

# PCC170 - Projeto e Análise de Experimentos Computacionais

Marco Antonio M. Carvalho

Departamento de Computação  
Instituto de Ciências Exatas e Biológicas  
Universidade Federal de Ouro Preto



## 1 Revisão de conceitos estatísticos - Parte 2

## Fonte

Este material é parcialmente baseado no conteúdo de:

- ▶ Felipe Campelo (2018), Lecture Notes on Design and Analysis of Experiments. Online: <http://git.io/v3Kh8> Version 2.12; Creative Commons BY-NC-SA 4.0.
- ▶ Marcelo Menezes Reis. Conceitos elementares de estatística. Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, 2003. Disponível em: <https://bitly.com/vrgmc>.
- ▶ Fernanda Peres. Estatística aplicada à vida real. 2022. Disponível em: <https://bitly.com/fMep1>.
- ▶ Silvia E Shimakura, 2006. CE003 - Estatística II. Departamento de Estatística, Universidade Federal do Paraná Disponível em: <https://bitly.com/xzevC>.

## Licença

Este material está licenciado sob a Creative Commons BY-NC-SA 4.0. Isto significa que o material pode ser compartilhado e adaptado, desde que seja atribuído o devido crédito, que o material não seja utilizado de forma comercial e que o material resultante seja distribuído de acordo com a mesma licença.

## Análise Estatística

Métodos estatísticos não provam nada, porém, possibilitam a definição objetiva de margens de plausibilidade para determinadas afirmações:

- ▶ Obtenção de conclusões objetivas a partir dos dados;
- ▶ **Testes de hipóteses e estimação dos intervalos de confiança;**
- ▶ Apresentação de resultados em termos de um modelo empírico;
- ▶ Análise de resíduo e validação do modelo.

## Hipóteses estatísticas

Em geral, intervalos de confiança são a forma mais informativa de apresentar os achados principais de um estudo.

Contudo, algumas vezes existe um particular interesse em decidir sobre a verdade ou não de uma hipótese específica, e.g., se dois grupos têm a mesma média ou não, ou se o parâmetro populacional tem um valor em particular ou não.

Os testes de hipóteses nos fornecem a estrutura para isto.

Intervalos de confiança e testes de hipóteses estão intimamente relacionados.

## Hipóteses estatísticas

As hipóteses estatísticas são definidas como declarações objetivas sobre parâmetros de uma ou mais populações.

É importante notar que as afirmações em hipóteses estatísticas são sobre parâmetros da população, não da amostra.

A abordagem do teste de significância de hipótese nula envolve o contraste entre hipóteses nula e alternativa.

## Hipótese Nula ( $H_0$ )

- ▶ Ausência de efeitos.
- ▶ Modelo conservador.

Exemplo:  $H_0 : \mu = 25$

## Hipótese Alternativa ( $H_a$ )

- ▶ Presença de algum efeito.
- ▶ Existência de algo “novo”.

Exemplo:  $H_a : \mu \neq 25$



## Hipóteses estatísticas

Determinação do valor de referência para a hipótese nula  $H_0$ :

- ▶ Conhecimento prévio sobre o processo (investigação de mudanças);
- ▶ Valor obtido a partir de teoria ou modelos (validação do modelo);
- ▶ Requisitos do projeto (investigação da conformidade do sistema);

## Teste de hipóteses

O teste de hipóteses envolve:

- ▶ Obtenção de uma amostra;
- ▶ Realização de testes estatísticos;
- ▶ Decisão com base no resultado dos testes;

## Teste de hipóteses

Suponha que vendemos ervilhas verdes para grandes clientes e que desejamos determinar se os sacos de 50 kg realmente contêm seu peso nominal, pelo menos em média.

Nesse caso, a hipótese nula poderia ser definida como “o peso líquido médio de um saco é de 50kg”, e a hipótese alternativa poderia ser expressa como a desigualdade complementar.

▶  $H_0 : \mu = 50kg;$

▶  $H_a : \mu \neq 50kg;$

Suponha ainda que  $n = 10$  sacos são amostrados aleatoriamente, cujos conteúdos são ponderados usando uma balança confiável.

## Teste de hipóteses

Observe que já estamos adotando uma série de premissas subjacentes, estatísticas e técnicas:

- ▶ A média é uma boa medida para nossa questão de interesse (suposições subjacentes: a variação é pequena o suficiente, os clientes compram quantidades maiores para que, em média, o negócio seja justo, etc);
- ▶ Os sacos amostrados são representativos da nossa população de interesse (ou seja, dos sacos que realmente serão vendidos);
- ▶ O conteúdo dos sacos são apenas ervilhas verdes (sem quantidades significativas de contaminantes).

## Teste de hipóteses

Como a média amostral  $\bar{X}$  é uma boa estimativa da média real  $\mu$ , o senso comum sugere que:

- ▶ Se  $\bar{X} \approx 50\text{kg}$ : corroboração de  $H_0$ ;
- ▶ Se  $\bar{X} \ll 50\text{kg}$  ou  $\bar{X} \gg 50\text{kg}$ : refutação de  $H_0$ ;

Podemos usar  $\bar{X}$  como base para um teste.

Entretanto, é necessário definir uma **região crítica** para a rejeição de  $H_0$ , de modo a evitar erros dos tipos I e II.

## Erro tipo I

Falso positivo. Rejeitar a hipótese nula quando ela é verdadeira.

A probabilidade de ocorrência de um falso positivo em qualquer procedimento de teste de hipóteses é geralmente conhecida como **nível de significância**  $\alpha$ .

Outro termo frequente é o **nível de confiança do teste**, dado por  $(1 - \alpha)$ .

## Erro tipo II

Falso negativo. Corroborar (falha em rejeitar) a hipótese nula quando ela é falsa.

O **poder** de um teste é a probabilidade de rejeitar a hipótese nula quando esta é de fato falsa.

Isto é igual a  $1 - \beta$ , em que  $\beta$  denota a probabilidade de um erro tipo II.

Em geral, quanto maior o tamanho da amostra, maior o poder do teste.

## Teste de hipóteses

As características dos tipos de erros levam à seguinte classificação das conclusões obtidas do teste de hipóteses:

- ▶ Rejeição de  $H_0$ : conclusão forte;
- ▶ Falha em rejeitar  $H_0$ : conclusão fraca (mas podemos reforçá-la);

É importante lembrar que não rejeitar  $H_0$  não significa que haja evidências a favor de  $H_0$  – apenas sugere que este é um modelo melhor que a alternativa.



## Teste de hipóteses

O restante do exemplo sobre os sacos de ervilhas verdes, bem como detalhes sobre os cálculos realizados, podem ser encontradas no Capítulo 5 de

- ▶ Felipe Campelo (2018), Lecture Notes on Design and Analysis of Experiments. Online: <http://git.io/v3Kh8> Version 2.12; Creative Commons BY-NC-SA 4.0.

*Spoiler:* Um  $t$ -test, supondo a normalidade da distribuição, é o teste estatístico adequado para uma situação simples.

## Testes estatísticos

Em uma pesquisa correlacional o pesquisador não influencia nenhuma variável, mas apenas as mede e procura por relações (correlações) entre elas, como pressão sanguínea e nível de colesterol.

Duas ou mais variáveis quaisquer estão relacionadas se em uma amostra de observações os valores dessas variáveis são distribuídos de forma consistente.

Em outras palavras, as variáveis estão relacionadas se seus valores correspondem sistematicamente uns aos outros para aquela amostra de observações.

## Nível de significância estatística

Considere que já tenha sido calculada uma medida da relação entre duas variáveis.

A próxima questão é “quão significativa é esta relação?”

Para determinar o nível de significância estatística torna-se necessária uma função que represente o relacionamento entre “magnitude” e “significância” das relações entre duas variáveis, dependendo do tamanho da amostra.

Tal função diria exatamente “quão provável é obter uma relação de dada magnitude (ou maior) de uma amostra de dado tamanho, assumindo que não há tal relação entre aquelas variáveis na população”.

## Nível de significância estatística

Em outras palavras, aquela função forneceria o nível de significância, e isso permitiria conhecer a probabilidade de erro envolvida em rejeitar a idéia de que a relação em questão não existe na população.

Esta hipótese, de que não há relação na população, é denotada por uma **hipótese nula**.

## Nível de significância estatística

As funções de probabilidade são complexas e variam, porém, em muitos casos estas são relacionadas a um tipo geral de função que é chamada de **Normal**.

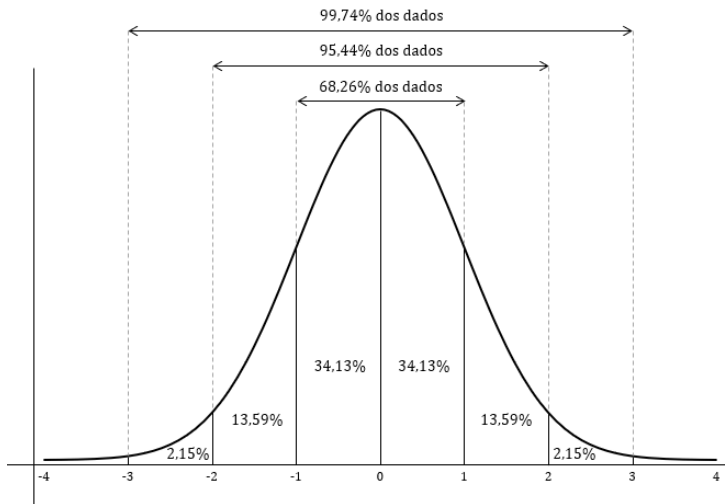
Por padrão, utiliza-se o nível de significância com valor de 95%.

## Distribuição Normal

Uma propriedade característica da distribuição Normal é que:

- ▶ 68% de todas as suas observações caem dentro de um intervalo de 1 desvio padrão da média;
- ▶ um intervalo de 2 desvios padrões inclui 95% dos valores;
- ▶ 99% das observações caem dentro de um intervalo de 3 desvios padrões da média.

# Revisão de conceitos estatísticos



fonte: <https://bityli.com/wFhvV>

## Distribuição Normal

Em uma distribuição Normal as observações que tem um *valor padronizado* de menos do que -2 ou mais do que +2 tem uma frequência relativa de 5% ou menos.

Valor padronizado significa que um valor é expresso em termos de sua diferença em relação à média, dividida pelo desvio padrão.



## Distribuição Normal

Qualquer que seja a distribuição da variável de interesse para grandes amostras, as médias amostrais serão aproximadamente normalmente distribuídas, e tenderão a uma distribuição Normal à medida que o tamanho de amostra crescer.

Podemos ter uma variável original com uma distribuição muito diferente da Normal, até mesmo ser discreta, porém, se tomarmos várias amostras grandes desta distribuição, e fizermos um histograma das médias amostrais, a forma se parecerá como uma curva Normal.

## Teorema Central do Limite

A aproximação para a Normal melhora à medida que o tamanho amostral cresce.

Este resultado é conhecido como o **Teorema Central do Limite** e é notável porque nos permite conduzir alguns procedimentos de inferência sem qualquer conhecimento da distribuição da população.

## *p-value*

A significância estatística de um resultado é uma medida estimada do grau em que este resultado é “verdadeiro”, no sentido de que seja realmente o que ocorre na população, ou seja no sentido de “representatividade da população”.

Mais tecnicamente, o *p-value* representa um índice decrescente da confiabilidade de um resultado.

## *p-value*

Quanto mais alto o *p-value*, menos se pode acreditar que a relação observada entre as variáveis na amostra é um indicador confiável da relação entre as respectivas variáveis na população.

Especificamente, o *p-value* representa a probabilidade de erro envolvida em aceitar o resultado observado como válido, isto é, como “representativo na população”.

# Revisão de conceitos estatísticos



Interpretação visual do  $p$ -value.  
 fonte: <https://bityli.com/ffHUp>

## *p-value*

Por exemplo, um *p-value* de 0,05 indica que há 5% de probabilidade de que a relação entre as variáveis, encontrada na amostra, seja um “feliz acaso”.

Por padrão, um *p-value* de 0,05 é costumeiramente tratado como um “limite aceitável” de erro.

Resultados com um *p-value* 0,01 são comumente considerados estatisticamente significantes, e com *p-value* 0,005 ou *p-value* 0,001 são frequentemente chamados “altamente” significantes.

Estas classificações, porém, são convenções arbitrárias e apenas informalmente baseadas em experiência geral de pesquisa.

## *p-value* e testes estatísticos

Nas próximas aulas serão apresentados testes de hipóteses para comparação de duas ou mais populações.

Cada teste possui suas particularidades e apresenta um conjunto de dados específico como resultado, entretanto, todos retornam o *p-value*.

Ao contrário de outros parâmetros apresentados, o *p-value* é sempre analisado diretamente.

## Ferramentas de estatística computacional

Existe uma ampla gama de ferramentas computacionais para estatística descritiva e inferência estatística.

- ▶ R;
- ▶ Minitab;
- ▶ Matlab;
- ▶ Python (*SciPy*, *matplotlib*, *seaborn*, *plot.ly*, *jupyter*, ou pacote Anaconda);
- ▶ Etc.



## Leitura recomendada

- ▶ Brady, T. *Reviewer's quick guide to common statistical errors in scientific papers*. Disponível em <https://goo.gl/IKvAYc>.
- ▶ Marcelo Menezes Reis. *Conceitos elementares de estatística*. Departamento de Informática e Estatística, Universidade Federal de Santa Catarina, 2003. Disponível em: <https://bityli.com/vrgmc>.

# Dúvidas?

