

# Conceitos Básicos em Estatística

Gilberto Pereira Sassi

Instituto de Matemática e Estatística  
Universidade Federal Fluminense

# Informações básicas sobre a disciplina

- Listas de exercícios e material complementar em <http://gpsassi.github.io>
- Programa do curso: [http://www.est.uff.br/images/meusarquivos/disciplinas/Programa-GET00053\\_apartir2005.pdf](http://www.est.uff.br/images/meusarquivos/disciplinas/Programa-GET00053_apartir2005.pdf)
- Nota final:  $\frac{P_1 + P_2}{2}$  + média de exercícios em aula, em que  $P_1$  é a nota da primeira prova e  $P_2$  a nota da segunda prova
- Data da primeira prova: 21 de Junho
- Data da segunda prova: 23 de Julho
- E-mail: [gilberto@est.uff.br](mailto:gilberto@est.uff.br)
- Horário para dúvidas: toda quinta das 9 às 10 no departamento de Estatística
- Localização: Valonguinho, IME – sétimo andar sala 13
- Bibliografia do curso: Barbetta, P.A. – Estatística Básica Aplicada às Ciências Sociais e Bussab, W.O., Morettin, P.A. – Estatística Básica

# Tipo de inferência

Na ciência, há dois tipos principais de inferir:

**Inferência Dedutiva** Se argumenta a partir das premissas para desenhar conclusões.

**Exemplo:** Premissa: todas as plantas fazem fotossíntese. A mangueira é uma planta, então a mangueira faz fotossíntese.

**Inferência Indutiva** Generaliza-se da parte para o todo.

**Exemplo:** Em pesquisas de intenção de votos, entrevistamos uma parte dos eleitores para inferir a intenção de votos de todos os eleitores.

*A inferência indutiva é o assunto desse curso.*

O conjunto de todos os elementos do estudo é chamado de população.

Uma parte da população é chamada de amostra.

# Análise Exploratória de Dados

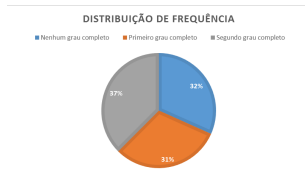
Antes de inferir sobre uma característica da população (por exemplo, percentual de eleitores de Lula) ou sobre uma hipótese (por exemplo, a renda dos eleitores do PSDB é dez vezes a renda dos eleitores do PT), construímos e calculamos algumas medidas para obter informações que não conseguimos encontrar nos dados brutos.

## Exemplo

Suponha que temos interesse em estudar o grau de escolaridade dos chefes de família do bairro Saco Grande II em Florianópolis – Santa Catarina.

	Grau de instrução do chefe de família	Localidade
1	Nenhum Grau Completo	Conjunto Residencial Monte Verde
2	Nenhum Grau Completo	Conjunto Residencial Monte Verde
3	Nenhum Grau Completo	Conjunto Residencial Monte Verde
4	Nenhum Grau Completo	Conjunto Residencial Monte Verde
5	Nenhum Grau Completo	Conjunto Residencial Monte Verde
6	Nenhum Grau Completo	Conjunto Residencial Monte Verde
7	Nenhum Grau Completo	Conjunto Residencial Monte Verde
8	Primeiro Grau Completo	Conjunto Residencial Monte Verde
9	Primeiro Grau Completo	Conjunto Residencial Monte Verde
10	Primeiro Grau Completo	Conjunto Residencial Monte Verde
11	Primeiro Grau Completo	Conjunto Residencial Monte Verde
12	Primeiro Grau Completo	Conjunto Residencial Monte Verde
13	Primeiro Grau Completo	Conjunto Residencial Monte Verde
14	Primeiro Grau Completo	Conjunto Residencial Monte Verde
15	Primeiro Grau Completo	Conjunto Residencial Monte Verde
16	Primeiro Grau Completo	Conjunto Residencial Monte Verde
17	Primeiro Grau Completo	Conjunto Residencial Monte Verde
18	Primeiro Grau Completo	Conjunto Residencial Monte Verde

(a) Dados Brutos.



(b) Gráfico de “pizza”

**Figura:** Os dados brutos versus gráfico de composição de setores para a variável grau de instrução.

# Tipos de variáveis

## Exemplo

Um investidor tem interesse em traçar o perfil dos funcionários da empresa XXX.

Para cada funcionário da empresa XXX, desejamos saber algumas características:

- Estado civil: solteiro ou casado
- Grau de Instrução: Ensino Fundamental, Ensino Médio ou superior
- Salário
- Número de filhos

# Tipos de Variáveis

Uma característica de um elemento da população é chamada de variável e podemos classificá-la em quatro categorias:

**Variável Qualitativa Nominal** Variável cujos os resultados são atributos sem hierarquia ou ordem

**Exemplo:** Sexo: Masculino, Feminino ou Outro

**Variável Qualitativa Ordinal** Variável cujos os resultados são atributos com hierarquia

**Exemplo:** Grau de Instrução: Ensino Fundamental, Ensino Médio ou Superior

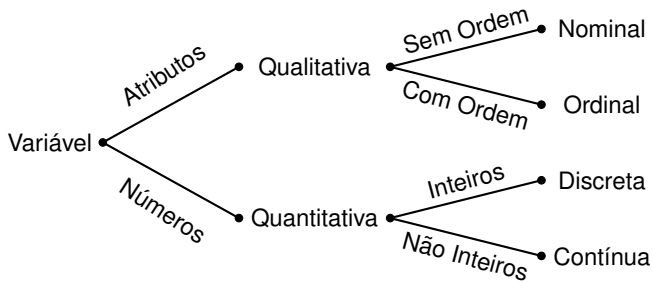
**Variável Quantitativa Discreta** Variável cujos resultados são números inteiros

**Exemplo:** Número de Filhos

**Variável Quantitativa Contínua** Variável cujos resultados são número possivelmente não inteiros

**Exemplo:** Salário

# Tipo de Variáveis



# Distribuição de Frequência para Variável Qualitativa

A primeira forma de resumir os dados é através de uma tabela chamada de Distribuição de Frequência.

## Exemplo

Na pesquisa sobre o Grau de Instrução dos chefes de família, temos 120 chefes de família com a seguinte distribuição de frequência.

**Tabela:** Distribuição de Frequência para a variável Grau de Instrução.

Distribuição de frequência			
Grau de Instrução	Frequência	Proporção	Porcentagem
Nenhum grau completo	38	0,3167	31,67
Primeiro grau completo	37	0,3083	30,83
Segundo grau completo	45	0,3750	37,50
Total	120	1,0000	100,00



## Distribuição de Frequência para Variável Quantitativa

Para construir distribuição de frequência para variáveis quantitativas, precisamos agrupar os dados em categorias.

### Exemplo

Distribuição de frequência dos salários dos funcionários da empresa XXX.

**Tabela:** Distribuição de Frequência para a variável quantitativa contínua Salário em salários mínimos.

Distribuição de frequência – Salário			
Classe de Salário	Frequência	Proporção	Porcentagem
4  —— 8	10	0,2778	27,78
8  —— 12	12	0,3333	33,33
12  —— 16	8	0,2222	22,22
16  —— 20	5	0,1389	13,89
20  —— 24	1	0,0278	2,78
Total	36	1,0000	100,00

# Distribuição de Frequência para Variável Quantitativa

Ressaltamos que a distribuição de frequência para variável quantitativa podemos perder informação. Por exemplo, sabemos que temos doze funcionários com salário entre 8 e 12 salários mínimos, mas não sabemos exatamente quanto é o salário de cada um deles. Além disso, a escolha das classes é arbitrária. Se escolhermos intervalos pequenos, resumiremos pouco os dados e se tomarmos intervalos grandes, perdemos muita informação. A decisão do tamanho do intervalo depende do objetivo do estudo e é especificada pelo responsável do estudo.