

## **Curso de Estatística I**

### **Ementa:**

1. Estatística Descritiva
2. Cálculo das Probabilidades
3. Variáveis Aleatórias
4. Distribuição de Poisson e Distribuição Binomial

### **Bibliografia**

**FARIAS, A.A. E OUTROS** Introdução à Estatística, 2ª ed., Rio de Janeiro; LTC, 2008.

**Triola, M.F.** Introdução à Estatística, 10ª ed., Rio de Janeiro, LTC, 2008

**MEYER, P.L.** Probabilidade, Aplicações à Estatística, Rio de Janeiro LTC, 1993.

**MORETTIN, L.G.:** Estatística Básica, Rio de Janeiro, Mackron Books.

**SPIEGEL, M.:** Probabilidade e Estatística, Rio de Janeiro, Mackron Books.

**LIPSCHUTZ, S.:** Probabilidade, Rio de Janeiro, Mackron Books.

### **INTRODUÇÃO**

- **Objetivo:** O principal objetivo é apresentar as idéias de variabilidade, incerteza e alguns métodos estatísticos simples, que permitem que se façam inferências a partir de dados observados.
- **O que é Estatística:** A estatística trata de idéias e métodos que visam a aperfeiçoar a obtenção de conclusões a partir de informações numéricas, na presença da incerteza, ou melhor, o de como utilizar informações incompletas para tomar decisões válidas e tirar conclusões satisfatórias.
- **Inferência Estatística:** Se propõe a tirar conclusões úteis a respeito de um conjunto, denominado população, com base num subconjunto de dados disponíveis, que é usualmente, chamado de amostra.

**Estatística Descritiva:** é um conjunto de técnicas que objetivam descrever, analisar e interpretar os dados numéricos de uma população ou amostra.

### **População e Amostra**

Entende-se por população o conjunto de elementos que têm, em comum, determinada característica. Todo subconjunto não vazio e com menor número de elementos do que a população constitui uma amostra dessa população.

### **Técnicas de Amostragem**

A amostra casual simples é composta por elementos retirados ao acaso da população. Então todo elemento da população tem igual probabilidade de ser escolhido para a amostra.

- **Amostra Sistemática**

Na amostra sistemática os elementos são escolhidos não por acaso, mas por um sistema. Exemplo: Um professor pode escolher todo aluno com número terminado em determinado dígito.

- **Amostra Estratificada**

A amostra estratificada é composta por elementos provenientes de todos os estratos da população. Devem ser obtidas amostras estratificadas sempre que a população for constituída por diferentes estratos. Exemplo: Se as pessoas que moram no vários bairros de uma cidade são diferentes, cada bairro é um estrato. Para obter uma amostra de pessoas dessa cidade, seria razoável obter uma amostra de cada bairro e depois reunir as informações numa amostra estratificada.

## Variáveis

Os dados coletados em uma pesquisa estatística referem-se a variáveis que são classificadas como qualitativas, ordinais e quantitativas.

- **Variável qualitativa:** Uma variável é qualitativa quando os dados podem ser distribuídos em categorias mutuamente exclusivas. Exemplos: sexo (feminino e masculino), cor (verde, amarelo, etc), causa de morte , grupo sanguíneo etc.
- **Variável ordinal:** Uma variável é ordinal quando os dados podem ser distribuídos em categorias mutuamente exclusivas que têm ordenação natural. Exemplos: grau de instrução (primário, secundário e superior), estágio social, estágio da doença etc.
- **Variável quantitativa:** Uma variável é quantitativa quando é expressa por números. Exemplos: idade, estatura, peso corporal etc.

## Apresentação de Dados

Depois dos dados coletados e apurados eles podem ser apresentados sob a forma de tabelas, quadros ou gráficos.

### Apresentação de Dados em Tabelas

São chamadas de séries estatísticas que também são classificadas em TEMPORAIS OU HISTÓRICAS, GEOGRÁFICAS , ESPECÍFICAS e as DISTRIBUIÇÕES DE FREQUÊNCIAS que veremos mais adiante. Os dados devem ser apresentados em tabelas construídas de acordo cm as normas técnicas ditadas pela Fundação IBGE.

### Componentes das Tabelas

As tabelas têm título, corpo, cabeçalho e coluna indicadora.

- Título: explica o que a tabela contém;
- Corpo: é formado pelas linhas e colunas de dados;
- Cabeçalho: especifica o conteúdo das colunas;
- Coluna indicadora: especifica o conteúdo das linhas.

- Rodapé: especifica a fonte.

Casos registrados de intoxicação humana, segundo a causa determinante. Brasil, 1993

Causa	Freqüência
Acidente	29.601
Abuso	2.604
Suicídio	7.965
Profissional	3.735
Outras	1.959
Ignoradas	1.103

Fonte: MS/FIOCRUZ/SINITOX

Na tabela acima, série específica, temos:

Título: Casos registrados de intoxicação humana, segundo a causa determinante. Brasil 1993.

Cabeçalho: É constituído pelas palavras:

Causa	Frequência
-------	------------

Coluna indicadora: É constituída pelas especificações:

Acidente
Abuso
Suicídio
Profissional
Outras
Ignorada

Corpo: É formado pelos números:

29.601
2.604
7.965
3.735
1.959
1.103

Toda tabela deve ser delimitada por traços horizontais. Podem ser feitos traços verticais para separar as colunas, mas não devem ser feitos traços verticais para delimitar a tabela. O cabeçalho é separado do corpo por um traço horizontal. As tabelas podem apresentar, além das frequências, as frequências relativas e o total.

Casos registrados de intoxicação humana, segundo a causa determinante. Brasil, 1993

Causa	Frequência	Frequência relativa
Acidente	29.601	64,23
Abuso	2.604	5,64
Suicídio	7.965	17,32
Profissional	3.735	8,12
Outras	1.959	4,26
Ignoradas	1.103	2,40
Total	45.974	

Fonte: MS/FIOCRUZ/SINITOX

### **Tabelas de contingência**

Muitas vezes os elementos da amostra ou da população são classificados de acordo com dois fatores. Os dados devem então ser apresentados em tabelas de contingência, isto é, em tabelas de dupla entrada, cada entrada relativa a um dos fatores. No exemplo abaixo temos dois fatores: o ano de registro e o sexo.

Nascidos vivos registrados segundo o ano de registro e o sexo

Ano de registro	Sexo		Total
	Masculino	Feminino	
1984	1 307 758	1 251 280	2 559 038
1985	1 339 059	1 280 545	2 619 604
1986	1 418 050	1 361 203	2 779 253

Fonte: IBGE (1988)

Nota: Nascimentos ocorridos no ano de registro

### **Tabelas de Distribuição de Frequências**

As tabelas com grande número de dados são cansativas e não dão ao leitor visão rápida e global do fenômeno. Para isso, é preciso que os dados estejam organizados em uma tabela de distribuição de frequências. Para construir uma tabela de distribuição de frequências precisamos das seguintes definições:

#### **Dados brutos**

É o conjunto dos dados numéricos obtidos após a crítica dos valores coletados. Assim 24 - 10 - 10 - 1 - 12 - 8 - 10 - 9 -3 - 7 -2 - 7 - 4 -5 -7 - 11 - 6 - 7 são exemplos de dados brutos.

#### **Rol**

É o arranjo dos dados brutos em ordem crescente ou decrescente. Assim 1 - 2 - 3 - 4 - 5 - 6 - 7 - 7 - 7 - 7 - 8 - 9 - 10 - 10 - 10 - 11 - 12 - 24.

#### **Amplitude Total**

É a diferença entre o maior e o menor valor observados. No exemplo acima temos  $A_t = 24 - 1 = 23$ .

#### **Frequência Absoluta ( $F_i$ )**

É o número de vezes que o elemento aparece na amostra, ou número de elementos pertencentes a uma classe. No exemplo temos  $F_{10} = 3$ .

#### **Número de Classes (C)**

Não há uma fórmula exata para o cálculo do número de classes. Se a amostra

for de tamanho menor ou igual a 25 teremos 5 classes, e se a amostra for de tamanho N maior do que 25, o número de classes será a raiz quadrada de N.

### Amplitude das Classes

$$A_C = A_t : C.$$

### Limite das Classes

- a)  $10 \vdash 12$  : compreende todos os valores entre 10 e 12 (inclusive 10 e 12);
- b)  $10 \vdash 12$  : compreende todos os valores entre 10 e 12 (excluindo o 12);
- c)  $10 \dashv 12$  : compreende todos os valores entre 10 e 12 (excluindo o 10).

### Pontos Médios das Classes

É a média aritmética entre o limite superior e o limite inferior da classe. Assim, se a classe for  $10 \vdash 12$ , teremos  $x_i = (10 + 12)/2 = 11$ , como ponto médio da classe.

### Frequência Relativa ( $f_i$ )

É dada pela fórmula  $f_i = F_i/N$ , ou seja, é a porcentagem daquele valor na amostra.

classe	$F_i$	$f_i$
45 $\vdash$ 55	15	0.15
55 $\vdash$ 65	30	0.30
65 $\vdash$ 75	35	0.35
75 $\vdash$ 85	15	0.15
85 $\vdash$ 95	5	0.50
$\Sigma$	100	1.00

Note que  $\Sigma f_i = 1$

### Frequência absoluta acumulada

É a soma das frequências dos valores inferiores ou iguais ao valor dado.

classe	$F_i$	$F_{ac}$
45 † 55	15	15
55 † 65	30	45
65 † 75	35	80
75 † 85	15	95
85 † 95	5	100
$\Sigma$	100	

Considerando as estaturas, em cm, de 40 estudantes de uma faculdade constatou-se que:

186 187 192 180 150 180 173 160 168 169  
165 173 174 160 165 155 150 187 174 173  
150 180 173 165 155 150 187 174 173 180  
151 163 155 172 180 176 172 160 175 154

Pede-se:

1. O rol;
2. A distribuição para dados não agrupados;
3. A distribuição para dados agrupados.