

# Análise da Variância (ANOVA)

A Análise da Variância (**AN**alysis **Of** **VA**riance) é procedimento para testar a hipótese de que **três ou mais médias populacionais são iguais** de modo que:

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

$H_1$ : pelo menos uma das médias é diferente das outras.

O teste é feito com base na análise das variâncias amostrais.

Utiliza-se a distribuição  $F$ .

# ANOVA de um fator

---

Um tratamento (ou fator) é uma propriedade, ou característica, que nos permite distinguir as diferentes populações umas das outras.

Estatística teste:

$$F_{teste} = \frac{\text{variância entre amostras}}{\text{variância dentro das amostras}}$$

# ANOVA de um fator

Cálculos com **tamanhos amostrais iguais**

$$F_{teste} = \frac{n s_{\bar{x}}^2}{s_p^2}$$

$s_{\bar{x}}^2$  é a **variância das médias** amostrais

$s_p^2$  é a **média das variâncias** amostrais

**Graus de liberdade:**

Numerador:  $gl_1 = k - 1$

Denominador:  $gl_2 = k (n - 1)$

k = número de amostras

n = tamanho amostral

# ANOVA de um fator

Cálculos com **tamanhos amostrais diferentes**

$$F_{teste} = \frac{\left[ \frac{\sum_{i=1}^k n_i (\bar{x}_i - \bar{\bar{x}})^2}{k-1} \right]}{\left[ \frac{\sum_{i=1}^k (n_i - 1) s_i^2}{\sum_{i=1}^k (n_i - 1)} \right]}$$

$\bar{\bar{x}}$  é a média de todos os valores amostrais combinados

## Grau de Liberdade

Numerador:  $gl_1 = k - 1$

Denominador:  $gl_2 = N - k$

$$N = \sum n_i$$

## Componentes-chave:

SQ(tratamento) Soma de quadrados  
SQ(erro) Representa a variação

MQ(...) = SQ(...)/número: **M**édia **Q**uadrática

# ANOVA de um fator

## Exemplo com **tamanhos amostrais diferentes**

Níveis de álcool no sangue após o consumo de cinco doses da bebida.

Grupos de homens

A	B	C
0,11	0,08	0,04
0,10	0,09	0,04
0,09	0,07	0,05
0,09	0,07	0,05
0,10	0,06	0,06
		0,04
		0,05

Os integrantes do grupo A foram testados uma hora depois; do grupo B, duas horas depois e do grupo C quatro horas depois

$n_i$   
 $\bar{x}_i$   
 $s_i^2$   
 $N$   
 $k$   
 $\bar{x}$

# ANOVA de dois fatores

Cálculos individuais em categorias de acordo com **dois fatores**. Em outras palavras: os valores amostrais são categorizados de duas maneiras. Ex. Na corrida de NY:

Fatores: idade e sexo.

Tempo (s) para corredores da Maratona NY

		Idade		
		21-29	30-39	40 ou mais
Sexo	Masculino	13615	14677	14528
		18784	16090	17034
		14256	14086	14935
		10905	16460	14996
		12077	20808	22146
	Feminino	16401	15357	17260
		14216	16771	25399
		15402	15036	18647
		15326	16297	15077
		12047	17636	25898

# ANOVA de dois fatores

Cálculos individuais em categorias de acordo com **dois fatores**. Em outras palavras, os valores amostrais são categorizados de duas maneiras. Ex. Na corrida de NY:

Fatores: idade e sexo. Subcategorias (células), neste caso seis células

Tempo (s) para corredores da Maratona NY

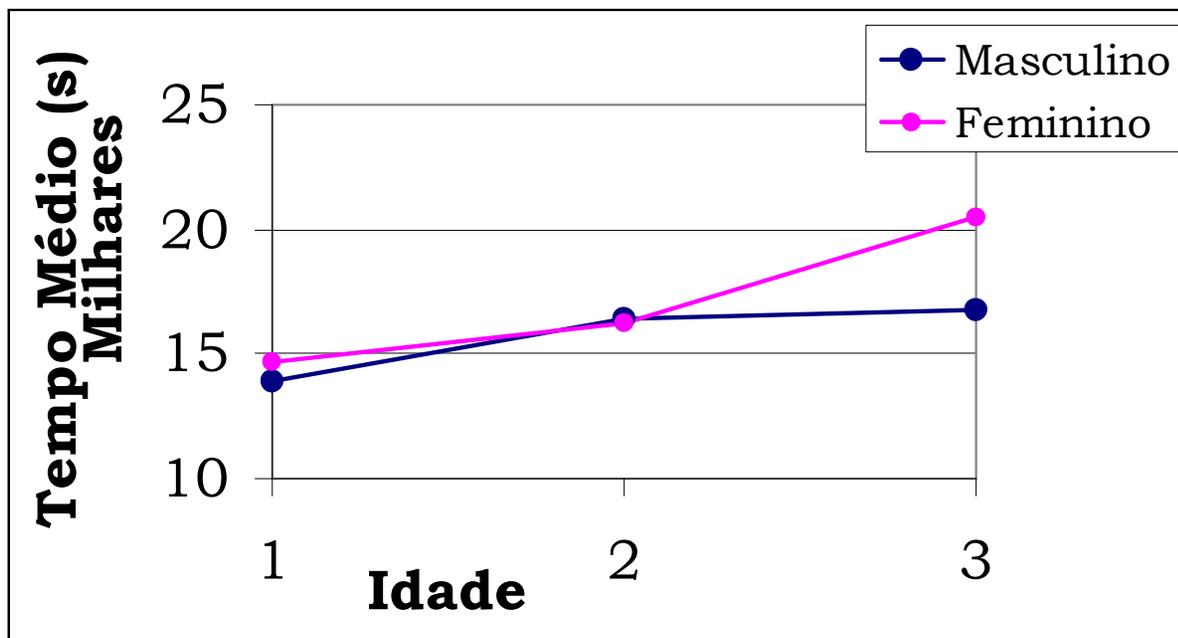
		Idade		
		21-29	30-39	40 ou mais
Sexo	Masculino	①	②	③
	Feminino	④	⑤	⑥

# ANOVA de dois fatores

Procedimento para o cálculo:

Passo 1. Efeito da interação: Teste  $H_0$ , de que não há interação entre os dois fatores. Um procedimento é o gráfico de interação de médias de células.

Verifique a estatística teste



$$F = \frac{MQ(\text{interação})}{MQ(\text{Erro})}$$

Se rejeitar  $H_0$ , para por aqui.

# ANOVA de dois fatores

$$\bar{\bar{x}} = \frac{\sum_{i=1}^L \sum_{j=1}^C \sum_{k=1}^{n'} x_{ijk}}{LCn'}$$

Média global  $n' =$  número de repetições em uma célula

$$\bar{x}_{i..} = \frac{\sum_{j=1}^C \sum_{k=1}^{n'} x_{ijk}}{Cn'}$$

Média do fator  $i$  (fatores em linhas)

$$\bar{x}_{.j.} = \frac{\sum_{i=1}^L \sum_{k=1}^{n'} x_{ijk}}{Ln'}$$

Média do fator  $j$  (fatores em colunas)

$$\bar{x}_{ij.} = \frac{\sum_{k=1}^{n'} x_{ijk}}{n'}$$

Média da célula  $i,j$

# ANOVA de dois fatores

Varição Total  $SQ(Total) = \sum_{i=1}^L \sum_{j=1}^C \sum_{k=1}^{n'} (x_{ijk} - \bar{\bar{x}})^2$

Varição Fator A  $SQ(L) = Cn' \sum_{i=1}^L (\bar{x}_{i..} - \bar{\bar{x}})^2$

Varição Fator B  $SQ(C) = Ln' \sum_{j=1}^C (\bar{x}_{.j.} - \bar{\bar{x}})^2$

Varição Decorrente da Interação

$$SQ(Interac\tilde{a}\tilde{o}) = n' \sum_{i=1}^L \sum_{j=1}^C (\bar{x}_{ij.} - \bar{x}_{i..} - \bar{x}_{.j.} + \bar{\bar{x}})^2$$

Erro  $SQ(Erro) = \sum_{i=1}^L \sum_{j=1}^C \sum_{k=1}^{n'} (x_{ijk} - \bar{x}_{ij.})^2$

# ANOVA de dois fatores

Anova: fator duplo com repetição

RESUMO	21-29	30-39	40 ou mais	Total
<i>Masculino</i>				
Contagem	5	5	5	15
Soma	69637	82121	83639	235397
Média	13927,4	16424,2	16727,8	15693,1
Variância	9087754,3	6962640,2	10125758,2	9165617,8
<i>Feminino</i>				
Contagem	5	5	5	15
Soma	73392	81097	102281	256770
Média	14678,4	16219,4	20456,2	17118
Variância	2762103,3	1115302,3	24117287,7	14392308,6

<i>Total</i>				
Contagem	10	10	10	
Soma	143029	163218	185920	
Média	14302,9	16321,8	18592	
Variância	5423270,3	3601847,5	19080511,1	

ANOVA

Fonte da variação	SQ	gl	MQ	F	Valor-P	F crítico
Amostra	15226837,6	1	15226837,6	1,69	0,21	4,26
Colunas	92087146,9	2	46043573,4	5,10	0,01	3,40
Interações	21040438,9	2	10520219,4	1,17	0,33	3,40
Dentro	216683384,0	24	9028474,3			
Total	345037807,4	29				

Tempo (s) para corredores da Maratona NY

		Idade		
		21-29	30-39	40 ou mais
Sexo	Masculino	13615	14677	14528
		18784	16090	17034
		14256	14086	14935
	Feminino	10905	16460	14996
		12077	20808	22146
		16401	15357	17260
		14216	16771	25399
		15402	15036	18647
		15326	16297	15077
		12047	17636	25898

MQ(sexo)  
MQ(idade)  
MQ(interação)  
MQ(erro)

**Excel:**

Ferramenta >

Análise de dados >

ANOVA Fator Duplo com repetição

# ANOVA de dois fatores

Procedimento para o cálculo (continuação)

Passo 2. Efeitos de **Linha**/Coluna

Linha: Teste  $H_0$ , “Não há qualquer efeito do fator linha” (As médias das linhas são iguais)

$$F = MQ(\text{sexo}) / MQ(\text{erro})$$

ANOVA								
<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>		
Sexo Amostra	15226837,6	1	15226837,6	1,69	0,21	4,26		
Idade Colunas	92087146,9	2	46043573,4	5,10	0,01	3,40		
Interação Interações	21040438,9	2	10520219,4	1,17	0,33	3,40		
Erro Dentro	216683384,0	24	9028474,3					
Total	345037807,4	29						

# ANOVA de dois fatores

Procedimento para o cálculo (continuação)

Passo 2. Efeitos de Linha/Coluna

Linha: Teste  $H_0$ , “Não há qualquer efeito do fator coluna” (As médias das colunas são iguais)

$$F = MQ(\text{idade}) / MQ(\text{erro})$$

ANOVA							
<i>Fonte da variação</i>	<i>SQ</i>	<i>gl</i>	<i>MQ</i>	<i>F</i>	<i>valor-P</i>	<i>F crítico</i>	
Sexo Amostra	15226837,6	1	15226837,6	1,69	0,21	4,26	
Idade Colunas	92087146,9	2	46043573,4	5,10	0,01	3,40	
Interação Interações	21040438,9	2	10520219,4	1,17	0,33	3,40	
Erro Dentro	216683384,0	24	9028474,3				
<b>Total</b>	<b>345037807,4</b>	<b>29</b>					

# ANOVA de dois fatores

---

## Conclusão do exemplo:

Com base nos dados amostrais, concluimos que os tempos parecem ter médias desiguais para diferentes categorias de idade, mas os tempos parecem ter médias iguais para ambos os sexos.