

NOTAS DE AULA DE ESTATÍSTICA
PROFESSOR: ATAUALPA MAGNO FERRAZ DE NOVAES

Prezados alunos, desejo que vocês possam desfrutar destas notas de aula.

Um agradecimento muito especial aos meus alunos Dionei Mascarenhas, Leonardo Leandro e Caio Moreira pelo incentivo e pelo constante auxílio que eles me prestam ao longo do tempo em que trabalho na UCSal.

Eles são não apenas meus amigos, mas também “amigos” do Curso de Informática da Católica, disponibilizando estas notas de aula para os colegas.

Desde que desejamos aprimorar este trabalho ao longo do tempo, sugestões e críticas serão bem vindas.

Email: amferraznovaes@ig.com.br ou ataualpa@im.ufba.br.

ESTATÍSTICA

CONCEITO → É a ciência que se preocupa com a coleta, análise, interpretação e apresentação dos dados e tem como objetivo fundamental o estudo de uma ou mais populações.

DIVISÃO: Estatística Descritiva e Estatística Indutiva (ou Inferencial)

Estatística Descritiva → é aquela que se preocupa com a coleta, análise, interpretação e apresentação dos dados

Estatística Indutiva (ou Inferencial ou ainda amostral) → É aquela que partindo de uma amostra, estabelece hipóteses sobre a população de origem e formula previsões fundamentando-se na Teoria das Probabilidades.

POPULAÇÃO → É todo conjunto, finito ou infinito, que possui ao menos uma característica em comum entre todos os seus elementos componentes.

CARACTERÍSTICAS: Se divide em Atributos e Variáveis.

Atributos → Características que não podem ser medidas. Exemplos: Cor, Religião, Estado Civil, Sexo, etc.

Variáveis → São as características que podem ser medidas. Dividem-se em:

Discretas: São aquelas variáveis que podem assumir apenas alguns valores num intervalo de observação.

Exemplo → Número de filhos

Contínuas → São aquelas variáveis que podem assumir quaisquer valores num intervalo de observação.

Exemplo: Temperatura

CENSO → Coleta exaustiva das informações de “N “ características da População.

AMOSTRAGEM → É o processo de coleta das informações de parte da População (chamada de amostra), mediante métodos adequados de seleção destas unidades. Amostra, em Estatística, é uma parte da População adequadamente selecionada de acordo com uma regra ou um plano de modo a refletir da forma mais fiel possível o comportamento de toda a População.

SÉRIES ESTATÍSTICAS: Série Estatística é qualquer coleção de dados colocada numa tabela e classificada segundo as variações do fenômeno observado.

DIVISÃO DAS SÉRIES ESTATÍSTICAS:

1) Séries de dados não grupados:

- Temporais (Variação do fenômeno em função do Tempo)
- Geográficas (Variação do fenômeno em função da Região)
- Especificativas (Variação do fenômeno em função da Espécie)
- Mistas (Variação do fenômeno em função de mais de uma das anteriores)

2) Séries de dados grupados:

- Distribuição de Frequências por Intervalos (Série Estatística na qual a variável observada está dividida em subintervalos do Intervalo total, **variável contínua**)
- Distribuição de Frequências por Pontos (Série Estatística na qual as frequências observadas estão associadas a um ponto real observado, **variável discreta**)

RESUMO

ESTATÍSTICA (ORIGEM LATINA) → “Ciência dos Negócios do Estado”.

A Estatística trabalha com métodos científicos para coleta, organização, resumo e apresentação de dados e também para obtenção de conclusões e tomadas de decisão.

Amostra é uma parte da População estudada.

As Fontes dos Dados são de grande importância para que a análise do pesquisador reflita, ao máximo a realidade. As fontes de dados se dividem em dois tipos:

- **A Fonte de Dados é Primária** quando as informações são colhidas diretamente pelo pesquisador ou por seus auxiliares.
- **A Fonte de Dados é Secundária** quando o pesquisador recorre a relatórios, revistas, livros ou dados coletados por instituições especializadas.

Seja S um espaço amostral; definimos variável aleatória a toda função real definida sobre S . O conjunto de valores que a variável aleatória X pode assumir pode ser discreto (por exemplo, apenas valores inteiros) ou contínuo (por exemplo, um intervalo de números reais). Observe as definições abaixo.

VARIÁVEL ALEATÓRIA DISCRETA

Suponha que uma instituição de ensino esteja interessada em saber qual o número de alunos presentes às aulas de um determinado professor, em certo período da vida escolar. Se X simbolizar esse número, então X será uma variável que só pode assumir valores inteiros, inclusive zero. Mais precisamente, diz-se que X será uma variável discreta quando a menor diferença não nula entre dois valores possíveis dessa variável for finita.

Em muitos casos a variável discreta resulta da contagem, razão pela qual seus valores são expressos através de números inteiros não negativos.

Do ponto de vista formal, dizemos que a variável aleatória é discreta quando o conjunto de valores que ela pode assumir for um conjunto enumerável (que é um conceito eminentemente matemático; não entraremos em maiores detalhes), alguns exemplos de conjuntos enumeráveis: N, Z e Q (e seus subconjuntos).

VARIÁVEL ALEATÓRIA CONTÍNUA

Diz-se que X é uma variável aleatória contínua quando, ao passar de um valor real a para outro valor real b, assume todos os valores intermediários entre a e b. Para se fazer uma idéia concreta da variável aleatória contínua, basta pensar em um filete de mercúrio de um termômetro; ao dilatar-se o filete ele passará de uma leitura a para uma leitura b, passando por todas as temperaturas intermediárias.

A variável aleatória contínua, normalmente, assume valores num intervalo real ou no próprio conjunto dos números reais R.

Seja S um espaço amostral; definimos variável aleatória a toda função real definida sobre S. O conjunto de valores que a variável aleatória X pode assumir pode ser discreto (por exemplo, apenas valores inteiros) ou contínuo (por exemplo, um intervalo de números reais). Observe as definições abaixo.

MEDIDAS ESTATÍSTICAS

(MÉDIAS, MODA, MEDIANA, DESVIO PADRÃO E VARIÂNCIA)

MEDIDAS DE POSIÇÃO

Vamos supor que uma prova seja aplicada aos alunos de uma sala de aula. Imaginemos que, fornecidos os resultados, um determinado aluno queira saber qual a *posição* de sua nota em relação às demais. Para esclarecer essas dúvidas, o aluno deve lançar mão de medidas denominadas medidas de posição: A MODA, A MEDIANA E A MÉDIA (OU MÉDIAS).

MÉDIAS:

1) Média Aritmética (Simples) : é o quociente da soma dos números dos quais se deseja tirar a média, pela quantidade deles.

EXEMPLOS:

Dados os números abaixo, determine a média aritmética deles:

a) 15 e 17

Como são dois números, temos: Média Aritmética = $\frac{15 + 17}{2} = 16$

b) 12, 4, 23 e 37

Como são quatro números, temos: Média Aritmética = $\frac{12 + 4 + 23 + 37}{4} = 19$

2) Média Aritmética Ponderada: É um caso particular da média aritmética simples. Veja o exemplo abaixo:

Calcular a média ponderada dos números: 2; 3; 3; 3; 4; 4; 8; 8; 8; 8; 12; 23; 23

Podemos resolver como se tratasse de uma média aritmética simples:

$$\frac{2 + 3 + 3 + 3 + 4 + 4 + 8 + 8 + 8 + 8 + 12 + 23 + 23}{13} = \frac{109}{13} \approx 8,38$$

Ou podemos fazer do seguinte modo:

Contar quantas vezes aparece cada número, multiplicar esse resultado pelo número e, por fim, dividir o resultado pelo total de números (considerando inclusive a quantidade de vezes que o número se repete), assim:

$$\frac{2 \times 1 + 3 \times 3 + 4 \times 2 + 8 \times 4 + 12 \times 1 + 23 \times 2}{13} = \frac{109}{13} \approx 8,38 \text{ ou podemos fazer do seguinte modo:}$$

☹ **Dá no mesmo, notou?**

3) **Média Geométrica:** Podemos dizer que se parece com o conceito de média aritmética, só que ao invés de divisão, usa o conceito de raiz (quadrada, cúbica, quinta. Depende da quantidade de números) da multiplicação dos números. Veja os exemplos abaixo:

a) Qual a média geométrica dos números 12 e 3.

Resolução: Como são dois números, usamos a raiz quadrada:

$$\text{Média Geométrica} = \sqrt{12 \times 3} = \sqrt{36} = 6$$

b) Qual a média geométrica dos números 2, 4 e 8.

Resolução: Como são três números, usamos a raiz cúbica:

$$\text{Média Geométrica} = \sqrt[3]{2 \times 4 \times 8} = \sqrt[3]{64} = 4$$

c) Qual a média geométrica dos números 3, 5, 27 e 125.

Resolução: Como são quatro números, usamos a raiz quarta:

$$\text{Média Geométrica} = \sqrt[4]{3 \times 5 \times 27 \times 125} = \sqrt[4]{3^4 \times 5^4} = 3 \times 5 = 15$$

4) **Média Harmônica:** É o inverso da média aritmética dos inversos dos números dados.

Veja os exemplos abaixo:

a) Qual a média harmônica dos números 1/5 e 1/9?

Primeiramente, invertamos os números dados, isto é, teremos 5 e 9, depois calculamos a média aritmética

$$\text{simples: } \frac{5 + 9}{2} = 7; \text{ finalmente, invertamos este último resultado: } 1/7 \text{ (resposta).}$$

b) Qual a média harmônica dos números 2; 3 e 4 ?

Primeiramente, invertemos os números dados, isto é, teremos $1/2$; $1/3$ e $1/4$, depois calculamos a média

aritmética simples: $\frac{\frac{1}{2} + \frac{1}{3} + \frac{1}{4}}{3} = \frac{\frac{13}{12}}{3} = \frac{13}{36}$; Finalmente, invertemos este último número: $36/13$.

MODA

Chamamos de moda (M_o) de uma distribuição ao valor que nela aparece com maior frequência.

EXEMPLOS:

a) Consideremos uma pesquisa na qual foram obtidos os resultados que constam na seguinte lista:

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

O resultado que aparece com maior frequência (6 vezes) é o 7, logo, Moda = $M_o = 7$.

b) Consideremos uma pesquisa na qual foram obtidos os resultados presentes na lista abaixo:

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

Os resultados que aparecem com maior frequência (3 vezes) são 1 e 6; logo, essa distribuição possui duas modas (distribuição bimodal) e elas são: Moda 1 = $M_{o1} = 1$ e Moda 2 = $M_{o2} = 6$

c) Se todos os valores de uma distribuição forem distintos, diremos que a distribuição não possui moda. Por exemplo, a distribuição abaixo não possui moda:

1, 2, 3, 4, 5, 6, 7, 8

d) Considere a distribuição por intervalos:

NÚMERO DE PONTOS	FREQÜÊNCIA
[0; 20 [	4
[20; 40 [	<u>21</u> (<i>Maior Frequência</i>)
[40; 60 [	16
[60; 80 [	6
[80; 100 [	3
TOTAL	50

Se conhecermos apenas a tabela acima, a moda é calculada do seguinte modo:

Para determinarmos a moda, em casos como este, procedemos do seguinte modo:

1º Passo: Observamos o intervalo de maior frequência que no caso é [20; 40 [(pois a maior frequência é 21, como visto na tabela acima)

2º Passo: Calculamos a seguir o ponto médio do intervalo acima, através da média aritmética dos extremos:

$$\frac{20 + 40}{2} = 30 . \text{ Portanto: } \text{Moda} = M_o = 30$$

MEDIANA

Chamamos de mediana (Md) de uma distribuição ao valor que ocupa a posição central, quando todos os valores são colocados em ordem.

EXEMPLOS:

a) Dados os números: 3, 4, 5, 6, 7; a mediana é o número 5, pois há 2 valores antes e dois valores depois dele, ou seja, 5 é o termo central. Em símbolos: $Md = 5$

b) Consideremos uma pesquisa na qual foram obtidos os resultados que constam na seguinte lista:

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

Observe que há 21 termos, no rol acima, o número 6 (sublinhado) é o termo central (11^o) e portanto é a mediana, isto é: Mediana = $Md = 6$.

c) Se o número de valores que aparece na lista for par, a mediana será a média aritmética dos dois valores centrais, quando todos eles são colocados em ordem. Veja o exemplo abaixo:

1, 2, 3, 4, 5, 6, 7, 8

Os dois valores centrais são 4 e 5, daí, a mediana será: $Md = \frac{4 + 5}{2} = \frac{9}{2} = 4,5$

d) Considere a distribuição por intervalos:

NÚMERO DE PONTOS	FREQÜÊNCIA
[0; 20 [	3
[20; 40 [	20
[40; 60 [	10
[60; 80 [	9
[80; 100 [	8
TOTAL	50

1^o Passo: Identificamos o intervalo em que cai a mediana , levando em conta que ela deve dividir ao meio o número de valores observados (50); a 25^a observação está no intervalo [40; 60 [. Antes desse intervalo foram feitas 23 observações (3 + 20) e depois dele 17 observações (9 + 8);

2^o Passo: Usando uma regra de três simples, fazemos uma interpolação como indicada abaixo:

Md – 40	_____	25 – 23
60 – 40	_____	33 – 23
Logo, $Md = 44$		

MEDIDAS DE DISPERSÃO

Consideremos a distribuição cujos resultados constam na seguinte lista:

4, 6, 7, 8, 10

Sabemos que a média aritmética dessa distribuição é: $M = \frac{4 + 6 + 7 + 8 + 10}{5} = 7$

Chamamos de desvio de cada valor à diferença entre esse valor e a média da distribuição. Assim:

O desvio do valor 4 é $4 - 7 = -3$

O desvio do valor 6 é $6 - 7 = -1$

O desvio do valor 7 é $7 - 7 = 0$

O desvio do valor 8 é $8 - 7 = 1$

O desvio do valor 10 é $10 - 7 = 3$

Com o auxílio da idéia de desvio, vamos definir três medidas denominadas medidas de dispersão: o desvio médio, a variância e o desvio padrão.

DESVIO MÉDIO

Chamamos de desvio médio (DM) de uma distribuição à média aritmética dos módulos dos desvios. No exemplo analisado, o desvio médio é:

$$DM = \frac{|-3| + |-1| + |0| + |1| + |3|}{5} = 1,6$$

VARIÂNCIA

Chamamos de variância (V) de uma distribuição à média aritmética dos quadrados dos desvios. Assim, no exemplo dado:

$$V = \frac{(-3)^2 + (-1)^2 + (0)^2 + (1)^2 + (3)^2}{5} = \frac{20}{5} = 4$$

DESVIO PADRÃO

Denominamos desvio padrão (DP) de uma distribuição à raiz quadrada da variância:

No exemplo analisado, o desvio padrão é:

$$DP = \sqrt{V} = \sqrt{4} = 2$$

RESUMO

MEDIDAS ESTATÍSTICAS

MEDIDAS DE TENDÊNCIA CENTRAL OU MEDIDAS DE POSIÇÃO:

- MÉDIA ARITMÉTICA SIMPLES (μ) $\rightarrow x_{\min} \leq \mu \leq x_{\max}$;

$$\mu = \frac{\sum x_i}{N} \text{ (dados não grupados) ou } \mu = \frac{\sum f_x}{N} \text{ (dados grupados)}$$

- **MEDIANA (M)** → É a medida estatística que divide exatamente ao meio um conjunto de valores observados.

Exemplos (dados não agrupados):

- 1) Valores observados → 2, 4, 5, 7, 10, 11, 12. **Mediana = 7**
- 2) Valores observados → 2, 4, 5, 7, 10, 11. **Mediana = $\frac{5+7}{2} = 6$**

- **MODA (m)** → É uma medida estatística que representa o valor ou os valores que mais se repetem:

Exemplos (dados não agrupados):

- 1) Valores observados → 4, 3, **8**, 9, **8**, 12. **Moda = 8** (Sequência de dados unimodal).
- 2) Valores observados → **5**, 4, **5**, 6, 7, 8, 9, 4, 3, 8, 9, **5**. **Moda = 5** (Sequência de dados unimodal)
- 3) Valores observados → **4**, **8**, **7**, **4**, 5, **7**, 9, **8**, 6. **Modas = 4, 7 e 8** (Sequência de dados trimodal)

OBSERVAÇÃO: Para uma sequência de dados agrupados há vários tipos de Moda: Moda de Czuber, Moda de King, Moda de Pearson, etc.

MEDIDAS DE DISPERSÃO:

- DESVIO PADRÃO
- VARIÂNCIA