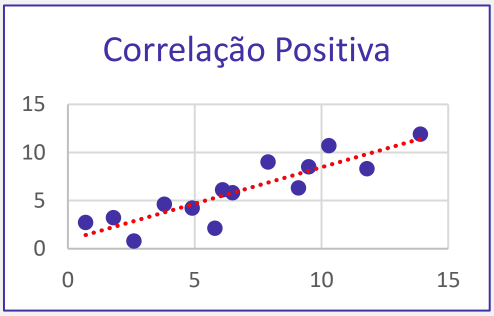
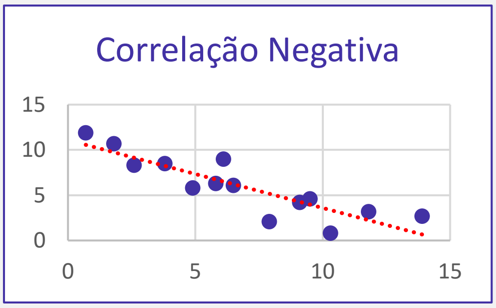
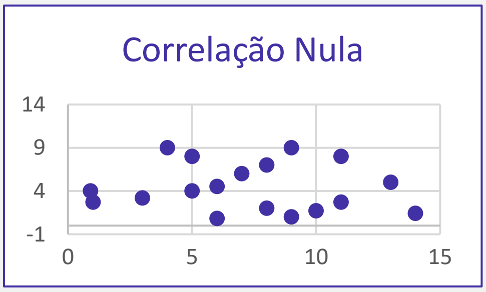
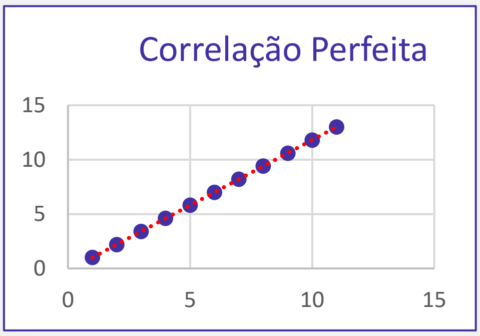
By @kakashi_copiador

RESUMO DA AULA

CORRELAÇÃO LINEAR

A **correlação** é usada para indicar a força que mantém unidos dois conjuntos de valores. A **correlação linear** pode ser:

Gráfico	Definição
	**Direta ou positiva** – quando temos dois fenômenos que variam no mesmo sentido. Se aumentarmos ou diminuirmos um deles, o outro também aumentará ou diminuirá;
	**Inversa ou negativa** – quando temos dois fenômenos que variam em sentido contrário. Se aumentarmos ou diminuirmos um deles, acontecerá o contrário com o outro, no caso, diminuirá ou aumentará;
	**Inexistente ou nula** – quando não existe correlação ou dependência entre os dois fenômenos. Nessa situação, o valor do coeficiente de correlação linear será zero ($r = 0$) ou um valor aproximadamente igual a zero ($r \cong 0$);
	**Perfeita** – quando os fenômenos se ajustam perfeitamente a uma reta.

Coeficiente de Correlação de Pearson

COEFICIENTE DE CORRELAÇÃO LINEAR DE PEARSON

É adotado para medir o quão forte é a RELAÇÃO linear entre duas VARIÁVEIS.

FÓRMULA

$$r = \frac{\sum_{i=1}^{n}[(X_i - \bar{X}) \times (Y_i - \bar{Y})]}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2 \times \sum_{i=1}^{n}(Y_i - \bar{Y})^2}}$$

FÓRMULAS ALTERNATIVAS

$$\sum_{i=1}^{n}[(X_i - \bar{X}) \times (Y_i - \bar{Y})] = \sum_{i=1}^{n}(X_i \times Y_i) - n \times \bar{X} \times \bar{Y}$$

$$\sum_{i=1}^{n}[(X_i - \bar{X})^2] = \sum_{i=1}^{n}(X_i^2) - n \times \bar{X}^2$$

$$\sum_{i=1}^{n}[(Y_i - \bar{Y})^2] = \sum_{i=1}^{n}(Y_i^2) - n \times \bar{Y}^2$$

Sobre a Coeficiente de Correlação de Pearson, podemos afirmar que:

- I – Pode assumir quaisquer valores entre 1 e -1, ou seja: $\boxed{-1 \leq r \leq 1}$.
- II – Quanto mais próximo r estiver de 0, menor será a relação linear entre as duas variáveis
- III – Quanto mais próximo r estiver de (1 ou -1), maior será a relação linear entre as duas variáveis.

Propriedades do Coeficiente de Correlação

1º Propriedade

- **O coeficiente de correlação não sofre alteração quando uma constante é adicionada a (ou subtraída de) uma variável.**

2º Propriedade

- **O coeficiente de correlação pode não sofrer alteração ou pode ter seu sinal alterado quando uma variável é multiplicada (ou dividida) por uma constante. Caso as constantes tenham o mesmo sinal, o valor do coeficiente de correlação não será alterado. Por outro lado, se as constantes tiverem sinais contrários, o coeficiente mudará de sinal, mas o valor permanecerá inalterado.**

REGRESSÃO LINEAR SIMPLES

REGRESSÃO LINEAR SIMPLES

Calcula a expressão matemática que relaciona Y (variável dependente) em função de X (variável independente).

Trata-se da equação que representa uma reta:

$$y = m \cdot x + b$$

- Propriedades

Sobre a Regressão Linear Simples, podemos afirmar que:

- I – O coeficiente m é conhecido como taxa de variação ou **coeficiente angular da reta.**
- II – O coeficiente angular é expresso por: $m = \frac{\Delta y}{\Delta x} = \frac{y - y_0}{x - x_0}$
- III – O coeficiente b é conhecido como coeficiente linear da reta e determina o ponto em que a reta intercepta o eixo y.
- IV – Quando a correlação linear não é perfeita, utilizamos a expressão $Y_i = \alpha + \beta X_i + \varepsilon_i$, para determinar a reta de regressão.

Método dos Mínimos Quadrados

MÉTODO DOS MÍNIMOS QUADRADOS

A reta a ser adotada deverá ser aquela que torna mínima a soma dos quadrados das distâncias da reta aos pontos experimentais, medidas no sentido da variação aleatória.

Esse método é empregado na obtenção dos estimadores α e β de um modelo de regressão linear:

$$Y_i = \alpha + \beta X_i + \varepsilon_i.$$

Expressão usada para determinar a reta de regressão é:

$$\hat{Y}_i = a + bX_i$$

Reta Passando pela Origem

MODELO DE REGRESSÃO QUE PASSA OBRIGATORIAMENTE PELA ORIGEM É:

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

$$Y_i = 0 + \beta X_i + \varepsilon_i$$

$$\boxed{Y_i = \beta X_i + \varepsilon_i.}$$

ANÁLISE DE VARIÂNCIA DA REGRESSÃO

ANÁLISE DE VARIÂNCIA DA REGRESSÃO

Estratégia para verificar se compensa ou não utilizar um modelo de regressão linear,
$Y_i = \alpha + \beta X_i + \varepsilon_i$

→

Observar a redução no resíduo (desvio) quando comparado com um modelo aproximadamente uniforme $Y_i = \mu + \varepsilon_i$.

Testar a hipótese:

$$\begin{cases} H_0: \beta = 0 \\ H_1: \beta \neq 0 \end{cases}$$

→

O resultado da análise de variância da regressão é uma tabela que resume várias medidas usadas no teste de hipóteses anterior.

Graus de Liberdade

O número de graus de liberdade do modelo de regressão é:

$$GL_{Modelo} = 2 - 1 = 1$$

O número de graus de liberdade dos resíduos é:

$$GL_{Resíduos} = n - 2$$

O número de graus de liberdade total é:

$$GL_{Total} = n - 1$$

Somas de Quadrados

A soma dos quadrados totais é calculada por meio das seguintes fórmulas:

$$SQT = SQM + SQR$$

$$SQT = \sum_{i=1}^{n} (Y_i - \bar{Y})^2$$

A soma dos quadrados do modelo de regressão é calculada mediante as seguintes fórmulas:

$$SQM = \sum_{i=1}^{n} \left(\hat{Y}_i - \bar{Y}\right)^2$$

$$SQM = b \times \sum_{i=1}^{n} [(X_i - \bar{X}) \times (Y_i - \bar{Y})]$$

$$SQM = b^2 \times \sum_{i=1}^{n} (X_i - \bar{X})^2$$

A soma dos quadrados dos resíduos é calculada pela fórmula:

$$SQR = \sum_{i=1}^{n} \left(Y_i - \hat{Y}_i\right)^2$$

Coeficiente de Determinação

O coeficiente de determinação é calculado pela fórmula:

$$R^2 = \frac{SQM}{SQT}$$

Em que R é o coeficiente de correlação linear, calculado pela expressão:

$$R = \sqrt{\frac{SQM}{SQT}}$$

O coeficiente de determinação também pode ser escrito da seguinte forma:

$$R^2 = \frac{SQM}{SQT} = \frac{SQT - SQR}{SQT} = 1 - \frac{SQR}{SQT}$$

Coeficiente de Determinação Ajustado

É obtida pela **divisão** de $\boldsymbol{SQR}$ e $\boldsymbol{SQT}$ pelos respectivos **graus de liberdade**:

$$\overline{R^2} = 1 - \frac{SQR/(n-2)}{SQT/(n-1)}$$

A **relação** entre o coeficiente de determinação **ajustado** ($\overline{\boldsymbol{R^2}}$) e o coeficiente de determinação **tradicional** ($\boldsymbol{R^2}$) é dada por:

$$\overline{R^2} = 1 - (1 - R^2) \times \frac{(n-1)}{(n-2)}$$

Quadrados Médios

Quadrado médio do **modelo (QMM)**: $QMM = \frac{SQM}{1}$

Quadrado médio dos **resíduos (QMR)**: $QMR = \frac{SQR}{n-2}$

Quadrado médio **total (QMT)**: $QMT = \frac{SQT}{n-1}$

Estatística F (Razão F)

Estatística $\boldsymbol{F}$ **(ou razão F)**: $F^* = \frac{QMM}{QMR}$

Se $F^* > F_{crítico}$, podemos **rejeitar a hipótese nula**;

Se $F^* < F_{crítico}$, **não** podemos **rejeitar a hipótese nula**.

Tabela de Análise de Variância da Regressão

Fonte de Variação	Graus de Liberdade	Soma dos Quadrados	Quadrados Médios	Estatística F (Razão F)
Modelo	1	SQM	$QMM = \dfrac{SQM}{1}$	$F^* = \dfrac{QMM}{QMR}$
Resíduos	$n - 2$	SQR	$QMR = \dfrac{SQR}{n-2}$	
Total	$\boldsymbol{n - 1}$	$\boldsymbol{SQT}$	$\boldsymbol{QMT = \dfrac{SQT}{n-1}}$	