

## 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

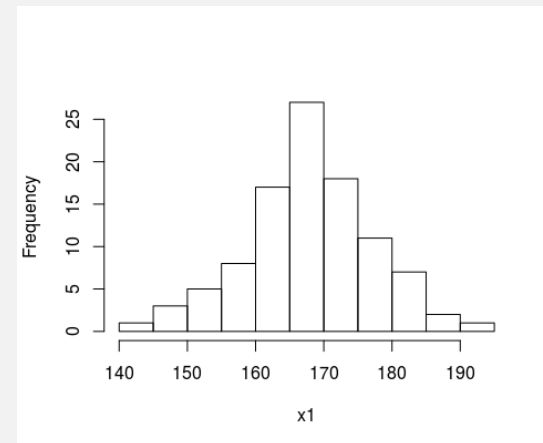
## A hipótese da normalidade

- Todos os métodos que vimos até aqui presumem que os dados são normalmente distribuídos
- Desvios da normalidade precisam ser contornados<sup>1</sup>
- Veremos duas maneiras: transformações e alternativas
- Mas antes, como identificar essa necessidade?

<sup>1</sup>há controvérsias:

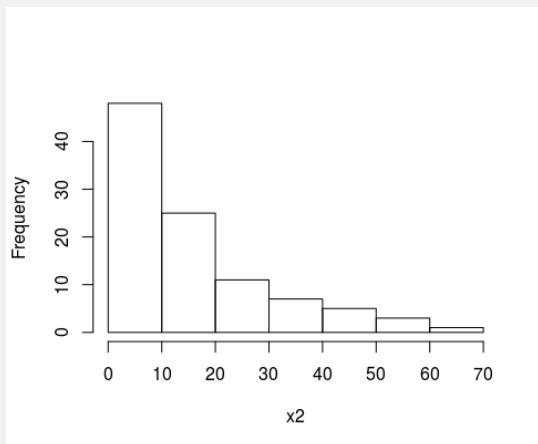
<http://www.r-bloggers.com/normality-tests-don%E2%80%99t-do-what-you-think-they-do/>

## Visualização - Histograma



Dados normais

## Visualização - Histograma



Dados não-normais

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

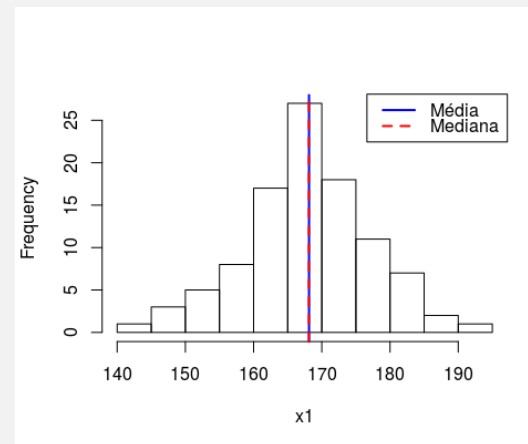
Normalidade

Transformações

Métodos não-paramétricos

Resumo

## Visualização - Histograma



Dados normais

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

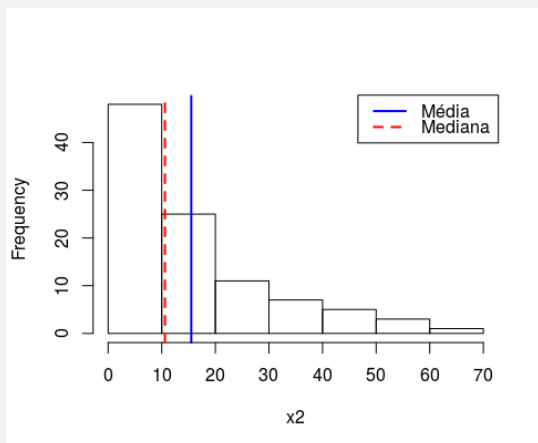
Normalidade

Transformações

Métodos não-paramétricos

Resumo

## Visualização - Histograma



Dados não-normais

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Visualização

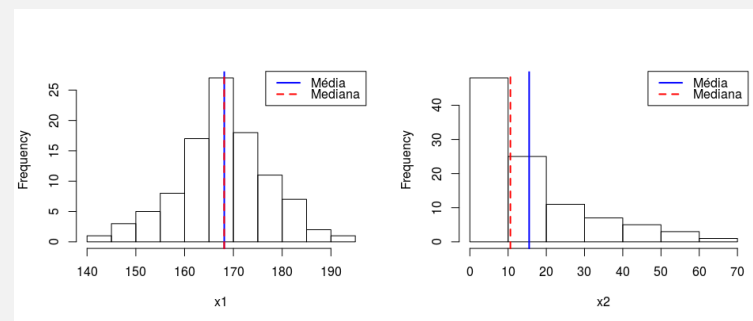
Normalidade

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

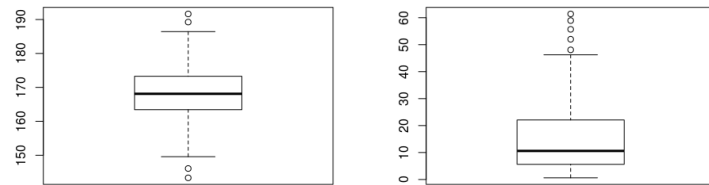
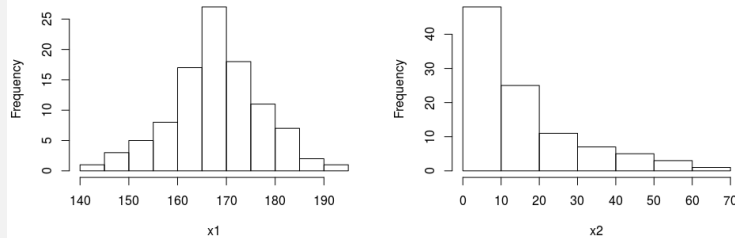
Felipe Figueiredo

Normalidade  
Visualização  
Normalidade

Transformações

Métodos não-paramétricos

Resumo



## Visualização - QQ plot



Métodos não-paramétricos

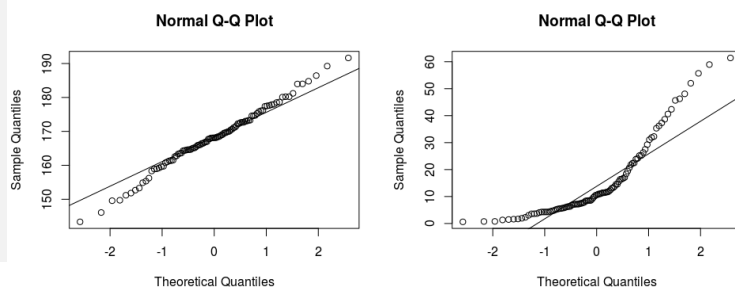
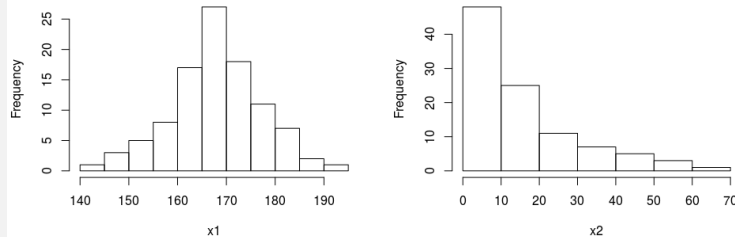
Felipe Figueiredo

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

- 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 hipótese<sup>2</sup>

<sup>2</sup>Lembre 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

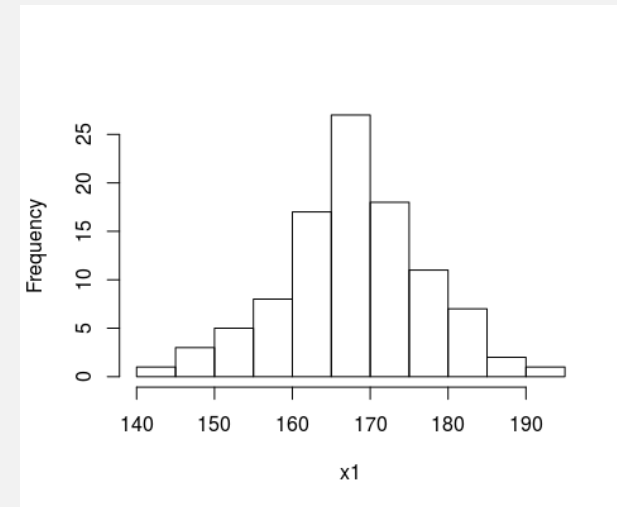
Visualização

Normalidade

Transformações

Métodos não-paramétricos

Resumo



p-value = 0.7766

## Shapiro-Wilk



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

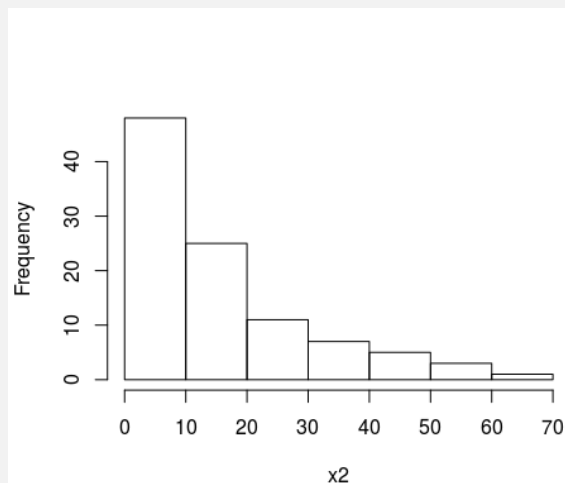
Visualização

Normalidade

Transformações

Métodos não-paramétricos

Resumo



p-value = 1.657e-09

## Transformações



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

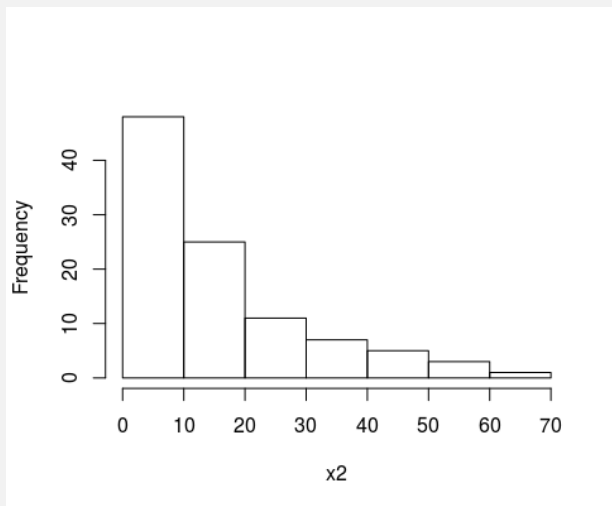
Exemplo

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 erro
- Hipóteses sobre o problema ou desenho experimental ajudam

## Exemplo



Transformação sugerida: logaritmo.

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

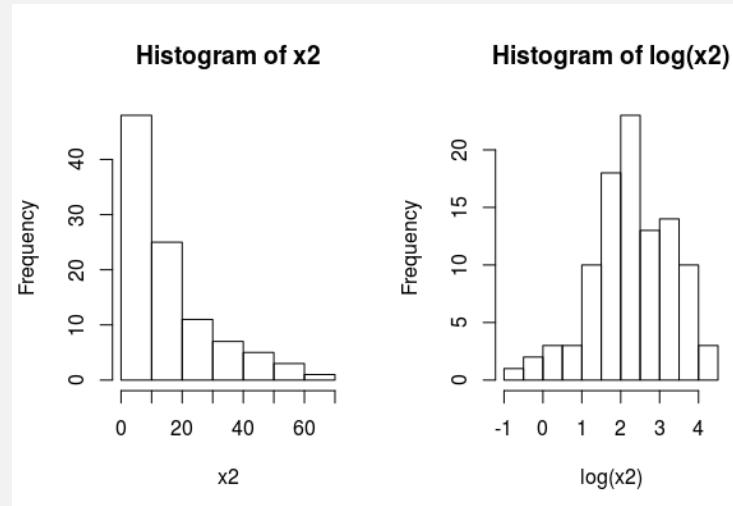
Transformações

Exemplo

Métodos não-paramétricos

Resumo

## Exemplo



Dados normais x dados transformados (log)

Métodos não-paramétricos

Felipe Figueiredo

Normalidade

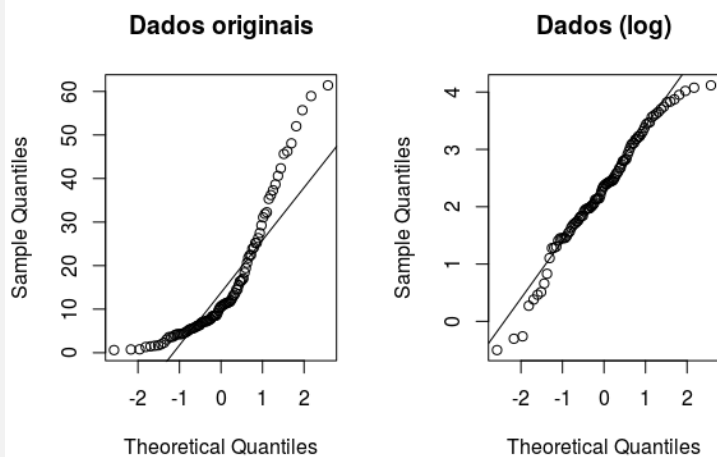
Transformações

Exemplo

Métodos não-paramétricos

Resumo

## Exemplo



Dados normais x dados transformados (log)

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

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<sup>a</sup>
- amostras pareadas  $\Rightarrow$  Wilcoxon (signed rank test)

<sup>a</sup>Também conhecido como Wilcoxon (rank sum test)

## Em termos práticos...



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

1 amostra

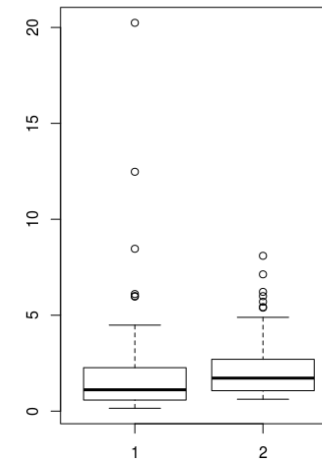
2 médias

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

- Assumindo<sup>3</sup> 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?

<sup>3</sup>pelo desenho experimental

## Novamente...



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

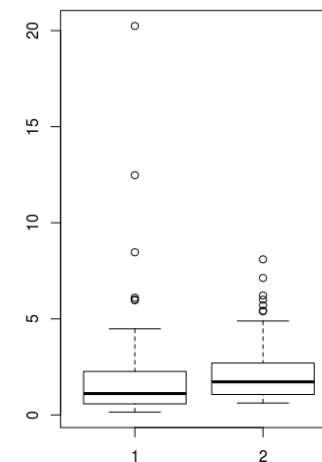
1 amostra

2 médias

3+ amostras

Correlação

Resumo



# Histogramas



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

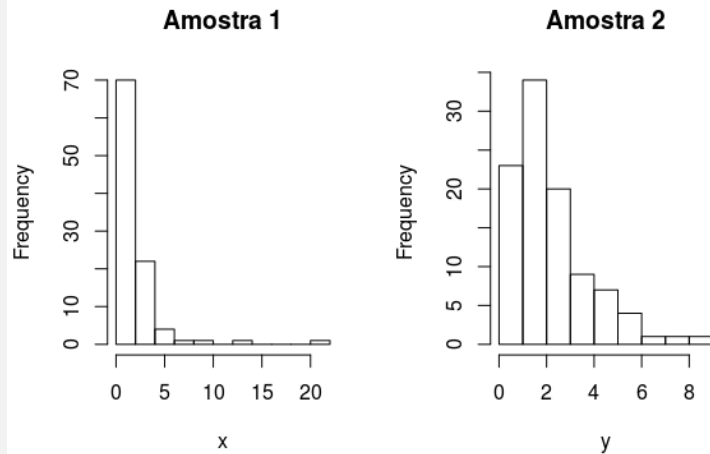
1 amostra

2 médias

3+ amostras

Correlação

Resumo



# QQ-plots



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

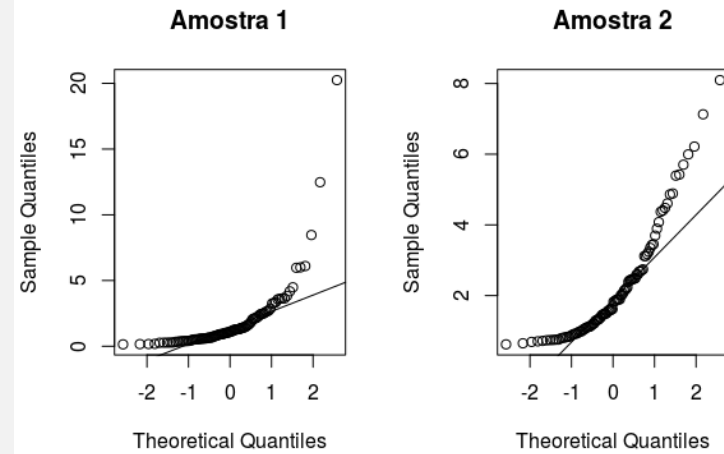
1 amostra

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

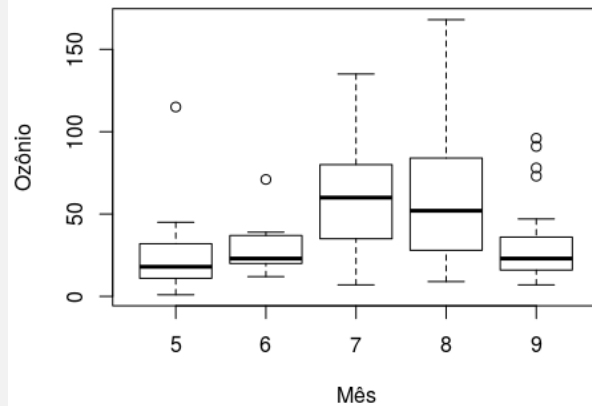
- Para testar se há diferença significativa em 3 ou mais amostras
  - Análise de Variâncias (ANOVA)
  - Leva em conta as variâncias entre os grupos (**inter**)
  - Leva em conta a variância em cada grupo (**intra**)
  - $H_0$  : Todos os grupos são =
  - $H_1$  : pelo menos um grupo é significativamente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

## 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): p-value = 2.79e-08
- Devemos rejeitar a hipótese de normalidade.
- Então o ANOVA **não é** apropriado!
- Substituto: teste de Kruskal-Wallis

### 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

## Resumo



Métodos não-paramétricos

Felipe Figueiredo

Normalidade

Transformações

Métodos não-paramétricos

Resumo

Paramétrico	Não-paramétrico
t-teste pareado	Wilcoxon
t-teste não-pareado	Mann-Whitney
ANOVA 1 fator	Kruskal-Wallis
Correlação de Pearson	Correlação de Spearman