



Prof^a.

Elaine Marques

Monitora:

Macileide Oliveira

Estatística Paramétrica e Não-Paramétrica com uso do software R



O que é ANOVA

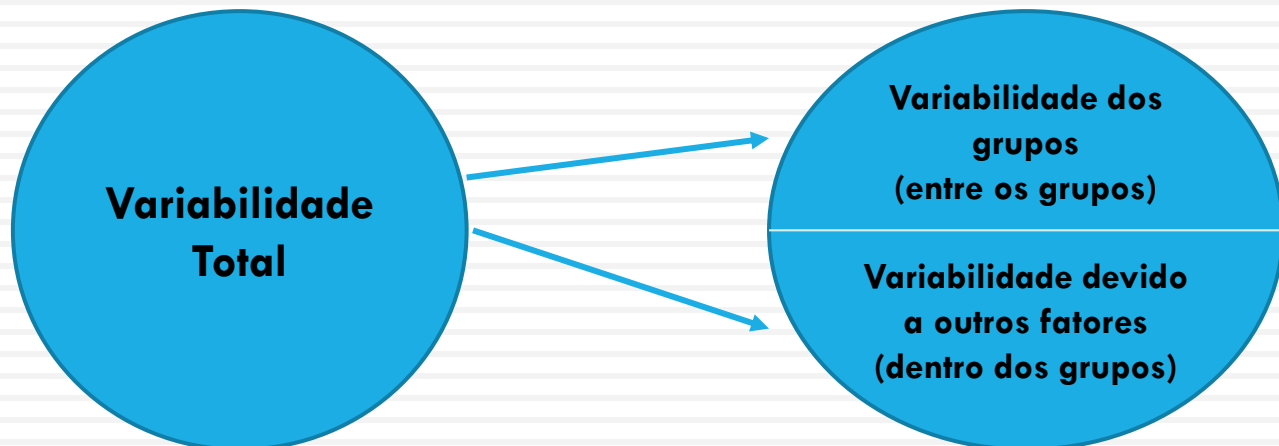
A Análise de Variância - ANOVA é um teste paramétrico utilizado para testar a igualdade de três ou mais médias populacionais, baseado na análise das variâncias amostrais.

Os dados amostrais são separados em grupos segundo uma característica (fator).



Para que serve a ANOVA?

A ANOVA é comumente utilizada para descrever um modelo de regressão linear, ou seja, serve para avaliar a qualidade do ajuste do modelo de regressão linear.





Como utiliza a ANOVA?

Como saber quais marcas de remédios apresenta menor tempo para fazer efeito?



Amostras do tempo de efeito para 4 tipos de antidepressivos, de marcas diferentes.



Teste de Hipótese

- H_0 (hipótese nula): As médias populacionais são iguais.
- H_1 (hipótese alternativa): Pelo menos uma das médias populacionais é diferente das demais.

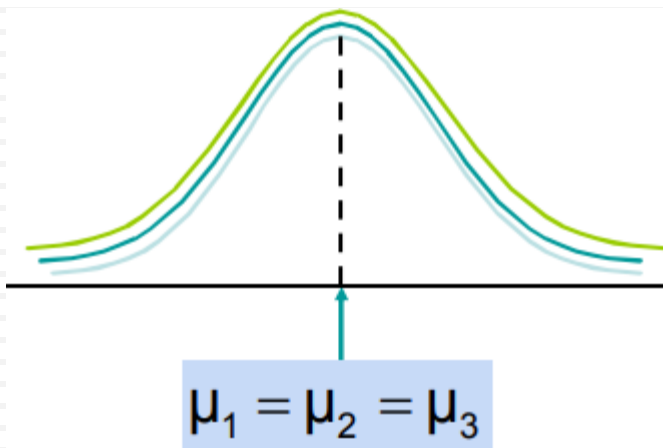
$$H_0: \mu_1 = \mu_2 = \mu_3 = \dots = \mu_k$$

H_1 : Nem todas as médias populacionais são iguais.



Teste de Hipótese

Interpretação: Não Rejeita H_0 (Hipótese Nula) e Rejeita H_1 (hipótese alternativa).



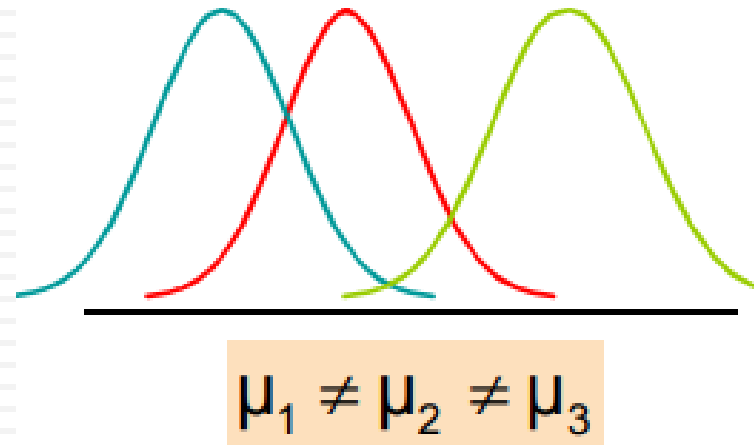
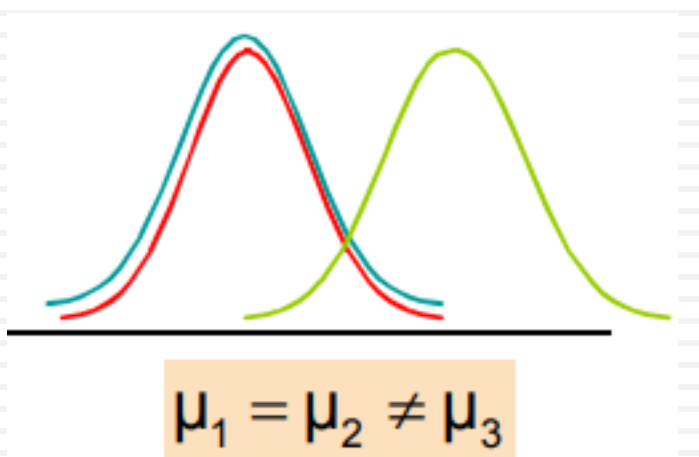
Todas as médias são iguais:
 H_0 é verdadeira
(Sem efeito do tratamento)



Teste de Hipótese

Interpretação: Rejeita H_0 (Hipótese Nula) e Não Rejeita H_1 (hipótese alternativa).

Pelo menos uma média é diferente: H_0 não é verdadeira
(Existe efeito do tratamento)





ANOVA One –Way (um fator)

PRESSUPOSTOS E SUPOSIÇÕES:

- Populações normalmente distribuída;
- Populações tem mesma variância (ou mesmo desvio padrão);
- Amostras são aleatórias e mutuamente independentes;
- As diferentes amostras são obtidas de populações classificadas em apenas uma categoria.



Análise de Variância

- Teste de Hipótese;
- Usada para comparar 3 ou mais grupos;
- Possui uma variável quantitativa e uma ou mais variáveis categóricas;
- Realiza comparações em todo o conjunto de dados;
- Procura a variação entre os grupos e compara a variação “dentro” dos grupos.



Exemplo de aplicação no R

Suponha que um farmacêutico realizou um experimento para verificar o tempo de efeito de 4 tipos de marcas de antidepressivo (A, B, C e D).





Exemplo de aplicação no R

Ele suspeita que o tempo de efeito do medicamento é decorrente dos diferentes tipos de marcas.

Ou seja, existe um medicamento que apresenta um tempo médio de efeito melhor que os demais. O tempo de efeito em cada tipo de marca foi a seguinte:



Exemplo de aplicação no R

O tempo de efeito em cada tipo de marca foi a seguinte:

A	B	C	D
25	31	22	33
26	25	26	29
20	28	28	31
23	27	25	34
21	24	29	28
23	27	26	31



Exemplo de aplicação no R

Temos as seguintes hipóteses formuladas:

H0: Não há diferença significativa entre as diferentes marcas de remédio em relação ao tempo médio que se leva pra fazer efeito.

H1: Existe diferença significativa entre as diferentes marcas de remédio em relação ao tempo médio que se leva pra fazer efeito.



ANOVA One Way NO R

Verificando o diretório de trabalho:

```
# Utilize a função getwd para verificar o diretório de trabalho:
getwd()
# Utilize a função setwd para mudar o seu diretório de trabalho:
setwd('C:/Users/Elaine/Desktop/CursoAnova')

# Usando seus dados no R
# Se estiver no formato .csv, use:
medicamento <- read.csv("anova1.csv", sep=";", header=T)
medicamento

# Se estiver no formato .txt, use:
# medicamento2 <- read.delim("anova.txt")

# OBS: lembre-se de especificar o argumento sep = ";"
# Caso a configuração do seu computador esteja com separador decimal como ",",
# chame a tabela da seguinte maneira no R:
# medicamento3 <- read.csv2("anova.csv", dec = ",")

# Visualize sua planilha de dados
View(medicamento)
```



ANOVA One Way NO R

A função do R que executa a ANOVA são as funções **aov** e **lm**. Para este exemplo utilizamos a função **aov**.

```
#A função do R que executa a ANOVA é a aov.  
# exemplo: aov(V.Dependente ~ V.Independente , data=dados)  
modeloanova <- aov(HORAS ~ REMEDIO, data =medicamento)  
modeloanova  
> modeloanova <- aov(HORAS ~ REMEDIO, data =medicamento)  
> modeloanova  
Call:  
  aov(formula = HORAS ~ REMEDIO, data = medicamento)  
  
Terms:  
                REMEDIO Residuals  
Sum of Squares    163.75     112.00  
Deg. of Freedom         3         16  
  
Residual standard error: 2.645751  
Estimated effects may be unbalanced
```



ANOVA One Way NO R

Com a função *summary* conseguimos mais informações:

```
# Função summary fornece mais informações em relação ao modelo  
summary(modeloanova)
```

```
> summary(modeloanova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)
REMEDIO	3	163.8	54.58	7.798	0.00198 **
Residuals	16	112.0	7.00		

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> |
```




Montando a tabela da ANOVA

One Way

Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrado Médio	Estatística F	Valor -P
Entre	k-1	SQE	$QME = SQE / (k-1)$	$F = QME / QMD$	$P(>F)$
Dentro	n-k	SQD	$QMD = SQD / (n-k)$		
Total	n-1	$SQT = SQE + SQD$	$QMT = SQT / (n-1)$		

Notação:

Fonte de Variação (FV): Descrição das diferentes somas de quadrados;

Graus de liberdade (GL): k= número de amostras; n = soma do número de elementos de todas as amostras;

SQE: Soma de Quadrados Entre grupos ou tratamentos: é uma medida da variação entre as médias amostrais combinados;

SQD: Soma de Quadrados Dentro dos grupos ou soma de quadrados dos resíduos;

SQT= Soma de Quadrados Total: é uma medida da variação total(em torno de x) em todos os dados amostrais combinados;



Montando a tabela da ANOVA

One Way

Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrado Médio	Estatística F	Valor -P
Entre	k-1	SQE	$QME = SQE / (k-1)$	$F = QME / QMD$	$P(>F)$
Dentro	n-k	SQD	$QMD = SQD / (n-k)$		
Total	n-1	$SQT = SQE + SQD$	$QMT = SQT / (n-1)$		

Notação:

QME: Quadrado médio entre: Variação entre as amostras;

QMD: Quadrado médio dentro: Variação dentro das amostras;

QMT: Quadrado médio total: é uma medida da variação total(em torno de x) em todos os dados amostrais combinados;

Estatística F: Teste F

Valor-p: valor-p para a estatística F.



Montando a tabela da ANOVA One Way

```
> summary(modeloanova)
```

```
      Df Sum Sq Mean Sq F value Pr(>F)
REMEDIO  3  163.8   54.58   7.798 0.00198 **
Residuals 16  112.0    7.00
```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
> |
```

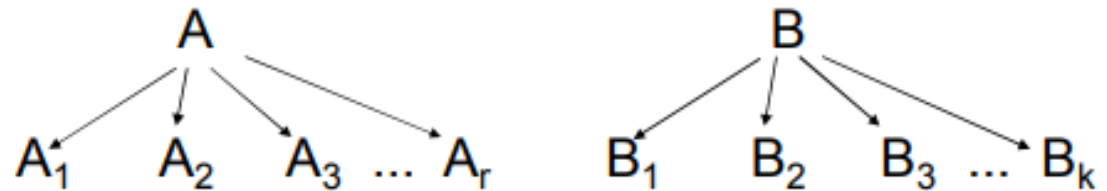
Fonte de Variação	Graus de liberdade	Soma de Quadrados	Quadrado Médio	Estatística F	Valor -P
Entre	3	163,8	54,58	7,798	0.00198
Dentro	16	112,0	7,00		
Total	19	275,75			



Anova Two Way – Dois fatores

SEXO	REMEDIO	HORAS
F	A	25
F	A	26
F	A	20
F	A	23
F	A	21
M	B	31
M	B	25
M	B	28
M	B	27
M	B	24
F	C	22
F	C	26
F	C	28
F	C	25
F	C	29
M	D	33
M	D	29
M	D	31
M	D	34
M	D	28

Dois fatores de interesse: A e B com vários níveis (categorias).



(A) O **fator marcas de antidepressivo**, que contém 4 categorias: A(A1), B(A2), C(A3) e D(A4).

(B) O **fator Sexo**, que contém duas categorias: Feminino (B1) e Masculino(B2).



Anova de Medidas Repetidas

A ANOVA de medidas repetidas é um teste estatístico para a análise de dados longitudinais pareados.

Esta técnica pode ser entendida como uma expansão da ANOVA ou um caso especial do Modelo Linear de Efeitos Mistos (LMM).



Anova de Medidas Repetidas

Os pressupostos deste teste são próximos aos discutidos em outros testes inferenciais:

- (i) Os dados são aleatórios e representativos da população
- (ii) A variável dependente é contínua
- (iii) Os resíduos do modelo são normalmente distribuídos
- (iv) Há esfericidade dos grupos



Montando a tabela da ANOVA com Medidas Repetidas

Organização das unidades experimentais em um planejamento com $k=4$ Juízes (itens) e $r = 6$ indivíduos.

	Repetição	Tratamentos ou medidas (j)				Total (i)	Média
		Juiz 1	Juiz 2	Juiz 3	Juiz 4		
Indivíduos (i)	1	2	4	3	3	12	3,00
	2	5	7	5	6	23	5,75
	3	1	3	1	2	7	1,75
	4	7	9	9	8	33	8,25
	5	2	4	6	1	13	3,25
	6	6	8	8	4	26	6,50
Total (j)		23	35	32	24	114	28,50

ANOVA com uma classificação com medidas repetidas.

Fonte de Variação	Soma de Quadrados	Graus de liberdade	Quadrado Médio	Razão F	Valor-P
Entre indivíduos (L)	122,50	5	24,50	19,92	3,85482E-06
Intra indivíduos (I)	36,00	18	2,00	1,63	0,17314376
Juiz (tratamento)	17,50	3	5,83	4,74	0,016122571
Resíduos	18,50	15	1,23	1,00	
Total	158,50	23	-		



Montando a tabela da ANOVA com Medidas Repetidas

Fonte de Variação	Soma de Quadrados	Graus de liberdade	Quadrado Médio	Razão F	Valor-P
Entre indivíduos (L)	SQL	$r-1$	QML	F_l	$P(F > F_l)$
Intra indivíduos (I)	SQI	$r(k-1)$	QMI	F_i	$P(F > F_i)$
Itens (tratamento)	SQA	$k-1$	QMA	F_a	$P(F > F_a)$
Resíduos	SQE	$(r-1)(k-1)$	QME		
Total	SQT	$rk-1$			

Fonte de Variação	Soma de Quadrados	Graus de liberdade	Quadrado Médio	Razão F	Valor-P
Entre indivíduos (L)	122,50	5	24,50	19,92	3,85482E-06
Intra indivíduos (I)	36,00	18	2,00	1,63	0,17314376
Juiz (tratamento)	17,50	3	5,83	4,74	0,016122571
Resíduos	18,50	15	1,23	1,00	
Total	158,50	23	-		



ANOVA com Medidas Repetidas NO R

```
#ANOVA DE MEDIDAS REPETIDAS
```

```
avalia<-c(2,5,1,7,2,6,4,7,3,9,4,8,3,5,1,9,6,8,3,6,2,8,1,4)
```

```
sujeito<-c(1:6, 1:6, 1:6, 1:6)
```

```
juiz<-c(rep("1",6), rep("2",6) , rep("3",6), rep("4",6))
```

```
b<-lm(avalia ~as.factor(sujeito) + as.factor(juiz))
```

```
anova(b)
```

```
#Alpha de Cronbach
```

```
library(ltm)
```

```
juiz1<-c(2,5,1,7,2,6)
```

```
juiz2<-c(4,7,3,9,4,8)
```

```
juiz3<-c(3,5,1,9,6,8)
```

```
juiz4<-c(3,6,2,8,1,4)
```

```
dados<-data.frame(juiz1,juiz2,juiz3, juiz4)
```

```
cronbach.alpha(dados, standardized = FALSE, CI = TRUE,
```

```
probs = c(0.025, 0.975), B = 1000, na.rm = FALSE)
```



ANOVA com Medidas Repetidas NO R

Analysis of Variance Table

Response: avalia

	Df	Sum Sq	Mean Sq
as.factor(sujeito)	5	122.5	24.5000
as.factor(juiz)	3	17.5	5.8333
Residuals	15	18.5	1.2333

	F value	Pr(>F)
as.factor(sujeito)	19.8649	3.922e-06
as.factor(juiz)	4.7297	0.01624
Residuals		

as.factor(sujeito) ***

as.factor(juiz) *

Residuals

Signif. codes:

0 '***' 0.001 '**' 0.01 '*' 0.05
'.' 0.1 ' ' 1

Cronbach's alpha for the 'dados' data-set

Items: 4

Sample units: 6

alpha: 0.95

Bootstrap 95% CI based on 1000 samples

2.5% 97.5%

0.757 0.994



Dúvidas?

