

ANOVA

(Analysis of Variance)

Prof. Dr. Guanys de Barros Vilela Junior

Para que serve a ANOVA?

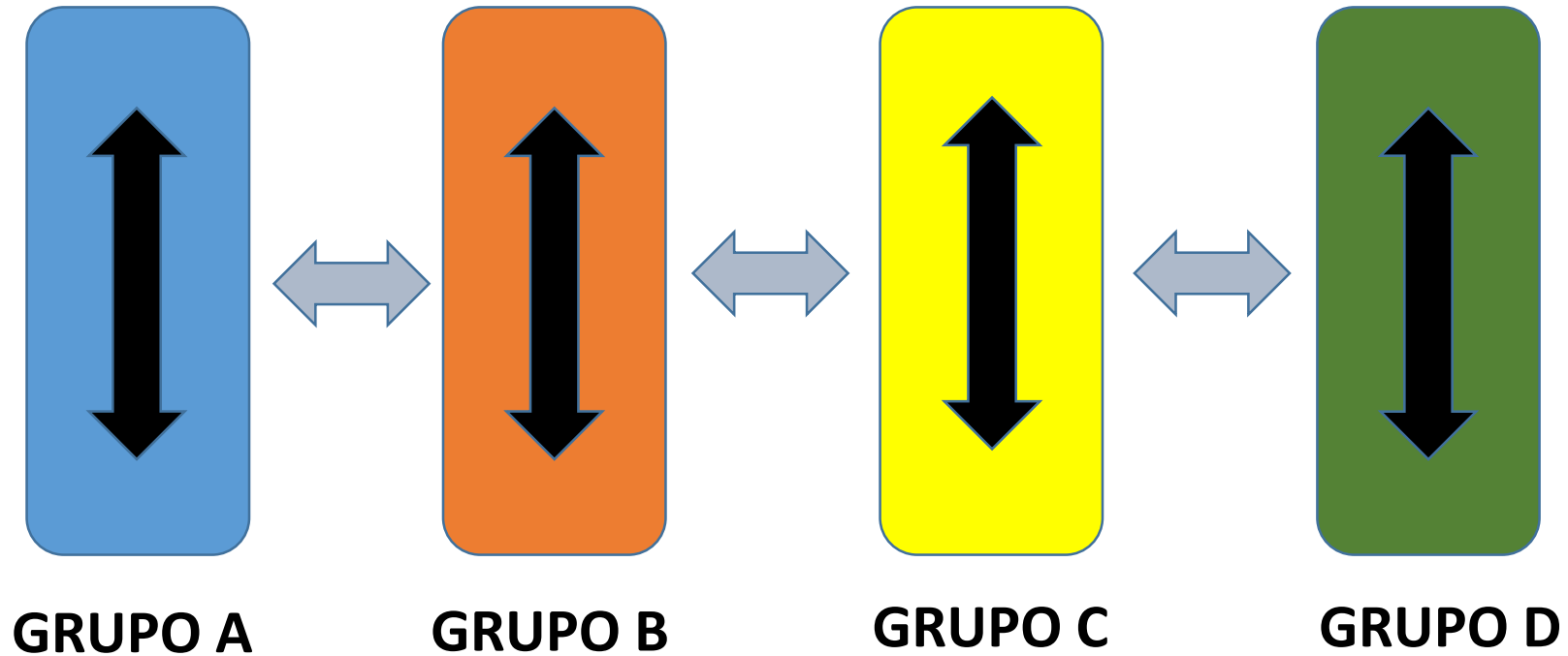
- Para comparar três ou mais variáveis ou amostras.
- Por exemplo, queremos testar os efeitos cardiorrespiratórios de 4 programas (A, B, C, D) de treinamento, através do $VO_{2máx}$, aplicados simultaneamente em 4 diferentes grupos de 20 sujeitos cada.
- Se (por absurdo) fossem comparados aos pares (AB; AC; AD; BC; BD; CD) cada comparação aceitando uma margem de erro de 5%, teríamos uma chance de erro Tipo I (rejeitar a hipótese verdadeira) de: $6 \times 5\% = 30\%$.
- H_0 : não existe diferenças entre as médias dos 4 programas de treinamento.
- H_1 : a média de pelo menos um dos grupos se difere das demais.


Para que serve a ANOVA?

- Se H_0 for rejeita isto mostra a existência de diferença entre os grupos e com a ANOVA é possível localizar onde está localizada está diferença.
- Para isto é necessário testar a variabilidade em cada grupo e também entre os grupos, ou seja, intragrupos e intergrupos.
- Para isto é calculado o F, pela fórmula:

$$F = \frac{\textit{variância entre os grupos}}{\textit{variância dentro dos grupos}}$$

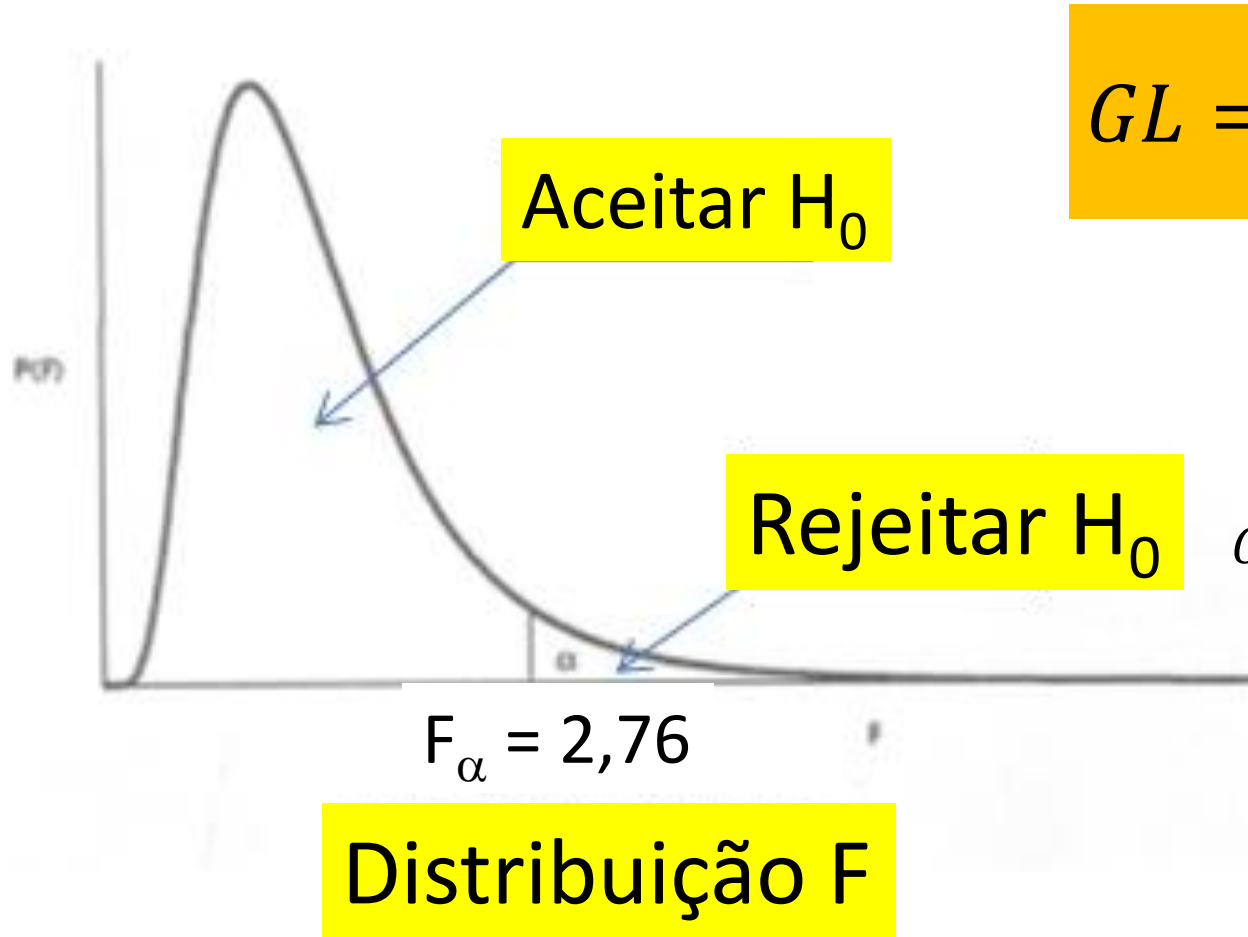
Variância intergrupos e intragrupos



 = variabilidade intragrupos

 = variabilidade intergrupos

Distribuição F e Graus de Liberdade



$$GL = \frac{\text{número de grupos} - 1}{\text{soma dos GL de cada grupo}}$$

No exemplo citado:

$$GL = \frac{4 - 1}{(20 - 1) + (20 - 1) + (20 - 1) + (20 - 1)}$$

$$GL = \frac{3}{76} \quad \alpha = 5\%$$

Ver Tabela da Distribuição F

Calculando F e comparando-o com o F_{α}

- Após calcular o F pela equação:

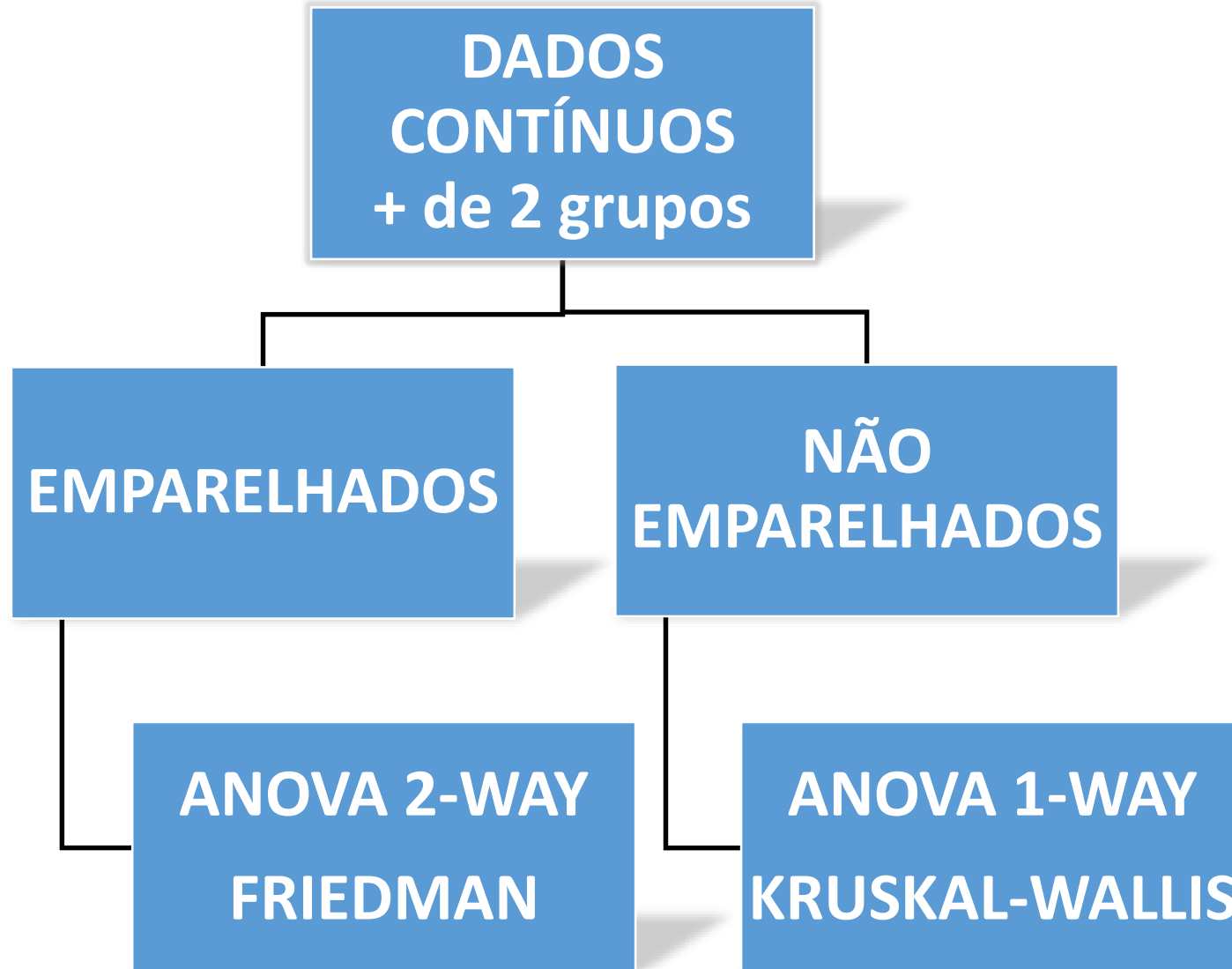
$$F = \frac{\textit{variância entre os grupos}}{\textit{variância dentro dos grupos}}$$

- Se o **F** calculado for **MAIOR** que F_{α} da tabela, rejeitamos H_0 , ou seja, existem diferenças entre dois ou mais grupos.

Sobre a ANOVA

- **ATENÇÃO**: A ANOVA informa apenas se existem diferenças entre dois ou mais grupos, mas não identifica quais são os grupos diferentes.
- Para identificar tais grupos é necessário fazer a comparação entre pares.
- Existem vários testes para isto, tais como: Teste de Fisher, Teste de Duncan, Teste de Tukeys, Teste de Scheffé, dentre outros.
- A distribuição deve ser normal para aplicar a ANOVA.
- ANOVA *one-way*: só um critério, por exemplo, o número de artigos publicados por diferentes grupos de pesquisadores.
- ANOVA *two-way*: dois critérios, por exemplo, o número de artigos e o sexo dos membros dos diferentes grupos de pesquisadores.

Finalizando



Exercício

A tabela ao lado mostra o número de artigos publicados num período de 2 anos de quatro grupos de pesquisadores. Cada grupo é formado por 7 pesquisadores.

Calcule o **F** (ANOVA) para averiguar se as diferenças na produção de artigos nestas universidades é estatisticamente significativa.

Para isto seguiremos o passa-a-passo:

G1	G2	G3	G4
2	5	5	8
1	7	4	4
3	9	6	6
3	8	5	8
4	9	5	5
2	7	6	6
3	10	4	7

PASSO-A-PASSO PARA REALIZAR A ANOVA

- **Passo 1:** calcular a soma e média de cada grupo
- **Passo 2:** calcular a média total
- **Passo 3:** calcular os desvios (**d**) de cada escore em relação à média (**d = x - \bar{x}**)
- **Passo 4:** calcular o quadrado dos desvios (**d²**)
- **Passo 5:** calcular a soma dos quadrados (**SQ_{intra}**) para cada grupo
- **Passo 6:** calcular os **d**'s para cada grupo a partir da média total dos grupos (**d_g**)
- **Passo 7:** calcular o quadrado dos desvios (**d_g²**) e sua somatória
- **Passo 8:** calcular a soma dos quadrados entre os grupos (**SQ_{entre}**)
- **Passo 9:** calcular os graus de liberdade (**GL**) entre os grupos e dentro dos grupos
- **Passo 10:** calcular o quadrado médio (**QM**) *dentro* e *entre* os grupos
- **Passo 11:** calcular o **F**, comparar com o valor crítico tabelado e aplicar teste de Scheffé para identificar quais grupos são estatisticamente diferentes.

Solução:

Obs: A realização da ANOVA possui vários detalhes e requer uma boa organização dos dados para a realização dos cálculos.

	G1	G2	G3	G4
	2	5	5	8
	1	7	4	4
	3	9	6	6
	3	8	5	8
	4	9	5	5
	2	7	6	6
	3	10	4	7

Passo 1

SOMA	18	55	35	44
MÉDIA	2,57	7,86	5,00	6,29

Solução:

Passo 2: calcular a média total

$$\Sigma_{\text{total}} = \Sigma_1 + \Sigma_2 + \Sigma_3 + \Sigma_4 = 18 + 55 + 35 + 44 = 152$$

$$\Sigma n = 7 + 7 + 7 + 7 = 28$$

$$\bar{X} = \frac{152}{28} = 5,43$$

Solução:

Passo 3: calcular os desvios (**d**) de cada escore em relação à média (**d = x - \bar{x}**)

Passo 4: calcular o quadrado dos desvios (**d²**)

d1	d1 ²	d2	d2 ²	d3	d3 ²	d4	d4 ²
-0,57	0,33	-2,86	8,16	0,00	0,00	1,71	2,94
-1,57	2,47	-0,86	0,73	-1,00	1,00	-2,29	5,22
0,43	0,18	1,14	1,31	1,00	1,00	-0,29	0,08
0,43	0,18	0,14	0,02	0,00	0,00	1,71	2,94
1,43	2,04	1,14	1,31	0,00	0,00	-1,29	1,65
-0,57	0,33	-0,86	0,73	1,00	1,00	-0,29	0,08
0,43	0,18	2,14	4,59	-1,00	1,00	0,71	0,51
$\Sigma d1^2 =$	5,71	$\Sigma d2^2 =$	16,86	$\Sigma d3^2 =$	4,00	$\Sigma d4^2 =$	13,43

Solução:

Passo 5: calcular a soma dos quadrados (SQ_{intra}) para cada grupo

$$SQ_{intra} = 5,71 + 16,86 + 4,00 + 13,43 = 40,0$$

Passo 6: calcular os d 's para cada grupo a partir da média total dos grupos (d_g)

Passo 7: calcular o quadrado dos desvios (d_g^2) e sua somatória

	\bar{X} grupo	\bar{X} total	$d_g = \bar{X}_{grupo} - \bar{X}_{total}$	d_g^2
Grupo 1	2,57	5,43	-2,86	8,163
Grupo 2	7,86	5,43	2,43	5,897
Grupo 3	5,00	5,43	-0,43	0,183
Grupo 4	6,29	5,43	0,86	0,734
			$\Sigma d_g^2 =$	14,979

Solução:

Passo 8: calcular a soma dos quadrados entre os grupos (SQ_{entre})

$$SQ_{\text{entre}} = \sum d_g^2 \cdot n = 14,979 \times 7 = 104,85$$

Passo 9: calcular os graus de liberdade (**GL**) entre os grupos e dentro dos grupos

$$GL_{\text{entre}} = \text{número de grupos} - 1 = 4 - 1 = 3$$

$$GL_{\text{intra}} = \text{número de observações} - \text{número de grupos} = 28 - 4 = 24$$

Solução:

Passo 10: calcular o quadrado médio (**QM**) *dentro* e *entre* os grupos, dividindo-os pelos respectivos **GL**.

$$QM_{dentro} = \frac{SQ_{dentro}}{GL_{dentro}} = 40,0/24 = 1,666$$

$$QM_{entre} = \frac{SQ_{entre}}{GL_{entre}} = \frac{104,85}{3} = 34,95$$

Solução:

Passo 11: calcular o **F** e comparar com o valor crítico tabelado.

$$F = \frac{QM_{entre}}{QM_{dentro}} = \frac{34,95}{1,66} = 21,05$$

Como $F > F_{\alpha}$ ($21,05 > 3,01$) com significância de 5%, podemos rejeitar a hipótese nula, ou seja, as diferenças na produção de artigos são estatisticamente significantes entre as 4 universidades, mas onde estão tais diferenças?

Solução:

Para identificarmos onde a(s) diferença(s) está (estão) localizada(s), é necessário realizar uma comparação *Post Hoc*.

Para isto, utilizaremos o Teste de Sheffé:

$$I = (K - 1) \cdot (F_{\alpha}) \cdot \left(\frac{2 \cdot QM}{n}\right)$$

Onde: **K** é o número de grupos

F_α é o F tabelado

QM é a soma dos quadrados médios dentro dos grupos

n é o tamanho dos grupos

$$I = (4 - 1) \cdot (3,01) \cdot \frac{2 \cdot 1,66}{7} = 4,28$$

O Intervalo (I) calculado (4,28) é o valor mínimo da diferença entre as médias de dois grupos para que esta possa ser considerada significativa.

Solução:

Para localizar a diferença usando este $l = 4,28$, optamos por uma tabela das diferenças entre as médias:

	G1	G2	G3	G4
G1	0	5,29	2,43	3,72
G2		0	2,86	1,57
G3			0	1,29
G4				0

$G2 - G1 = 7,86 - 2,57 = 5,29 > 4,28$,
portanto a diferença só ocorre entre G1 e G2.

UFA !!!

Referências

- BARROS, M.V.G. & REIS, R.S. Análise de dados em atividade física e saúde. Londrina: Mediograf, 2003.
- DORIA FILHO, U. Introdução à bioestatística. São Paulo: Negócio, 1999.
- THOMAS, J.R. & NELSON, J.K. Métodos de pesquisa em atividade física. Porto Alegre: Artmed, 2002.
- VILELA JUNIOR, G.B. Conteúdos ministrados na disciplina Estatística em cursos de graduação e pós graduação. Campinas: UNIMEP / Metrocamp, 2003 – 2014.