



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Métodos não paramétricos

Ou: o que fazer caso seus dados não sejam normais?

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

Sumário

- 1 Discussão da aula passada
 - Discussão da aula passada
- 2 Normalidade
 - Visualização
 - Testes contra a normalidade
- 3 Transformações
 - Transformações
- 4 Métodos não paramétricos
 - Introdução
 - Teste para 1 amostra
 - Testes para 2 amostras
 - Teste para 3 ou mais amostras
 - Correlação
- 5 Resumo
- 6 Aprofundamento
 - Aprofundamento



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Discussão da aula passada



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Discussão da leitura obrigatória da aula passada

A hipótese da normalidade



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

- Todos os métodos que vimos até aqui presumem que os dados são normalmente distribuídos
- Desvios da normalidade precisam ser contornados¹
- Veremos duas maneiras: transformações e alternativas

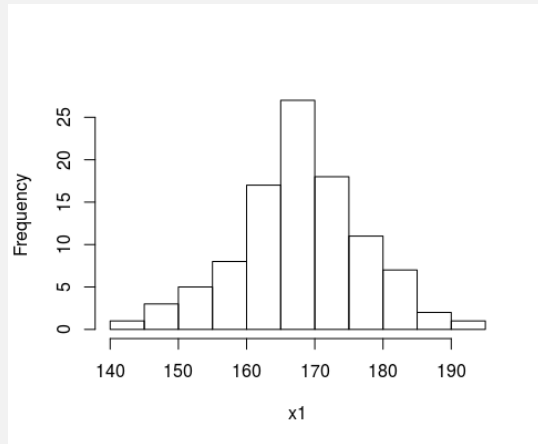
Mas antes...

... como identificar essa necessidade?

¹há controvérsias:

<https://www.r-bloggers.com/normality-tests-don't-do-what-you-think-they-do/>

Visualização - Histograma



Dados normais

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

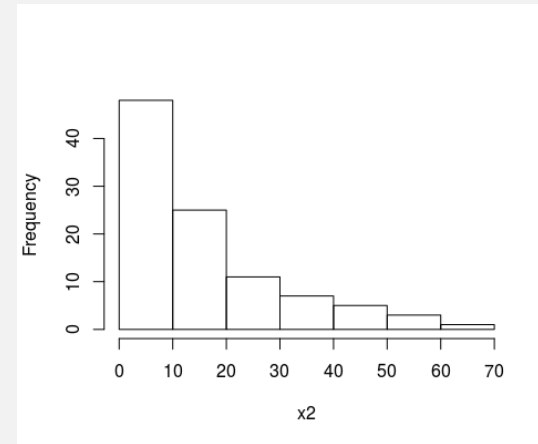
Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Visualização - Histograma



Dados não normais

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

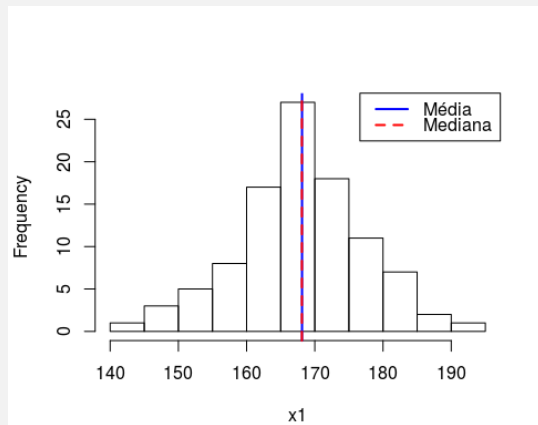
Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Visualização - Histograma



Dados normais

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

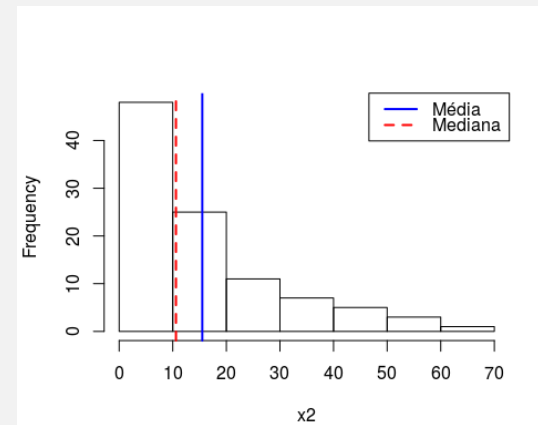
Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Visualização - Histograma



Dados não normais

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

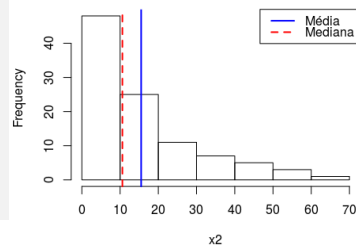
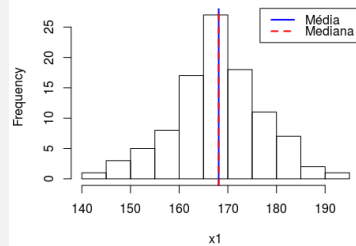
Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Visualização - Histograma



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

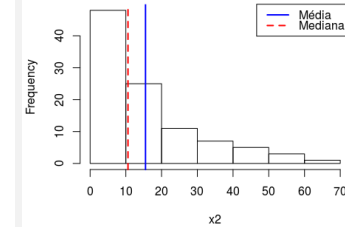
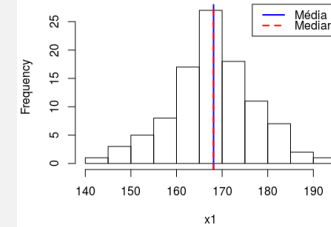
Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Visualização - boxplot



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

O Q-Q plot



- Gráfico que compara os quantis da amostra com os quantis teóricos
- Adicionalmente uma reta “ideal” é sobreposta, como referência
- Dados normalmente distribuídos ficam próximos da reta

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

Transformações

Métodos não paramétricos

Resumo

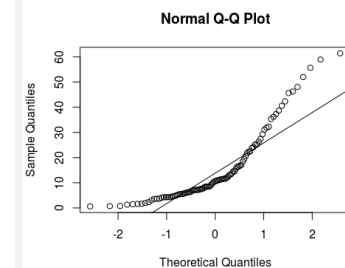
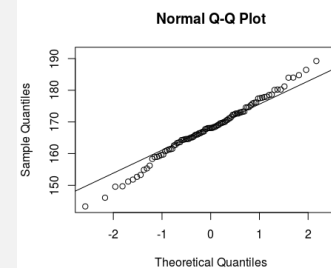
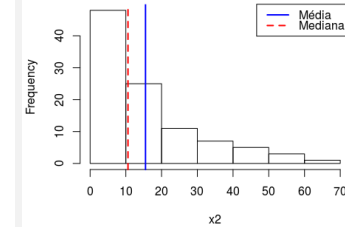
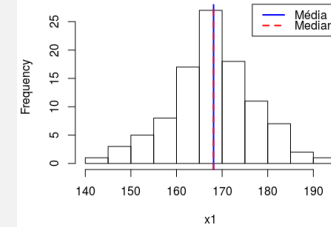
Aprofundamento

Princípio

Quanto maior o desvio da normalidade...

... maior a distância à reta

Visualização - QQ plot



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Visualização

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

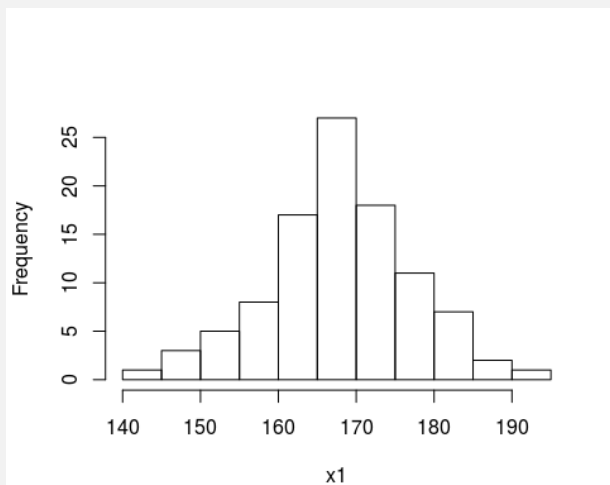
- Objetivo: é possível **determinar** se uma amostra veio de uma população normalmente distribuída?
- Resposta curta: **NÃO**.
- Resposta longa: podemos examinar se há evidências para “aceitar” esta hipótese²

²Lembre: **nunca** aceitamos uma hipótese – apenas deixamos de rejeitar H_0 .

Alguns testes contra a normalidade

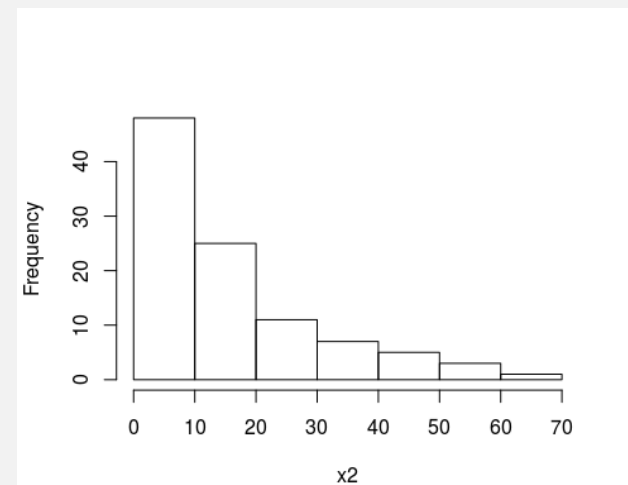
- **Shapiro-Wilk**
- Anderson-Darling
- Kolmogorov-Smirnov

Shapiro-Wilk – Rejeitamos a H_0 de normalidade?



p-value = 0.7766

Shapiro-Wilk – Rejeitamos a H_0 de normalidade?



p-value = 1.657e-09

Transformações



- Podemos aplicar uma transformação nos dados, para coagi-los a se aproximar das premissas requeridas
- Transformações usuais incluem:
 - logaritmo
 - exponencial
 - raiz quadrada
 - potências
- Geralmente envolve tentativa e erro ³
- **Hipóteses sobre o problema ou desenho experimental ajudam**

³Mas a transformação de Box-Cox pode ajudar!

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Exemplo



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

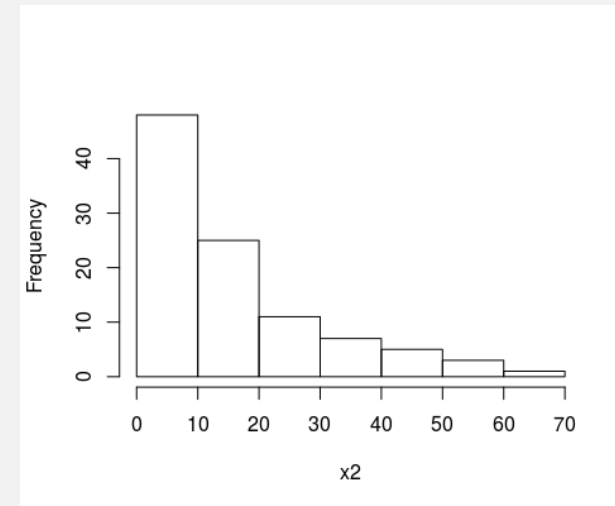
Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento



Transformação sugerida: logaritmo.

Exemplo



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

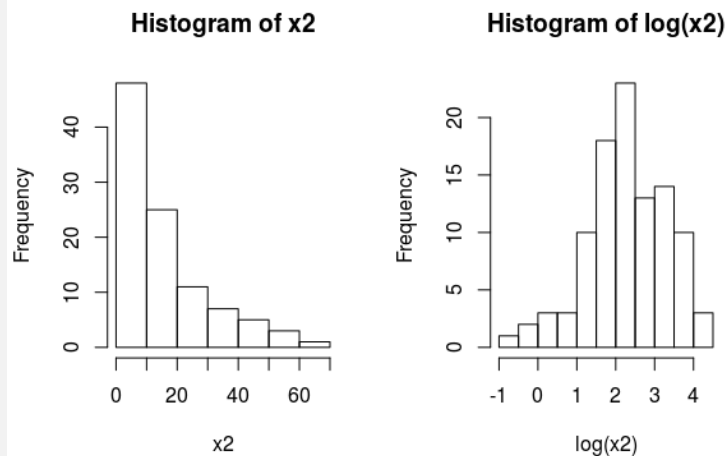
Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento



Dados normais x dados log-transformados

Exemplo



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

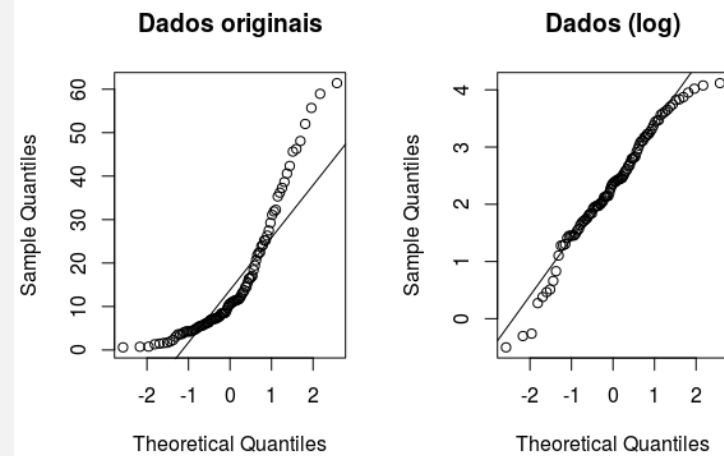
Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento



(p-valor S-W: 1.657e-09) x (p-valor S-W: 0.05032)

Introdução aos métodos não paramétricos



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Métodos paramétricos

usam a distribuição dos dados^a...

...para possibilitar cálculos simples como média e DP.

^aGeralmente distribuição Normal

Métodos não paramétricos

Não presumem nada sobre a distribuição dos dados.

Introdução aos métodos não paramétricos



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Sem média e DP⁴, a única coisa que resta para comparar...

... é a **ordem** dos dados (*ranks*).

⁴tendência central e dispersão, respectivamente

Teste para 1 amostra



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

- Desvios da normalidade severos impactam os testes paramétricos

- Nesses casos, tenta-se transformar os dados, se possível
- Caso não seja, deve-se usar um teste não paramétrico⁵

Teste para uma amostra

Ao invés do teste t, usar o teste de Wilcoxon (Capítulo 25)

⁵Sem transformação!

Quais são as variáveis?



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

- Dependente:
 - categórica ordinal
 - numérica discreta
 - numérica contínua (não-normal)
- Independente: parâmetro fixo

Exemplo

escore HHS mediano ~ 70

Exemplo

escore ASA mediano ~ II

Testes para 2 amostras



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Dados normais

- amostras independentes \Rightarrow t-teste não pareado
- amostras pareadas \Rightarrow t-teste pareado

Dados não normais

- amostras independentes \Rightarrow **Mann-Whitney** (Capítulo 24)
- amostras pareadas \Rightarrow Wilcoxon (Capítulo 25)

Quais são as variáveis?



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

- Dependente:
 - categórica ordinal
 - numérica discreta
 - numérica contínua (não-normal)
- Independente:
 - categórica ordinal
 - numérica discreta
 - numérica contínua (não-normal)

Esta relação pode ser expressa como

escore HHS tratamento \sim escore HHS controle

Em termos práticos...



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

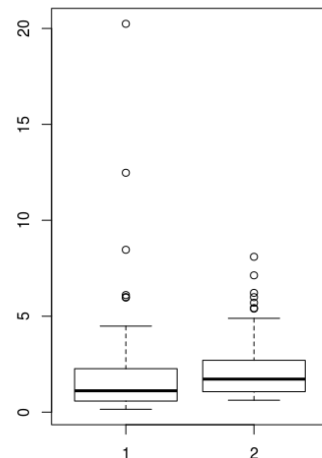
3+ amostras

Correlação

Resumo

Aprofundamento

P: Estas amostras são significativamente diferentes?



Exemplo



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

- Assumindo⁶ que elas são
 - **normalmente distribuídas**, e
 - independentes,poderíamos fazer um teste t não pareado.

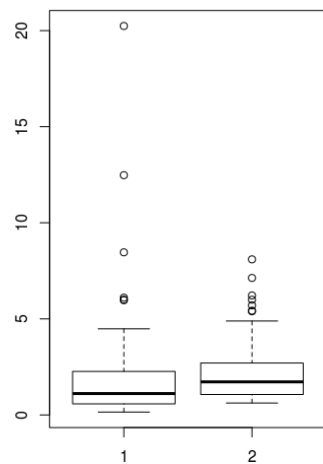
- Resultado: p-valor = **0.259**

Pergunta

Isto significa que as amostras não são significativamente diferentes?

⁶pelo desenho experimental

Novamente...



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

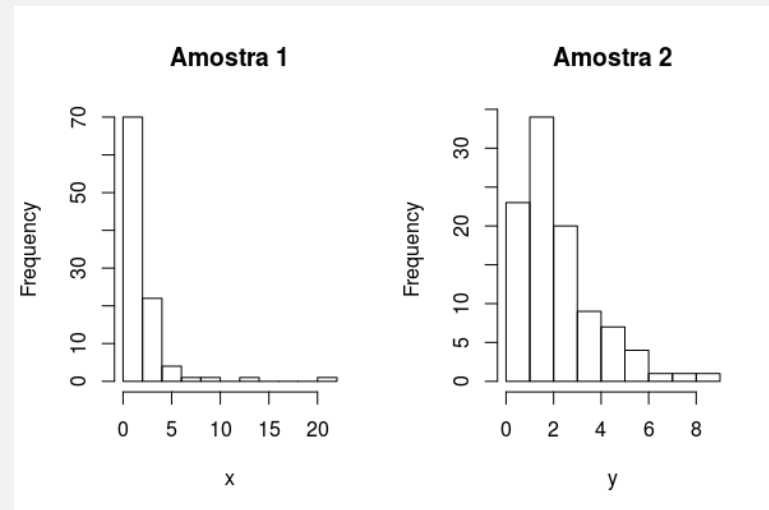
3+ amostras

Correlação

Resumo

Aprofundamento

Histogramas



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

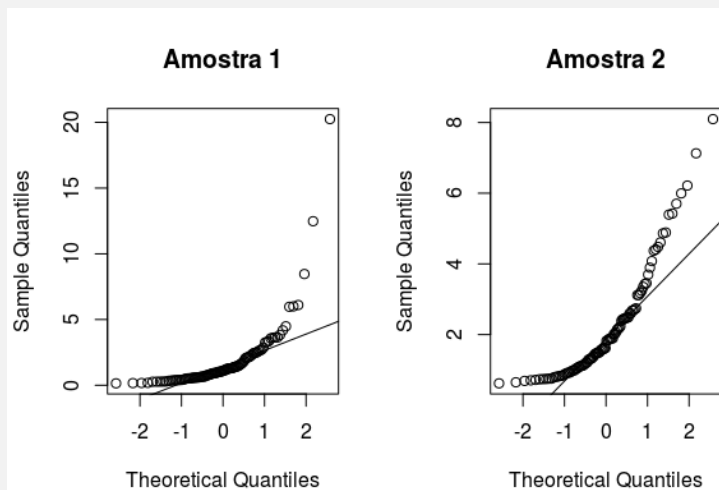
3+ amostras

Correlação

Resumo

Aprofundamento

QQ-plots



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Mann-Whitney



Teste t

p-valor = 0.259 (não significativo)

- Aplicando o teste de Shapiro-Wilk em x e y
 - x: p-valor = 5.515e-16
 - y: p-valor = 5.274e-09

- Devemos rejeitar a hipótese de normalidade.
- Então o teste t **não é** apropriado!
- Substituto: teste de Mann-Whitney

Teste de Mann-Whitney

p-value = 0.0001346 (significativo)

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Relembrando



- Para testar a diferença nas médias de 3 ou mais amostras
 - Análise de Variâncias (ANOVA)
 - Leva em conta as variâncias entre os grupos (**inter**)
 - Leva em conta a variância em cada grupo (**intra**)
 - H_0 : Todos os grupos são =
 - H_1 : pelo menos um grupo é significativamente \neq

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

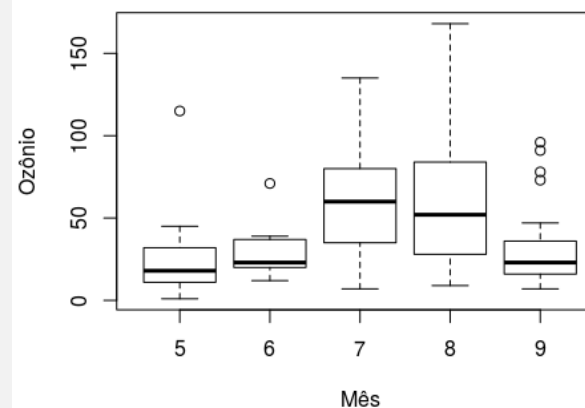
Aprofundamento

Em termos práticos...



P: Estas amostras são significativamente diferentes?

Medições de qualidade do ar em NY



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Quais são as variáveis?



- Dependente:
 - categórica ordinal
 - numérica discreta
 - numérica contínua (não-normal)
- Independente:
 - grupo (categórica nominal – 3+ níveis)

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

Aprofundamento

Esta relação pode ser expressa como

Ozônio ~ Mês

Kruskal-Wallis



ANOVA

p-value = 0.0776 (não significativo)

- Shapiro-Wilk (Ozônio por mês (Maio – Setembro):
< 0.0001, 0.0628, 0.86689, 0.090325, < 0.0001
- Devemos rejeitar a hipótese de normalidade.
- Então o ANOVA **não é** apropriado!
- Substituto: teste de Kruskal-Wallis (Capítulo 30)

Teste de Kruskal-Wallis

p-value = 6.901e-06 (significativo)

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Intro

1 amostra

2 médias

3+ amostras

Correlação

Resumo

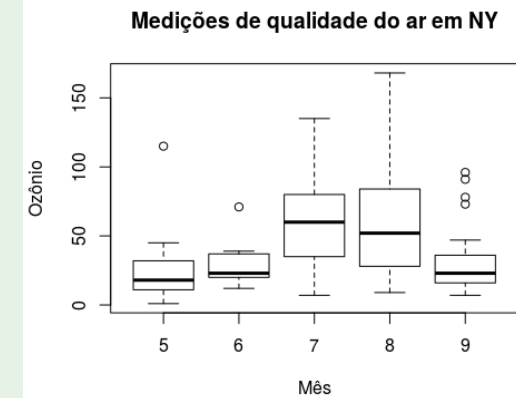
Aprofundamento

Mais quais são os meses diferentes?

Pós-teste de Wilcoxon

Mês x Mês (correção de Bonferroni)

- 5 x 6: $p = 1.0000$
- 5 x 7: $p = 0.0003$
- 5 x 8: $p = 0.0012$
- 5 x 9: $p = 1.0000$
- 6 x 7: $p = 0.1414$
- 6 x 8: $p = 0.2591$
- 6 x 9: $p = 1.0000$
- 7 x 8: $p = 1.0000$
- 7 x 9: $p = 0.0074$
- 8 x 9: $p = 0.0325$



Correlação não-paramétrica

- A correlação de Pearson
 - associa dados numéricos (contínuos);
 - mede a direção e força desta associação.

Correlação de Spearman

Ao invés da correlação linear de Pearson...

... usar a correlação de ranks de Spearman (Capítulo 17).

Número de resultados no PUBMED⁷

- t-test: 61488
- ANOVA: 431252
- Wilcoxon: 19881
- Mann-Whitney: 25571
- Kruskal-Wallis: 11943
- Shapiro-Wilk: 519
- Kolmogorov-Smirnov: 0
- Anderson-Darling: 49
- Chi-square: 107277
- OR: 221034
- RR: 344996

⁷ Levantamento feito em 2017-11-30

Resumo (teste oftálmico)



Table 37.1. Selecting a Statistical Test

Goal	Type of Data			
	Measurement (from Gaussian Population)	Rank, Score, or Measurement (from Non-Gaussian Population)	Binomial (Two Possible Outcomes)	Survival Time
Describe one group	Mean, SD	Median, interquartile range	Proportion	Kaplan Meier survival curve
Compare one group to a hypothetical value	One-sample t test	Wilcoxon test	Chi-square or Binomial test**	—
Compare two unpaired groups	Unpaired t test	Mann-Whitney test	Fisher's test (chi-square for large samples)	Log-rank test or Mantel-Haenszel*
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test	Conditional proportional hazards regression**
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Cox proportional hazard regression*
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test	Cochrane Q**	Conditional proportional hazards regression**
Quantify association between two variables	Pearson correlation	Spearman correlation	Contingency coefficients**	—
Predict value from another measured variable	Simple linear regression or Nonlinear regression	Nonparametric regression**	Simple logistic regression*	Cox proportional hazard regression*
Predict value from several measured or binomial variables	Multiple linear regression* or Multiple nonlinear regression**	—	Multiple logistic regression*	Cox proportional hazard regression*

*Only briefly mentioned in this book.

**Not discussed in this book.

Resumo (agora sim)



Goal	Measurement (from Gaussian Population)	Rank, Score, or Measurement (from Non-Gaussian Population)
Describe one group	Mean, SD	Median, interquartile range
Compare one group to a hypothetical value	One-sample t test	Wilcoxon test
Compare two unpaired groups	Unpaired t test	Mann-Whitney test
Compare two paired groups	Paired t test	Wilcoxon test
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test
Compare three or more matched groups	Repeated-measures ANOVA	Friedman test
Quantify association between two variables	Pearson correlation	Spearman correlation

Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Aprofundamento



Métodos não paramétricos

Felipe Figueiredo

Discussão da aula passada

Normalidade

Transformações

Métodos não paramétricos

Resumo

Aprofundamento

Aprofundamento

Leitura obrigatória

- Capítulo 37
- Capítulo 38

Exercícios selecionados

Não há.

Leitura recomendada

- **Parte VI – Designing Clinical Trials**
- Trechos de testes não paramétricos que pulamos dos caps:
 - 17
 - 24
 - 25
 - 30