REGRESSÃO LINEAR

ASPECTOS GERAIS

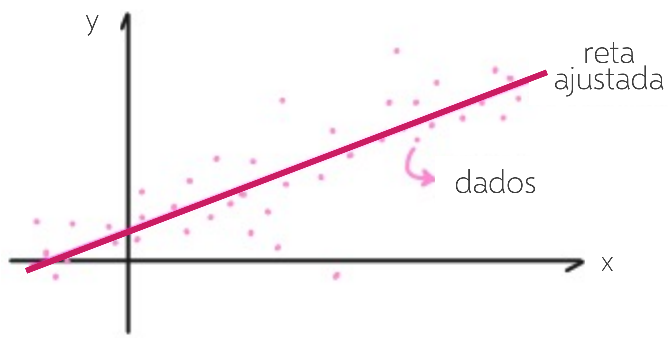
= calcular a expressão matemática que relaciona duas variáveis x e y:

$$Y = p + m \,.\, x$$

p= coeficiente linear
m= coeficiente angular

$$m = \frac{\Delta y}{\Delta x}$$

Exemplo:



COEFICIENTE DE CORRELAÇÃO LINEAR DE *PEARSON*

$$r = \frac{\sum[(X_i - \bar{X}) \,.\, (Y_i - \bar{Y})]}{\sqrt{\sum[(X_i - \bar{X})^2 \,.\, (Y_i - \bar{Y})^2]}}$$

Mede a força da relação linear entre 2 variáveis x e y

$-1 \leq r \leq 1:$
- Mais próximo de -1 ou 1 → Correlação mais forte
- Mais próximo de zero → Correlação mais fraca (não há relação linear entre as variáveis)

Em questões, é útil usar:

$$\sum[(X_i - \bar{X}) \,.\, (Y_i - \bar{Y})] = \sum(X_i \,.\, Y_i) - n \,.\, \bar{X} \,.\, \bar{Y}$$

MODELO ESTATÍSTICO

$$Y_i = \alpha + \beta X_i + \mu_i$$

x = variável independente
y = variável dependente
n = número de pares de valores observados
μ_i = erro ou desvio
$\alpha + \beta X_i$ = componente de y que varia linearmente com x

RETA DE REGRESSÃO ESTIMADA

$$\hat{Y} = a + bXi$$

$$e = Y_i - \hat{Y}_i$$ → desvio

MÉTODO DOS MÍNIMOS QUADRADOS

= determina as estimativas de a e b dos parâmetros minimizando a soma dos quadrados dos desvios

Após calcular *b*, basta substituir na equação $\bar{Y}$ = a+b $\bar{X}$ para determinar *a*

$$b = \frac{\sum[(X_i - \bar{X}) \,.\, (Y_i - \bar{Y})]}{\sum(X_i - \bar{X})^2}$$

RETA QUE PASSA PELA ORIGEM

- $\alpha = 0$

$$Y_i = \beta X_i + \varepsilon_i$$ → $$\beta = \frac{\sum X \,.\, Y}{\sum X^2}$$

REGRESSÃO LINEAR

ANÁLISE DA VARIÂNCIA DA REGRESSÃO

$$SQT = SQM + SQR$$

- SQT: Soma dos quadrados total
- SQM: Soma dos quadrados do modelo
- SQR: Soma dos quadrados dos resíduos

FÓRMULAS IMPORTANTES

$$SQM = b.\sum[(X_i - \bar{X}).(Y_i - \bar{Y})]$$

Soma dos quadrados do modelo de regressão

ou

$$SQM = b^2.\sum(X_i - \bar{X})^2$$

$$SQR = \sum(Y_i - \hat{Y})^2$$

É a parte do desvio total não explicada pelo modelo de regressão

TABELA DA ANOVA

DECORE!

FONTE DE VARIAÇÃO	GRAUS DE LIBERDADE (gl)	SOMA DE QUADRADOS	QUADRADOS MÉDIOS	F
MODELO	1	SQM	$QMM = \frac{SQM}{1}$	$F_{teste} = \frac{SMM}{QMR}$
RESÍDUOS	n-2	SQR	$QMR = \frac{SQR}{n-2}$	
TOTAL	n-1	SQT	$QMT = \frac{SQT}{n-1}$	

n = tamanho da amostra

(QMR = Estimativa de variância σ^2 residual)

COEFICIENTES

IMPORTANTE!

- Coeficiente de correlação: $R = \sqrt{\frac{SQM}{SQT}}$
- Coeficiente de determinação: $R^2 = 1 - \frac{SQR}{SQT}$

  Exprime a proporção da variação total de y que é explicada pela reta de regressão

$(0 \leq R^2 \leq 1)$