

- Dispersão e correlação linear
- Ajuste Linear, quadrático, exponencial e hiperbólico

É comum relacionarmos grandezas tais como preço e quantidade produzida, preço e faturamento, custo e quantidade produzida, etc.

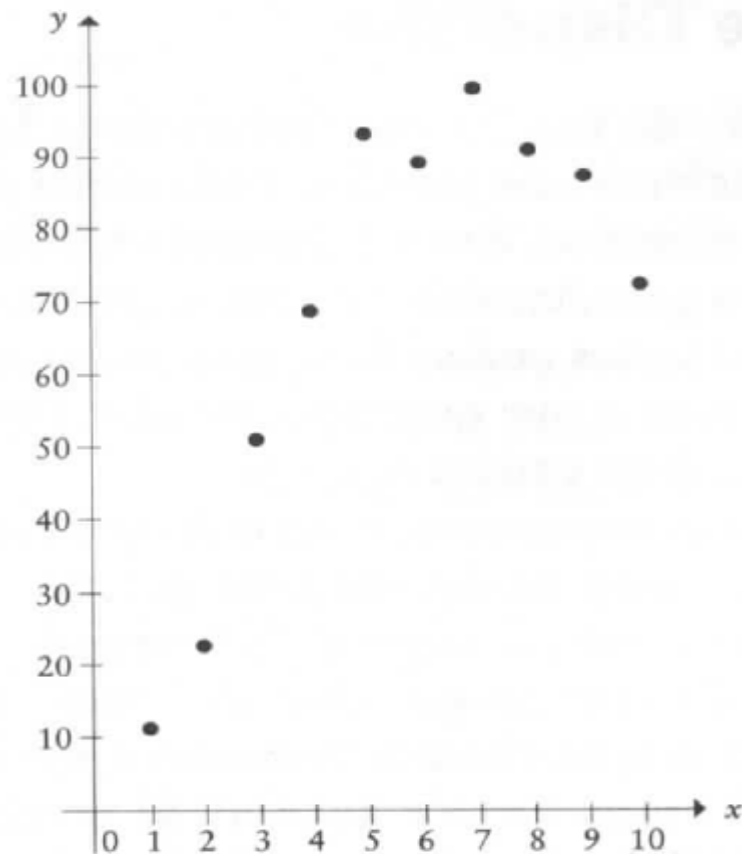
Notamos que essas relações podem ficar mais bem caracterizadas quando definimos variáveis para o que queremos estudar. Desse ponto de vista, é comum trabalhar com duas variáveis, x e y .

Naturalmente, ao estudar as relações entre as variáveis, estamos interessados em analisar e interpretar o comportamento das grandezas relacionadas.

Uma maneira inicial de analisar e interpretar a relação entre as variáveis é a elaboração de diagramas de dispersão, que permitem a visualização do comportamento entre as variáveis estabelecidas.

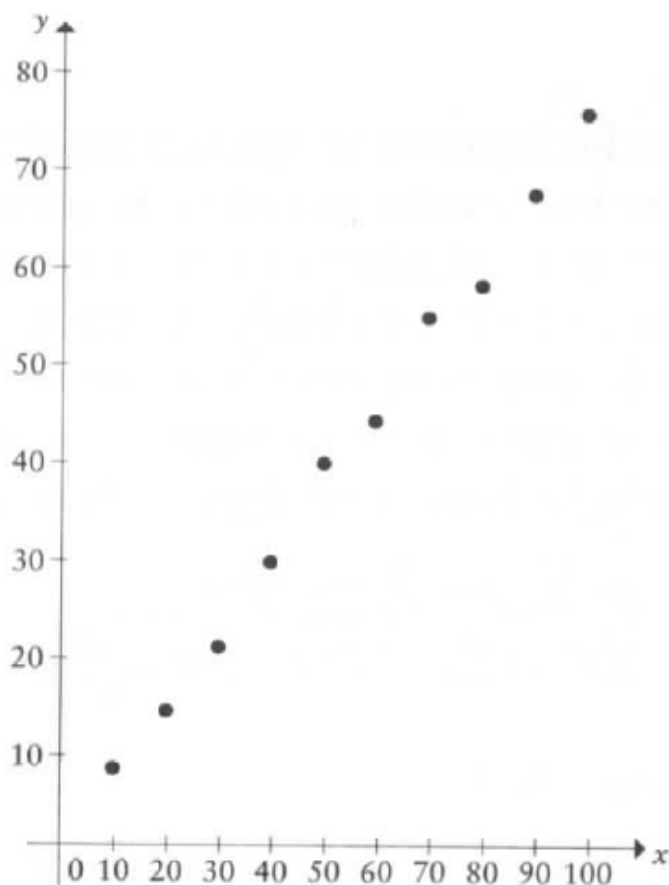
Considerando, por exemplo, a produção de um tecido em relação a quantidade de insumo utilizada em sua produção, conforme a tabela seguinte, podemos esboçar o diagrama de dispersão para a produção de tecido.

Insumo (x)	1	2	3	4	5	6	7	8	9	10
Produção (y)	10	21	49	67	91	87	97	89	85	70

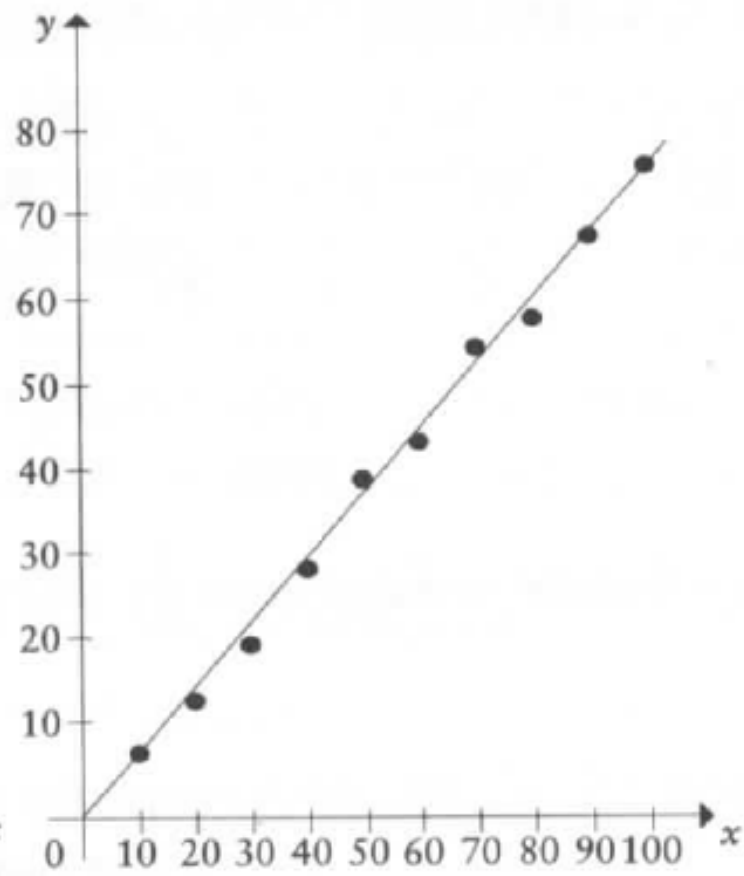
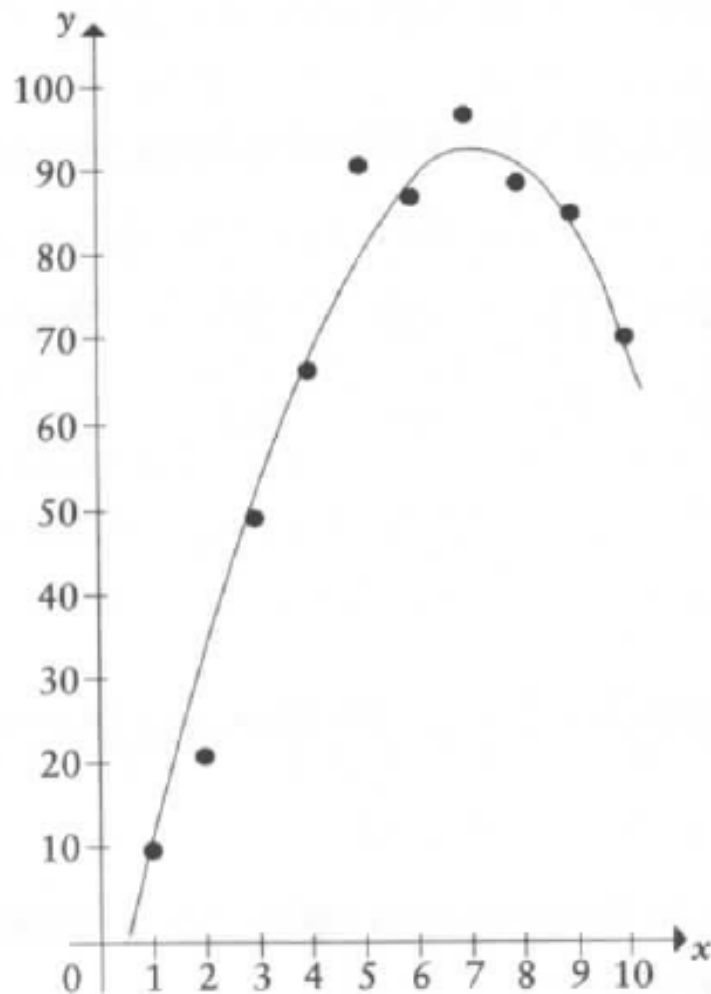


Ou ainda, se considerarmos o custo para o transporte de tecidos relacionado as distâncias a serem percorridas, conforme a tabela seguinte, podemos esboçar o diagrama de dispersão do custo para o transporte.

Distância (x)	10	20	30	40	50	60	70	80	90	100
Custo (y)	8	15	21	30	40	44	55	59	68	75



Observando a disposição dos pontos nos diagramas anteriores, podemos traçar a mão uma curva e uma reta que se aproxima dos pontos distribuídos.



Correlação Linear

Como os pontos do segundo diagrama se aproximam de uma reta, dizemos que existe uma **correlação linear** entre as variáveis que originam tal diagrama.

Nesses casos, podemos calcular o valor da correlação linear entre tais variáveis e, de acordo com o valor obtido, avaliar se os pontos se aproximam pouco ou muito de uma reta; em outras palavras, por meio desse cálculo, podemos avaliar se existe forte ou fraca correlação linear.

O coeficiente de correlação linear será representado por r e é dado por:

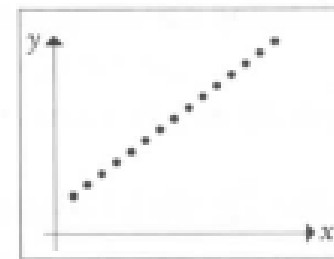
$$r = \frac{n \cdot \sum xy - (\sum x) \cdot (\sum y)}{\sqrt{[n \cdot \sum x^2 - (\sum x)^2] \cdot [n \cdot \sum y^2 - (\sum y)^2]}}$$

e varia entre o intervalo $[-1,1]$

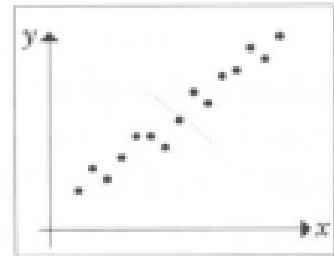
Algumas possibilidades
para r e sua interpretação
gráfica

Teremos uma reta perfeita
quando
 $r = +1$ ou $r = -1$

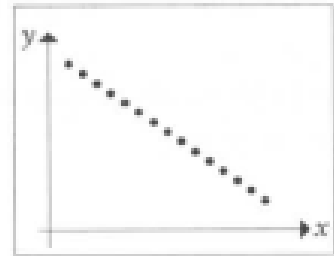
- $r = +1$
Perfeitamente Positiva



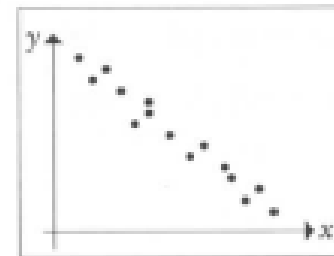
- $r \approx 0,7$
Moderadamente Positiva



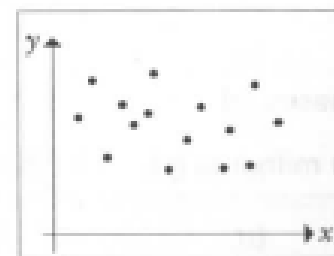
- $r = -1$
Perfeitamente Negativa



- $r \approx -0,7$
Moderadamente Negativa



- $r = 0$
Correlação Linear Nula



Passos para o ajuste do modelo de regressão **linear simples** (reta)

1) Coleta de dados

x_i	y_i	$x_i \cdot y_i$	x_i^2
x_1	y_1	$x_1 \cdot y_1$	x_1^2
x_2	y_2	$x_2 \cdot y_2$	x_2^2
x_3	y_3	$x_3 \cdot y_3$	x_3^2
.	.	.	.
.	.	.	.
.	.	.	.
Σx	Σy	Σxy	Σx^2

2) Cálculo da média simples das variáveis x e y

$$\bar{x} = \frac{\Sigma x}{n} \quad \bar{y} = \frac{\Sigma y}{n}$$

3) Cálculo do coeficiente angular da reta

$$\hat{a} = \frac{\Sigma xy - n \cdot \bar{x} \cdot \bar{y}}{\Sigma x^2 - n \cdot (\bar{x})^2}$$

4) Cálculo do coeficiente linear da reta

$$\hat{b} = \bar{y} - \hat{a} \cdot \bar{x}$$

Depois disto, tem-se então o modelo linear dado por

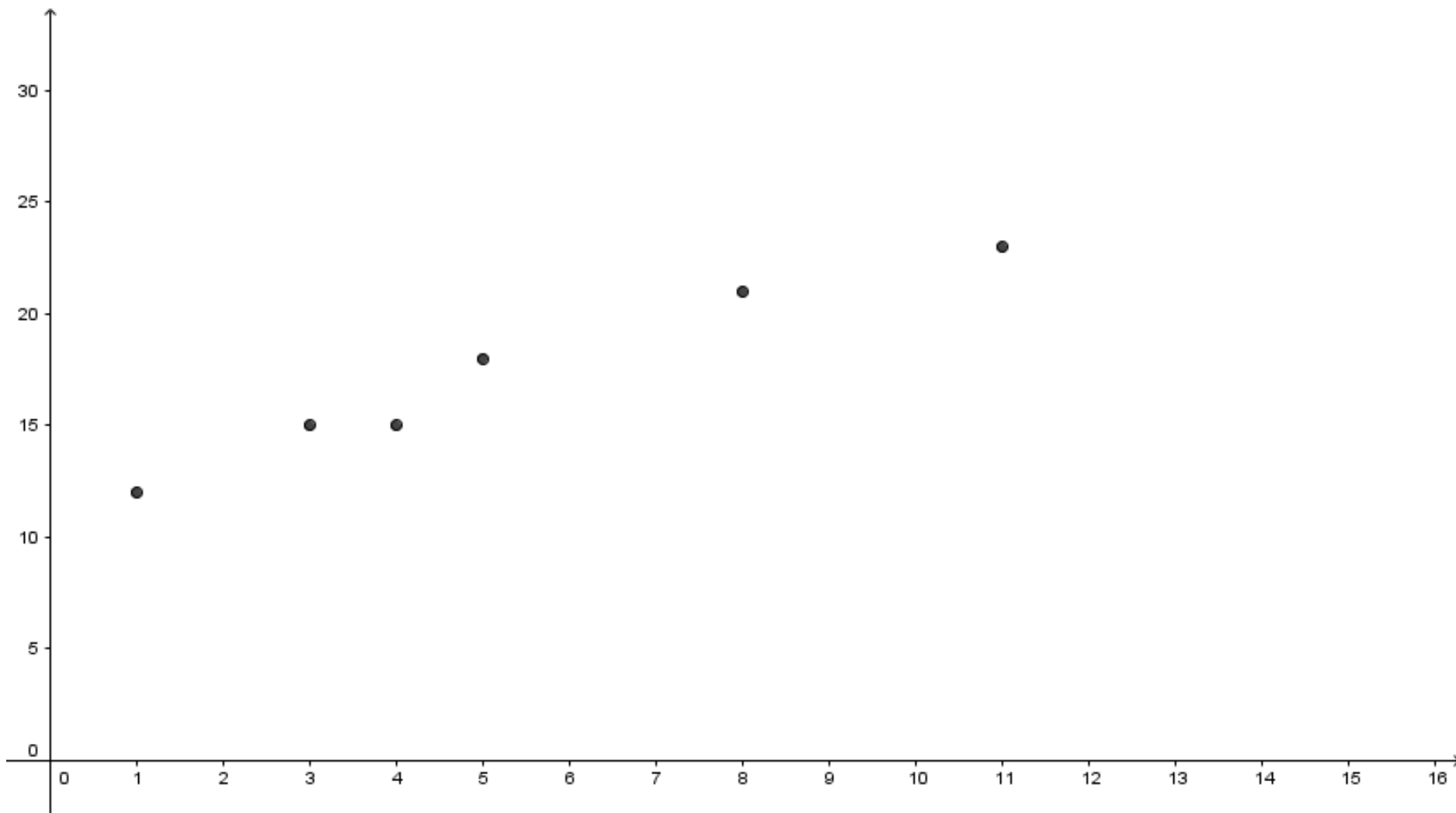
$$\hat{y} = \hat{a} \cdot x + \hat{b}$$

Exemplo: Considere o seguinte conjunto de dados

x	y
1	12
3	15
4	15
5	18
8	21
11	23

Obtenha o gráfico de dispersão e ajuste uma reta aos dados em questão

Gráfico de dispersão



1) Coleta de dados

2) Cálculo da média simples das variáveis x e y

$$\bar{x} = \frac{\sum x}{n} = \frac{32}{6} \cong 5,3$$
$$\bar{y} = \frac{\sum y}{n} = \frac{104}{6} \cong 17,3$$

x	y	xy	x^2
1	12	12	1
3	15	45	9
4	15	60	16
5	18	90	25
8	21	168	64
11	23	253	121
Soma			
32	104	628	236

3) Cálculo do coeficiente angular da reta

$$\hat{a} = \frac{\sum xy - n.\bar{x}.\bar{y}}{\sum x^2 - n(\bar{x})^2} \cong \frac{628 - 6.5,3.17,3}{236 - 6.(5,3)^2} \cong 1,2$$

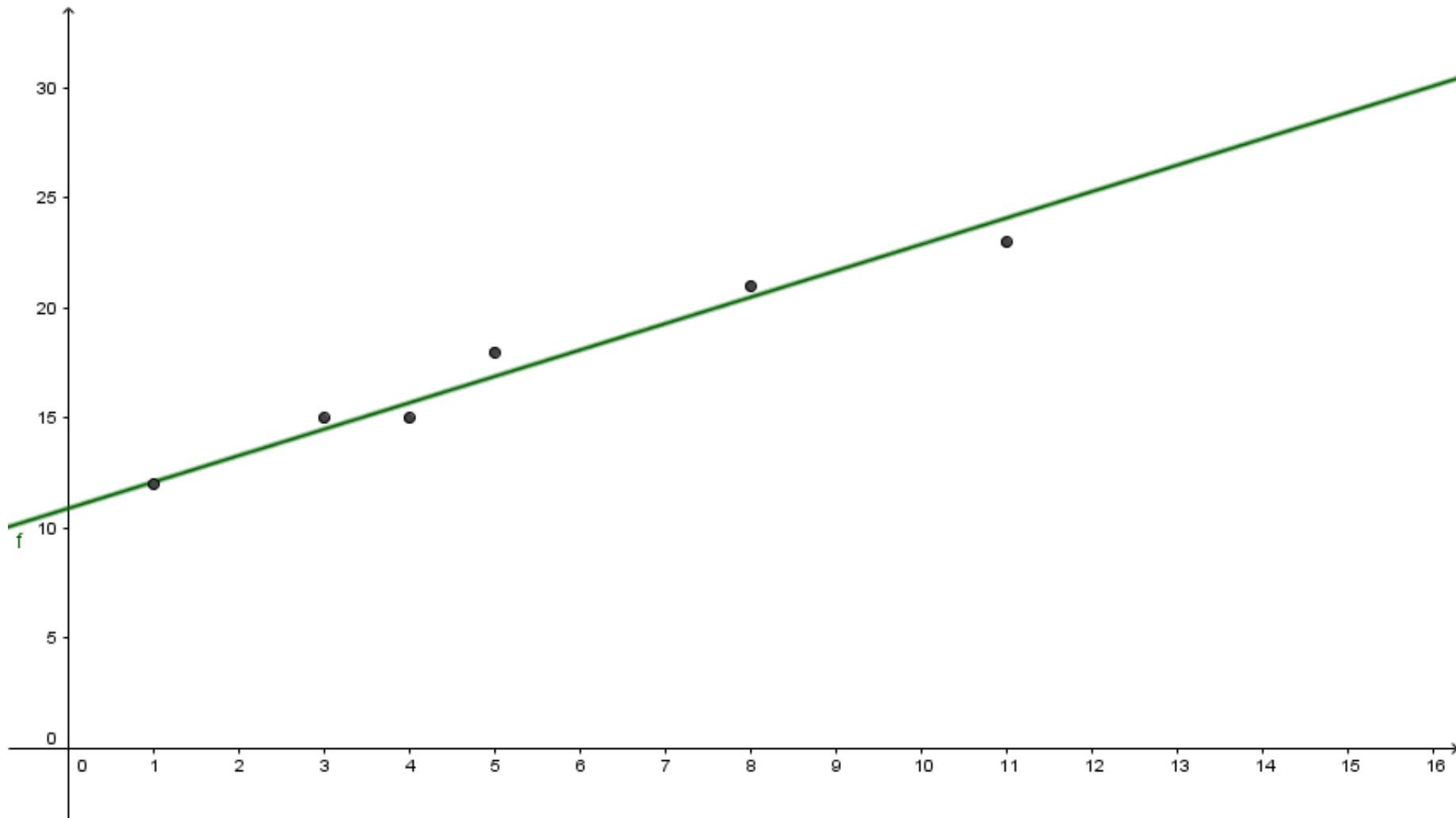
4) Cálculo do coeficiente linear da reta

$$\hat{b} = \bar{y} - \hat{a}\bar{x} \cong 17,3 - 1,2.5,3 \cong 10,9$$

O modelo aproximado será então

$$y = 1,2x + 10,9$$

Gráfico de dispersão com o ajuste



Exercício proposto 1

Uma empresa de embalagens plásticas, preocupada com a demanda (y_i) de seu produto, resolveu elaborar um estudo sobre as variações dos preços de venda (x_i). Após esse estudo e levantamento de dados, obteve as informações condensadas na tabela a seguir.

Meses	Jan.	Fev.	Mar.	Abr.	Mai	Jun.	Jul.	Ago.	Set.
Preço de venda (x_i)	16	18	20	23	26	28	30	33	35
Demanda (y_i)	1.200	1.150	950	830	800	760	700	690	670

A partir dessas informações, responda às questões relativas aos itens:

- Construindo o diagrama de dispersão, podemos afirmar, quanto à sua evolução, que o sistema se comporta de forma aproximadamente linear?
- Após ter construído o diagrama de dispersão, os pontos apresentam um comportamento linear crescente ou decrescente?
- As variáveis demanda e preço de mercado caminham, em termos de evolução, no mesmo sentido ou em sentidos contrários?
- Calcule e interprete o coeficiente de correlação linear.
- Estabeleça a equação de regressão de y (demanda) em relação a x (preço).
- Represente em um mesmo sistema de eixos a dispersão dos dados (x ; y) e a reta de regressão.
- Qual a previsão da demanda, quando os preços atingirem os patamares de $x = 40$ e $x = 50$?

Passos para o ajuste do modelo de regressão **quadrática** (parábola)

1) Coleta de dados

X_i	y_i	$x_i \cdot y_i$	x_i^2	x_i^3	x_i^4	$x_i^2 \cdot y_i$
X_1	y_1	$x_1 \cdot y_1$	x_1^2	x_1^3	x_1^4	$x_1^2 \cdot y_1$
X_2	y_2	$x_2 \cdot y_2$	x_2^2	x_2^3	x_2^4	$x_2^2 \cdot y_2$
X_3	y_3	$x_3 \cdot y_3$	x_3^2	x_3^3	x_3^4	$x_3^2 \cdot y_3$
\vdots	\vdots	\vdots	\vdots	\vdots	\vdots	\vdots
Σx	Σy	Σxy	Σx^2	Σx^3	Σx^4	$\Sigma x^2 \cdot y$

2) Resolver o sistema abaixo, encontrando os parâmetros da função quadrática

$$\begin{cases} \hat{a} \cdot \Sigma x^4 + \hat{b} \cdot \Sigma x^3 + \hat{c} \cdot \Sigma x^2 = \Sigma x^2 y \\ \hat{a} \cdot \Sigma x^3 + \hat{b} \cdot \Sigma x^2 + \hat{c} \cdot \Sigma x = \Sigma xy \\ \hat{a} \cdot \Sigma x^2 + \hat{b} \cdot \Sigma x + \hat{c} \cdot n = \Sigma y \end{cases}$$

Depois disto, tem-se então o modelo quadrático dado por

$$y = \hat{a} \cdot x^2 + \hat{b} \cdot x + \hat{c}$$

Exercício proposto 2

Um fazendeiro, cuja principal cultura é o café, recentemente resolveu industrializar e comercializar a produção e vem acompanhando atentamente, ao longo de seis trimestres, a evolução da demanda de seu produto, com o objetivo de melhoria de resultados nas vendas. Ele solicitou ao departamento responsável de sua empresa dados relativos à demanda observada em função da ocorrência desses seis trimestres. Rapidamente, os resultados foram gerados e alocados em uma tabela.

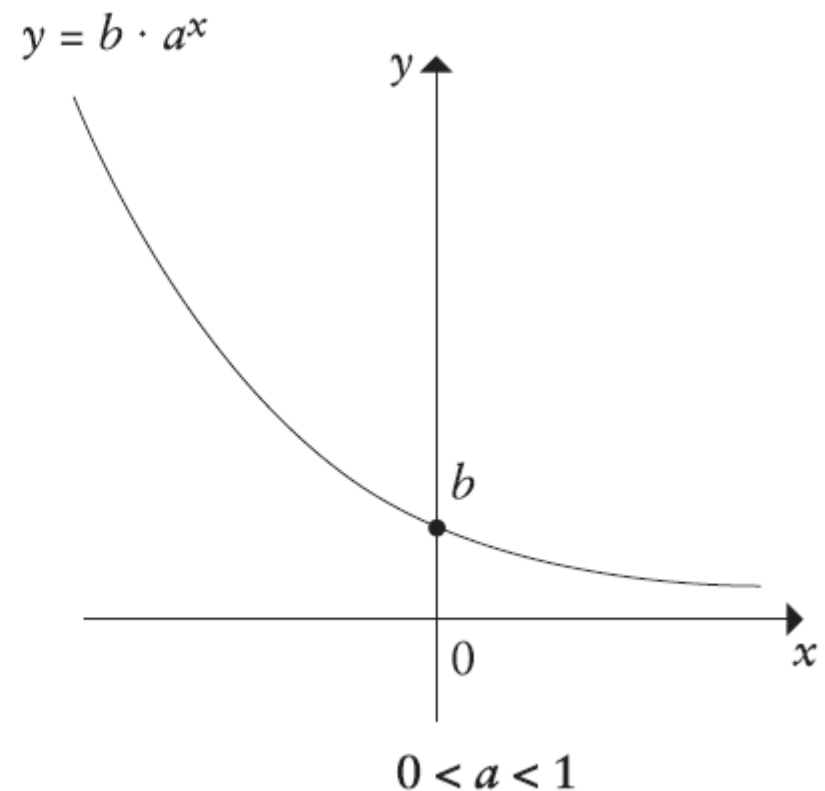
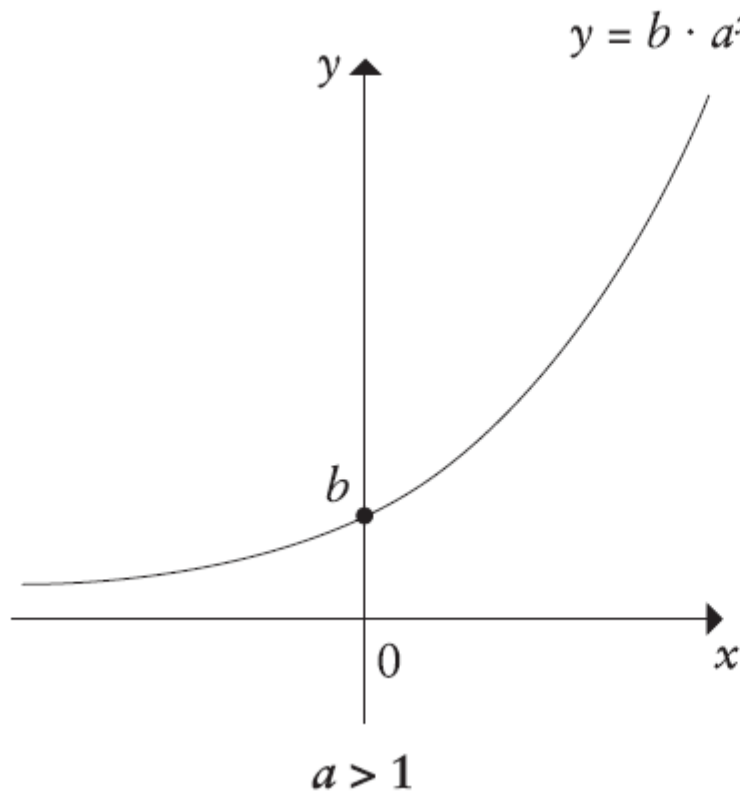
Trimestres (x_i)	1	2	3	4	5	6
Demanda observada (y_i) (em unidades)	4.800	3.500	3.850	5.200	7.300	10.950

A partir dos dados levantados:

- Construa o sistema de dispersão para a demanda em função dos trimestres e verifique se o comportamento desse sistema se aproxima de uma curva parabólica.
- Calcule e comente o coeficiente de correlação (r).
- Procure ajustar uma curva parabólica ($\hat{y} = \hat{a} \cdot x^2 + \hat{b} \cdot x + \hat{c}$) aos dados coletados.
- Uma vez ajustada a curva do 2º grau, calcule o vértice e interprete o resultado obtido.
- Segundo a função obtida no item (c), determine os intervalos de crescimento e decréscimo para a demanda.
- Estime a demanda para os dois trimestres seguintes.
- Partindo do pressuposto de que a demanda alcança um nível de 15.000 unidades, determine quando isso ocorre.

Passos para o ajuste do modelo de regressão exponencial

Observação: Uma função exponencial tem a característica de apresentar comportamento curvilíneo.



1) Coleta de dados

x_i	y_i	$\ln y_i$	$x_i \cdot \ln y_i$	x_i^2
x_1	y_1	$\ln y_1$	$x_1 \cdot \ln y_1$	x_1^2
x_2	y_2	$\ln y_2$	$x_2 \cdot \ln y_2$	x_2^2
x_3	y_3	$\ln y_3$	$x_3 \cdot \ln y_3$	x_3^2
.
.
.
Σx	Σy	$\Sigma \ln y$	$\Sigma x \ln y$	Σx^2

2) Cálculo dos parâmetros $\hat{a} = e^A$ e $\hat{b} = e^B$ onde

$$A = \frac{\sum (x \cdot \ln y) - \frac{\sum x \sum \ln y}{n}}{\sum x^2 - \frac{(\sum x)^2}{n}} \quad \text{e} \quad B = \frac{\sum \ln y}{n} - \frac{A \sum x}{n}$$

Depois disto, tem-se então o modelo exponencial dado por

$$y = \hat{b} \cdot \hat{a}^x$$

Exercício proposto 3

Uma empresa, observando o crescimento da oferta de seu produto em relação aos preços praticados no mercado em que atua, disponibilizou esses dados na tabela a seguir, de acordo com a variação de um período de tempo.

Meses	Jan.	Fev.	Mar.	Abr.	Mai	Jun.	Jul.	Ago.	Set.
Preço (x) (reais)	10	12	14	15	17	19	21	23	26
Oferta (y) (unidades)	100	120	135	167	198	220	268	310	390

De acordo com tais informações:

- Construa o sistema de dispersão das variáveis x e y .
- Observando o gráfico do sistema de dispersão, verifica-se um crescimento da oferta em função dos preços com tendência exponencial?
- Diante das observações dos itens (a) e (b), estabeleça a regressão exponencial de y sobre x .
- Construa, em um mesmo sistema de eixos, a dispersão e a curva exponencial ajustada no item (c).

Exercício proposto 4

Um índice econômico está evoluindo de acordo com sua variação anual. Esses dados levantados estão expressos na tabela a seguir:

Anos (x)	1	2	3	4	5	6	7
Índice acumulado (y)	100	120	150	318	622	870	1.450

De acordo com os dados:

- Represente graficamente o diagrama de dispersão das variáveis x e y . Observando o diagrama de dispersão, pode-se afirmar que o sistema tem características de um modelo exponencial?
- Estabeleça a regressão exponencial para o índice acumulado (y) em função dos anos (x).
- Em um mesmo sistema de eixos, construa o diagrama de dispersão e o gráfico da função exponencial ajustada.
- Projete o índice acumulado para os anos 8 e 9.
- Determine a taxa de crescimento anual para esse índice.