

Métodos não-paramétricos

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

Felipe Figueiredo

Instituto Nacional de Traumatologia e Ortopedia

Sumário

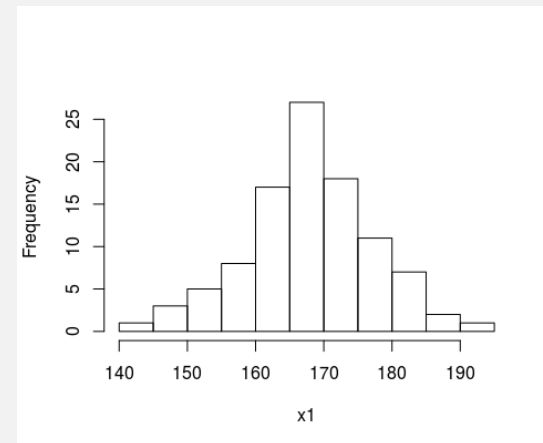
- 1 Normalidade
 - Visualização
 - Testes contra a normalidade
- 2 Transformações
 - Transformações
 - Exemplo
- 3 Métodos não-paramétricos
 - Teste para 1 amostra
 - Testes para 2 amostras
 - Teste para 3 ou mais amostras
 - Correlação
- 4 Resumo

A hipótese da normalidade

- 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?

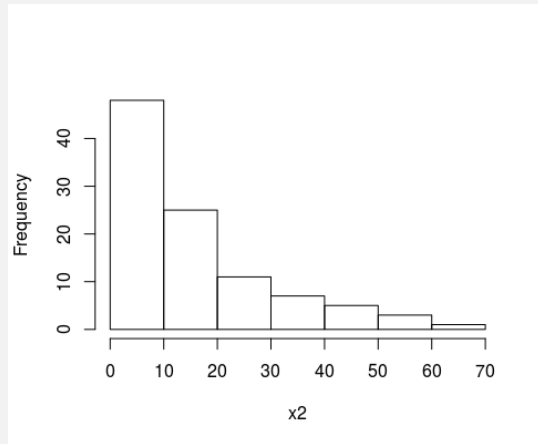
¹há controvérsias: <https://www.r-bloggers.com/normality-tests-don't-do-what-you-think-they-do/>

Visualização - Histograma



Dados normais

Visualização - Histograma



Dados não-normais

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

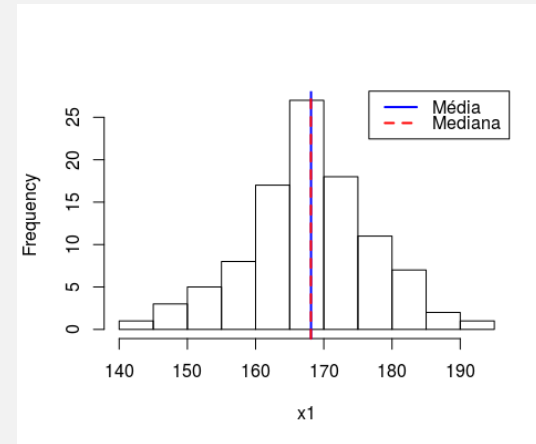
Normalidade

Transformações

Métodos não-paramétricos

Resumo

Visualização - Histograma



Dados normais

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

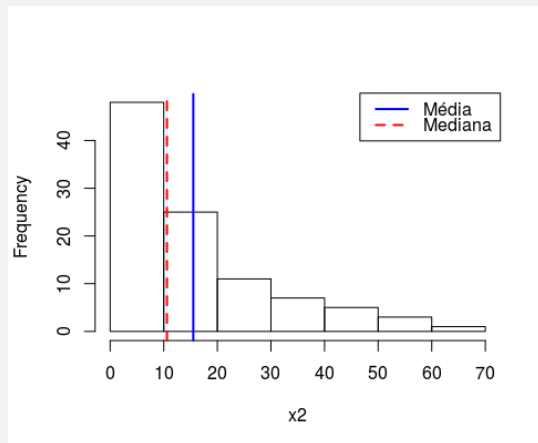
Normalidade

Transformações

Métodos não-paramétricos

Resumo

Visualização - Histograma



Dados não-normais

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

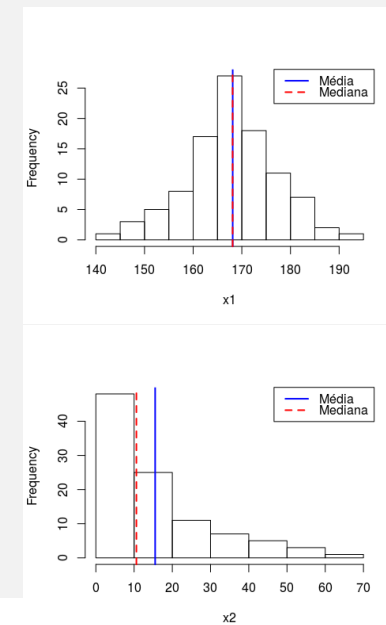
Normalidade

Transformações

Métodos não-paramétricos

Resumo

Visualização - Histograma



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

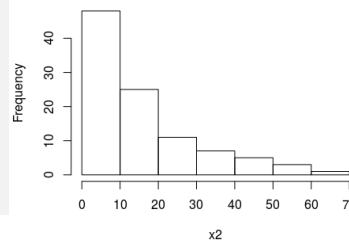
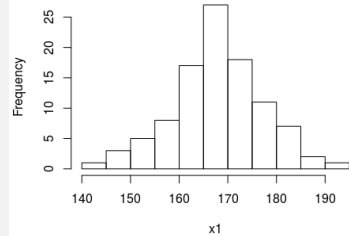
Normalidade

Transformações

Métodos não-paramétricos

Resumo

Visualização - boxplot



Métodos não-paramétricos

Felipe Figueiredo

Normalidade
Visualização
Normalidade

Transformações

Métodos não-paramétricos

Resumo

O Q-Q plot



Métodos não-paramétricos

Felipe Figueiredo

Normalidade
Visualização
Normalidade

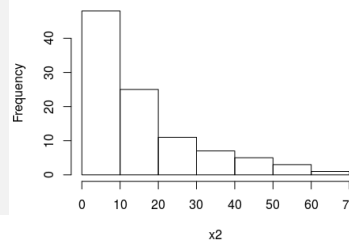
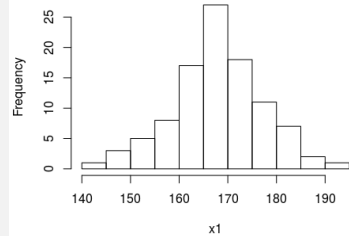
Transformações

Métodos não-paramétricos

Resumo

- 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
- Quanto maior o desvio da normalidade, maior a distância à reta

Visualização - QQ plot



Métodos não-paramétricos

Felipe Figueiredo

Normalidade
Visualização
Normalidade

Transformações

Métodos não-paramétricos

Resumo



Métodos não-paramétricos

Felipe Figueiredo

Normalidade
Visualização
Normalidade

Transformações

Métodos não-paramétricos

Resumo

- 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 que **nunca** aceitamos uma hipótese – apenas deixamos de rejeitar sua recíproca.

Alguns testes de normalidade



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

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

Normalidade

Transformações

Métodos não-paramétricos

Resumo

Shapiro-Wilk



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

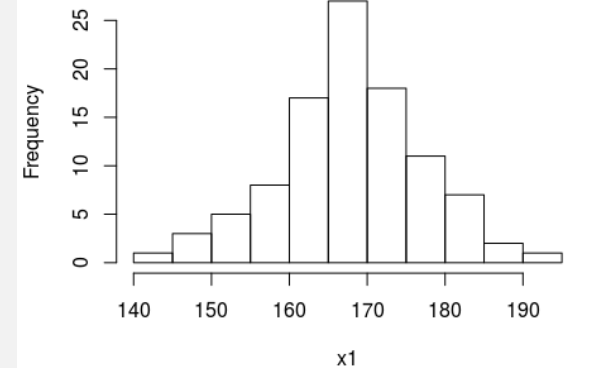
Visualização

Normalidade

Transformações

Métodos não-paramétricos

Resumo



p-value = 0.7766

Shapiro-Wilk



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

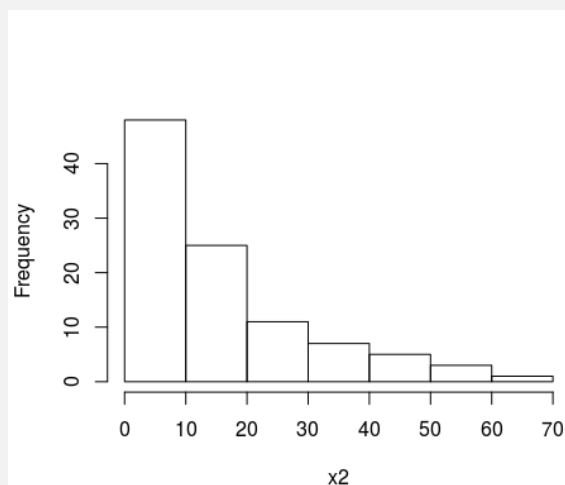
Visualização

Normalidade

Transformações

Métodos não-paramétricos

Resumo



p-value = 1.657e-09

Transformações



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Exemplo

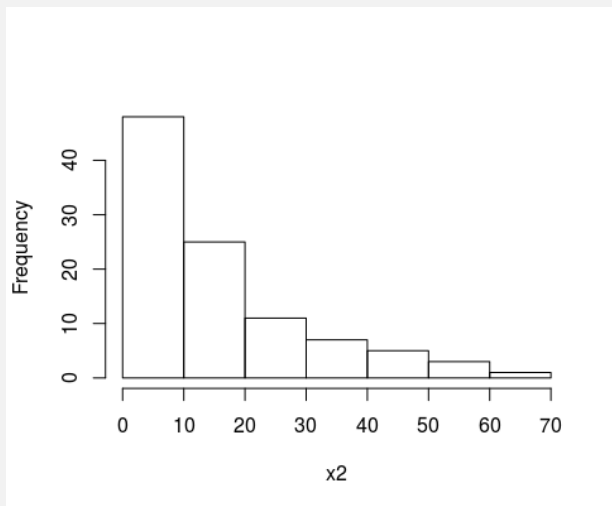
Métodos não-paramétricos

Resumo

- Algumas vezes, podemos aplicar uma transformação nos dados, para que eles se adequem às premissas requeridas
- Transformações comuns incluem:
 - logaritmo
 - exponencial
 - raiz quadrada
 - potências
- Geralmente envolve tentativa e erro³
- Hipóteses sobre o problema ou desenho experimental ajudam

³Mas a transformação de Box-Cox pode ajudar!

Exemplo



Transformação sugerida: logaritmo.

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

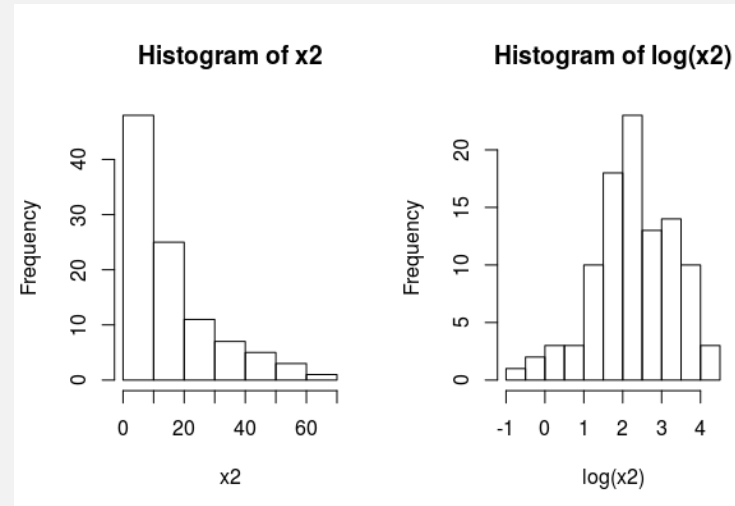
Transformações

Exemplo

Métodos não-paramétricos

Resumo

Exemplo



Dados normais x dados transformados (log)

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

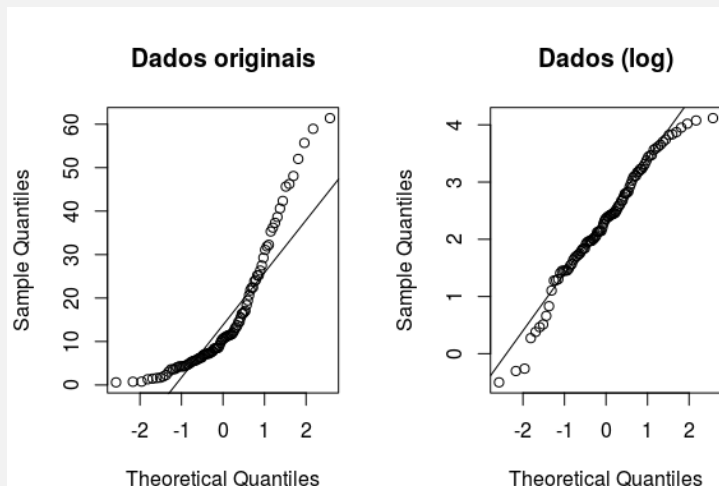
Transformações

Exemplo

Métodos não-paramétricos

Resumo

Exemplo



Dados normais (p-valor Shapiro-Wilk: 1.657e-09) x dados log-transformados (p-valor Shapiro-Wilk: 0.05032)

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Exemplo

Métodos não-paramétricos

Resumo

Teste para 1 amostra



- Desvios da normalidade severos impactam os testes paramétricos
- Nesses casos, deve-se transformar os dados, se possível
- Caso não seja, deve-se usar um teste não-paramétrico

Teste para uma amostra

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

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

Testes para 2 amostras



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

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)

Em termos práticos...



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

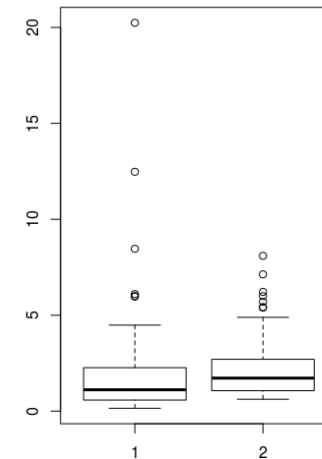
Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

P: Estas amostras são significativamente diferentes?



Exemplo



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

- Assumindo⁴ 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?

⁴pelo desenho experimental

Novamente...



Métodos não-paramétricos

Felipe Figueiredo

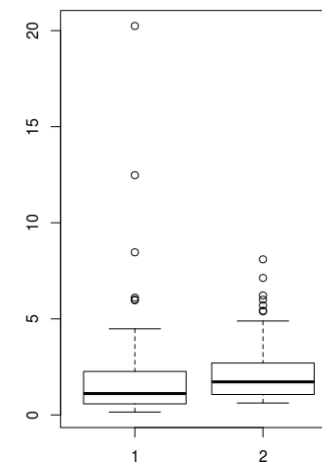
Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo



Histogramas



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

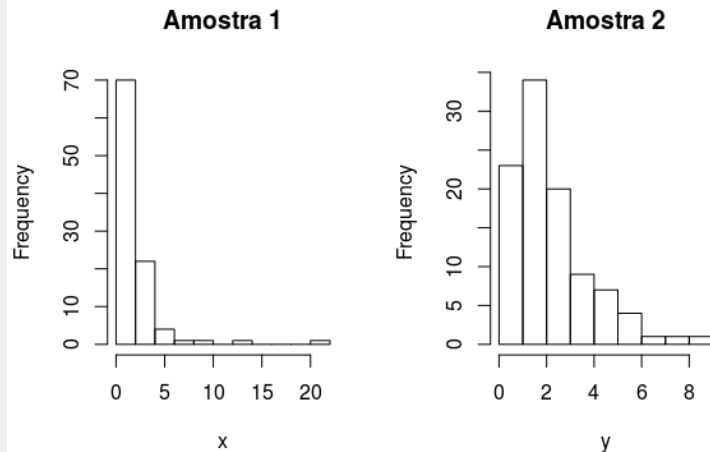
1 amostra

2 médias

3+ amostras

Correlação

Resumo



QQ-plots



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

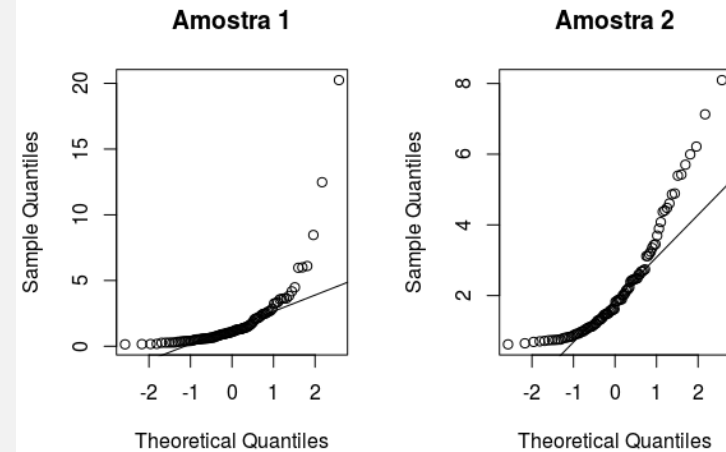
1 amostra

2 médias

3+ amostras

Correlação

Resumo



Mann-Whitney



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra

2 médias

3+ amostras

Correlação

Resumo

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)

Relembrando



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra

2 médias

3+ amostras

Correlação

Resumo

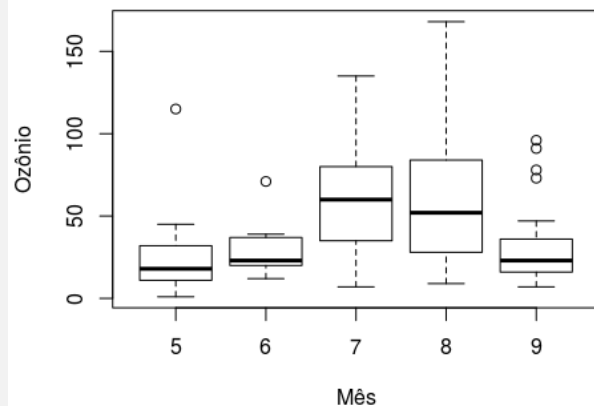
- Para testar se há diferença significativa em 3 ou mais amostras
 - Análise de Variâncias (ANOVA)
 - Leva em conta as variâncias entre os grupos (**inter**)
 - Leva em conta a variância em cada grupo (**intra**)
 - H_0 : Todos os grupos são =
 - H_1 : pelo menos um grupo é significativamente \neq

Em termos práticos...



P: Estas amostras são significativamente diferentes?

Medições de qualidade do ar em NY



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

Kruskal-Wallis



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

ANOVA

p-valor = 0.0776 (não significativo)

- Shapiro-Wilk (Ozônio por mês):
< 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)

Relembrando



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra
2 médias
3+ amostras
Correlação

Resumo

- A correlação de Pearson associa dados numéricos
- Mede a direção e força desta associação

Correlação

Ao invés da correlação linear de Pearson, usar a correlação de ranks de Spearman (Capítulo 17).

Número de resultados no PUBMED



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

Resumo

- 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

Resumo (teste oftálmico)



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

Resumo

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* Conditional proportional hazards regression**
Compare two paired groups	Paired t test	Wilcoxon test	McNemar's test	Cox proportional hazard regression*
Compare three or more unmatched groups	One-way ANOVA	Kruskal-Wallis test	Chi-square test	Conditional proportional hazards 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)



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

Resumo

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

Pós-aula



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

Resumo

Leitura obrigatória

- Capítulo 37
- Capítulo 38

Leitura recomendada

Seções de métodos não-paramétricos dos capítulos mencionados na aula.