

ESTATÍSTICA

CORRELAÇÃO E REGRESSÃO LINEAR

Josimar Padilha

CORRELAÇÃO LINEAR SIMPLES

Em pesquisas, frequentemente, procura-se verificar se existe relação entre duas ou mais variáveis, isto é, saber se as alterações sofridas por uma das variáveis são acompanhadas por alterações nas outras. Por exemplo, peso vs. idade, consumo vs. renda, altura vs. peso, de um indivíduo.

O termo correlação significa relação em dois sentidos (co + relação), e é usado em estatística para designar a força que mantém unidos dois conjuntos de valores. A verificação da existência e do grau de relação entre as variáveis é o objeto de estudo da correlação.

Uma vez caracterizada esta relação, procura-se descrevê-la sob forma matemática, através de uma **função**.

A estimação dos parâmetros dessa função matemática é o objeto da regressão. Os pares de valores das duas variáveis poderão ser colocados num diagrama cartesiano chamado “diagrama de dispersão”. A vantagem de construir um diagrama de dispersão está em que, muitas vezes sua simples observação já nos dá uma ideia bastante boa de como as duas variáveis se relacionam.

Em teoria da probabilidade e estatística, **correlação**, também chamada de coeficiente de correlação, indica a força e a direção do relacionamento linear entre duas variáveis aleatórias. No uso estatístico geral, correlação ou co-relação se refere a medida da relação entre duas variáveis, embora correlação não implique **CAUSALIDADE**.

Neste sentido geral, existem vários coeficientes medindo o grau de correlação, adaptados à natureza dos dados.

Vários coeficientes são utilizados para situações diferentes. O mais conhecido é o coeficiente de correlação de Pearson, o qual é obtido dividindo a covariância de duas variáveis pelo produto de seus desvios padrão.

Variável: características ou itens de interesse de cada elemento de uma população ou amostra, podendo ser chamada de parâmetro.

Duas variáveis estão relacionadas se a mudança de uma provoca a mudança na outra.

Exemplos: Velocidade x consumo combustível

Índice de criminalidade x grau de escolaridade

Idade x condicionamento físico.

Correlação

Correlação entre duas variáveis

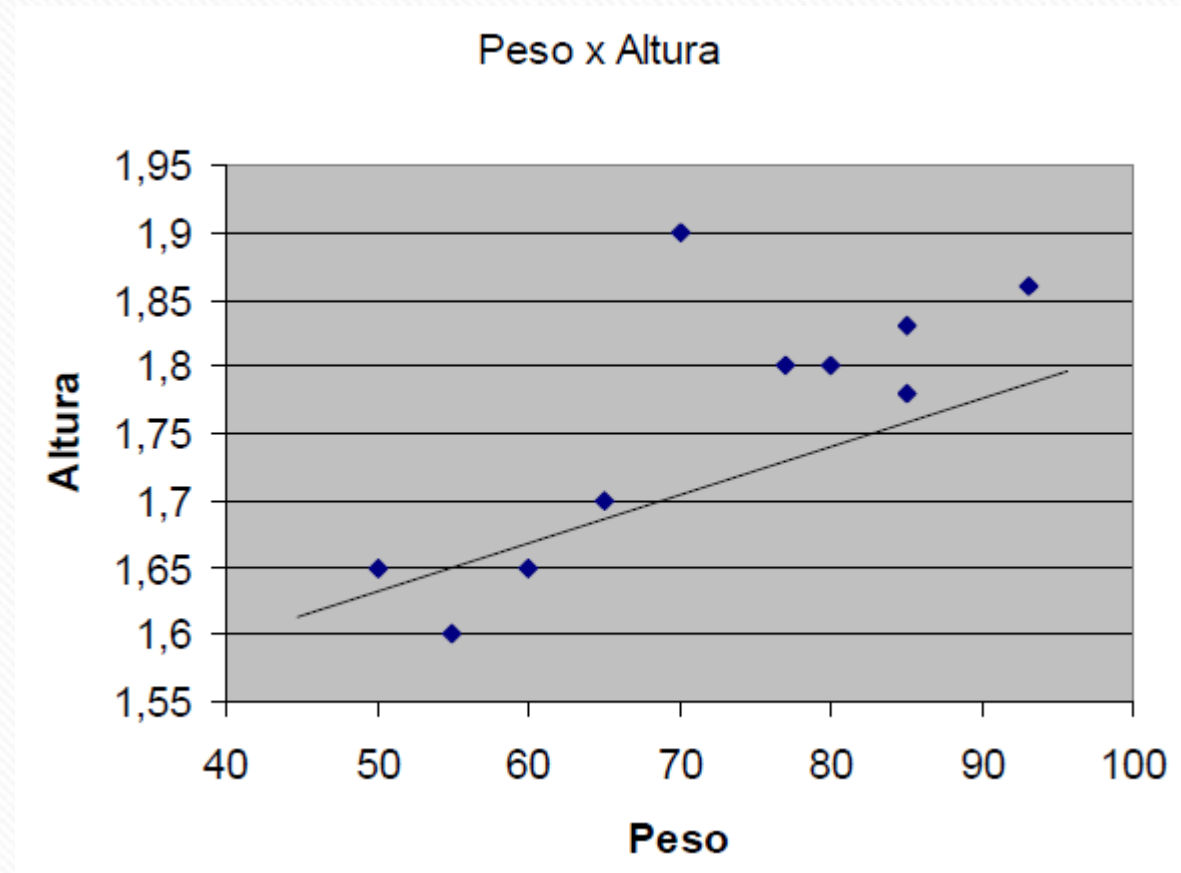
1. Quando uma delas está, de alguma forma, relacionada com a outra.
2. Quando a alteração no valor de uma variável (dita independente) provoca alterações no valor da outra variável (dita dependente)

Diagramas de Dispersão

- Um diagrama de dispersão mostra a relação entre duas variáveis quantitativas, medidas sobre os mesmos indivíduos.
- Os valores de uma variável aparecem no eixo horizontal, e os da outra, no eixo vertical.
- Comumente, coloca-se no eixo x um parâmetro.
- Cada indivíduo aparece como o ponto do gráfico definido pelos valores de ambas as variáveis para aquele indivíduo.

EXEMPLO DE CORRELAÇÃO ENTRE AS VARIÁVEIS: PESO E ALTURA

Peso (kg)	Altura (m)
80	1,80
85	1,83
50	1,65
70	1,90
55	1,60
77	1,80
85	1,78
93	1,86
65	1,70
60	1,65



IMPORTANTE:

Eixo das ABCISSAS : x

Variável que é alterada por uma modificação no processo (variável independente)

Eixo das ORDENADAS : y

Variável que pode mudar de acordo com a mudança da variável em ' x ' (variável dependente)

Diagramas de Dispersão

Análise – fatores a serem considerados:

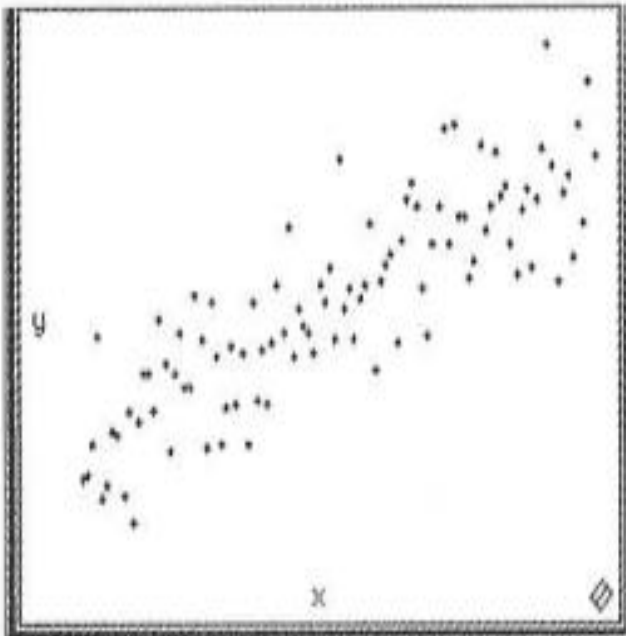
DIREÇÃO:

- Crescente;
- Decrescente.

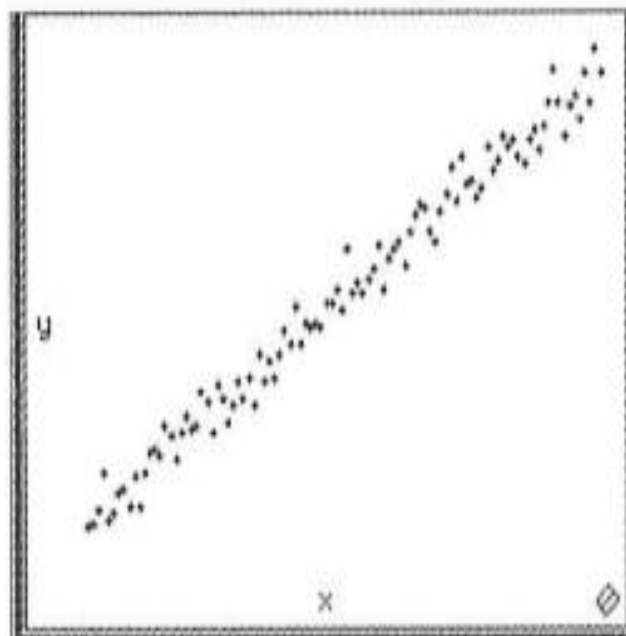
FORMA:

- Linear;
- Não-linear;
- Aglomerados.

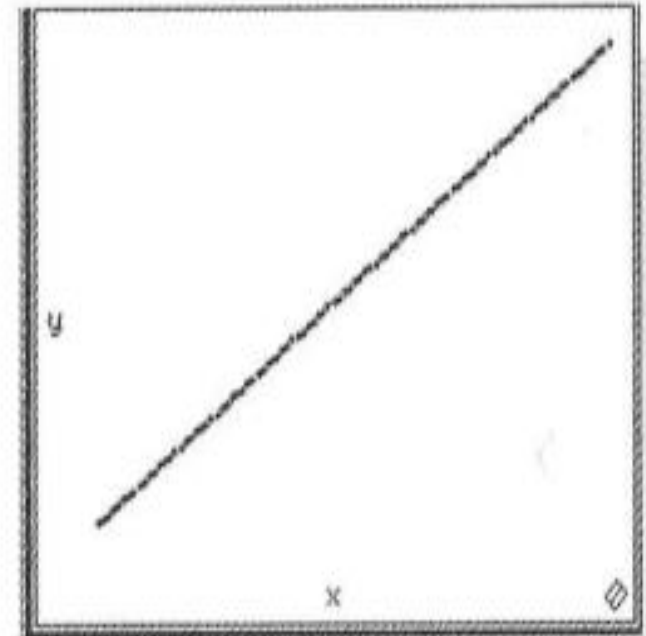
ALGUMAS SITUAÇÕES QUE PODEM SE APRESENTAR OS DIAGRAMAS DE DISPERSÃO



(a) Correlação positiva:
 $r = 0,851$

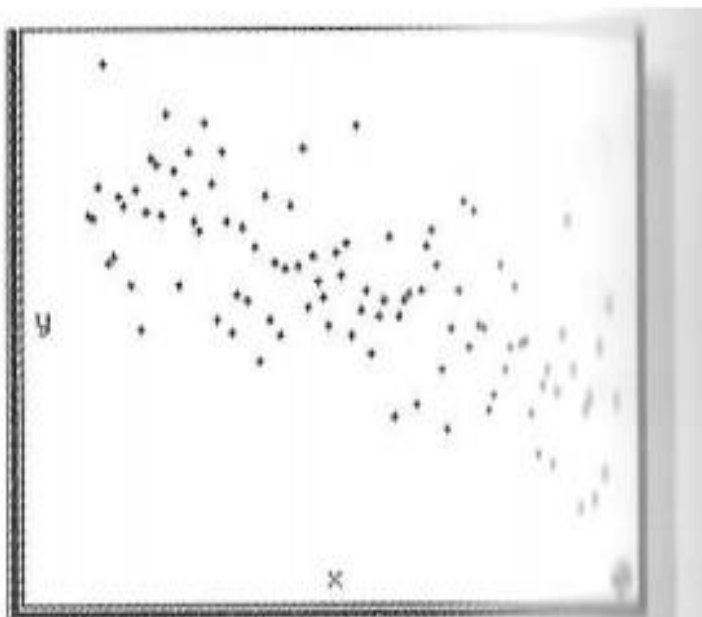


(b) Correlação positiva:
 $r = 0,991$

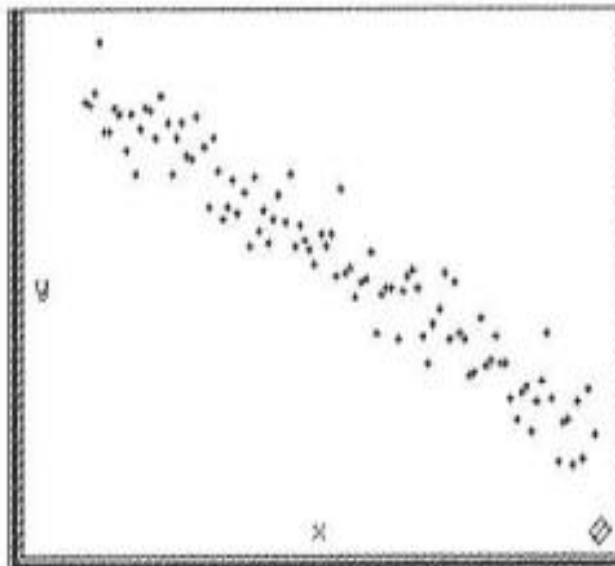


(c) Correlação positiva perfeita:
 $r = 1$

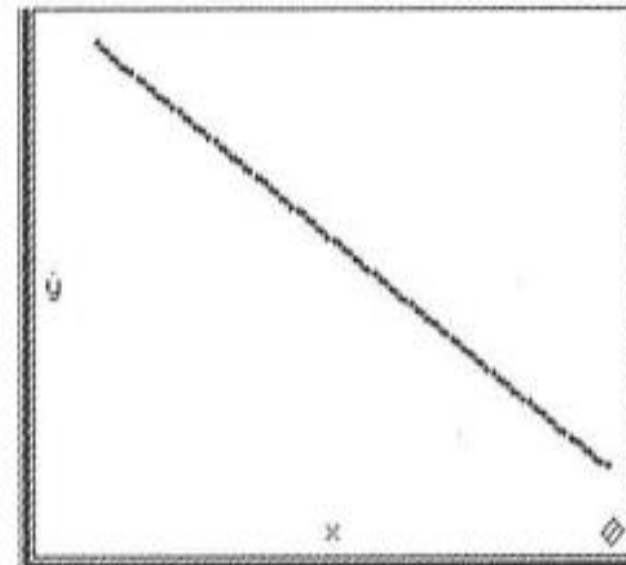
Algumas situações que podem se apresentar os diagramas de dispersão



(d) Correlação negativa:
 $r = -0,702$

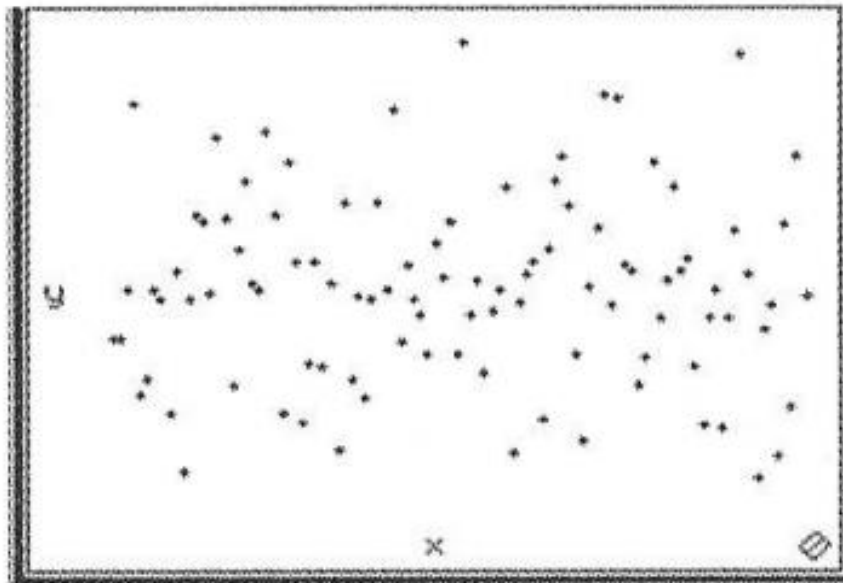


(e) Correlação negativa:
 $r = -0,965$

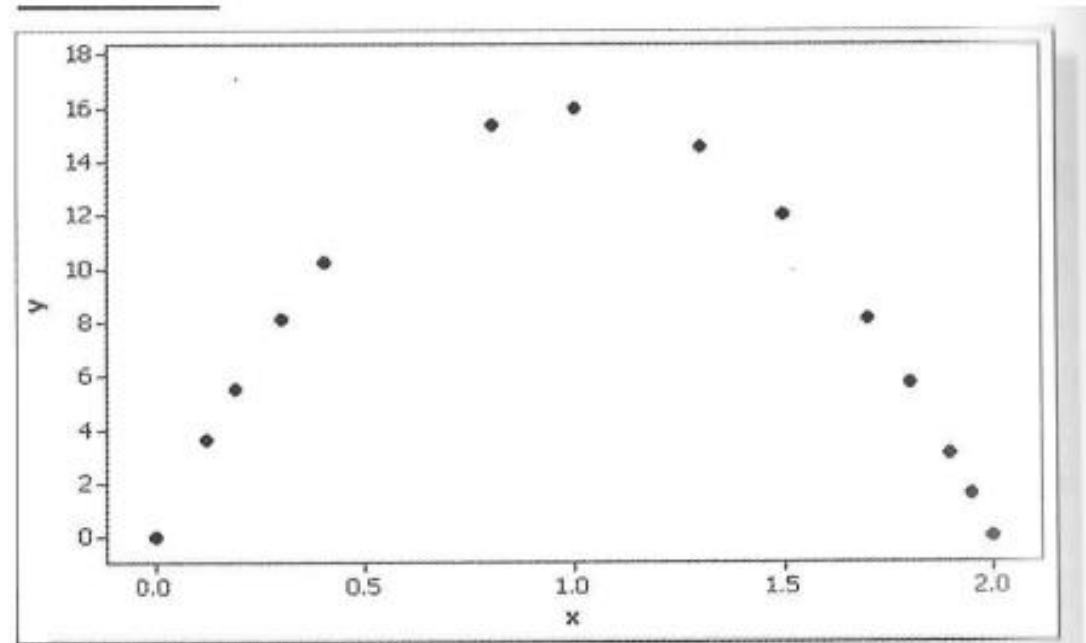


(f) Correlação negativa perfeita:
 $r = -1$

DIAGRAMAS DE DISPERSÃO (não há correlação linear)



(g) Nenhuma correlação: $r = 0$



(h) Relação não-linear: $r = -0,087$