4. Ajuste de curvas

4.1 Relações entre variáveis.

4.2 Regressão linear simples.

4.3 Qualidade do ajuste.

4.4 Regressão linear múltipla.

4.5 Formas alternativas de estimar os parâmetros.

4.6 Diferença entre regressão e interpolação.

4.7 Estudos de caso:

- Tensão-deformação de aço.
- Produto iônico da água.

4.8 Exercícios.

Relações entre variáveis

- Relacionar, por meio de um modelo matemático, a variável resposta (ou dependente) com o conjunto de variáveis explicativas (ou independentes).

- Para ter controle, determinar algum parâmetro ou mesmo fazer previsão acerca do comportamento da variável resposta.

- Variação da leitura de uma variável:

  - erros de medida experimentais;

  - variáveis cujos valores se alteram durante o experimento.

- Tipos de relações entre as variáveis

  1. determinísticas,

  2. semideterminísticas e

  3. empíricas.

Relações determinísticas

- Variáveis relacionadas entre si por uma lei expressa por fórmula matemática precisa.

- Variação nas observações é atribuída a erros experimentais.

- Por exemplo, se r reais forem investidos durante m meses a uma taxa de juros j, ao final do prazo ter-se-á v reais.

- As variáveis r, m, j e v estão relacionadas pela expressão exata fornecida pela Matemática Financeira

  $$v = r(1 + j)^m,$$

  que é a lei dos juros compostos.

- Qualquer análise adicional é desnecessária para relacionar estas variáveis.

Relações semideterminísticas

- Teoria prescreve forma para a relação.
- Mas não os valores particulares dos parâmetros que aparecem na relação.
- É necessário realizar experimentos para obter informações acerca desses parâmetros.
- Precisão limitada dos instrumentos de medida.
- Perturbações incontroláveis dos experimentos.
- Outros fatores introduzem erros nos dados.
- Causam perturbação na verdadeira relação.
- Por exemplo, a concentração c de uma substância após um tempo t em uma reação química de primeira ordem é

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

  c_0: concentração inicial e k: constante de velocidade *de uma reação específica*.
- A constante k é obtida experimentalmente.

Relações empíricas

- Relação entre as variáveis envolvidas não são conhecidas.

- Determinar uma fórmula matemática que relacione essas variáveis.

- Gráfico feito com valores observados dessas variáveis fornece uma idéia da relação entre elas com algumas variações aleatórias.

- Por exemplo, deseja-se conhecer em um experimento agrícola qual a relação entre a produção p de uma lavoura de feijão e a dosagem d de um certo fertilizante.

- Outros fatores influentes como acidez do solo, umidade e controle de pragas são mantidos constantes tanto quanto possível.

Relações empíricas cont.

- Experimento consiste em aplicar dosagens diferentes do fertilizante em áreas distintas.
- Anotar a produção de feijão em cada uma delas.
- Dosagens diferentes do fertilizante induzirão à produção de quantidades diferentes.
- Não esperar que a relação obtida siga uma fórmula matemática precisa, dada a complexidade do problema.
- Ter suficiente conhecimento sobre uma relação empírica.
- Desenvolver a teoria que conduza a uma fórmula matemática.
- Caso semideterminístico.

Regressão linear simples

- Relações mais simples entre duas variáveis são as relações lineares.
- A 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.

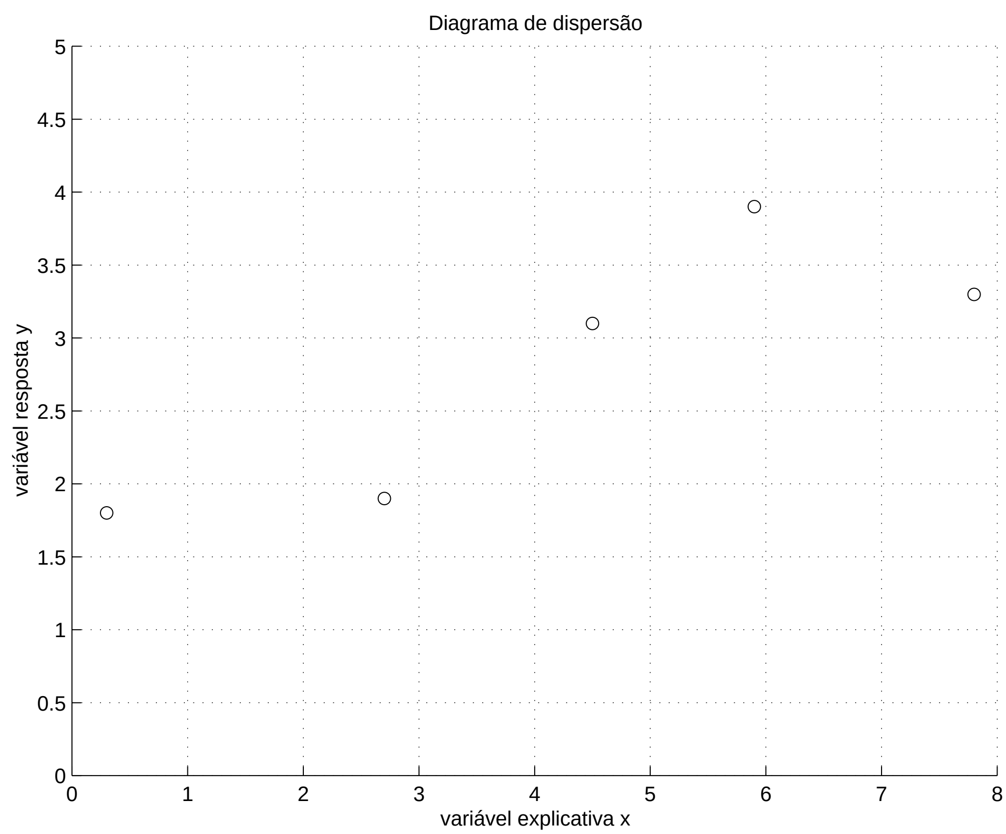
Diagrama de dispersão

- 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

.

- Diagrama de dispersão dos dados



Retas de regressão

- Modelo simples que relaciona as variáveis x e y

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

- β_0 e β_1 são os parâmetros a serem estimados.
- ϵ contém os 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 meio de polinômio interpolador linear.
- 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

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

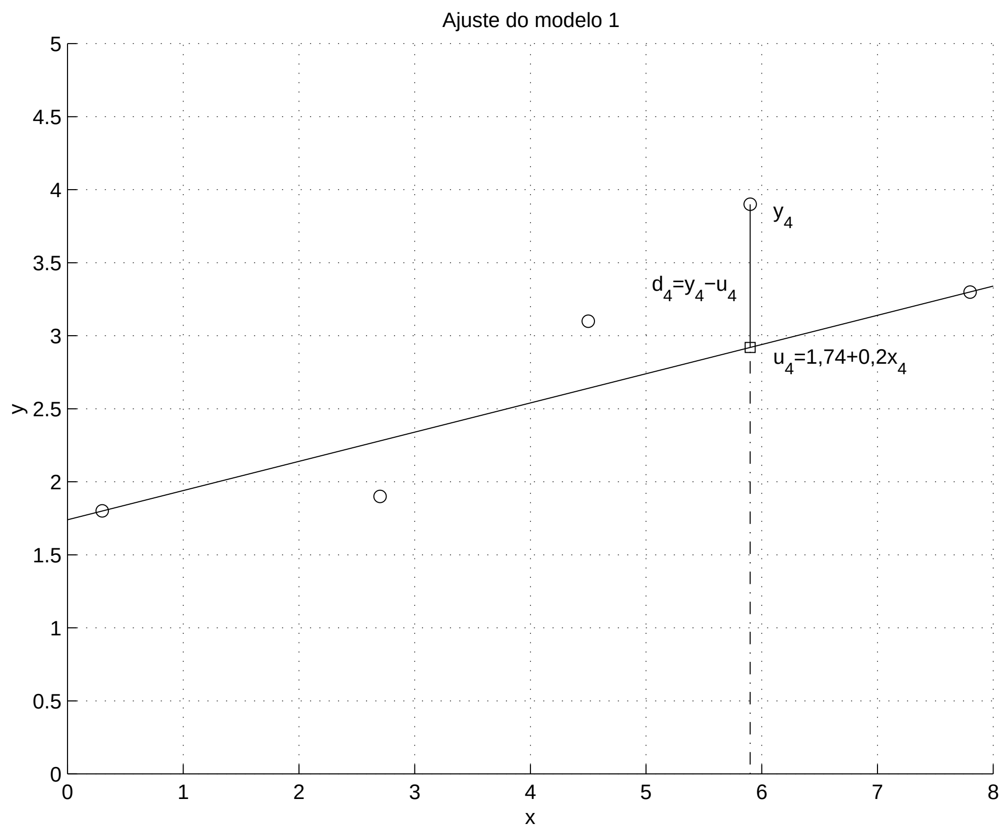
$$u(x) = 1{,}8 + 0{,}2(x - 0{,}3) \rightsquigarrow u(x) = 1{,}74 + 0{,}2x.$$

- Distância vertical d_i entre o i-ésimo ponto dado y_i e o ponto $u_i = 1{,}74 + 0{,}2x_i$ de mesma abscissa x_i

$$d_i = y_i - u_i.$$

Gráfico do modelo 1

❏ modelo 1: $u = 1{,}74 + 0{,}2x$.



Qualidade do modelo 1

- Qualidade do ajuste

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

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

- Resultados do ajuste pelo modelo 1

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 interpolador linear.
- Reta traçada por dois pontos quaisquer.
- Pontos escolhidos não pertencentes ao diagrama de dispersão.
- Por exemplo, escolhendo os pontos

x	2	6
y	2	3

.

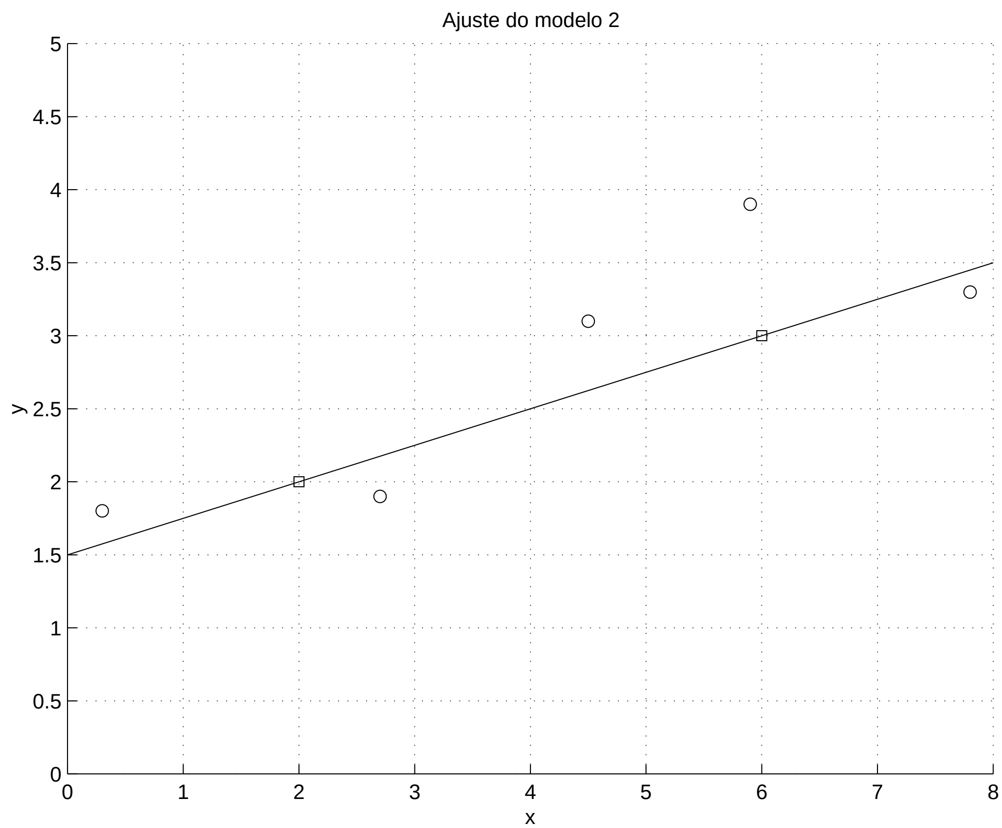
- Equação da reta $u(x)$

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

$$2 + 0{,}25(x-2) \leadsto u(x) = 1{,}5 + 0{,}25x.$$

Gráfico do modelo 2

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



Qualidade do modelo 2

- Resultados do ajuste pelo 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 dentre aqueles 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

$$D(\beta_0,\beta_1)=\sum_{i=1}^{n}(y_i-u_i)^2=\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)^2.$$

- Derivadas parciais

$$\frac{\partial D(\beta_0,\beta_1)}{\partial\beta_0}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i),$$

$$\frac{\partial D(\beta_0,\beta_1)}{\partial\beta_1}=-2\sum_{i=1}^{n}(y_i-\beta_0-\beta_1x_i)x_i.$$

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

$$-2\sum_{i=1}^{n}(y_i-b_0-b_1x_i)=0\rightarrow\sum_{i=1}^{n}b_0+\sum_{i=1}^{n}b_1x_i=\sum_{i=1}^{n}y_i,$$

$$-2\sum_{i=1}^{n}(y_i-b_0-b_1x_i)x_i=0\rightarrow\sum_{i=1}^{n}b_0x_i+\sum_{i=1}^{n}b_1x_i^2=\sum_{i=1}^{n}x_iy_i.$$

Reta de quadrados mínimos

- Na forma matricial e simplificando a notação

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

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

$$u(x) = b_0 + b_1 x,$$

$$\boxed{b_1 = \frac{\sum x_i \sum y_i - n \sum x_i y_i}{(\sum x_i)^2 - n \sum x_i^2}},$$

$$\boxed{b_0 = \frac{\sum y_i - b_1 \sum x_i}{n}}.$$

Exemplo de quadrados mínimos

- Calcular a reta de quadrados mínimos usando

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_iy_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

- Solução de quadrados mínimos

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

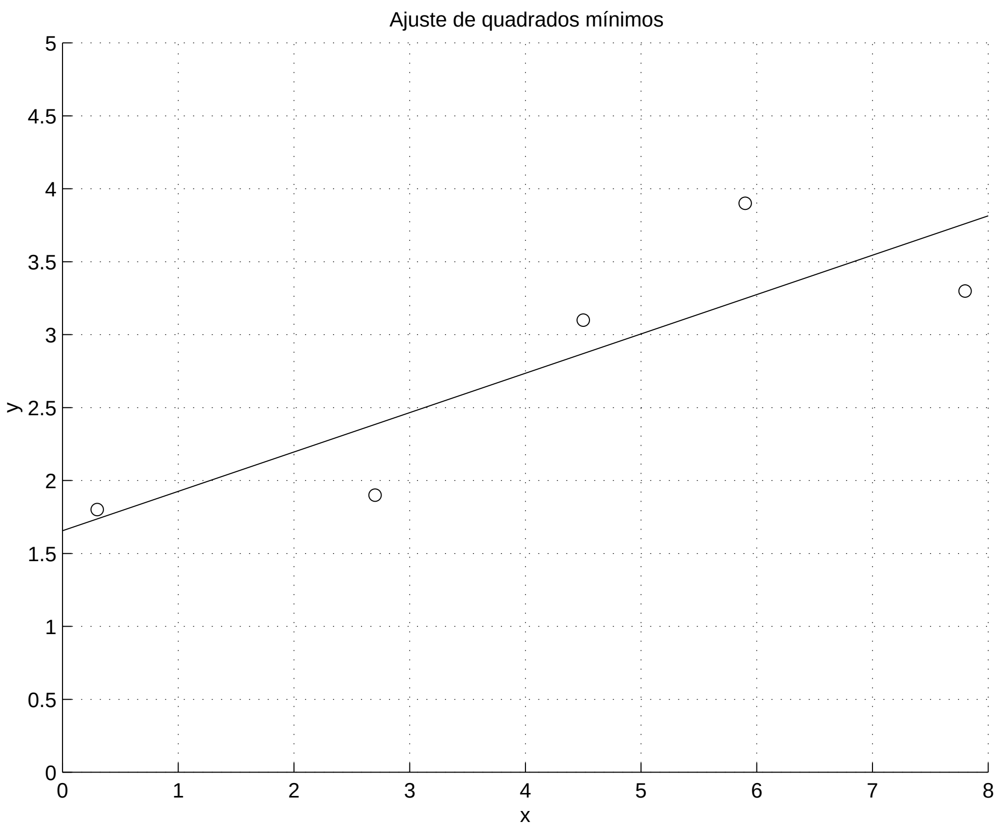
$$\rightsquigarrow b_1 = 0{,}2698;$$

$$b_0 = \frac{\sum y_i - b_1 \sum x_i}{n} = \frac{14{,}0 - 0{,}2698 \cdot 21{,}2}{5}$$

$$\rightsquigarrow b_0 = 1{,}6560.$$

Reta de quadrados mínimos

- Reta $u = 1{,}6560 + 0{,}2698x$



Qualidade do modelo

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

- Tomando o quadrado em ambos os termos

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

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

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

- Pode-se mostrar que

$$\sum_{i=1}^{n}(y_i - u_i)(u_i - \bar{y}) = 0.$$

- Conseqüentemente

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

Cálculo de r^2

- Soma dos quadrados

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

- 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)

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

- Qualidade do ajuste do modelo 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, $0\leq r^2\leq 1$.

Cálculo de r^2 cont.

❑ Considerando

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

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

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

❑ Coeficiente de determinação

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

❑ 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

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

- $D(b_0, b_1)$: somatório dos desvios, 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 irão diminuir se forem introduzidos mais parâmetros no modelo.
- Redução global de σ^2 define se mais parâmetros devem ou não ser incorporados ao modelo.

Exemplo

- 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

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

.

- Cálculo dos parâmetros

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

$$b_0=\frac{\sum y_i-b_1\sum x_i}{n}=\frac{102{,}0-1{,}7918\cdot 42{,}4}{8}\leadsto b_0=3{,}2535;$$

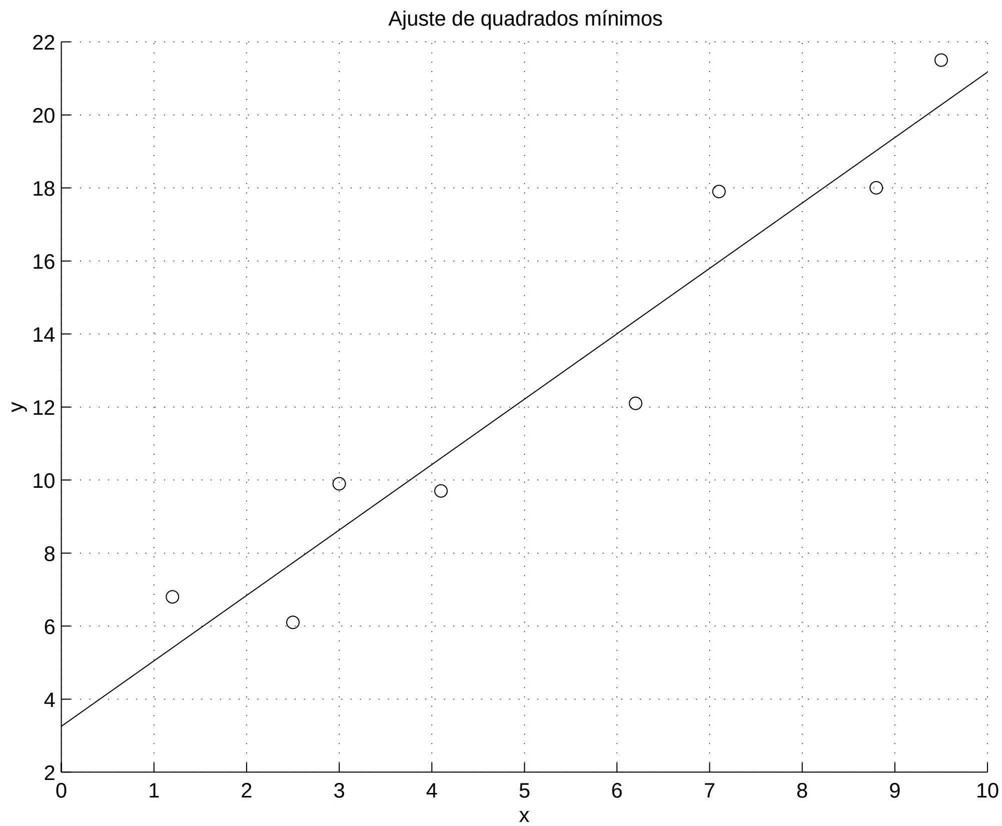
$$r^2=1-\frac{D(b_0,b_1)}{\sum y_i^2-\frac{1}{n}(\sum y_i)^2}=1-\frac{18{,}4085}{1528{,}62–(102{,}0)^2/8}\leadsto r^2=0{,}9193;$$

$$\sigma^2=\frac{18{,}4085}{8-2}\leadsto\sigma^2=3{,}0681.$$

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

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_p x_p + \epsilon,$$

- β_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

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

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

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

Método dos quadrados mínimos

- Derivadas parciais de D

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_0} =$$

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

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_1} =$$

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

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_2} =$$

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

$$\vdots$$

$$\frac{\partial D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)}{\partial \beta_p} =$$

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

Mínimo de $D(\beta_0, \beta_1, \beta_2, \ldots, \beta_p)$

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

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

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - \ldots - b_px_{ip}) = 0 \leadsto$$

$$\sum_{i=1}^{n} b_0 + \sum_{i=1}^{n} b_1x_{i1} + \sum_{i=1}^{n} b_2x_{i2} + \ldots + \sum_{i=1}^{n} b_px_{ip} = \sum_{i=1}^{n} y_i,$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - \ldots - b_px_{ip})x_{i1} = 0 \leadsto$$

$$\sum_{i=1}^{n} b_0x_{i1} + \sum_{i=1}^{n} b_1x_{i1}x_{i1} + \sum_{i=1}^{n} b_2x_{i2}x_{i1} + \ldots + \sum_{i=1}^{n} b_px_{ip}x_{i1} = \sum_{i=1}^{n} x_{i1}y_i,$$

$$\vdots$$

$$-2\sum_{i=1}^{n}(y_i - b_0 - b_1x_{i1} - b_2x_{i2} - \ldots - b_px_{ip})x_{ip} = 0 \leadsto$$

$$\sum_{i=1}^{n} b_0x_{ip} + \sum_{i=1}^{n} b_1x_{i1}x_{ip} + \sum_{i=1}^{n} b_2x_{i2}x_{ip} + \ldots + \sum_{i=1}^{n} b_px_{ip}x_{ip} = \sum_{i=1}^{n} x_{ip}y_i.$$

Equações normais

- Equações normais

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

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

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

- Coeficiente de determinação

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

- Variância residual

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

Regressão polinomial

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

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

- Equações normais

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

Algoritmo: regressão linear múltipla

```
Algoritmo Regressão_múltipla
{ Objetivo: Calcular parâmetros de quadrados mínimos }
parâmetros de entrada n, v, p, x, y
parâmetros de saída b, r2, sigma2
  se v > 1 e v + 1 ≠ p então, escreva modelo inválido, abandone, fim se
  para i ← 1 até n faça  { inclusão de uma coluna de 1's relativa à b0 }
    para j ← v + 1 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 reg. polinomial gera potências de x }
    para j ← 2 até p − 1 faça
      para i ← 1 até n faça, x(i,j + 1) ← 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 ← 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 = L^T }
  fim para
  b ← Substituições_Retroativas(p,U,t)  { coeficientes }
  D ← 0; Sy2 ← 0
  para i ← 1 até n faça
    Soma ← 0
    para j ← 1 até p faça
      Soma ← Soma + b(j) ∗ x(i,j)
    fim para
    u(i) ← Soma; d(i) ← y(i) − u(i); D ← D + d(i)^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
```

Exemplo

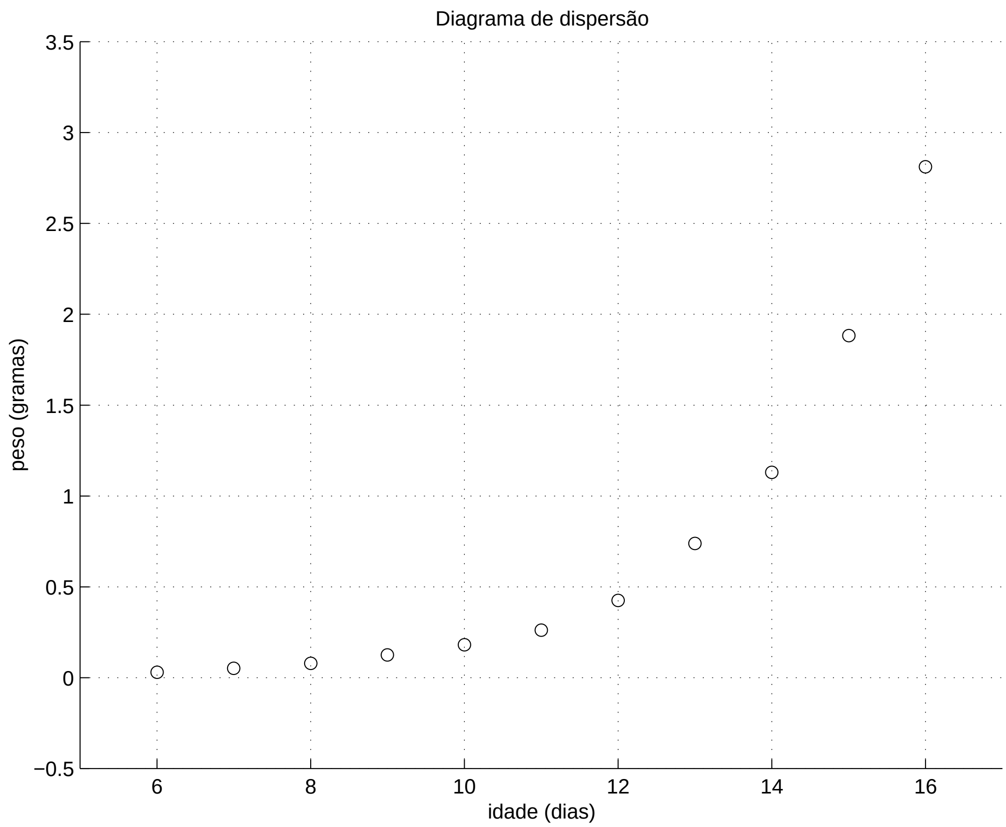
- Dados relacionando o peso y de embriões de frangos desidratados, em gramas, com a sua idade x, em dias.

i	x_i	y_i
1	6	0,029
2	7	0,052
3	8	0,079
4	9	0,125
5	10	0,181
6	11	0,261
7	12	0,425
8	13	0,738
9	14	1,130
10	15	1,882
11	16	2,812

Diagrama de dispersão

- O ajuste não deve ser feito por um polinômio de grau 1.

- Usar um polinômio de grau mais elevado.



Resultados

- Valores do coeficiente de determinação r^2 e da variância residual σ^2 para o modelo polinomial

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

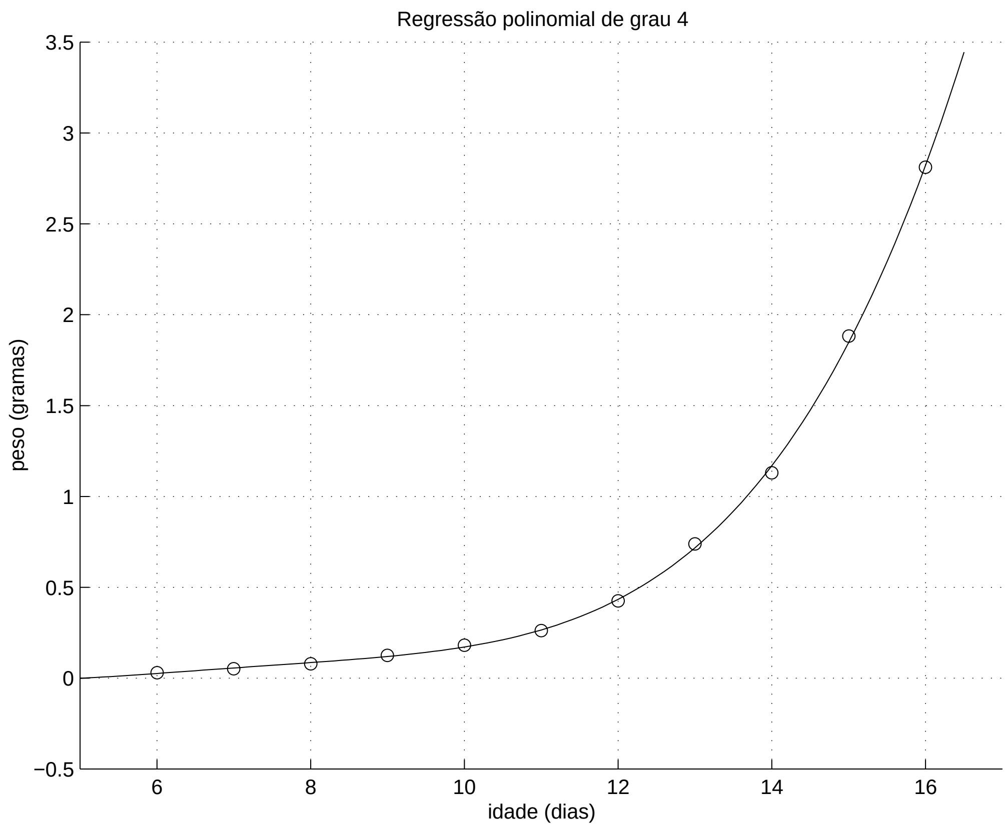
g	r^2	σ^2
1	0,74418	2,32178×10^{-1}
2	0,96984	3,07961×10^{-2}
3	0,99883	1,36642×10^{-3}
4	0,99957	5,86451×10^{-4}
5	0,99962	6,21028×10^{-4}
6	0,99966	7,04092×10^{-4}

.

- r^2 aumenta quando o grau do polinômio de quadrados mínimos é aumentado.
- σ^2 apresenta o menor valor para o grau $g = 4$.
- Este deve ser o grau escolhido para o ajuste polinomial.

Polinômio de regressão

- Polinômio de grau 4 traçado no diagrama de dispersão.



Transformações não lineares

- Modelos não lineares nos parâmetros podem ser transformados em modelos lineares.

- Faz-se uma simples substituição de variáveis por funções dessas variáveis.

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

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

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

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

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

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

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

Malcondicionamento

- 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 via equações normais.
- Coeficiente de determinação r^2 e número de condição espectral $\kappa_2(X^T X)$ (embriões de frango)

g	r^2	$\kappa_2(X^T X)$
1	0,74418	1,74040×10³
2	0,96984	3,93510×10⁶
3	0,99883	1,15846×10¹⁰
4	0,99957	4,12715×10¹³
5	0,99962	1,75113×10¹⁷
6	0,99966	4,83023×10¹⁹
7	0,99972	3,30131×10²¹

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

Formas alternativas de estimar os parâmetros

- Modelo de regressão linear múltipla

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

- 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 à β_0,
- β: vetor $((p+1) \times 1)$ dos parâmetros a serem estimados e
- ϵ: vetor $(n \times 1)$ dos erros aleatórios

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

Estimativa do vetor β

- Minimizar a função

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

- Pelas regras de diferenciação matricial

$$\begin{aligned}
\frac{\partial f(\beta)}{\partial \beta^T} &= \frac{\partial f(\beta)}{\partial (y - X\beta)^T} \frac{\partial (y - X\beta)}{\partial \beta^T}, \\
&= 2(y - X\beta)^T(-X) = -2(y - X\beta)^T X.
\end{aligned}$$

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

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

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

- Equações normais na forma matricial

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

Equações normais

- Matriz $X^T X$ tem elementos reais, é não singular, e é definida positiva.
- O ponto $f(b)$ é, de fato, um mínimo de $f(\beta)$.
- Equações normais formam um sistema malcondicionado.
- Processos alternativos para a estimativa de β que evitam a formação da matriz $X^T X$.

Decomposição QR

- O vetor b deve minimizar a soma de quadrados residual

$$\|y - Xb\|_2^2.$$

- Decomposição QR da matriz X $(n \times (p+1))$

$$X = QR,$$

- Q: matriz ortogonal $(n \times n)$ e
- R: matriz triangular superior $(n \times (p+1))$ da forma

$$R = \begin{bmatrix} R_1 \\ 0 \end{bmatrix},$$

- R_1: matriz quadrada triangular superior de ordem $p+1$.

Estimativa dos parâmetros

- Soma de quadrados residual

$$\|y - Xb\|_2^2 = \|y - QRb\|_2^2.$$

- Como Q é uma matriz ortogonal, $Q^TQ = I$

$$\|y - Xb\|_2^2 = \|Q^Ty - Q^TQRb\|_2^2 \rightsquigarrow$$

$$\|y - Xb\|_2^2 = \|Q^Ty - Rb\|_2^2.$$

- Definindo

$$Q^Ty = c = \begin{bmatrix} c_1 \\ c_2 \end{bmatrix},$$

- c_1: vetor $(p+1)$ e c_2: vetor $(n-p-1)$

$$Rb = \begin{bmatrix} R_1 \\ 0 \end{bmatrix} b = \begin{bmatrix} R_1 b \\ 0 \end{bmatrix},$$

$$\|y-Xb\|_2^2 = \left\| \begin{bmatrix} c_1 \\ c_2 \end{bmatrix} - \begin{bmatrix} R_1 b \\ 0 \end{bmatrix} \right\|_2^2 = \|c_1 - R_1 b\|_2^2 + \|c_2\|_2^2.$$

Estimativa dos parâmetros cont.

- Soma de quadrados residual será mínima quando

$$\boxed{R_1 b - c_1 = 0},$$

- b for a solução do sistema triangular superior

$$R_1 b = c_1.$$

- Soma de quadrados residual SQRes

$$D(b_0, b_1, \ldots, b_p) = \|c_2\|_2^2 = c_2^T c_2.$$

- Valores preditos

$$u = Xb = QRb = Q \begin{bmatrix} R_1 b \\ 0 \end{bmatrix} \rightsquigarrow \boxed{u = Q \begin{bmatrix} c_1 \\ 0 \end{bmatrix}}.$$

- Vetor dos desvios $d = y - Xb$

$$d = Qc - QRb = Q(c - Rb) = Q \begin{bmatrix} c_1 - R_1 b \\ c_2 - 0b \end{bmatrix},$$

$$\boxed{d = Q \begin{bmatrix} 0 \\ c_2 \end{bmatrix}}.$$

Decomposição do valor singular

- Decomposição do valor singular de uma matriz X $(n \times (p+1))$

$$X = USV^T,$$

- 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 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

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

- Sendo

$$U^T y = a = \begin{bmatrix} a_1 \\ a_2 \end{bmatrix}, \tilde{b} = V^T b$$

- a_1: vetor $(p+1)$ e a_2: vetor $(n-p-1)$

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

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

$$\|y - Xb\|_2^2 = \|a_1 - S_1\tilde{b}\|_2^2 + \|a_2\|_2^2.$$

Estimativa dos parâmetros cont.

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

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

- Pela ortogonalidade de V

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

- Soma de quadrados residual

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

- Valores preditos

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

- Vetor desvio $d = y - Xb$

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

Comparação dos métodos para RLM

- Equações normais: vantagens
  - Maior velocidade com que podem ser formadas e resolvidas.
  - Com 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^TX é o quadrado daquele da matriz X.
  - Difícil computar X^TX e X^Ty, exatamente.
  - Perturbações feitas no problema básico podem ter conseqüências desastrosas.

- Métodos de ortogonalização: vantagens
  - Superiores propriedades numéricas.
  - Grande quantidade de memória disponível a custo baixo.

- Métodos de ortogonalização: 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

$$P_{n-1}(x) = a_0 + a_1 x + a_2 x^2 + \ldots + a_{n-1} x^{n-1}.$$

- Possui n coeficientes a_i, $i = 0, 1, \ldots, n-1$.
- O número de pontos utilizados para gerar o polinômio interpolador é igual ao número de coeficientes do polinômio.
- Polinômio de regressão de grau g, usando n pontos

$$U_g(x) = b_0 + b_1 x + b_2 x^2 + \ldots + b_g x^g,$$

- sendo $g \leq n-1$.
- 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

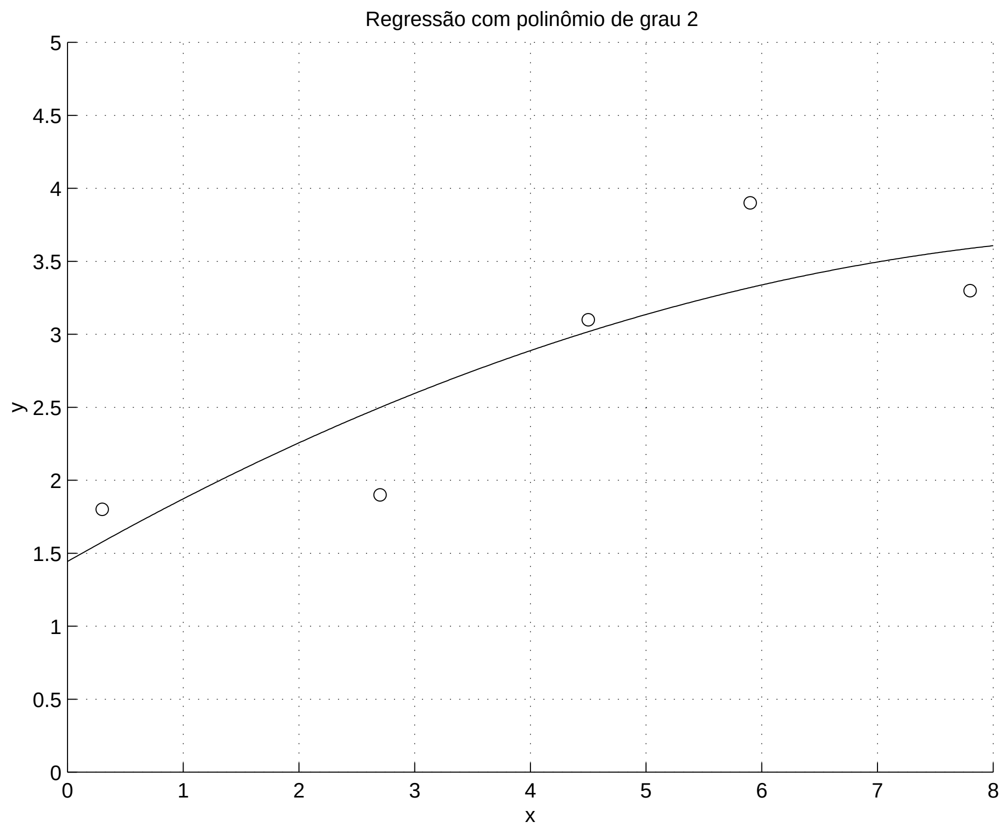
$$\begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \end{bmatrix} = \begin{bmatrix} 1 & 1 \\ x_1 & x_2 \end{bmatrix} \begin{bmatrix} y_1 \\ y_2 \end{bmatrix} \leadsto$$

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

- O sistema linear é idêntico às equações normais, para $n = 2$, utilizadas para calcular os parâmetros de uma regressão linear simples.

Regressão polinomial quadrática

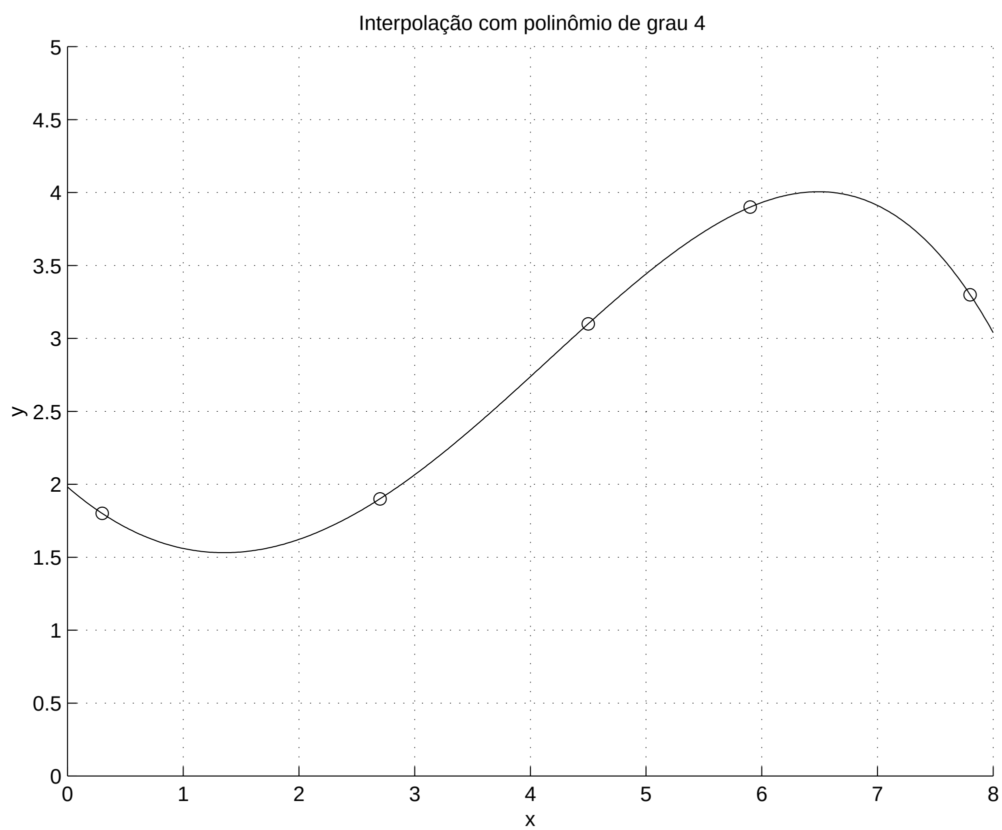
- Polinômio de regressão de grau $g = 2$ com $n = 5$ pontos.



Regressão idêntica à interpolação

- Quando o polinômio de regressão possuir grau $g = n - 1 = 4$ ele se torna idêntico a um polinômio interpolador de mesmo grau.

- O polinômio passa por todos os pontos do diagrama de dispersão.



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 tende ao infinito à medida que o número de parâmetros p do modelo se aproxima do número de pontos n

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