Algoritmos Numéricos 2^a edição

Capítulo 4: Ajuste de curvas

Capítulo 4: Ajuste de curvas

Relações entre variáveis

- Relações entre variáveis envolvidas em um experimento: determinísticas, semideterminísticas e empíricas.
- Determinísticas: variáveis relacionadas entre si por algum tipo de lei expressa por uma fórmula matemática precisa.
- Variação nas observações é atribuída a erros experimentais.
- Semideterminísticas: alguma teoria prescreve uma forma para a relação entre as variáveis.
- Realizar experimentos para obter informações acerca de parâmetros do modelo.
- Empíricas: relações entre variáveis envolvidas não são conhecidas.

Relações entre variáveis cont.

- Determinar uma fórmula matemática que relacione as variáveis.
- Diagrama de dispersão: gráfico com valores observados das variáveis fornece idéia da relação entre elas com algumas variações aleatórias.
- Somente com suficiente conhecimento sobre uma relação empírica é possível desenvolver uma teoria que conduza a uma fórmula matemática (semideterminística).

Perturbação na verdadeira relação

- Precisão limitada dos instrumentos de medida.
- Perturbações incontroláveis das condições experimentais.
- Fatores que introduzem erros nos dados.
- Variação das leituras de uma variável: erros de medida experimentais e a outras variáveis, cujos valores se alteram durante um experimento.

Objetivo do ajuste de curvas

- Relacionar, por um modelo matemático, a variável resposta (ou dependente) com o conjunto de variáveis explicativas (ou independentes).
- Para ter controle.
- Determinar algum parâmetro.
- Fazer previsão acerca do comportamento da variável resposta.

Análise de regressão

- Conjunto de métodos:
- Estimativa de parâmetros.
- Análise de variância e de resíduos.
- Testes de hipóteses.
- Lidam com formulação de modelos matemáticos que descrevem relações entre variáveis.
- Modelos com propósito de predição e outras inferências estatísticas.
- Ajuste de curvas: determinação de parâmetros de modelos semideterminísticos.

Regressão linear simples

- Relações mais simples entre duas variáveis são as relações lineares.
- Variável independente ou explicativa x é relacionada com a variável dependente ou resposta y por meio de um modelo linear

$$y = b_0 + b_1 x.$$

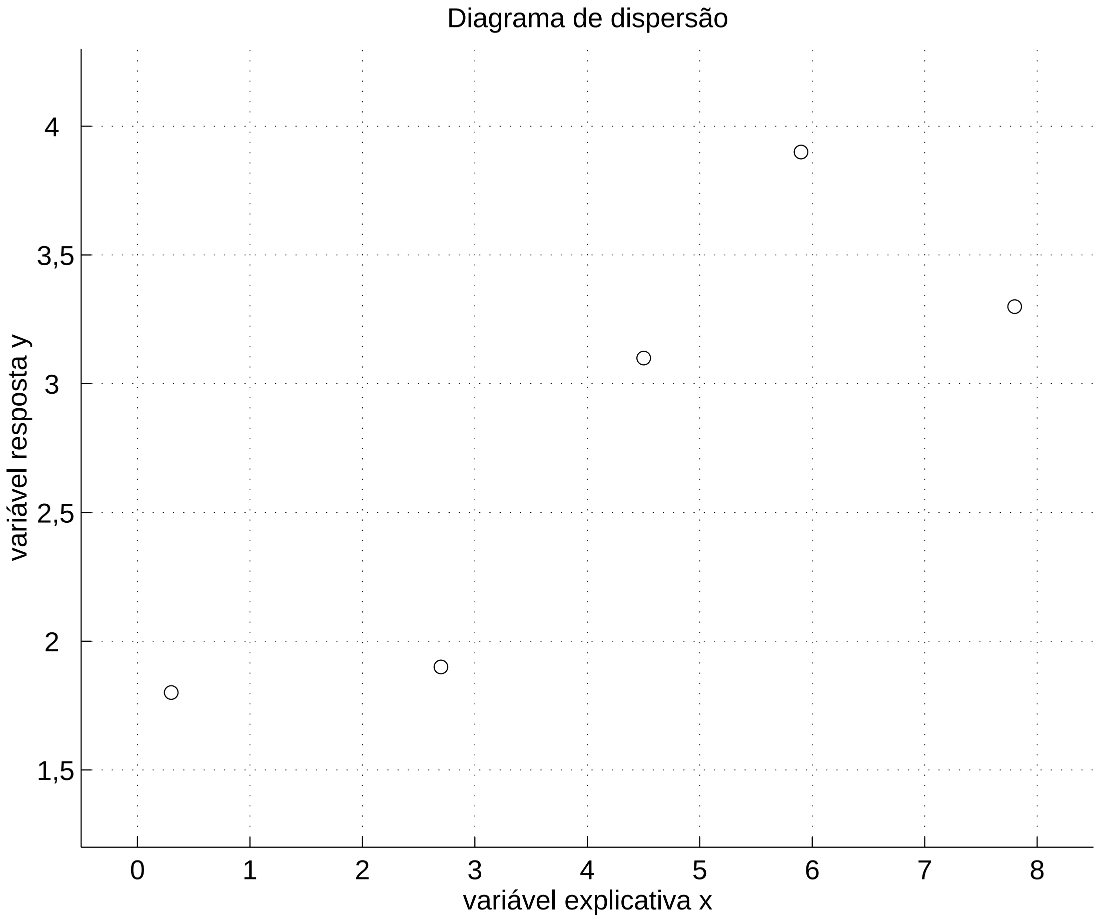
- Esboçar os dados em um gráfico de coordenadas cartesianas denominado diagrama de dispersão.
- Diagrama mostra a natureza da relação intrínseca entre as duas variáveis estudadas.

Diagrama de dispersão

- Variáveis explicativas x e as respostas y

x	0,3	2,7	4,5	5,9	7,8
y	1,8	1,9	3,1	3,9	3,3

$\Vdash\!\Leftarrow$



Retas de regressão

- Modelo simples que relaciona duas variáveis x e y

$$y = \beta_0 + \beta_1 x + \epsilon.$$

- β_0 e β_1: parâmetros a serem estimados.
- ϵ: componentes desconhecidos e aleatórios de erro que se sobrepõem à verdadeira relação linear.
- Como estimar os parâmetros β_0 e β_1?

Modelo 1

- Primeira tentativa: obtida por um polinômio interpolador linear.
- Não se pode traçar uma única reta que passe por todos os pontos simultaneamente.
- Reta esboçada a partir de dois pontos quaisquer, por exemplo, o primeiro e o último

x	0,3	7,8
y	1,8	3,3

.

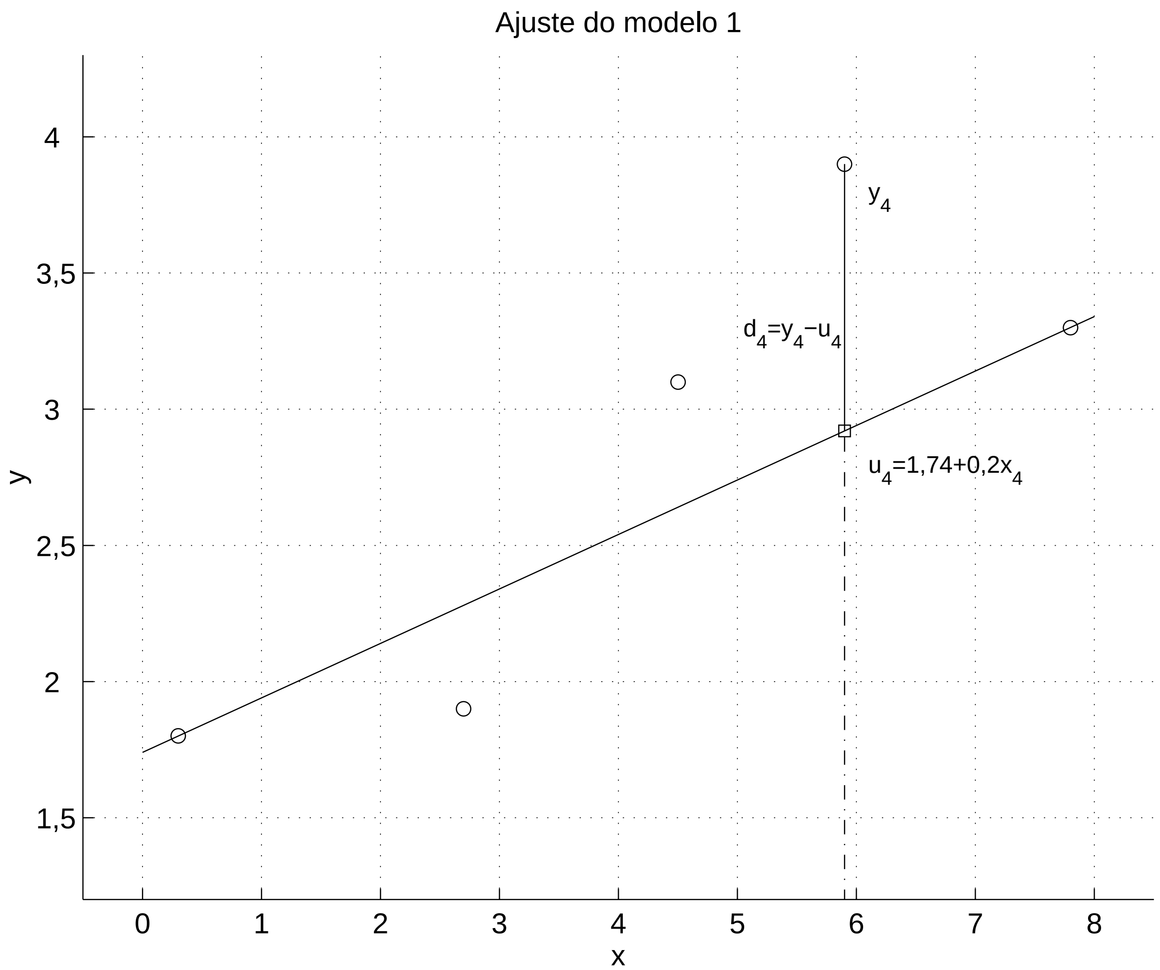
- Equação da reta $u(x)$ que passa por estes dois pontos

$$u(x)=y_0+\frac{y_1-y_0}{x_1-x_0}(x-x_0)=1{,}8+\frac{3{,}3-1{,}8}{7{,}8-0{,}3}(x-0{,}3)=1{,}8+0{,}2(x-0{,}3)\rightsquigarrow$$

$$u(x)=1{,}74+0{,}2x.$$

Gráfico do modelo 1

- Modelo 1: $u = 1{,}74 + 0{,}2x$.
- $d_i = y_i - u_i$: distância vertical entre o i-ésimo ponto dado y_i e o ponto $u_i = 1{,}74 + 0{,}2x_i$ de mesma abscissa x_i.



Qualidade do modelo 1

- Qualidade do ajuste: soma de todas as n distâncias verticais de y_i aos pontos da reta $u_i = 1{,}74 + 0{,}2x_i$, considerando valores positivos de d_i

$$D(b_0, b_1) = \sum_{i=1}^{n}(y_i - u_i)^2 = \sum_{i=1}^{n}(y_i - b_0 - b_1x_i)^2 = \sum_{i=1}^{n} d_i^2,$$

$$D(1{,}74; 0{,}2) = \sum_{i=1}^{5}(y_i - 1{,}74 - 0{,}2x_i)^2.$$

- Resultados do ajuste pelo modelo 1: $u = 1{,}74 + 0{,}2x$.

i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,80	0,00
2	2,7	1,9	2,28	−0,38
3	4,5	3,1	2,64	0,46
4	5,9	3,9	2,92	0,98
5	7,8	3,3	3,30	0,00
$D(1{,}74; 0{,}2) = 1{,}3164$				

Modelo 2

- Segunda tentativa: também obtida por polinômio interpoladaor linear.
- Traçar a reta por dois pontos quaisquer.
- Pontos escolhidos não pertencem ao diagrama de dispersão, por exemplo,

x	2	6
y	2	3

.

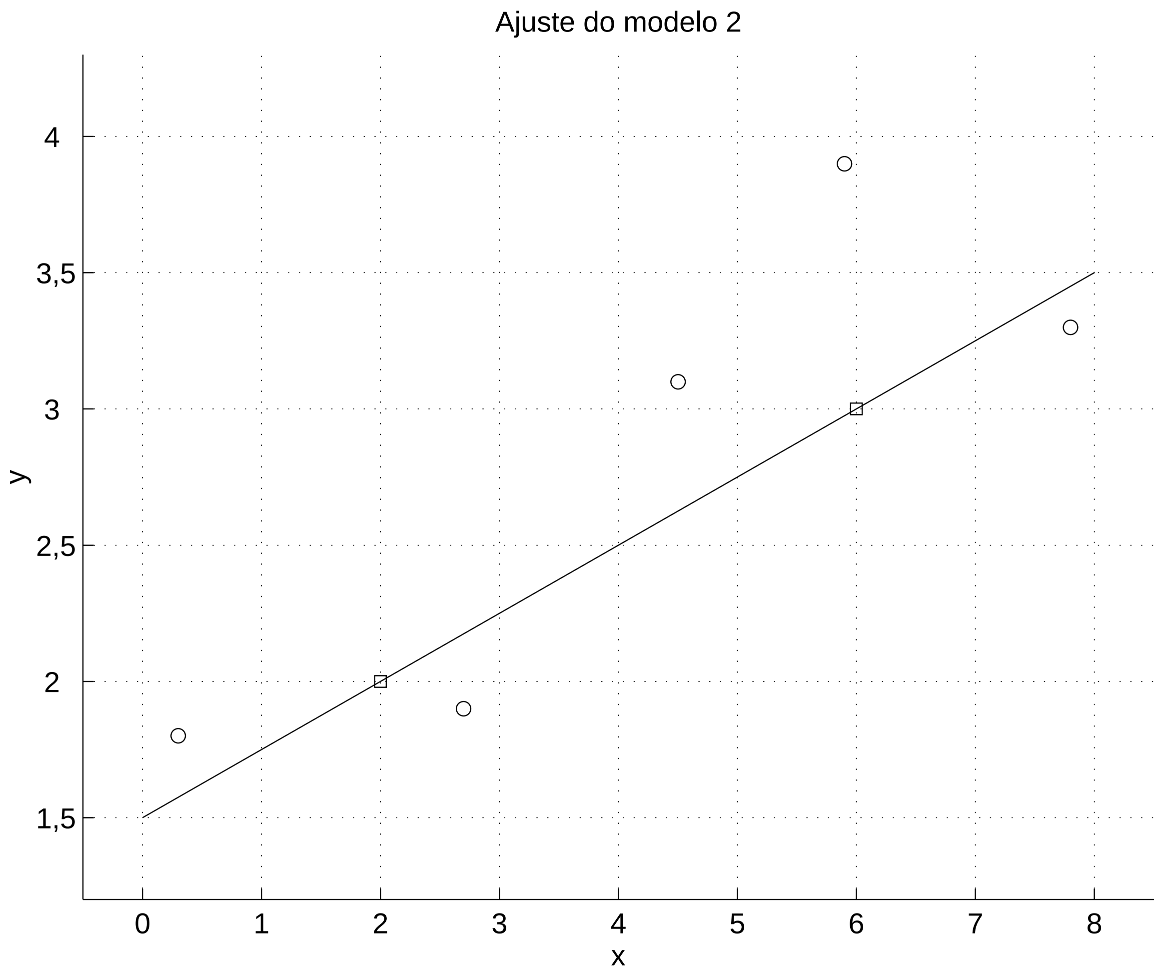
- Equação da reta $u(x)$ que passa por esses pontos

$$u(x) = y_0 + \frac{y_1 - y_0}{x_1 - x_0}(x - x_0) = 2 + \frac{3-2}{6-2}(x-2) = 2 + 0{,}25(x-2) \rightsquigarrow$$

$$u(x) = 1{,}5 + 0{,}25x.$$

Gráfico do modelo 2

- Modelo 2: $u = 1{,}5 + 0{,}25x$.



Qualidade do modelo 2

- Resultados do ajuste do modelo 2

i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,575	0,225
2	2,7	1,9	2,175	−0,275
3	4,5	3,1	2,625	0,475
4	5,9	3,9	2,975	0,925
5	7,8	3,3	3,450	−0,150
$D(1{,}5; 0{,}25) = 1{,}2300$				

.

- Modelo 2 é mais adequado

$$D(1{,}5; 0{,}25) = 1{,}2300 < D(1{,}74; 0{,}2) = 1{,}3164.$$

Método dos quadrados mínimos

- Qualidade do ajuste depende da equação da reta escolhida.
- Reta que não passa por dois pontos do diagrama de dispersão produziu resultado melhor.
- Por onde se deve traçar a reta de modo a obter o menor valor do desvio D?
- Método dos quadrados mínimos consiste em encontrar uma estimativa da reta

$$u = \beta_0 + \beta_1 x.$$

- Produzir o menor valor possível do desvio

$$D(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

Dedução dos quadrados mínimos

- Função desvio

$$D(\beta_0, \beta_1) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i)^2.$$

- Derivadas parciais

$$\frac{\partial D(\beta_0, \beta_1)}{\partial \beta_0} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) \ \text{ e}$$

$$\frac{\partial D(\beta_0, \beta_1)}{\partial \beta_1} = -2 \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_i) x_i.$$

Mínimo da função $D(\beta_0, \beta_1)$

- Valores para os quais a função $D(\beta_0, \beta_1)$ possui um mínimo: onde as derivadas parciais se anulam.
- Se $D(b_0, b_1)$ for o ponto de mínimo de $D(\beta_0, \beta_1)$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i) = 0 \longrightarrow \sum_{i=1}^{n} b_0 + \sum_{i=1}^{n} b_1 x_i = \sum_{i=1}^{n} y_i \ \text{ e}$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_i)x_i = 0 \longrightarrow \sum_{i=1}^{n} b_0 x_i + \sum_{i=1}^{n} b_1 x_i^2 = \sum_{i=1}^{n} x_i y_i.$$

Reta de quadrados mínimos

- Na forma matricial e simplificando a notação

$$\begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix}. \tag{1}$$

- Valores em que $D(\beta_0, \beta_1)$ apresenta um mínimo são obtidos pela solução do sistema linear (1), denominado equações normais.

- Utilizando as operações l-elementares

$$\begin{bmatrix} n & \sum x_i \\ 0 & -\frac{1}{n}(\sum x_i)^2 + \sum x_i^2 \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \end{bmatrix} = \begin{bmatrix} \sum y_i \\ -\frac{1}{n}\sum x_i \sum y_i + \sum x_i y_i \end{bmatrix}.$$

- Parâmetros da reta de quadrados mínimos $u(x) = b_0 + b_1 x$

$$\boxed{b_1 = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2}} \quad \text{e} \quad \boxed{b_0 = \frac{\sum y_i - b_1 \sum x_i}{n}}. \tag{2}$$

Exemplo: ajuste linear por quadrados mínimos

Exemplo 1 Calcular a reta de quadrados mínimos utilizando dados da tabela

x	0,3	2,7	4,5	5,9	7,8
y	1,8	1,9	3,1	3,9	3,3

.

- Valores dos somatórios

i	x_i	y_i	x_i^2	$x_i y_i$	y_i^2
1	0,3	1,8	0,09	0,54	3,24
2	2,7	1,9	7,29	5,13	3,61
3	4,5	3,1	20,25	13,95	9,61
4	5,9	3,9	34,81	23,01	15,21
5	7,8	3,3	60,84	25,74	10,89
$\sum$	21,2	14,0	123,28	68,37	42,56

. $\Vert\!\Leftarrow$

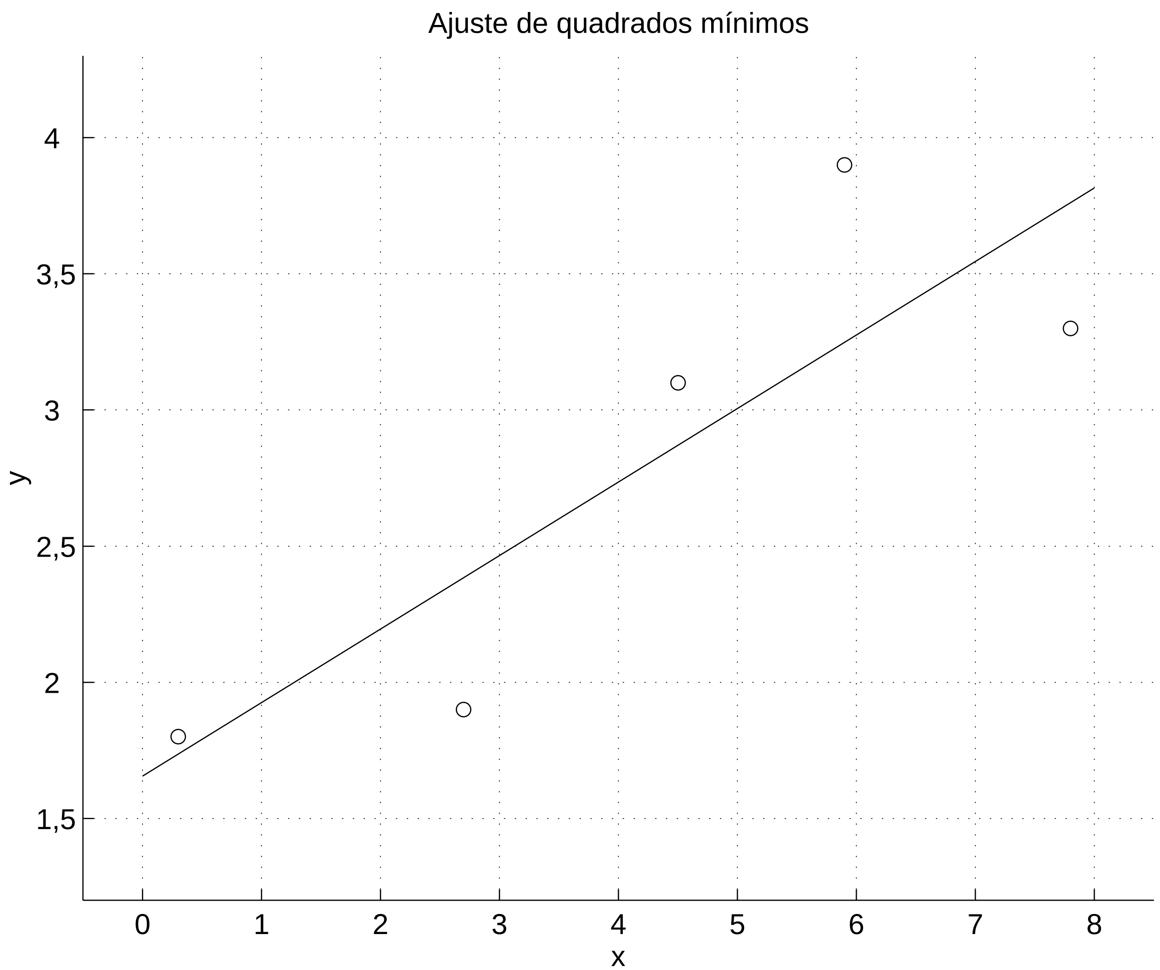
- Solução de quadrados mínimos

$$b_1 = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2} = \frac{21{,}2 \times 14{,}0 - 5 \times 68{,}37}{(21{,}2)^2 - 5 \times 123{,}28} \rightsquigarrow b_1 = 0{,}2698 \text{ e}$$

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n} = \frac{14{,}0 - 0{,}2698 \times 21{,}2}{5} \rightsquigarrow b_0 = 1{,}6560.$$

Ajuste por quadrados mínimos

- Reta de quadrados mínimos: $u = 1{,}6560 + 0{,}2698x$



Qualidade do modelo

- Resultados do ajuste por quadrados mínimos

i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,7369	0,0631
2	2,7	1,9	2,3845	−0,4845
3	4,5	3,1	2,8701	0,2299
4	5,9	3,9	3,2478	0,6522
5	7,8	3,3	3,7604	−0,4604
$D(1{,}6560; 0{,}2698) = 0{,}9289$				

.

- Melhor dos três modelos propostos

$$D(1{,}6560; 0{,}2698) = 0{,}9289 < D(1{,}5; 0{,}25) = 1{,}2300 < D(1{,}74; 0{,}2) = 1{,}3164.$$

Coeficiente de determinação

- Seja a expressão para o i-ésimo ponto

$$y_i - \bar{y} = (y_i - u_i) + (u_i - \bar{y}),$$

- sendo $u_i = b_0 + b_1 x_i$ e $\bar{y} = \dfrac{1}{n}\left(\displaystyle\sum_{i=1}^{n} y_i\right)$.

- Tomando o quadrado em ambos os termos da igualdade

$$(y_i - \bar{y})^2 = (y_i - u_i)^2 + (u_i - \bar{y})^2 + 2(y_i - u_i)(u_i - \bar{y}).$$

- Calculando o somatório para $i = 1, 2, \ldots, n$

$$\sum_{i=1}^{n}(y_i - \bar{y})^2 = \sum_{i=1}^{n}(y_i - u_i)^2 + \sum_{i=1}^{n}(u_i - \bar{y})^2 + 2\sum_{i=1}^{n}(y_i - u_i)(u_i - \bar{y}). \quad (3)$$

- Pode-se mostrar que $\displaystyle\sum_{i=1}^{n}(y_i - u_i)(u_i - \bar{y}) = 0$.

Soma de quadrados

- Somatórios

$$\sum_{i=1}^{n}(y_i-\bar{y})^2=\sum_{i=1}^{n}(y_i-u_i)^2+\sum_{i=1}^{n}(u_i-\bar{y})^2.$$

- SQTot (soma de quadrados total)

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

- SQRes (soma de quadrados residual)

$$\sum_{i=1}^{n}(y_i-u_i)^2.$$

- SQReg (soma de quadrados devido à regressão)

$$\sum_{i=1}^{n}(u_i-\bar{y})^2.$$

Cálculo de r^2

- Qualidade do ajuste do modelo $u = b_0 + b_1x$ aos dados

$$r^2 = \frac{\text{SQReg}}{\text{SQTot}} = \frac{\text{SQTot} - \text{SQRes}}{\text{SQTot}} \rightsquigarrow r^2 = 1 - \frac{\text{SQRes}}{\text{SQTot}}.$$

- r^2: coeficiente de determinação

$$0 \leq r^2 \leq 1.$$

- Quanto mais próximo r^2 for da unidade, melhor será o ajuste.

Cálculo de r^2 cont.

- Considerando que

$$\boxed{D(b_0, b_1) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} d_i^2}, \tag{4}$$

$$\sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n} y_i^2 - 2\bar{y}\sum_{i=1}^{n} y_i + n\bar{y}^2 \leadsto \sum_{i=1}^{n}(y_i-\bar{y})^2 = \sum_{i=1}^{n} y_i^2 - \frac{1}{n}\left(\sum_{i=1}^{n} y_i\right)^2.$$

- Coeficiente de determinação

$$\boxed{r^2 = 1 - \frac{D(b_0, b_1)}{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2}}. \tag{5}$$

- Proporção da variação total dos dados em torno da média $\bar{y}$ que é explicada pelo modelo de regressão.

Variância residual

- Variância residual σ^2

$$\boxed{\sigma^2 = \frac{D(b_0, b_1)}{n - p}}. \tag{6}$$

- $D(b_0, b_1)$: somatório dos desvios (dado por (4).
- n: número de pontos e
- p: número de parâmetros estimados.
- No caso de regressão linear simples $u = b_0 + b_1 x$: $p = 2$.
- Tanto o numerador quanto o denominador de (6) irão diminuir se forem introduzidos mais parâmetros no modelo.
- Redução global de σ^2 que definirá se mais parâmetros devem ou não ser incorporados ao modelo.

Exemplo: qualidade do ajuste

Exemplo 2 Calcular o coeficiente de determinação e a variância residual da tabela do exemplo anterior

i	x_i	y_i	u_i	d_i
1	0,3	1,8	1,7369	0,0631
2	2,7	1,9	2,3845	−0,4845
3	4,5	3,1	2,8701	0,2299
4	5,9	3,9	3,2478	0,6522
5	7,8	3,3	3,7604	−0,4604
$D(1{,}6560; 0{,}2698) = 0{,}9289$				

- Coeficiente de determinação

$$r^2 = 1 - \frac{D(b_0, b_1)}{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2} = 1 - \frac{0{,}9289}{42{,}56 - (14{,}0)^2/5} \rightsquigarrow r^2 = 0{,}7235.$$

- Variância residual

$$\sigma^2 = \frac{D(b_0, b_1)}{n - p} = \frac{0{,}9289}{5 - 2} \rightsquigarrow \sigma^2 = 0{,}3096.$$

Exemplo: reta de quadrados mínimos

Exemplo 3 Calcular a reta de quadrados mínimos

x	1,2	2,5	3,0	4,1	6,2	7,1	8,8	9,5
y	6,8	6,1	9,9	9,7	12,1	17,9	18,0	21,5

.

- Dispositivo para regressão linear simples por quadrados mínimos

i	x_i	y_i	x_i^2	x_iy_i	y_i^2	u_i	d_i	d_i^2
1	1,2	6,8	1,44	8,16	46,24	5,4037	1,3963	1,9497
2	2,5	6,1	6,25	15,25	37,21	7,7330	−1,6330	2,6667
3	3,0	9,9	9,00	29,70	98,01	8,6289	1,2711	1,6157
4	4,1	9,7	16,81	39,77	94,09	10,5999	−0,8999	0,8098
5	6,2	12,1	38,44	75,02	146,41	14,3627	−2,2627	5,1198
6	7,1	17,9	50,41	127,09	320,41	15,9753	1,9247	3,7045
7	8,8	18,0	77,44	158,40	324,00	19,0213	−1,0213	1,0431
8	9,5	21,5	90,25	204,25	462,25	20,2756	1,2244	1,4992
$\sum$	42,4	102,0	290,04	657,64	1528,62	102,0003	−0,0004	18,4085

.

Exemplo: reta de quadrados mínimos cont.

- Cálculo dos parâmetros

$$b_1 = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2} = \frac{42{,}4 \times 102{,}0 - 8 \times 657{,}64}{(42{,}4)^2 - 8 \times 290{,}04} \rightsquigarrow b_1 = 1{,}7918$$

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n} = \frac{102{,}0 - 1{,}7918 \times 42{,}4}{8} \rightsquigarrow b_0 = 3{,}2535.$$

- Coeficiente de determinação

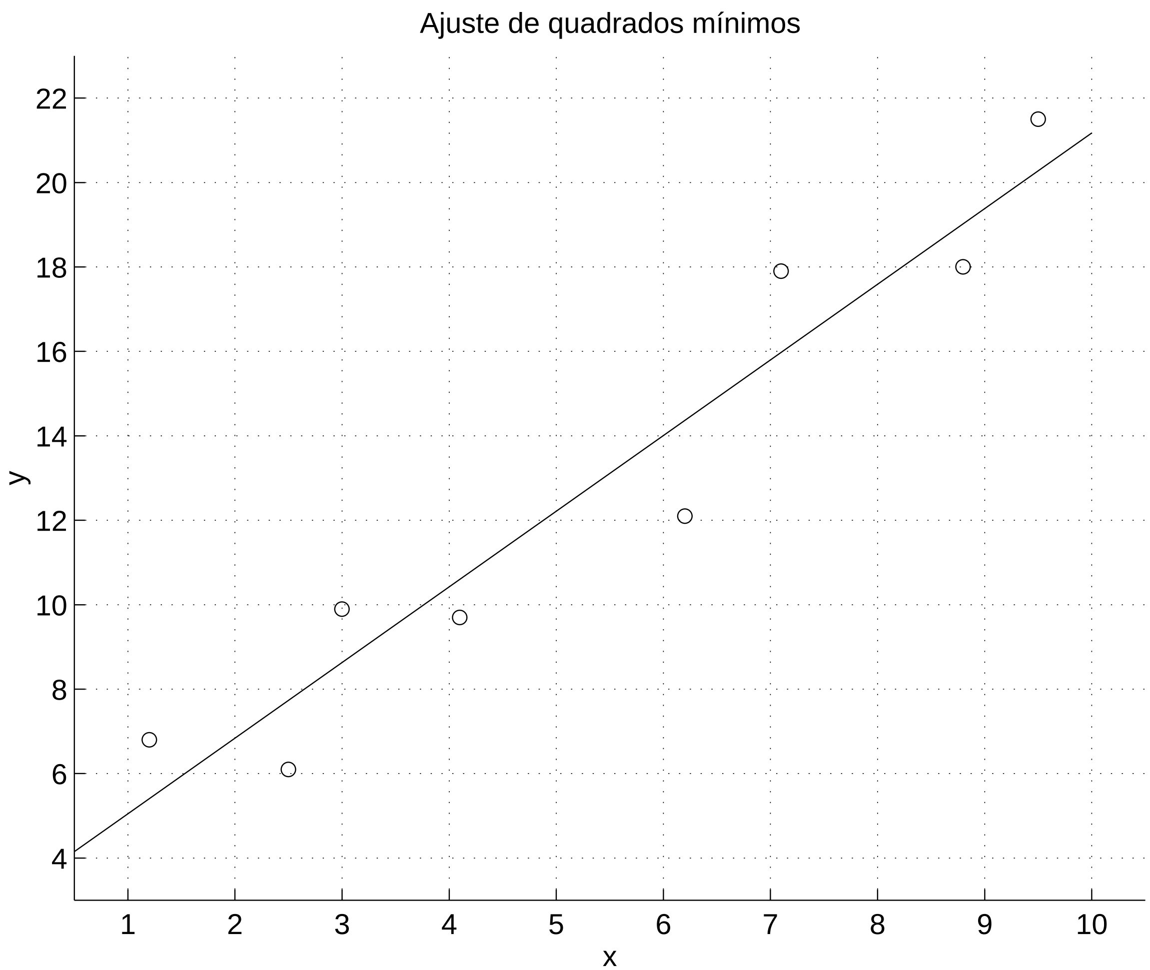
$$r^2 = 1 - \frac{D(b_0, b_1)}{\sum y_i^2 - \frac{1}{n}(\sum y_i)^2} = 1 - \frac{18{,}4085}{1528{,}62 - (102{,}0)^2/8} \rightsquigarrow r^2 = 0{,}9193.$$

- Variância residual

$$\sigma^2 = \frac{D(b_0, b_1)}{n - p} = \frac{18{,}4085}{8 - 2} \rightsquigarrow \sigma^2 = 3{,}0681.$$

Gráfico da reta de quadrados mínimos

- Equação de quadrados mínimos: $u = 3{,}2535 + 1{,}7918x$.



Regressão linear múltipla

- Modelo mais completo que relaciona a variável resposta y com as p variáveis explicativas x_i

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon, \tag{7}$$

- $\beta_i,\ i = 0, 1, \ldots, p$: parâmetros a serem estimados e
- ϵ: variável aleatória desconhecida que interfere na verdadeira relação linear.
- Método dos quadrados mínimos utilizado para estimar os $p + 1$ parâmetros β_i

$$D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p) = \sum_{i=1}^{n} (y_i - u_i)^2 = \sum_{i=1}^{n} (y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip})^2,$$

- sendo x_{ij} a i-ésima observação da j-ésima variável explicativa.

Método dos quadrados mínimos

- Derivadas parciais da função desvio $D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_0} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip}),$$

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_1} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip})x_{i1},$$

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_2} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip})x_{i2},$$

$$\vdots$$

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_p} = -2\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_{i1} - \beta_2 x_{i2} - \ldots - \beta_p x_{ip})x_{ip}.$$

Mínimo da função desvio D

- Se $D(b_0, b_1, b_2, \ldots, b_p)$ for o ponto de mínimo da função $D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$

$$\frac{\partial D(b_0, b_1, b_2, \ldots, b_p)}{\partial \beta_i} = 0, \; i = 0, 1, \ldots, p$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \ldots - b_p x_{ip}) = 0 \leadsto \sum_{i=1}^{n} b_0 + \sum_{i=1}^{n} b_1 x_{i1} + \sum_{i=1}^{n} b_2 x_{i2} + \ldots + \sum_{i=1}^{n} b_p x_{ip} = \sum_{i=1}^{n} y_i,$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \ldots - b_p x_{ip})x_{i1} = 0 \leadsto \sum_{i=1}^{n} b_0 x_{i1} + \sum_{i=1}^{n} b_1 x_{i1} x_{i1} + \sum_{i=1}^{n} b_2 x_{i2} x_{i1} + \ldots + \sum_{i=1}^{n} b_p x_{ip} x_{i1} = \sum_{i=1}^{n} x_{i1} y_i,$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \ldots - b_p x_{ip})x_{i2} = 0 \leadsto \sum_{i=1}^{n} b_0 x_{i2} + \sum_{i=1}^{n} b_1 x_{i1} x_{i2} + \sum_{i=1}^{n} b_2 x_{i2} x_{i2} + \ldots + \sum_{i=1}^{n} b_p x_{ip} x_{i2} = \sum_{i=1}^{n} x_{i2} y_i,$$

$$\vdots$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1 x_{i1} - b_2 x_{i2} - \ldots - b_p x_{ip})x_{ip} = 0 \leadsto \sum_{i=1}^{n} b_0 x_{ip} + \sum_{i=1}^{n} b_1 x_{i1} x_{ip} + \sum_{i=1}^{n} b_2 x_{i2} x_{ip} + \ldots + \sum_{i=1}^{n} b_p x_{ip} x_{ip} = \sum_{i=1}^{n} x_{ip} y_i.$$

Equações normais

- Sistema de equações lineares na forma matricial

$$\begin{bmatrix} n & \sum x_{i1} & \sum x_{i2} & \cdots & \sum x_{ip} \\ \sum x_{i1} & \sum x_{i1}x_{i1} & \sum x_{i2}x_{i1} & \cdots & \sum x_{ip}x_{i1} \\ \sum x_{i2} & \sum x_{i1}x_{i2} & \sum x_{i2}x_{i2} & \cdots & \sum x_{ip}x_{i2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_{ip} & \sum x_{i1}x_{ip} & \sum x_{i2}x_{ip} & \cdots & \sum x_{ip}x_{ip} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_p \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_{i1}y_i \\ \sum x_{i2}y_i \\ \vdots \\ \sum x_{ip}y_i \end{bmatrix}. \quad (8)$$

- Vetor solução b $((p+1) \times 1)$ do sistema de equações lineares fornece os parâmetros para a equação de quadrados mínimos

$$u = b_0 + b_1x_1 + b_2x_2 + \ldots + b_px_p.$$

Regressão linear múltipla: qualidade do ajuste

- Coeficiente de determinação

$$\boxed{r^2 = 1 - \frac{D(b_0, b_1, \ldots, b_p)}{\sum y_i^2 - \frac{1}{n}\left(\sum y_i\right)^2}}. \tag{9}$$

- Variância residual

$$\boxed{\sigma^2 = \frac{D(b_0, b_1, b_2, \ldots, b_p)}{n-p}}. \tag{10}$$

Regressão polinomial

- Caso particular da regressão linear múltipla.
- Relaciona a variável resposta y com uma variável explicativa x

$$y = \beta_0 + \beta_1 x + \beta_2 x^2 + \ldots + \beta_g x^g + \epsilon.$$

- Equações normais

$$\boxed{\begin{bmatrix} n & \sum x_i & \sum x_i^2 & \cdots & \sum x_i^g \\ \sum x_i & \sum x_i^2 & \sum x_i^3 & \cdots & \sum x_i^{g+1} \\ \sum x_i^2 & \sum x_i^3 & \sum x_i^4 & \cdots & \sum x_i^{g+2} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \sum x_i^g & \sum x_i^{g+1} & \sum x_i^{g+2} & \cdots & \sum x_i^{2g} \end{bmatrix} \begin{bmatrix} b_0 \\ b_1 \\ b_2 \\ \vdots \\ b_g \end{bmatrix} = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \\ \sum x_i^2 y_i \\ \vdots \\ \sum x_i^g y_i \end{bmatrix}}. \qquad (11)$$

Algoritmo: regressão linear múltipla e polinomial via equações normais

```
Algoritmo Regressão_linear_EN
{ Objetivo: Calcular parâmetros de quadrados mínimos de modelo linear múltiplo }
{           via equações normais }
parâmetros de entrada n, v, p, x, y
  { número de pontos, número de variáveis, número de parâmetros, }
  { variáveis explicativas e variáveis respostas }
parâmetros de saída b, r2, sigma2, CondErro
  { coef. de regressão, coef. de determinação, variância residual e condição de erro }
  se v > 1 e v + 1 ≠ p então CondErro ← 1, abandone, fim se
  CondErro ← 0; vp1 ← v + 1; pm1 ← p − 1
  para i ← 1 até n faça   { inclusão de uma coluna de 1's relativa à b0 }
    para j ← vp1 até 2 passo −1 faça x(i, j) ← x(i, j−1) fim para; x(i, 1) ← 1
  fim para
  se v = 1 e p > 2 então   { se regressão polinomial, então gera potências de x }
    para j ← 2 até pm1 faça
      jp1 ← j + 1
      para i ← 1 até n faça x(i, jp1) ← x(i, 2)^j, fim para
    fim para
  fim se
  { equações normais }
  para i ← 1 até p faça
    para j ← 1 até p faça
      Soma ← 0; para k ← 1 até n faça, Soma ← Soma + x(k, i) ∗ x(k, j), fim para
      Sxx(i, j) ← Soma   { matriz dos coeficientes }
    fim para
    Soma ← 0; para k ← 1 até n faça, Soma ← Soma + x(k, i) ∗ y(k), fim para
    Sxy(i) ← Soma   { vetor dos termos independentes }
  fim para
  [L, Det, CondErro] ← Cholesky(p, Sxx)  { decomposição de Cholesky }
  t ← Substituições_Sucessivas(p, L, Sxy)
  para i ← 1 até p faça
    para j ← 1 até i faça U(j, i) ← L(i, j), fim para; { U = transposta de L }
  fim para
  b ← Substituições_Retroativas(p, U, t)  { coeficientes }
  D ← 0; Sy2 ← 0
  para i ← 1 até n faça
    u ← 0; para j ← 1 até p faça, u ← u + b(j) ∗ x(i, j), fim para
    d ← y(i) − u; D ← D + d^2; Sy2 ← Sy2 + y(i)^2
  fim para
  r2 ← 1 − D/(Sy2 − Sxy(1)^2/n)   { coeficiente de determinação }
  sigma2 ← D/(n − p)   { variância residual }
fim algoritmo
```

|⇐

Definição de modelo no algoritmo

- Modelos permitidos e correspondentes valores de v e p

$$u = b_0 + b_1 x \rightsquigarrow v = 1,\ p = 2,$$

$$u = b_0 + b_1 x_1 + b_2 x_2 \rightsquigarrow v = 2,\ p = 3 \text{ e}$$
$$u = b_0 + b_1 x + b_2 x^2 \rightsquigarrow v = 1,\ p = 3.$$

- Modelos não permitidos *neste* algoritmo: $v > 1$ **e** $v + 1 \neq p$

$$u = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_2^2 \rightsquigarrow v = 2,\ p = 4.$$

Complexidade: regressão linear múltipla e polinomial via equações normais

Operações	Complexidade
adições	$(p^2+3p+2)n+5$
multiplicações	$(p^2+2p+2)n+1$
divisões	3

Regressão linear múltipla

Operações	Complexidade
adições	$(p^2+2p+4)n+p+3$
multiplicações	$(\frac{3}{2}p^2+\frac{1}{2}p+3)n+1$
divisões	3

Regressão polinomial

- n: número de pontos.
- p: número de parâmetros.
- Potenciação contada como multiplicação.
- Desconsideradas operações para solução do sistema linear.

Exemplo: produto interno bruto dos Estados Unidos de 1947 a 1962

Exemplo 4 Ajustar os dados da tabela ao modelo

$$u = b_0 + b_1 x_1 + b_2 x_2,$$

usando o algoritmo

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
x_{i1}	60,3	61,1	60,2	61,2	63,2	63,6	65,0	63,8	66,0	67,9	68,2	66,5	68,7	69,6	69,3	70,6
x_{i2}	108	109	110	112	112	113	115	116	117	119	120	122	123	125	128	130
y_i	234	259	258	285	329	347	365	363	396	419	443	445	483	503	518	555

,

- x_1: total de empregos (milhões),
- x_2: população com 14 anos ou mais (milhões) e
- y: o PIB americano (bilhões de dólares).

Exemplo: produto interno bruto dos Estados Unidos de 1947 a 1962 cont.

```
% Os parametros de entrada
n = 16
v = 2
p = 3
x =
   60.3000  108.0000
   61.1000  109.0000
   60.2000  110.0000
   61.2000  112.0000
   63.2000  112.0000
   63.6000  113.0000
   65.0000  115.0000
   63.8000  116.0000
   66.0000  117.0000
   67.9000  119.0000
   68.2000  120.0000
   66.5000  122.0000
   68.7000  123.0000
   69.6000  125.0000
   69.3000  128.0000
   70.6000  130.0000
```

Exemplo: produto interno bruto dos Estados Unidos de 1947 a 1962 cont.

```
y =
   234
   259
   258
   285
   329
   347
   365
   363
   396
   419
   443
   445
   483
   503
   518
   555
% produzem os resultados
coeficientes de regressao
b(0) = -1.40740e+03
b(1) =  1.34511e+01
b(2) =  7.80271e+00
coeficiente de determinacao = 0.99267
variancia residual          = 8.37581e+01
condicao de erro            = 0
```

- Equação de quadrados mínimos $u = -1{,}40740_{\times}10^3 + 1{,}34511_{\times}10^1 x_1 + 7{,}80271 x_2$.
- Coeficiente de determinação $r^2 = 0{,}99267$.
- Variância residual $\sigma^2 = 83{,}7581$. $|\Leftarrow$

Exemplo: raiz quadrada de x para $0{,}01 \leq x \leq 1$

Exemplo 5 A partir da tabela, que compila valores de $\sqrt{x}$ para $0{,}01 \leq x \leq 1$, determinar o polinômio de quadrados mínimos de grau $g = 3$ utilizando o algoritmo

i	x_i	$\sqrt{x_i}$
1	0,01	0,1000
2	0,10	0,3162
3	0,20	0,4472
4	0,30	0,5477
5	0,40	0,6325
6	0,50	0,7071
7	0,60	0,7746
8	0,70	0,8367
9	0,80	0,8944
10	0,90	0,9487
11	1,00	1,0000

.

Exemplo: raiz quadrada de x para $0{,}01 \leq x \leq 1$ cont.

```
% Os parametros de entrada
n = 11
v = 1
p = 4
x =
    0.0100
    0.1000
    0.2000
    0.3000
    0.4000
    0.5000
    0.6000
    0.7000
    0.8000
    0.9000
    1.0000
y =
    0.1000
    0.3162
    0.4472
    0.5477
    0.6325
    0.7071
    0.7746
    0.8367
    0.8944
    0.9487
    1.0000
```

Exemplo: raiz quadrada de x para $0{,}01 \leq x \leq 1$ cont.

```
% fornecem os resultados
coeficientes de regressao
b(0) =  1.01126e-01
b(1) =  2.06854e+00
b(2) = -2.17822e+00
b(3) =  1.01865e+00
coeficiente de determinacao = 0.99738
variancia residual          = 2.96085e-04
condicao de erro            = 0
```

- Polinômio de quadrados mínimos que aproxima $\sqrt{x}$ para $0{,}01 \leq x \leq 1$

$$u = 1{,}01865x^3 - 2{,}17822x^2 + 2{,}06854x + 0{,}10113.$$

- Coeficiente de determinação $r^2 = 0{,}99738$.
- Variância residual $\sigma^2 = 2{,}96085{\times}10^{-4}$.

Exemplo: população do Brasil

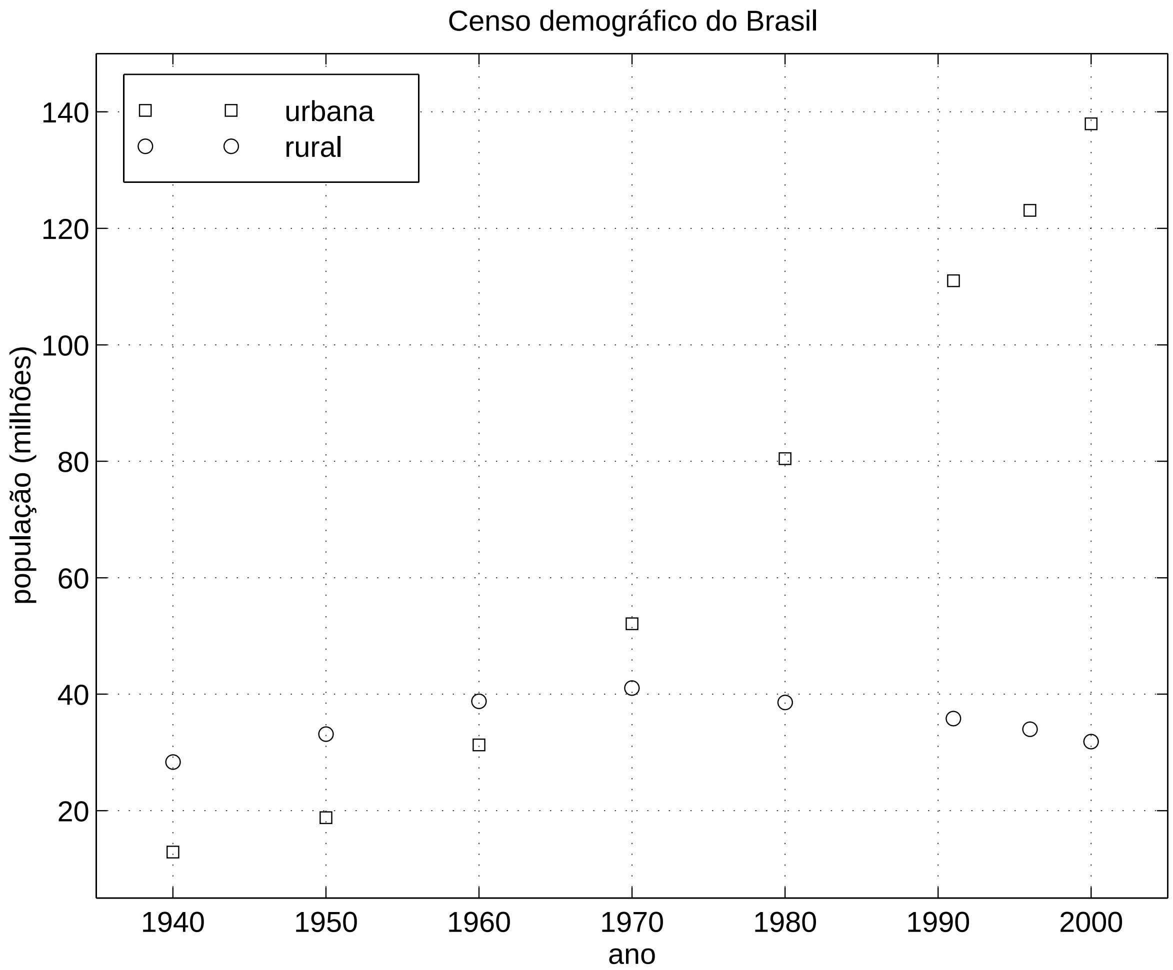
Exemplo 6 Sejam os dados históricos dos censos demográficos do Brasil, de acordo com o Instituto Brasileiro de Geografia e Estatística (IBGE), apresentados na tabela

Ano	Urbana	Rural
1940	12.880.182	28.356.133
1950	18.782.891	33.161.506
1960	31.303.034	38.767.423
1970	52.084.984	41.054.053
1980	80.436.409	38.566.297
1991	110.990.990	35.834.485
1996	123.076.831	33.993.332
2000	137.953.959	31.845.211

. $|\Leftarrow$

Deseja-se determinar qual o melhor grau para uma regressão polinomial.

Exemplo: diagrama de dispersão



- Ajuste **`urbana x ano`** não deve ser feito por um polinômio de grau 1.
- Polinômio de grau mais elevado.

Exemplo: escolha do grau do polinômio de quadrados mínimos

g	r^2	σ^2
1	0,96602	$9{,}58219{\times}10^{1}$
2	0,99776	$7{,}59261{\times}10^{0}$
3	0,99939	$2{,}57228{\times}10^{0}$
4	0,99940	$3{,}42930{\times}10^{0}$
5	0,99980	$1{,}65615{\times}10^{0}$
6	0,99998	$4{,}17203{\times}10^{-1}$

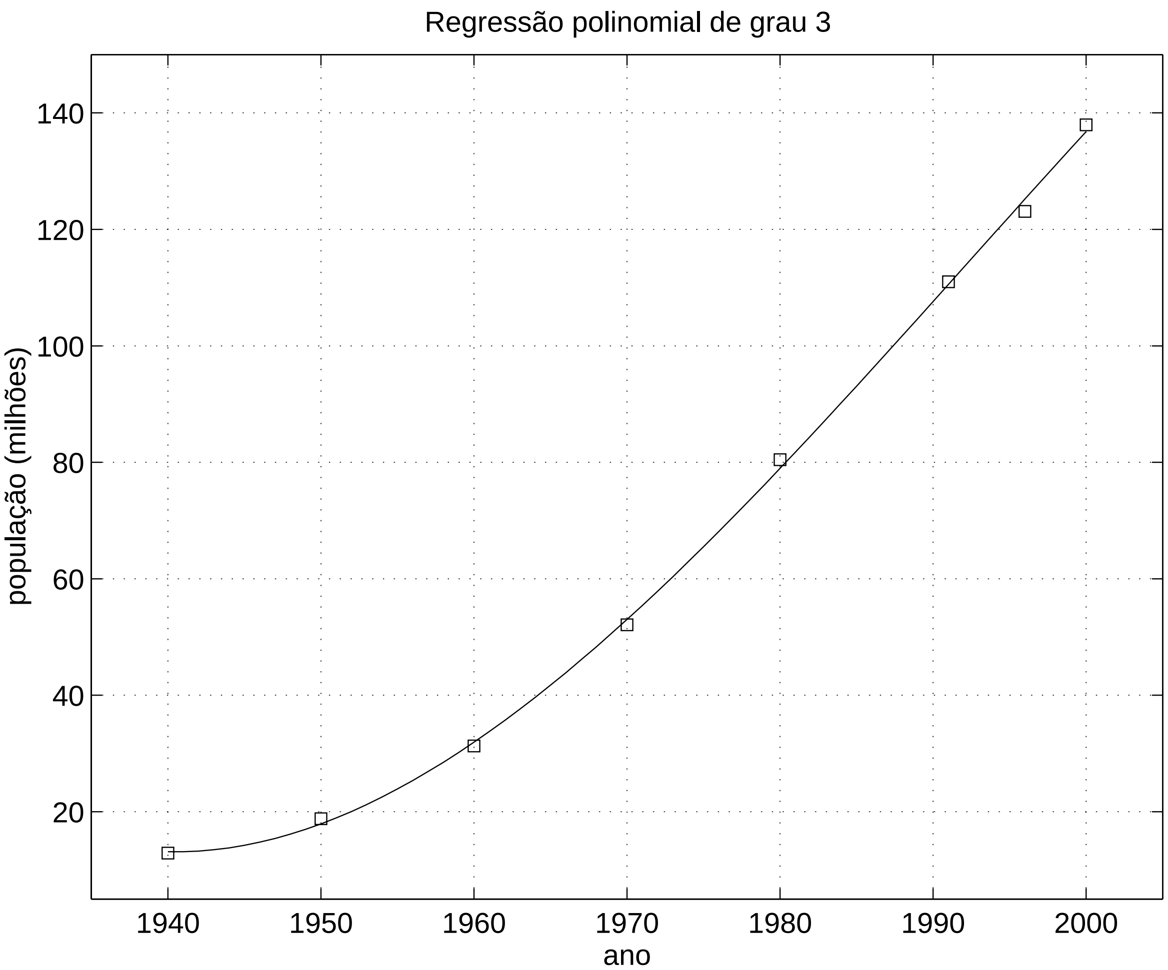
- g: grau do polinômio.
- r^2: coeficiente de determinação.
- σ^2: variância residual.
- Mudanças de variáveis para reduzir erros de arredondamento.
- Variável explicativa centrada de modo que $x = \text{Ano} - 1970$.
- Variável resposta dada em milhões de habitantes $y = \text{Urbana}{\times}10^{-6}$.

Exemplo: escolha do grau do polinômio de quadrados mínimos cont.

g	r^2	σ^2
1	0,96602	$9{,}58219{\times}10^{1}$
2	0,99776	$7{,}59261{\times}10^{0}$
3	0,99939	$2{,}57228{\times}10^{0}$
4	0,99940	$3{,}42930{\times}10^{0}$
5	0,99980	$1{,}65615{\times}10^{0}$
6	0,99998	$4{,}17203{\times}10^{-1}$

- r^2 aumenta quando grau do polinômio de quadrados mínimos aumenta.
- σ^2 vai reduzindo até grau $g = 3$ e depois começa a oscilar.
- Grau escolhido para o ajuste polinomial.
- Evitar grau elevado.
- Ideal seria $n \gg p$.

Exemplo: polinômio de grau 3



urbana x ano

Transformações não lineares

- Modelos não lineares nos parâmetros transformados em modelos lineares.
- Substituição de variáveis por funções dessas variáveis.

$$y = ax^b \rightsquigarrow \log_e(y) = \log_e(a) + b\log_e(x);$$

$$y = ab^x \rightsquigarrow \log_e(y) = \log_e(a) + \log_e(b)x;$$

$$y = ae^{bx} \rightsquigarrow \log_e(y) = \log_e(a) + bx;$$

$$y = e^{a+bx_1+cx_2} \rightsquigarrow \log_e(y) = a + bx_1 + cx_2;$$

$$y = ax_1^b x_2^c \rightsquigarrow \log_e(y) = \log_e(a) + b\log_e(x_1) + c\log_e(x_2);$$

$$y = \frac{1}{a+bx_1+cx_2} \rightsquigarrow \frac{1}{y} = a + bx_1 + cx_2;$$

$$y = \frac{1}{1+e^{a+bx_1+cx_2}} \rightsquigarrow \log_e\left(\frac{1}{y} - 1\right) = a + bx_1 + cx_2.$$

Exemplo: cinética de reação química de primeira ordem

Exemplo 7 Em uma reação química de primeira ordem, a constante k de velocidade se relaciona com a concentração c e o tempo t pela expressão

$$c = c_0 e^{-kt},$$

onde c_0 é a concentração inicial de um reagente. Usando os dados da tabela e o algoritmo, calcular a constante de velocidade

i	1	2	3	4	5
t	0,1	0,2	0,3	0,4	0,5
c	0,56	0,32	0,21	0,11	0,08

,

- t: tempo (segundos).
- c: concentração (M).
- Para tal,

$$c = c_0 e^{-kt} \rightsquigarrow \log_e(c) = \log_e(c_0) - kt.$$

Exemplo: cinética de reação química de primeira ordem cont.

```
% Os parametros de entrada
n = 5
v = 1
p = 2
t =
    0.1000
    0.2000
    0.3000
    0.4000
    0.5000
logc =
   -0.5798
   -1.1394
   -1.5606
   -2.2073
   -2.5257
% fornecem os resultados
b(0) = -1.14650e-01
b(1) = -4.95970e+00
coeficiente de determinacao = 0.99179
variancia residual          = 6.78379e-03
condicao de erro            = 0
```

- Equação de regressão: $\log_e(c) = -1{,}14650{\times}10^{-1} - 4{,}95970t \rightsquigarrow c_0 = e^{-1{,}14650{\times}10^{-1}} = 0{,}89168\ M$.
- $k = 4{,}95970$ segundos^{-1}.
- Coeficiente de determinação $r^2 = 0{,}99179$.
- Variância residual $\sigma^2 = 6{,}78379{\times}10^{-3}$.

Malcondicionamento das equações normais

- Seja a equação de regressão polinomial

$$u = b_0 + b_1 x + b_2 x^2 + \ldots + b_g x^g,$$

- parâmetros b_i calculados pelas equações normais.
- Dados `(ano, urbana)` da tabela com número de pontos $n = 8$.

Malcondicionamento das equações normais cont.

- g: grau do polinômio de regressão.
- r^2: coeficiente de determinação.
- $\kappa_2(X^T X)$: número de condição em norma-2 da matriz dos coeficientes $X^T X$ das equações normais.

g	r^2	$\kappa_2(X^T X)$
1	0,96602	$4{,}514_{\times}10^{2}$
2	0,99776	$8{,}298_{\times}10^{5}$
3	0,99939	$6{,}498_{\times}10^{8}$
4	0,99940	$8{,}591_{\times}10^{11}$
5	0,99980	$7{,}418_{\times}10^{14}$
6	0,99998	$9{,}923_{\times}10^{17}$
7	1,00000	$6{,}753_{\times}10^{20}$

.

- À medida que o grau g do polinômio aumenta, $r^2 \longrightarrow 1$.
- $\kappa_2(X^T X) \longrightarrow \infty$.
- Equações normais possuem matriz dos coeficientes malcondicionada.

Ajuste via decomposição em valores singulares

- Modelo de regressão linear múltipla na forma matricial

$$y = X\beta + \epsilon,$$

- y: vetor $(n \times 1)$ contendo as n observações da variável resposta,
- X: matriz $(n \times (p+1))$, $n \geq p+1$, contendo os n valores das p variáveis explicativas, além da primeira coluna de 1's relativa a β_0,
- β: vetor $((p+1) \times 1)$ dos parâmetros a serem estimados e
- ϵ: vetor $(n \times 1)$ dos erros aleatórios.

Forma matricial

$$y = \begin{bmatrix} y_1 \\ y_2 \\ y_3 \\ y_4 \\ \vdots \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \cdots & x_{1p} \\ 1 & x_{21} & x_{22} & \cdots & x_{2p} \\ 1 & x_{31} & x_{32} & \cdots & x_{3p} \\ 1 & x_{41} & x_{42} & \cdots & x_{4p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \text{ e } \epsilon = \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \epsilon_3 \\ \epsilon_4 \\ \vdots \\ \epsilon_n \end{bmatrix}.$$

Cálculo dos parâmetros

- Método dos quadrados mínimos: minimizar a função

$$f(\beta) = \|y - X\beta\|_2^2 = (y - X\beta)^T (y - X\beta).$$

- Pelas regras de diferenciação matricial

$$\frac{\partial f(\beta)}{\partial \beta^T} = \frac{\partial f(\beta)}{\partial (y - X\beta)^T} \frac{\partial (y - X\beta)}{\partial \beta^T} = 2(y - X\beta)^T(-X) = -2(y - X\beta)^T X.$$

$$\frac{\partial f(\beta)}{\partial \beta} = -2X^T(y - X\beta).$$

- A função $f(\beta)$ apresenta um mínimo em $f(b)$, onde b é o ponto em que a derivada se anula

$$\frac{\partial f(b)}{\partial \beta} = -2X^T(y - Xb) = 0 \rightsquigarrow (X^T X)b = X^T y.$$

- Equações normais (8) na forma matricial.

Equações normais

- Segunda derivada

$$\frac{\partial(\partial f(\beta)/\partial\beta)}{\partial\beta^T} = \frac{\partial(-2X^Ty + 2X^TX\beta)}{\partial\beta^T} = 2X^TX.$$

- Matriz X^TX tem elementos reais, é não singular e definida positiva.
- O ponto $f(b)$ é, de fato, um mínimo de $f(\beta)$.
- O sistema $(X^TX)b = X^Ty$ apresenta uma única solução.
- Equações normais formam um sistema malcondicionado.
- Processo alternativo para a estimativa de β que evita a formação da matriz X^TX.

Decomposição em valores singulares

- Decomposição em valores singulares consiste em fatorar uma matriz X $(n \times (p+1))$, tal que
$$X = USV^T, \tag{12}$$
- U: matriz ortogonal $(n \times n)$,
- V: matriz ortogonal $((p+1) \times (p+1))$ e
- S: matriz diagonal $(n \times (p+1))$ da forma
$$S = \begin{bmatrix} S_1 \\ 0 \end{bmatrix},$$
- S_1: matriz quadrada diagonal de ordem $p+1$.

Estimativa dos parâmetros

- Soma de quadrados residual

$$\|y - Xb\|_2^2 = \|y - USV^T b\|_2^2.$$

- Matriz ortogonal U^T não altera o valor da norma

$$\|y - Xb\|_2^2 = \|U^T y - U^T USV^T b\|_2^2 \rightsquigarrow \|y - Xb\|_2^2 = \|U^T y - SV^T b\|_2^2.$$

- Definindo

$$U^T y = a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tag{13}$$

- a_1: vetor de tamanho $p+1$ e a_2: vetor de tamanho $n-p-1$,

$$\tilde{b} = V^T b, \tag{14}$$

$$S\tilde{b} = \begin{bmatrix} S_1 \\ 0 \end{bmatrix} \tilde{b} = \begin{bmatrix} S_1\tilde{b} \\ 0 \end{bmatrix},$$

$$\|y - Xb\|_2^2 = \left\| \begin{bmatrix} a_1 \\ a_2 \end{bmatrix} - \begin{bmatrix} S_1\tilde{b} \\ 0 \end{bmatrix} \right\|_2^2 \rightsquigarrow \|y - Xb\|_2^2 = \|a_1 - S_1\tilde{b}\|_2^2 + \|a_2\|_2^2.$$

Estimativa dos parâmetros cont.

- Soma de quadrados residual será mínima quando $\tilde{b}$ for a solução do sistema diagonal

$$S_1\tilde{b} = a_1.$$

- Em vista de (14) e da ortogonalidade de V,

$$\boxed{b = V\tilde{b}}. \tag{15}$$

- Soma de quadrados residual

$$D(b_0, b_1, \ldots, b_p) = \|a_2\|_2^2 = a_2^T a_2.$$

- Valores preditos

$$u = Xb = USV^Tb = US\tilde{b} = U\begin{bmatrix} S_1\tilde{b} \\ 0 \end{bmatrix} \rightsquigarrow \boxed{u = U\begin{bmatrix} a_1 \\ 0 \end{bmatrix}}.$$

- Vetor desvio d

$$\boxed{d = U\begin{bmatrix} 0 \\ a_2 \end{bmatrix}}. \tag{16}$$

Exemplo: regressão linear via DVS

Exemplo 8 Calcular os parâmetros da reta de quadrados mínimos do Exemplo 1 utilizando a decomposição em valores singulares.

- Matriz X e vetor y

$$X = \begin{bmatrix} 1 & 0{,}3 \\ 1 & 2{,}7 \\ 1 & 4{,}5 \\ 1 & 5{,}9 \\ 1 & 7{,}8 \end{bmatrix} \quad \text{e} \quad y = \begin{bmatrix} 1{,}8 \\ 1{,}9 \\ 3{,}1 \\ 3{,}9 \\ 3{,}3 \end{bmatrix}.$$

Exemplo: regressão linear via DVS cont.

- Decomposição em valores singulares de X

$$U = \begin{bmatrix} 0{,}0414 & -0{,}8144 & -0{,}3904 & -0{,}3366 & -0{,}2635 \\ 0{,}2513 & -0{,}4559 & 0{,}1351 & 0{,}3940 & 0{,}7453 \\ 0{,}4087 & -0{,}1871 & 0{,}8174 & -0{,}2246 & -0{,}2816 \\ 0{,}5311 & 0{,}0219 & -0{,}2412 & 0{,}6611 & -0{,}4714 \\ 0{,}6972 & 0{,}3057 & -0{,}3209 & -0{,}4939 & 0{,}2712 \end{bmatrix},$$

$$S = \begin{bmatrix} 11{,}2679 & 0 \\ 0 & 1{,}1467 \\ 0 & 0 \\ 0 & 0 \\ 0 & 0 \end{bmatrix} \text{ e } V = \begin{bmatrix} 0{,}1713 & -0{,}9852 \\ 0{,}9852 & 0{,}1713 \end{bmatrix}.$$

Exemplo: regressão linear via DVS cont.

- Por (13)

$$a = U^T y = \begin{bmatrix} 6{,}1910 \\ -1{,}8179 \\ 0{,}0883 \\ 0{,}3949 \\ -0{,}8747 \end{bmatrix}.$$

- Vetor $\tilde{b}$ é a solução do sistema diagonal $S_1\tilde{b} = a_1$

$$\begin{bmatrix} 11{,}2679 & 0 \\ 0 & 1{,}1467 \end{bmatrix} \begin{bmatrix} \tilde{b}_1 \\ \tilde{b}_2 \end{bmatrix} = \begin{bmatrix} 6{,}1910 \\ -1{,}8179 \end{bmatrix} \leadsto \tilde{b} = \begin{bmatrix} 0{,}5494 \\ -1{,}5853 \end{bmatrix}.$$

- Vetor b dos coeficientes é obtido por (15)

$$b = V\tilde{b} \leadsto b = \begin{bmatrix} 1{,}6559 \\ 0{,}2697 \end{bmatrix}.$$

- (ver exemplo).

Algoritmo: regressão linear múltipla e polinomial via DVS

```
Algoritmo Regressão_linear_DVS
{ Objetivo: Calcular parâmetros de quadrados mínimos de modelo linear múltiplo }
{           via decomposição em valores singulares }
parâmetros de entrada n, v, p, x, y
   { número de pontos, número de variáveis, número de parâmetros, }
   { variáveis explicativas e variáveis respostas }
parâmetros de saída b, r2, sigma2, CondErro
   { coef. de regressão, coef. de determinação, variância residual e condição de erro }
   se v > 1 e v + 1 ≠ p então CondErro ← 1, abandone, fim se
   CondErro ← 0; vp1 ← v + 1; pm1 ← p − 1
   para i ← 1 até n faça   { inclusão de uma coluna de 1's relativa à b_0 }
      para j ← vp1 até 2 passo −1 faça x(i, j) ← x(i, j − 1) fim para
      x(i, 1) ← 1
   fim para
   se v = 1 e p > 2 então   { se regressão polinomial, então gera potências de x }
      para j ← 2 até pm1 faça
         jp1 ← j + 1
         para i ← 1 até n faça x(i, jp1) ← x(i, 2)^j, fim para
      fim para
   fim se
   [U, S, V] ← dvs(x)   { chamada da rotina para decomposição em valores singulares }
   para i ← 1 até n faça   { Cálculo do vetor auxiliar a = U^T y }
      a(i) ← 0
      para j ← 1 até n faça a(i) ← a(i) + U(j, i) ∗ y(j) fim para
   fim para
   para i ← 1 até p faça   { Cálculo do vetor dos coeficientes b = V S_1^{-1} a_1 }
      b(i) ← 0
      para j ← 1 até p faça
         b(i) ← b(i) + V(i, j)/S(j, j) ∗ a(j)
      fim para
   fim para
   D ← 0   { Cálculo do desvio }
   para i ← p + 1 até n faça, D ← D + a(i)^2, fim para
   Sy ← 0; Sy2 ← 0   { Cálculo dos somatórios auxiliares }
   para i ← 1 até n faça
      Sy ← Sy + y(i); Sy2 ← Sy2 + y(i)^2
   fim para
   r2 ← 1 − D/(Sy2 − Sy^2/n)   { coeficiente de determinação }
   sigma2 ← D/(n − p)   { variância residual }
fim algoritmo
```

|⇐

Definição de modelo no algoritmo

- Modelos permitidos e correspondentes valores de v e p

$$u = b_0 + b_1 x \rightsquigarrow v = 1,\ p = 2,$$

$$u = b_0 + b_1 x_1 + b_2 x_2 \rightsquigarrow v = 2,\ p = 3 \text{ e}$$

$$u = b_0 + b_1 x + b_2 x^2 \rightsquigarrow v = 1,\ p = 3.$$

- Modelos não permitidos *neste* algoritmo: $v > 1$ **e** $v + 1 \neq p$

$$u = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_2^2 \rightsquigarrow v = 2,\ p = 4.$$

Complexidade: regressão linear múltipla e polinomial via DVS

Operações	Complexidade
adições	$n^2+(p+2)n+p^2-p+5$
multiplicações	n^2+2n+p^2-p+1
divisões	p^2+3

Regressão linear múltipla

Operações	Complexidade
adições	n^2+4n+p^2+3
multiplicações	$n^2+(\frac{1}{2}p^2-\frac{3}{2}p+3)n+p^2-p+1$
divisões	p^2+3

Regressão polinomial

- n: número de pontos,
- p: número de parâmetros.
- Potenciação tratada como multiplicações.
- Desconsideradas operações para decomposição em valores singulares.

Exemplo: uso do algoritmo

Exemplo 9 Ajustar os dados do Exemplo 4 ao modelo $u = b_0 + b_1x_1 + b_2x_2$ usando o algoritmo.

```
% Os parametros de entrada
n = 16
v = 2
p = 3
x =
   60.3000  108.0000
   61.1000  109.0000
   60.2000  110.0000
   61.2000  112.0000
   63.2000  112.0000
   63.6000  113.0000
   65.0000  115.0000
   63.8000  116.0000
   66.0000  117.0000
   67.9000  119.0000
   68.2000  120.0000
   66.5000  122.0000
   68.7000  123.0000
   69.6000  125.0000
   69.3000  128.0000
   70.6000  130.0000
```


Exemplo: uso do algoritmo cont.

```
y =
   234
   259
   258
   285
   329
   347
   365
   363
   396
   419
   443
   445
   483
   503
   518
   555
% produzem os resultados
coeficientes de regressao
b(0) = -1.40740e+03
b(1) =  1.34511e+01
b(2) =  7.80271e+00
coeficiente de determinacao = 0.99267
variancia residual          = 8.37581e+01
condicao de erro            = 0
```

- Equação de quadrados mínimos: $u = -1{,}40740{\times}10^3 + 1{,}34511{\times}10^1 x_1 + 7{,}80271 x_2$,
- $r^2 = 0{,}99267$ e $\sigma^2 = 83{,}7581$ (ver exemplo).

Comparação dos métodos computacionais para RLM

- Equações normais: vantagens
  - Maior velocidade com que podem ser formadas e resolvidas.
  - Com o uso de *precisão dupla*, a diferença de exatidão dos dois métodos, poucas vezes, valerá a pena ser considerada.
- Equações normais: desvantagens
  - Número de condição da matriz $X^T X$ é o quadrado daquele da matriz X.
  - Difícil computar as matrizes $X^T X$ e $X^T y$, exatamente.
  - Perturbações feitas no problema básico podem ter conseqüências desastrosas.

Comparação dos métodos computacionais para RLM cont.

- Decomposição em valores singulares: vantagens
  - Superiores propriedades numéricas.
  - Grande quantidade de memória disponível a um custo relativamente baixo.
- Decomposição em valores singulares: desvantagens
  - Requerem maior quantidade de memória.
  - Complexidade computacional é maior que a da decomposição de Cholesky.

Diferença entre regressão e interpolação

- Polinômio interpolador de grau $n-1$ construído de modo a passar por n pontos
$$P_{n-1}(x) = a_0 + a_1x + a_2x^2 + \ldots + a_{n-1}x^{n-1}.$$
- Possui n coeficientes $a_i,\ i = 0, 1, \ldots, n-1$.
- Número de pontos utilizados para gerar o polinômio interpolador é igual ao número de coeficientes do polinômio.
- Polinômio de regressão de grau g, utilizando n pontos
$$U_g(x) = b_0 + b_1x + b_2x^2 + \ldots + b_gx^g,$$
- tal que $g \leq n-1$.
- Quando $g = n-1$ o polinômio de regressão será idêntico ao polinômio interpolador.

Sistema linear e equações normais

- Polinômio interpolador de grau $g = 1$ que passa por $n = 2$ pontos (x_1, y_1) e (x_2, y_2).
- Coeficientes obtidos pela solução do sistema linear

$$\begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 \\ y_2 \end{bmatrix}.$$

- Pré-multiplicando pela transposta da matriz dos coeficientes

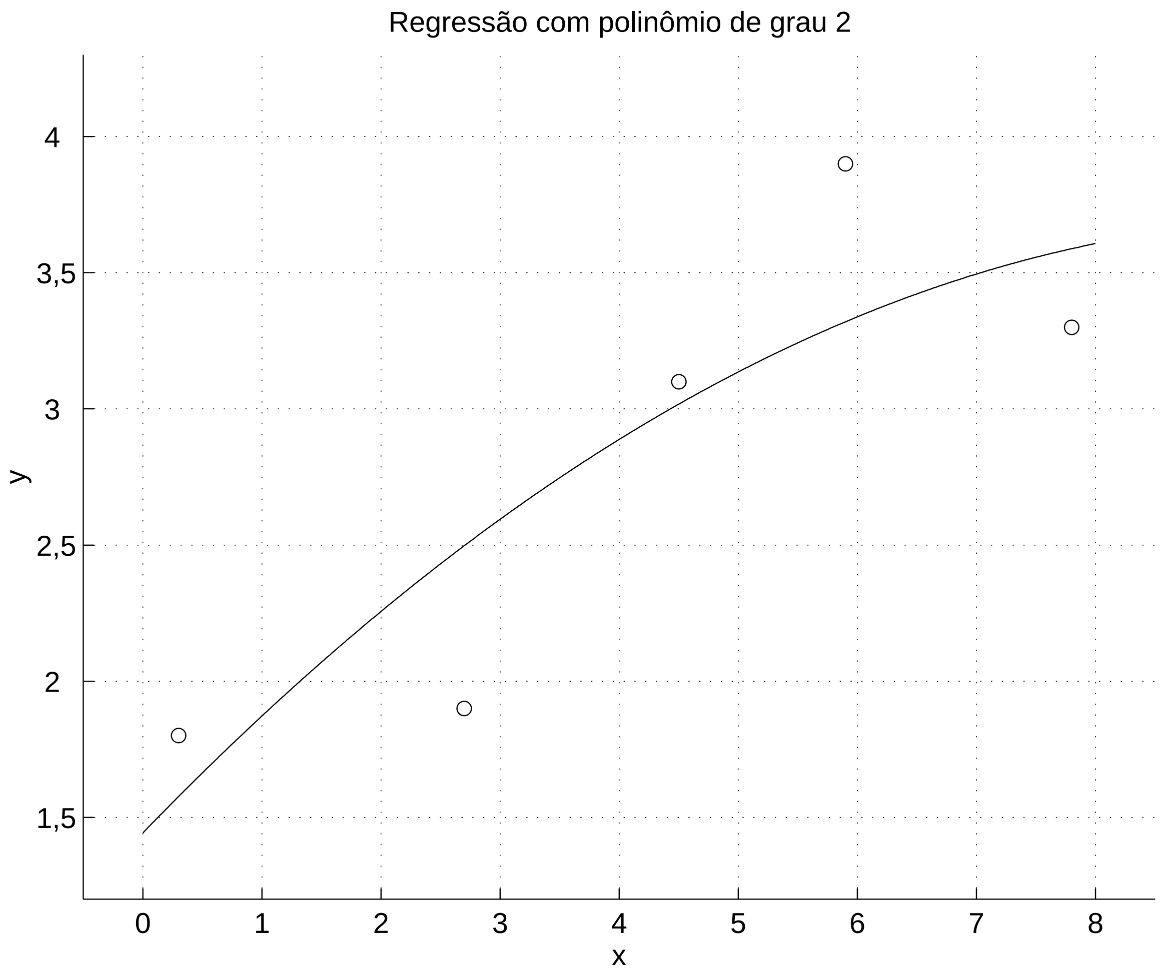
$$\begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \rightsquigarrow$$

$$\begin{bmatrix} 2 & x_1 + x_2 \\ x_1 + x_2 & x_1^2 + x_2^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} y_1 + y_2 \\ x_1 y_1 + x_2 y_2 \end{bmatrix}.$$

- Sistema linear idêntico às equações normais (1), para $n = 2$, utilizadas para calcular os parâmetros de uma regressão linear simples.

Regressão polinomial quadrática

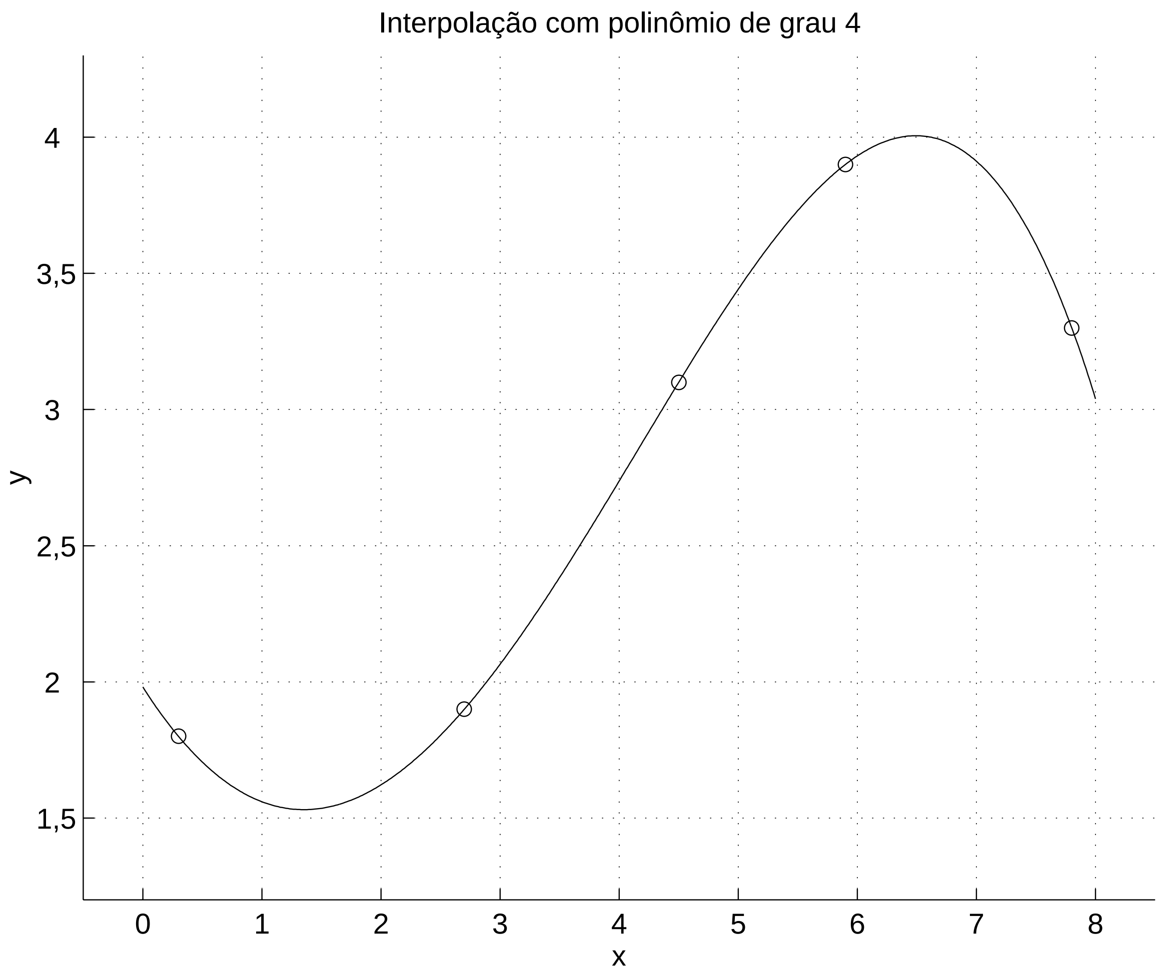
- Dados da tabela.



$$n = 5 \text{ e } g = 2$$

Regressão idêntica à interpolação

- Polinômio passa por todos os pontos do diagrama de dispersão.



$$n = 5 \text{ e } g = n - 1 = 4$$

Uso da regressão e da interpolação

- Em termos de complexidade computacional, a interpolação é um processo mais simples que a regressão polinomial.
- A interpolação deve ser utilizada quando se necessita de um valor intermediário não constante de uma tabela.
- A regressão tem que ser utilizada quando se deseja estimar um parâmetro de um modelo semideterminístico e/ou prever um valor dado por esse modelo.
- A variância residual torna-se indefinida quando o número de parâmetros p do modelo for igual ao número de pontos n

$$\sigma^2 = \frac{D(b_0, b_1, b_2, \ldots, b_p)}{n - p}.$$

Fim

Capítulo 4: Ajuste de curvas