

ANOVA e MANOVA

Análise de Variância

ACH2036 – Métodos Quantitativos Aplicados à Adm. de Empresas I

Prof. Regis Rossi A. Faria

2º sem. 2020

Breve histórico

- A Análise de Variância é considerada por muitos o método estatístico de maior repercussão na pesquisa científica.
- Juntamente com a análise de covariância, foi desenvolvida por Ronald A. Fisher em 1925, quando este trabalhava em uma estação de pesquisa agrícola na Inglaterra, e passaram a constituir um instrumental fundamental para a interpretação de resultados de experimentos controlados

Introdução

- O interesse da técnica está em detectar diferenças entre grupos (também chamados *tratamentos*) tomando por base a análise de variância para uma variável (dependente e métrica) que estejamos observando como se comporta em cada um de g grupos (ou k tratamentos) de um estudo.
- É uma *técnica de dependência* que mede as diferenças para uma *variável dependente métrica*, com base em um conjunto de *variáveis independentes categóricas*

Introdução

- A análise multivariada de variância (MANOVA) é uma extensão da análise de variância (ANOVA) para acomodar mais de uma variável dependente.
- É uma *técnica de dependência* que mede as diferenças para 2 ou mais *variáveis dependentes métricas*, com base em um conjunto de *variáveis categóricas* (não-métricas) que atuam como variáveis independentes.

Introdução

- MANOVA é chamada de procedimento multivariado porque a usamos para avaliar diferenças de grupos para múltiplas variáveis dependentes métricas simultaneamente.
- Em MANOVA, cada grupo de tratamento é observado em duas ou mais variáveis dependentes de interesse

Formas gerais

- ANOVA e MANOVA podem ser enunciadas nas seguintes formas gerais:

Análise de Variância

$$Y_1 = X_1 + X_2 + X_3 + \dots + X_n$$

(métrica)

(não-métrica)

Análise Multivariada de Variância

$$Y_1 + Y_2 + Y_3 + \dots + Y_n = X_1 + X_2 + X_3 + \dots + X_n$$

(métrica)

(não-métrica)

Uso

- ANOVA é uma generalização do teste t para duas amostras, sendo desenvolvida para análise dos dados de experimentos que possuem mais de dois tratamentos.
→ *Consequentemente, ela permite comparar a média de várias amostras simultaneamente*
- ANOVA e MANOVA fornecem as ferramentas necessárias para julgar se uma diferença observada (entre grupos) ocorre devido a um efeito dos (diferentes) tratamentos ou à variabilidade de amostragem aleatória.

Uso

- MANOVA tem também um papel em planejamentos não-experimentais (ex: em levantamentos de informações) onde grupos de interesse (ex: sexo, comprador/ não-comprador) são definidos e então as diferenças em qualquer número de variáveis métricas (p.ex., atitudes, satisfação, taxa de compras) são avaliadas quanto à significância estatística.

Uso

- Ex: sexo (grupo M e F)
- Diferença nas variáveis métricas atitude, satisfação, entre os grupos → se tem significância estatística

Análise de Variância

- Teste de comparação de médias de mais de duas populações
- Requisitos:
 - Exige que a variável dependente seja quantitativa
 - Exige que a distribuição da variável dependente siga uma distribuição normal dentro de cada grupo
 - Exige que as variâncias dos grupos sejam semelhantes

Situação de interesse

- Um pesquisador expôs 2 grupos de respondentes a diferentes anúncios e em seguida pediu que avaliassem o nível de apelo dos anúncios em uma escala de 10 pontos.
- Suponha que estejamos interessados em avaliar 3 mensagens, em vez de 2. Os respondentes seriam aleatoriamente designados a um dos 3 grupos, e teríamos 3 médias de amostras para comparar.
- Para analisar esses dados, poderíamos ser tentados a conduzir *testes t* separados para a diferença entre cada par de médias:
 - grupo 1 versus grupo 2
 - grupo 1 versus grupo 3
 - grupo 2 versus grupo 3

Situação de interesse

- Mas múltiplos *testes t* aumentam a taxa de erro do Tipo I → a análise de variância (ANOVA) evita essa inflação do erro Tipo I ao fazerem múltiplas comparações de grupos de tratamento, determinando, *em um único teste*, se o conjunto inteiro de médias de amostras sugere que as amostras foram obtidas a partir da mesma população geral.

Vejamos como funciona a técnica
por meio de exemplos

Exemplo 1

- Verificar se há diferença no volume de urina entre 4 grupos de pacientes que tomaram 4 diferentes diuréticos A, B, C, D,

Exemplo 1

	A	B	C	D	
	11	8	5	4	
	8	5	7	4	
	5	2	3	2	
	8	5	3	0	
	8	5	7	0	
média	8	5	5	2	média geral = 4

Qual o melhor teste de hipóteses?

- Vamos procurar determinar qual dentre possíveis testes de hipóteses podemos aplicar, um que seja adequado a este exemplo, respondendo às questões abaixo:

1) Determinação da variável em estudo

volume de urina

2) Tipo da variável dependente

quantitativa contínua

3) N° de Amostras

4 amostras

4) Relacionamento entre as amostras

Independentes

TABELA DE ORIENTAÇÃO NA ESCOLHA DE TESTES ESTATÍSTICOS

Tipo da variável dependente	Uma variável					Duas variáveis
	Uma amostra	Duas amostras		Mais de duas amostras		Medidas de correlação
		<u>relacionadas</u>	<u>independentes</u>	<u>relacionadas</u>	<u>independentes</u>	
Qualitativa nominal ou ordinal	<u>binomial ou X^2</u>	<u>McNemar</u>	X^2 ou Fischer	Prova Q de Cochran	X^2 para várias amostras	<u>coeficiente de contingência C</u>
Quantitativa discreta ou contínua (dados não seguem curva de Gauss)	Kolmogorov Smirnov	<u>Wilcoxon ou Prova dos sinais</u>	Mann-Whitney Ou Prova da Mediana	Prova de Friedman	<u>Kruskal-Wallis ou Prova da mediana</u>	<u>correlação de Spearman</u>
Quantitativa discreta ou contínua (dados seguem curva de Gauss)	<u>teste de proporções</u>	<u>teste t de Student pareado</u>	<u>teste t de Student para amostras independentes</u>	ANOVA para medidas repetidas	<u>ANOVA para grupos independentes</u>	<u>correlação de Pearson</u>



Testes não-paramétricos

Exemplo 2

- Verificar se há diferença na taxa de glicose entre 3 grupos de ratos submetidos a 3 diferentes cirurgias

Exemplo 2

Tipo de cirurgia →	1	2	3	
	96,0	90,0	86,0	
	95,0	93,0	85,0	
	100,0	89,0	105,0	
	108,0	88,0	105,0	
	120,0	87,0	90,0	
	110,5	92,5	100,0	
	97,0	87,5	95,0	
	92,5	85,0	95,0	
Médias	102,4	89,0	95,1	média geral= 95,5
variâncias	90,8	7,5	61,6	
Tamanho da amostra	8	8	8	

Qual o melhor teste de hipóteses?

1) Determinação da variável em estudo

taxa de glicemia

2) Tipo da variável dependente

quantitativa contínua

3) N° de Amostras

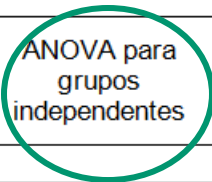
3 amostras

4) Relacionamento entre as amostras

Independentes

TABELA DE ORIENTAÇÃO NA ESCOLHA DE TESTES ESTATÍSTICOS

Tipo da variável dependente	Uma variável					Duas variáveis
	Uma amostra	Duas amostras		Mais de duas amostras		Medidas de correlação
		<u>relacionadas</u>	<u>independentes</u>	<u>relacionadas</u>	<u>independentes</u>	
Qualitativa nominal ou ordinal	<u>binomial ou X^2</u>	<u>McNemar</u>	X^2 ou Fischer	Prova Q de Cochran	X^2 para várias amostras	<u>coeficiente de contingência C</u>
Quantitativa discreta ou contínua (dados não seguem curva de Gauss)	Kolmogorov Smirnov	<u>Wilcoxon ou Prova dos sinais</u>	Mann-Whitney Ou Prova da Mediana	Prova de Friedman	<u>Kruskal-Wallis ou Prova da mediana</u>	<u>correlação de Spearman</u>
Quantitativa discreta ou contínua (dados seguem curva de Gauss)	<u>teste de proporções</u>	<u>teste t de Student pareado</u>	<u>teste t de Student para amostras independentes</u>	ANOVA para medidas repetidas	<u>ANOVA para grupos independentes</u>	<u>correlação de Pearson</u>



Testes não-paramétricos

Lógica por trás do teste

- Duas estimativas independentes da variância para a variável dependente são comparadas
 - A primeira reflete a variabilidade geral de respondentes dentro dos grupos (SQR) → quadrado médio intra-grupo
 - A segunda representa as diferenças entre grupos atribuíveis aos efeitos de tratamento (SQTr) → quadrado médio entre grupos

Lógica por trás do teste

- A razão entre SQTr e SQR é uma medida de quanta variância é atribuível aos diferentes tratamentos versus a variância esperada a partir de amostragem aleatória.
- Esta razão é conceitualmente semelhante ao valor t, mas neste caso nos dá um valor para a estatística F

$$F = \frac{SQTr}{SQR}$$

Lógica por trás do teste

- A razão entre SQTr e SQR é uma medida de quanta variância é atribuível aos diferentes tratamentos versus a variância esperada a partir de amostragem aleatória.
- Esta razão é conceitualmente semelhante ao valor t, mas neste caso nos dá um valor para a estatística F

$$F = \frac{SQTr}{SQR}$$

variabilidade das médias dos grupos de tratamento (em relação à var. dependente usada)

estimativa da variabilidade média dos respondentes dentro dos grupos

Exemplo 3: hemoglobina glicada em gestantes

• Situação

Um pesquisador na área da endocrinologia acredita que a hemoglobina glicada é diferente entre 3 grupos de gestantes:

- com diabetes (CD)
- com diabetes gestacional (DG)
- sem diabetes (SD)

• Evidência amostral

Para verificar se o pesquisador está correto, foram selecionadas 30 gestantes:

- 10 com diabetes (CD)
- 10 com diabetes gestacional (DG)
- 10 sem diabetes (SD)

	Grupo SD	Grupo DG	Grupo CD	
	7,86	6,20	9,67	
	6,38	7,82	8,08	
	6,90	8,50	9,25	
	7,78	6,50	8,20	
	8,17	8,09	8,64	
	6,26	6,90	9,67	
	6,30	7,82	9,23	
	7,86	7,45	10,43	
	7,42	7,75	9,97	
	8,63	7,43	9,59	
Média	7,36	7,45	9,27	média geral=8,02
Variância	0,717	0,516	0,580	
Nº pacientes	10	10	10	

Qual o melhor teste de hipóteses?

1) Determinação da variável dependente

hemoglobina glicada

2) Tipo da variável dependente

quantitativa contínua

3) N° de Amostras

3 amostras

4) Relacionamento entre as amostras

Independentes

TABELA DE ORIENTAÇÃO NA ESCOLHA DE TESTES ESTATÍSTICOS

Tipo da variável dependente	Uma variável					Duas variáveis
	Uma amostra	Duas amostras		Mais de duas amostras		Medidas de correlação
		<u>relacionadas</u>	<u>independentes</u>	<u>relacionadas</u>	<u>independentes</u>	
Qualitativa nominal ou ordinal	<u>binomial ou X^2</u>	<u>McNemar</u>	X^2 ou Fischer	Prova Q de Cochran	X^2 para várias amostras	<u>coeficiente de contingência C</u>
Quantitativa discreta ou contínua (dados não seguem curva de Gauss)	Kolmogorov Smirnov	<u>Wilcoxon ou Prova dos sinais</u>	Mann-Whitney Ou Prova da Mediana	Prova de Friedman	<u>Kruskal-Wallis ou Prova da mediana</u>	<u>correlação de Spearman</u>
Quantitativa discreta ou contínua (dados seguem curva de Gauss)	<u>teste de proporções</u>	<u>teste t de Student pareado</u>	<u>teste t de Student para amostras independentes</u>	ANOVA para medidas repetidas	<u>ANOVA para grupos independentes</u>	<u>correlação de Pearson</u>



Testes não-paramétricos

Exemplo 3: hemoglobina glicada em gestantes

Hipóteses estatísticas:

$$\begin{cases} H_0 : \mu_{SD} = \mu_{DG} = \mu_{CD} \\ H_1 : \text{as médias não são todas iguais entre si} \end{cases}$$

Fixa-se α

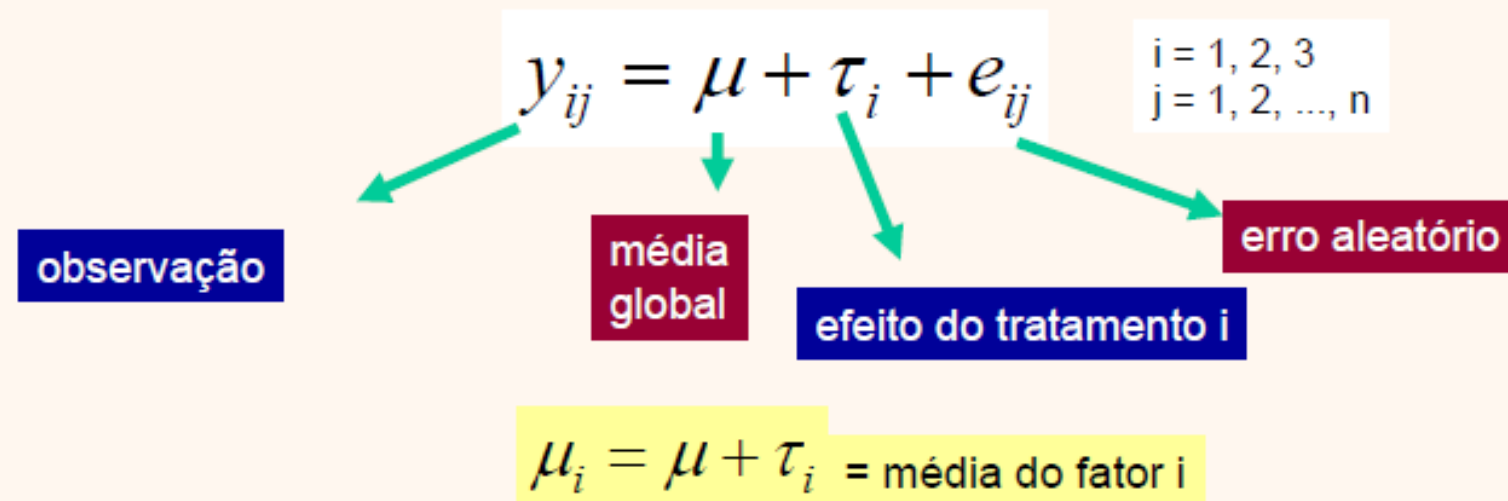
	Grupo SD	Grupo DG	Grupo CD	
	7,86	6,20	9,67	
	6,38	7,82	8,08	
	6,90	8,50	9,25	
	7,78	6,50	8,20	
	8,17	8,09	8,64	
	6,26	6,90	9,67	
	6,30	7,82	9,23	
	7,86	7,45	10,43	
	7,42	7,75	9,97	
	8,63	7,43	9,59	
Média	7,36	7,45	9,27	média geral=8,02
Variância	0,717	0,516	0,580	
Nº pacientes	10	10	10	

	Grupo SD	Grupo DG	Grupo CD	
	$x_{SD,1} = 7,86$	$x_{DG,1} = 6,20$	$x_{CD,1} = 9,67$	
	$x_{SD,2} = 6,38$	$x_{DG,2} = 7,82$	$x_{CD,2} = 8,08$	
	$x_{SD,3} = 6,90$	$x_{DG,3} = 8,50$	$x_{CD,3} = 9,25$	
	$x_{SD,4} = 7,78$	$x_{DG,4} = 6,50$	$x_{CD,4} = 8,20$	
	$x_{SD,5} = 8,17$	$x_{DG,5} = 8,09$	$x_{CD,5} = 8,64$	
	$x_{SD,6} = 6,26$	$x_{DG,6} = 6,90$	$x_{CD,6} = 9,67$	
	$x_{SD,7} = 6,30$	$x_{DG,7} = 7,82$	$x_{CD,7} = 9,23$	
	$x_{SD,8} = 7,86$	$x_{DG,8} = 7,45$	$x_{CD,8} = 10,43$	
	$x_{SD,9} = 7,42$	$x_{DG,9} = 7,75$	$x_{CD,9} = 9,97$	
	$x_{SD,10} = 8,63$	$x_{DG,10} = 7,43$	$x_{CD,10} = 9,59$	
média	$\overline{x_{SD}} = 7,36$	$\overline{x_{DG}} = 7,45$	$\overline{x_{CD}} = 9,27$	$\overline{x_{geral}} = 8,02$
variância	$s^2_{SD} = 0,717$	$s^2_{DG} = 0,516$	$s^2_{CD} = 0,580$	

Modelo da ANOVA

$g = 3$ grupos

tratamento			
(1)	(2)	(3)	
y_{11}	y_{21}	y_{31}	
y_{12}	y_{22}	y_{32}	
...	
y_{1n}	y_{2n}	y_{3n}	Média global:
Média:	$\bar{y}_{1.}$	$\bar{y}_{2.}$	$\bar{y}_{3.}$ $\bar{y}_{..}$



Análise de variância (ANOVA) com um fator

Replicação	Tratamento				
	1	2	...	g	
1	y_{11}	y_{21}	...	y_{g1}	
2	y_{12}	y_{22}	...	y_{g2}	
...	
n	y_{1n}	y_{2n}	...	y_{gn}	
Soma	$y_{1.}$	$y_{2.}$...	$y_{g.}$	$y_{..} = \sum_i y_{i.}$
Média	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{g.}$	$\bar{y}_{..} = \frac{1}{g} \sum_i y_{i.}$

Soma de quadrados total:

$$SQ_{Tot} = \sum_{i=1}^g \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$$

Graus de liberdade:

$$gl = N - 1$$

onde: $N = ng$

SQ_{Tot} : Soma dos Quadrados Total

Análise de variância (ANOVA) com um fator

Replicação	Tratamento				
	1	2	...	g	
1	y_{11}	y_{21}	...	y_{g1}	
2	y_{12}	y_{22}	...	y_{g2}	
...	
n	y_{1n}	y_{2n}	...	y_{gn}	
Soma	$y_{1.}$	$y_{2.}$...	$y_{g.}$	$y_{..} = \sum_i y_{i.}$
Média	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{g.}$	$\bar{y}_{..} = \frac{1}{g} \sum_i y_{i.}$

Soma de quadrados dos tratamentos:

$$SQ_{Trat} = \sum_{i=1}^g \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 = n \sum_{i=1}^g (\bar{y}_{i.} - \bar{y}_{..})^2$$

Graus de liberdade:

$$gl = g - 1$$

SQ_{Trat} : Soma dos **Q**uadrados entre **T**ratamentos (ou grupos)

Análise de variância (ANOVA) com um fator

Replicação	Tratamento				
	1	2	...	g	
1	y_{11}	y_{21}	...	y_{g1}	
2	y_{12}	y_{22}	...	y_{g2}	
...	
n	y_{1n}	y_{2n}	...	y_{gn}	
Soma	$y_{1.}$	$y_{2.}$...	$y_{g.}$	$y_{..} = \sum_i y_{i.}$
Média	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{g.}$	$\bar{y}_{..} = \frac{1}{g} \sum_i y_{i.}$

Soma de quadrados do erro:

$$SQ_{Erro} = \sum_{i=1}^g \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

Graus de liberdade:

$$gl = N - g$$

SQ_{Erro} : Soma de Quadrados do Erro (intra tratamentos ou resíduos)

Análise de variância de um fator

Fonte de variação	Soma de quadrados	gl: graus de liberdade	Quadrados médios	Razão F
Entre tratamentos	SQ_{Trat}	$g - 1$	$QM_{\text{Trat}} = SQ_{\text{Trat}} / (g - 1)$	$F = QM_{\text{Trat}} / QM_{\text{Erro}}$
Erro (intra-tratamentos ou resíduos)	SQ_{Erro}	$N - g$	$QM_{\text{Erro}} = SQ_{\text{Erro}} / (N - g)$	
Total	SQT	$N - 1$		

Estatística do teste: F



Exemplo 3: hemoglobina glicada em gestantes

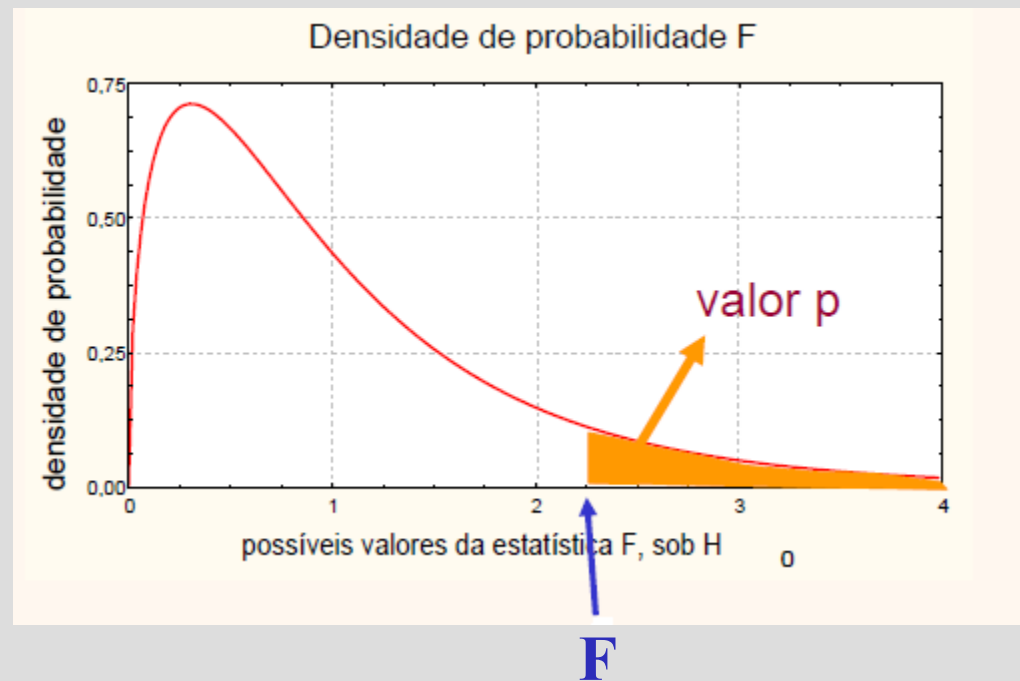
$$F = \frac{\frac{\text{desvio quadráticos entre grupos}}{\text{número de grupos} - 1}}{\frac{\text{desvio quadrático intra os grupos}}{\text{número total de sujeitos} - \text{número de grupos}}}$$

$$F = \frac{\frac{SQTr}{k - 1}}{\frac{SQR}{n - k}}$$

F corresponde à razão entre a variância entre-grupos e a variância intra-grupos

Teste F

A estatística F tem distribuição F com $(g-1)$ graus de liberdade no numerador e $(N-g)$ graus de liberdade no denominador*



* Ver livro Barbetta – Tabela 6

Voltando ao exemplo das gestantes

Grupos	Sem Diabetes	Diabetes Gestacional	Com Diabetes	
	$x_{SD,1} = 7,86$	$x_{DG,1} = 6,20$	$x_{CD1} = 9,67$	
	$x_{SD,2} = 6,38$	$x_{DG,2} = 7,82$	$x_{CD,2} = 8,08$	
	$x_{SD,3} = 6,90$	$x_{DG,3} = 8,50$	$x_{CD,3} = 9,25$	
	$x_{SD,4} = 7,78$	$x_{DG,4} = 6,50$	$x_{CD,4} = 8,20$	
	$x_{SD,5} = 8,17$	$x_{DG,5} = 8,09$	$x_{CD,5} = 8,64$	
	$x_{SD,6} = 6,26$	$x_{DG,6} = 6,90$	$x_{CD,6} = 9,67$	
	$x_{SD,7} = 6,30$	$x_{DG,7} = 7,82$	$x_{CD,7} = 9,23$	
	$x_{SD,8} = 7,86$	$x_{DG,8} = 7,45$	$x_{CD,8} = 10,43$	
	$x_{SD,9} = 7,42$	$x_{DG,9} = 7,75$	$x_{CD,9} = 9,97$	
	$x_{SD,10} = 8,63$	$x_{DG,10} = 7,43$	$x_{CD,10} = 9,59$	
Média	$\overline{x}_{SD} = 7,36$	$\overline{x}_{DG} = 7,45$	$\overline{x}_{CD} = 9,27$	$\overline{x}_{geral} = 8,02$
Variância	$s^2_{SD} = 0,717$	$s^2_{DG} = 0,516$	$s^2_{CD} = 0,580$	

Soma dos Quadrados Total

Análise de variância (ANOVA) com um fator

Replicação	Tratamento				
	1	2	...	g	
1	y_{11}	y_{21}	...	y_{g1}	
2	y_{12}	y_{22}	...	y_{g2}	
...	
n	y_{1n}	y_{2n}	...	y_{gn}	
Soma	$y_{1.}$	$y_{2.}$...	$y_{g.}$	$y_{..} = \sum_i y_{i.}$
Média	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{g.}$	$\bar{y}_{..} = \frac{1}{g} \sum_i y_{i.}$

Soma de quadrados total:

$$SQ_{Tot} = \sum_{i=1}^g \sum_{j=1}^n (y_{ij} - \bar{y}_{..})^2$$

Graus de liberdade:

$$gl = N - 1$$

onde: $N = ng$

Cálculo de SQ_{total} : soma dos quadrados total

SD	DG	CD	SD	DG	CD
$(7,86-8,02)^2$	$(6,20-8,02)^2$	$(9,67-8,02)^2$	0,03	3,33	2,71
$(6,3-8,02)^2$	$(7,82-8,02)^2$	$(8,08-8,02)^2$	2,71	0,04	0,00
6,90,,,	8,50,,,	9,25,,,	1,27	0,23	1,50
7,78,,,	6,50,,,	8,20,,,	0,06	2,33	0,03
8,17,,,	8,09,,,	8,64,,,	0,02	0,00	0,38
6,26,,,	6,90,,,	9,67,,,,	+ 3,12	+ 1,27	2,71 +
6,30,,,	7,82,,,	9,23,,,,	2,98	0,04	1,45
7,86,,,	7,45,,,	10,43,,,,	0,03	0,33	5,78
7,42,,,	7,75,,,	9,97,,,,	0,37	0,08	3,78
8,63,,,	7,43,,,	9,59,,,,	0,37	0,35	2,45
Média geral= 8,02			10,93	8,00	20,79
$SQ_{total}=39,72$					

Soma de Quadrados dos Tratamentos (grupos)

Análise de variância (ANOVA) com um fator

Replicação	Tratamento				
	1	2	...	g	
1	y_{11}	y_{21}	...	y_{g1}	
2	y_{12}	y_{22}	...	y_{g2}	
...	
n	y_{1n}	y_{2n}	...	y_{gn}	
Soma	$y_{1.}$	$y_{2.}$...	$y_{g.}$	$y_{..} = \sum_i y_{i.}$
Média	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{g.}$	$\bar{y}_{..} = \frac{1}{g} \sum_i y_{i.}$

Soma de quadrados dos tratamentos:

$$SQ_{Trat} = \sum_{i=1}^g \sum_{j=1}^n (\bar{y}_{i.} - \bar{y}_{..})^2 = n \sum_{i=1}^g (\bar{y}_{i.} - \bar{y}_{..})^2$$

Graus de liberdade:

$$gl = g - 1$$

	SD	DG	CD
	7,86	6,20	9,67
	6,38	7,82	8,08
	6,90	8,50	9,25
	7,78	6,50	8,20
	8,17	8,09	8,64
	6,26	6,90	9,67
	6,30	7,82	9,23
	7,86	7,45	10,43
	7,42	7,75	9,97
	8,63	7,43	9,59
Média grupos	7,36	7,45	9,27

média geral = 8,02

$$SQ_{\text{trat}} = 10 \cdot (7,36 - 8,02)^2 + 10 \cdot (6,20 - 8,02)^2 + 10 \cdot (9,67 - 8,02)^2 = 23,40$$

Soma de Quadrados do Erro

Análise de variância (ANOVA) com um fator

Replicação	Tratamento				
	1	2	...	g	
1	y_{11}	y_{21}	...	y_{g1}	
2	y_{12}	y_{22}	...	y_{g2}	
...	
n	y_{1n}	y_{2n}	...	y_{gn}	
Soma	$y_{1.}$	$y_{2.}$...	$y_{g.}$	$y_{..} = \sum_i y_{i.}$
Média	$\bar{y}_{1.}$	$\bar{y}_{2.}$...	$\bar{y}_{g.}$	$\bar{y}_{..} = \frac{1}{g} \sum_i y_{i.}$

Soma de quadrados do erro:

$$SQ_{Erro} = \sum_{i=1}^g \sum_{j=1}^n (y_{ij} - \bar{y}_{i.})^2$$

Graus de liberdade:

$$gl = N - g$$

Soma de Quadrados dos Erros

SD	DG	CD	SD	DG	CD
$(7,86-7,36)^2$	$(6,20-7,45)^2$	$(9,67-9,27)^2$	0,25	1,55	0,16
$(6,3-7,36)^2$	$(7,82-7,45)^2$	$(8,08-9,27)^2$	0,95	0,14	1,42
$(6,90-7,36)^2$	$(8,50-7,45)^2$	$(9,25-9,27)^2$	0,21	1,11	0,00
$(7,78-7,36)^2$	$(6,50-7,45)^2$	$(8,20-9,27)^2$	0,18	0,89	1,15
$(8,17-7,36)^2$	$(8,09-7,45)^2$	$(8,64-9,27)^2$	0,66 +	0,41 +	0,40 +
$(6,26-7,36)^2$	$(6,90-7,45)^2$	$(9,67-9,27)^2$	1,20	0,30	0,16
$(6,30-7,36)^2$	$(7,82-7,45)^2$	$(9,23-9,27)^2$	1,12	0,14	0,00
$(7,86-7,36)^2$	$(7,45-7,45)^2$	$(10,43-9,27)^2$	0,25	0,00	1,34
$(7,42-7,36)^2$	$(7,75-7,45)^2$	$(9,97-9,27)^2$	0,00	0,09	0,49
$(8,63-7,36)^2$	$(7,43-7,45)^2$	$(9,59-9,27)^2$	1,62	0,00	0,10
7,36	7,45	9,27	6,45	4,64	5,22

$$\text{SQR} = 6,45 + 4,64 + 5,22 = 16,32$$

Exemplo 3

$$F = \frac{\frac{SQTr}{k-1}}{\frac{SQR}{n-k}}$$
$$F = \frac{\frac{23,40}{(3-1)}}{\frac{16,32}{(30-3)}} = 19,5$$

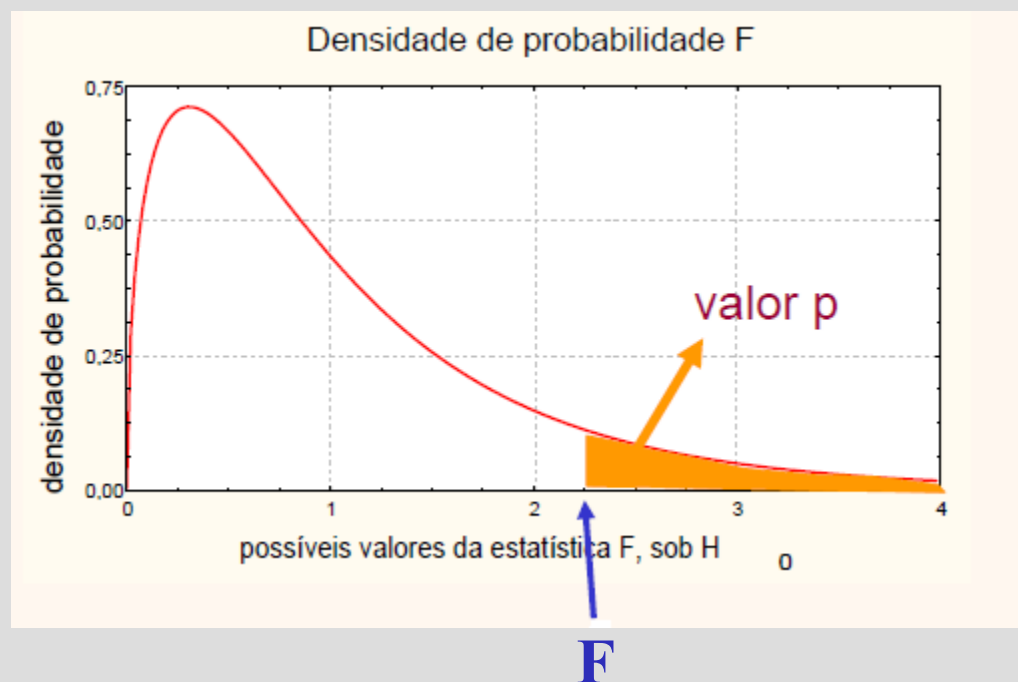
graus de liberdade entre tratamentos (numerador) = $(3-1) = 2$
graus de liberdade intra tratamentos (denominador) = $(30-3) = 27$

Agora precisamos comparar este valor com o valor obtido para o nível de significância sob 2 graus de liberdade no numerador e 27 no denominador, que vai fornecer:

$$F_{\alpha=0,05 ; 2,27} = 3,35$$

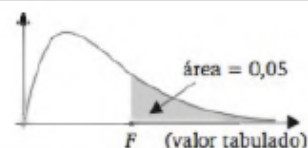
Teste F

A estatística F tem distribuição F com $(g-1)$ graus de liberdade no numerador e $(N-g)$ graus de liberdade no denominador *



* Consultar tabelas de valores padronizados para a distribuição F. Ver tabela VI (Bussab/Morettin); tabela 6 (Barbetta).

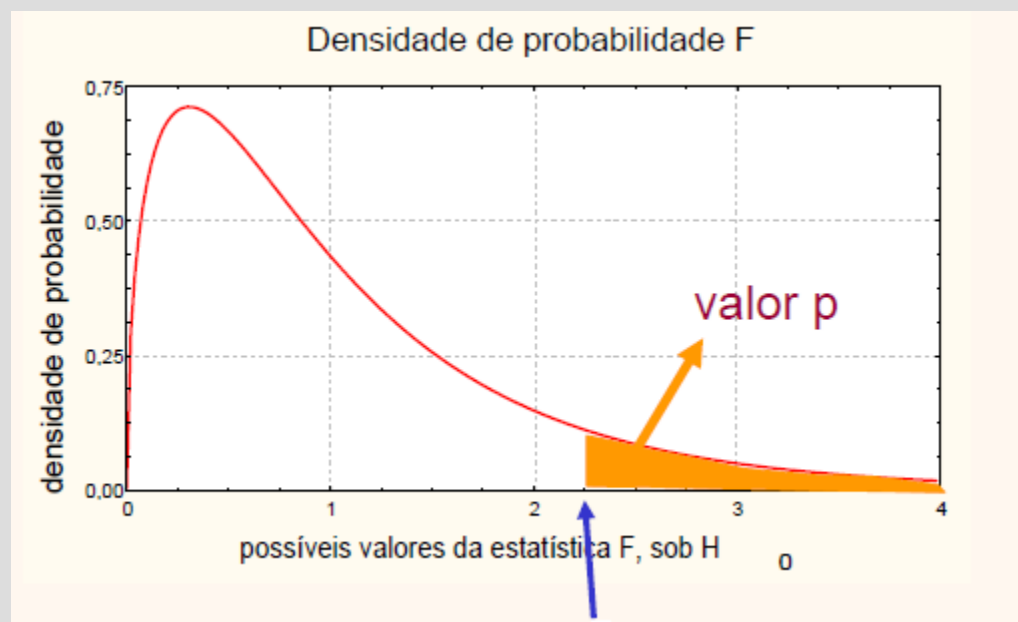
Tabela 6 (Continuação).
 $\alpha = 0,05$



gl denom.	Graus de liberdade no numerador									
	1	2	3	4	5	6	7	8	9	10
1	161,45	199,50	215,71	224,58	230,16	233,99	236,77	238,88	240,54	241,88
2	18,51	19,00	19,16	19,25	19,30	19,33	19,35	19,37	19,38	19,40
3	10,13	9,55	9,28	9,12	9,01	8,94	8,89	8,85	8,81	8,79
4	7,71	6,94	6,59	6,39	6,26	6,16	6,09	6,04	6,00	5,96
5	6,61	5,79	5,41	5,19	5,05	4,95	4,88	4,82	4,77	4,74
6	5,99	5,14	4,76	4,53	4,39	4,28	4,21	4,15	4,10	4,06
7	5,59	4,74	4,35	4,12	3,97	3,87	3,79	3,73	3,68	3,64
8	5,32	4,46	4,07	3,84	3,69	3,58	3,50	3,44	3,39	3,35
9	5,12	4,26	3,86	3,63	3,48	3,37	3,29	3,23	3,18	3,14
10	4,96	4,10	3,71	3,48	3,33	3,22	3,14	3,07	3,02	2,98
11	4,84	3,98	3,59	3,36	3,20	3,09	3,01	2,95	2,90	2,85
12	4,75	3,89	3,49	3,26	3,11	3,00	2,91	2,85	2,80	2,75
13	4,67	3,81	3,41	3,18	3,03	2,92	2,83	2,77	2,71	2,67
14	4,60	3,74	3,34	3,11	2,96	2,85	2,76	2,70	2,65	2,60
15	4,54	3,68	3,29	3,06	2,90	2,79	2,71	2,64	2,59	2,54
16	4,49	3,63	3,24	3,01	2,85	2,74	2,66	2,59	2,54	2,49
17	4,45	3,59	3,20	2,96	2,81	2,70	2,61	2,55	2,49	2,45
18	4,41	3,55	3,16	2,93	2,77	2,66	2,58	2,51	2,46	2,41
19	4,38	3,52	3,13	2,90	2,74	2,63	2,54	2,48	2,42	2,38
20	4,35	3,49	3,10	2,87	2,71	2,60	2,51	2,45	2,39	2,35
21	4,32	3,47	3,07	2,84	2,68	2,57	2,49	2,42	2,37	2,32
22	4,30	3,44	3,05	2,82	2,66	2,55	2,46	2,40	2,34	2,30
23	4,28	3,42	3,03	2,80	2,64	2,53	2,44	2,37	2,32	2,27
24	4,26	3,40	3,01	2,78	2,62	2,51	2,42	2,36	2,30	2,25
25	4,24	3,39	2,99	2,76	2,60	2,49	2,40	2,34	2,28	2,24
26	4,23	3,37	2,98	2,74	2,59	2,47	2,39	2,32	2,27	2,22
27	4,21	3,35	2,96	2,73	2,57	2,46	2,37	2,31	2,25	2,20
28	4,20	3,34	2,95	2,71	2,56	2,45	2,36	2,29	2,24	2,19

Teste F

A estatística F tem distribuição F com (g-1) graus de liberdade no numerador e (N-g) graus de liberdade no denominador *



F

$$F_{\alpha=0,05; 2,27} = 3,35 < F=19,5$$

Exemplo 3

$$F = \frac{\frac{SQTr}{k-1}}{\frac{SQR}{n-k}}$$
$$F = \frac{\frac{23,40}{(3-1)}}{\frac{16,32}{(30-3)}} = 19,5$$

graus de liberdade entre tratamentos (numerador) = $(3-1) = 2$
graus de liberdade intra tratamentos (denominador) = $(30-3) = 27$
 $F_{\alpha=0,05; 2,27} = 3,35$

Conclusão: há evidências, a um nível de significância de 5%, que as médias dos grupos não são iguais, ou seja, rejeita-se H_0 .

Referências

- Hair Jr., J. F; *et al.* Análise Multivariada de Dados. 6^a ed. Porto Alegre: Bookman, 2009.
- Créditos: Prof^a. Ana Amélia Benedito Silva (*slides*)

Fim