($X_{\min}$)	1,70	1,58	1,57
Amplitude ($X_{\max} - X_{\min}$)	0,09	0,40	0,40

Amplitude

Exemplo

No exemplo das notas de estatística, o menor valor observado foi 0 e o maior foi 10.

$$x_{\text{máx}} = 10$$

$$x_{\text{mín}} = 0$$

$$A = x_{\text{máx}} - x_{\text{mín}} = 10 - 0 = 10$$

A amplitude nesse caso é 10.

Amplitude

Exemplo

Com relação às idades dos alunos:

$$x_{\text{máx}} = 32$$

$$x_{\text{mín}} = 17$$

$$A = x_{\text{máx}} - x_{\text{mín}} = 32 - 17 = 15$$

A amplitude nesse caso é 15.

Amplitude do intervalo interquartil

- Como vimos, amplitude depende completamente de apenas dois valores extremos da distribuição
- Em aulas passadas, introduzimos o conceito de quartis
- Ao invés de usar valores extremos, podemos usar valores que determinam um conjunto de valores centrais

Definição

A amplitude do intervalo interquartil (A/IQ) é dada pela distância entre o primeiro e o terceiro quartis:

$$A/IQ = Q_3 - Q_1$$

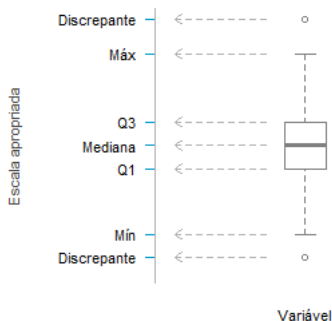
Amplitude do intervalo interquartil

Exemplo

Vamos determinar a amplitude do intervalo interquartil dos dados de altura dos indivíduos dos 3 grupos.

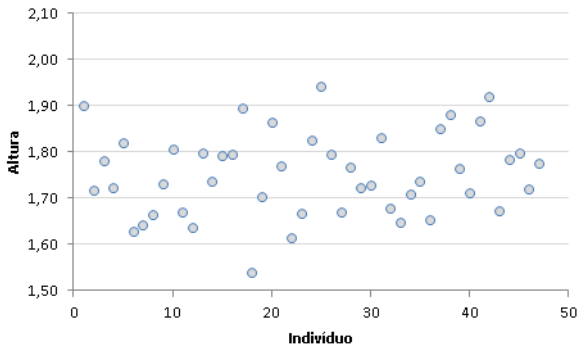
Amplitude do intervalo interquartil

Amplitude do intervalo interquartil e o *Boxplot*



Variância e Desvio Padrão

- Medida de dispersão dos dados em torno da média.

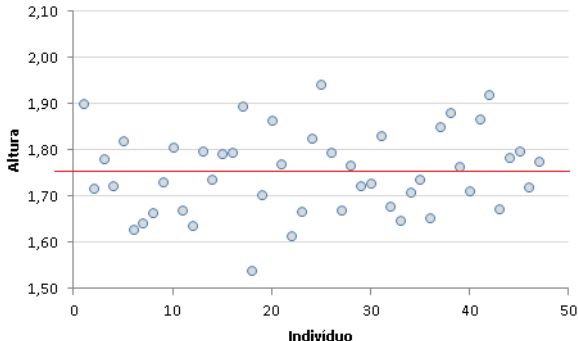


- Suponha que a média do conjunto de dados retratado acima é $\bar{x} = 1,75$
- O afastamento de cada observação com relação à média é chamado de desvio

⇒ A Variância é uma medida do padrão dos desvios com relação à média.

Variância e Desvio Padrão

- Medida de dispersão dos dados em torno da média.

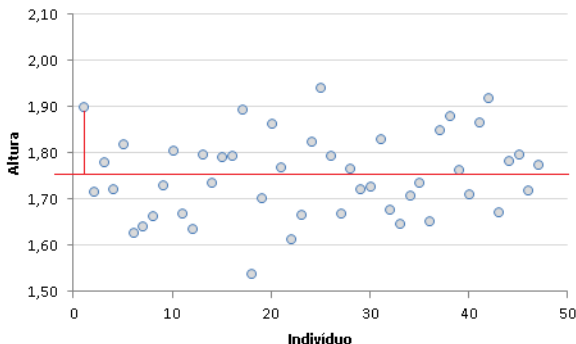


- Suponha que a média do conjunto de dados retratado acima é $\bar{x} = 1,75$
- O afastamento de cada observação com relação à média é chamado de desvio

⇒ A Variância é uma medida do padrão dos desvios com relação à média.

Variância e Desvio Padrão

- Medida de dispersão dos dados em torno da média.

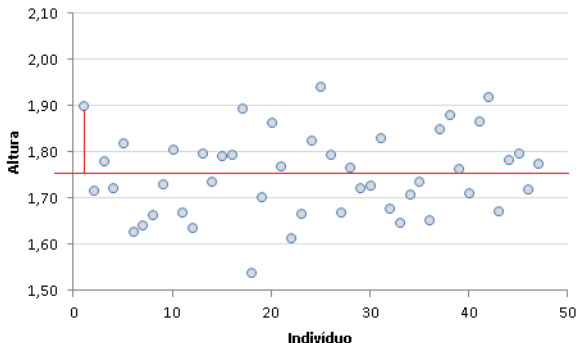


- Suponha que a média do conjunto de dados retratado acima é $\bar{x} = 1,75$
- O afastamento de cada observação com relação à média é chamado de desvio

⇒ A Variância é uma medida do padrão dos desvios com relação à média.

Variância e Desvio Padrão

- Medida de dispersão dos dados em torno da média.



- Suponha que a média do conjunto de dados retratado acima é $\bar{x} = 1,75$
- O afastamento de cada observação com relação à média é chamado de desvio

⇒ A Variância é uma medida do padrão dos desvios com relação à média.



Variância e Desvio Padrão

Variância

Para um conjunto de dados com n observações $x_1, x_2, \dots, x_n$, de uma variável X , a **Variância** é definida por

$$Var(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}, \quad (1)$$

Onde $\bar{x}$ é a média amostral.

Desvio Padrão

Para um conjunto de dados com n observações $x_1, x_2, \dots, x_n$ de uma variável X , o **Desvio Padrão** é definido por:

$$DP(X) = \sqrt{Var(X)} \quad (2)$$

Onde $\bar{x}$ é a média amostral.



Variância e Desvio Padrão

Procedimento para calcular a Variância:

- 1 Calcula-se a média $\bar{x}$
- 2 Para cada observação x_i , calcula-se o seu desvio com relação à média, isto é:

$$(x_i - \bar{x})$$

- 3 Eleva-se os termos acima ao quadrado, obtendo os desvios ao quadrado:

$$(x_i - \bar{x})^2$$

- 4 Soma-se os desvios ao quadrado:

$$\sum_{i=1}^n (x_i - \bar{x})^2$$

- 5 Divide-se a soma acima por n :

$$Var(X) = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}$$

Variância e Desvio Padrão

Exemplo

Vamos calcular a variância e o desvio padrão das alturas dos indivíduos do Grupo 1.

Variância e Desvio Padrão

Exemplo

Abaixo temos as alturas dos 10 primeiros indivíduos de cada grupo. Vamos calcular à mão o **desvio padrão** das alturas dos 10 indivíduos do grupo 1.

Indivíduo	Grupo 1
1	1,78
2	1,71
3	1,75
4	1,70
5	1,76
6	1,75
7	1,75
8	1,72
9	1,78
10	1,74

Variância e Desvio Padrão

Exemplo

Abaixo temos as alturas dos 10 primeiros indivíduos dos grupos. Para esses grupos, determine:

a) Amplitude

b) Média

c) Variância

d) Desvio padrão

▷ Em qual grupo os dados apresentam maior **dispersão**?

▷ Qual o grupo mais homogêneo?

Indivíduo	Grupo 1	Grupo 2	Grupo 3
1	1,78	1,84	1,57
2	1,71	1,77	1,70
3	1,75	1,79	1,69
4	1,70	1,69	1,68
5	1,76	1,64	1,88
6	1,75	1,69	1,69
7	1,75	1,77	1,66
8	1,72	1,75	1,89
9	1,78	1,69	1,70
10	1,74	1,72	1,77

Variância e Desvio Padrão

Cálculo da variância para dados agrupados em tabela de frequência

Exemplo

Suponha que numa prova com 100 alunos, o número de questões certas no teste de múltipla escolha tenha se distribuído como na tabela abaixo:

Nota	Frequência
1	1
2	6
3	10
4	18
5	29
6	19
7	8
8	4
9	3
10	2
Total	100

Como calcular a variância nesse caso?

Variância e Desvio Padrão

Cálculo da variância para dados agrupados em tabela de frequência

Basta observar o procedimento original, ponderando os desvios correspondentes a cada valor pela frequência observada do valor. Isso corresponde a aplicar a seguinte fórmula:

$$Var(X) = \frac{\sum_{i=1}^k (x_i - \bar{x})^2 \times f_i}{\sum_{i=1}^k f_i}$$

Onde k é o número de distintos valores observados da variável x .

► Observe que $\sum_{i=1}^k f_i = n$, ou seja, o número total de observações no conjunto de dados.

Variância e Desvio Padrão

Cálculo da variância para dados agrupados em tabela de frequência

Roteiro para o cálculo da variância para os dados agrupados do exemplo anterior:

- 1 Observe que há 10 valores distintos, ou seja, $k = 10$.
- 2 Calcule a média:

$$\begin{aligned}\bar{x} &= \frac{\sum_{i=1}^{10} x_i \times f_i}{\sum_{i=1}^{10} x_i} \\&= \frac{(1 \times f_1) + (2 \times f_2) + \dots + (10 \times f_{10})}{f_1 + f_2 + \dots + f_{10}} \\&= \frac{(1 \times 1) + (2 \times 6) + \dots + (10 \times 2)}{1 + 6 + \dots + 2} \\&= \frac{509}{100} \\&= 5,09\end{aligned}$$

Variância e Desvio Padrão

Calculando a variância quando se tem dados agrupados numa tabela de frequência

- 2 Para cada nota x_i , calcula-se o seu desvio com relação à média, isto é:

$$(x_i - \bar{x}) = (x_i - 5,09)$$

- 3 Eleva-se os termos acima ao quadrado, obtendo os desvios ao quadrado:

$$(x_i - 5,09)^2$$

- 4 Soma-se os desvios ao quadrado ponderados pelas frequências observadas:

$$\sum_{i=1}^n (x_i - 5,09)^2 \times f_i$$

- 5 Divide-se a soma acima pelo número total de observações, ou seja, por 100 (observe que $\sum_{i=1}^{10} f_i = 100$):

$$Var(X) = \frac{\sum_{i=1}^{10} (x_i - 5,09)^2 \times f_i}{100}$$

Variância e Desvio Padrão

Exercício

Um levantamento de idades de 100 mães grávidas obteve a seguinte tabela de frequências:

Idade da mãe	f_i
15 a 20	7
20 a 25	36
25 a 30	23
30 a 35	18
35 a 40	10
40 a 45	6
Total	100

- Calcule o desvio padrão desses dados, considerando o ponto médio dos intervalos como *valor representativo* do intervalo.
- Dica: use o procedimento do exemplo anterior, substituindo os valores x_i pelos pontos médios dos intervalos ($\hat{x}_i$)

Coeficiente de Variação

- A variância ou do desvio padrão medem a dispersão em relação à média de forma absoluta!
- Vamos comparar os seguintes contextos:
 - ▶ desvio padrão de 10 g, num conjunto de dados cujo peso médio é 1.000 g
 - ▶ desvio padrão de 5 g, num conjunto de dados cujo peso médio é 50 g
- Qual das duas distribuições acima apresenta maior *variabilidade*?
- É necessário avaliar a variabilidade de maneira relativa...

Coeficiente de Variação

Coeficiente de variação (CV)

$$CV = 100 \times \frac{DP}{\bar{X}},$$

onde:

- DP é o desvio padrão
- $\bar{x}$ é a média