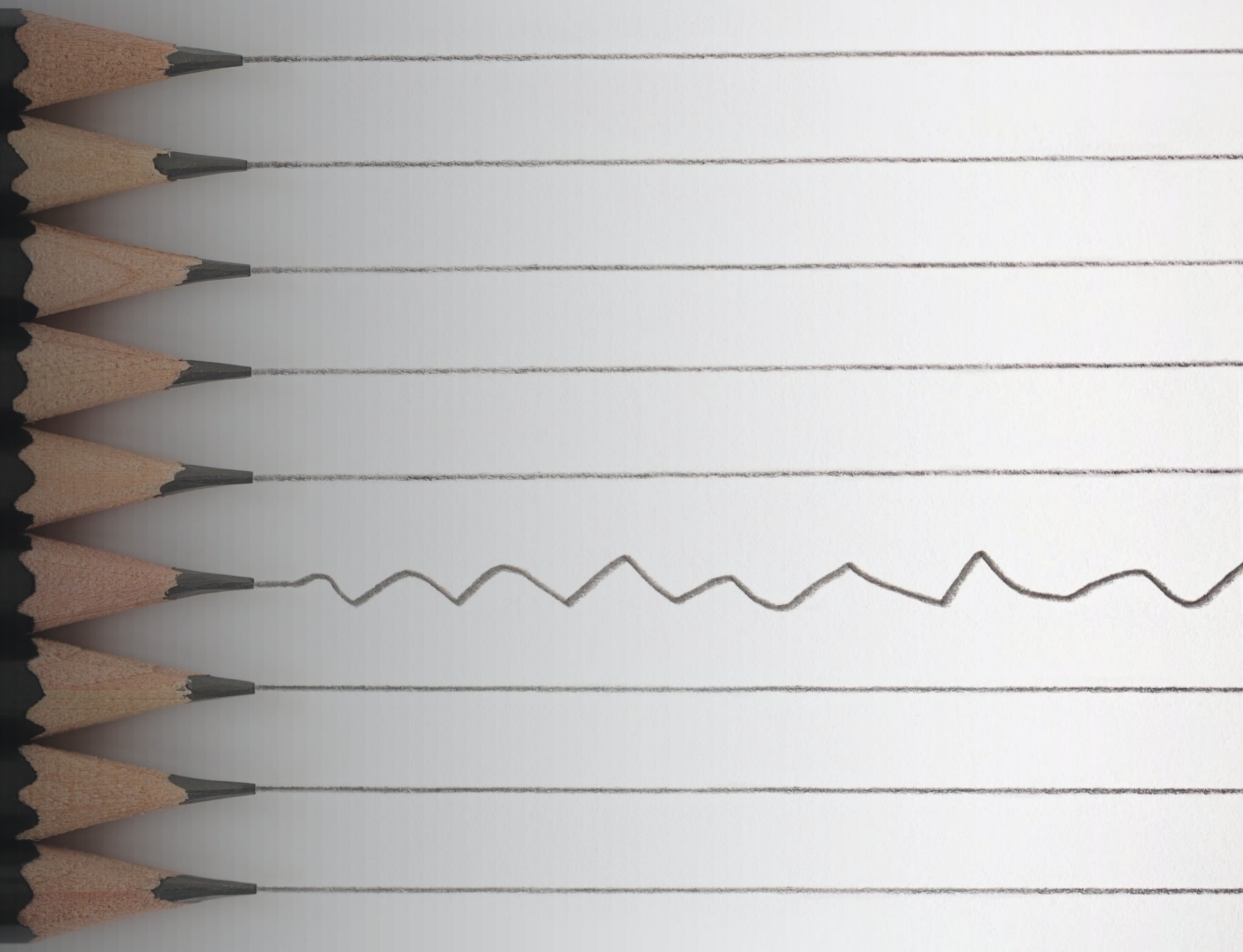
Estatística Aplicada a Ciências Ambientais

Testes de Hipótese

Daniel Detzel
detzel@ufpr.br



Agenda

Testes de hipótese

fundamentos

Testes de hipótese

testes para diferenças entre grupos

Aplicações em R

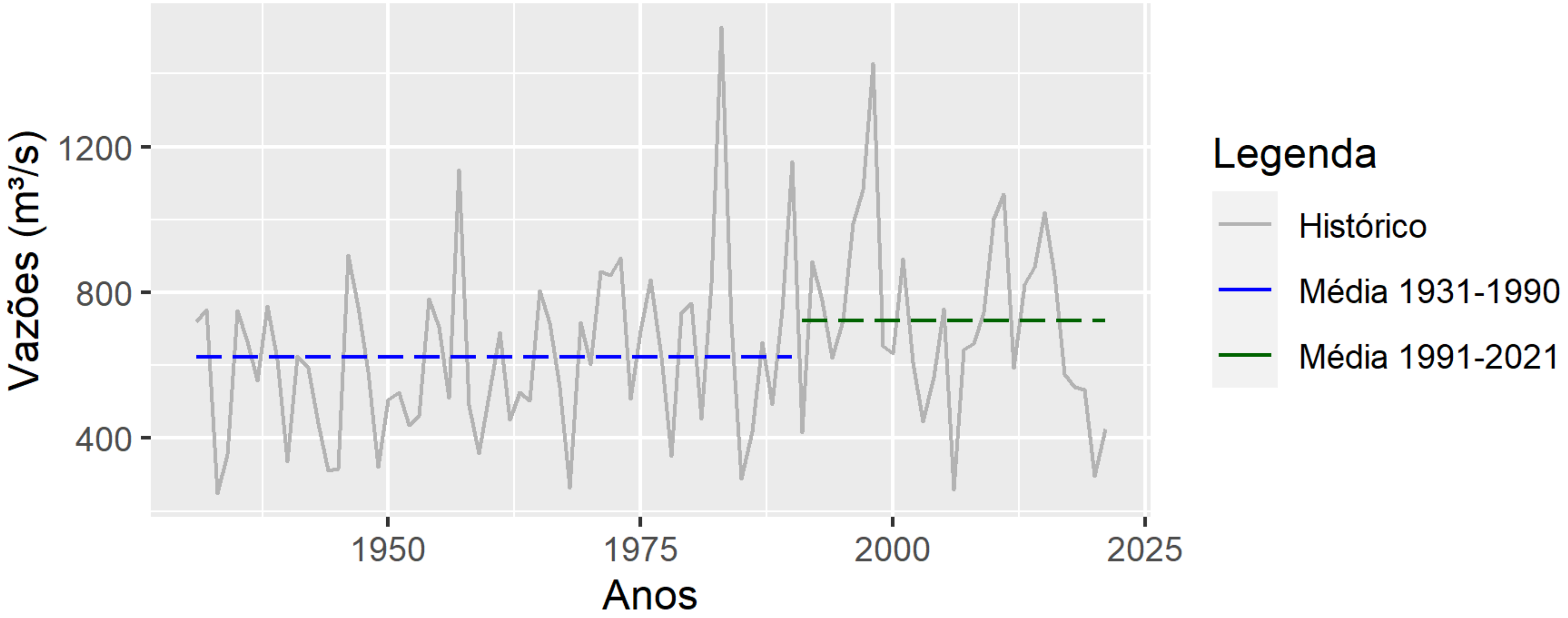
Testes de hipótese

fundamentos

Testes de hipótese | fundamentos

Considere a série de vazões médias anuais afluentes a Foz do Areia, entre 1931 e 2020.

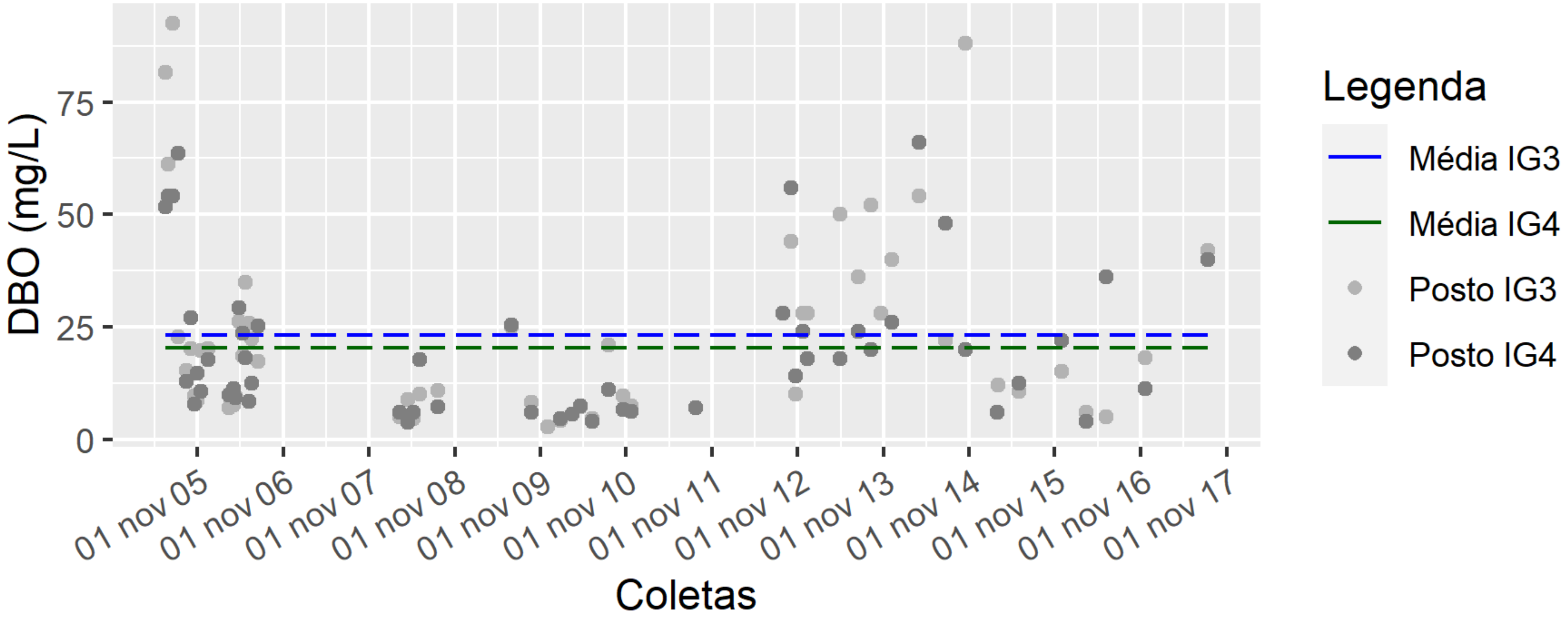
pergunta-se: houve mudança na média de longo termo dos últimos 30 anos?



Testes de hipótese | fundamentos

Considere agora a série DBO medida em dois pontos diferentes do rio Iguaçu (IG3 e IG4), entre 2005 e 2017

pergunta-se: os valores de DBO são iguais nos dois pontos?



Testes de hipótese | fundamentos

Por só se ter conhecimento de amostras e não da população, os valores (números) estimados das amostras podem ser diferentes, mas ainda sim representar estatisticamente as mesmas grandezas populacionais

Os testes de hipótese permitem obter uma resposta às perguntas levantadas

Teste de Hipótese

Regra de decisão para rejeitar, ou não, uma hipótese estatística com base na amostra

Testes de hipótese | fundamentos

Tudo se iniciou com uma xícara de chá...

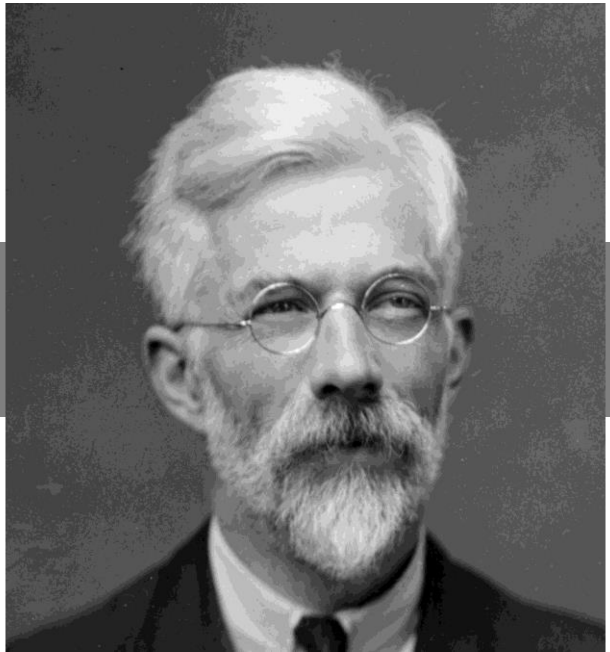
O experimento *Lady Tasting Tea* (a senhora que tomava chá)

Muriel Bristol ("the lady")

Imagem: https://mujeresconciencia.com/2017/04/21/muriel-bristol-algologa/

Sir Ronald Fisher

Imagem: http://www.42evolution.org/ronald-a-fisher/



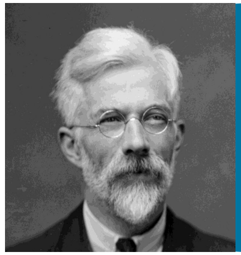
Chá inglês (com leite!)

Imagem: https://www.pinterest.co.uk/pin/412783122066184575/

Testes de hipótese | fundamentos

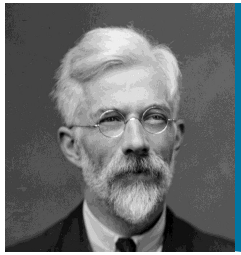
Em uma bela tarde, na hora do chá, Muriel e Fisher conversavam...

"Fisher, você sabia que em uma xícara de chá eu consigo saber quando o leite foi colocado primeiro, ou quando o chá foi colocado primeiro?"



"Ah, tá... çei..."

"É sério!! Não acredita em mim?"



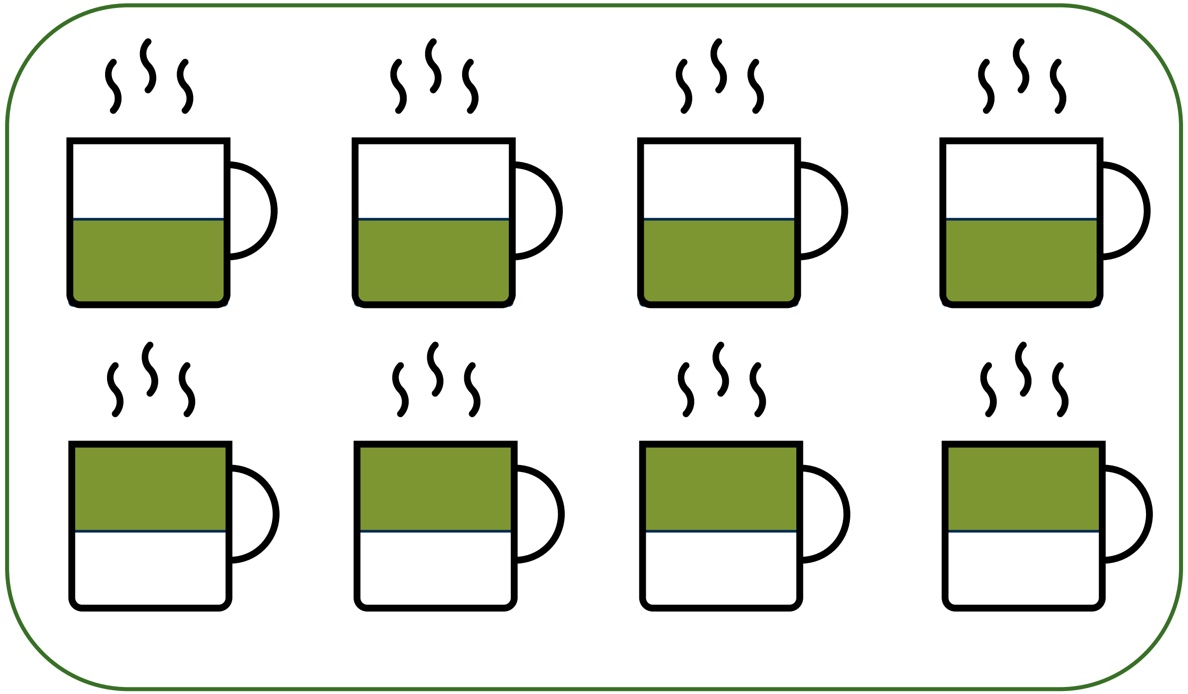
"Não... mas eu posso testar isso."

Testes de hipótese | fundamentos

O experimento:

8 xícaras: 4 com chá primeiro, 4 com leite primeiro

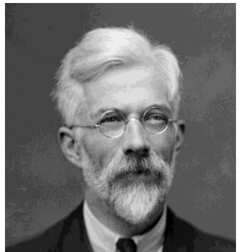
apresentar a Muriel em ordem aleatória



Análise combinatória: 70 possibilidades de arranjos

Se ela acertar, posso concluir que ela é boa mesmo?

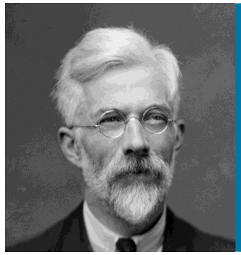
Pode ter sido chute!



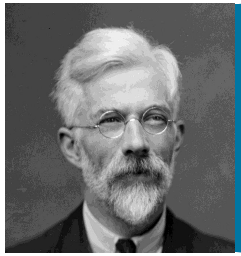
Testes de hipótese | fundamentos

No caso de a Muriel acertar, são duas as explicações:

1. Ela chutou
2. Ela é realmente boa



"Se eu adotar como hipótese que ela irá chutar, a probabilidade de acerto será de 1/70 (~1,4%). É uma probabilidade baixa..."



"Bom, então se ela acertar, posso concluir que ela é realmente boa"

Registros históricos afirmam que Muriel acertou.

"Viu só?"

Testes de hipótese | fundamentos

O raciocínio de Fisher é derivado do método hipotético-dedutivo de Karl Popper

Neste método, procuram-se evidências empíricas para rejeitar uma hipótese

A hipótese estatística é uma afirmação sobre os parâmetros de uma população

Quando a rejeição não é possível, a hipótese é válida, mas não (ou nunca) definitivamente confirmada

fatos novos podem invalidar uma hipótese a qualquer momento

Testes de hipótese | fundamentos

Popper sugere desafiar repetidamente uma hipótese

caso ela permaneça válida, diz-se que ela adquire certo grau de confiança

Em um teste estatístico, a hipótese desafiada é chamada hipótese nula (H_0)

a hipótese complementar é chamada hipótese alternativa (H_A)

O teste determinará a probabilidade de os dados terem sido observados assumindo que H_0 seja verdadeira

$$P(\text{Dados} \mid H_0 \text{ verdadeira}) = \text{p-valor}$$

Testes de hipótese | fundamentos

Se o p-valor for:

baixo → rejeita-se a hipótese nula

alto → falha-se em rejeitar a hipótese nula

De acordo com essa filosofia, nunca uma hipótese é aceita!

aceitar ≠ não rejeitar

Créditos para as explicações sobre a teoria de Popper: prof. Fernando de Pol Mayer (LEG/DEST/UFPR)

Testes de hipótese | fundamentos

Passo-a-passo para aplicação de testes de hipótese:

1. Escolha o teste apropriado
2. Formule as hipóteses nula (H_0) e alternativa (H_A)
3. Escolha um nível de significância
4. Calcule a estatística do teste com base nos dados
5. Calcule o p-valor
6. Rejeite, ou não, a hipótese H_0

Testes de hipótese | escolha o teste apropriado

1. Escolha o teste apropriado

Existem 2 grandes grupos de testes:

paramétricos:

trabalham diretamente com os dados da amostra

requerem que a distribuição de probabilidades da amostra seja conhecida (e, frequentemente, normal)

não paramétricos:

trabalham com ranques (ou índices) atribuídos aos dados da amostra

a distribuição de probabilidades da amostra não é importante

Testes de hipótese | escolha o teste apropriado

Muitos dos mais robustos e eficientes testes envolvendo média e variância são paramétricos e requerem normalidade

Caso a amostra não tenha distribuição normal, os testes perdem poder

poder do teste: probabilidade de tomar a decisão correta em rejeitar H_0 quando ela for, de fato, falsa

Contudo, há ainda a possibilidade de transformar os dados para aproximá-los de uma distribuição normal

Testes de hipótese | escolha o teste apropriado

Uma opção é a transformação de Box-Cox:

$$y_t = \begin{cases} \dfrac{x^{\lambda} - 1}{\lambda} & ; \lambda \neq 0, x > 0 \\ \ln x & ; \lambda = 0, x > 0 \end{cases}$$

No R: biblioteca "MASS"

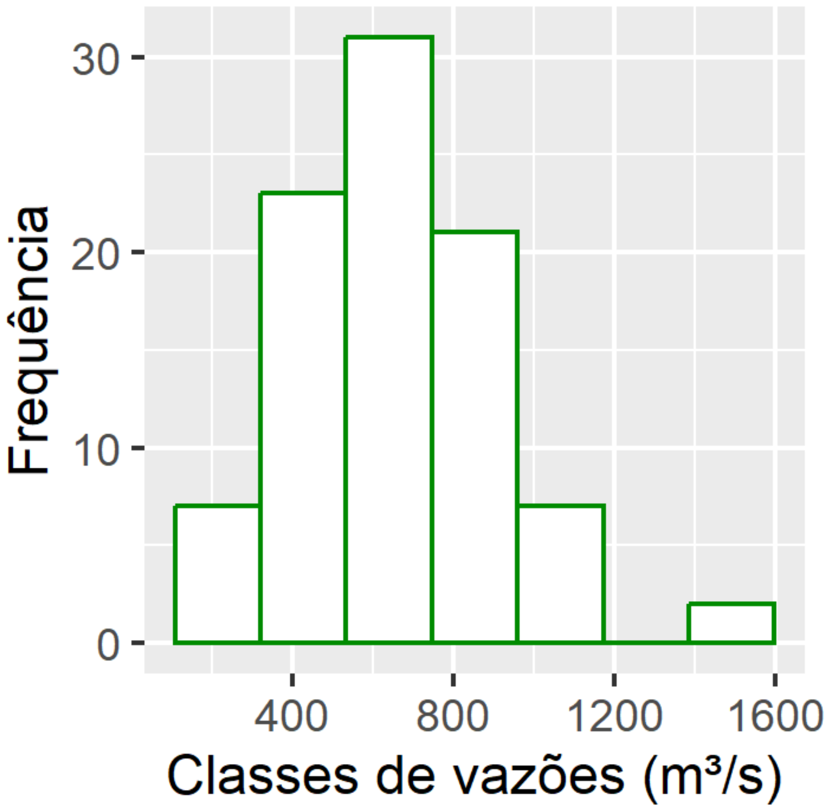
```
nome <- boxcox(lm(serie ~ 1))
```

onde λ é o parâmetro da transformação.

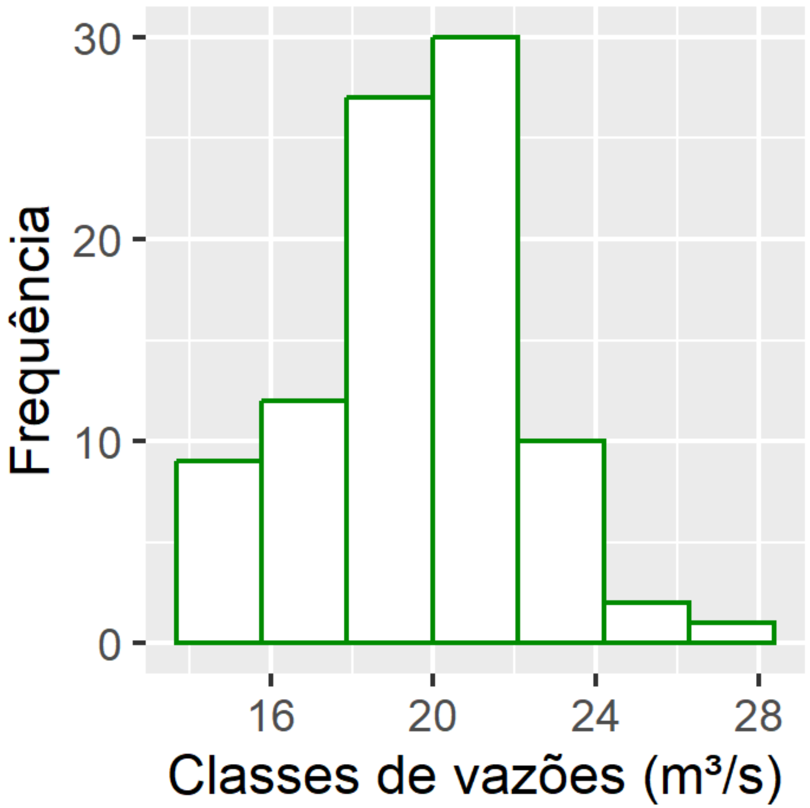
O parâmetro λ pode ser estimado por tentativa e erro, até que a amostra se aproxime satisfatoriamente de uma normal

Testes de hipótese | escolha o teste apropriado

Ex.: para Foz do Areia, antes e após a transformação ($\lambda = 0{,}3$)



Antes



Após

Testes de hipótese | formule as hipóteses

2. Formule as hipóteses nula (H_0) e alternativa (H_A)

Idealmente, as hipóteses devem ser formuladas antes da coleta ou análise dos dados, para evitar vieses

H_0: é o que se assume por verdade (do ponto de vista do analista!)

ex.: “não há diferenças entre os grupos”; “não há tendências”; etc.

H_A: é o que se assume por verdade caso H_0 seja rejeitada

ex.: “há diferenças entre os grupos”; “há tendências”; etc.

Testes de hipótese | formule as hipóteses

H_A pode ser do tipo unilateral ou bilateral

unilateral: quando a direção ou sentido da evidência estudada (maior, ou menor; positivo ou negativo) é determinante para a rejeição de H_0
ex.: a média da DBO em dois pontos diferentes do rio são iguais

bilateral: quando a direção ou sentido não é relevante para a rejeição de H_0
ex.: a média da DBO em um ponto do rio é maior (ou menor) do que do outro ponto do mesmo rio

Essa convenção é extrapolada para o teste em si (ex.: teste unilateral ou bilateral)

Testes de hipótese | escolha um nível de significância

3. Escolha um nível de significância

Epicetus (século II d.C.) disse:

> “
>
> As aparências para a mente são de quatro tipos:
>
> As coisas ou são o que parecem ser
> ou não são e nem parecem ser
> ou são e não parecem ser
> ou não são e parecem ser
>
> Identificar corretamente todos esses casos é a tarefa do homem sábio.

Testes de hipótese | escolha um nível de significância

Não conhecemos

		Realidade das “coisas”	
		São	Não são
O que parece	Parece ser	OK	Erro
	Não parece ser	Erro	OK

Testes de hipótese | escolha um nível de significância

Mecanismo dos testes: tipos de erros

		Realidade	
		H_0 verdadeira	H_0 falsa
Decisão	Não rejeitar H_0	OK $(1-\alpha)$ Confiança	Erro tipo II (β)
	Rejeitar H_0	Erro tipo I (α) Significância	OK $(1-\beta)$ Poder

Testes de hipótese | escolha um nível de significância

Nos testes de hipótese, o analista fixa um valor para α

o valor de β fica intrínseco e geralmente não é avaliado

Contudo, os erros tipo I e II estão relacionados

reduzir α conduz a um aumento de β

Valores típicos adotados para α são de 1%, 5% ou 10%, dependendo da aplicação

Testes de hipótese | calcule a estatística do teste

4. Calcule a estatística do teste com base nos dados

Cada teste de hipótese traz em sua formulação uma equação para a definição da estatística do teste

Esse valor é obtido com base na amostra e deve ser comparado a um valor teórico para que se possa tomar a decisão de rejeitar, ou não, H_0

Usualmente, os valores teóricos têm por base uma distribuição de probabilidades (t-Student, F-Snedecor-Fisher, normal, etc.)

Testes de hipótese | calcule o p-valor

5. Calcule o p-valor

O p-valor é a probabilidade que está associada à credibilidade de H_0

quanto menor o seu valor, maior a evidência para rejeitar H_0

no exemplo “Lady Tasting Tea”, o p-valor era 1,4%

O p-valor traz informações quanto ao poder da evidência científica e permite com que o analista decida sobre o nível de risco que ele está propenso a correr

o nível de risco é o próprio nível de significância α

Testes de hipótese | rejeite, ou não, H_0

6. Rejeite, ou não, a hipótese H_0

$$\textbf{p-valor} \neq \boldsymbol{\alpha}$$

Calculado usando a amostra

Atribuído pelo analista

Os valores devem ser comparados para que a decisão de rejeitar, ou não, H_0 seja tomada. Em geral:

quando p-valor $\leq \alpha$, rejeita-se H_0

quando p-valor $> \alpha$, falha-se em rejeitar H_0

Os termos são importantes!

nunca se "aceita" H_0

Testes de hipótese | fundamentos

Observação:

Testes de hipótese frequentemente têm premissas que devem ser atendidas para que eles sejam aplicados sem perda de poder

independência dos elementos de uma amostra

independência entre amostras

distribuição de probabilidades da(s) amostra(s)

Para o próximo item, essas premissas serão relaxadas por motivos didáticos. Elas serão retomadas na parte 2 do assunto Testes de Hipótese

TESTES DE HIPÓTESE

testes para diferenças entre dois grupos

Testes de hipótese | diferenças entre dois grupos

Esses testes de hipótese são aplicados para avaliar a diferença entre duas amostras

Podem também ser aplicados em uma só amostra, bastando que ela seja dividida em duas subamostras

Usualmente, as diferenças são avaliadas em termos de medidas de tendência central e de dispersão

Testes de hipótese | diferenças entre dois grupos

Objetivo (H_0)	**Teste**	**Classe**	**Distribuição amostral**	**Estimador de diferença**
Não há diferenças entre as médias	Teste t para duas amostras	Paramétrico	Normal	Média
Não há diferenças entre as amostras	Wilcoxon (*rank-sum*)	Não paramétrico	-	Mediana
Não há diferenças nas dispersões das amostras	Teste F para duas amostras	Paramétrico	Normal	Variância
	Fligner-Killeen	Não paramétrico	-	Valor Absoluto dos Resíduos

Testes de hipótese | teste t

Teste t para duas amostras:

Premissa: requer normalidade

Hipóteses:

H_0: $\mu_1 = \mu_2$ | as médias dos grupos 1 e 2 são idênticas

H_A: $\mu_1 \neq \mu_2$ | as médias dos grupos 1 e 2 são diferentes (bilateral)

H_A: $\mu_1 > \mu_2$ | a média do grupo 1 é maior do que a média do grupo 2 (unilateral)

H_A: $\mu_1 < \mu_2$ | a média do grupo 2 é maior do que a média do grupo 1 (unilateral)

Testes de hipótese | teste t

Estatística do teste:

$$t_0 = \frac{\bar{X}_1 - \bar{X}_2}{\sqrt{\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}}}$$

onde

$\bar{X}_1, \bar{X}_2$	médias das amostras 1 e 2
s_1^2, s_2^2	variâncias das amostras 1 e 2
n_1, n_2	tamanho das amostras 1 e 2

Testes de hipótese | teste t

Rejeita-se H_0, se:

$|t_0| > t_{\alpha/2,\upsilon}$ para teste bilateral
$|t_0| > t_{\alpha,\upsilon}$ para teste unilateral

onde

$t_{\alpha/2,\upsilon}, t_{\alpha,\upsilon}$ obtidos da distribuição t-Student com significância α e υ graus de liberdade:

$$\upsilon = \frac{\left(\frac{s_1^2}{n_1} + \frac{s_2^2}{n_2}\right)^2}{\frac{s_1^2/n_1}{n_1 - 1} + \frac{s_2^2/n_2}{n_2 - 1}}$$

Testes de hipótese | teste t

Nota: a versão original do teste t possuía uma limitação adicional de que as amostras precisavam ter variâncias iguais. Isso foi ajustado posteriormente a partir da equação do número de graus de liberdade mostrada (conhecida como correção de Welch)

No R:

```
t.test(amostra1, amostra2, alternative = “two.sided”, var.equal = FALSE)
```

Para o caso unilateral em que se testa $\mu_1 > \mu_2$:

```
t.test(amostra1, amostra2, alternative = “greater”, var.equal = FALSE)
```

Para o caso unilateral em que se testa $\mu_1 < \mu_2$:

```
t.test(amostra1, amostra2, alternative = “less”, var.equal = FALSE)
```

Testes de hipótese | teste de Wilcoxon

Teste de Wilcoxon (rank-sum ou Mann-Whitney):

Sem premissas quanto a distribuições

Hipóteses:

H_0: $\tilde{\mu}_1 = \tilde{\mu}_2$ | as medianas dos grupos 1 e 2 são idênticas

H_A: $\tilde{\mu}_1 \neq \tilde{\mu}_2$ | as medianas dos grupos 1 e 2 são diferentes (bilateral)

H_A: $\tilde{\mu}_1 > \tilde{\mu}_2$ | a mediana do grupo 1 é maior do que a mediana do grupo 2 (unilateral)

H_A: $\tilde{\mu}_1 < \tilde{\mu}_2$ | a mediana do grupo 2 é maior do que a mediana do grupo 1 (unilateral)

Testes de hipótese | teste de Wilcoxon

Procedimento para duas amostras de tamanhos n_1 e n_2, sendo $n_1 \leq n_2$:

1. Ranquear conjuntamente os elementos das amostras, resultando em $r_k = 1,2, \dots, (n_1 + n_2)$. Identificar as amostras de origem para cada ranque
2. Somar os índices amostra de menor tamanho. Para amostras de tamanhos iguais, somar os índices de uma das amostras (não importa qual)

A estatística do teste é a própria soma do passo 2:

$$W = \sum_{i=1}^{n_1} r_i$$

Testes de hipótese | teste de Wilcoxon

Procedimento para subamostras retiradas de uma amostra:

1. Dividir a amostra em duas subamostras de tamanhos n_1 e n_2 ($n_1 \leq n_2$)
2. Ordenar os valores da amostra completa, atribuindo índices $(1, 2, \ldots, n_1 + n_2)$ e identificando as subamostras de origem
3. Somar os índices de cada amostra, obtendo-se m_1 e m_2

A estatística do teste é dada por:

$$W = \min(m_1, m_2)$$

Testes de hipótese | teste de Wilcoxon

Estatística do teste:

$$z = \frac{W - n_1(n_1 + n_2 + 1)/2}{\sqrt{n_1 n_2 (n_1 + n_2 + 1)/12}}$$

Rejeita-se H_0, se:

$|z| > z_{\alpha/2}$ para teste bilateral

$|z| > z_{\alpha}$ para teste unilateral

onde

$z_{\alpha/2}, z_{\alpha}$ obtidos da distribuição normal padrão

Testes de hipótese | teste de Wilcoxon

No R:

```
wilcox.test(amostra1, amostra2, exact = FALSE, alternative = "two.sided‟)
```

Para o caso unilateral em que se testa $\tilde{\mu}_1 > \tilde{\mu}_2$:

```
wilcox.test(amostra1, amostra2, exact = FALSE, alternative = "greater")
```

Para o caso unilateral em que se testa $\tilde{\mu}_1 < \tilde{\mu}_2$:

```
wilcox.test(amostra1, amostra2, exact = FALSE, alternative = ‟less")
```

Testes de hipótese | teste F

Teste F para duas amostras:

Premissa: requer normalidade

Hipóteses:

H_0: $\sigma_1^2 = \sigma_2^2$ | as variâncias dos grupos 1 e 2 são idênticas

H_A: $\sigma_1^2 \neq \sigma_2^2$ | as variâncias dos grupos 1 e 2 são diferentes (bilateral)

H_A: $\sigma_1^2 > \sigma_2^2$ | a variância do grupo 1 é maior do que a variância do grupo 2 (unil.)

H_A: $\sigma_1^2 < \sigma_2^2$ | a variância do grupo 2 é maior do que a variância do grupo 1 (unil.)

Testes de hipótese | teste F

Estatística do teste:

$$F = \frac{s_1^2}{s_2^2}$$

onde

s_1^2, s_2^2 variâncias das amostras 1 e 2

Testes de hipótese | teste F

Rejeita-se H_0, se:

$F < F_{1-\frac{\alpha}{2},n_1-1,n_1-2}$ e $F > F_{\frac{\alpha}{2},n_1-1,n_1-2}$ para teste bilateral

$F < F_{1-\alpha,n_1-1,n_1-2}$ e $F > F_{\alpha,n_1-1,n_1-2}$ para teste unilateral

onde

$F_{1-\frac{\alpha}{2},n_1-1,n_1-2}, F_{\frac{\alpha}{2},n_1-1,n_1-2}$ obtidos da distribuição F-Fisher-Snedecor com significância α

Testes de hipótese | teste F

A distribuição F-Snedecor-Fisher é assimétrica, o que justifica a consulta a dois valores teóricos para se chegar à regra de decisão.

No R:

```
var.test(amostra1, amostra2, alternative = “two.sided”)
var.test(amostra1, amostra2, alternative = “greater”)
var.test(amostra1, amostra2, alternative = “less”)
```

Testes de hipótese | teste de Fligner-Killeen

Teste de Fligner-Killeen:

Sem premissas quanto a distribuições

Hipóteses:

H_0: $\sigma_1^2 = \sigma_2^2 = \cdots = \sigma_k^2$ | as variâncias das k amostras são idênticas

H_A: $\sigma_i^2 \neq \sigma_1^2 = \cdots = \sigma_k^2$ | a variância de ao menos 1 amostra é diferente

Testes de hipótese | teste de Fligner-Killeen

A base do teste está na determinação do valor absoluto dos resíduos (AVR) resultantes dos desvios das observações de uma amostra em relação à sua mediana

Sendo $j = 1, \dots, k$ amostras e $i = 1, \dots, n_k$, calcula-se:

$$AVR_{ij} = \left| x_{ij} - \tilde{x}_j \right|$$

onde

x_{ij} elemento i da amostra j

$\tilde{x}_j$ mediana da amostra j

Testes de hipótese | teste de Fligner-Killeen

O teste, então, ranqueia os AVR_{ij} e aplica pesos para cada ranque

Na sequência, testa se o valor médio do resultado é estatisticamente igual para todos os grupos

caso não seja, possivelmente haja diferença na dispersão entre as amostras

Testes de hipótese | teste de Fligner-Killeen

No R:

```
fligner.test(amostras, identificador)
```

O `identificador` é um vetor que identifica os grupos no vetor `amostras`

Alternativamente:

```
fligner.test(amostras ~ grupos, data.frame)
```

O símbolo ~ é usado para mapear os `grupos` de cada `amostra` contida no `data.frame`

TESTES DE HIPÓTESE

aplicações em R

Testes de hipótese | aplicações em R

No início da aula, foram levantados dois questionamentos:

1. As vazões médias anuais dos últimos 30 anos em Foz do Areia estão diferentes em relação aos valores observados anteriormente?
2. As DBOs medidas nos os postos IG3 e IG4 são diferentes?

Essas respostas serão obtidas por meio da aplicação dos testes mostrados utilizando o R (ver código fornecido a partir da linha 107)

Revisão

Testes de hipótese são regras para rejeitar, ou não, uma hipótese estatística sobre uma amostra

Cada teste possui objetivos e premissas que as amostras devem atender

atentar para as premissas dos testes paramétricos

O nível de significância α é escolhido pelo analista e representa o risco que ele está disposto a correr em rejeitar H_0, quando ela é verdadeira

O p-valor é obtido dos dados e representa a probabilidade de confirmar a hipótese nula

A decisão é tomada comparando-se o p-valor com α

UFPR
UNIVERSIDADE FEDERAL DO PARANÁ

PPGERHA

Estatística Aplicada a Ciências Ambientais

Daniel Detzel
detzel@ufpr.br