Preditiva.ai

Noções de Inferência Estatística

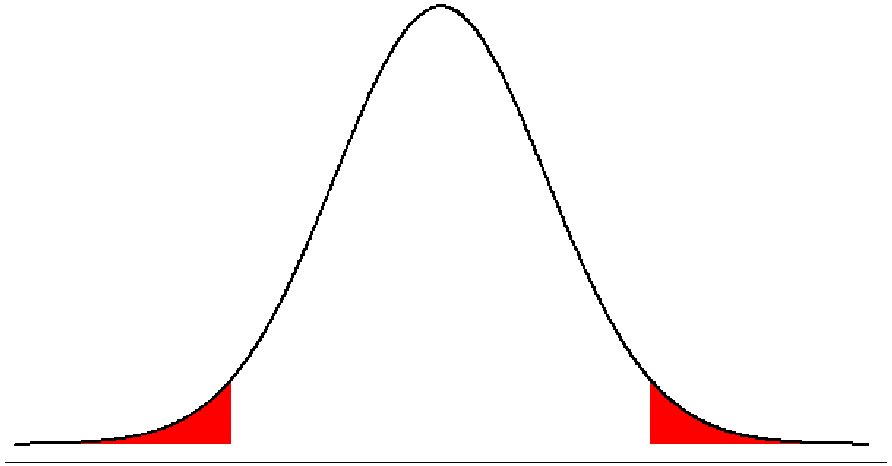
Introdução aos Testes de Hipóteses

Noções de Inferência Estatística

Introdução aos Testes de Hipóteses

Os **Testes de Hipóteses** são uma técnica de **Inferência** muito importante para realizar **comparações** dos dados amostrais com valores de referência ou dados de outras amostras.

Uma característica fundamental nos **Testes de Hipóteses** é a de realizar essas comparações considerando, além das medidas resumo, a **distribuição de probabilidades** do fenômeno estudado e as **incertezas** relacionadas a utilização de uma amostra.



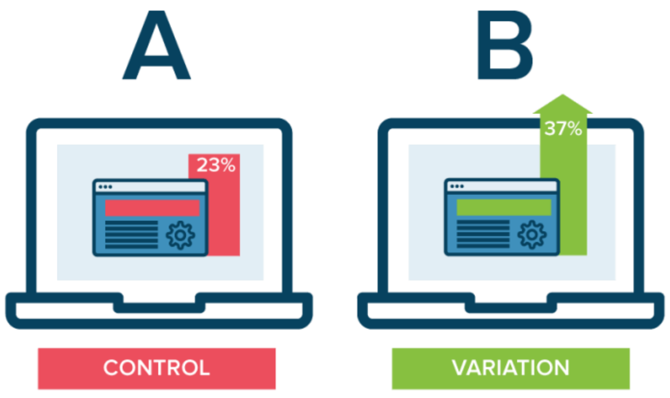
Noções de Inferência Estatística

Introdução aos Testes de Hipóteses

Por essa grande relevância, os **Testes de Hipóteses** estão muito presentes em nosso dia a dia, em aplicações como por exemplo:

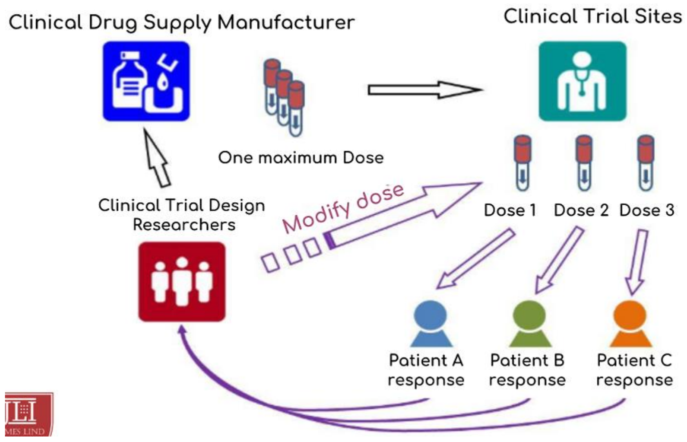
Testes A/B

Comparação do efeito de duas campanhas diferentes de marketing.



Estudos Clínicos

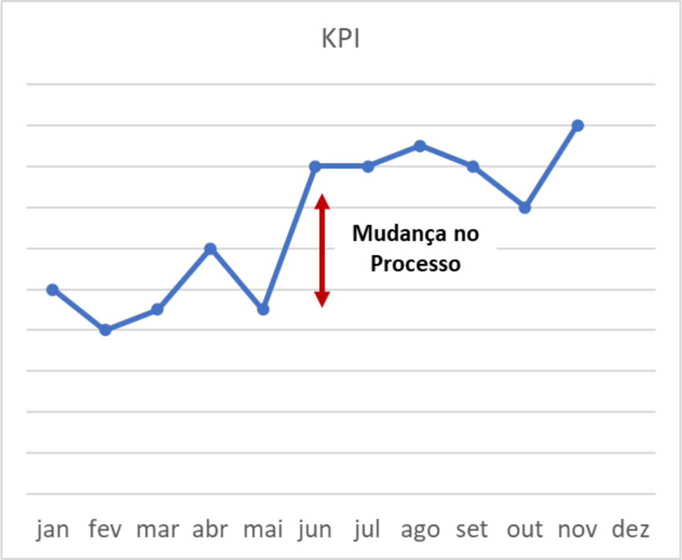
Medicamento vs. Placebo, Tratamento A vs. Tratamento B.



Fonte: https://www.jliedu.com/blog/adaptive-design-clinical-trials/

Melhoria Contínua

Desempenho do novo processo vs. processo antigo.



Noções de Inferência Estatística

Testes de Hipóteses

Média de uma população

Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

Preditiva.ai

Uma empresa responsável por realizar testes de qualidade nas águas das represas que abastecem a região metropolitana da cidade de São Paulo coletou **100 amostras de água da represa Billings**. Após realizar a análise de pH, a empresa obteve os seguintes dados:

Represa	Nº Amostras	pH	
		Média	Desvio Padrão
Billings	100	5,8	1,4

Como a Legislação Brasileira define que apenas águas com o **pH entre 6 e 9 sejam utilizadas no abastecimento**, surgiu a seguinte pergunta:

Deve-se **suspender o fornecimento de água** das regiões abastecidas pela represa Billings?

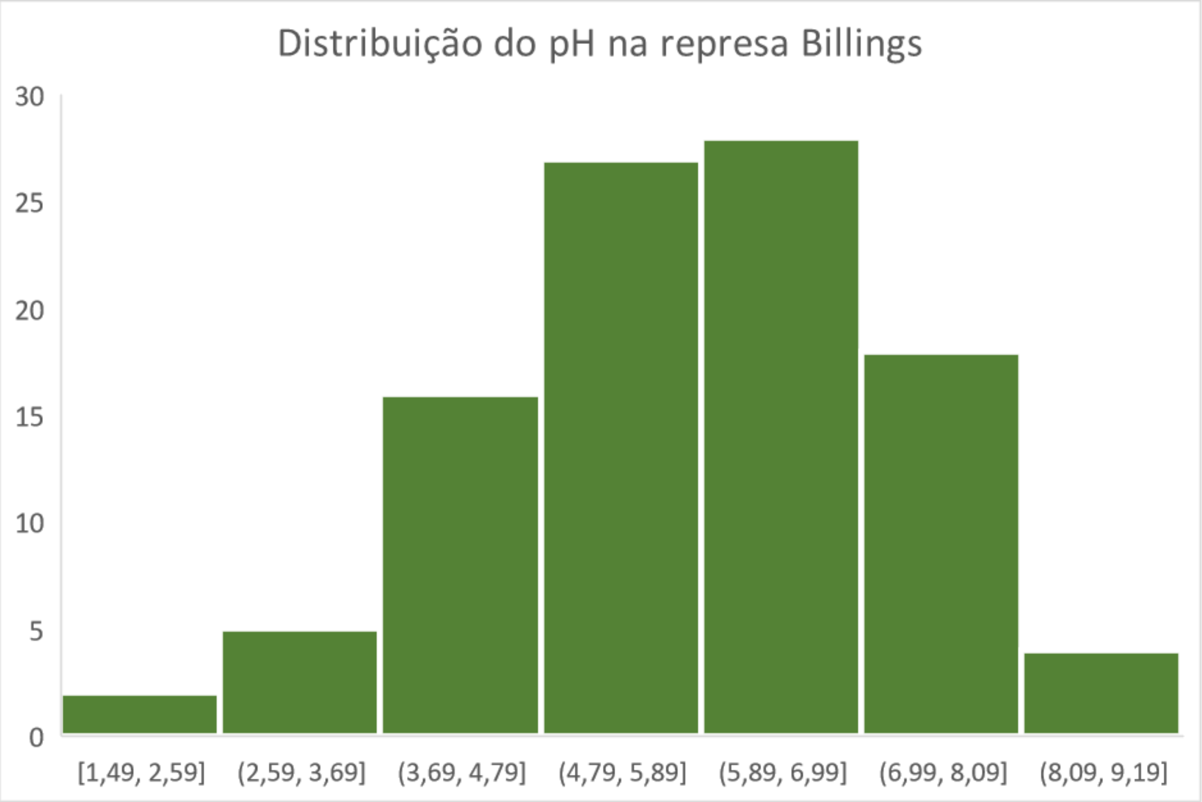
Noções de Inferência Estatística
Teste de Hipóteses: Média de uma população

Preditiva.ai

Lembre-se: como estamos analisando **dados de uma amostra**, devemos utilizar corretamente as **técnicas de inferência** para tomarmos a **melhor decisão**. E neste caso, devemos utilizar os Testes de Hipóteses para realizar essa comparação corretamente.

Represa	Nº Amostras	pH	
		Média	Desvio Padrão
Billings	100	5,8	1,4

Como queremos **comparar a Média do pH** medido nas 100 amostras com um **valor de referência** e o histograma do pH se assemelha a uma **distribuição Normal**, o Teste de Hipóteses mais adequado é o **Teste-t**.



Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

Para responder a pergunta se deve-se ou não suspender o abastecimento de água, precisamos saber se a **média do pH medido a partir das 100 amostras** é **estatisticamente** igual a 6 ou inferior a 6.

A **Hipótese Nula** sempre apresentará uma **igualdade**, pois é sob essa condição que calculamos a **estatística do teste**. Já a **Hipótese Alternativa** sempre apresentará uma **desigualdade**.

Dessa forma, teremos as seguintes hipóteses **nula** e **alternativa**:

H_0: O pH das águas da represa Billings é igual a 6, ou: **pH = 6**

H_1: O pH das águas da represa Billings é inferior a 6, ou: **pH < 6**

Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

Preditiva.ai

2º Passo: Calcular a estatística do teste

Cada **Teste de Hipóteses** possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

Como utilizaremos o **Teste-t**, sua **estatística de teste** é definida pela equação abaixo:

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S}$$

n: tamanho da amostra

$\boldsymbol{\bar{X}}$: média amostral

$\boldsymbol{\mu_0}$: média sob a hipótese H_0

S: desvio padrão amostral

Realizando o cálculo com os dados que obtivemos das 100 amostras, temos:

$$T = \frac{\sqrt{n}(\bar{X} - \mu_0)}{S} = \frac{\sqrt{100}(5{,}8 - 6)}{1{,}4} = -1{,}43$$

Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

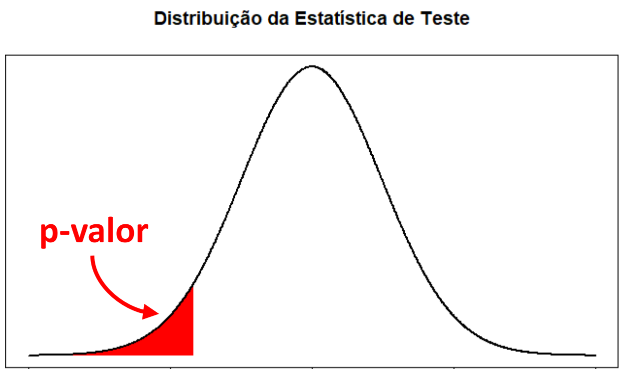
3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

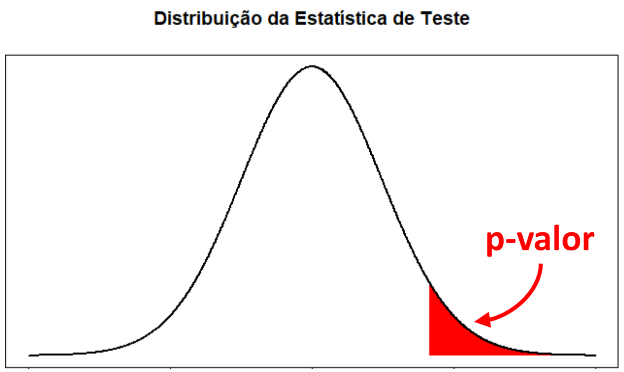
A **estatística do Teste-t** possui uma **distribuição** de probabilidade **t-Student** com n-1 graus de liberdade. É essa distribuição que utilizaremos para calcular o **quão plausível** é a **hipótese nula**.

Além da distribuição da estatística do teste, usamos também a **hipótese alternativa**, pois **dependendo da desigualdade** definida, calcularemos regiões diferentes da distribuição t-Student.

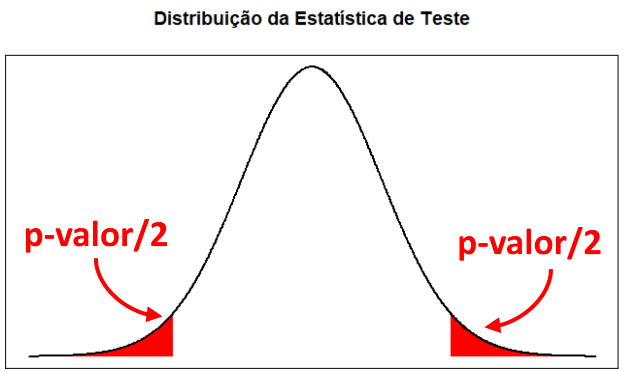
$\boldsymbol{H_1}: <$



$\boldsymbol{H_1}: >$



$\boldsymbol{H_1}: \neq$



Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

Preditiva.ai

3º Passo: Calcular o p-valor

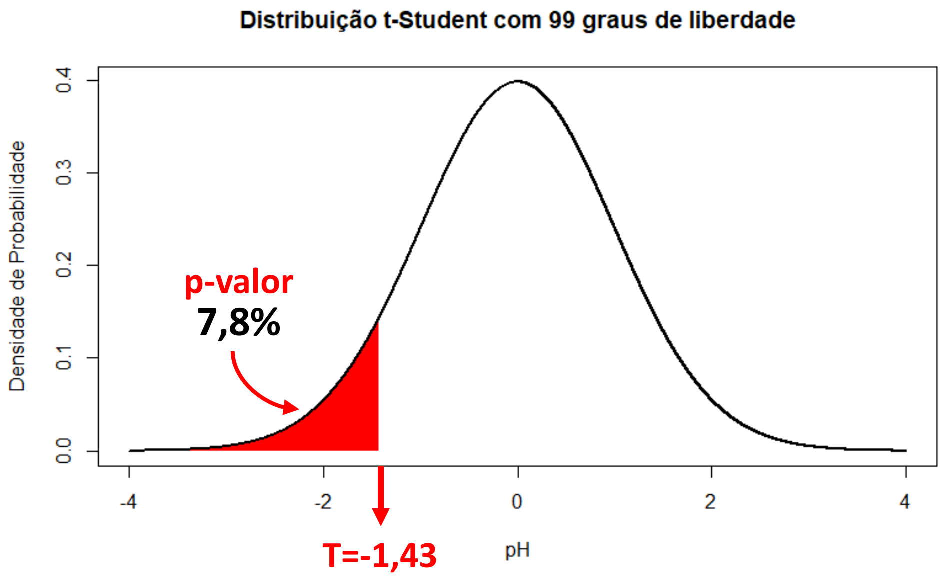
Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

Relembrando as hipóteses definidas:

H_0: O pH das águas da represa Billings é igual a 6, ou: **pH = 6**

H_1: O pH das águas da represa Billings é inferior a 6, ou: **pH < 6**

Como a hipótese alternativa foi definida como "<", calcularemos o **p-valor** na cauda esquerda da **distribuição** de probabilidade **t-Student** com 99 graus de liberdade.



Para o cálculo, vamos utilizar a função **DIST.T** do Excel:

DIST.T(T; n-1; VERDADEIRO)

DIST.T(-1,43; 99; VERDADEIRO) = 0,078

Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

Preditiva.ai

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese H_0.

Agora vamos comparar o **p-valor** calculado de **7,8%** com a **Escala de significância de Fisher**:

Nível significância	10%	5%	2,5%	1%	0,5%	0,1%
Evidências contra H_0	Marginal	Moderada	Substancial	Forte	Muito forte	Fortíssima

O valor de **7,8%** fica localizado entre "**Marginal**" e "**Moderada**", o que significa que as **evidências contra H_0 não são tão fortes**.

Como regra geral, **compara-se o p-valor com o nível de significância de 5%**:

- **P-valor inferior** a 5%: **rejeição** da hipótese nula (H_0)
- **P-valor superior** a 5%: **não rejeição** da hipótese nula (H_0)

Noções de Inferência Estatística
Teste de Hipóteses: Média de uma população

Preditiva.ai

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese **H_0**.

Como o **p-valor** de **7,8%** é maior do que o **nível de significância** de 5%, podemos concluir então que **não existem evidências estatísticas suficientes contra H_0**, ou seja, **não rejeitamos H_0**.

Relembrando as hipóteses definidas:

H_0: O pH das águas da represa Billings é igual a 6, ou: **pH = 6**

H_1: O pH das águas da represa Billings é inferior a 6, ou: **pH < 6**

E como **não rejeitamos H_0**, podemos dizer que **não existem evidências estatísticas** de que o **pH das águas da represa Billings não seja igual a 6**.

Portanto, **não é necessário suspender o fornecimento de água** das regiões abastecidas pela represa Billings.

Noções de Inferência Estatística
Estrutura dos Testes de Hipóteses

Preditiva.ai

Os **4 passos** que acabamos de realizar constituem o método de aplicação dos **Testes de Hipóteses**, e serão sempre os mesmos para qualquer **Teste de Hipóteses**.

Entender o conceito de cada passo te possibilitará utilizar qualquer **Teste de Hipóteses**, afinal existem muitos, cada um adequado para atender uma necessidade de análise.

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

2º Passo: Calcular a estatística do teste

Cada **Teste de Hipóteses** possui sua própria **estatística do teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é H_0**.

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese H_0.

Noções de Inferência Estatística

Teste de Hipóteses: Média de uma população

Preditiva.ai

O que significa rejeitar ou não rejeitar H_0?

2 Hipóteses:

Culpado ou Inocente?

Levantamento:

Conjunto de provas (evidências)

Decisão:

Evidências suficientes → Culpado

Não há evidências suficientes → Inocente

Noções de Inferência Estatística
Testes de Hipóteses
Média e Variância de duas populações

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações

A mesma empresa responsável pelos testes de qualidade nas águas das represas coletou agora **110 amostras de água da represa Guarapiranga**, realizou as mesmas análises de pH, e obteve os dados descritos na tabela abaixo:

Represa	Nº Amostras	pH	
		Média	Desvio Padrão
Billings	100	5,8	1,4
Guarapiranga	110	6,3	1,5

O órgão de fiscalização suspeita que o pH das águas da represa Guarapiranga seja **superior** ao pH das águas da represa Billings, e solicitou a empresa um estudo para confirmação.

Noções de Inferência Estatística

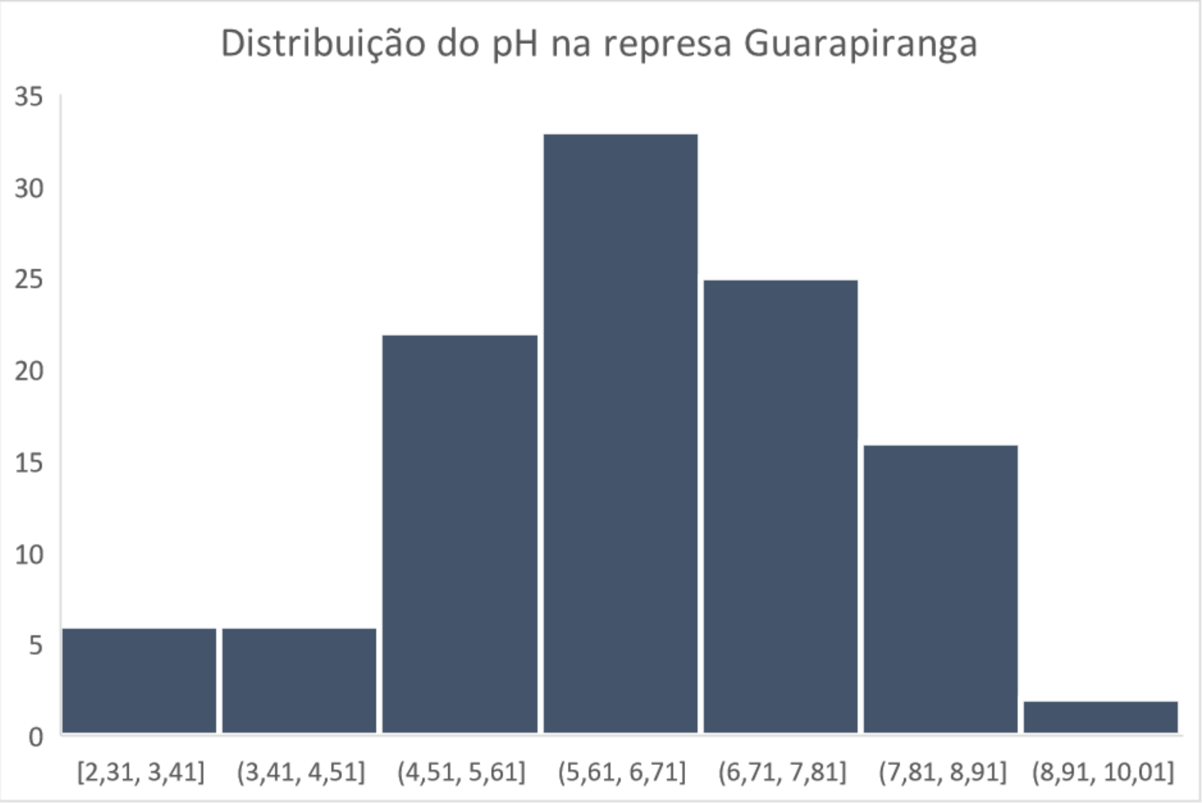
Teste de Hipóteses: Média de duas populações

Preditiva.ai

Por estarmos trabalhando com **amostras** e querermos fazer a **comparação** dos valores dessas populações, precisamos utilizar os Testes de Hipóteses para dar uma **resposta confiável** para a pergunta: as águas da Guarapiranga possuem um pH superior as águas da Billings?

Represa	Nº Amostras	pH	
		Média	Desvio Padrão
Billings	100	5,8	1,4
Guarapiranga	110	6,3	1,5

Como o **histograma das amostras** da represa Guarapiranga também se assemelha a uma **distribuição Normal**, vamos utilizar o **Teste-t para duas populações**.



Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

Para responder a pergunta precisamos saber se a **média do pH da represa Guarapiranga** é **estatisticamente maior** do que a **média do pH da represa Billings**.

Como já sabemos que a **Hipótese Nula** sempre apresentará uma **igualdade** e a **Hipótese Alternativa** uma **desigualdade**, definimos as seguintes hipóteses **nula** e **alternativa**:

H_0: O pH da Guarapiranga é **igual** ao pH da Billings, ou: $pH_G = pH_B$

H_1: O pH da Guarapiranga é **superior** ao pH da Billings, ou: $pH_G > pH_B$

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações

Preditiva.ai

2º Passo: Calcular a estatística do teste

Cada Teste de Hipóteses possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

No caso do **Teste-t para 2 populações**, o Excel possui a função **TESTE.T** que calcula a **estatística de teste** e o **p-valor** em apenas 1 passo.

pH Billings	pH Guarapiranga
4,59	7,88
5,64	5,43
3,28	6,52
4,63	6,75
6,93	7,11
7,30	7,23
7,89	3,95
[illegible]	[illegible]

TESTE.T(matriz 1; matriz 2; caudas; tipo)

caudas: 1: unicaudal ou 2: bicaudal

tipo: 1: par, 2: variâncias iguais ou 3: variâncias diferentes

Em **caudas**, quando H_1 é "<" ou ">" escolhemos **unicaudal**, e quando H_1 é "≠" escolhemos **bicaudal**. Como o H_1 definido é ">", utilizaremos "**1**" para indicar a opção **unicaudal**.

Já em **tipo**, indicamos qual o tipo do teste. A opção "par" não se aplica neste caso (veremos mais adiante), mas vamos precisar escolher entre "**variâncias iguais**" e "**variâncias diferentes**".

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações

Preditiva.ai

2º Passo: Calcular a estatística do teste

Cada Teste de Hipóteses possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

No caso do **Teste-t para 2 populações**, o Excel possui a função **TESTE.T** que calcula a **estatística de teste** e o **p-valor** em apenas 1 passo.

pH Billings	pH Guarapiranga
4,59	7,88
5,64	5,43
3,28	6,52
4,63	6,75
6,93	7,11
7,30	7,23
7,89	3,95

TESTE.T(matriz 1; matriz 2; caudas; tipo)

caudas: 1: unicaudal ou 2: bicaudal

tipo: 1: par, 2: variâncias iguais ou 3: variâncias diferentes

Para podermos escolher a **opção correta** correta, precisaremos fazer um Teste de Hipóteses para testar se as **variâncias do pH** das águas das duas represas **podem ser consideradas iguais**. Para isso, vamos fazer uma pausa neste teste de média e aplicar o **Teste-F** que é utilizado para avaliar se as **duas variâncias podem ser consideradas iguais** quando as amostras possuem **distribuição** semelhante a **Normal**.

Preditiva.ai

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

Como queremos saber exatamente se as **variâncias do pH** das águas das duas represas **podem ser consideradas iguais**, definimos as seguintes hipóteses **nula** e **alternativa**:

H_0: As variâncias do pH da Billings e Guarapiranga são **iguais**, ou: $\mathbf{Var_{pHG} = Var_{pHB}}$

H_1: As variâncias do pH da Billings e Guarapiranga são **diferentes**, ou: $\mathbf{Var_{pHG} \neq Var_{pHB}}$

Noções de Inferência Estatística

Teste de Hipóteses: Variância de duas populações

Preditiva.ai

2º Passo: Calcular a estatística do teste

Cada Teste de Hipóteses possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é H_0**.

No caso do **Teste-F** , o Excel possui a função **TESTE.F** que calcula a **estatística de teste** e o **p-valor** em apenas 1 passo.

pH Billings	pH Guarapiranga
4,59	7,88
5,64	5,43
3,28	6,52
4,63	6,75
6,93	7,11
7,30	7,23
7,89	3,95
[illegible]	[illegible]

TESTE.F(matriz 1; matriz 2)

TESTE.F(matriz 1 ; matriz 2) = 38,3%

O resultado da função **TESTE.F** foi **38,3%**. Como o **p-valor** indica o **quão plausível é H_0**, podemos antecipar que neste teste **H_0 se mostra bastante plausível**.

Noções de Inferência Estatística

Teste de Hipóteses: Variância de duas populações

Preditiva.ai

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese **H_0**.

Como o **p-valor** de **38,3%** é maior que o **nível de significância** de **5%**, podemos concluir que **não existem evidências estatísticas suficientes contra H_0**, ou seja, **não rejeitamos H_0**.

Relembrando as hipóteses definidas:

H_0: As variâncias do pH da Billings e Guarapiranga são **iguais**, ou: **$Var_{pHG} = Var_{pHB}$**

H_1: As variâncias do pH da Billings e Guarapiranga são **diferentes**, ou: **$Var_{pHG} \neq Var_{pHB}$**

E como **não rejeitamos H_0**, podemos dizer que **não existem evidências estatísticas** de que a **variância do pH das águas das duas represas não sejam iguais**.

Portanto, devemos escolher a opção "**2: variâncias iguais**".

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações

Preditiva.ai

2º Passo: Calcular a estatística do teste

Cada Teste de Hipóteses possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

Voltando a comparação das medias, sabemos agora que o tipo a ser informado é o **2: variâncias iguais**.

pH Billings	pH Guarapiranga
4,59	7,88
5,64	5,43
3,28	6,52
4,63	6,75
6,93	7,11
7,30	7,23
7,89	3,95
[illegible]	[illegible]

TESTE.T(matriz 1; matriz 2; caudas; tipo)

caudas: 1: unicaudal ou 2: bicaudal

tipo: 1: par, 2: variâncias iguais ou 3: variâncias diferentes

TESTE.T(matriz 1 ; matriz 2 ; 1 ; 2) = 0,7%

O resultado da função **TESTE.F** foi **0,7%**. Como o **p-valor** indica o **quão plausível é H_0**, podemos antecipar que neste teste **H_0 se mostra pouco plausível**.

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações

Preditiva.ai

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese **H_0**.

Como o **p-valor** de **0,7%** é menor que o **nível de significância** de **5%**, podemos concluir que **existem evidências estatísticas suficientes contra H_0**, ou seja, **rejeitamos H_0**.

Relembrando as hipóteses definidas:

H_0: O pH da Guarapiranga é **igual** ao pH da Billings, ou: **$pH_G = pH_B$**

H_1: O pH da Guarapiranga é **superior** ao pH da Billings, ou: **$pH_G > pH_B$**

E como **rejeitamos H_0**, podemos dizer que **existem evidências estatísticas** de que a **média do pH** das águas da **Guarapiranga é maior** do que a **média do pH** das águas da **Billings**.

Noções de Inferência Estatística
Testes de Hipóteses
Média de duas populações pareadas

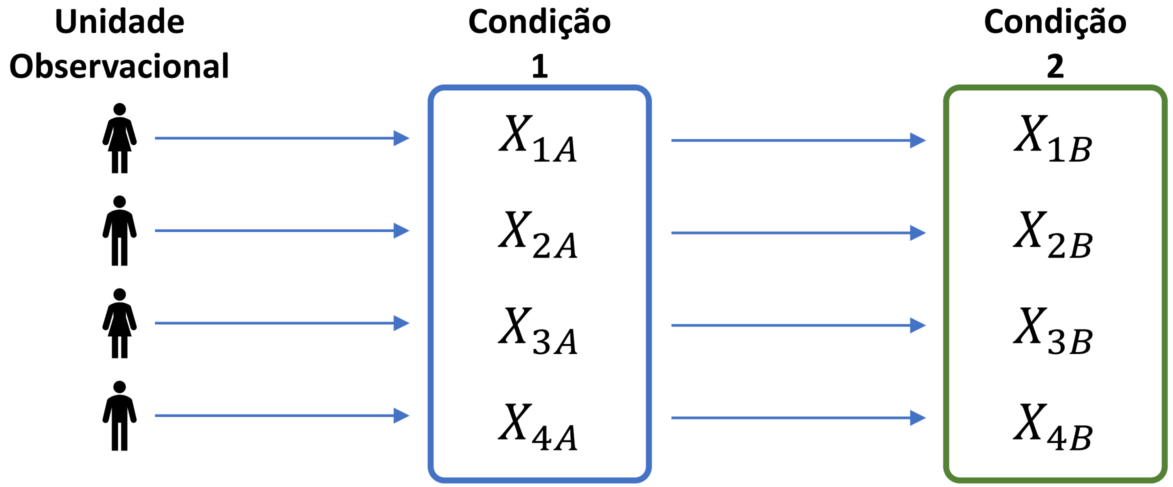
Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações pareadas

Um outro **Teste de Hipóteses** bastante útil é o **Teste-t Pareado**, muito utilizado para avaliar as mesmas unidades observacionais em diferentes condições. Por exemplo:

- Pessoas **antes** e **depois** de um tratamento.
- Desempenho de máquinas **antes** e **depois** de um ajuste.
- Produtividade de colaboradores em **home office** e no **escritório.**

Ou seja, o objetivo é avaliar se a **diferença entre as médias** nas diferentes condições **é igual a zero**.



Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações pareadas

Buscando reduzir custos e melhorar a qualidade de vida de seus colaboradores, uma empresa de atendimento decidiu fazer um **teste** para avaliar se a **adoção do home office** para os operadores produziria algum **efeito negativo** na qualidade dos atendimentos.

Para o teste foram **selecionados aleatoriamente** 30 operadores que tiveram as **médias das notas de avaliação** (0 a 10) dos clientes registradas nos **2 locais de trabalho**.

Local de Trabalho	Nº Operadores	Avaliação dos clientes	
		Média	Desvio Padrão
Escritório	30	7,8	1,2
Home office		7,4	1,5

Com os **dados obtidos**, como podemos **responder a pergunta:** "Operadores trabalhando em home office possuem pior avaliação?"

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações pareadas

Preditiva.ai

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

Para responder a pergunta precisamos saber se a **média das notas dos clientes para os operadores trabalhando no escritório** é **estatisticamente maior** do que a **média das notas dos clientes para os mesmos operadores trabalhando em home office**.

Relembrando, a **Hipótese Nula** sempre apresentará uma **igualdade** e a **Hipótese Alternativa** uma **desigualdade**. Portanto definimos as seguintes hipóteses **nula** e **alternativa**:

H_0: As notas no Escritório são **iguais** as notas em Home Office, ou: $N_E = N_{HO}$

H_1: As notas no Escritório são **maiores** que as notas em Home Office, ou: $N_E > N_{HO}$

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações pareadas

2º Passo: Calcular a estatística do teste

Cada Teste de Hipóteses possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o nível descritivo

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

No caso do **Teste-t para 2 populações pareadas**, o Excel possui a função **TESTE.T** que calcula a **estatística de teste** e o **p-valor** em apenas 1 passo.

Operador	Escritório	Home office
1	8,9	6,9
2	8,3	9,8
3	6,4	7,9
4	8,5	9,6
5	6,2	6,3
6	7,2	5,7
7	6,6	7,0
8	6,0	7,1

TESTE.T(matriz 1; matriz 2; caudas; tipo)

caudas: 1: unicaudal ou 2: bicaudal

tipo: 1: par, 2: variâncias iguais ou 3: variâncias diferentes

TESTE.T(matriz 1 ; matriz 2 ; 1 ; 1) = 12%

Em **caudas**, como o H_1 definido é "**>**", utilizaremos "**1**" para indicar a opção **unicaudal**.

Já em **tipo**, utilizaremos a opção "**1**" para indicar que as populações são **pareadas**.

Já sabemos que **p-valor** indica o **quão plausível é** H_0, e neste caso, o valor calculado de **12%** indica que **H_0 é plausível**.

Noções de Inferência Estatística

Teste de Hipóteses: Média de duas populações pareadas

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese H_0.

Como o **p-valor** de **12%** é maior que o **nível de significância** de **5%**, podemos concluir que **não existem evidências estatísticas suficientes contra H_0**, ou seja, **não rejeitamos H_0**.

Relembrando as hipóteses definidas:

H_0: As notas no Escritório são **iguais** as notas em Home Office, ou: $N_E = N_{HO}$

H_1: As notas no Escritório são **maiores** que as notas em Home Office, ou: $N_E > N_{HO}$

E como **não rejeitamos H_0**, podemos dizer que **não existem evidências estatísticas** de que a **média das notas** dos operadores no **Escritório não seja igual** a **média das notas** dos mesmos operadores em **Home Office**.

Noções de Inferência Estatística
Testes de Hipóteses
Proporção de duas populações

Noções de Inferência Estatística

Teste de Hipóteses: Proporção de duas populações

Preditiva.ai

Com todos os esforços para combater o COVID-19, surgiram diversas iniciativas para desenvolver uma vacina **eficaz para reduzir o impacto dos sintomas**. Um laboratório em suas pesquisas chegou a **duas estratégias** viáveis:

1. Vírus enfraquecido
2. RNA mensageiro

Para decidir por qual estratégia seguir, realizou um estudo clínico com 87 pessoas distribuídas **aleatoriamente** entre as duas estratégias.

Com os resultados obtidos, como podemos **responder**: " Existe diferença na estratégia em relação à proporção de sintomas graves?"

Estratégia	Nº Amostras	Sintomas	
		Graves	Leves
Vírus enfraquecido	38	16%	84%
RNA mensageiro	50	28%	72%

Imagem: https://www.news-medical.net/news/20201124/25283/Portuguese.aspx

Noções de Inferência Estatística

Teste de Hipóteses: Proporção de duas populações

Preditiva.ai

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

Para responder a pergunta precisamos saber se a **proporção** de pessoas com sintomas graves tendo sido vacinadas com o **vírus enfraquecido (VE)** é **estatisticamente diferente** da **proporção** de pessoas com sintomas graves que foram vacinadas **com o RNA mensageiro (RNA)**.

Definimos então as seguintes hipóteses **nula** e **alternativa**:

H_0: A proporção sintomas graves VE é **igual** a proporção sintomas graves RNA, ou: $\mathbf{p_{VE} = p_{RNA}}$

H_1: A proporção sintomas graves VE é **diferente** da proporção sintomas graves RNA, ou: $\mathbf{p_{VE} \neq p_{RNA}}$

Noções de Inferência Estatística

Teste de Hipóteses: Proporção de duas populações

2º Passo: Calcular a estatística do teste

Cada **Teste de Hipóteses** possui sua própria **estatística de teste**, que é um cálculo realizado com os **dados da amostra**.

Para esta comparação utilizaremos o **Teste-Z para 2 populações**, que possui a seguinte **estatística de teste**:

$$Z = \frac{(\hat{p}_{VE} - \hat{p}_{RNA})}{\sqrt{\hat{p}(1-\hat{p})\left(\frac{1}{n_{VE}} + \frac{1}{n_{RNA}}\right)}}$$

$$\hat{p} = \frac{\hat{p}_{VE} \cdot n_{VE} + \hat{p}_{RNA} \cdot n_{RNA}}{n_{VE} + n_{RNA}}$$

$\hat{p}_{VE}$: proporção sintomas graves - vírus enfraquecido

$\hat{p}_{RNA}$: proporção sintomas graves - RNA mensageiro

$\hat{p}$: proporção geral de sintomas graves

n_{VE}: número de pessoas - vírus enfraquecido

n_{RNA}: número de pessoas - RNA mensageiro

Noções de Inferência Estatística

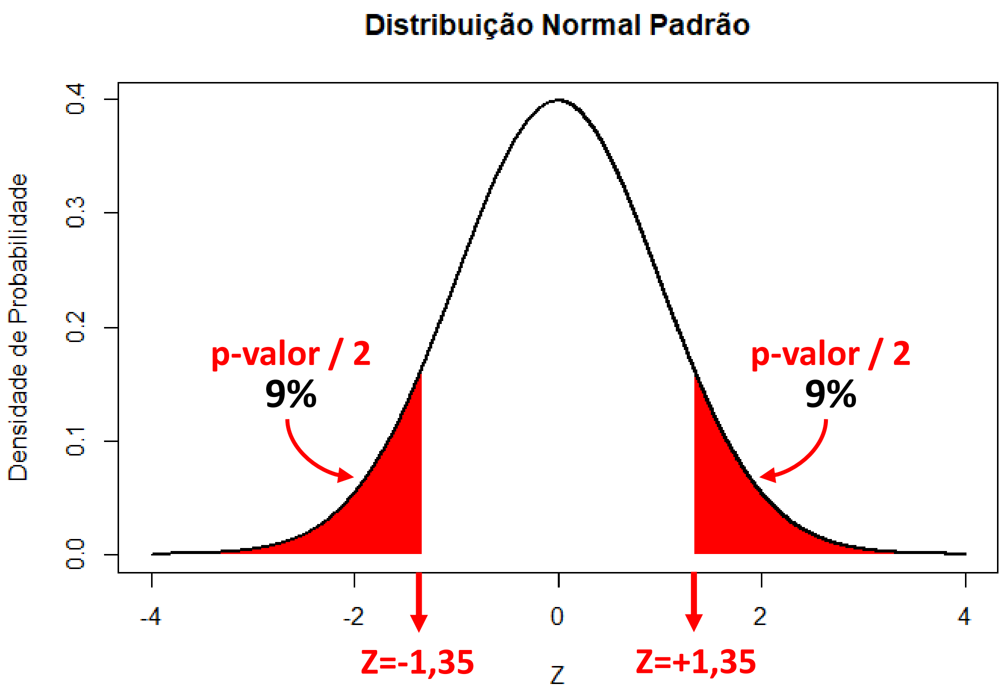
Teste de Hipóteses: Proporção de duas populações

Preditiva.ai

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é H_0**.

A **estatística do Teste-Z para 2 populações** possui uma **distribuição** de probabilidade **Normal Padrão**. E como a **hipótese alternativa**, foi definida como " ≠ ", então o **p-valor** será calculado **somando** as probabilidades nas **duas caudas**.



DIST.NORMP.N(Z; VERDADEIRO)

DIST.NORMP.N(-1,35; VERDADEIRO) = 9%

Multiplicando o **p-valor** calculado por 2, obtemos **18%**.

Noções de Inferência Estatística

Teste de Hipóteses: Proporção de duas populações

Preditiva.ai

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese **H_0**.

Como o **p-valor** de **18%** é maior que o **nível de significância** de **5%**, podemos concluir que **não existem evidências estatísticas suficientes contra H_0**, ou seja, **não rejeitamos H_0**.

Relembrando as hipóteses definidas:

H_0: A proporção sintomas graves VE é **igual** a proporção sintomas graves RNA

H_1: A proporção sintomas graves VE é **diferente** da proporção sintomas graves RNA

E como **não rejeitamos H_0**, podemos dizer que **não existem evidências estatísticas** de que a **proporção** de pessoas com sintomas graves **não é igual** nos 2 tipos de vacinas.

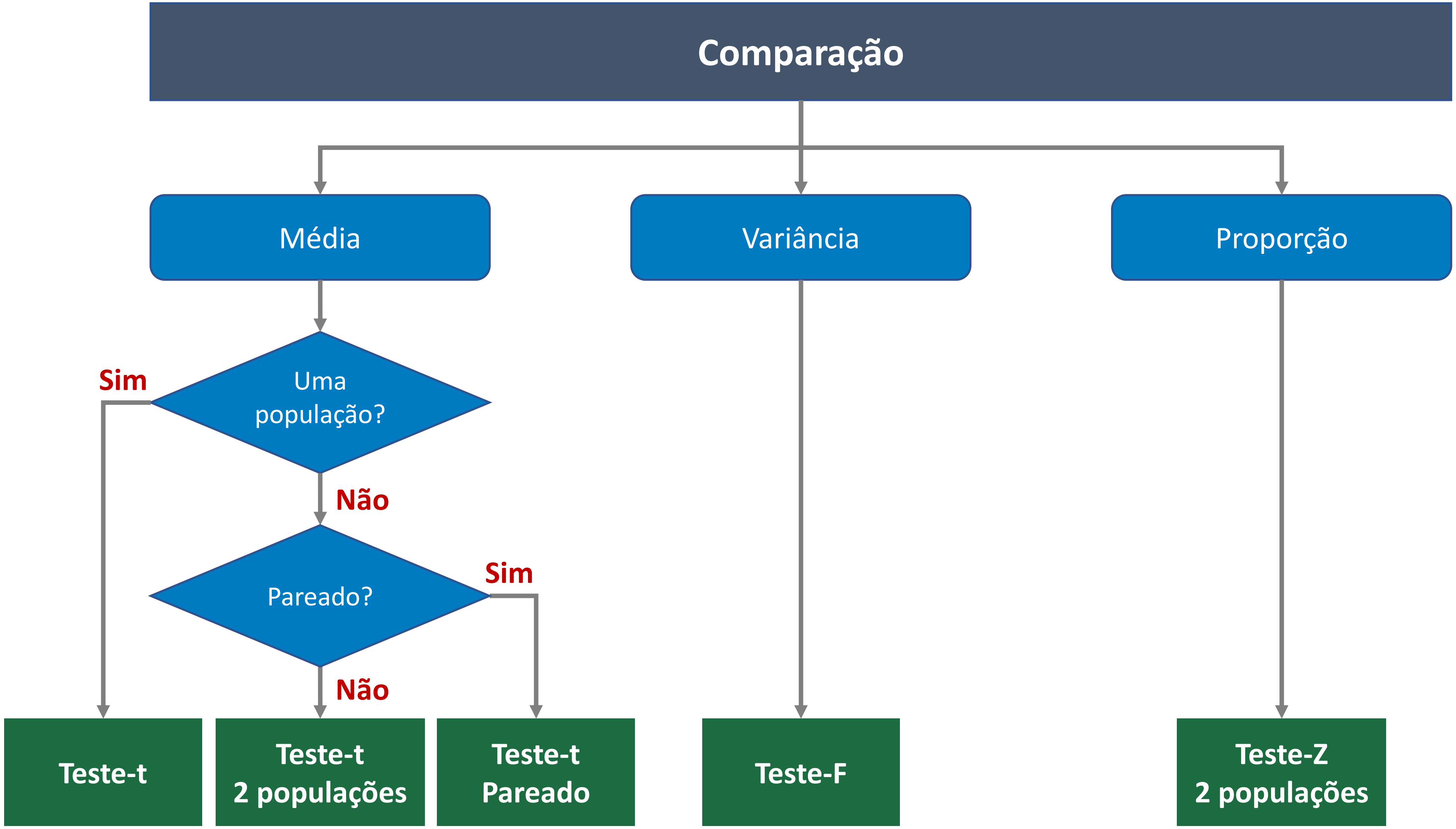
Noções de Inferência Estatística

Testes de Hipóteses

Revisão

Noções de Inferência Estatística

Aplicações dos Testes de Hipóteses



Noções de Inferência Estatística

Estrutura dos Testes de Hipóteses

Preditiva.ai

Entender o conceito de cada passo te possibilitará utilizar qualquer **Teste de Hipóteses**, afinal existem muitos, cada um adequado para atender uma necessidade de análise.

1º Passo: Definir as Hipóteses

A partir da pergunta que desejamos responder, definimos as **hipóteses que serão testadas**.

Teremos sempre:
H_0 ou **Hipótese nula**
H_1 ou **Hipótese alternativa**

2º Passo: Calcular a estatística do teste

Cada **Teste de Hipóteses** possui sua própria **estatística do teste**, que é um cálculo realizado com os **dados da amostra**.

3º Passo: Calcular o p-valor

Cada **estatística de teste** possui uma **distribuição de probabilidades**. A partir dela e das hipóteses calculamos o **p-valor**, que é uma medida de **quão plausível é** H_0.

4º Passo: Comparar e tomar decisão

Por fim, comparamos o **p-valor** com um **nível de significância** para concluir se **rejeitamos** ou **não rejeitamos** a hipótese H_0.

Noções de Inferência Estatística
Método Científico

Noções de Inferência Estatística
Método Científico

O **Método Científico** é um processo utilizado para **construção de conhecimento** baseado em **5 principais etapas**:

1. **Definição de uma questão**

Pode ser algo muito específico: "Por que o céu é azul?" ou algo mais aberto: "Como aumento as vendas em Dezembro?". Nesta etapa também são realizadas pesquisas para identificar **experimentos já realizados** e **conhecimento já existente**. **Definir bem a questão é fundamental** para o resultado do processo!

O **Método Científico** é um processo utilizado para **construção de conhecimento** baseado em **5 principais etapas**:

2. Formulação de uma hipótese

Uma hipótese é uma **explicação** ou uma **tentativa de resposta** a questão definida no passo anterior. **Ela precisa ser testável**, ou seja, deve ser possível realizar um experimento que **comprove a hipótese formulada ou a contrarie**.

Fonte: https://xkcd.com/892/

O **Método Científico** é um processo utilizado para **construção de conhecimento** baseado em **5 principais etapas**:

3. Previsão dos resultados

Antecipe o que acontecerá **se a hipótese estiver correta**. Podem haver várias previsões, e **quanto mais improvável que uma previsão esteja correta apenas por acaso, mais convincente ela será** caso se concretize. A evidência também é mais forte se a resposta para a previsão não for conhecida, pois evita o **viés de retrospectiva**.

Imagem: https://www.subpng.com/png-c40cr1/

Noções de Inferência Estatística

Método Científico

O **Método Científico** é um processo utilizado para **construção de conhecimento** baseado em **5 principais etapas**:

4. Teste ou Experimentação

Planeje um experimento para investigar se o **mundo real se comporta como previsto** na etapa 3. **Karl Popper** recomenda que os **cientistas tentem "derrubar" suas hipóteses**. Além disso, deve-se tomar todo o cuidado com **falhas que podem comprometer o resultado dos experimentos** como erros de medida, vieses humanos entre outros.

Noções de Inferência Estatística
Método Científico

O **Método Científico** é um processo utilizado para **construção de conhecimento** baseado em **5 principais etapas**:

5. Análise dos resultados

Nesta última etapa deve-se **analisar os resultados da experimentação** e concluir se a hipótese definida na etapa 2 é refutada ou não. Se a **hipótese for refutada**, deve-se definir uma **nova hipótese**. Se os resultados da experimentação **suportam a hipótese atual**, mas as evidências não são fortes, deve-se **conduzir novos experimentos**.

Noções de Inferência Estatística

Método Científico

Vamos praticar a utilização do **Método Científico**:

Um fabricante de automóveis quer descobrir **quem é o público alvo para o seu novo modelo de veículo de passeio**.

Com essa informação ele poderá **concentrar seus recursos e esforços,** e **aumentar o volume de vendas** de seu produto para as pessoas que realmente têm **interesse na compra**.

Noções de Inferência Estatística

Método Científico

Preditiva.ai

Exemplo de aplicação do **Método Científico**:

1. Questão	2. Hipóteses	3.Previsões	4.Experimentação	5.Análises
Quem é o público alvo para meu novo modelo?	Por se tratar de um sedan espaçoso, imaginamos que o público alvo sejam famílias com 4+ pessoas	Uma divulgação direcionada a esse público alvo aumentaria nosso volume de vendas	Divulgação do novo modelo em diversos meios, sem segmentar para o público alvo	Avaliar se os compradores estão caracterizados como famílias com 4+ pessoas

Resultados

Compradores com um perfil de famílias com 4+ pessoas

- Não há indícios contra a hipótese formulada: pode estar **correta**
- Se as **evidências forem fracas** (pouca diferença), podem ser necessários **novos experimentos**

Compradores sem um perfil muito definido

- Há indícios **contra a hipótese formulada**: provavelmente está **equivocada**
- Necessário **reiniciar o processo** a partir de uma **nova hipótese**

Noções de Inferência Estatística
Método Científico

Preditiva.ai

Principais **vantagens** da utilização do **Método Científico** em analytics:

1. Baseado em **evidências empíricas** e não em “achismos”.
2. Permite **replicação dos resultados** por utilizar um **processo estruturado.**
3. É um **processo iterativo**, ou seja, as **respostas** para a 1ª questão podem gerar **novas questões.**
4. Construção **cumulativa de conhecimento**, ou seja, a cada iteração **novos conhecimentos são gerados.**

Noções de Inferência Estatística
Método Científico aplicado a Dados

Uma abordagem bastante utilizada para organizar o processo de análise de dados é **CRISP-DM**.

Entendimento do negócio
- Objetivo
- Premissas
- Risco envolvidos
- Custos
- Critérios de sucesso
- Planejamento

Entendimento dos dados
- Coleta
- Descrição dos dados
- Análise Exploratória
- Sanity Check

Preparação dos dados
- Seleção de variáveis
- Limpeza dos dados
- Formatação dos dados

Análise / Modelagem
- Escolha da técnica
- Desenvolvimento do estudo ou modelo
- Finalização

Validação
- Verificação dos critérios de sucesso
- Revisão do processo/Validação do modelo
- Determinação dos próximos passos

Deploy (Implantação)
- Plano de implantação
- Monitoramento do modelo/KPI's
- Criação da documentação

Preditiva.ai