

# Especialização em *Data Science* e Estatística Aplicada

## Módulo II - Análise estatística de várias populações

Profa. Dra. Tatiane F N Melo

Goiânia, 2024

**IME**

INSTITUTO DE  
MATEMÁTICA E  
ESTATÍSTICA

**FEN**

FACULDADE DE  
ENFERMAGEM



**UFG**

UNIVERSIDADE  
FEDERAL DE GOIÁS



# Aula 4 - Parte 2

## 1. Análise de variância de um fator (One-Way ANOVA)

### ■ Execução da ANOVA

## 2. Referências Bibliográficas

# Análise de Variância de um fator (ANOVA)

Como vimos anteriormente, antes de realizar a ANOVA precisamos ficar atentos aos pressupostos:

- Normalidade dos dados.
- Homogeneidade das variâncias.
- Independência dos dados (Não pareados).

# Análise de Variância de um fator (ANOVA)

- Com relação à normalidade podemos usar o teste `shapiro.test()` e/ou técnicas gráficas (histogramas e QQplots).
- Um teste muito utilizado para a homogeneidade das variâncias é o teste de Bartlett (`bartlett.test()`).
  - Este teste assume que os dados seguem uma distribuição normal dentro de cada grupo.
  - Por causa dessa suposição, o teste de Bartlett é sensível a desvios de normalidade.
  - Quando os dados não seguem uma distribuição normal, os resultados do teste podem não ser confiáveis.

# Análise de Variância de um fator (ANOVA)

- Se rejeitamos a hipótese de normalidade dos dados, então podemos usar o Teste de Levene (`leveneTest()`).
  - O teste de Levene verifica a igualdade de variâncias entre grupos, mas não requer a suposição de normalidade tão rigorosamente quanto o teste de Bartlett.
- Se rejeitarmos a hipótese de igualdade de variâncias entre grupos, então a ANOVA tradicional (definida anteriormente) não pode ser usada.
  - Neste caso, podemos usar, por exemplo, a ANOVA de Welch (WELCH, 1951), que é uma versão robusta da ANOVA que não exige homogeneidade das variâncias. Ela ajusta os graus de liberdade com base nas variâncias dos grupos, tornando-a adequada quando as variâncias são desiguais.

# Análise de Variância de um fator (ANOVA)

- Já para verificação do pressuposto de independência dos dados não existe um teste estatístico direto como existem para os pressupostos de homogeneidade de variâncias ou normalidade.
  - A independência dos dados significa que a observação de uma variável em um grupo não deve influenciar as observações de outra variável no mesmo grupo ou em outros grupos.
  - Ou seja, cada observação deve ser obtida de forma que uma observação não afete as outras (isto pode ser visto no momento da amostragem).

# Análise de Variância de um fator (ANOVA)

## Resumindo

Para realizar uma análise de variância de um fator, precisamos:

- Carregar os dados no R.
- Verificar se os pressupostos da ANOVA são atendidos.
- Realizar a ANOVA para comparar as médias entre os grupos.

Se rejeitarmos a hipótese de igualdade de médias da ANOVA, podemos, se for de interesse do pesquisador, realizar um teste de comparação múltipla para identificar quais grupos têm diferenças significativas entre si.

# Teste para comparação múltipla

- Na literatura, existem vários testes para comparação múltipla, após realizarmos a ANOVA, e a mesma indicar que há diferença entre as médias dos grupos.
- Um teste muito utilizado, após a ANOVA para comparar as médias entre todos os pares de grupos, é o teste de Tukey.
  - É um teste ideal para quando muitas comparações estão sendo feitas, mas depende da normalidade dos dados e da homogeneidade das variâncias para ser válido.



# Teste para comparação múltipla

- Uma alternativa ao teste de Tukey, quando os grupos têm variâncias diferentes, ou seja, quando o pressuposto de homogeneidade das variâncias é violado é o teste de Games-Howell (GAMES & HOWELL, 1974).
  - Este teste é utilizado após a ANOVA de Welch (que não assume homogeneidade das variâncias). Ele é ideal para realizar comparações múltiplas entre grupos quando o pressuposto de variâncias homogêneas é violado.

# Execução da ANOVA com um fator

## Voltando ao Exemplo 1

Exemplo 1 no R.

# Aplicação à dados reais

## Exemplo 3

Vamos considerar a Base de dados SINASC (Sistema de Informações sobre Nascidos Vivos), no município de São Paulo, em 2023. Suponha que queremos verificar se existe uma diferença significativa entre as idades médias das mães nos diferentes locais de nascimento dos filhos. As variáveis usadas são: "Idade das mães" e "Local de nascimento", cujos níveis são: hospital, Outros estabelecimentos de saúde, Domicílio e Outros.

# Aplicação à dados reais

## Continuação do Exemplo 3

A hipótese nula é:

$H_0$  : As médias das idades das mães são iguais em todos os locais de nascimento, ou seja,

$$H_0 : \mu_1 = \mu_2 = \mu_3 = \mu_4,$$

onde

- $\mu_1$  é a média da idade das mães cujo local de nascimento é "Hospital";
- $\mu_2$  é a média da idade das mães cujo local de nascimento é "Outros estabelecimentos de saúde";
- $\mu_3$  é a média da idade das mães cujo local de nascimento é "Domicílio";
- $\mu_4$  é a média da idade das mães cujo local de nascimento é "Outros".

# Aplicação à dados reais

## Continuação do Exemplo 3

A hipótese alternativa é:

$H_1$  : Pelo menos uma das médias das idades das mães é diferente entre os locais de nascimento.

# Aplicação à dados reais

## Continuação do Exemplo 3

Exemplo 3 no R.

# Referência bibliográfica

1. Daniel, W. W., & Cross, C. L. *Biostatistics: A Foundation for Analysis in the Health Sciences* (11th ed.). Hoboken, NJ: Wiley, 2018.
2. Games, P. A., & Howell, F. J. Pairwise Multiple Comparison Procedures with Unequal N's and/or Variances: A Monte Carlo Study. *Journal of Educational and Behavioral Statistics*, 1(2), 113-125, 1976, <https://doi.org/10.3102/10769986001002113>
3. Ministério da Saúde - Vacinômetro COVID-19. [https://infoms.saude.gov.br/extensions/SEIDIGI\\_DEMAS\\_Vacina\\_C19/SEIDIGI\\_DEMAS\\_Vacina\\_C19.html](https://infoms.saude.gov.br/extensions/SEIDIGI_DEMAS_Vacina_C19/SEIDIGI_DEMAS_Vacina_C19.html), último acesso: 17/09/2024.
4. Welch, B. L. On the comparison of several mean values: An alternative approach. *Biometrika*, 38(3/4), 330-336, 1951, <https://doi.org/10.2307/2332579>

# Especialização em *Data Science* e Estatística Aplicada

## Módulo II - Análise estatística de várias populações

Profa. Dra. Tatiane F N Melo

[tmelo@ufg.br](mailto:tmelo@ufg.br)

**IME**

INSTITUTO DE  
MATEMÁTICA E  
ESTATÍSTICA

**FEN**

FACULDADE DE  
ENFERMAGEM



**UFG**

UNIVERSIDADE  
FEDERAL DE GOIÁS

