

# PCC170 - Projeto e Análise de Experimentos Computacionais

Marco Antonio M. Carvalho

Departamento de Computação  
Instituto de Ciências Exatas e Biológicas  
Universidade Federal de Ouro Preto



## 1 Testes paramétricos

- Student's t-test
- Repeated Measures One-Way ANOVA
- Pairwise t-test
- One-Way ANOVA
- Tukey's HSD test

## Fonte

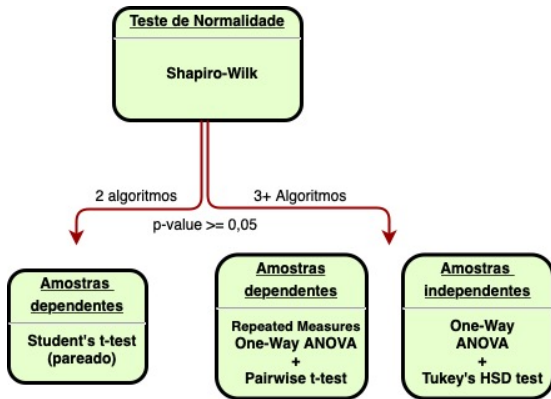
Este material é parcialmente baseado no conteúdo de

- ▶ Alboukadel Kassambara. *Statistical tools for high-throughput analysis*. 2022. Disponível em <https://bityli.com/ixKGg>
- ▶ Chi Yau. *R tutorial: An R introduction to statistics*. 2022. Disponível em <https://bityli.com/qSzEd>
- ▶ Zach Bobbitt. *How to Perform a Repeated Measures ANOVA in R*. 2020. Disponível em <https://bityli.com/hFwuC>

## Licença

Este material está licenciado sob a Creative Commons BY-NC-SA 4.0. Isto significa que o material pode ser compartilhado e adaptado, desde que seja atribuído o devido crédito, que o material não seja utilizado de forma comercial e que o material resultante seja distribuído de acordo com a mesma licença.

# Testes paramétricos



Fluxograma de escolha de testes paramétricos.

## Introdução

Muitos dos procedimentos estatísticos assumem que os dados são normalmente distribuídos.

Esses testes são chamados de testes **paramétricos**, pois sua validade depende da distribuição dos dados.

Um erro frequente é usar testes estatísticos que assumem uma distribuição normal em dados que são realmente assimétricos.

## Student's t-test

O *Student's t-test* é um método utilizado, entre outras coisas, para comparar duas amostras, no sentido de verificar se existem diferenças significativas entre os seus resultados.

As distribuições testadas devem ser **normais** e no caso de amostras pareadas, estas são consideradas **dependentes**.

A hipótese nula é a de que não há diferença entre as médias das populações emparelhadas.

# Testes paramétricos

## Como executar o teste

No *R*, crie uma série de dados para cada algoritmo (*Método A* e *Método B*) e atribua os valores de cada um.

O parâmetro *less* avalia se a média do primeiro método é menor do que a do segundo.

Em outras palavras, está sendo testado se os dados são significativamente diferentes e se os dados do seu algoritmo são melhores do que os dados do outro algoritmo, considerando um problema de minimização.

```
1 > MetodoA <- c(214, 159, 13, 356, 789, 123)
2 > MetodoB <- c(159, 135, 123, 543, 12, 345)
3 > t.test(MetodoA, MetodoB, paired=TRUE, alternative="less")
```



# Testes paramétricos

## Paired t-test

data: MetodoA and MetodoB

$t = 0.37177$ ,  $df = 5$ ,  $p\text{-value} = 0.6373$

alternative hypothesis: true difference in means  
is less than 0

95 percent confidence interval:

-Inf 360.6001

sample estimates:

mean of the differences  
56.16667


## Análise

Se o *p-value* for menor do que um valor crítico dado pelo nível de significância  $\alpha$  (normalmente 0,05), então o pressuposto de inexistência de diferença significativa é rejeitado no nível de significância  $\alpha$ .


Ou seja, há evidência de que os dados testados diferem entre si.

Caso contrário, não se pode afirmar que os dados não são idênticos.

Embora sejam reportados, os valores de  $t$  e  $df$  não são interpretados diretamente.

 The parametric *Student's t-test* (Student, 1908) was applied to verify whether there is a significant difference between the compared results. The test indicated that there is a significant difference between the results of the methods for the instances of the IPMTC-I set, and that BRKGA has the best mean values ( $t = -4.5271$ ,  $df = 11$ ,  $p = 0.0004309$ ) with a significance level of 0.05.

Fonte: Soares, Leonardo Cabral R., and Marco Antonio M. Carvalho. *Biased random-key genetic algorithm for scheduling identical parallel machines with tooling constraints*. European Journal of Operational Research 285.3 (2020): 955-964.

 The *Student's t-test* (Student, 1908) was applied to verify whether there is a significant difference between the methods compared and indicated that there is no a significant difference between the results compared ( $t = 0.88192$ ,  $df = 4$ ,  $p = 0.7862$ ) for a significance level of 0.05.

Fonte: Soares, Leonardo CR, et al. *Heuristic methods to consecutive block minimization*. Computers & Operations Research 120 (2020): 104948.

## ANOVA

A *Análise de Variância* (ANOVA) é uma coleção de modelos estatísticos que visa verificar se há diferença significativa entre as médias de três ou mais diferentes populações ao mesmo tempo, i.e., é um teste de comparações múltiplas.

Este é um teste paramétrico, portanto, as populações devem possuir distribuição normal.

No teste ANOVA **unidirecional**, os dados são organizados em vários grupos com base em uma única variável de agrupamento, também chamada de **variável de fator**.

## Repeated Measures One-Way ANOVA

O *Repeated Measures One-Way ANOVA*, ou ANOVA para medidas repetidas, considera que as amostras são **dependentes**, como os valores de solução de diferentes métodos para as mesmas instâncias.

A hipótese nula é a de que as médias populacionais são iguais, e a hipótese alternativa é a de que pelo menos uma das médias é diferente das demais.

Note que isso não quer dizer que todas as médias são diferentes.

Adicionalmente, o teste não indica qual população possui média diferente.

## Preparando os dados

No *R*, crie uma série de dados *data* em formato de matriz com três colunas: **instancia**, **metodo** e **resultado**.

Os dados da coluna **resultado** devem ser agrupados por instância.

No exemplo a seguir, há cinco métodos (de *A* a *E*) e os resultados de cada um deles para sete instâncias.

# Testes não paramétricos

```
1 > data <- data.frame(instancia = rep(1:7, each=5),
2                       metodo = rep(c("A", "B", "C", "D", "E"),
3                                   times=7),
4                       resultado = c(33.08, 38.69, 38.64,
5                                     33.66, 39.20, 60.83, 59.57, 60.33,
6                                     56.57, 60.35, 66.47, 62.65, 64.31,
7                                     59.48, 67.55, 69.35, 64.23, 66.23,
6                                     66.37, 70.93, 70.59, 65.61, 71.54,
7                                     73.35, 79.60, 72.90, 66.17, 71.91,
6                                     68.08, 84.33, 64.37, 60.23, 63.67,
5                                     60.22, 70.15))
4
5 > data
6
7 > boxplot(data$resultado~data$metodo)
```



# Testes não paramétricos

1	1	A	33.08
2	1	B	38.69
3	1	C	38.64
4	1	D	33.66
5	1	E	39.20
6	2	A	60.83
7	2	B	59.57
8	2	C	60.33
9	2	D	56.57
10	2	E	60.35
11	3	A	66.47
12	3	B	62.65
13	3	C	64.31
14	3	D	59.48
15	3	E	67.55

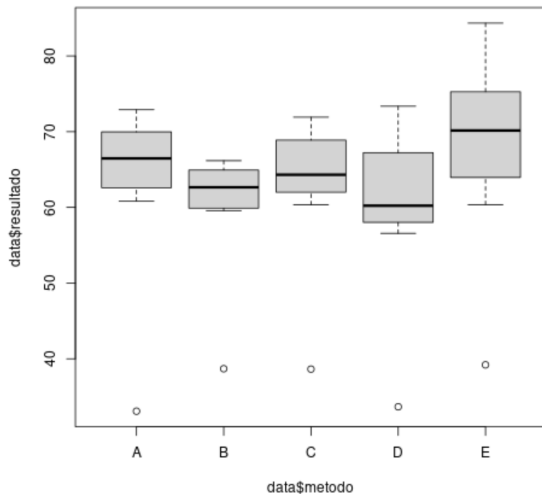
# Testes não paramétricos

16	4	A	69.35
17	4	B	64.23
18	4	C	66.23
19	4	D	66.37
20	4	E	70.93
21	5	A	70.59
22	5	B	65.61
23	5	C	71.54
24	5	D	73.35
25	5	E	79.60
26	6	A	72.90
27	6	B	66.17
28	6	C	71.91
29	6	D	68.08
30	6	E	84.33

# Testes não paramétricos

31	7	A	64.37
32	7	B	60.23
33	7	C	63.67
34	7	D	60.22
35	7	E	70.15

# Testes não paramétricos



# Testes não paramétricos

## Como executar o teste

No *R*, utilize a matriz criada para definir os parâmetros do *Repeated Measures One-Way ANOVA*:

- y*: O vetor de variáveis de resposta é dado pela coluna *resultado*.
- groups*: A variável de agrupamento, que indica a qual grupo uma observação pertence, é representada pela coluna *metodo*.
- blocks*: A variável de blocagem é representada pela coluna *instancia*.

# Testes paramétricos

```
1 > teste <- aov(resultado~factor(metodo)+Error(factor(  
    instancia)), data = data)  
2  
3 > summary(teste)
```

# Testes paramétricos

Error: factor(instancia)

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
Residuals	6	4507	751.1		

Error: Within

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
factor(metodo)	4	285.1	71.27	8.496	0.000203 ***
Residuals	24	201.3	8.39		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
' ' 1

## Análise

Entre os dados retornados consta a coluna  $Pr(>F)$ , que corresponde ao *p-value*.

Se o *p-value* for menor do que nível de significância  $\alpha$  (normalmente 0,05) é possível concluir que há diferença significativa entre as populações.

Caso contrário, não é possível afirmar que há diferença significativa entre os resultados dos métodos.

Caso haja diferença significativa, realizamos um *Pairwise t-test* para descobrir onde reside a diferença.



## Como reportar

*"The repeated measures ANOVA was performed to compare the effect of methods on the instances. According to the The p-value of  $X.0e-X$ , there **[was or was not]** a statistically significant difference in results between at least two methods."*

## Pairwise t-test

O *pairwise t-test* realiza comparações par a par entre as médias das populações com correções para comparações múltiplas.

Este teste é utilizado comumente em conjunto com o *repeated measures ANOVA*.

É utilizada a correção de *Bonferroni* no *p-value*, para contornar o problema de múltiplas comparações e evitar resultados falsos-positivos.

# Testes paramétricos

## Como executar o teste

O *repeated measures one-way ANOVA* seguido do *pairwise t-test* é exemplificado a seguir.

```
1 > teste <- aov(resultado~factor(metodo)+Error(factor(  
    instancia)), data = data)  
2  
3 > summary(teste)  
4  
5 > pairwise.t.test(data$resultado, data$metodo, p.adjust="  
    bonferroni")
```

## Análise

Como resultado é reportada uma matriz triangular dos *p-values* entre todos os grupos comparados.

A análise é realizada para cada combinação linha  $\times$  coluna.

Um *p-value* menor do que o nível de significância  $\alpha$  (normalmente 0,05) nos permite concluir que há diferença significativa entre as populações.

# Testes paramétricos

Pairwise comparisons using t tests with pooled SD

```
data: data$resultado and data$metodo
```

	A	B	C	D
B	0.7306852	-		-
C	0.9507733	0.1502305	-	-
D	0.1904008	0.1001987	0.1014083	-
E	0.1731804	0.0054761	0.9775033	0.9337718

P value adjustment method: bonferroni

The ANOVA test was applied to evaluate if there is significant difference among the removal operators. The p-value of  $6.62 \cdot 10^{-05}$  indicates that there is such a difference. The pairwise t-test was applied to analyze where this difference lies. According to the test,  $UR_E$  significantly differs from UR, and RR significantly differs from UR,  $ER_L$ , and  $ER_R$ . It is worthwhile to mention that the new proposed operators ( $ER_L$  and  $ER_R$ ) are competitive as they do not show significant difference from the UR and  $UR_E$  operators.

Fonte: Santos, Vinícius Gandra Martins, and Marco Antonio Moreira de Carvalho. *Tailored heuristics in adaptive large neighborhood search applied to the cutwidth minimization problem*. European Journal of Operational Research 289.3 (2021): 1056-1066.

## One-Way ANOVA

O *One-Way ANOVA* é um teste de comparações múltiplas de populações consideradas normalmente distribuídas.

O *One-Way ANOVA* possui o mesmo objetivo do *repeated measures ANOVA*, porém, a diferença é que no *One-Way ANOVA* as amostras são **independentes**.

Este teste deve ser utilizado quando os dados **não são** pareados por instância, como indicadores de desempenho de diferentes componentes de um mesmo algoritmo, e.g., operadores de busca local.

A hipótese nula é de que as populações possuem distribuições idênticas.

## Preparando os dados

No *R*, crie uma série de dados *data* em formato de matriz com três colunas: **instancia**, **metodo** e **resultado**.

Os dados da coluna **resultado** podem ser agrupados por instância.

No exemplo a seguir, novamente há cinco métodos (de *A* a *E*) e os resultados para sete instâncias, porém, os valores diferem do exemplo anterior.



# Testes paramétricos

```
1 teste <- aov(resultado~metodo, data = data)
2
3 summary(teste)
```

# Testes paramétricos

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
metodos	4	1836	459.1	3.257	0.0248 *
Residuals	30	4229	141.0		

---

Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1  
' ' 1

## Análise

Entre os dados retornados consta a coluna  $Pr(>F)$ , que corresponde ao *p-value*.

Se o *p-value* for menor do que nível de significância  $\alpha$  (normalmente 0,05) é possível concluir que há diferença significativa entre as populações.

Caso contrário, não é possível afirmar que há diferença significativa entre os resultados dos métodos.

Caso haja diferença significativa, realizamos um *Tukey's test* para descobrir onde reside a diferença.

## Como reportar (1)

*"The ANOVA test was applied to evaluate if there is significant difference among the methods. The p-value of  $X.0e-X$  indicates that there is such a difference."*

## Como reportar (2)

*The ANOVA test was applied to evaluate if there is significant difference among the methods. The p-value of  $X.0e-X$  indicates that there is not such a difference.*

## Tukey's HSD test

O *Tukey's HSD* (*honestly significant difference*) test é um teste de comparações múltiplas, utilizado em conjunto com o ANOVA.

Este teste compara todos os possíveis pares de médias e pode ser usado para encontrar médias que são significativamente diferentes entre si.

Os grupos associados a cada média no teste devem ser normalmente distribuídos.

A hipótese nula é a de que todas as médias comparadas pertencem a mesma população.

## Como executar o teste

Usando os mesmos dados preparados anteriormente, o *ANOVA* seguido do *pairwise t-test* é exemplificado a seguir.

```
1 > teste = aov(resultados ~ metodos, data=data)
2 > TukeyHSD(teste, conf.level = 0.95)
```

## Análise

Como resultado é apresentada uma tabela cuja coluna *p adj* corresponde ao *p-value* de cada comparação par a par, identificados no início de cada linha.

Um *p-value* menor do que o nível de significância  $\alpha$  (normalmente 0,05) indica que há diferença significativa entre os resultados do par correspondente à linha da tabela.

# Testes paramétricos

Tukey multiple comparisons of means

95% family-wise confidence level

Fit: aov(formula = resultados ~ metodos, data = data)

\$metodos

	diff	lwr	upr	p adj
B-A	-2.92000000	-22.34316	16.50316	0.0054761
C-A	-0.13714286	-19.56030	19.28602	0.0114083
D-A	-2.83714286	-22.26030	16.58602	0.1502305
E-A	4.93142857	-14.49173	24.35459	0.1001987
C-B	2.78285714	-16.64030	22.20602	0.0545846
D-B	0.08285714	-19.34030	19.50602	0.1731804
E-B	7.85142857	-11.57173	27.27459	0.7306852
D-C	-2.70000000	-22.12316	16.72316	0.0021710
E-C	5.06857143	-14.35459	24.49173	0.0019408
E-D	7.76857143	-11.65459	27.19173	0.9507733



# Testes não paramétricos

## Como executar o teste

Também é possível gerar um gráfico que exhibe as diferenças entre todos os métodos, ideal para ser utilizado ao reportar o teste.

```
1 > teste = aov(resultados ~ metodos, data=data)
2 > grafico = TukeyHSD(teste, conf.level = 0.95)
3 > plot (grafico)
```

# Testes paramétricos

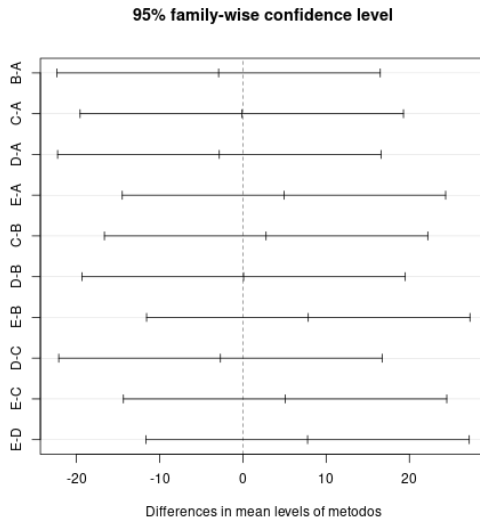


Gráfico dos níveis de diferença entre as médias.

## Como reportar

*"The Tukey's test was applied to analyze where this difference lies. According to the test, method A significantly differs from method B, and method C significantly differs from methods D and E. There is not a significant difference among the other combinations of methods."*

## Leitura recomendada

- ▶ Chi Yau. *Population Mean Between Two Matched Samples*. 2022. Disponível em <https://bityli.com/PkoPc>.
- ▶ Chi Yau. *Analysis of Variance*. 2022. Disponível em <https://bityli.com/DXsGF>.
- ▶ Shapiro, S. S., & Wilk, M. B. (1965). *An analysis of variance test for normality (complete samples)*. Biometrika, 52(3/4), 591-611.
- ▶ Student, 1908. *The probable error of a mean*. Biometrika 6 (1), 1-25. (1908).
- ▶ Fisher, Ronald A. (1921). *On the Probable Error of a Coefficient of Correlation Deduced from a Small Sample*. Metron. 1: 3-32.
- ▶ Tukey, John (1949). *Comparing Individual Means in the Analysis of Variance*. Biometrics. 5 (2): 99-114.

# Dúvidas?

