



Escola Politécnica de Pernambuco
Especialização em Ciência de Dados e Analytics

Estatística Computacional

Aula 2.1 – Testes de Hipóteses e ANOVA – PARTE II

Prof. Dr. Rodrigo Lins Rodrigues

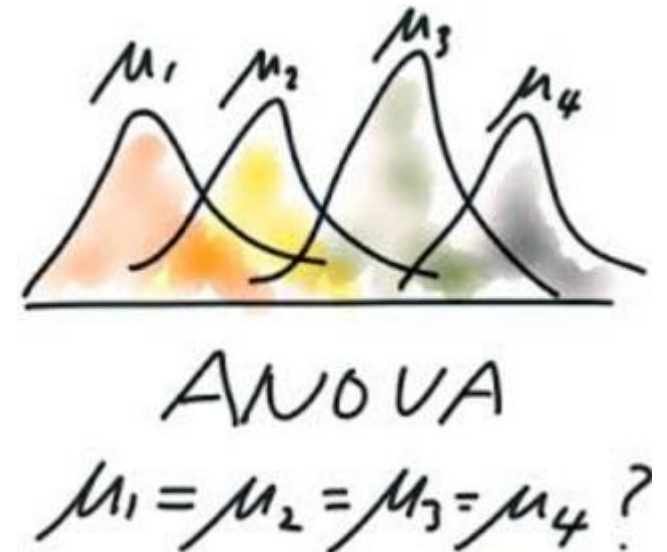
rodrigolins.rodrigues@ufrpe.br

A young man with dark hair, wearing a red and white striped shirt, is holding a magnifying glass over a chalkboard. The chalkboard is filled with various mathematical formulas, graphs, and diagrams, including a bar chart on the left and a line graph on the right. The man has a focused expression, looking through the magnifying glass.

Análise de Variância ANOVA

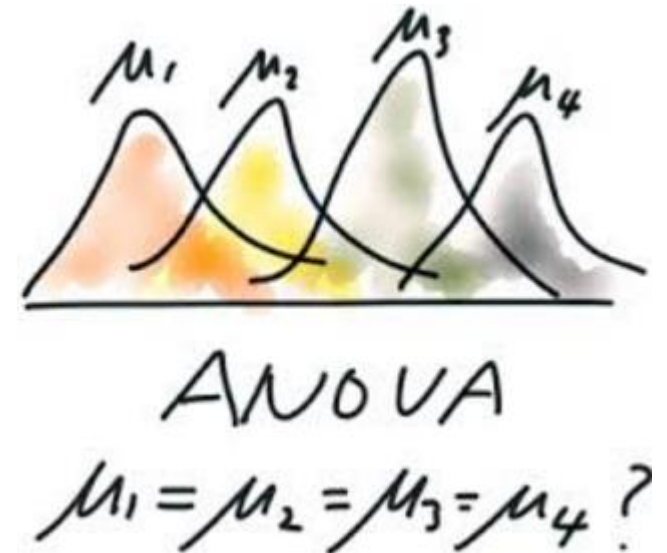
Análise de Variância ANOVA

- Todos os **testes de hipóteses** que vimos até agora trataram de no **máximo duas variáveis**;
- A análise de variância, ou *ANOVA*, é utilizada para comparar médias de **três ou mais variáveis**;



Análise de Variância ANOVA

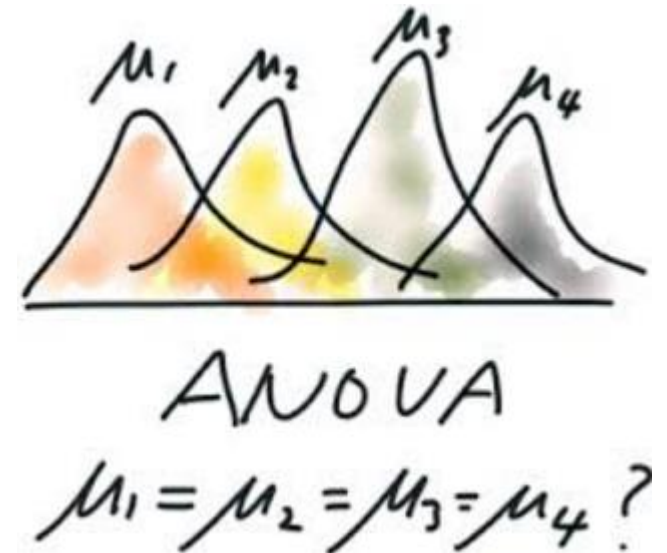
- É muito utilizada para identificar **diferenças** entre grupos;
- É empregada em situações que buscam **identificar a eficácia** de determinados tratamentos;
- É baseada no **teste *F***.



Análise de Variância ANOVA

- Compara o quanto os grupos diferem entre si em relação à quantidade de variabilidade dentro de cada grupo;
- Neste caso a hipótese formulada:

$$\left\{ \begin{array}{l} H_0: \mu_1 = \mu_2 \dots \mu_n \\ H_1: \text{pelo menos duas médias diferentes} \end{array} \right.$$



Análise de Variância ANOVA

- De forma genérica os dados para uma análise de ANOVA são:

Amostras ou Grupos			
1	2	...	k
y_{11}	y_{12}	...	y_{1k}
y_{21}	y_{22}	...	y_{2k}
...
y_{n1}	y_{n2}	...	y_{nk}

Análise de Variância ANOVA

- Etapas para análise de variância (ANOVA)

1. Verifique as condições de análise de variância utilizando os dados coletados de cada uma das k populações;
2. Estabeleça as hipóteses:

$$H_0: \mu_1 = \mu_2 \dots \mu_n$$

H_1 : pelo menos duas médias diferentes

3. Colete dados de k amostras aleatórias, uma de cada população.
4. Realize o *teste-F* para os dados do passo 3 e encontre o *p-valor*.
5. Se o *p-valor* for inferior a 0,05 conclui-se que pelo menos duas das médias populacionais são diferentes (H_1).

Análise de Variância ANOVA

- **Exemplo:**

- ✓ Uma amostra de 32 produtos foi coletada para analisar a qualidade do mel de três fornecedores;
- ✓ Uma das medidas de qualidade do mel é a porcentagem de sacarose que normalmente varia de 0,25 a 6,5%;
- ✓ A tabela a seguir apresenta a porcentagem de sacarose para a amostra coletada de cada um dos três fornecedores;
- ✓ É necessário verificar se há diferenças significativas entre os três fornecedores.

Análise de Variância ANOVA

- **Exemplo:**

Fornecedor I	Fornecedor II	Fornecedor III
0,33	1,54	1,47
0,79	1,11	1,69
1,24	0,97	1,55
1,75	2,57	2,04
0,94	2,94	2,67
2,42	3,44	3,07
1,97	3,02	3,33
0,87	3,55	4,01
0,33	2,04	1,52
0,79	1,67	2,03
1,24		
3,12		

Análise de Variância ANOVA

- **Passo 1:** Primeiramente verifica-se os pressupostos de normalidade para cada grupo;

Tests of Normality							
Fornecedor		Kolmogorov-Smirnov ^a			Shapiro-Wilk		
		Statistic	df	Sig.	Statistic	df	Sig.
Sacarose	1,00	,202	12	,189	,915	12	,246
	2,00	,155	10	,200*	,929	10	,438
	3,00	,232	10	,137	,883	10	,142

- **Passo 2:** Formulando a hipótese:

$$H_0: \mu_1 = \mu_2 \dots \mu_n$$

H_1 : pelo menos duas médias diferentes

Análise de Variância ANOVA

- **Passo 3:** Fixar o nível de significância em 5%;
- **Passo 4:** Realizar o calculo da estatística F_{cal} ;

$$F_{cal} = 4,676$$

- **Passo 5:** Verificar o valor na tabela F:

$$F_t = F_{2,29(5\%)} = 3,33$$

- **Passo 6:** Como o valor calculado é maior do que o valor tabelado ($F_{cal} > F_t$), a hipótese nula é rejeitada.

Agora é com vocês!

- Qual o objetivo em utilizar ANOVA ?
- Qual a diferença entre ANOVA e teste T;
- Como é formulada a hipótese para um teste de ANOVA?



Praticando ANOVA em R

Praticando ANOVA em R

- **Exemplo:**

- ✓ Quero comparar técnicas de extração de atributos para a classificação de imagens de peixes;
- ✓ Desejo verificar se existe diferença significativa entre as três técnicas, de acordo com os valores de desempenho de cada uma delas;

Praticando ANOVA em R

- **Solução:**

1. Verifica as condições de análise de variância;
2. Estabelece as hipóteses:

$$H_0: \mu_1 = \mu_2 \dots \mu_n$$

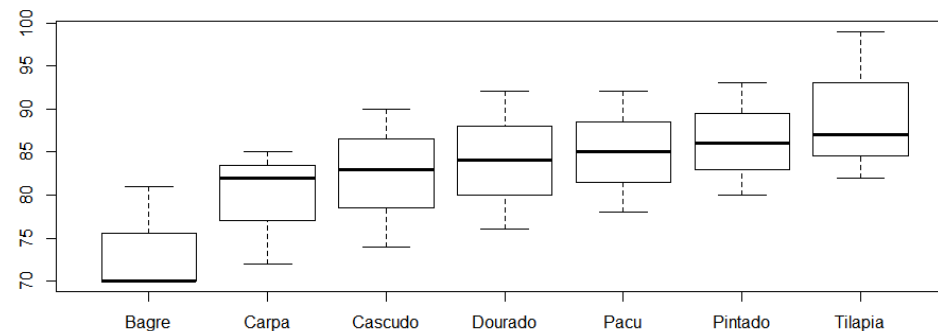
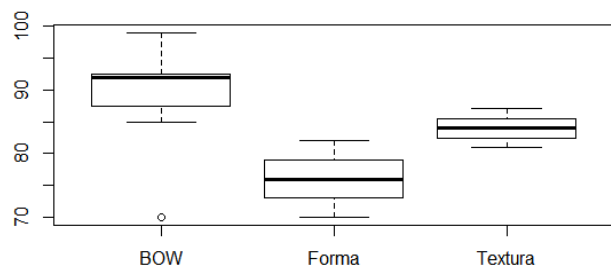
H_1 :: pelo menos duas médias diferentes

3. Colete dados de k amostras aleatórias, uma de cada população.
4. Realize o *teste-F* para os dados do passo 3 e encontre o *p*-valor.
5. Se o *p*-valor for inferior a 0,05 conclui-se que pelo menos duas das médias populacionais são diferentes (H_1).

Praticando ANOVA em R

- Solução:

```
3 # Ler os dados de um arquivo (interagindo com o usuário)
4 dados <- read.table(file.choose(),header=TRUE)
5
6 # Mostra boxplots das técnicas, lado a lado, em relação ao desempenho
7 boxplot(dados$desempenho ~ dados$tecnica)
8
9 # Mostra boxplots das classes, lado a lado, em relação ao desempenho
10 boxplot(dados$desempenho ~ dados$classe)
```



Praticando ANOVA em R

- Solução:

```
16 dados.anova <- aov(dados$desempenho ~ dados$tecnica + dados$classe)
17
18 # Mostra a tabela ANOVA
19 summary(dados.anova)
```

	Df	Sum Sq	Mean Sq	F value	Pr(>F)	
dados\$tecnica	2	578.4	289.19	18.562	0.000212	***
dados\$classe	6	464.5	77.41	4.969	0.008914	**
Residuals	12	187.0	15.58			

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

Dúvidas



- **Contatos:**

- ✓ Email: rodrigo.linsrodrigues@ufrpe.br
- ✓ Facebook: [/rodrigomuribec](https://www.facebook.com/rodrigomuribec)