



Matemática Computacional

Teórica 7

Departamento de Matemática
Instituto Superior de Engenharia do Porto

2º Semestre 20-21

Conteúdo

- 1 Regressão Linear Simples
- 2 Inferências
- 3 Anova
- 4 Regressão linear simples no Excel

Regressão Linear Simples

RLS

No processo de tomada de decisões é, muitas vezes, necessário fazer previsões.

É muito mais fácil tomar decisões sobre determinada variável, quando é possível estabelecer uma relação entre esta (a variável dependente) e uma outra variável (variável independente ou explicativa) cujo comportamento se conhece.

Regressão Linear Simples

Exemplo

Um diretor de pessoal de uma empresa descobre que existe uma lógica estreita entre a produtividade dos operários e os resultados de um teste de aptidões efetuado antes da sua entrada.

A partir do momento em que o diretor de pessoal estabelece o tipo de relação existente entre as duas variáveis - produtividade dos operários e resultados do teste de aptidões a que foram submetidos - poderá fazer previsões sobre a produtividade de um novo candidato com base no resultado do teste.

Regressão Linear Simples

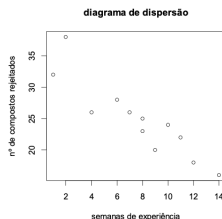
Modelo matemático

$$Y_i = E(Y|x_i) + \varepsilon_i$$

$$Y_i = a + bx_i + \varepsilon_i$$

- $E(Y|x_i) = a + bx_i$ - componente determinístico.
- x_i (*constante*)- valor da observação i da variável independente.
- Y_i -variável aleatória dependente.
- a -ordenada na origem e b -declive da reta são os parâmetros de regressão a determinar.
- $\varepsilon_i \sim N(0, \sigma^2)$ - erro da observação i
- Os erros ε_i assumem-se estatisticamente independentes.

Os dados onde vamos aplicar a RLS são tabelas de pares de valores. Cada observação é um par de valores, um para cada variável. Construímos o **diagrama de dispersão** das observações.



Semanas de experiência	Compostos rejeitados
9	20
6	28
14	16
8	23
12	18
10	24
4	26
2	38
11	22
1	32
8	25

Há uma forte indicação de que o conjunto de pontos se dispersa aleatoriamente sobre uma reta

Estimação dos parâmetros

Para estimar a e b habitualmente utilizamos o Método dos Mínimos Quadrados (MMQ).

- Estimativas que minimizam a soma dos quadrados dos erros ou desvios verticais entre \hat{y}_i e y_i . Estes erros são os resíduos $\hat{e}_i = y_i - \hat{y}_i$. Pretende-se minimizar a soma dos quadrados dos resíduos SE .

$$SE = \sum_{i=1}^n \hat{e}_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n \left[y_i - (\hat{a} + \hat{b}x_i) \right]^2$$

Estimação dos parâmetros

Mostra-se que os minimizantes de SE são obtidos igualando a 0 as derivadas parciais de SE em ordem a \hat{a} e a \hat{b} :

$$\begin{cases} \frac{\partial SE}{\partial \hat{a}} = -2 \sum_{i=1}^n (y_i - \hat{a} - \hat{b}x_i) = 0 \\ \frac{\partial SE}{\partial \hat{b}} = -2 \sum_{i=1}^n x_i (y_i - \hat{a} - \hat{b}x_i) = 0 \end{cases}$$

Obtemos as **equações normais**

$$\begin{cases} n\hat{a} + \hat{b} \sum_{i=1}^n x_i = \sum_{i=1}^n y_i \\ \hat{a} \sum_{i=1}^n x_i + \hat{b} \sum_{i=1}^n x_i^2 = \sum_{i=1}^n x_i y_i \end{cases}$$

Estimação dos parâmetros

Resolvendo o sistema de equações normais em ordem a \hat{a} e a \hat{b} , vem

$$\begin{cases} \hat{b} = \frac{n \sum_{i=1}^n x_i y_i - \left(\sum_{i=1}^n x_i \right) \left(\sum_{i=1}^n y_i \right)}{n \sum_{i=1}^n x_i^2 - \left(\sum_{i=1}^n x_i \right)^2} \\ \hat{a} = \frac{1}{n} \sum_{i=1}^n y_i - \hat{b} \frac{1}{n} \sum_{i=1}^n x_i \end{cases}$$

Estimação dos parâmetros

\hat{a} e \hat{b} , podem ser apresentados em função de somas de quadrados e somas cruzadas.

$$S_{xx} = \sum_{i=1}^n (x_i - \bar{x})^2 = \sum_{i=1}^n x_i^2 - n\bar{x}^2$$

$$S_{yy} = \sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n y_i^2 - n\bar{y}^2$$

$$S_{xy} = \sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^n x_i y_i - n\bar{x}\bar{y}$$

Estimação dos parâmetros

As somas anteriormente definidas, permitem reescrever:

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

$$\hat{b} = \frac{S_{xy}}{S_{xx}}$$

$$SE = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = S_{yy} - \frac{S_{xy}^2}{S_{xx}} = S_{yy} - \hat{b}S_{xy} = S_{yy} - \hat{b}^2 S_{xx}$$

Coefficientes de Determinação e Correlação

Coeficiente de Determinação

O valor r^2 da estatística

$$R^2 = \frac{S_{xy}^2}{S_{xx}S_{yy}}$$

é o coeficiente de determinação e $0 \leq r^2 \leq 1$

- $r^2 = 1$ - todas as observações encontram-se sobre a reta de regressão (ajuste perfeito).
- $r^2 = 0$ - reta de regressão horizontal (ajuste inútil)
- $r^2 \cong 1$ - o grau de linearidade entre x e Y aumenta.

Coefficientes de Determinação e Correlação

Coeficiente de Correlação amostral

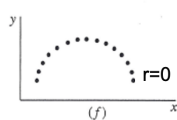
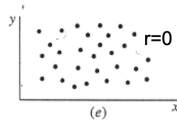
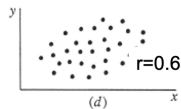
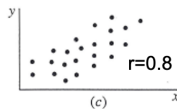
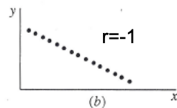
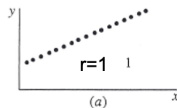
O valor r da estatística

$$R = \frac{S_{xy}}{\sqrt{S_{xx}} \sqrt{S_{yy}}}$$

é o coeficiente de correlação amostral e $-1 \leq r \leq 1$

Obs: O sinal de r é o sinal de \hat{b}

Coefficientes de Determinação e Correlação



Exemplo

25 pares de observações onde
Y = quantidade de vapor usado por mês
X = temperatura em graus Fahrenheit

Nº obs.	Y	X	Nº obs.	Y	X
1	10.98	35.3	13	11.88	28.1
2	11.13	29.7	14	9.57	39.1
3	12.51	30.8	15	10.94	46.8
4	8.40	58.8	16	9.58	48.5
5	9.27	61.4	17	10.09	59.3
6	8.73	71.3	18	8.11	70.0
7	6.36	74.4	19	6.83	70.0
8	8.50	76.7	20	8.88	74.5
9	7.82	70.7	21	7.68	72.1
10	9.14	57.5	22	8.47	58.1
11	8.24	46.4	23	8.86	44.6
12	12.19	28.9	24	10.36	33.4
			25	11.08	28.6

Exemplo

- $\bar{x} = 52.60 \quad \bar{y} = 9.424$

- $S_{xy} = \sum_{i=1}^{25} x_i y_i - 25 \times \bar{x} \bar{y} = -571.128$

- $S_{xx} = \sum_{i=1}^{25} x_i^2 - 25 \times \bar{x}^2 = 7154.42$

- $S_{yy} = \sum_{i=1}^{25} y_i^2 - 25 \times \bar{y}^2 = 63.8158$

- $r^2 = 0.7144$

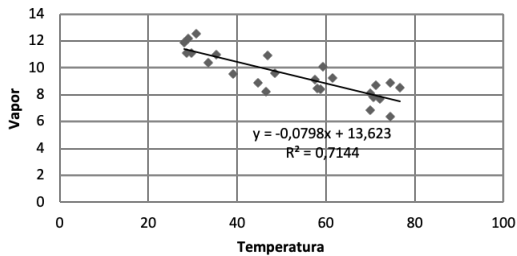
- $\hat{a} = 13.623 \quad \hat{b} = -0.0798$

Equação da reta de regressão dos mínimos quadrados

$$\hat{y} = \hat{a} + \hat{b}x = 13.6230 - 0.0798x$$

Exemplo

Gráfico de dispersão



Distribuição Amostral de \hat{Y}_0

Segundo o modelo de regressão linear

$$Y_i \sim N(a + bx_i, \sigma^2), i = 1, 2, \dots, n$$

A distribuição amostral de \hat{Y}_0 : $\hat{Y}_0 \sim N(\mu_{\hat{Y}_0}, \sigma_{\hat{Y}_0}^2)$ onde

$$\mu_{\hat{Y}_0} = E(Y_0) = a + bx_0 \text{ e } \sigma_{\hat{Y}_0}^2 = \sigma^2(Y_0) = \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

Como habitualmente $\sigma_{\hat{Y}_0}^2$ é desconhecida, substituímos na expressão anterior σ^2 por S^2 e obtemos

$$S_{\hat{Y}_0}^2 = S^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}} \right]$$

Intervalos de confiança

É possível mostrar que $T = \frac{\hat{Y}_0 - E(Y_0)}{S_{\hat{Y}_0}} \sim t_{n-2}$.

Intervalo de confiança de $E(Y_0)$

O intervalo de confiança a $(1 - \alpha) \times 100\%$ para a **resposta média** $E(Y_0)$ dado um valor x_0 é dado por

$$\left[\hat{y}_0 - t_c s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{y}_0 + t_c s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

onde $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ e $t_c = t_{1-\alpha/2}$

Intervalos de confiança

No caso de se pretender estimar o verdadeiro valor de Y para um dado valor da variável explicativa x_0 , recorre-se a um resultado semelhante ao anterior.

$$T = \frac{\hat{Y}_0 - Y_0}{S \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}} \sim t_{n-2}.$$

Intervalo de previsão(confiança) de Y_0

O **intervalo de previsão** a $(1 - \alpha) \times 100\%$ da **resposta** Y_0 dado um valor x_0 é dado por

$$\left[\hat{y}_0 - t_c s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{y}_0 + t_c s \sqrt{1 + \frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

onde $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ e $t_c = t_{1-\alpha/2}$

Intervalos de confiança

O mesmo se pode aplicar aos parâmetros \hat{a} e \hat{b} .

$$T_a = \frac{\hat{a} - a}{S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{n-2}$$

Intervalo de confiança do parâmetro a

O intervalo de confiança a $(1 - \alpha) \times 100\%$ para o parâmetro a é dado por

$$\left[\hat{a} - t_c s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}, \hat{a} + t_c s \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}} \right]$$

onde $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ e $t_c = t_{1-\alpha/2}$

Intervalos de confiança

$$T_b = \frac{\hat{b} - b}{S/\sqrt{S_{xx}}} \sim t_{n-2}$$

Intervalo de confiança do parâmetro b

O **intervalo de confiança** a $(1 - \alpha) \times 100\%$ para o parâmetro b é dado por

$$\left] \hat{b} - t_c s \sqrt{\frac{1}{S_{xx}}}, \hat{b} + t_c s \sqrt{\frac{1}{S_{xx}}} \right[$$

onde $s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ e $t_c = t_{1-\alpha/2}$

Testes de Hipóteses

Testes estatísticos sobre os parâmetros e a capacidade explicativa do modelo. Por exemplo para o parâmetro a :

$$H_0 : a = a_0 \text{ v.s. } H_1 : a \neq a_0$$

A estatística de teste a aplicar

$$T_a = \frac{\hat{a} - a_0}{S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{n-2}$$

Decisão:

Com um nível de significância α , rejeitar H_0 quando $|t| > t_{1-\alpha/2}(n-2)$

Testes de Hipóteses

Para o parâmetro b :

$$H_0 : b = b_0 \text{ v.s. } H_1 : b \neq b_0$$

A estatística de teste a aplicar

$$T_b = \frac{\hat{b} - b_0}{S/\sqrt{S_{xx}}} \sim t_{n-2}$$

Decisão:

Com um nível de significância α , rejeitar H_0 quando $|t| > t_{1-\alpha/2}(n-2)$

Tabela t-student

Tabela da Distribuição t-Student										
Valores Críticos t (p, n)				p =	P(T > t)					
Percentis				1-p =	P(T ≤ t)					
p	0,005	0,01	0,025	0,05	0,1	0,2	0,25	0,3	0,4	0,45
1-p	0,995	0,99	0,975	0,95	0,9	0,8	0,75	0,7	0,6	0,55
1	63,657	31,821	12,706	6,314	3,078	1,376	1,000	0,727	0,325	0,158
2	9,925	6,965	4,303	2,920	1,886	1,061	0,816	0,617	0,289	0,142
3	5,841	4,541	3,182	2,353	1,638	0,978	0,765	0,584	0,277	0,137
4	4,604	3,747	2,776	2,132	1,533	0,941	0,741	0,569	0,271	0,134
5	4,032	3,365	2,571	2,015	1,476	0,920	0,727	0,559	0,267	0,132
6	3,707	3,143	2,447	1,943	1,440	0,906	0,718	0,553	0,265	0,131
7	3,499	2,998	2,365	1,895	1,415	0,896	0,711	0,549	0,263	0,130
8	3,355	2,896	2,306	1,860	1,397	0,889	0,706	0,546	0,262	0,130
9	3,250	2,821	2,262	1,833	1,383	0,883	0,703	0,543	0,261	0,129
10	3,169	2,764	2,228	1,812	1,372	0,879	0,700	0,542	0,260	0,129
11	3,106	2,718	2,201	1,796	1,363	0,876	0,697	0,540	0,260	0,129
12	3,055	2,681	2,179	1,782	1,356	0,873	0,695	0,539	0,259	0,128
13	3,012	2,650	2,160	1,771	1,350	0,870	0,694	0,538	0,259	0,128
14	2,977	2,624	2,145	1,761	1,345	0,868	0,692	0,537	0,258	0,128
15	2,947	2,602	2,131	1,753	1,341	0,866	0,691	0,536	0,258	0,128
16	2,921	2,583	2,120	1,746	1,337	0,865	0,690	0,535	0,258	0,128
17	2,898	2,567	2,110	1,740	1,333	0,863	0,689	0,534	0,257	0,128
18	2,878	2,552	2,101	1,734	1,330	0,862	0,688	0,534	0,257	0,127
19	2,861	2,539	2,093	1,729	1,328	0,861	0,688	0,533	0,257	0,127
20	2,845	2,528	2,086	1,725	1,325	0,860	0,687	0,533	0,257	0,127
21	2,831	2,518	2,080	1,721	1,323	0,859	0,686	0,532	0,257	0,127
22	2,819	2,508	2,074	1,717	1,321	0,858	0,686	0,532	0,256	0,127
23	2,807	2,500	2,069	1,714	1,319	0,858	0,685	0,532	0,256	0,127
24	2,797	2,492	2,064	1,711	1,318	0,857	0,685	0,531	0,256	0,127
25	2,787	2,485	2,060	1,708	1,316	0,856	0,684	0,531	0,256	0,127

Exemplo I.C.

Intervalo de confiança a 95% para a quantidade média de vapor quando a temperatura é de 35 graus

Equação da reta de regressão dos mínimos quadrados

$$\hat{y} = \hat{a} + \hat{b}x = 13.6230 - 0.0798x$$

$$\hat{y}(35) = 13.6230 - 0.0798 \times 35 \approx 10.83$$

O intervalo de confiança a $(1 - \alpha) \times 100\%$ para a **resposta média** $E(Y_0)$, dado um valor x_0 é dado por

$$\left[\hat{y}_0 - t_c s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}, \hat{y}_0 + t_c s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} \right]$$

$$\text{onde } s^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \text{ e } t_c = t_{1-\alpha/2}$$

Exemplo I.C.

- $\hat{y} = \hat{a} + \hat{b}x = 13.6230 - 0.0798x$
 $\hat{y}(35) = 13.6230 - 0.0798 \times 35 \approx 10.83$
- $I.C. (E(Y_0)) =]10.83 - \Delta, 10.83 + \Delta[$
- $\Delta = t_c s \sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}}$
- $t_{0.975} = 2.07$
- $s = 0.89$
- $\sqrt{\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{xx}}} = 0.29$

$$I.C. (E(Y_0)) =]10.30, 11.36[$$

Exemplo T.H.

Verificar se é possível admitir que $a \neq 0$ usando um nível de significância de 5%.

$$H_0 : a = 0 \text{ v.s. } H_1 : a \neq 0$$

A estatística de teste a aplicar

$$T_a = \frac{\hat{a} - a_0}{S \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{S_{xx}}}} \sim t_{n-2}$$

$$t_a = \frac{13.6230}{0.89 \times \sqrt{\frac{1}{25} + \frac{52.6^2}{7154.42}}} = 23.43$$

Decisão:

Como $t_a > 2.07$, rejeita-se H_0 . O teste é conclusivo, o que quer dizer que a um nível de significância de 5% há evidência estatística suficiente para se afirmar que a ordenada na origem é diferente de zero.

Análise de variância: Anova

$$\hat{Y} = \hat{a} + \hat{b}X$$

Método de análise da qualidade do modelo de regressão baseado na divisão da variação total de Y (variável dependente).

Desvio total

$$Y_i - \bar{Y} = (\hat{Y}_i - \bar{Y}) + (Y_i - \hat{Y}_i)$$

Desvio total = Desvio explicado + Desvio não explicado

Análise de variância: Anova

Variação total

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

Variação total = Variação explicada + Variação não explicada

Variação total

Variabilidade total do conjunto de observações medida pela soma dos quadrados dos desvios totais

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

Análise de variância: Anova

Variação explicada

A soma dos quadrados dos desvios explicados pela regressão

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

Variação não explicada

A soma dos quadrados dos resíduos (erros de ajuste)

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

Análise de variância: Anova

$$ST = SR + SE$$

SE - parte da variabilidade das observações de Y que se mantem mesmo conhecendo o valor de x .

SR - parte da variabilidade das observações de Y que é eliminada quando se usa o conhecimento da variável independente para prever Y .

ST - variabilidade total das observações de Y .

Análise de variância: Anova

Graus de liberdade

$$ST = SR + SE$$

ST - $n - 1$ graus de liberdade resultantes de n desvios $Y_i - \bar{Y}$
menos uma restrição $\sum_{i=1}^n (Y_i - \bar{Y}) = 0$

SR - 1 grau de liberdade resultante de 2 estimadores de a e b
menos uma restrição $\hat{a} = \bar{Y} - \hat{b}\bar{x}$

SE - $n - 2$ graus de liberdade resultantes de n resíduos $Y_i - \hat{Y}_i$
menos duas restrições para a e b .

$$g.l.(ST) = g.l.(SR) + g.l.(SE)$$

Análise de variância: Anova

Os cálculos ficam sintetizados numa tabela:

Fonte de variação	Soma de quadrados	Graus de liberdade	Média quadrática	Estatística de teste f
Regressão	SR	1	$MSR = \frac{SR}{1}$	$\frac{MSR}{MSE}$
Erro	SE	$n - 2$	$MSE = \frac{SE}{n-2}$	
Total	ST	$n - 1$		

F- distribuição F de Snedecor $F_{\alpha,1,n-2}$

Análise de variância: Anova

Teste da significância da regressão

$$H_0 : b = b_0 \text{ v.s. } H_1 : b \neq b_0$$

Estatística de teste: $F = \frac{MSR}{MSE}$

Decisão: Rejeita-se H_0 se $f_{obs} > f_{\alpha,1,n-2}$.

Neste caso, podemos concluir que é aceitável admitir que a função de regressão é linear.

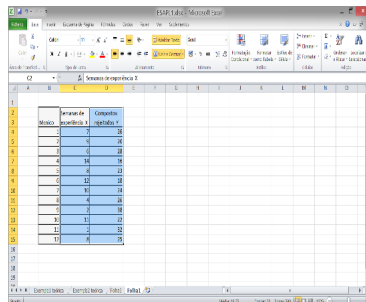
F- distribuição F de Snedecor $F_{\alpha,1,n-2}$

Regressão linear simples no Excel

Passo 1. Introduza os dados numa folha excel

	Semanas de experiência X	Compostos rejeitados Y
1	7	26
2	9	20
3	6	28
4	14	16
5	8	23
6	12	18
7	10	24
8	4	26
9	2	38
10	11	22
11	1	32
12	8	25

Passo 2. Selecione as colunas X e Y



Regressão linear simples no Excel

Passo 3. Selecione Inserir --- gráfico ----Dispersão

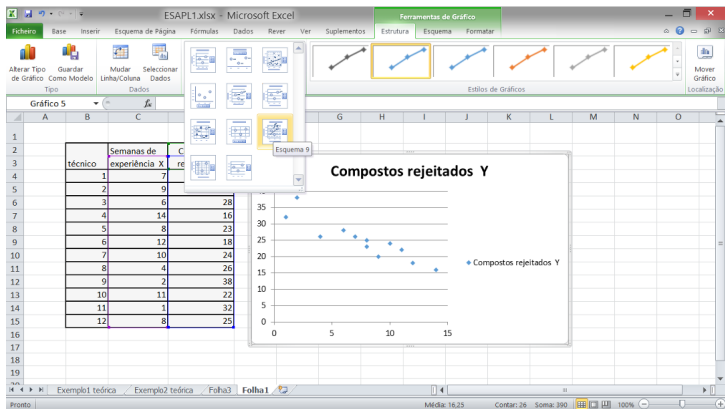
The screenshot shows the Microsoft Excel interface with the 'Inserir' (Insert) tab selected. The 'Dispersão' (Scatter) chart type is highlighted in the 'Gráficos' (Charts) group. The spreadsheet data is as follows:

	A	B	C	D	E
1					
2					
3					
4					
5					
6					
7					
8					
9					
10					
11					
12					
13					
14					
15					
16					
17					
18					
19					

The data table is located in the range C3:D15. The columns are labeled 'Semanas de experiência X' (Column C) and 'Compostos rejeitados Y' (Column D). The rows contain numerical data points.

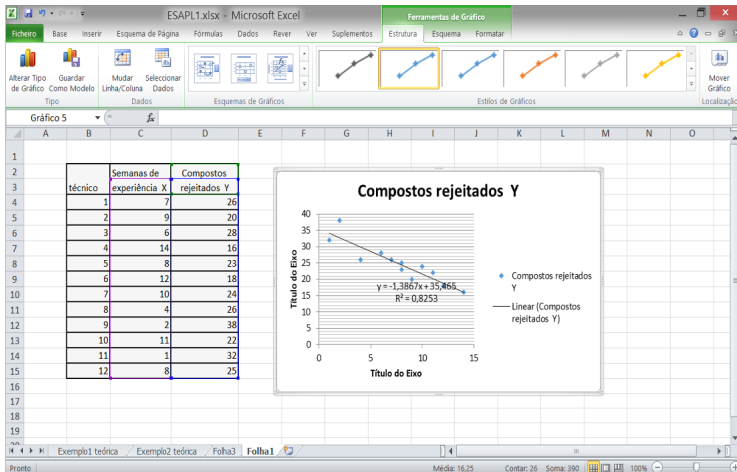
Regressão linear simples no Excel

Passo 4. Dispersão --- esquema de gráficos ----esquema 9



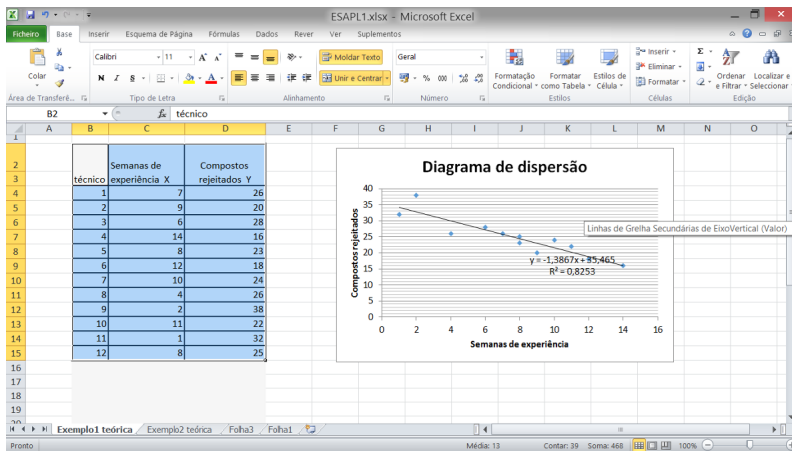
Regressão linear simples no Excel

Passo 5. Selecione os títulos dos eixos



Regressão linear simples no Excel

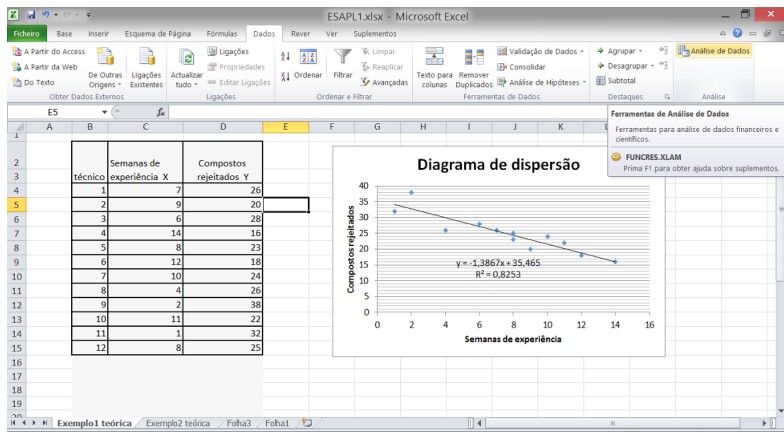
Passo 6. Apresente o gráfico de uma forma clara
Reta de regressão e coeficiente de determinação



Regressão linear simples no Excel

Uso da ferramenta **Análise de dados**

Obs. Se esta ferramenta não estiver disponível, carregue o suplemento **Analysis ToolPack**.
Selecione **Ficheiro--Opções--Suplementos-- Analysis Toolpak**.



Regressão linear simples no Excel

Uso da ferramenta **Análise de dados**

Análise de dados----Regressão

The screenshot shows the Microsoft Excel interface with the 'Dados' (Data) tab selected. A data table is visible in the worksheet, and the 'Analisar dados' (Data Analysis) task pane is open on the right side.

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1															
2															
3			técnico	Semanas de experiência X	Compostos rejeitados Y										
4			1	7	26										
5			2	9	20										
6			3	6	28										
7			4	14	16										
8			5	8	23										
9			6	12	18										
10			7	10	24										
11			8	4	26										
12			9	2	38										
13			10	11	22										
14			11	1	32										
15			12	8	25										
16															
17															
18															
19															

The 'Analisar dados' (Data Analysis) task pane is open, showing the following options:

- Teste F: duas amostras para variâncias
- Análise de Fourier
- Histograma
- Média móvel
- Geração de número aleatório
- Ordem e percentil
- Regressão** (highlighted)
- Amostragem
- Teste T: duas amostras emparelhadas para médias
- Teste T: duas amostras com variâncias iguais

Regressão linear simples no Excel

Uso da ferramenta **Análise de dados**

Análise de dados—Regressão

Preencher caixa de diálogo de regressão

- Intervalo Y: selecione as células da coluna Y-compostos rejeitados
- Intervalo X: selecione as células da coluna X- semanas de experiência
- Residuais: Desenho de valores residuais
- Intervalo de saída : selecione uma célula qualquer da folha de cálculo, fora da tabela de valores, ou nova folha de cálculo

ESAPL1.xlsx - Microsoft Excel

Regressão

Entrada

Intervalo Y:

Intervalo X:

☐ Bêta

☐ Constante é zero

☒ Nível de confiança 95 %

Opções de saída

☒ Intervalo de saída:

☐ Nova folha de cálculo:

☐ Novo livro

Residuais

☒ Resíduos

☐ Resíduos padronizados

☒ Desenho de valores residuais

☐ Desenho de linha ajustada

Probabilidade normal

☐ Desenho de probabilidade normal

técnico	Semanas de experiência X	Compostos rejeitados Y
1	7	26
2	9	20
3	6	28
4	14	16
5	8	23
6	12	18
7	10	24
8	4	26
9	2	38
10	11	22
11	1	32
12	8	25

Regressão linear simples no Excel

Uso da ferramenta **Análise de dados**

Sumário dos resultados

$$r^2 = 0,8253$$

Tabela Anova

$$\hat{a} = 35,465$$

$$\hat{b} = -1,387$$

$$I.C._{95\%}(a) = [31,6238; 39,3059]$$

$$I.C._{95\%}(b) = [-1,8363; -0,9372]$$

Obs: gl – graus de liberdade

SQ – soma de quadrados SR, SE e ST

MQ – média quadrática

