ESTATÍSTICA BÁSICA

CONCEITOS E EXERCÍCIOS

EDGARD GONÇALVES CARDOSO

Ementa

- Noções de estatística
- Histórico
- Coletas de dados
- Planejando a coleta de dados
- Medidas de ordenamento ou posição
- Moda, Média, Mediana, Variância e Desvio Padrão
- Exercícios de fixação de conceitos

NOÇÕES DE ESTATÍSTICA

- A Estatística durante séculos foi usada inconscientemente pelos povos como um caráter meramente descritivo e de registro de ocorrências.

- As primeiras atividades foram por volta de 2000 a.C. e foram usados no recenseamento das populações agrícolas chinesas.

- No início do século XIX, os grandes matemáticos entraram em cena, como exemplo, o francês Simon Laplace e o alemão Carl Friedrich Gauss (1777–1855), este último surge com aplicações da “distribuição normal" para modelagem de erros de medição.

- A teoria da distribuição normal foi usada pelo astrônomo e matemático belga Adolphe Quételet (1796 –1874), no estudo estatístico de diversas características das populações humanas: altura, peso, natalidade, mortalidade, renda mensal etc.

- Ronald Aylmer Fisher (1890 – 1962), estatístico britânico, foi o gênio que criou a moderna teoria da estatística.

- Na Estatística, Fisher trabalhou com ajustes de curvas de frequências, com coeficientes de correlação, os chamados coeficientes de Fisher, na análise de variância (ANOVA) e nas técnicas de estimação dos parâmetros.

- Influenciado pelos trabalhos de Karl Pearson, outro importante estatístico britânico.

- Fisher utilizou os resultados que obteve na Estatística como ferramentas para aplicação nos seus estudos de genética, sendo hoje considerado um dos maiores nomes na Teoria de Estatística e na Estatística aplicada à Biologia.

- Em geral, manipulamos um conjunto de dados com o objetivo de extrairmos informação sobre o comportamento de um processo ou produto.

- A Estatística utiliza a variabilidade presente nos dados para obter tal informação.

- A variabilidade está presente em todo lugar. Por exemplo, a posição de um carro estacionado em uma garagem não é a mesma ao longo dos dias.

- Neste caso, a posição do carro apresenta uma variação. Nossa estratégia consiste em avaliar as variações e obter informações através dela.

A aplicação de técnicas estatísticas envolve várias etapas:

- Coleta de dados;
- Exposição dos dados;
- Modelos Estatísticos.

- Não se deve coletar dados sem que antes se tenha definido claramente o problema ou situação a ser enfrentada, bem como os objetivos com relação aos mesmos;

COLETA DE DADOS

- Os sistemas de medição (instrumento, operadores, método, meio) que serão utilizados devem ser avaliados e ter capacidade de medição suficiente;

- Os cálculos e leituras devem ser feitos com muita atenção para evitar distorções;

- Devem ser utilizados métodos adequados para coleta de dados de acordo com o problema estudado.

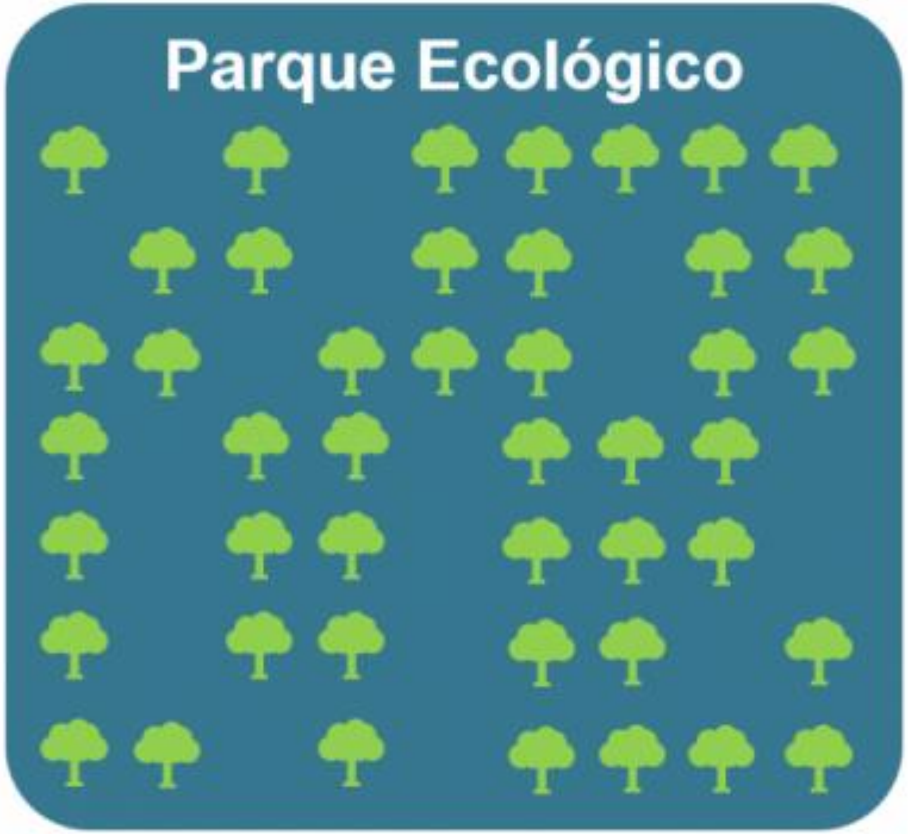
- Uma amostra é uma parcela de uma população que pode conter informações sobre esta população.

- Outra definição importante (para a escolha da técnica estatística e das interpretações dos resultados) é a classificação dos dados.

População e Amostra

População: Conjunto de elementos que possui a mesma característica de interesse.

Amostra: Qualquer subconjunto de elementos da população de interesse.




Dados
Quantitativos
Qualitativos
Nominais
Ordinais
Discretos
Contínuos

- Como a variável sexo é QUALITATIVA NOMINAL, isto é, não há uma ordem natural em suas categorias, a ordem das linhas da tabela pode ser qualquer uma.

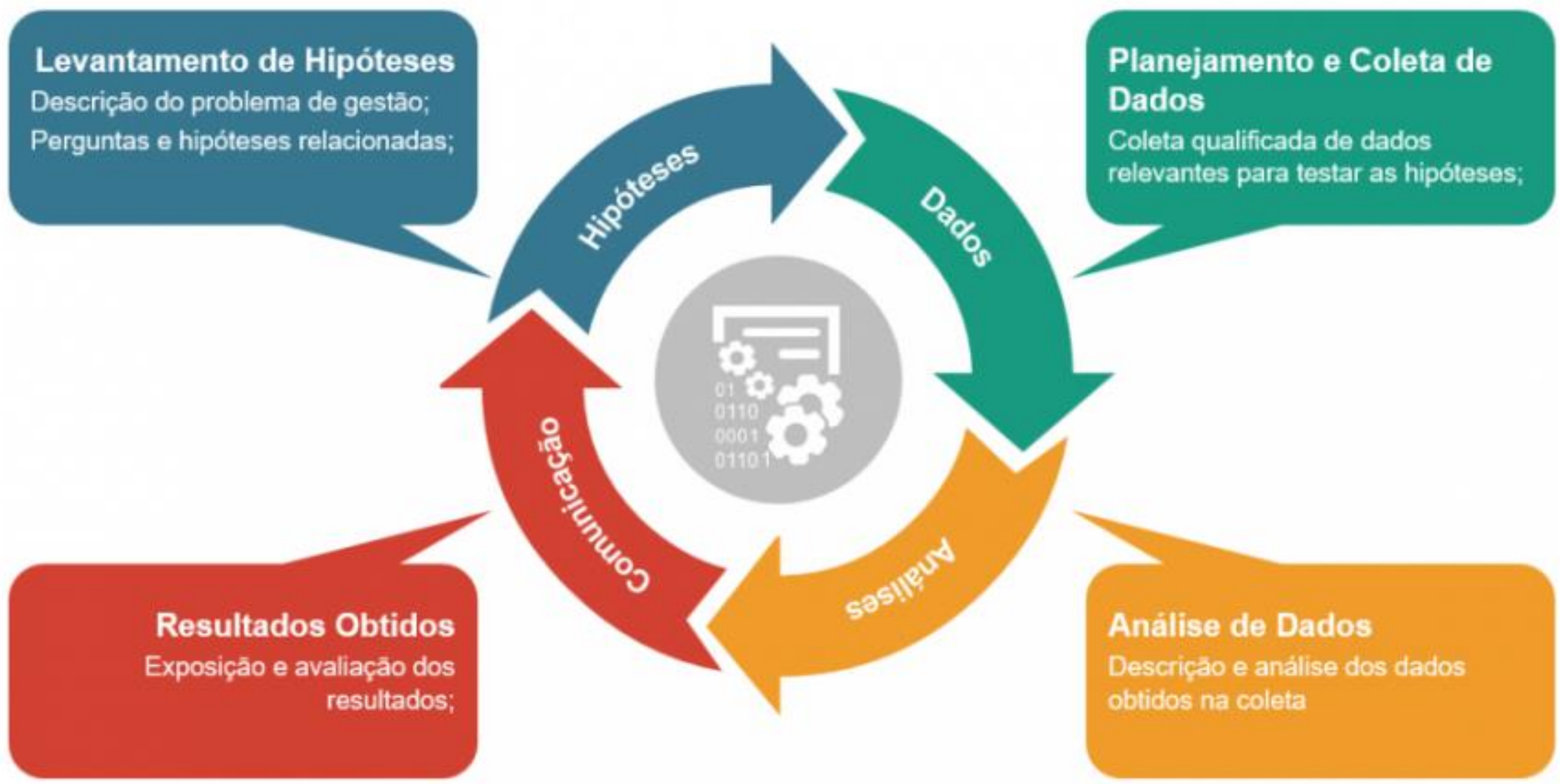
Sexo	Frequência Absoluta	Frequência Relativa
Feminino	35	36%
Masculino	62	64%
Total	97	100%

- Quando a variável tabelada for do tipo QUALITATIVA ORDINAL, as linhas da tabela de frequências devem ser dispostas na ordem existente para as categorias.

Mês de Observação	**Frequência Absoluta**	**Frequência Relativa**	**Frequência Absoluta Acumulada**	**Frequência Relativa Acumulada**
Abril	8	8%	8	8%
Maio	6	6%	14	14%
Junho	6	6%	20	21%
Julho	11	11%	31	32%
Agosto	23	24%	54	56%
Setembro	20	21%	74	76%
Outubro	14	14%	88	91%
Novembro	9	9%	97	100%
Total	**97**	**100%**		

PLANEJANDO A COLETA DE DADOS

- Para estudarmos adequadamente uma população através de uma amostra, devemos planejar a coleta de dados. Com este objetivo, formulamos algumas perguntas:
  - Com que frequência ocorrem os problemas?
  - Quais são as causas potenciais do problema?


Levantamento de Hipóteses
Descrição do problema de gestão;
Perguntas e hipóteses relacionadas;
Planejamento e Coleta de Dados
Coleta qualificada de dados relevantes para testar as hipóteses;
Hipóteses
Dados
Análises
Comunicação
01
0110
0001
01101
Resultados Obtidos
Exposição e avaliação dos resultados;
Análise de Dados
Descrição e análise dos dados obtidos na coleta

- Um bom planejamento para coleta de dados deve considerar as seguintes perguntas:
  - Qual a pergunta a ser respondida?
  - Como comunicar a resposta obtida?
  - Qual ferramenta de análise pretendemos usar e como utilizar os resultados?
  - Qual tipo de dado é necessário para utilizar as ferramentas desejadas e responder a pergunta?
  - Como coletar esses dados com o mínimo de esforço e erro?
  - Onde acessar estes dados?
  - Quem pode nos fornecer os dados?
  - Qual o período em que os dados serão coletados?

- Tendo as respostas para estas perguntas, devemos:
  - Construir uma metodologia para nos certificar de que todas as informações estão definidas;
  - Coletar os dados de forma consistente e honesta;
  - Certificar-se de que existe tempo suficiente para a coleta de dados;
  - Definir quais informações adicionais serão necessárias para estudos futuros, referências ou reconhecimento.

http://www.portalaction.com.br/en/node/87

MEDIDAS DE ORDENAMENTO OU POSIÇÃO

Medidas de Tendência Central

VALORES CENTRAIS OU MÉDIAS DE UMA AMOSTRA: INDICAM POSIÇÃO DE CENTRALIDADE, OU O PONTO CENTRAL DA DISTRIBUIÇÃO.

MÉDIA ARITMÉTICA SIMPLES: QUOCIENTE DA SOMA DOS VALORES OBSERVADOS PELO NÚMERO TOTAL DE VALORES.

Alunos	Idades dos alunos
1	23
2	21
3	21
4	20
5	19
6	20
7	19
8	20
9	20
10	20
11	22
12	35
13	28
14	19
15	18
16	20
17	18

Idades	Frequência
18	2
19	3
20	6
21	2
22	1
23	1
28	1
35	1

- Média Aritmética Ponderada: Quando há valores que se repetem.

$$\bar{X} = \frac{\sum_{i=1}^{k} X_m \cdot f_i}{\sum_{i=1}^{k} f_i}$$

Idades	Frequência
18	2
19	3
20	6
21	2
22	1
23	1
28	1
35	1

$$\bar{X} = \frac{18x2 + 19x3 + 20x6 + 21x2 + 23x1 + 28x1 + 35x1}{17}$$

$$\bar{X} = 20{,}06$$

Em probabilidade esta média é chamada ***Esperança Matemática****.*

- Mediana

Mediana: Medida de posição central.

- A mediana é o valor que ocupa a posição central (meio) da distribuição.

- Série de valores com quantidade impar de termos
  - Mediana = $Md=((n+1))/2$
  - Para 9 (nove) termos: $Md=((9+1))/2=10/2=5$
  - Portanto, a mediana é o 5º termo
  - Conforme se segue: 2, 4, 6, 8, 10, 12, 14, 16, 18 (9 termos)
  - $Md=10$

- Utilizada quando a distribuição apresenta resultados extremos muito discrepantes.
- A mediana não sofre a influência de valores extremos.

Moda

- Moda: Valor dominante de uma distribuição.
  - Aquele que numa série de valores se apresenta com a maior frequência. Um conjunto de valores pode apresentar mais de uma moda: plurimodal.

Exemplo I: {18, 18, 18, 20, 20, 21, 23, 26, 29, 31, 33}: Moda= 18

Exemplo II: {18, 18, 18, 24, 24, 24, 26, 26, 28, 30, 35}: Bimodal= 18 e 24

Exemplo III: {10,10, 20, 20, 30, 30, 35, 40, 45, 50, 50}: Plurimodal: 10, 20, 30, 50

Exemplo IV: {1, 3, 5, 7, 9, 11, 13, 15}: Amodal: não há Moda

MEDIDAS DE DISPERSÃO

Medidas de Variabilidade

- Índices que indicam o grau de concentração ou dispersão de uma distribuição em torno da média.

- Principais índices de variabilidade:
  - Amplitude total
  - Desvio médio
  - Variância
  - Desvio padrão

Amplitude

- Amplitude Total (Intervalo Total):
  - É a diferença entre o maior e o menor valor de uma série.

- $A = Max - Min$
- $A = 36 - 19$
- $A = 17$

Alunos	Idades dos alunos de um determinado curso de tecnologia
1	27
2	28
3	20
4	28
5	31
6	23
7	23
8	36
9	29
10	23
11	23
12	19
13	20
14	21
15	21
16	27
17	33

Mediana: Medida de posição central.

Série de valores com quantidade par de termos

Mediana = $Md = \frac{(n)}{2}$ e Mediana = $Md = \frac{(n)}{2} + 1$

Para 10 (dez) termos:

$Md = \frac{(10)}{2} = 5$ e Mediana = $Md = \frac{(10)}{2} + 1 = 6$

Portanto, a mediana está entre o 5º e o 6º termo

Conforme se segue: 2, 4, 6, 8, 10, 12, 14, 16, 18, 20 (10 termos)

Mediana = $Md = \frac{(10+12)}{2} = 11$

Desvio médio

- Desvio Médio é Média aritmética dos afastamentos (ou desvios), tomados em valor absoluto, entre cada valor e a média aritmética.

$$DM = \frac{\sum f_i . |X_i - \bar{X}|}{n}$$

$$DM = \frac{\sum |X_i - \bar{X}|}{n}$$

Exemplo

Considere as notas 2, 8, 5, 6 obtidas por 4 alunos, numa avaliação de biologia. Determine o desvio médio.

Inicialmente, calcularemos a média:

$$\bar{X} = \frac{\sum_{i=1}^{k} X_m \cdot f_i}{\sum_{i=1}^{k} f_i}$$

$$\bar{X} = \frac{2+8+5+6}{4}$$

$$\bar{X} = 5{,}25$$

- Agora, calculamos o desvio médio, lembrando que fi = 1, visto que cada um dos quatro valores apareceu uma única vez.

$$DM = \frac{\sum f_i . |X_i - \bar{X}|}{n}$$

$$DM = \frac{|2 - 5{,}25| + |8 - 5{,}25| + |5 - 5{,}25| + |6 - 5{,}25|}{4}$$

$$DM = \frac{|-3{,}25| + |2{,}75| + |-0{,}25| + |0{,}75|}{4}$$

$$DM = \frac{7}{4} = 1{,}75$$

$$DM = \frac{\sum f_i . |X_i - \bar{X}|}{n}$$

$$DM = \frac{2 * |48 - 54{,}4| + 3 * |(51 - 54{,}4| + 5 * |55 - 54{,}4| + 4 * |58 - 54{,}4| + 1 * |60 - 54{,}4|}{15}$$

$$DM = \frac{12{,}8\ +\ 10{,}2\ +\ 3{,}0\ +\ 14{,}4\ +\ 5{,}6}{15}$$

$$DM = 3{,}07$$

Item	Frequência
48	2
51	3
55	5
58	4
60	1

Utilização: Indica o quanto, em média, os valores se afastam do ponto central (média) numa distribuição do tipo ***Curva de Gaus***

Variância

- Considerando-se uma amostra de dados, cada dado isolado pode ter um desvio (dispersão) em relação à média da amostra.
- Essa dispersão é a diferença entre o valor individual e a média da amostra de dados.

- Para se avaliar o grau de dispersão de toda a amostra de dados utiliza-se a variância que é a soma dos quadrados dos desvios dividido pelo tamanho da amostra, menos 1.

$$s^2 = \frac{\sum_{i=1}^{n} f_i . (X_i - \bar{X})^2}{n - 1}$$

$$s^2 = \frac{f_1 * [(X_1 - \bar{X})^2] + f_2 * [(X_2 - \bar{X})^2] + f_n * [(X_n - \bar{X})^2]}{n - 1}$$

$$s^2 = \frac{81{,}92 + 34{,}68 + 1{,}80 + 51{,}84 + 31{,}36}{14}$$

$$s^2 = 14{,}4$$

Desvio padrão

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

Afastamento quadrático médio ou afastamento padrão. É a raiz quadrada da variância.

Desvio padrão dos dados isolados ponderados com frequências distintas:

- Afastamento quadrático médio ou afastamento padrão. É a raiz quadrada da variância.

$$s = \sqrt{s^2}$$

Exemplo

$$s^2 = \frac{(81{,}92 + 34{,}68 + 1{,}80 + 51{,}84 + 31{,}36)}{14}$$

Exemplo:

$$s^2 = \frac{81,92 + 34,68 + 1,80 + 51,84 + 31,36}{14}$$

$$s = \sqrt{14,4}$$

$$S = 3,79$$

Utilização: é a medida mais usada com medida de variabilidade, principalmente quando a distribuição for normal

∑: símbolo de somatório. Indica que temos que somar todos os termos, desde a primeira posição (i=1) até a posição n

xi: valor na posição i no conjunto de dados

MA: média aritmética dos dados

n: quantidade de dados

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

Em uma equipe de remo os atletas possuem as seguintes alturas:

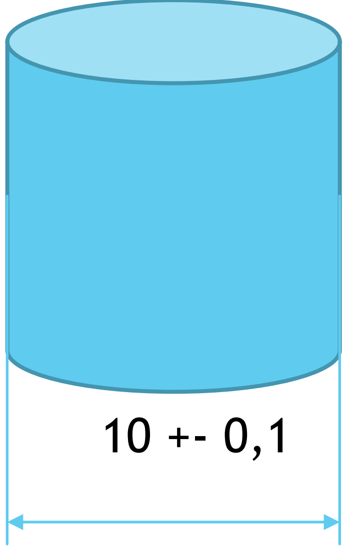
1,55 m;	1,75 m; e	1,80 m.

Qual é o valor da média e do desvio padrão da altura desta equipe?

Cálculo da média, sendo n = 3

$$\bar{x} = \frac{1{,}55 + 1{,}75 + 1{,}80}{3} = 1{,}68\, m$$

- Cálculo do desvio padrão



10 +- 0,1
9,9 - 10,1

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

$$s = \sqrt{\frac{(1{,}55 - 1{,}68)^2 + (1{,}70 - 1{,}68)^2 + (1{,}80 - 1{,}68)^2}{3}}$$

$$s = 0{,}1027$$

Alunos	Idades dos alunos
1	23
2	21
3	21
4	20
5	19
6	20
7	19
8	20
9	20
10	20
11	22
12	35
13	28
14	19
15	18
16	20
17	18

MEDIDAS DE ORDENAMENTO OU POSIÇÃO

Idades	Frequência
18	2
19	3
20	6
21	2
22	1
23	1
28	1
35	1

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

$$s = \sqrt{\frac{2*(18-20)^2 + 3*(19-20)^2 + 6*(20-20)^2 + 2*(21-20)^2 + 1*(22-20)^2 + 1*(23-20)^2 + 1*(28-20)^2 + 1*(35-20)^2}{17}}$$

$$s = \sqrt{\frac{8+3+0+2+1+4+9+64+225}{17}}$$

$$s = 4{,}33$$