



SHIFT

 FIAP





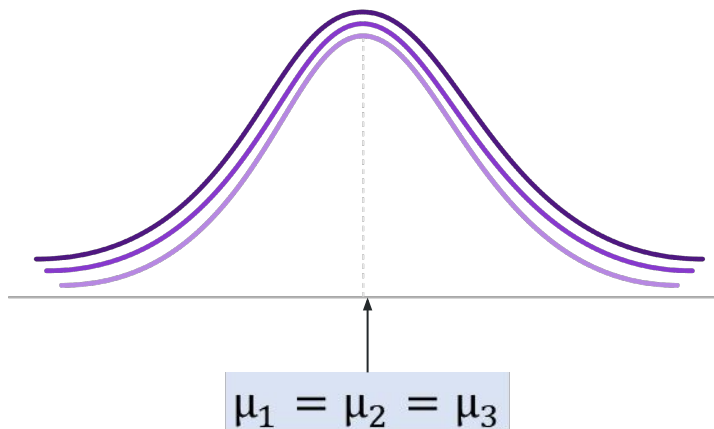
PYTHON JOURNEY

MACHINE & DEEP LEARNING

A BASE

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

H_1 : Nem todos os μ_k são iguais



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



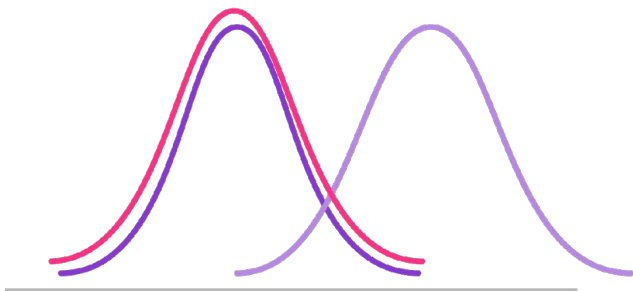
A BASE

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

H_1 : Nem todos os μ_k são iguais

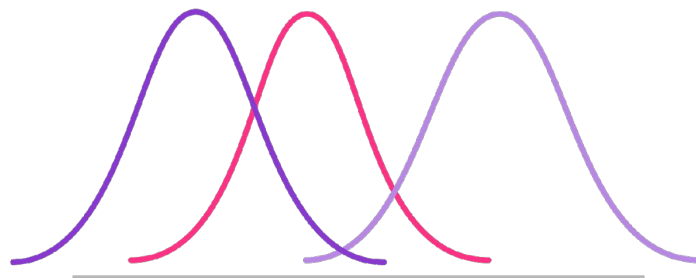
Ao menos uma média é diferente:

H_0 NÃO é verdadeira
(existe efeito do tratamento)



$$\mu_1 = \mu_2 \neq \mu_3$$

or



$$\mu_1 \neq \mu_2 \neq \mu_3$$



A BASE

Hipóteses do ANOVA de um critério

Hipótese Nula: a média de todas as populações são iguais, ou seja, o tratamento (fator) não tem efeito (nenhuma variação em média entre os grupos).

Hipótese Alternativa: nem todas as médias populacionais são iguais, ou seja, pelo menos uma média é diferente, isto é, existe efeito do tratamento. Não quer dizer que todas as médias são diferentes (alguns pares podem ser iguais).

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

H_1 : Nem todas as médias populacionais são iguais.



COMENTÁRIOS INICIAIS

- O problema consiste em comparar três médias.
- Uma solução intuitiva seria comparar as médias duas a duas.
- Essa solução, no entanto, não é a mais eficiente.
- É possível mostrar que ela não garante o mesmo nível de significância utilizado nas comparações duas a duas.
- A solução mais eficiente é dada por um método chamado Análise de Variância (ANOVA).



PREMISSAS

- Amostras independentes extraídas de populações com variâncias iguais:

$$\sigma_1^2 = \sigma_2^2 = \sigma_3^2 = \sigma^2$$

- Duas estimativas independentes:

- Estimativa “dentro”:

$$s_d^2 = \frac{(s_1^2 + s_2^2 + \dots + s_k^2)}{k}$$

- Estimativa “entre”:

$$s_e^2 = n \times s_{\bar{X}}^2$$



PREMISSAS

- Se as médias são iguais:

$$s_e^2 \cong s_d^2$$

- Se as médias são diferentes:

$$s_e^2 \neq s_d^2$$

$$\begin{cases} H_0; \mu_a = \mu_b = \mu_c \\ H_1; \mu_i \neq \mu_j \quad \text{Para pelo menos um par (i,j)} \end{cases}$$



$$\begin{cases} H_0; \sigma_e^2 = \sigma_d^2 \\ H_1; \sigma_e^2 \neq \sigma_d^2 \end{cases}$$



EXEMPLO

- Uma revista especializada dos Estados Unidos fez um teste comparando o consumo de gasolina de três diferentes marcas de carro: A, B e C.
- Foram selecionados 21 motoristas profissionais e os carros foram aleatoriamente distribuídos entre eles.
- 7 motoristas ficaram com carros da marca A; outros 7, com carros da marca B; e os 7 restantes, com carros da marca C.



EXEMPLO

- Todos rodaram apenas em área urbana e após algum tempo calculou-se o consumo em milhas por galão.
- A revista está interessada em saber se o consumo médio é diferente entre os diferentes tipos de carro.



EXEMPLO

Marca A	Marca B	Marca C
22,2	24,6	22,7
19,9	23,1	21,9
20,3	22,0	23,3
21,4	23,5	24,1
21,2	23,6	22,1
21,0	22,1	23,4
20,3	23,5	22,6

$n = 7$ (quantidade de elementos em cada amostra).

$k = 3$ (quantidade de amostras).



EXEMPLO

Marca A	Marca B	Marca C
22,2	24,6	22,7
19,9	23,1	21,9
20,3	22,0	23,3
21,4	23,5	24,1
21,2	23,6	22,1
21,0	22,1	23,4
20,3	23,5	22,6

RESUMO				
Grupo	Contagem	Soma	Média	Variância
Marca A	7	146,3	20,9	0,626667
Marca B	7	162,4	23,2	0,826667
Marca C	7	160,1	22,87	0,602381



EXEMPLO

RESUMO				
Grupo	Contagem	Soma	Média	Variância
Marca A	7	146,3	20,9	0,626667
Marca B	7	162,4	23,2	0,826667
Marca C	7	160,1	22,87	0,602381
		Variância	1,55	

$$\left. \begin{aligned} s_d^2 &= \frac{0,626 + 0,827 + 0,602}{3} = 0,685 \\ s_e^2 &= 7 \times 1,55 = 10,83 \end{aligned} \right\} \text{Qual a conclusão?}$$



EXEMPLO

ANOVA			
Fonte de variação	SQ	gl	MQ
Entre grupos	21,664	2	10,832
Dentro dos grupos	12,334	18	0,685

$gl_{entre} = k - 1$

$gl_{dentro} = (n - 1)k$

MQ = Média de quadrados.

SQ = Soma de quadrados.

$$MQ = \frac{SQ}{gl} \quad \therefore \quad \begin{cases} MQ_{entre} = \frac{SQ_{entre}}{gl_{entre}} = s_e^2 \\ MQ_{dentro} = \frac{SQ_{dentro}}{gl_{dentro}} = s_d^2 \end{cases}$$



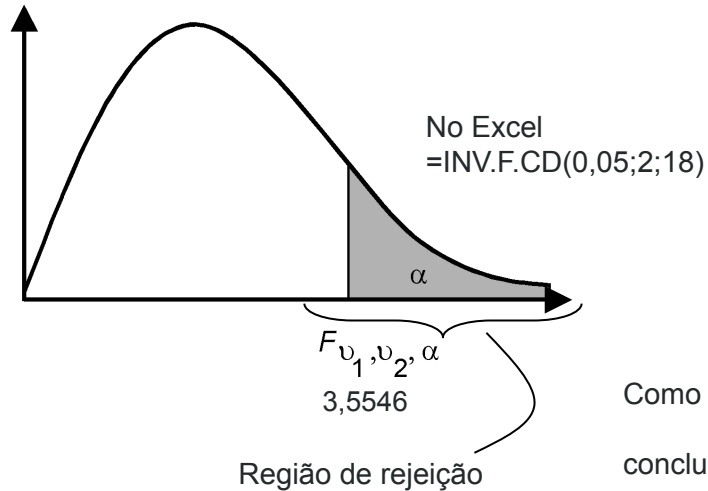
EXEMPLO

ANOVA						
Fonte de variação	SQ	gl	MQ	F	valor -P	F crítico
Entre grupos	21,664	2	10,832	15,808	0,000109	3,5546
Dentro dos grupos	12,334	18	0,685			
TOTAL	33,998	20				

Adotando $\alpha=5\%$



EXEMPLO



$$s_e^2 = MQ_{\text{entre}}$$

$$s_d^2 = MQ_{\text{dentro}}$$

$$F = \frac{s_e^2}{s_d^2} = \frac{MQ_{\text{entre}}}{MQ_{\text{dentro}}}$$

Como **F** > **Fcrítico** (ou **valor-p** < α), Rejeitamos H_0 , ou seja, concluímos que pelo menos um par de médias é diferente.



EXERCÍCIO

- Uma empresa deseja abrir uma loja e está analisando três regiões possíveis. Para auxiliar na decisão, ela pesquisou a renda familiar nas proximidades de cada região. Os dados obtidos estão na tabela abaixo.
- Podemos concluir que existe diferença na renda média familiar usando um nível de significância de 5%?

Região A	Região B	Região C
34	44	45
38	41	50
40	39	46
30	40	48



EXERCÍCIO

Região A	Região B	Região C
34	44	45
38	41	50
40	39	46
30	40	48



EXERCÍCIO

Anova: fator único

RESUMO

Grupo	Contagem	Soma	Média	Variância
Região A	4	142	35,5	19,67
Região B	4	164	41	4,67
Região C	4	189	47,25	4,92

ANOVA

Fonte de variação	SQ	gl	MQ	F	valor -P	F crítico
Entre grupos	276,5	2	138,25	14,179	0,001653	4,25649
Dentro dos grupos	87,75	9	9,75			
TOTAL	364,25	11				



EXERCÍCIO

Fonte da Variação	SQ	gl	Variância	Razão F
Entre	SQE	k - 1	$S^2_{\text{entre}} = \frac{SQE}{k - 1}$	$F = \frac{S^2_{\text{entre}}}{S^2_{\text{dentro}}}$
Dentro	SQD	n - k	$S^2_{\text{dentro}} = \frac{SQD}{n - k}$	
Total	$SQT = SQE + SQD$	n - 1		

k = número de amostras (grupos)

n = soma do número de elementos de todas as amostras

gl = graus de liberdade



OBRIGADO

 /andresilvadecarvalho



[/lattes.cnpq.br/6876528572507972](https://lattes.cnpq.br/6876528572507972)

FIAP

Copyright © 2021 | Professor André Silva de Carvalho

Todos os direitos reservados. Reprodução ou divulgação total ou parcial deste documento, é expressamente proibido sem consentimento formal, por escrito, do professor/autor



SHIFT

 FIAP

