

# Aula 6 - Testes estatísticos

**Fizemos uma breve revisão dos testes paramétricos que tínhamos visto e começamos com o princípio da ANOVA**

## ANOVA

Vamos reviver a história da patricinha? Que nos ajudou a gerar sentimentos para a variância e desvio padrão?

A patricinha tem o gasto dela, e chegamos na conclusão de que a média seria o melhor chute para esse contexto de negócio, certo?

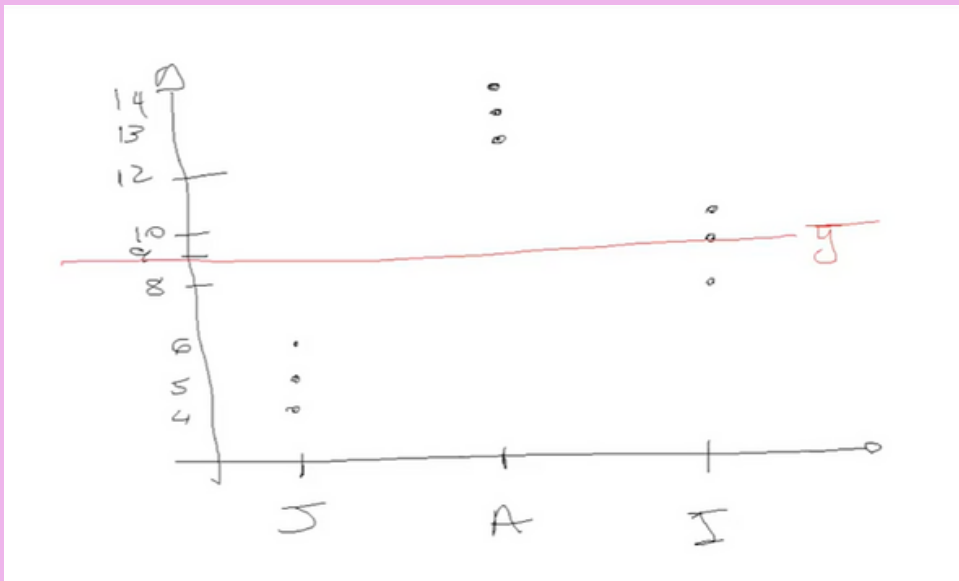
A média é tão sensual, pois é a única medida que consegue está perto de todos os pontos ao mesmo tempo

Após termos essa discussão em cima da média, foi adicionada uma nova variável para o cenário da patricinha: o perfil-idade do cliente

$y$ gasto	perfil-idade
4	S
5	S
6	S
12	A
13	A
14	A
8	I
9	I
10	I

Relembramos que o vamos estimar é nosso **Y** e a variável que nos dá suporte, é nosso **X**  
Neste caso, o gasto é o Y e o perfil\_idade é nosso X

Vimos que o gráfico que melhor nos mostra o comportamento dessa variável Y para cada um desses X é o **boxplot** e com isso, tiramos alguns sentimentos!



Começamos a ver o princípio de um modelo e entender que a ANOVA é um modelo e também é um teste. Com isso, fizemos nosso chute sofisticado, que nada mais é que a média de cada perfil

Y gasto	X perfil-idade	$\hat{\mu}$
4	J	5
5	J	5
6	J	5
12	A	13
13	A	13
14	A	13
8	I	9
9	I	9
10	I	9

Como isso é um princípio de modelo, vamos fazer uma equação para calcular nosso chute. E para essa equação, vamos usar o princípio das **Dummies**.

A variável Dummy é utilizada para verificar a **presença** ou **ausência** de algo. é uma variável binária ( 0 ou 1 )

y	X	$\hat{y}$	seJ	seA	seI
4	J	5	1	0	0
5	J	5	1	0	0
6	J	5	1	0	0
12	A	13	0	1	0
13	A	13	0	1	0
14	A	13	0	1	0
8	I	9	0	0	1
9	I	9	0	0	1
10	I	9	0	0	1

Com isso, chegamos na seguinte equação:

$$\rightarrow \hat{y} = 5 * seJ + 13 * seA + 9 * seI$$

Porém, vimos que ela não está correta, pois existe o **Princípio da Parcimônia**, onde precisamos fazer a mesma coisa menos, ou seja, usaremos menos dummies (sempre n-1). Essa premissa na Estatística é **REGRA**

**Durante a aula, ficamos um bem tempo tentando e entendendo como fazer as equações. É de extrema importância entender bem esse conceito, por isso, reveja a aula para pegar o sentimento!!!**

Depois de algumas discussões, chegamos nas seguintes equações:

Tirando a Dummy Jovem:

$$\hat{g} = 5 + 4 De-I + 8 De-A$$

Tirando a Dummy Idoso:

$$\hat{g} = 9 + 4 De-A - 4 De-J$$

Tirando a Dummy Adulto:

$$\hat{g} = 13 - 8 De-J - 4 De-I$$

Portanto, vimos que o teste estatístico da ANOVA 1 fator é:

$$\begin{cases} H_0: \mu_1 = \mu_2 = \dots = \mu_k \\ H_1: \text{pelo menos um } \mu \neq \end{cases}$$

## Mas ainda não acabou!! Em relação à parte do modelo, foram feitas mais três perguntas:

- 1. E se eu perguntasse o quanto meu modelo “média” erra? Que conta eu poderia fazer?

$$\sum (y - \bar{y})^2 = SST$$

- 2. O quanto o meu modelo sofisticado é melhor do que a média geral?

$$\sum (\hat{y} - \bar{y})^2 = SSR$$

- 3. O quanto o meu modelo erra?

$$\sum (y - \hat{y})^2 = SSE$$

■ **É fato que:**

$$SQT = SQM + SQE$$

**No final da aula, criamos uma medida que nos diz o quão bom o nosso modelo é, ela é o  $R^2$ , que é dado por:**

$$R^2 = SQM / SQT$$



Esse número varia de 0 a 1, e quanto maior, melhor é nosso modelo.

**Mas para frente, em Regressão Linear, vamos entender melhor essas medidas, então, fique tranquilo!!  
Reveja essa aula e sinta as medidas :)**