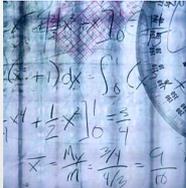


Estatística Aplicada

Noções Introdutórias e utilização do computador



Terceira aula - Estatística Descritiva II

Descrição dos dados

Resumindo numericamente

Para resumir numericamente dados quantitativos o objetivo é escolher medidas apropriadas de **locação** ("qual o tamanho dos números envolvidos?") e de **dispersão** ("quanta variação existe?") para os tipos de dados.

Estatística Descritiva

Pedro Menezes - 2008 2

- Medidas de tendência central
- **Medidas de dispersão**
- Separatrizes

Estatística Descritiva

Pedro Menezes - 2008 3

Estatística Descritiva

1. Medidas de dispersão absoluta
2. Medidas de dispersão relativas

Pedro Menezes - 2008 4

Estatística Descritiva

1. Medidas de dispersão absoluta

- Amplitude total
- Desvio Médio
- Variância
- Desvio Padrão

Pedro Menezes - 2008 5

Estatística Descritiva

Medidas de dispersão: Amplitude Total

Amplitude total (A_T): é a diferença entre o maior e o menor valor observado.

$A_T = x_{\max} - x_{\min}$	Grupo A = 71 - 69 = 2
	Grupo B = 83 - 60 = 23

Pedro Menezes - 2008 6

Estatística Descritiva

Medidas de dispersão: Desvio Médio

O **Desvio médio (D_M)** Analisa todos os desvios ou distâncias em relação a média aritmética.

$d_i = (x_i - x)$ onde, d_i = desvio ou distância x_i = valores observados x = média aritmética	A soma de todos os desvios em relação a média aritmética é igual a zero. $\sum d_i = \sum (x_i - x) = 0$
--	---

Pedro Menezes - 2008 7

Estatística Descritiva

Medidas de dispersão: Desvio Médio

Cálculo dos d_i :		Para eliminar a soma zero, coloca-se os desvios em módulo:	
Grupo A $d_1 = 70 - 70 = 0$ $d_2 = 71 - 70 = +1$ $d_3 = 69 - 70 = -1$ $d_4 = 70 - 70 = 0$ $d_5 = 70 - 70 = 0$ $\sum d_i = 0$	Grupo B $d_1 = 69 - 71 = -11$ $d_2 = 80 - 71 = +9$ $d_3 = 70 - 71 = -1$ $d_4 = 62 - 71 = -9$ $d_5 = 83 - 71 = +12$ $\sum d_i = 0$	Grupo A $d_1 = 0 = 0$ $d_2 = 1 = 1$ $d_3 = -1 = 1$ $d_4 = 0 = 0$ $d_5 = 0 = 0$ $\sum d_i = 2$	Grupo B $d_1 = -11 = 11$ $d_2 = 9 = 9$ $d_3 = -1 = 1$ $d_4 = -9 = 9$ $d_5 = 12 = 12$ $\sum d_i = 42$

Pedro Menezes - 2008 8

Estatística Descritiva

Medidas de dispersão: Desvio Médio

Dessa forma, é possível calcular a média dos desvios por:

$D_M = \frac{\sum d_i }{n} = \frac{\sum x_i - x }{n}$	Grupo A $D_M = \frac{\sum d_i }{n} = \frac{2}{5} = 0,4$	Grupo B $D_M = \frac{\sum d_i }{n} = \frac{42}{5} = 8,4$
---	---	--

Pedro Menezes - 2008 9

Estatística Descritiva

Medidas de dispersão: Variância

Repare nas duas amostras seguintes, que embora tenham a mesma média, têm uma dispersão bem diferente:

Pedro Menezes - 2008 10

Estatística Descritiva

Medidas de dispersão: Variância

Como a medida de localização mais utilizada é a média, será relativamente a ela que se define a principal medida de dispersão - a variância, apresentada a seguir.

Pedro Menezes - 2008 11

Estatística Descritiva

Medidas de dispersão: Variância amostral

Define-se a variância, e representa-se por s^2 , como sendo a medida que se obtém somando os quadrados dos desvios das observações da amostra, relativamente à sua média, e dividindo pelo número de graus de liberdade (n-1):

$$s^2 = \frac{(x_1 - \bar{x})^2 + (x_2 - \bar{x})^2 + (x_3 - \bar{x})^2 + \dots + (x_n - \bar{x})^2}{(n - 1)} = \frac{\sum_{i=1}^n (x_i - \bar{x})^2}{(n - 1)}$$

- Grandeza em duas dimensões

Pedro Menezes - 2008 12

Estatística Descritiva

Medidas de dispersão: Variância amostral

Os graus de liberdade indicam os espaços entre os dados; e são iguais a (n-1) porque os espaços entre eles estão sempre uma unidade abaixo do número dos próprios dados. São estes graus que representam a variação. n=1 não varia.

Pedro Menezes - 2008 13

Estatística Descritiva

Medidas de dispersão: Desvio Padrão

Uma vez que a variância envolve a soma de quadrados, a unidade em que se exprime não é a mesma que a dos dados. Assim, para obter uma medida da variabilidade ou dispersão com as mesmas unidades que os dados, tomamos a raiz quadrada da variância e **obtemos o desvio padrão**:

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

- Grandeza em uma dimensão

Pedro Menezes - 2008 14

Estatística Descritiva

Medidas de dispersão: Desvio Padrão

O desvio padrão é uma medida que **só pode assumir valores não negativos** e quanto maior for, maior será a dispersão dos dados.

Algumas propriedades do desvio padrão, que resultam imediatamente da definição, são:

- o desvio padrão é sempre não negativo e será tanto maior, quanta mais variabilidade houver entre os dados.
- se s = 0, então não existe variabilidade, isto é, os dados são todos iguais.

Pedro Menezes - 2008 15

Exemplo 1:

Exemplo: Na 2ª classe de certa escola o professor deu uma tarefa constituída por um certo número de contas para os alunos resolverem. Pretendendo determinar a dispersão dos tempos de cálculo, observam-se 10 alunos durante a realização da tarefa, tendo-se obtido os seguintes valores:

Aluno	Tempo (minutos)	$(x_i - \bar{x})$	$(x_i - \bar{x})^2$
1	13	-3,9	15,21
2	15	-1,9	3,61
3	14	-2,9	8,41
4	18	1,1	1,21
5	25	8,1	65,61
6	14	-2,9	8,41
7	16	-0,9	0,81
8	17	0,1	0,01
9	20	3,1	9,61
10	17	0,1	0,01
Σ	169	0,0	112,90

Pedro Menezes - 2008 16

Resposta 1:

Na tabela anterior juntamos duas colunas auxiliares, uma para colocar os desvios das observações em relação à média e a outra para escrever os quadrados destes desvios. A partir da coluna das observações calculamos a soma dessas observações, que nos permitiu calcular a **média = 16,9**. Uma vez calculada a média foi possível calcular a coluna dos desvios. Repare-se que, como seria de esperar, a soma dos desvios é igual a zero.

A soma dos quadrados dos desvios permite-nos calcular a **variância**

$$s^2 = \frac{112,9}{9} = 12,54$$

$$s = 3,54$$

Pedro Menezes - 2008 17

Medidas de dispersão: Desvio Padrão

Na representação gráfica ao lado visualizamos os **desvios das observações** relativamente à média (valores do exemplo anterior):

Pedro Menezes - 2008 18

Estadística Descritiva

Medidas de dispersão: Desvio Padrão

Propriedades para dados com distribuição aproximadamente normal:

Uma propriedade que se verifica se os dados se distribuem de forma aproximadamente **normal**, ou seja, quando o histograma apresenta uma forma característica com uma classe média predominante e as outras classes se distribuem à volta desta de forma aproximadamente simétrica e com freqüências a decrescer à medida que se afastam da classe média, é a seguinte:

Aproximadamente 68% dos dados estão no intervalo $[\bar{x} - s, \bar{x} + s]$



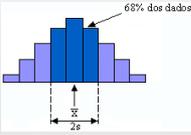
Pedro Menezes - 2008 19

Estadística Descritiva

Medidas de dispersão: Desvio Padrão

Propriedades para dados com distribuição aproximadamente normal:

1 - Aproximadamente 68% dos dados estão no intervalo $[\bar{x} - s, \bar{x} + s]$

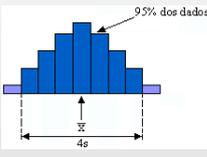


Pedro Menezes - 2008 20

Estadística Descritiva

Medidas de dispersão: Desvio Padrão

2 - Aproximadamente 95% dos dados estão no intervalo $[\bar{x} - 2s, \bar{x} + 2s]$



Pedro Menezes - 2008 21

Medidas de dispersão: Desvio Padrão

Propriedades para dados com distribuição aproximadamente normal:
Aproximadamente **100% dos dados** estão no intervalo

$$[\bar{x} - 3s, \bar{x} + 3s]$$

Pedro Menezes - 2008 22

Medidas de dispersão: Desvio Padrão

Executar Programa

Pedro Menezes - 2008 23

Medidas de dispersão: Desvio Padrão

Excel

Pedro Menezes - 2008 24

Estatística Descritiva

1. Medidas de dispersão absoluta

2. Medidas de dispersão relativas

Pedro Menezes - 2008 26

Estatística Descritiva

2. Medidas de dispersão relativa

Coefficiente de variação (CV) É uma medida relativa de dispersão útil para a comparação em termos relativos do grau de concentração em torno da média de séries distintas.

Amostra
s

$$CV = \frac{s}{\bar{X}} \cdot 100$$

Pedro Menezes - 2008 26

Estatística Descritiva

2. Medidas de dispersão relativa: CV

O coeficiente de variação é expresso em porcentagem. Duas maneiras de analisar o CV :

Pequena dispersão: CV ≤ 10%	Baixa dispersão: CV ≤ 15%
Média dispersão: 10% < CV < 20%	Média dispersão: 15% < CV < 30%
Grande dispersão: CV ≥ 20%	Grande dispersão: CV ≥ 30%

Pedro Menezes - 2008 27

Estatística Descritiva

- Medidas de tendência central
- Medidas de dispersão
- **Separatrizes**

Pedro Menezes - 2008 28

Estatística Descritiva

Medidas Separatrizes

- Medidas que dividem as distribuições em partes iguais.
- Servem para descrever posições em uma distribuição de dados.

Pedro Menezes - 2008 29

Estatística Descritiva

Medidas Separatrizes

- Quartil
- Decil
- Percentil

Pedro Menezes - 2008 30

Medidas Separatrizes: Quartil

Divide os dados em quatro partes iguais

Q1, Q2, Q3

Pedro Menezes - 2008 31

Medidas Separatrizes: Decil

Divide os dados em dez partes iguais

D1, D2, D3, D4..., D9

Pedro Menezes - 2008 32

Medidas Separatrizes: Percentil

Divide os dados em cem partes iguais

P1, P2, P3, P4, P5..., P99

Pedro Menezes - 2008 33

Estatística Descritiva

Medidas Separatrizes

Turma A

2	3	4	4	5	6	7	7	7	8	
Q_1	Q_2	Q_3					Q_4	P_{75}		
P_{25}	mediana									

Pedro Menezes - 2008 34

Estatística Descritiva

Exercício 1:

Dados:
3, 4, 1, 8, 5, 5, 9, 8, 7, 2, 6

Ache o Q2

Pedro Menezes - 2008 35

Estatística Descritiva

Resposta 1:

Dados:
1, 2, 3, 4, 5, 5, 6, 7, 8, 8, 9

Q2 = 5

Pedro Menezes - 2008 36

Exercício 2:

Dados:
2, 3, 4, 10, 1, 1, 9, 3, 5, 9, 8

Ache o Q3

Pedro Menezes - 2008 37

Resposta 2:

Dados:
1, 1, 2, 3, 3, 4, 5, 8, 9, 9, 10

Q3 = 9

Pedro Menezes - 2008 38

Medidas Separatrizes: Cálculo Quartil

$Q_n = n (N+1)/4$

Ex. 1, 2, 3, 4, 5, 6, 7, 8, 10, 10, 11

$Q_3 = 3 (11+1)/4 = 9^o = 10$

Pedro Menezes - 2008 39

Estadística Descriptiva

Medidas Separatrizes: Cálculo Decil

$D_n = n(N)/10$

Ex. 1, 2, 3, 3, 5, 6, 7, 8, 9, 10

$D_4 = 4(10)/10 = 4^\circ = 3$

Pedro Menezes - 2008 40

Estadística Descriptiva

Medidas Separatrizes: Cálculo Percentil

$P_n = n(N)/100$

Ex. 1, 2, 3, 4, 5, 6, 7, 8, 9, 10

$P_{50} = 50(10)/100 = 5^\circ = 5$

Pedro Menezes - 2008 41

Estadística Descriptiva

Medidas Separatrizes

Excel

Pedro Menezes - 2008 42
