Capítulo 9 - Introdução à regressão linear simples

9.1 Modelos de regressão

São modelos utilizados para compreender a relação entre uma variável dependente (Y) e uma ou mais variáveis independentes (X ou $X_1, X_2, \ldots$)

Exemplos:

1) Y - pressão atmosférica (variável dependente ou resposta)

x - altitude (variável independente ou explicativa)

Modelo determinístico:

$$Y = \beta_0 + \beta_1 x$$

Modelos de regressão linear simples:

$$E(Y|x) = \beta_0 + \beta_1 x$$

ou

$$Y = \beta_0 + \beta_1 x + \varepsilon$$

onde Y - variável aleatória

x - valores fixos (medições sem erro ou com erro desprezável)

ε - variável aleatória tal que $E(\varepsilon) = 0$

2) Y - pressão atmosférica

x_1 - altitude

x_2 - temperatura

Modelo de regressão linear múltipla (não faz parte do programa):

$$Y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \varepsilon$$

O termo <u>linear</u> refere-se a linearidade nos parâmetros (β_0, β_1, etc.), assim, os modelos da mesma forma que os anteriores mas onde em vez de x e y figuram funções destas variáveis também são considerados modelos de regressão linear. Por exemplo

$$\log Y = \beta_0 + \beta_1 \log x + \varepsilon$$

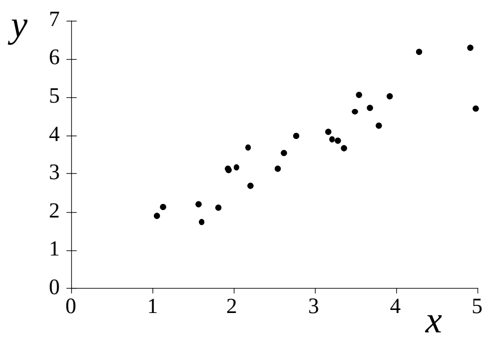
ou

$$Y = \beta_0 + \beta_1 x^2 + \varepsilon$$

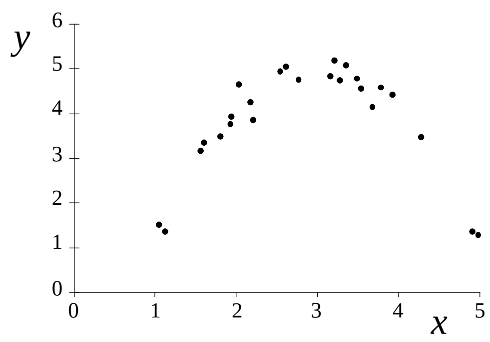
Um problema de regressão linear simples começa pela recolha de uma amostra de pares de pontos

$$(x_i, y_i), \qquad i = 1, \ldots, n$$

O primeiro passo é sempre a representação gráfica destes pontos (podemos verificar aqui se a relação entre x e y parece ser bem modelada por uma recta ou se há necessidade de alguma transformação):



Modelo $Y = \beta_0 + \beta_1 x + \varepsilon$ parece adequado



Modelo $Y = \beta_0 + \beta_1 x + \varepsilon$ não parece adequado

Por vezes há considerações de natureza teórica que ajudam a formular o modelo.

Exemplo (EAA): Experiência realizada com o objectivo de determinar o doseamento do ferro em duas amostras de cereal Kellogg's Special K através da técnica de espectrofotometria de absorção atómica (EAA). Esta técnica assenta na chamada lei de Lambert-Beer, que relaciona a absorvância (*A*) de uma dada espécie, com a sua concentração (*C*) através da expressão

$$A = abC$$

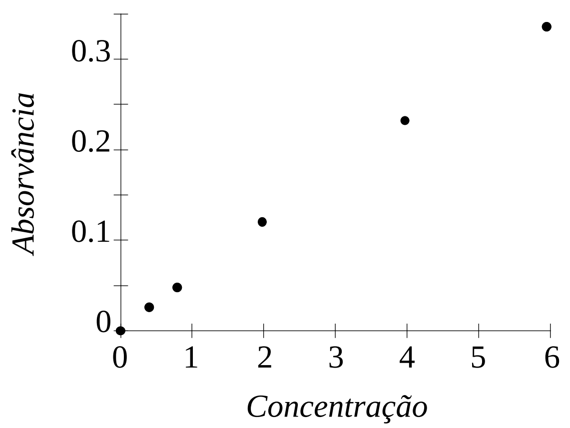
onde *a* é a absortividade molecular (constante para cada espécie de solvente) e *b* a espessura do meio absorvente (no caso uma chama). Para medir a absorvância é necessário medir a intensidade da

radiação que atinge a amostra (I_0) e a que dela emerge (*I*) e usar a expressão $A = -\log\left(I/I_0\right)$

Como as constantes *a* e *b* são na maioria dos casos desconhecidas ou mal determinadas torna-se necessário determinar experimentalmente a relação entre *A* e *C* (usando a indicação fornecida pelo modelo teórico de que se trata de uma relação linear). Para isso mediu-se a absorvância para seis soluções com concentrações conhecidas de Fe^{3+}:

Solução	Concentração (ppm)	Absorvância
Branco	0	0
I	0.40	0.026
II	0.79	0.048
III	1.98	0.120
IV	3.97	0.232
V	5.95	0.336

Representação gráfica:



O gráfico também indica que um modelo do tipo

$$A = \beta_0 + \beta_1 C + \varepsilon$$

parece adequado.

9.2 Regressão linear simples (estimação de parâmetros)

<u>Modelo</u>: $$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1,\ldots,n$$

onde ε_i são v.a. <u>não correlacionadas</u> tais que

$$E\left(\varepsilon_i\right) = 0$$

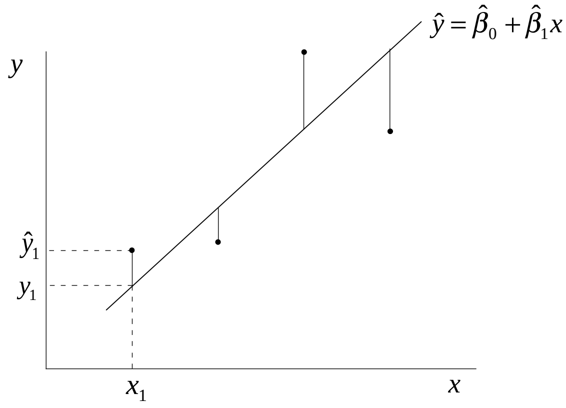
e

$$V\left(\varepsilon_i\right) = \sigma^2$$

Como estimar os parâmetros β_0 e β_1?

Método dos mínimos quadrados:

Encontrar os valores $\hat{\beta}_0$ e $\hat{\beta}_1$ que minimizam a soma dos quadrados dos desvios verticais entre os pontos observados e a recta $\hat{y} = \hat{\beta}_0 + \hat{\beta}_1 x$:



Para cada valor de x_i observado o ponto sobre a recta tem ordenada $\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$

Seja

$$Q = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2 = \sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2$$

Pretende-se então determinar $\hat{\beta}_0$ e $\hat{\beta}_1$ que minimizam Q.

$$\begin{cases} \dfrac{\partial Q}{\partial \beta_0} = -2\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) = 0 \\ \dfrac{\partial Q}{\partial \beta_1} = -2\sum_{i=1}^{n} \left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right) x_i = 0 \end{cases}$$

$$\Leftrightarrow \begin{cases} n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^{n} x_i = \sum_{i=1}^{n} y_i \\ \hat{\beta}_0 \sum_{i=1}^{n} x_i + \hat{\beta}_1 \sum_{i=1}^{n} x_i^2 = \sum_{i=1}^{n} x_i y_i \end{cases}$$

trata-se de um sistema de duas equações lineares a duas incógnitas que tem como solução as **estimativas dos mínimos quadrados:**

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x} \quad \text{e} \quad \hat{\beta}_1 = \frac{\sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y}}{\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2}$$

onde $\bar{y} = \dfrac{\sum_{i=1}^{n} y_i}{n}$ e $\bar{x} = \dfrac{\sum_{i=1}^{n} x_i}{n}$.

<u>Notas</u>:

1) Pode-se mostrar, pelo determinante da matriz Hessiana, que se existir solução ela corresponde sempre a um mínimo.

2) Existe solução se e só se $\sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 \neq 0$, ou seja, se na amostra existirem pelo menos dois valores distintos de x.

3) Às diferenças $e_i = y_i - \hat{y}_i$, $i = 1, \ldots, n$, chamam-se resíduos. A análise dos resíduos permite avaliar se o modelo assumido é adequado (ver secção 9.7).

4) Notações alternativas:

$$S_{XX} = \sum_{i=1}^{n} x_i^2 - n\,\bar{x}^2 = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)^2$$

$$S_{XY} = \sum_{i=1}^{n} x_i y_i - n\,\bar{x}\,\bar{y} = \sum_{i=1}^{n} \left(x_i - \bar{x}\right)\left(y_i - \bar{y}\right) =$$

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

$$\hat{\beta}_1 = \frac{S_{XY}}{S_{XX}}$$

Exemplo (EAA) (cont.):

<u>Modelo</u>: $A_i = \beta_0 + \beta_1 C_i + \varepsilon_i, \quad i = 1, \ldots, 6,$

$V(\varepsilon_i) = \sigma^2 \quad E(\varepsilon_i) = 0 \quad \mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \ (i \neq j)$

($x_i \equiv C_i \qquad y_i \equiv A_i$)

$$\sum_{i=1}^{6} x_i = \sum_{i=1}^{6} C_i = 13.09$$

$$\sum_{i=1}^{6} y_i = \sum_{i=1}^{6} A_i = 0.762$$

$$\sum_{i=1}^{6} x_i^2 = \sum_{i=1}^{6} C_i^2 = 55.8679$$

$$\sum_{i=1}^{6} x_i y_i = \sum_{i=1}^{6} C_i A_i = 3.20616$$

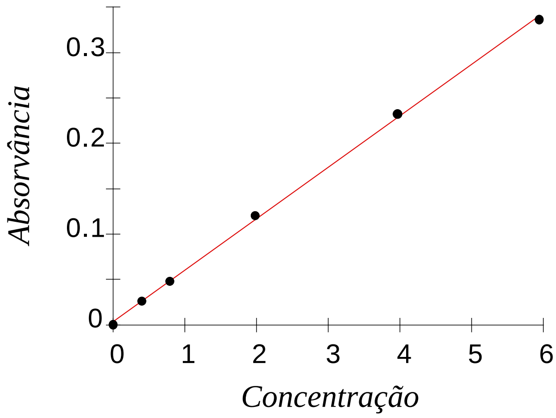
$$\hat{\beta}_1 = \frac{3.20616 - 6 \times \frac{13.09}{6} \times \frac{0.762}{6}}{55.8679 - 6 \times \left(\frac{13.09}{6}\right)^2} = 0.0565$$

$$\hat{\beta}_0 = \frac{0.762}{6} - 0.056526422 \times \frac{13.09}{6} = 0.0037$$

Recta estimada: $\hat{A} = 0.0037 + 0.0565C$

<u>Nota</u>: nos cálculos efectuados "à mão" devem usar-se, nos passos intermédios, o máximo de algarismos possíveis, para evitar a acumulação de erros de arredondamento. Só na apresentação de resultados é que se fazem arredondamentos.

Representação gráfica dos pontos observados e da recta estimada (ou ajustada):



9.3 Propriedades dos estimadores dos mínimos quadrados e estimação de σ^2

Como se viu para cada x_i, Y_i é uma v.a., da qual dispomos de um valor observado y_i. $\hat{\beta}_0$ e $\hat{\beta}_1$ podem também ser vistos como variáveis aleatórias (os estimadores). Vamos verificar que são estimadores centrados dos parâmetros β_0 e β_1, e ainda calcular as suas variâncias.

$$E\left(\hat{\beta}_1\right) = E\left(\frac{S_{XY}}{S_{XX}}\right) = \frac{1}{S_{XX}} E\left[\sum_{i=1}^{n} Y_i (x_i - \bar{x})\right] =$$

$$= \frac{1}{S_{XX}} E\left[\sum_{i=1}^{n} (x_i - \bar{x})(\beta_0 + \beta_1 x_i + \varepsilon_i)\right] =$$

$$= \frac{1}{S_{XX}} \left[\beta_0 \sum_{i=1}^{n} (x_i - \bar{x}) + \beta_1 \sum_{i=1}^{n} x_i (x_i - \bar{x}) + \right.$$
$$\left. + \sum_{i=1}^{n} (x_i - \bar{x}) E(\varepsilon_i)\right] = \frac{\beta_1}{S_{XX}} \left(\sum_{i=1}^{n} x_i^2 - n\bar{x}^2\right) = \beta_1$$

$$V\left(\hat{\beta}_1\right) = V\left(\frac{S_{XY}}{S_{XX}}\right) = \frac{1}{S_{XX}^2} V\left[\sum_{i=1}^{n} Y_i (x_i - \bar{x})\right]$$

$$V(\varepsilon_i) = \sigma^2 \Rightarrow V(Y_i) = \sigma^2$$

$$\mathrm{Cov}(\varepsilon_i, \varepsilon_j) = 0 \ (i \neq j) \Rightarrow \mathrm{Cov}(Y_i, Y_j) = 0 \ (i \neq j)$$

logo

$$V\left(\hat{\beta}_1\right) = \frac{1}{S_{XX}^2} \sigma^2 \sum_{i=1}^{n} (x_i - \bar{x})^2 = \frac{\sigma^2}{S_{XX}}$$

Do mesmo modo se poderia mostrar que

$$E\left(\hat{\beta}_0\right) = \beta_0 \quad \text{e} \quad V\left(\hat{\beta}_0\right) = \sigma^2 \left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right)$$

e que $$\mathrm{Cov}\left(\hat{\beta}_0, \hat{\beta}_1\right) = -\frac{\sigma^2 \bar{x}}{S_{XX}}$$

(para isso era necessário primeiro mostrar que $\mathrm{Cov}\left(\bar{Y}, \hat{\beta}_1\right) = 0$)

Para estimar as variâncias associadas às estimativas $\hat{\beta}_0$ e $\hat{\beta}_1$ torna-se necessário estimar σ^2 (variância do erro aleatório), isso é feito usando os resíduos.

Seja $$SS_E = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} \left(y_i - \hat{y}_i\right)^2$$

Pode mostrar-se que $E\left(SS_E\right) = (n-2)\sigma^2$

logo
$$\hat{\sigma}^2 = \frac{SS_E}{n-2}$$

é um estimador centrado de σ^2.

Tem-se então que o <u>erro padrão</u> (desvio padrão) <u>estimado</u> de $\hat{\beta}_1$ é

$$se\left(\hat{\beta}_1\right) = \sqrt{\frac{\hat{\sigma}^2}{S_{XX}}}$$

e que o <u>erro padrão estimado</u> de $\hat{\beta}_0$ é

$$se\left(\hat{\beta}_0\right) = \sqrt{\hat{\sigma}^2\left(\frac{1}{n} + \frac{\bar{x}^2}{S_{XX}}\right)}$$

Nota: Outras formas de escrever SS_E:

$$SS_E = \sum_{i=1}^{n}\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i\right)^2 = \sum_{i=1}^{n}\left(y_i - \bar{y} + \hat{\beta}_1\bar{x} - \hat{\beta}_1 x_i\right)^2 =$$

$$= \sum_{i=1}^{n}\left[\left(y_i - \bar{y}\right) - \hat{\beta}_1\left(x_i - \bar{x}\right)\right]^2 =$$

$$= S_{YY} + \frac{S_{XY}^2}{S_{XX}^2} S_{XX} - 2\frac{S_{XY}}{S_{XX}} S_{XY} = S_{YY} - \frac{S_{XY}}{S_{XX}} S_{XY} =$$

$$= S_{YY} - \hat{\beta}_1 S_{XY} = S_{YY} - \hat{\beta}_1^2 S_{XX}$$

Exemplo (EAA) (cont.):

$$S_{YY} = \sum_{i=1}^{n} y_i^2 - n\bar{y}^2 = 0.1841 - 6\times\left(\frac{0.762}{6}\right)^2 = 0.0873$$

$$S_{XX} = 55.8679 - 6\times\left(\frac{13.09}{6}\right)^2 = 27.3099$$

$$SS_E = 0.0873 - 0.0565^2 \times 27.3099 = 6.4466\times 10^{-5}$$

$$\hat{\sigma} = \sqrt{\frac{SS_E}{4}} = 4.0145\times 10^{-3}$$

$$se\left(\hat{\beta}_1\right) = \hat{\sigma}\sqrt{\frac{1}{S_{XX}}} = 7.682\times 10^{-4}$$

$$se\left(\hat{\beta}_0\right) = \hat{\sigma}\sqrt{\frac{1}{6} + \frac{\bar{x}^2}{S_{XX}}} = 2.344\times 10^{-3}$$

9.4 Alguns abusos do modelo de regressão

- Escolha incorrecta das variáveis explicativas pode levar a conclusões absurdas (uma associação entre duas variáveis não implica necessariamente uma relação de causa-efeito).
- As relações extraídas dos modelos de regressão só são válidas dentro do domínio dos dados originais. **Não se devem fazer extrapolações**.

9.5 Testes de hipóteses em regressão linear simples

Para fazer testes de hipóteses em relação aos parâmetros β_0 e β_1, é necessário admitir que os erros têm distribuição normal (esta hipótese de trabalho deve ser verificada *a posteriori* por análise dos resíduos).

Assim, supondo que $\varepsilon_i \underset{iid}{\sim} N\left(0;\sigma^2\right)$ conclui-se imediatamente que

$$Y_i \sim N\left(\beta_0 + \beta_1 x_i;\sigma^2\right)$$

e sendo os estimadores $\hat{\beta}_0$ e $\hat{\beta}_1$ combinações lineares das v.a. Y_i, terão também distribuições normais, ou seja;

$$\hat{\beta}_0 \sim N\left(\beta_0;\sigma^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}\right)\right)$$

e

$$\hat{\beta}_1 \sim N\left(\beta_1;\frac{\sigma^2}{S_{XX}}\right)$$

Pode ainda mostrar-se que $\hat{\beta}_0$ e $\hat{\beta}_1$ são independentes de $\hat{\sigma}^2$ e que

$$\frac{(n-2)\hat{\sigma}^2}{\sigma^2} = \frac{SS_E}{\sigma^2} \sim \chi^2_{n-2}$$

pelo que

$$\frac{\hat{\beta}_0 - \beta_0}{\sqrt{\hat{\sigma}^2\left(\frac{1}{n}+\frac{\bar{x}^2}{S_{XX}}\right)}} = \frac{\hat{\beta}_0 - \beta_0}{se\left(\hat{\beta}_0\right)} \sim t_{n-2}$$

e

$$\frac{\hat{\beta}_1 - \beta_1}{\sqrt{\frac{\hat{\sigma}^2}{S_{XX}}}} = \frac{\hat{\beta}_1 - \beta_1}{se\left(\hat{\beta}_1\right)} \sim t_{n-2}$$

Então, para testar $H_0\colon\ \beta_1 = \beta_{1,0}$ contra $H_1\colon\ \beta_1 \neq \beta_{1,0}$ usa-se a estatística

$$T_0 = \frac{\hat{\beta}_1 - \beta_{1,0}}{se\left(\hat{\beta}_1\right)} \underset{sob\ H_0}{\sim} t_{n-2}$$

e rejeita-se H_0 ao nível de significância α se

$$|t_0| > a, \quad \text{onde} \quad a\colon\ P\left(T_{n-2} > a\right) = \alpha/2.$$

O procedimento em relação ao parâmetro β_0 é similar.

Um teste de especial importância (por vezes chamado <u>teste à significância da regressão</u>) é o seguinte:

$$H_0\colon\ \beta_1 = 0 \quad \text{contra} \quad H_1\colon\ \beta_1 \neq 0$$

Não rejeitar H_0 ("aceitar" H_0) é equivalente a concluir que não há associação linear entre *x* e *y* (ou entre as funções de *x* e *y* em causa), ou seja, não há associação ou a associação não é linear.

Exemplo (EAA) (cont.):

$$se\left(\hat{\beta}_1\right) = \hat{\sigma}\sqrt{\frac{1}{S_{XX}}} = 7.682\times 10^{-4}$$

$$se\left(\hat{\beta}_0\right) = \hat{\sigma}\sqrt{\frac{1}{6}+\frac{\bar{x}^2}{S_{XX}}} = 2.344\times 10^{-3}$$

Teste de $H_0\colon\ \beta_1 = 0$ contra $H_1\colon\ \beta_1 \neq 0$

$$t_0 = \frac{0.0565 - 0}{7.682\times 10^{-4}} = 73.58$$

Para um nível de significância de 5% e $n-2=4$, obtém-se $t_{4,0.975} = 2.776$.

Como 73.58 > 2.776 rejeita-se H_0.

Teste de $H_0\colon\ \beta_0 = 0$ contra $H_1\colon\ \beta_0 \neq 0$

$$t_0 = \frac{0.0037 - 0}{2.344\times 10^{-3}} = 1.57$$

Como - 2.776 < 1.57 < 2.776 não se rejeita H_0, ao nível de significância de 5%, o que dado o modelo teórico era esperado.

9.6 Intervalos de confiança

Podem construir-se intervalos de confiança para β_0 e β_1 usando as distribuições deduzidas na secção anterior:

$$\frac{\hat{\beta}_0 - \beta_0}{se(\hat{\beta}_0)} \sim t_{n-2} \quad \text{e} \quad \frac{\hat{\beta}_1 - \beta_1}{se(\hat{\beta}_1)} \sim t_{n-2}$$

obtendo-se

$$I.C._{100\times(1-\alpha)\%}(\beta_0) = \left[\hat{\beta}_0 - a \times se(\hat{\beta}_0); \hat{\beta}_0 + a \times se(\hat{\beta}_0)\right]$$

$$I.C._{100\times(1-\alpha)\%}(\beta_1) = \left[\hat{\beta}_1 - a \times se(\hat{\beta}_1); \hat{\beta}_1 + a \times se(\hat{\beta}_1)\right]$$

onde a: $P(T_{n-2} > a) = \alpha/2$.

Exemplo (EAA) (cont.):

$\hat{\beta}_1 = 0.0565$ $\hat{\beta}_0 = 0.0037$

$se(\hat{\beta}_1) = 7.682 \times 10^{-4}$ $se(\hat{\beta}_0) = 2.344 \times 10^{-3}$

$a = t_{4,0.975} = 2.776$

$I.C._{95\%}(\beta_1) = [0.05439; 0.05865]$

$I.C._{95\%}(\beta_0) = [-0.00283; 0.01018]$

****

Por vezes é também necessário construir intervalos de confiança para

$$E(Y|x_0) = \mu_{Y|x_0} = \beta_0 + \beta_1 x_0$$

Um estimador centrado de $\mu_{Y|x_0}$ é

$$\hat{\mu}_{Y|x_0} = \hat{\beta}_0 + \hat{\beta}_1 x_0$$

e a sua variância é

$$V(\hat{\mu}_{Y|x_0}) = V(\hat{\beta}_0 + \hat{\beta}_1 x_0) =$$

$$= V(\hat{\beta}_0) + x_0^2 V(\hat{\beta}_1) + 2x_0 \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1) =$$

$$= \sigma^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}\right]$$

donde

$$se(\hat{\mu}_{Y|x_0}) = \sqrt{\hat{\sigma}^2 \left[\frac{1}{n} + \frac{(x_0 - \bar{x})^2}{S_{XX}}\right]}$$

e

$$\frac{\hat{\mu}_{Y|x_0} - (\beta_0 + \beta_1 x_0)}{se(\hat{\mu}_{Y|x_0})} \sim t_{n-2}$$

Então um I.C. a $100 \times (1-\alpha)\%$ para $\mu_{Y|x_0} = \beta_0 + \beta_1 x_0$ é

$$\left[\hat{\mu}_{Y|x_0} - a \times se(\hat{\mu}_{Y|x_0}); \hat{\mu}_{Y|x_0} + a \times se(\hat{\mu}_{Y|x_0})\right]$$

9.7 Coeficiente de determinação e análise empírica de resíduos

<u>Coeficiente de determinação</u>:

$$R^2 = 1 - \frac{SS_E}{S_{YY}} = 1 - \frac{\sum_{i=1}^{n}(y_i - \hat{y}_i)^2}{\sum_{i=1}^{n}(y_i - \bar{y})^2}$$

$\hat{y}_i = y_i$, $\forall_i \Leftrightarrow R^2 = 1$

(todos os pontos estão sobre a recta, o modelo de regressão explica toda a variabilidade observada em *y*). Modelo muito bom!

$\hat{\beta}_1 = 0 \Leftrightarrow \hat{\beta}_0 = \bar{y} \Leftrightarrow \hat{y}_i = \bar{y}$, $\forall_i \Leftrightarrow R^2 = 0$

(o modelo de regressão não explica nada da variabilidade observada em *y*). Modelo muito mau!

$$0 \le R^2 \le 1$$

Pode dizer-se que R^2 representa a percentagem da variabilidade de *y* que é explicada pela regressão.

Nota:

$$R^2 = 1 - \frac{\left(S_{YY} - \frac{S_{XY}^2}{S_{XX}}\right)}{S_{YY}} = \frac{S_{XY}^2}{S_{XX}S_{YY}} =$$

$$= \left[\frac{\sum_{i=1}^{n}(x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^{n}(x_i - \bar{x})^2 \sum_{i=1}^{n}(y_i - \bar{y})^2}} \right]^2$$

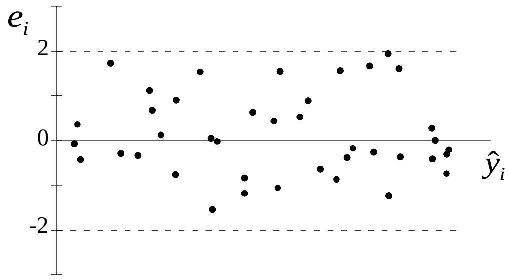
ou seja, o coeficiente de determinação coincide com o quadrado do coeficiente de correlação (empírico) entre *x* e *y*.

Exemplo (EAA) (cont.):

$$R^2 = 1 - 6.4466 \times 10^{-5} / 0.0873 = 0.9993 = 99.93\%$$

Análise empírica de resíduos:

- Gráfico $e_i = y_i - \hat{y}_i$ versus $\hat{y}_i$. Se as hipóteses de trabalho estiverem correctas deve ter o aspecto de uma mancha de pontos dispersos ao acaso numa faixa horizontal.



- O histograma dos resíduos deve ter um aspecto simétrico de forma semelhante à densidade da distribuição normal. (Se *n* for grande pode fazer-se um teste do χ^2).

Exemplo (EAA) (conclusão):

Voltemos então ao objectivo inicial. Com as seis observações de concentrações conhecidas obteve-se a chamada recta de calibração do aparelho:

$$\hat{A} = \hat{\beta}_0 + \hat{\beta}_1 C$$

com $\hat{\beta}_1 = 0.0565$ $\quad$ $\hat{\beta}_0 = 0.0037$

$se(\hat{\beta}_1) = 7.682 \times 10^{-4}$ $\quad$ $se(\hat{\beta}_0) = 2.344 \times 10^{-3}$

$\hat{\sigma} = 4.0145 \times 10^{-3}$

Vamos agora aplicar a relação obtida experimentalmente para estimar a concentração em ferro de uma amostra do cereal. Medida a absorvância obteve-se $A_0 = 0.114$, invertendo a relação anterior tem-se $\hat{C} = \dfrac{A_0 - \hat{\beta}_0}{\hat{\beta}_1} = 1.95$ (ppm).

Pela expressão $Teor = \dfrac{m(Fe)}{m(cereal)} = \dfrac{\hat{C}V}{m(cereal)}$

onde $m(cereal) = 2.054$ g e V = 100 ml obtém-se

Teor = 0.095 mg Fe / g cereal

Análise de erros:

Erro associado à estimativa $\hat{C}$: usando a expressão deduzida no Exemplo 3 da Secção 5.7 (Erros e propagação de erros):

$$\sigma_X^2 \approx \frac{1}{\beta_1^2}\left[\sigma^2 + \sigma_0^2 + \sigma_1^2 E(X)^2 + 2\sigma_{01}E(X)\right]$$

onde $\sigma^2 = V(Y)$, $\quad \sigma_0^2 = V(\hat{\beta}_0)$, $\quad \sigma_1^2 = V(\hat{\beta}_1)$, $\sigma_{01} = \mathrm{Cov}(\hat{\beta}_0, \hat{\beta}_1)$ e substituindo todos os valores desconhecidos pelas respectivas estimativas temos

$\hat{\sigma}_C^2 = \hat{\sigma}_X^2 \approx 5.8943 \times 10^{-3}$ $\qquad$ $\hat{\sigma}_C \approx 0.077$

Exercício: Estimar σ_{Teor} sabendo que $\sigma_{m(cereal)} = 0.0005$ e $\sigma_V = 0.1$.