

Universidade Federal do Rio Grande do Norte  
Instituto Metr pole Digital  
**IMD0601 - Bioestat stica**

# ANOVA

---

Prof. Dr. Tetsu Sakamoto  
Instituto Metr pole Digital - UFRN  
Sala A224, ramal 182  
Email: [tetsu@imd.ufrn.br](mailto:tetsu@imd.ufrn.br)



## Baixe a aula (e os arquivos)

- Para aqueles que não clonaram o repositório:

```
> git clone https://github.com/tetsufmbio/IMD0601.git
```

- Para aqueles que já tem o repositório local:

```
> cd /path/to/IMD0601
```

```
> git pull
```

# Aula passada

- **ANOVA**

- **Comparação de médias de 3 ou mais grupos.**

Fonte de Variação	Soma dos Quadrados (SQ)	gl	Quadrados das Médias	Estatística F
Entre os grupos	$SQ_{TRAT} = \sum_{i=1}^k n_i \left( \frac{y_i}{n_i} - \frac{\bar{y}_{...}}{N} \right)^2$	$k - 1$	$QM_{TRAT} = \frac{SQ_{TRAT}}{k-1}$	$F = \frac{QM_{TRAT}}{QM_{ERRO}}$
Dentro dos Grupos	$SQ_{ERRO} = SQ_{Total} - SQ_{TRAT}$	$n - k$	$QM_{ERRO} = \frac{SQ_{ERRO}}{n-k}$	-
Total	$SQ_{Total} = \sum_{i=1}^k \sum_{j=1}^n \left( y_{ij} - \frac{\bar{y}_{...}}{N} \right)^2$	$n - 1$	-	-

# Nesta aula

- **ANOVA**
  - **Comparações múltiplas**
    - **Teste de Tukey**
- **Método não paramétrico**
  - **Kruskal-Wallis**
  - **Teste de comparação múltipla de Nemenyi**

# ANOVA

Hipóteses do ANOVA:

$$H_0 : \bar{x}_1 = \bar{x}_2 = \dots = \bar{x}_i$$

$H_A : \textit{pelo menos uma média é diferente}$

Se a hipótese nula foi rejeitada, então quais das médias são os que diferem?

- Teste-T par a par
- Múltiplas comparações (Teste de Tukey)

# Teste-T par a par

Realiza teste-T entre pares de grupos.

```
pairwise.t.test(<vetor de resposta>,<vetor de grupo>)
```

	A	B	C	Control	D
B	0.00773	-	-	-	-
C	1.5e-11	5.0e-13	-	-	-
Control	0.00058	0.33658	2.9e-13	-	-
D	1.7e-08	1.5e-10	0.00534	5.1e-11	-
E	8.6e-10	1.5e-11	0.09415	6.8e-12	0.33658

# Desvantagens de se utilizar teste-T par a par

Quando realizamos o teste T → Comparações são independentes;

Para cada comparação feita → 5% de chance de falso positivo (erro tipo I);

Se você faz 20 comparações, qual a probabilidade de você não encontrar nenhum falso positivo?

$$0.95^{20} = 0.358$$

Qual a probabilidade de encontrar pelo menos um falso positivo?

$$1 - 0.358 = 0.642$$

Quanto mais comparações → maiores as chances de encontrar falso positivo.

# Comparações múltiplas

Procedimentos que procuram corrigir o erro estatístico aplicado quando se realiza múltiplos testes em um determinado dado.

Exemplos:

- Teste de Tukey;
- FDR (False Discovery Rate);
- Correção de Bonferroni;
- etc;



# Teste de Tukey

$$HSD = \frac{M_i - M_j}{\sqrt{\frac{MS_w}{n_h}}}$$

- $M_i - M_j \rightarrow$  diferença das médias;
- $MS_w \rightarrow$  Quadrado das médias dentro do grupo (QMerro)
- $n_h \rightarrow$  número de amostras no tratamento

Teste T para duas amostras

$$t = \frac{\bar{X} - \bar{Y}}{\sqrt{\frac{s_X^2}{n} + \frac{s_Y^2}{m}}}$$

De todas as comparações que você fez, existe a chance de 5% deles serem falso positivos.

# Teste de Tukey

$$HSD = \frac{M_i - M_j}{\sqrt{\frac{MS_w}{n_h}}}$$

Pressuposto para utilização do teste:

- As observações são independentes dentro e entre os grupos;
- Os grupos devem ser normalmente distribuídos;
- A variância dentro do grupo deve ser constante.

# Será que os nossos dados satisfazem aos requisitos do ANOVA?

Distribuição normal nos grupos

- Verificar se os resíduos seguem uma distribuição normal;
  - QQplot;
  - Teste de Shapiro-Wilk;

Homogeneidade das variâncias

- Teste de Barlett;
- Teste de Levene;
- Teste de Fligner-Killeen.

# E quando os dados não atendem umas das condições?

Utiliza-se testes não paramétricos:

- ANOVA → **Kruskal-Wallis**
- Teste de Tukey → **Teste de comparação múltipla de Nemenyi**

# Teste de Kruskal-Wallis

Extensão do teste de Wilcoxon para duas amostras (Mann-Whitney);

Consiste em ordenar os valores, ranqueá-los e testar se as “medianas” entre os grupos são iguais;

- $H_0$ : Não existe diferença entre as “medianas” dos grupos.
- $H_1$ : Há pelo menos uma “mediana” diferente.

# Teste de Kruskal-Wallis

```
> kruskal.test(Growth~factor(Treatment))
```

```
Kruskal-Wallis rank sum test
```

```
data: Growth by factor(Treatment)
```

```
Kruskal-Wallis chi-squared = 27.21, df = 5, p-value = 5.19e-05
```

P-valor menor que 0,05 → Rejeita  $H_0$ ;

# Como saber em quais solos o peso do tomate difere?

## Teste de comparação múltipla de Nemenyi

- Nemenyi (1963);
- Teste de comparação múltipla utilizado após a aplicação de um teste não paramétrico com três ou mais fatores;
- Consiste em fazer comparação aos pares;

```
library(PMCMR)
```

```
posthoc.kruskal.nemenyi.test(Growth ~ factor(Treatment))
```

# Exercício

Tomateiros foram plantados em solos com diferentes composições nutricionais. Existe diferença no peso dos tomates que foram cultivados nesses diferentes solos?

Grupo	Peso dos tomates italianos colhidos (em gramas)					
Solo "A"	128,5	162,8	111,4	128,5	205,7	128,5
Solo "B"	128,5	102,8	102,8	179,9	248,5	137,1
Solo "C"	162,8	137,1	68,6	98,5	77,1	137,1
Solo "D"	85,7	102,8	111,4	77,1	60	68,6