UFT

Mestrado em
Desenvolvimento Regional

Nivelamento em Estatística

Medidas de Dispersão

Prof. Dr. Adriano N. da Paixão

Tópico: Medidas de tendência central

- Referências:

  – HOFFMAN, R. **Estatística para economistas.** 4ª Edição Revisada e ampliada. São Paulo: Pioneira Thomson Learning, 2006. – Capítulo 5.

Organização da aula

1 – Introdução

2 – Amplitude, variância e desvio padrão

3 - Coeficiente de variação

4 – Exercício

1 – Amplitude, variância e desvio padrão

- Consideremos os seguintes conjuntos:

$$A = \{5,5,5,5,5\}$$

$$B = \{3,4,5,6,7\}$$

$$C = \{13,14,15,16,17\}$$

$$D = \{1,3,5,7,9\}$$

$$E = \{3,5,5,5,7\}$$

$$F = \{3,3,4,4,5,5,6,6,7,7\}$$

1 – Amplitude, variância e desvio padrão

- A média aritmética dos elementos do conjunto C é 15, e para todos os demais conjuntos a média aritmética dos seus elementos é 5.
- Com base no simples exame desses conjuntos, podemos fazer as seguintes afirmativas, relativas à dispersão dos valores dos elementos de cada conjunto:

1 – Amplitude, variância e desvio padrão

- 1º) O conjunto A apresenta dispersão nula;
- 2º) Os conjuntos B e C apresentam a mesma dispersão, só diferindo quanto à média;
- 3º) a dispersão de D é maior do que a dispersão de B, ainda, como para o conjunto D a diferença entre dois valores consecutivos é sempre igual a 2 e para o conjunto B é sempre igual a 1, parece lícito afirmar que a dispersão de D, é em certo sentido, igual ao dobro da dispersão de B;

1 – Amplitude, variância e desvio padrão

- 4º) A dispersão de E é maior do que a dispersão de A e menor do que a dispersão de B.
- 5º) o conjunto F apresenta dispersão igual à de B, pois esses dois conjuntos só diferem quanto ao número de elementos (o conjunto F é simplesmente uma duplicação de B).

1 – Amplitude, variância e desvio padrão

- Uma medida de dispersão é a amplitude total (*h*) que é a diferença entre o maior e o menor valor observado.
- *h* = 0, para o conjunto A;
- *h*= 4, para os conjuntos B, C, E e F ;
- *h* = 8 para o conjunto D.
- Nota-se que os valores da amplitude não obedecem à quarta afirmativa, pois eles indicam, erroneamente que os conjuntos B e E apresentam o mesmo grau de dispersão. Isso acontece porque a amplitude só leva em consideração os valores extremos.

1 – Amplitude, variância e desvio padrão

- Uma medida de dispersão de dispersão deve levar em consideração todos os dados, não apenas o maior e o menor valor observado.
- Os valores dos desvios em relação à média mostram o grau de dispersão dos dados. Para os conjuntos em análise, os conjuntos de desvios são:

$$e_i = X_i - \overline{X}$$

$$A' = \{0,0,0,0,0\}; B' = \{-2,-1,0,1,2\}$$
$$C' = \{-2,-1,0,1,2\}; D' = \{-4,-2,0,2,4\}$$
$$E' = \{-2,0,0,0,2\}; F' = \{-2,-2,-1,-1,0,0,1,1,2,2\}$$

1 – Amplitude, variância e desvio padrão

- É interessante notar que ao considerarmos os desvios em relação à média, eliminamos as diferenças entre os conjuntos B e C.

- No entanto, não se pode usar a soma dos desvios como medida de dispersão. Por que?

$$\sum (X_i - \overline{X}) = \sum X_i - n\overline{X} = 0$$

1 – Amplitude, variância e desvio padrão

- Então, para caracterizar a dispersão dos dados, devemos considerar os desvios independentemente do sinal, o que se pode obter tomando os desvios ao quadrado.

$$\sum_{i=1}^{n} e_i^2$$

$$A': \sum_{i=1}^{n} e_i^2 = 0; \quad B' = C': \sum_{i=1}^{n} e_i^2 = 10;$$

$$D': \sum_{i=1}^{n} e_i^2 = 40; \quad E': \sum_{i=1}^{n} e_i^2 = 8; \quad \mathrm{F}': \sum_{i=1}^{n} e_i^2 = 20;$$

1 – Amplitude, variância e desvio padrão

- Entretanto, $\sum_{i=1}^{n} e_i^2$ não pode ser usado como medida de dispersão, porque tende a crescer com o número de observações; embora os conjuntos B e F apresentem a mesma dispersão, o valor de $\sum_{i=1}^{n} e_i^2$ para F é duas vezes maior do que para B.

1 – Amplitude, variância e desvio padrão

- Uma forma de resolver esse problema é dividir a soma do quadrado dos desvios pelo número de observações, para obtermos a *variância* dos dados:

$$\sigma^2 = \frac{\sum_{i=1}^{n} e_i^2}{n} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}$$

Variância amostral

$$s^2 = \frac{\sum_{i=1}^{n} e_i^2}{n-1} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n-1}$$

1 – Amplitude, variância e desvio padrão

- A variância, como medida de dispersão, tem o inconveniente de apresentar unidade de medida igual ao quadrado da unidade de medida dos dados. Assim, por exemplo, se X é medido em kg, a variância é medida em kg^2.
- O desvio padrão é por definição, a raiz quadrada, com sinal positivo, da variância. A unidade de medida do desvio padrão é igual a unidade de medida dos dados.

$$\sigma = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n} = \frac{\sum_{i=1}^{n} \left(X_i - \overline{X}\right)^2}{n}}$$

1 – Amplitude, variância e desvio padrão

- Na tabela 1 são apresentados os valores da amplitude, da variância (σ^2) e do desvio padrão (σ) para os seis conjuntos considerados. Note que o valor do desvio padrão obedece as cinco afirmativas relativas ao grau de dispersão nesses seis conjuntos, feitas no início desta seção.

1 – Amplitude, variância e desvio padrão

Conjunto	Amplitude	Variância σ^2	Desvio padrão σ
A	0	0	0
B	4	2	$\sqrt{2}$
C	4	2	$\sqrt{2}$
D	8	8	$2\sqrt{2}$
E	4	1,6	$\sqrt{1,6}$
F	4	2	$\sqrt{2}$

1 – Amplitude, variância e desvio padrão

A seguinte igualdade, cuja demonstração deixamos para o leitor, pode ser bastante útil ao cálculo da soma dos quadrados dos desvios:

$$\sum_{i=1}^{n}\left(X_i - \overline{X}\right)^2 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} X_i\right)^2$$

No caso dos dados estarem agrupados em uma distribuição de freqüência, cada valor distinto ou valor central da classe $\left(X_j, \text{com } j = 1,2,\ldots,k\right)$, deve ser ponderado pela respectiva freqüência (f_j), e as expressões para calculo da variância ficam:

1 – Amplitude, variância e desvio padrão

$$\sigma^2 = \frac{\sum_{j=1}^{k} e_j^2 f_j}{n} = \frac{\sum_{j=1}^{k} \left(X_j - \overline{X}\right)^2 f_j}{n}$$

$$\sum_{j=1}^{k} \left(X_j - \overline{X}\right)^2 f_j = \sum_{j=1}^{k} X_j^2 - \frac{1}{n}\left(\sum_{j=1}^{k} X_j f_j\right)^2 \text{ com } n = \sum_{j=1}^{k} f_j$$

Observação: existem ainda duas outras medidas de dispersão: o Desvio Médio e a Diferença Média, as quais não trataremos neste curso.

2 – Coeficiente de variação

- O desvio padrão por si só não diz muita coisa. Assim, um $\sigma=2$ pode ser considerado pequeno para uma série de valores cujo valor médio é 200; no entanto, se a média for 20, o mesmo não pode ser dito.
- Além disso, o fato de o desvio padrão ser expresso na mesma unidade dos dados limita o seu emprego quando desejamos comparar duas ou mais séries de valores, relativamente à sua dispersão ou variabilidade, quando expressas em unidades diferentes.

2 – Coeficiente de variação

- Para contornar essa dificuldade e limitações, podemos caracterizar a dispersão ou variabilidade dos dados em termos relativos ao seu valor médio, medida essa denominada coeficiente de variação (CV).

$$CV = \frac{\sigma}{\overline{\overline{X}}} \text{x}\,100$$

2 – Coeficiente de variação

Conjunto	Desvio padrão σ	Média	Coeficiente de Variação (CV)
A	0	5	0
B	$\sqrt{2}$	5	28,28%
C	$\sqrt{2}$	15	9,43%
D	$2\sqrt{2}$	5	56,57%
E	$\sqrt{1,6}$	5	25,30%
F	$\sqrt{2}$	5	28,28%

1 – Amplitude, variância e desvio padrão

Conjunto	Amplitude	Variância σ^2	Desvio padrão σ
A	0	0	0
B	4	2	$\sqrt{2}$
C	4	2	$\sqrt{2}$
D	8	8	$2\sqrt{2}$
E	4	1,6	$\sqrt{1,6}$
F	4	2	$\sqrt{2}$

Exercício

1. São dados seis valores observados em uma amostra da variavel X: 9, 13, 6, 7,10 e 9. Determine, para essa amostra, a) os valores da média de X, b) a variância de X, c) o desvio padrão de X e d) a amplitude.
2. Ache a média, a mediana e a(s) moda(s) da seguinte amostra de vendas diárias para 20 varejistas:

7	11	8	6	11
8	10	9	10	5
10	10	8	6	7
4	6	11	9	6

3. Determine a média, a mediana e a(s) moda(s) da seguinte distribuição de frequência

Valor Observado	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**
Frequência	1	0	2	4	1	0	1	0	5	1

Exercício

4. A tabela seguir dá os valores observados e as respectivas frequencias de uma amostra de 15 observações de uma variável X.

X	Freqüência
9	1
8	5
0	1
2	2
6	1
3	4
4	1

Determine, para essa amostra, a) os valores da média de X, b) a variância de X, c) o desvio padrão de X e d) a amplitude.