

**FACULDADE DE ESTUDOS INTERDISCIPLINARES**  
**Ciência de Dados e Inteligência Artificial**

## Problema

Após fazer a revisão da literatura na procura por proxies da evasão escolar, Horácio necessita coletar dados para poder treinar modelos de aprendizagem de máquina. No entanto, ele ainda não se sente seguro de que um modelo baseado nestes dados pode efetivamente identificar estudantes que estejam em risco de evasão.

Horácio decide então fazer um experimento para entender melhor os dados. Para simplificar, Horácio decidiu tentar primeiramente identificar os estudantes desmotivados, assumindo que este grupo possui maior risco de evasão. Mais precisamente, Horácio está interessado em comparar o grau de satisfação dos estudantes de quatro níveis de renda (A, B, C, D) com a motivação com o curso. Horácio aplica um questionário de 40 itens projetado para medir o grau de motivação com o curso. A pontuação 40 indica um aluno com a motivação mais alta e 0, a mais baixa.

A tabela a seguir consolida os dados obtidos.

Classe A	Classe B	Classe C	Classe D
$X_1$	$X_2$	$X_3$	$X_4$
19	23	16	5
35	30	5	6
15	20	9	4
26	20	10	5
30	27	5	0

Horácio estabelece as suas hipóteses:

$H_0$ . A renda familiar do estudante não tem relação com sua motivação com o curso

$H_1$ . A renda familiar do estudante tem relação com sua motivação com o curso

Como Horácio pode testar suas hipóteses?

## ANOVA

### Passo

Calcule a média para cada amostra.

$$\bar{X}_1 = \frac{\sum X_1}{N_1} = \frac{125}{5} = 25 \quad (1)$$

$$\bar{X}_2 = \frac{\sum X_2}{N_2} = \frac{120}{5} = 24 \quad (2)$$

$$\bar{X}_3 = \frac{\sum X_3}{N_3} = \frac{45}{5} = 9 \quad (3)$$

$$\bar{X}_4 = \frac{\sum X_4}{N_4} = \frac{20}{5} = 4 \quad (4)$$

### Passo

Calcule a soma dos escores e a soma dos escores quadrados.

$X_1$	$X_1^2$	$X_2$	$X_2^2$	$X_3$	$X_3^2$	$X_4$	$X_4^2$
19	361	23	529	16	256	5	25
35	1225	30	900	5	25	6	36
15	225	20	400	9	81	4	16
26	676	20	400	10	100	5	25
30	900	27	729	5	25	0	0
$\sum X_1 = 125$	$\sum X_1^2 = 3387$	$\sum X_2 = 120$	$\sum X_2^2 = 2958$	$\sum X_3 = 45$	$\sum X_3^2 = 487$	$\sum X_4 = 20$	$\sum X_4^2 = 102$

Calcule o número total de participantes da amostra e a média para todos os grupos combinados.

$$\sum X_{total} = \sum X_1 + \sum X_2 + \sum X_3 + \sum X_4 = 125 + 120 + 45 + 20 = 310 \quad (5)$$

$$\sum X_{total}^2 = \sum X_1^2 + \sum X_2^2 + \sum X_3^2 + \sum X_4^2 = 3387 + 2958 + 487 + 102 = 6934 \quad (6)$$

$$N_{total} = \sum N_1 + \sum N_2 + \sum N_3 + \sum N_4 = 5 + 5 + 5 + 5 = 20 \quad (7)$$

$$\bar{X}_{total} = \frac{\sum X_{total}}{N_{total}} = \frac{310}{20} = 15,5 \quad (8)$$

## Passo

Calcule a soma dos quadrados totais.

$$SS_{total} = \sum X_{total}^2 - N_{total} \bar{X}_{total}^2 = 6934 - (20)(15,5)^2 = 6934 - (20)(240,25) = 6934 - 4805 = 2129 \quad (9)$$

## Passo

Calcule a soma dos quadrados entre grupos.

$$SS_{entre} = \sum N_{grupo} \bar{X}_{grupo}^2 - N_{total} \bar{X}_{total}^2 = [(5)(25)^2 + (5)(24)^2 + (5)(9)^2 + (5)(4)^2] - (20)(15,5)^2 = 6490 - 4805 = 1685 \quad (10)$$

## Passo

Calcule a soma dos quadrados intra-grupos.

$$SS_{intra} = \bar{X}_{total}^2 - \sum N_{grupo} \bar{X}_{grupo}^2 = 6934 - [(5)(25)^2 + (5)(24)^2 + (5)(9)^2 + (5)(4)^2] = 6934 - 6490 = 444 \quad (11)$$

## Passo

Calcule os graus de liberdade entre grupos.

$$df_{entre} = k - 1 = 4 - 1 = 3 \quad (12)$$

## Passo

Calcule os graus de liberdade intra-grupos.

$$df_{intra} = N_{total} - k = 20 - 4 = 16 \quad (13)$$

## Passo

Calcule o quadrado médio entre grupos.

$$MS_{entre} = \frac{SS_{entre}}{df_{entre}} = \frac{1685}{3} = 561,66 \quad (14)$$

## Passo

Calcule o quadrado médio intra-grupos.

$$MS_{intra} = \frac{SS_{intra}}{df_{intra}} = \frac{444}{16} = 27,5 \quad (15)$$

## Passo

Calcule a Razão F.

$$F = \frac{MS_{entre}}{MS_{intra}} = \frac{561,66}{27,5} = 20,424 \quad (16)$$

Valor do F na tabela de significância para  $\alpha = 0,05$  e  $df = 3$  e  $df = 16$ : 3,24.

Como  $20,424 \gg 3,24$  temos que a classe (nível de renda) afeta a motivação dos estudantes.