

Módulo 3: A pesquisa e o software Excel

O software Excel; Introdução aos métodos estatísticos; Descrição e exploração dos dados; Variáveis qualitativas (Tabela de frequências); Variáveis quantitativas (Média aritmética, Mediana, Moda, Variância, Desvio padrão, Teste T).

O software Excel

- O *Microsoft Office Excel* é um *software* da *Microsoft (Windows)* “que fornece ferramentas para organizar, analisar e interpretar dados” (BERNAL; SILVA, 2012, p. 4).
- Ele possui 4 funções principais (BERNAL; SILVA, 2012):

Planilha eletrônica	Linhas numeradas e colunas identificáveis por letras, onde é possível identificar as células e as informações contidas nelas.
Banco de dados	Possui ferramentas de bancos de dados: consultar, buscar, ordenar, filtrar e administrar grande quantidade de dados.
Gráficos	Permite a representação gráfica de dados contidos em uma planilha eletrônica.
Ferramentas estatísticas	Dispõe de ferramentas para análises estatísticas mais simples.

Introdução aos métodos estatísticos

Variáveis qualitativas ou categóricas

- São medidas em categorias. Ex: Sexo, Profissão, Município etc.

Variáveis quantitativas

- Podem ser medidas por escalas quantitativas (unidades de medida).
Ex: Renda Familiar, Faturamento de uma Empresa, Peso, Altura etc.

Descrição e exploração dos dados

- Os dados deverão ser organizados, resumidos e apresentados de forma a permitir sua interpretação, conforme os objetivos da pesquisa e o tipo de variável.
- Um primeiro passo é avaliar cada variável separadamente ([análise univariada](#)).

Para variáveis qualitativas, utilizar-se-ão tabelas de frequência ou gráficos (gráfico de barras e gráfico pizza).

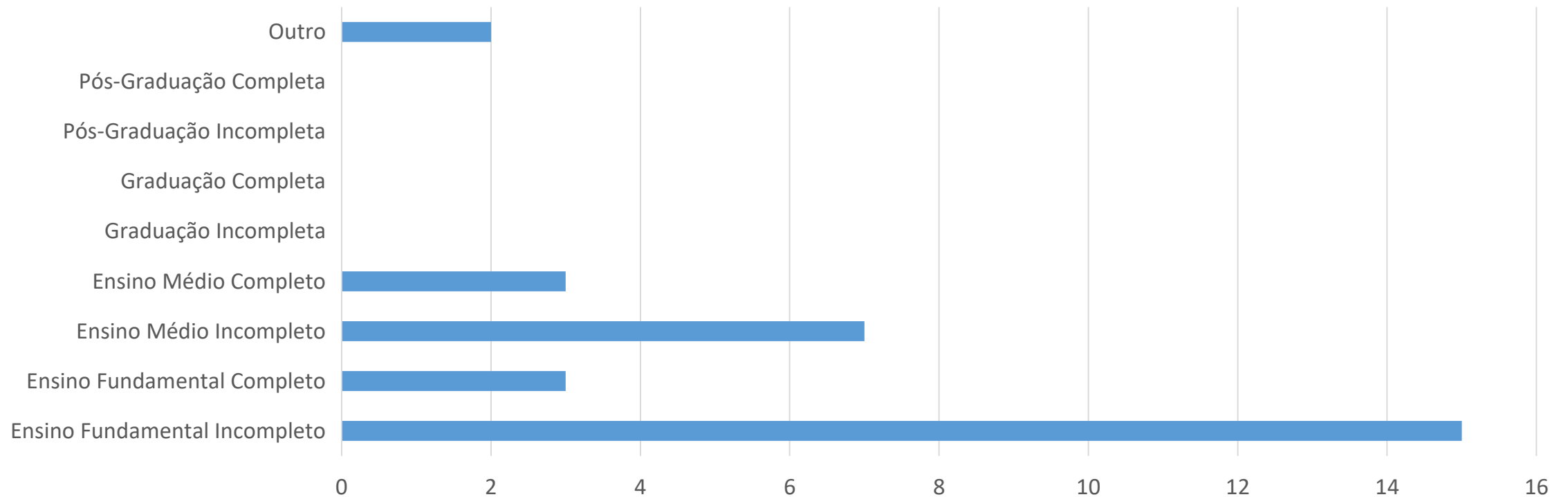
Para variáveis quantitativas, utilizar-se-ão estatísticas descritivas (média, mediana, moda, desvio padrão...) ou gráficos (histograma e *box plot*).

Variáveis qualitativas: Tabela de frequências

- **Distribuição de frequências** é o método estatístico para agrupamento de dados em classes; um arranjo de valores.
- **Como fazer no Excel:** Criar uma tabela com as variáveis e dados a serem analisados. Os dados corresponderão à frequência (repetição) de determinada condição. Nas planilhas em que haja transcrição de questionários com perfil de indivíduos, por exemplo, é possível organizar tabelas de “resumo”, com distribuição de frequência por categoria, utilizando-se a função estatística **=CONT.SE(intervalo,”critérios”)**.

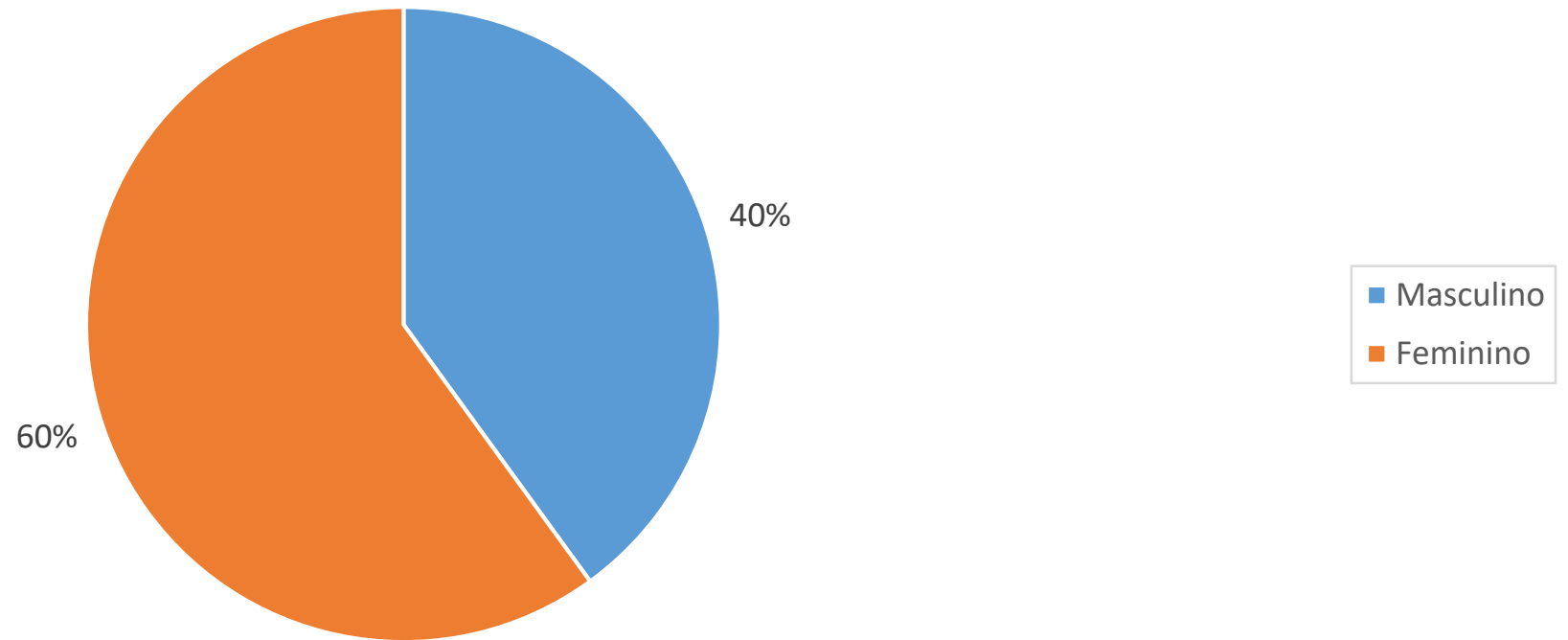
Variáveis qualitativas: Gráfico em Barras

Gráfico em Barras referente ao grau de escolaridade dos respondentes do questionário A



Variáveis qualitativas: Gráfico Pizza

Gráfico Pizza referente ao gênero dos respondentes do questionário A



Variáveis quantitativas: Média aritmética

- A **média aritmética**, ou simplesmente média, pode ser interpretada como um modelo para resumir dados (FIELD, 2009, p. 34).

$$\bar{x} = \sum \frac{(x_1 + x_n)}{n}$$

A média aritmética (\bar{x}) é a soma de todas as observações de um determinado grupo de dados dividida pelo número total de observações ($\sum \frac{(x_1 + x_n)}{n}$).

- **Como fazer no Excel:** Em uma planilha de dados, é possível calcular a média de uma categoria a partir da função estatística **=MÉDIA(intervalo)**.

Variáveis quantitativas: Mediana

- A mediana indica o valor central que separa um conjunto de dados em duas metades.
 - Quando o conjunto for composto por um número ímpar de elementos, a mediana será o valor central/intermediário.
 - Quando o conjunto for composto por um número par de elementos, a mediana será a média aritmética dos dois valores centrais.
- Como fazer no Excel: A mediana pode ser definida pela função =MED(intervalo).

Variáveis quantitativas: Moda

- Moda indicará o valor mais frequente (que mais se repete) em um conjunto de dados.
- Como fazer no Excel: Para identificar a moda de um conjunto de dados, utilizar-se-á a função estatística =**MODO(intervalo)** ou =**MODO.ÚNICO(intervalo)**.

Variáveis quantitativas: Variância

- A variância é a medida que estimará o erro na população, com base na amostra observada.

$$V = \frac{\sum (x_i - \bar{x})^2}{n - 1}$$

A variância (V) é a soma dos quadrados da diferença de cada observação pela média ($\sum (x_i - \bar{x})^2$) dividida pelo número de observações menos uma ou, como denominamos, graus de liberdade ($n - 1$). A variância é expressa em unidades quadradas.

- Como fazer no Excel: Utiliza-se a função `=VAR(intervalo)`.

Variáveis quantitativas: Desvio Padrão

- O desvio padrão expressa o erro médio na mesma unidade da variável. Portanto, “é uma medida que representa quão bem a média representa os dados” (FIELD, 2009, p. 35).

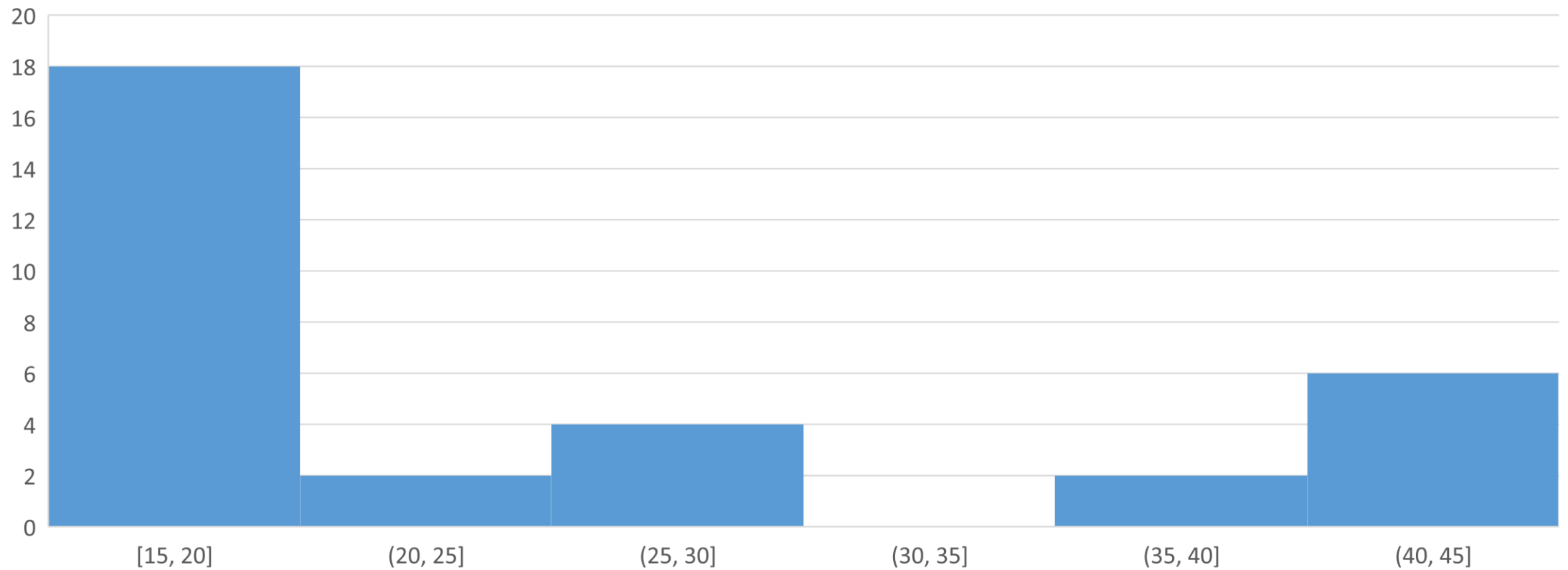
$$s = \sqrt{V} \text{ ou } s = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n-1}}$$

O desvio padrão é a raiz quadrada da variância (V), ou seja, permite a interpretação direta da variação do conjunto de dados, por ser expresso na mesma unidade que a variável. Podemos definir que o desvio padrão (s) é a média dos valores absolutos dos desvios.

- **Como fazer no Excel:** Pode-se calcular a raiz quadrada da variância ou utilizar a função estatística `=DESVPAD(intervalo)` ou `=DESVPAD.A(intervalo)`, sendo o segundo para cálculo de amostras.

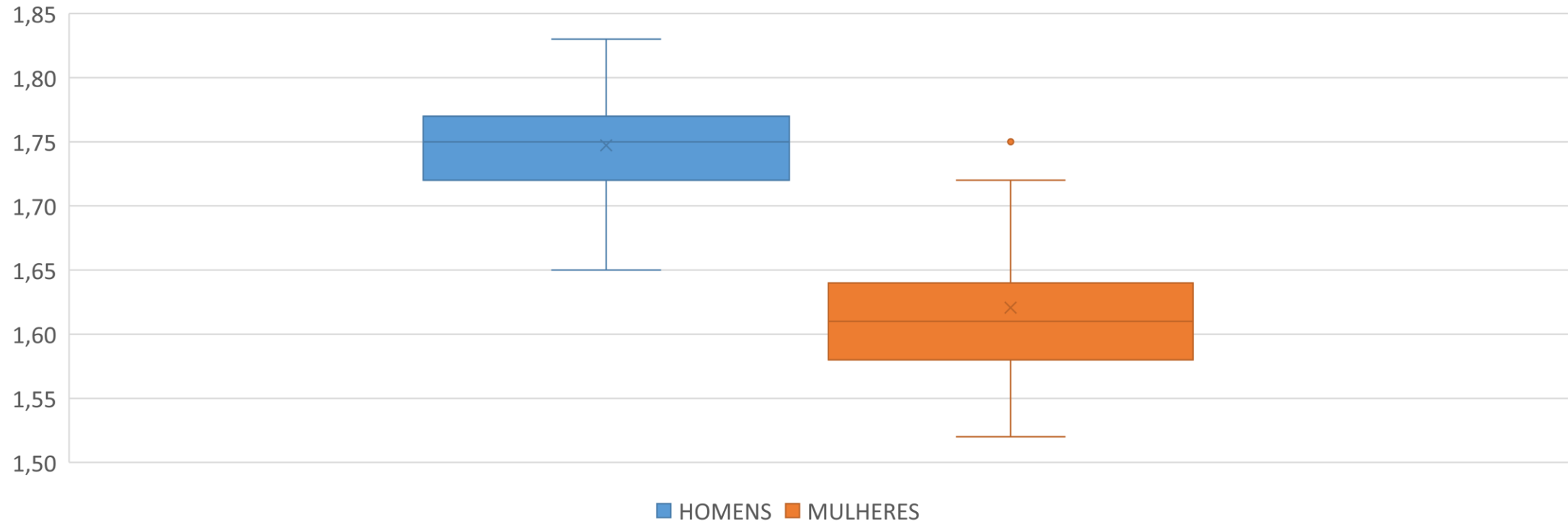
Variáveis quantitativas: Gráfico tipo Histograma

Gráfico do tipo Histograma referente à faixa etária dos respondentes do questionário B



Variáveis quantitativas: Gráfico tipo *Box Plot*

Gráfico do tipo *Box Plot* referente à altura dos respondentes do questionário C, conforme gênero



Variáveis quantitativas: Teste T

- O Teste T (ou Teste *T-student*) “é apropriado para comparar as médias de uma variável quantitativa entre dois grupos independentes” (MUNDSTOCK; FACHEL; CAMEY; AGRANONIK, 2010, p. 37), mas também para a comparação de um mesmo grupo sobre uma variável quantitativa em dois períodos de tempo diferentes.
- Ou seja, o Teste T permite os seguintes tipos de comparação:
 - De uma mesma amostra em dois períodos de tempo diferentes; ou
 - De duas amostras independentes.

Variáveis quantitativas: Teste T

- Para se interpretar o Teste T, é necessário compreender o nível de significância (**p-valor**) da variável, que, em Ciências Sociais Aplicadas, é de **0,05** (equivalente a 5%). Assim, temos a seguinte relação:

$$t < 0,05$$

- há indícios de diferença significativa entre as médias;

$$t > 0,05$$

- não há indícios de diferença significativa entre as médias (similaridade).

Variáveis quantitativas: Teste T

- **Importante!** Para se compreender o Teste T e o p-valor, é essencial que sejam definidas hipóteses (hipótese nula e hipótese alternativa) sobre os grupos que se está analisando. Assim, o p-valor é a probabilidade de rejeição da hipótese nula, como por exemplo: Um professor aplicou, em uma turma de alunos, uma prova de estatística no primeiro bimestre e outra no segundo bimestre. O professor quer verificar se houve progresso nas notas médias atingidas pela classe. Logo, haverá duas possibilidades:
 - H_0 : Houve melhoria significativa nas notas médias dos alunos.
 - H_1 : Não houve melhoria nas notas médias dos alunos.

Variáveis quantitativas: Teste T

- Como fazer no Excel: Pode-se calcular o Teste T a partir da função estatística `=TESTE.T(matriz1;matriz2;caudas;tipo)`, onde:
 - matriz1: dados da amostra 1;
 - matriz2: dados da amostra 2;
 - caudas: o tipo de distribuição, sendo 1 para unicaudal e 2 para bicaudal;
 - tipo: variação das amostras, sendo 1 para variação par, 2 para variação igual e 3 para variação desigual.

Referências

- AGRANONIK, M.; HIRAKATA, V. N.; CAMEY, S. A. **Introdução à análise estatística utilizando o SPSS 18.0.** Porto Alegre, 2010. Disponível em: <http://www.mat.ufrgs.br/~camey/HCPA/cursos/Poligrafo%20-%20SPSS_Introduct%F3rio.pdf> Acesso em: 16 ago. 2017
- FIELD, Andy. **Descobrimo a estatística usando o SPSS.** 2. ed. Porto Alegre: Artmed Editora, 2009.