

ANÁLISE DE ASSOCIAÇÃO

Associação entre variáveis quantitativas

Correlação

Estudar a relação entre duas variáveis quantitativas

Exemplos:

- Quantidade de vapor e temperatura num processo industrial
- Preço de ação da Telebrás e Índice da Bolsa de Valores de São Paulo
- Tempo de estudo e nota na prova
- Taxa de transportadores rodoviários de carga geral e Distância
- Processamento anual de armazéns e nível médio de estoque

Investigaremos a presença ou ausência de **relação linear** quantificando a força dessa relação: correlação.

Representação gráfica de duas variáveis quantitativas:
Diagrama de dispersão

Exemplo 1: nota da prova e tempo de estudo

X : tempo de estudo (em horas)

Y : nota da prova

Pares de observações (X_i, Y_i) para cada estudante

Tempo (X) Nota (Y)

3, 0 4, 5

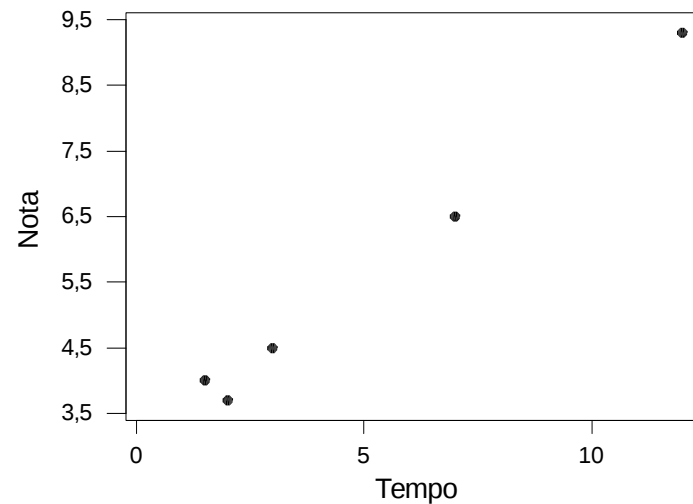
7, 0 6, 5

2, 0 3, 7

1, 5 4, 0

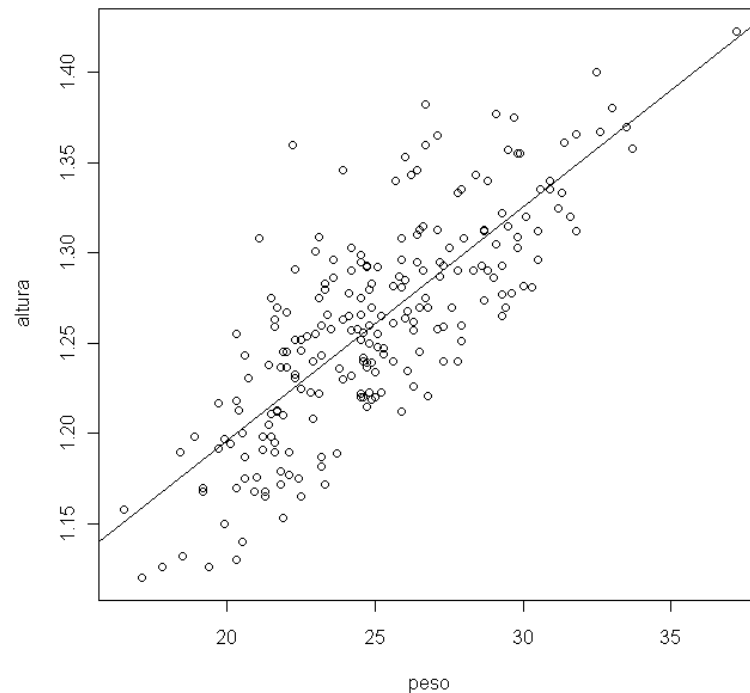
12, 0 9, 3

Diagrama de Dispersão



Coeficiente de correlação linear

É uma medida que avalia o quanto a “nuvem de pontos” no diagrama de dispersão aproxima-se de uma reta.



○ **coeficiente de correlação linear de *Pearson*** é calculado por:

$$r = \frac{\sum_{i=1}^n (X_i - \bar{X})(Y_i - \bar{Y})}{(n-1)S_X S_Y}$$

sendo que,

$\bar{X}$ e $\bar{Y}$ são as médias amostrais de X e Y , respectivamente,
 S_X e S_Y são os desvios padrão de X e Y , respectivamente.

Propriedade: $-1 \leq r \leq 1$

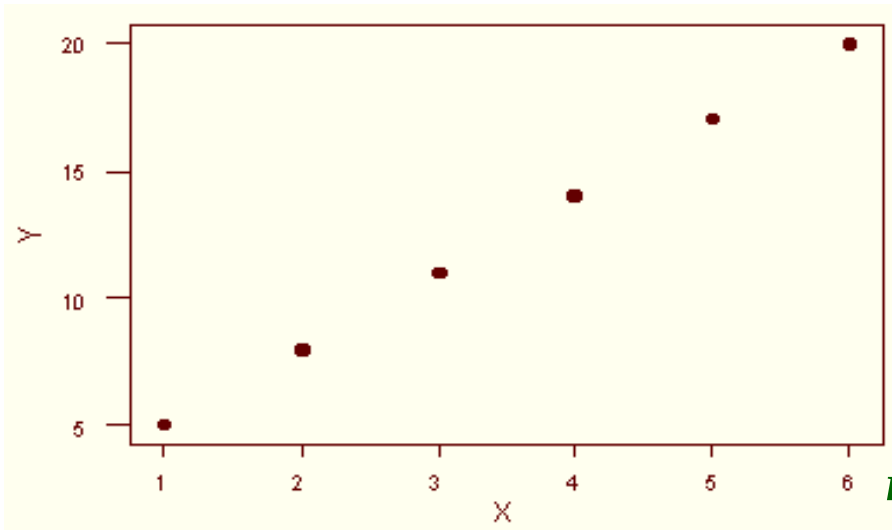
Casos particulares:

$r = 1 \Rightarrow$ correlação linear positiva e perfeita

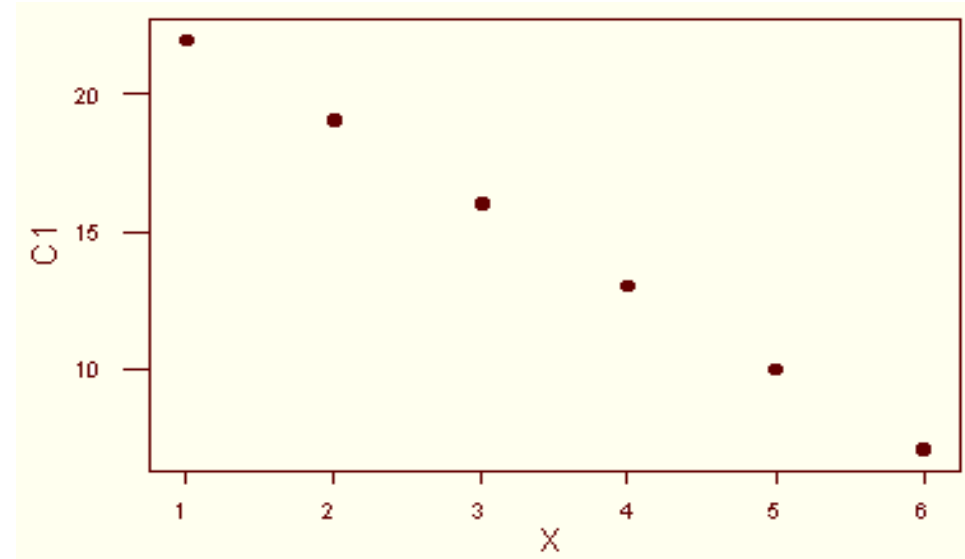
$r = -1 \Rightarrow$ correlação linear negativa e perfeita

$r = 0 \Rightarrow$ inexistência de correlação linear

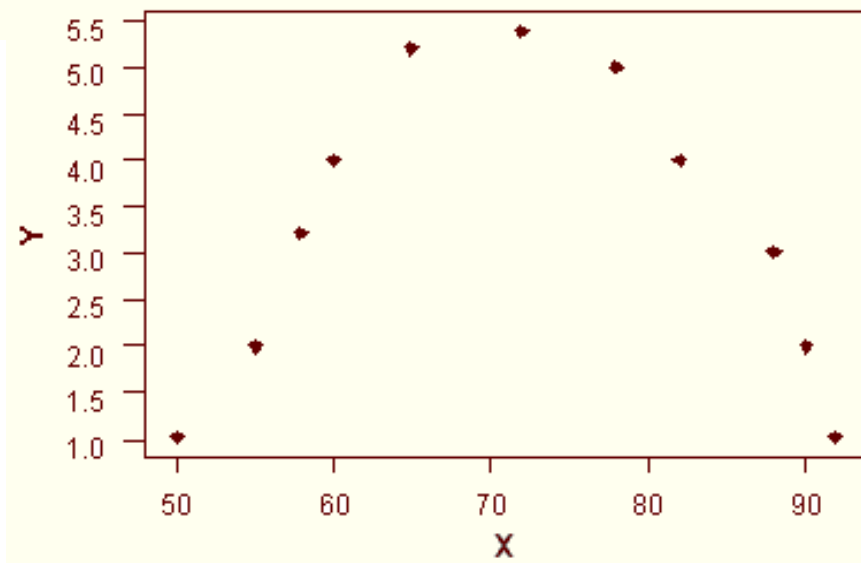
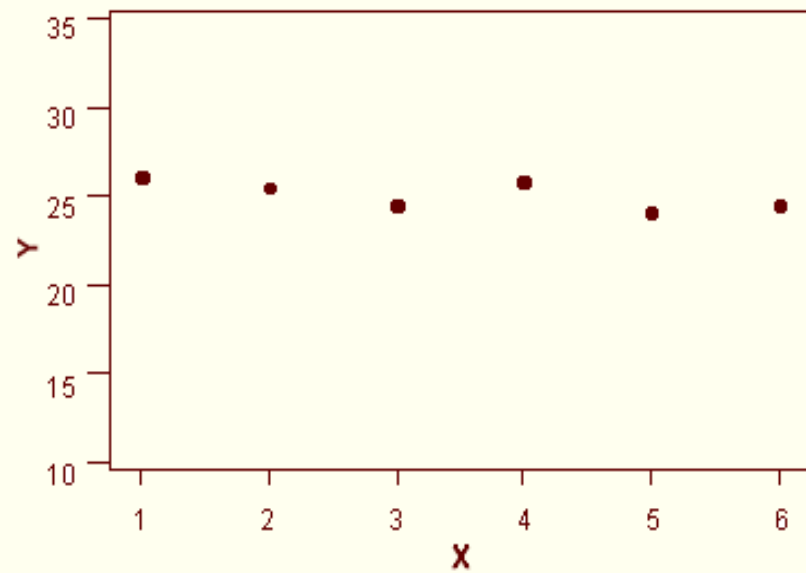
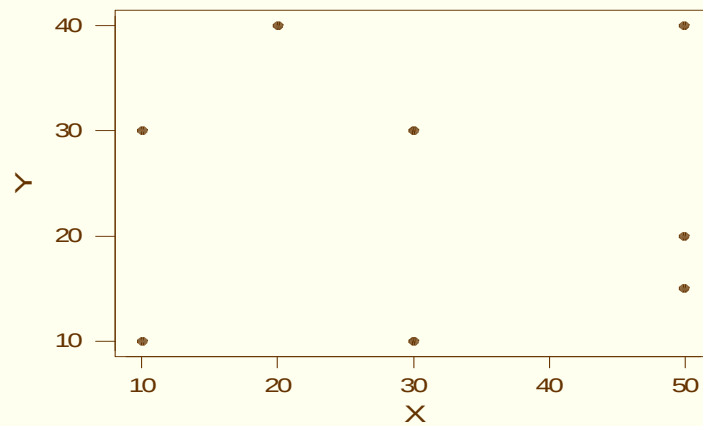
$r = 1$, correlação linear positiva e perfeita



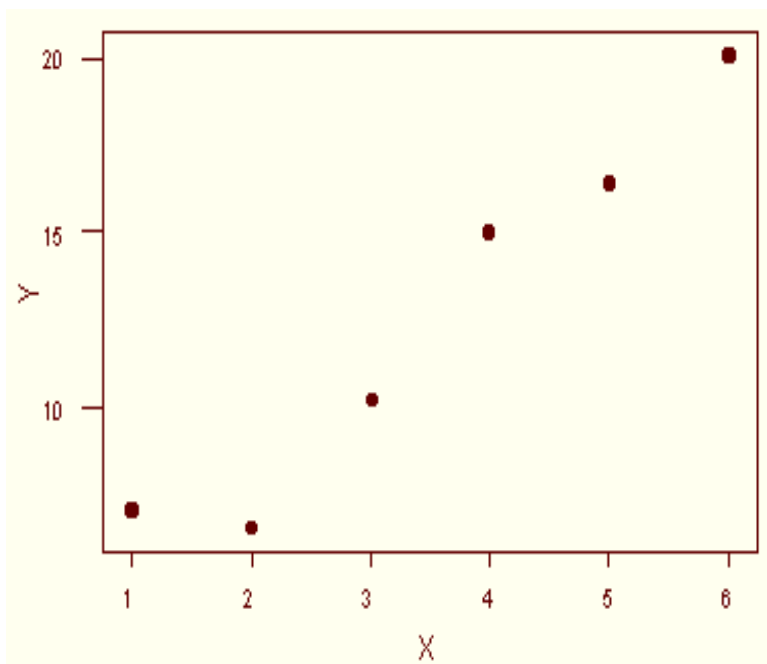
$r = -1$, correlação linear negativa e perfeita



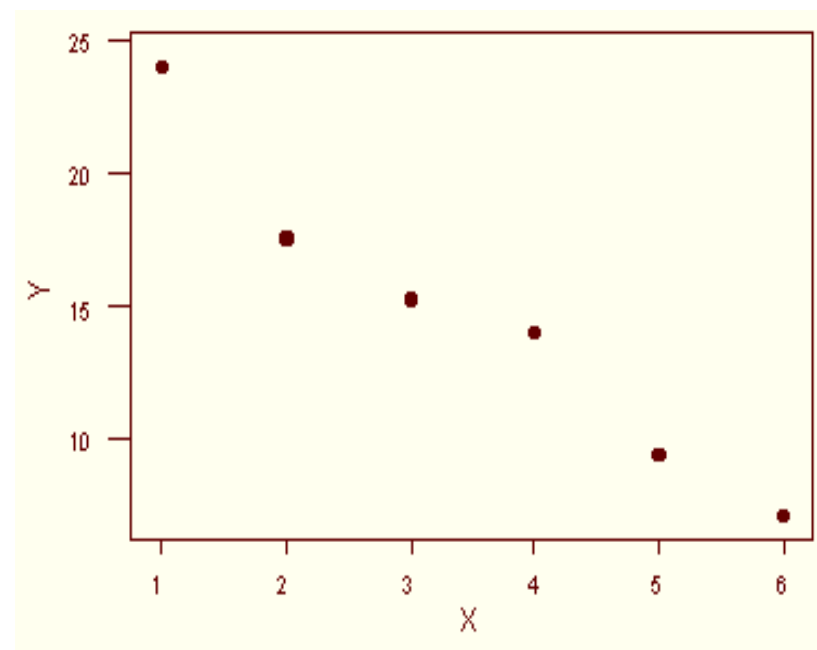
$$r \cong 0$$



$$r \cong 1$$



$$r \cong -1$$



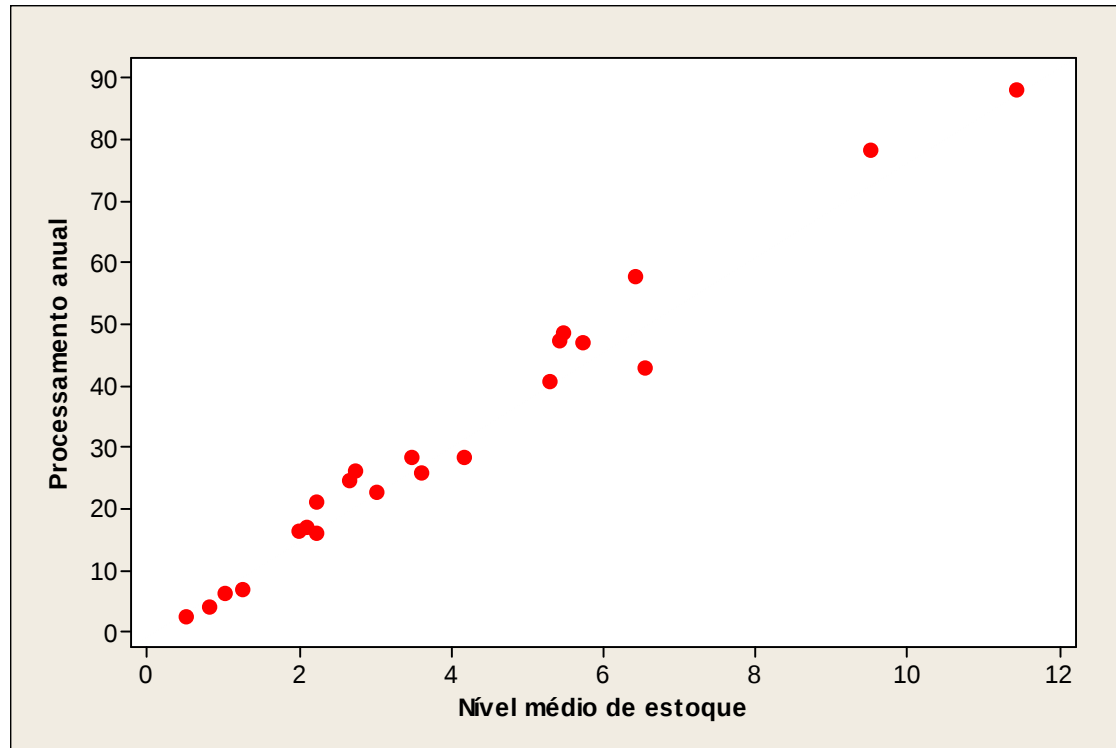
Exemplo 2: processamento anual do armazém (em \$) e nível médio do estoque (em \$)

Considere as duas variáveis observadas em 24 armazéns.

Y : processamento anual do armazém

X : nível médio do estoque

Diagrama de dispersão



Podemos notar que, conforme aumenta o nível médio de estoque (X), o processamento anual (Y) tende a aumentar. Nota-se também uma tendência linear. O valor de r é 0,986.

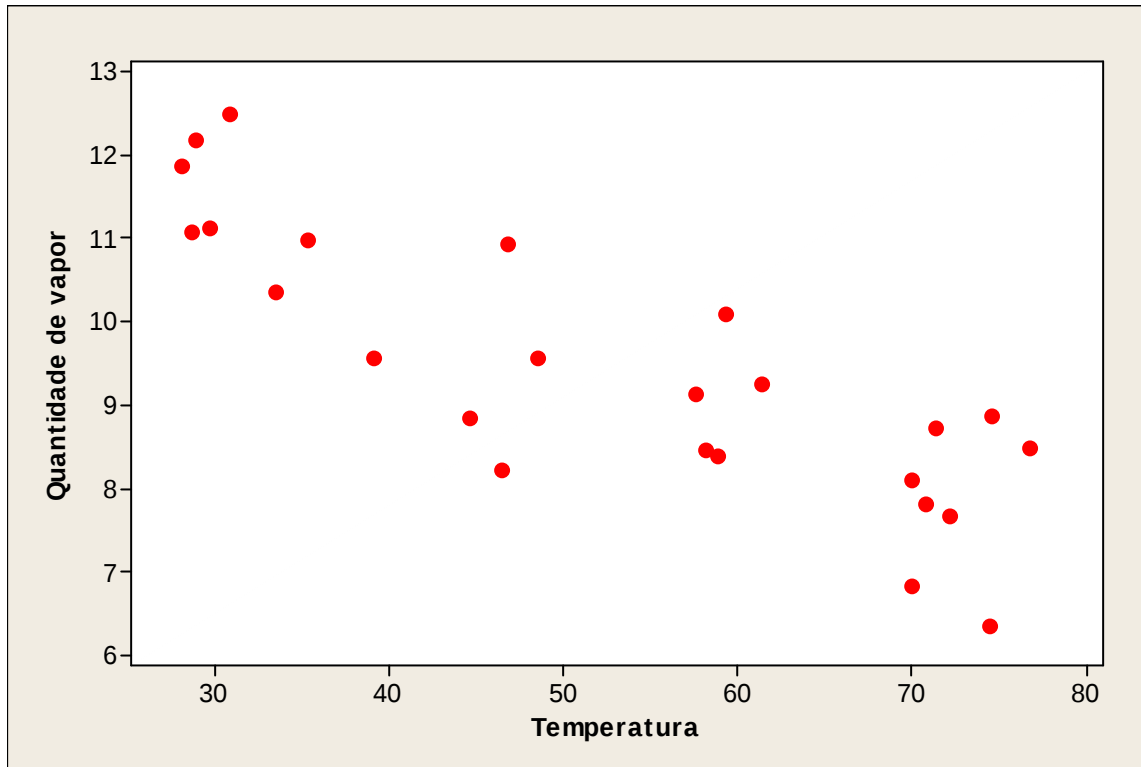
Exemplo 3: quantidade de vapor e temperatura (grau F)
de um processo industrial

Considere as duas variáveis observadas em 25 dias.

Y : quantidade de vapor

X : temperatura

Diagrama de dispersão



Podemos notar que, conforme aumenta a temperatura (X), a quantidade de vapor (Y) tende a diminuir. Nota-se também uma tendência linear. O valor de r é -0,845.

Limitações do coeficiente de correlação de *Pearson* r :

- quantifica somente o grau de **relacionamento linear** entre 2 variáveis quantitativas (podemos ter $r=0$ e existir um relacionamento não-linear significativo);
- a correlação estimada não pode ser extrapolada $\Rightarrow$ a relação entre X e Y pode mudar fora da região;
- alta correlação não implica necessariamente uma relação de causa e efeito;
- importante fazer o gráfico e não somente calcular r .

Comentário: Na interpretação do coeficiente de correlação é importante visualizar o diagrama de dispersão.

Suponha o seguinte exemplo: 6 variáveis são medidas em 11 indivíduos

Id	X	Y1	Y2	Y3	X4	Y4
1	10	8,04	9,14	7,46	8	6,58
2	8	6,95	8,14	6,77	8	5,76
3	13	7,58	8,74	12,74	8	7,71
4	9	8,81	8,77	7,11	8	8,84
5	11	8,33	9,26	7,81	8	8,47
6	14	9,96	8,10	8,84	8	7,04
7	6	7,24	6,13	6,08	8	5,25
8	4	4,26	3,10	5,39	19	12,50
9	12	10,84	9,13	8,15	8	5,56
10	7	4,82	7,26	6,42	8	7,91
11	5	5,68	4,74	5,73	8	6,89

Correlação de Pearson entre X e $Y1 = 0,816$

Correlação de Pearson entre X e $Y2 = 0,816$

Correlação de Pearson entre X e $Y3 = 0,816$

Correlação de Pearson entre $X4$ e $Y4 = 0,817$

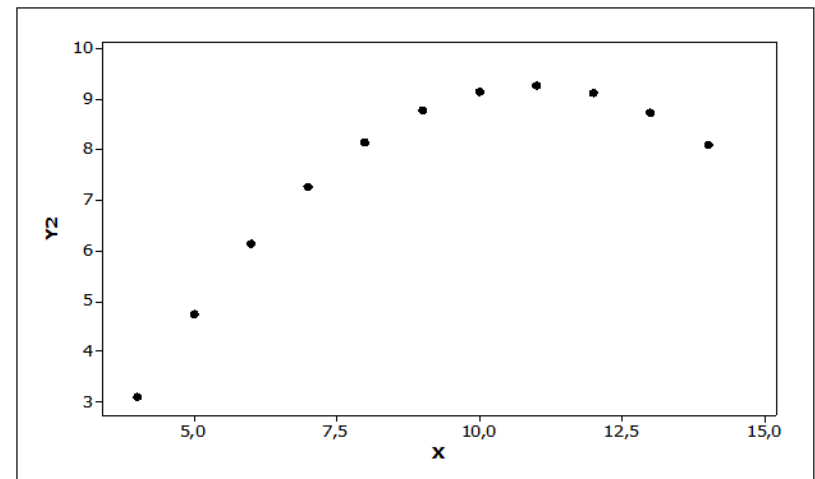
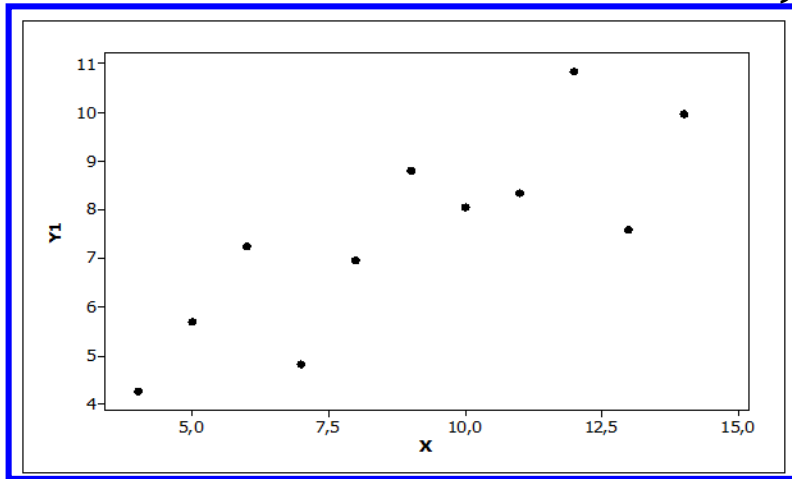
⇒ Mesmos valores de correlação.

⇒ Qual é a forma esperada da dispersão conjunta destas variáveis?

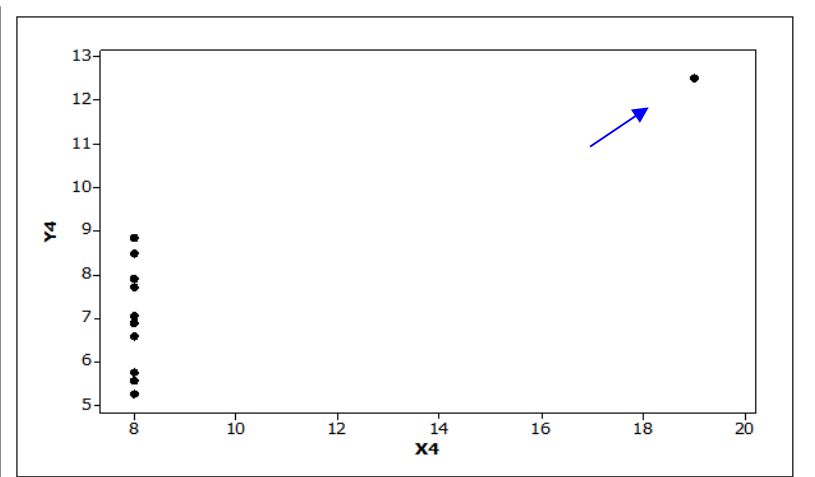
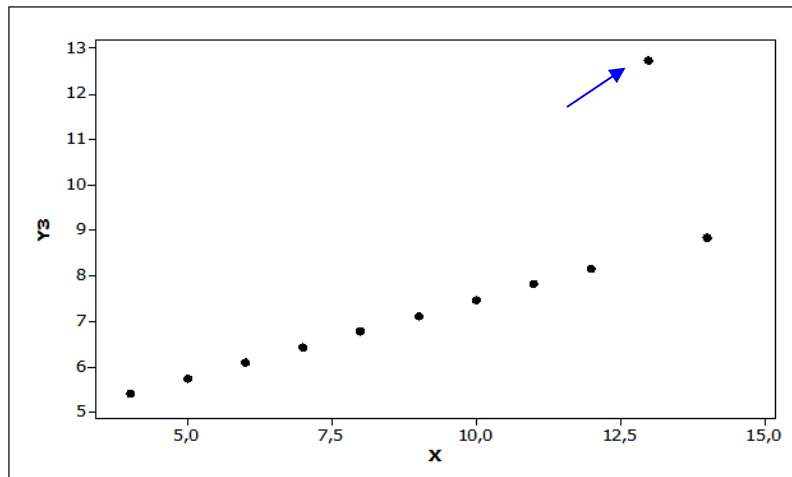
Diagramas de Dispersão e Coeficientes de Correlação

$$r = 0,816$$

Dispersão
esperada!



Pontos
influentes!



Associação entre variáveis qualitativas

Tabelas de Contingência

Podemos construir tabelas de frequências conjuntas (*tabelas de contingência*), relacionando duas variáveis qualitativas.

Exemplo 1:

Variáveis qualitativas do estudo:

- Curso (completo);
- Anos de empresa: categorizada por “até 3 anos” e “mais de 3 anos”).

Tabela de Contingência

Há indícios de associação entre Curso e Anos de empresa?

Curso	Anos de Empresa		Total
	Até 3	Mais de 3	
Fundamental	6	1	7
Médio	4	3	7
Superior	0	6	6
Total	10	10	20

Qual é o significado dos valores desta tabela?

Tabela de Contingência

Verificar associação através da:

- porcentagem segundo as colunas, ou
- porcentagem segundo as linhas.

Curso	Anos_cat		Total
	Ate 3	Mais que 3	
F	85.7	14.3	100
M	57.1	42.9	100
S	0.0	100.0	100
Total	50.0	50.0	100

Comparando as porcentagens de cada uma das linhas, observamos uma grande diferença com relação à porcentagem total.

Aparentemente, há associação entre Curso e Anos de empresa.

Teste de hipótese estatístico para verificar associação entre variáveis qualitativas

➤ Teste Qui-quadrado de Independência

Objetivo: Verificar se existe independência entre duas variáveis medidas nas mesmas unidades experimentais.

Hipótese testada num **teste de independência:**

H_0 (hipótese nula): A e B são variáveis independentes

Teste de independência

Regra de decisão:

Baseada no nível descritivo do teste, **P** ou **valor- P** , obtido dos dados e usando a distribuição de probabilidade Qui-quadrado, fornecida pelos pacotes computacionais.

→ Estabelece-se um valor para **α , o nível de significância do teste**. Em geral, α é fixado em 1%, 5% ou 10%.

→ Se, para **α fixado**,

$P \leq \alpha \Rightarrow$ rejeitamos a hipótese H_0 de independência;

$P > \alpha \Rightarrow$ não rejeitamos a hipótese H_0 .