**Anotações**

**Estatística descritiva**

**Análises Exploratórias - Aula 03**

**Medidas Separatrizes:**

* **Mínimo:** mínimo valor do conjunto de dados
* **1\* Quartil: representa o conjunto de dados situados em um intervalo entre o valor mínimo e o valor máximo.O primeiro quartil se situa 25% abaixo da mediana!**
* **Mediana:** Ela separa os conjuntos de dados ordenados em 50% para baixo e 50% para cima, ao meio!
* **3\* Quartil: representa o conjunto de dados situados em um intervalo entre o valor mínimo e o valor máximo.O terceiro quartil se situa 25% acima da mediana!**
* **Máximo:** Máximo valor do conjunto de dados

**Amplitude Total -** é a diferença entre o valor máximo e valor mínimo

**Amplitude Interquartílica(AIQ)**  = (Q3 - Q1)

LI = Q1 - (1.5 \* AIQ)

Abaixo do valor do LI é considerado um OUTLIER("o diferentão")

LS = Q3 + (1.5 \* AIQ)

Acima do valor do L3 é também considerado um OUTLIER("o diferentão)

**BOXPLOT** - É a melhor forma de identificar um OUTLIER, mas não é a única!

**Medidas de Associação:**

1 - **Coeficiente de correlação de Pearson:** Server para tentar ou explicar a variação de alguma coisa a partir da variação de outra coisa

**Variável dependente** - Variável a ser explicada ( variável Y )

**Variável independente** - Variável explicativa ( variável X )

Fórmula:

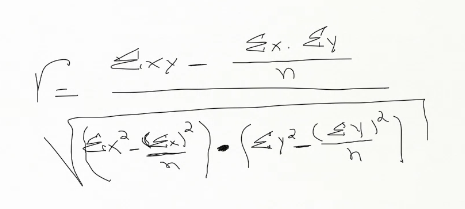
Covariância(xy)

r = —----------------------

Desvio(x) \* Desvio(y)

**O resultado dessa correlação tem infinitas possibilidades entre ( -1 e 1 )**

| **TOTAL** | **propaganda(x)** | **Vendas(Y)** | **x\*x** | **y\*Y** | **x\*Y** |
| --- | --- | --- | --- | --- | --- |
|  | 1 | 1 | 1 | 1 | 1 |
|  | 2 | 2 | 4 | 4 | 4 |
|  | 3 | 4 | 9 | 16 | 12 |
|  | 4 | 5 | 16 | 25 | 20 |
|  | 5 | 8 | 25 | 64 | 40 |
| **TOTAL** | **15** | **20** | **55** | **110** | **77** |



77 - (15\*20)/5 17

r = —----------------------------- = —---------------- = 0,9815 ao quadrado = 0,9633% ou 96,33%!

Rq [(55 - 45) \* (110 - 80))] Rq 300

**( r² ) - É necessário elevar ao quadrado para chegar ao valor real.**

* Como o resultado chegou próximo de 1,mostrou-se que as vendas têm alta correlação de imersão de propagandas! Assim eu posso dizer que 96,33% das vendas tem relação com a imersão das propagandas!
* **Quanto mais próximo de zero, mais fraco é a correlação, menor o poder de explicação de uma variável a partir de outra coisa!**
* **Correlação positiva: quando duas variáveis que possuem correlação crescem ou decrescem juntas, ou seja, que possuem uma relação direta;**
* **Correlação negativa: quando duas variáveis que possuem correlação mas quando uma variável cresce a outra decresce, ou vice-versa;**
* **Não ter correlação: quando o crescimento ou decrescimento de uma variável não tem efeito sobre outra variável.**

**Como podemos interpretar os valores ?**

* ρ = 0,9 a 1 (positivo ou negativo): correlação muito forte;
* ρ = 0,7 a 09 (positivo ou negativo): correlação forte;
* ρ = 0,5 a 0,7 (positivo ou negativo): correlação moderada;
* ρ = 0,3 a 0,5 (positivo ou negativo): correlação fraca;
* ρ = 0 a 0,3 (positivo ou negativo): não possui correlação.

Regressão linear simples:

Y^(vendas estimadas) = alfa + (beta\*x)