Métodos Quantitativos

André Amorim
Finanças Corporativas

contato@andreamorim.com.br
www.andreamorim.com.br

Anhanguera

1

8ª Aula – Medidas de dispersão

MQA

medidas de dispersão

André Amorim
Finanças Corporativas

Anhanguera

2

8ª Aula – Medidas de dispersão

MQA

Medidas de dispersão

Essas medidas buscam dimensionar quanto os dados estão distantes da média, por exemplo. Com o auxílio delas podemos decidir, por exemplo, se a média pode ser utilizada como representante de um conjunto.

André Amorim
Finanças Corporativas

Anhanguera

3

8ª Aula – Medidas de dispersão

MQA

Medidas de dispersão

"Dispersão (ou variabilidade) de um conjunto refere-se à maior ou menor diversificação dos valores de uma variável em torno de um valor de tendência central tomado como ponto de comparação".

Carlos Augusto de Medeiros, chefe da Unidade de Administração Geral
da Fundação Universidade Aberta do Distrito Federal

André Amorim
Finanças Corporativas

Anhanguera

4

8ª Aula – Medidas de dispersão

MQA

Desvio

Denominamos desvio a diferença de um valor do conjunto com relação à média.

André Amorim
Finanças Corporativas

Anhanguera

5

8ª Aula – Medidas de dispersão

MQA

Desvio

Tabela 2.14 | Desvios dos conjuntos de dados

i	Valores do conjunto			Desvios		
	x_i	y_i	z_i	$x_i - \bar{x}$	$y_i - \bar{y}$	$z_i - \bar{z}$
1	90	86	30	$90 - 90 = 0$	$86 - 90 = -4$	$30 - 90 = -60$
2	90	88	60	$90 - 90 = 0$	$88 - 90 = -2$	$60 - 90 = -30$
3	90	90	90	$90 - 90 = 0$	$90 - 90 = 0$	$90 - 90 = 0$
4	90	92	120	$90 - 90 = 0$	$92 - 90 = 2$	$120 - 90 = 30$
5	90	94	150	$90 - 90 = 0$	$94 - 90 = 4$	$150 - 90 = 60$
Total	$\Sigma x = 450$	$\Sigma y = 450$	$\Sigma z = 450$	$\sum(x_i - \bar{x}) = 0$	$\sum(y_i - \bar{y}) = 0$	$\sum(z_i - \bar{z}) = 0$

André Amorim
Finanças Corporativas

Anhanguera

6

8ª Aula – Medidas de dispersão

MQA

Desvio

- Observe que para as amostras das variáveis X, Y e Z a soma de todos os desvios e igual a zero.

- Isso não ocorre somente para estes conjuntos, mas para todos os conjuntos de dados.

- Desse modo, qualquer tentativa de utilizar a soma dos desvios $\sum(x_i - \bar{x})$ para dimensionar a variabilidade dos dados será frustrada

André Amorim
Finanças Corporativas

Anhanguera

7

8ª Aula – Medidas de dispersão

MQA

Desvio

- Isso ocorre pois os desvios negativos neutralizam os positivos, tornando o total igual a zero.

- Para driblar esse contratempo, os estatísticos se utilizam de um artificio matemático, o valor absoluto.

André Amorim
Finanças Corporativas

Anhanguera

8

8ª Aula – Medidas de dispersão

MQA

Desvio

- O valor absoluto de um número corresponde à distância que este se encontra do 0 (zero). A distância é sempre um valor positivo ou zero.
- Na prática, o valor absoluto de um número: (a) negativo é ele próprio com sinal trocado; (b) não negativo é ele próprio.

André Amorim
Finanças Corporativas

Anhanguera

9

8ª Aula – Medidas de dispersão

MQA

Desvio

O valor absoluto de:

- –1, simbolizado por |–1| é igual a 1, ou seja, |–1| = 1;
- 2, simbolizado por |2| é igual a 2, ou seja, |2| = 2;
- 0, simbolizado por |0| é igual a 0, ou seja, |0| = 0.

André Amorim
Finanças Corporativas

Anhanguera

10

8ª Aula – Medidas de dispersão

MQA

Desvio

Utilizando o valor absoluto, podemos refazer os cálculos como na Tabela 2.15.

Tabela 2.15 | Valores absolutos dos desvios

i	Valores do conjunto			Valor absoluto dos desvios		
	x_i	y_i	z_i	$\lvert x_i - \bar{x}\rvert$	$\lvert y_i - \bar{y}\rvert$	$\lvert z_i - \bar{z}\rvert$
1	90	86	30	0	4	60
2	90	88	60	0	2	30
3	90	90	90	0	0	0
4	90	92	120	0	2	30
5	90	94	150	0	4	60
Total	$\Sigma x = 450$	$\Sigma y = 450$	$\Sigma z = 450$	$\sum\lvert x_i - \bar{x}\rvert = 0$	$\sum\lvert y_i - \bar{y}\rvert = 12$	$\sum\lvert z_i - \bar{z}\rvert = 180$

Fonte: O autor (2015).

11

8ª Aula – Medidas de dispersão

MQA

Desvio

Também podemos simbolizar a soma dos valores absolutos dos desvios por:

$$\sum\lvert x_i - \bar{x}\rvert$$

André Amorim
Finanças Corporativas

Anhanguera

12

8ª Aula – Medidas de dispersão

MQA

Desvio médio

Desvio médio, simbolizado por Dm, é uma medida de dispersão calculada por meio da média aritmética dos valores absolutos dos desvios. Para as variáveis X, Y e Z, temos:

$$Dm(X) = \frac{\sum|x - \bar{x}|}{n} = \frac{0}{5} = 0 \qquad Dm(Y) = \frac{\sum|y - \bar{y}|}{n} = \frac{12}{5} = 2{,}4 \qquad Dm(Z) = \frac{\sum|z - \bar{z}|}{n} = \frac{180}{5} = 36$$

André Amorim
Finanças Corporativas

Anhanguera

13

8ª Aula – Medidas de dispersão

MQA

Desvio médio

Quanto menor o desvio médio, menor a dispersão; quanto maior o desvio médio, maior a dispersão dos dados.

O menor desvio médio possível é 0 (zero) e ocorre quando os dados são totalmente homogêneos.

André Amorim
Finanças Corporativas

Anhanguera

14

8ª Aula – Medidas de dispersão

MQA

Desvio médio

Outra maneira de neutralizar o efeito do sinal negativo ocorrido na Tabela 2.14 é elevar cada desvio ao quadrado, como mostra a Tabela 2.16.

Tabela 2.16 | Quadrado dos desvios

i	$(x_i - \bar{x})^2$	$(y_i - \bar{y})^2$	$(z_i - \bar{z})^2$
1	$0^2 = 0$	$(-4)^2 = 16$	$(-60)^2 = 3600$
2	$0^2 = 0$	$(-2)^2 = 4$	$(-30)^2 = 900$
3	$0^2 = 0$	$0^2 = 0$	$0^2 = 0$
4	$0^2 = 0$	$2^2 = 4$	$30^2 = 900$
5	$0^2 = 0$	$4^2 = 16$	$60^2 = 3600$
Total	$\Sigma(x_i - \bar{x})^2 = 0$	$\Sigma(y_i - \bar{y})^2 = 40$	$\Sigma(z_i - \bar{z})^2 = 9000$

Fonte: O autor (2015).

André Amorim
Finanças Corporativas

Anhanguera

15

8ª Aula – Medidas de dispersão

MQA

Desvio médio

A partir da Tabela 2.16 definimos nossa segunda medida de dispersão:

Variância

André Amorim
Finanças Corporativas

Anhanguera

16

8ª Aula – Medidas de dispersão

MQA

Variância

A variância, simbolizada por Var, é uma medida de dispersão calculada por meio da média aritmética dos quadrados dos desvios. Para as variáveis X, Y e Z, temos:

$$Var(X) = \frac{\sum(x-\bar{x})^2}{n} = \frac{0}{5} = 0 \qquad Var(Y) = \frac{\sum(y-\bar{y})^2}{n} = \frac{40}{5} = 8$$

$$Var(Z) = \frac{\sum(z-\bar{z})^2}{n} = \frac{9000}{5} = 1800$$

André Amorim
Finanças Corporativas

Anhanguera

17

8ª Aula – Medidas de dispersão

MQA

Variância

Imagine que os valores observados para as variáveis X, Y e Z sejam idades.

Quando elevamos os desvios ao quadrado para o cálculo da variância, obtemos um valor que, teoricamente, tem unidade de medida $idade^2$

Como isso pode causar confusão e dificuldade de interpretação, definimos a terceira medida de dispersão.

André Amorim
Finanças Corporativas

Anhanguera

18

8ª Aula – Medidas de dispersão

MQA

Desvio padrão

O desvio padrão, simbolizado por Dp, é uma medida de dispersão definida como a raiz quadrada da variância. Para as variáveis X, Y e Z, temos:

$$Dp(X) = \sqrt{Var(X)} = \sqrt{0} = 0 \qquad Dp(Y) = \sqrt{Var(Y)} = \sqrt{8} \cong 2{,}8$$

$$Dp(Z) = \sqrt{Var(Z)} = \sqrt{1800} \cong 42{,}4$$

André Amorim
Finanças Corporativas

Anhanguera

19

8ª Aula – Medidas de dispersão

MQA

Variância

Ao calcularmos o desvio padrão retornamos à unidade de medida do conjunto de dados, ou seja, se o conjunto de dados é medido em:

- idade, a variância é medida em idade^2 e o desvio padrão é medido em idade;
- m (metros), a variância é medida em m^2 e o desvio padrão é medido em m;
- R$ (reais), a variância é medida em R^2 e o desvio padrão é medido em R$.

André Amorim
Finanças Corporativas

Anhanguera

20

8ª Aula – Medidas de dispersão

MQA

Variância

As medidas apresentadas até aqui estão de forma absoluta (não percentual). Por esse motivo, ao calculá-las nem sempre conseguimos inferir muita coisa sobre a dispersão de um conjunto de dados.

Por exemplo, o valor é muito ou pouco?

Se não tivermos outro valor para que possamos compará-lo fica difícil fazer alguma afirmação.

Por causa disso, definimos nossa quarta medida de dispersão.

André Amorim
Finanças Corporativas

Anhanguera

21

8ª Aula – Medidas de dispersão

MQA

Variância

As medidas apresentadas até aqui estão de forma absoluta (não percentual). Por esse motivo, ao calculá-las nem sempre conseguimos inferir muita coisa sobre a dispersão de um conjunto de dados.

Por exemplo, o valor $Dp(Y) \cong 2{,}8$ é muito ou pouco?

Se não tivermos outro valor para que possamos compará-lo fica difícil fazer alguma afirmação.

Por causa disso, definimos nossa quarta medida de dispersão.

André Amorim
Finanças Corporativas

Anhanguera

22

8ª Aula – Medidas de dispersão

MQA

Coeficiente de variação

simbolizado por CV, é uma medida de dispersão definida como a razão entre o desvio padrão e a média de um conjunto de dados. Para as variáveis X, Y e Z, temos:

$$CV(X) = \frac{Dp(X)}{\bar{x}} = \frac{0}{90} = 0 \qquad CV(Y) = \frac{Dp(Y)}{\bar{y}} = \frac{2{,}8}{90} \cong 0{,}031 \qquad CV(Z) = \frac{Dp(Z)}{\bar{z}} = \frac{42{,}4}{90} \cong 0{,}471$$

André Amorim
Finanças Corporativas

Anhanguera

23

8ª Aula – Medidas de dispersão

MQA

Coeficiente de variação

Também podemos indicar os valores de forma percentual, como a seguir:

$$CV(X) = 0 \cdot 100\% = 0\% \qquad CV(Y) = 0{,}031 \cdot 100\% = 3{,}1\% \qquad CV(Z) = 0{,}471 \cdot 100\% = 47{,}1\%$$

André Amorim
Finanças Corporativas

Anhanguera

24

8ª Aula – Medidas de dispersão

MQA

Coeficiente de variação

O coeficiente de variação permite uma comparação do desvio padrão com a média do conjunto de dados.

Por exemplo, o desvio padrão de Y corresponde a 3,1% do valor médio do conjunto; o desvio padrão de Z corresponde a 47,1% do valor médio do conjunto.

André Amorim
Finanças Corporativas

Anhanguera

25

8ª Aula – Medidas de dispersão

MQA

Coeficiente de variação

Alguns autores costumam utilizar o coeficiente de variação para classificar um conjunto de dados quanto à dispersão dos valores em torno da média. Essa classificação é feita conforme Tabela 2.17.

Tabela 2.17 | Classificação de um conjunto de dados

Classificação	Critério
Baixa dispersão	$CV \leq 15\%$
Média dispersão	$15\% < CV < 30\%$
Alta dispersão	$CV \geq 30\%$

Fonte: O autor (2015)

André Amorim
Finanças Corporativas

Anhanguera

26

8ª Aula – Medidas de dispersão

MQA

Exemplificando

Considerando os conjuntos $A = \{2, 3, 6, 9\}$ e , $B = \{959, 1065, 1090\}$ qual deles possui os dados mais dispersos em torno da média?

Primeiramente calculamos $\bar{a}$, $\bar{b}$, $Var(A)$, $Var(B)$, $Dp(A)$, $Dp(B)$, $CV(A)$ e $CV(B)$.

$$\bar{a} = \frac{\sum a}{n_a} = \frac{2+3+6+9}{4} = 5;$$

$$\bar{b} = \frac{\sum b}{n_b} = \frac{959+1065+1090}{3} = 1038$$

André Amorim
Finanças Corporativas

Anhanguera

27

8ª Aula – Medidas de dispersão

MQA

Exemplificando

Considerando os conjuntos $A = \{2, 3, 6, 9\}$ e , $B = \{959, 1065, 1090\}$ qual deles possui os dados mais dispersos em torno da média?

$$Var(A) = \frac{\sum(a-\bar{a})^2}{n_a} \; \frac{(2-5)^2+(3-5)^2+(6-5)^2+(9-5)^2}{4} = 7{,}5$$

$$Var(B) = \frac{\sum(b-\bar{b})^2}{n_b} \; \frac{(959-1038)^2+(1065-1038)^2+(1090-1038)^2}{3}$$

$$Var(B) = \frac{6241+729+2704}{3} \cong 3224{,}67$$

André Amorim
Finanças Corporativas

Anhanguera

28

8ª Aula – Medidas de dispersão

MQA

Exemplificando

Considerando os conjuntos $A = \{2, 3, 6, 9\}$ e , $B = \{959, 1065, 1090\}$ qual deles possui os dados mais dispersos em torno da média?

$$Dp(A) = \sqrt{Var(A)} = \sqrt{7{,}5} \cong 2{,}74$$

$$Dp(B) = \sqrt{Var(B)} = \sqrt{3224{,}67} \cong 56{,}79$$

$$CV(A) = \frac{Dp(A)}{\bar{a}} = \frac{2{,}74}{5} = 0{,}548 = 54{,}8\%$$

$$CV(B) = \frac{Dp(B)}{\bar{b}} = \frac{56{,}79}{1038} \cong 0{,}055 = 5{,}5\%$$

Como $CV(A) > CV(B)$, concluímos que o conjunto A é mais disperso que o conjunto B. Além disso, poderíamos acrescentar que A possui alta dispersão e B, baixa dispersão.

André Amorim Finanças Corporativas

Anhanguera

29

8ª Aula – Medidas de dispersão

MQA

FIM

André Amorim
Finanças Corporativas

Anhanguera

30