ESTATÍSTICA

Introdução à Estatística

Prof. Dr. Ricardo da Silva Manca

Medidas Descritivas

Medidas Descritivas

- Medidas de Tendência Central
- Medidas Separatrizes
- Medidas de Dispersão ou Variabilidade

Medidas de Tendência Central

- Servem para termos uma idéia acerca dos valores médios da variável em estudo.

- São usados para sintetizar em um único número os dados observados.

- São exemplos de medidas de tendência central: Média, Moda e Mediana.

- A escolha de qual medida usar, depende…

Média Amostral

- Se os dados consistem de *n* observações $x_1, x_2,\ldots,x_n$, a média é dada pela soma das observações dividida pelo o número de observações. Por exemplo, se os dados são $x_1=2$, $x_2=3$, $x_3=1$, então a média é $(2+3+1)/3=2$.

- A média amostral é definida por :

$$\overline{X} = \frac{x_1 + x_2 + x_3 + \ldots + x_n}{n}$$

Média Amostral - Exemplo

Turma A : 2 3 4 4 5 6 7 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 7 8 9

Objetivo: Obter a média de cada turma:

Turma A

(2+3+4+4+5+6+7+7+7+7+8) / 11 = 60/11

Média turma A = 5,45

Turma B

(2+3+4+4+4+5+6+7+7+8+9)/11 = 59/ 11

Média turma B = 5,36

Mediana

- Divide uma distribuição ordenada de dados em duas partes iguais.

- A mediana (Md) á a observação central, depois de ordenada a amostra.

- Se a amostra tiver dimensão ímpar, a mediana coincide com a observação central.

  Exemplo: Na amostra
  1.2; 1.7; 2.1; 2.2; 2.4 a mediana é 2.1

- Se a amostra tiver dimensão par, a mediana toma o valor da média das duas observações mais centrais.

Mediana

- Para calcularmos a mediana é preciso ordenarmos os dados: $x_{(1)}, x_{(2)}, \ldots, x_{(n)}$.

- A mediana de um conjunto de dados é:

  Md = $x_{(n+1/2)}$, se n é ímpar

  Md = $[x_{(n/2)} + x_{(n/2+1)}]/2$, se n é par

- A mediana é mais robusta que a média a erros ou a observações afastadas.

Mediana - Exemplo

Exemplo 1:Turma A **:** 2 3 4 4 5 6 7 7 7 7 8

Turma B **:** 2 3 4 4 4 5 6 7 7 8 9

Turma A : *Mediana* = 6

Turma B : Mediana = 5

Exemplo 2**:** Turma A : 2 3 4 4 5 6 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 8 9

Turma A : *Mediana* = (5+6)/2=5,5

Turma B : Mediana = (4+5)/2=4,5

Mediana - Exemplo

Caso	1	2	3	4	5	6	7	8
x_i	x_1	x_2	x_3	x_4	x_5	x_6	x_7	x_8
Valores	2	4	5	5	7	9	10	30

Qual a média e a mediana ? Resposta: 6 e 5

Qual a média e a mediana ao acrescentarmos a observação 8?

Resposta: 9 e 6

Moda

- Valor que ocorre com maior freqüência.
- Obtida por inspeção da tabela de distribuição de freqüências.
- Ao contrário do que acontece com a mediana e a média, uma amostra pode possuir mais do que uma moda.

Moda - Exemplo

Turma A : 2 3 4 4 5 6 7 7 7 7 8

Turma B : 2 3 4 4 4 5 6 7 7 8 9

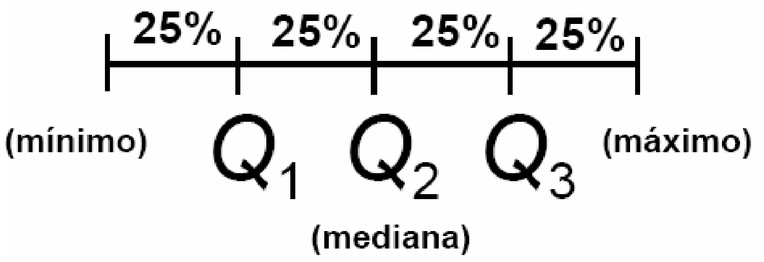
- *Moda turma A* = 7
- *Moda turma B* = 4

Medidas Separatrizes

- Medidas que separam a distribuição em partes iguais.
  - **Quartis**
  - **Decis**
  - **Percentis**

Quartis

- Quartis são os valores (Q_1, Q_2 e Q_3) que dividem a amostra, depois de ordenada, em quatro partes iguais (ou o mais iguais possível).



- Obtendo os quartis
  - Ordena-se os dados;
  - Calcula-se a posição do quartil através da fórmula: $P_{Qi} = i \cdot \frac{n}{4}$
  - O quartil será o valor que ocupa a posição calculada anteriormente.

Decis

- Dividem um conjunto de dados em dez partes iguais

- Encontra-se o valor do decil desejado, procedendo-se como no caso dos quartis, sendo a posição do decil, encontrada por:

$$P_{Di} = i \cdot \frac{n}{10}$$

Percentis

- Dividem um conjunto de dados em cem partes iguais

- Procede se como no caso dos quartis, sendo que para o cálculo da posição do percentil , a fórmula será:

$$P_{Pi} = i \cdot \frac{n}{100}$$

Medidas Sepatrizes - Exemplo

Turma A

2 3 4 4 5 5 7 7 7 8 8

$\Downarrow$ $\Downarrow$ $\Downarrow$

Q_1 Q_2 Q_3

P_{25} P_{50} P_{75}

Md

Medidas de Variabilidade

- Medidas de tendência central são descritores insuficientes de uma amostra.
- São necessárias medidas que reflitam a variação dentro de um conjunto de dados (medidas de variabilidade).
- Essas medidas serão pequenas se os dados forem próximos e grandes se eles estiverem muito espalhados.
- Além disso, tais medidas devem permitir comparar amostras de diferentes tamanhos e determinar se uma amostra é mais variável (ou heterogênea) que a outra.

Exemplo

- Os dados abaixo referem-se aos pesos dos pacientes em dois grupos:

	Grupo A	Grupo B
	78	65
	80	69
	82	78
	85	85
	85	85
	85	93
	86	96
	88	98
Soma	669	669
Média	83,6	83,6
Mediana	85	85
Moda	85	85
N	8	8

Amplitude Total

- Diferença entre o maior e o menor valor do conjunto de dados.

<u>Grupo A</u>

AMPLITUDE TOTAL = 88 – 78 = 10

<u>Grupo B</u>

AMPLITUDE TOTAL = 98 – 65 = 33

AT (grupo A) < AT (grupo B)

Variância

- É um indicativo da dispersão de um conjunto de dados em relação à média.

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

- A variância populacional é denotada por σ^2. Usualmente σ^2 é desconhecida.
- A variância amostral é denotada por S^2. Desvantagem - não é expressa na unidade de medida do dado original.

Desvio Padrão

- Corresponde à raiz quadrada da variância, tendo portanto a mesma unidade da variável que está sendo estudada. O desvio padrão será denotado por *S*.

- É a medida mais usada na comparação de diferenças entre grupos.

- Fornece um número que permite especificar quão acima ou quão abaixo da média está um determinado valor.

- Quanto maior o desvio padrão, maior a variabilidade dos dados.

Coeficiente de Variação

- Muitas vezes o desvio padrão pode ser considerado grande ou pequeno dependendo da ordem de grandeza da variável.
- Pode-se obter um índice relativo de dispersão:

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

- Alguns analistas consideram:

Baixa dispersão: CV ≤ 15%

Média dispersão: 15% < CV < 30%

Alta dispersão: CV ≥ 30%