MLG

Curso de Modelos Lineares Generalizado - DEST/UFMG
Marcos Oliveira Prates

4 de agosto de 2015

Informação do Curso

- e-mail: marcosop@est.ufmg.br
- sala: 4060 / ICEx
- telefone: 3409-5932
- site: www.est.ufmg.br/˜marcosop

Informação do Curso

- Modelo de Regressão Linear
  - Regressão Linear Simples
  - Regressão Linear Múltipla (forma matricial)
  - Análise de Resíduos
  - Transformação de Variáveis
  - Regressão Não Linear

Informação do Curso

- Modelos Lineares Generalizados
  - Família exponencial. Ajuste pelo método de Newton Raphson. Inferência. Seleção de Variáveis e Análise de resíduos.
  - Regressão Logística
  - Regressão de Poisson
  - Regressão Gama
- Ponto de vista clássico e Bayesiano

Informação do Curso

- Não haverá aula nos dias 08/03 e 10/03.
- Teremos 02 (duas) provas no valor de 20 pontos cada.
- Teremos a apresentação de 03 (três) trabalhos:
  1- Explicação e exemplo de uso de alguma função de GLM em R (10 pontos);
  2- Leitura de um artigo, resumo e apresentação (15 pontos);
  3- Trabalho prático (25 pontos);
- 10 pontos para listas, frequencia e participação.
- As datas da prova serão definidas oficalmente durante o semestre.

Informação do Curso

- Referências bibliográficas:
  - Ravishanker N. and Dey D. K. (2000). A First Course in Linear Model Theory. Chapman & Hall.
  - Seber G. A. F. and Lee A. J. (2003). Linear Regression Analysis. Wiley.
  - McCullagh P. and Nelder J. A. (1989). Generalized Linear Models. Chapman & Hall.
  - Neter J., Kutner M. H., Nachtsheim, C. J. and Wasserman W. (1996). Applied Linear Statistical Models. McGraw-Hill.
  - Cordeiro G. M. (2007). Modelos Lineares Generalizados. Mini curso RBRAS.
  - Paula G. A. (2004). MODELOS DE REGRESSAO com apoio computacional. www.ime.usp.br/˜giapaula.
  - **R**, Wikipedia, etc.

Modelos Estatísticos

- Modelos estatísticos devem ser capazes de acomodar a variabilidade inerente aos dados.
- Assim, de forma geral, um modelo estatístico pode ser escrito da seguinte forma:

$$Y = \text{componente determinística} + \text{componente aleatória}$$

- Como iremos ver no decorrer do curso, existem diversas maneiras de específicar essas componentes.
- Começaremos com o caso mais simples de uma regressão linear simples.

- Uma regressão linear simples tem como objetivo aproximar uma variável de resposta Y através de uma função linear de uma variável de interesse, ou seja,

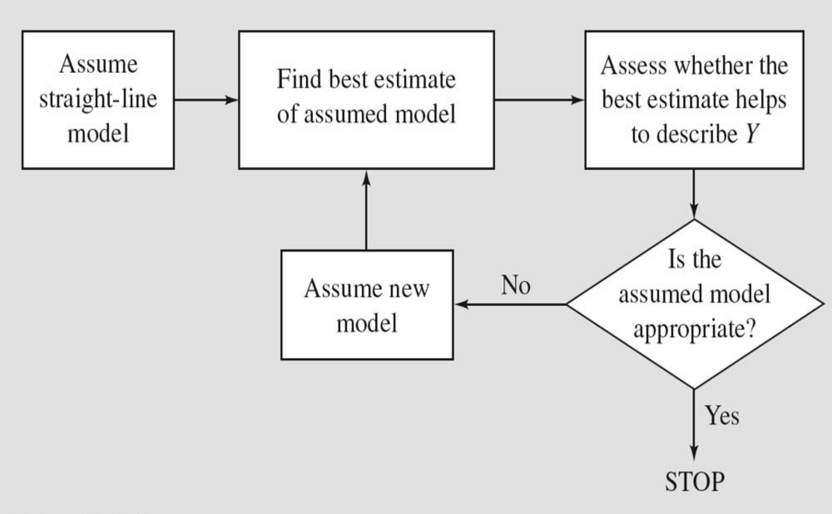
$$Y = f(X,\beta) + \varepsilon = \beta_0 + \beta_1 X + \varepsilon$$

- No qual assume-se que:
  - $E(\varepsilon) = 0$
  - $V(\varepsilon) = \sigma^2$ (Homocedasticidade)
  - $Cov(\varepsilon_i, \varepsilon_j) = 0$,

  em outras palavras, os erros tem média zero, variância constante e são não correlacionados.

- As variável preditora X pode vir de diversas fontes
  - inputs quantitativos (valores reais, medidas)
  - tranformação de variável quantitativas ($\log, \sqrt{}$,etc)
  - inputs qualitativos ("dummy" e.x. genero, classes)
- Dessa forma, um modelos de regressão consiste em 4 passos:
  1. Escolher o componente determinístico do modelo;
  2. Utilizar os dados para estimar os parâmetros do modelo;
  3. Especificar a distribuição do erro;
  4. Avaliar o modelo estatístico;

Diagrama de um Modelo de Regressão



- Após a escolha do modelo, devemos utilizar a informação nos dados para fazer a estimação dos parâmetros.
- O método mais utilizado é conhecido como método de mínimos quadrados, e é dados por:

$$
\begin{aligned}
SEQ(\beta) &= \sum_{i=1}^{n}(Y_i - f(X_i,\beta))^2 \\
&= \sum_{i=1}^{n}(Y_i - \beta_0 - X_i\beta_1)^2,
\end{aligned}
$$

  Os estimadores de β_0 e β_1 são aqueles que minimizam o *SEQ*
- Diferenciando *SEQ* em relação a β chegamos as equações normais:

$$
\frac{\partial SEQ}{\partial \beta_0} = -2\sum_{i=1}^{n}(Y_i - \beta_0 - X_i\beta_1)
$$

$$
\frac{\partial SEQ}{\partial \beta_1} = -2\sum_{i=1}^{n}X_i(Y_i - \beta_0 - X_i\beta_1)
$$

- Quando essas derivadas parciais são igualadas a zero e resolvidas, temos que $\hat{\beta}_0$ e $\hat{\beta}_1$ são os valores que minimizam o *SEQ*:

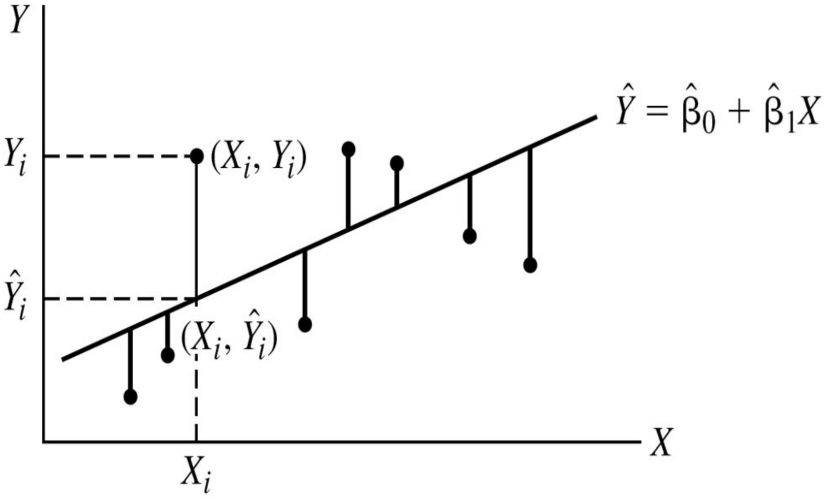
$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - X_i\hat{\beta}_1) = 0$$

$$\sum_{i=1}^{n}X_i(Y_i - \hat{\beta}_0 - X_i\hat{\beta}_1) = 0$$

- Assim resolvendo ambas equações simultanemanete temos que

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{n}(X_i - \bar{X})(Y_i - \bar{Y})}{\sum_{i=1}^{n}(X_i - \bar{X})^2} = \frac{\sum_{i=1}^{n} X_i Y_i - n\bar{X}\bar{Y}}{\sum_{i=1}^{n} X_i^2 - n\bar{X}^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1\bar{X}$$


Y
$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$
Y_i
(X_i, Y_i)
$\hat{Y}_i$
$(X_i, \hat{Y}_i)$
X
X_i

Regressão Linear Simples

- Dada a estimativa dos parâmetros $(\hat{\beta}_0, \hat{\beta}_1)$ temos o modelo ajustado
$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X$$
Note que por construção o ponto $(\bar{Y}, \bar{X})$ sempre cai exatamente no modelo ajustado
- Para os valores de $X = X_i$, um dos valores do banco de dados, os valor estimado é da forma
$$\hat{Y}_i = \hat{\beta}_0 + \hat{\beta}_1 X_i, \quad i = 1, \ldots, n.$$
- Os resíduos, e_i, são definidos como a diferença entre os valores observados e os valores estimados, assim:
$$e_i = Y_i - \hat{Y}_i = Y_i - (\hat{\beta}_0 + \hat{\beta}_1 X_i) \quad i = 1, \ldots, n.$$

- Logo temos que

$$\begin{aligned} SEQ &= \sum_{i=1}^{n}(Y_i - f(X_i,\beta))^2 = \sum_{i=1}^{n}(Y_i - \beta_0 - X_i\beta_1)^2 \\ &= \sum_{i=1}^{n}(Y_i - \hat{Y}_i)^2 = \sum_{i=1}^{n} e_i^2. \end{aligned}$$

- Propriedades dos erros
  Como (β_0 e $\hat{\beta}_1$) são estimados pelo método de mínimos quadrados temos

$$\sum_{i=1}^{n}(Y_i - \hat{\beta}_0 - X_i\hat{\beta}_1) = 0$$

$$\sum_{i=1}^{n}X_i(Y_i - \hat{\beta}_0 - X_i\hat{\beta}_1) = 0$$

portanto,

$$\sum_{i=1}^{n} e_i = \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - X_i\hat{\beta}_1) = 0 = \sum_{i=1}^{n} (Y_i - \hat{Y}_i)$$

$$\sum_{i=1}^{n} X_i e_i = \sum_{i=1}^{n} X_i(Y_i - \hat{\beta}_0 - X_i\hat{\beta}_1) = 0 = \sum_{i=1}^{n} X_i(Y_i - \hat{Y}_i)$$

- Até esse momento ainda não usamos a suposição de nenhuma distribuição para o erro
- Comumente adota-se que o erro $\varepsilon \sim N(0, \sigma^2)$, isso garante que todas as suposições do modelo sejam atendidas e possibilita inferência sobre a modelagem
- Com a suposição de normalidade temos que σ^2 pode ser estimado

$$\hat{\sigma}^2 = \text{MEQ} = \frac{1}{n-2}\text{SRQ} = \frac{1}{n-2}\sum_{i=1}^{n} e_i^2,$$

- Suponha que os dados (y_i, x_i), sejam modelados por uma regressão linear simples e assuma que $\varepsilon_i \overset{iid}{\sim} N(0, \sigma^2)$ temos que:

$$L(\beta_0, \beta_1, \sigma^2) = \prod_{i=1}^{n} \frac{-1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(y_i - \beta_0 - \beta_1 x_i)^2\right)$$

Logo,

$$\ln L = l(\beta_0, \beta_1, \sigma^2) = -\frac{n}{2}\ln 2\pi - \frac{n}{2}\ln \sigma^2 - \frac{1}{2\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2$$

diferenciando em relação aos β's temos

$$\frac{\partial l}{\partial \beta_0} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$\frac{\partial l}{\partial \beta_1} = \frac{1}{\sigma^2}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)x_i = 0$$

Resolvendo o sistema de equações, temos os mesmos estimadores do método de Mínimos Quadrados, porém

$$\frac{\partial l}{\partial \sigma^2} = -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}\sum_{i=1}^{n}(y_i - \beta_0 - \beta_1 x_i)^2 = 0$$

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^{n}(y_i - \hat{\beta_0} - \hat{\beta_1} x_i)^2}{n} = \frac{1}{n}\sum_{i=1}^{n} e_i^2$$

- Assumindo que o erro possui distribuição normal independetes, podemos executar testes de hipóteses para β_0 e β_1

$$\hat{\beta_0} \sim N\left(\beta_0, \sigma^2\left(\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)\right)$$

$$\hat{\beta_1} \sim N\left(\beta_1, \sigma^2 \frac{1}{\sum_{i=1}^{n}(x_i - \bar{x})^2}\right)$$

- Logo para testar, $i = 0, 1$

$$\begin{cases} H_0 : \beta_i = 0 \\ H_1 : \beta_i \neq 0 \end{cases} \tag{1}$$

seguimos de forma equivalente a testes de hipóteses normais

- Assim

$$\begin{cases} H_0 : \beta_0 = 0 \\ H_1 : \beta_0 \neq 0 \end{cases} \tag{2}$$

é dado por

$$t_{obs} = \frac{\hat{\beta}_0 - 0}{s(\hat{\beta}_0)} \sim t_{n-2}.$$

Portanto, reijeita-se H_0 se $|t_{obs}| > t_{\frac{\alpha}{2},n-2}$

- De forma equivalente temos que para $beta_1$

$$\begin{cases} H_0 : \beta_1 = 0 \\ H_1 : \beta_1 \neq 0, \end{cases} \tag{3}$$

$$t_{obs} = \frac{\hat{\beta}_1 - 0}{s(\hat{\beta}_1)} \sim t_{n-2}.$$

Portanto, reijeita-se H_0 se $|t_{obs}| > t_{\frac{\alpha}{2},n-2}$

- Antes de seguirmos com a análise de regressão linear, vamos dar uma pausa para re-ver ou ver algumas propriedades matriciais importantes para a continuação do curso.
- **Espaço Vetorial**: O espaço vetorial é um conjunto

$$\mathcal{V}_n = \{v_i = (v_{i1}, \ldots, v_{in}),\ v_{ij} \in \Re,\ i = 1, \ldots, l\}$$

  no qual é fechado sobre a adição, multiplicação por escalar, e contém o vetor **0**.
- **Vetores Dependentes e Vetore Independentes**: Seja $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ um conjunto de vetores n-dimensionais pertencentes a $\mathcal{V}_n$. Esse m vetores são ditos linearmente dependentes **se e somente se** existir escalares $c_1, \ldots, c_m$, com pelo menos um diferente de zero, tal que $\sum_{i=1}^{m} c_i \boldsymbol{v}_i = 0$.
- Se $c_1 = \ldots = c_m = 0$ para que $\sum_{i=1}^{m} c_i \boldsymbol{v}_i = 0$, então $\{\boldsymbol{v_1}, \ldots, \boldsymbol{v_m}\}$ são ditos independetes.

Propriedades Matricial

- Regras de Adição de matrizes. Sejam as matrizes A, B, C $m \times n$ e a, b, c escalares:
  1. $(A+B)+C = A+(B+C)$
  2. $A+B = B+A$
  3. $A+(-A) = (-A)+A = 0$
  4. $A+0 = 0+A$
  5. $c(A+B) = cA+cB$
  6. $(a+b)C = aC+bC$
  7. $abC = a(bC) = b(aC)$
  8. $0A = 0$
  9. $1A = A$

- Regras de Multiplicação de matrizes. Seja A uma matriz $m \times n$ e B, C matrizes com a dimensão apropriada:
  1. $(AB)C = A(BC)$
  2. $A(B + C) = AB + AC$
  3. $(A + B)C = AC + BC$
  4. $a(BC) = (aB)C = B(aC)$
  5. $I_m A = AI_n = A$
  6. $0_m A = A0_n = 0$
- É importante frizar que no caso matricial $AB \neq BA$ em geral, em alguns casos podemos ter AB definido e BA não.

Propriedades Matricial

- Regras de Transposição de matrizes. Seja A e B conformes sobre a adição e A e C conformes com a multiplicação:
  1. $(A')' = A$
  2. $(aA + bB)' = aA' + bB'$
  3. $(cA)' = cA'$
  4. $A' = B'$ se e somente se $A = B$
  5. $(AC)' = C'A'$
- **Matriz simétrica**: Uma matriz é dita simétrica se $A = A'$.

Propriedades Matricial

- **Traço da Matriz**: O traço da matriz é definido como a soma dos elementos da diagonal de A. Seja $A_{n\times n}$ então $tr(A) = \sum_{i=1}^{n} a_{ii}$.
- Propriedades do traço:
  1. $tr(I_n) = n$
  2. $tr(aA + bB) = atr(A) + btr(B)$
  3. $tr(AB) = tr(BA)$
  4. $tr(ABC) = tr(CAB) = tr(BCA)$
  5. $tr(A) = tr(A')$
  6. $tr(AA') = tr(A'A) = \sum_{i,j=1}^{n} a_{ij}^2$

- **Determinante de Matriz**: O determinante de uma matriz $A_{n\times n}$ é um escalar dado porque

$$|A| = \sum_{j=1}^{n} a_{ij}(-1)^{i+j}|M_{ij}|, \text{ para qual quer } i \text{ fixado}$$

onde M_{ij} é matriz obtida removendo-se a i-ésima linha e j-ésima coluna de A.

- Propriedades do determinante:
  1. $|A| = |A'|$
  2. $|cA| = c^n|A|$
  3. $|AB| = |A||B|$
  4. Se A é diagonal ou triangular inferior (superior) então $|A| = \prod_{i=1}^{n} a_{ii}$
  5. Se duas linhas ou colunas de A são iguais, então $|A| = 0$
  6. Se tem linhas ou colunas iguas a 0, então $|A| = 0$

- Propriedades do determinante:
  7. Se as linhas (colunas) são linearmente dependentes, então $|A| = 0$
  8. Seja B obtido multiplicando uma linha ou coluna A por um escalar c, então $|B| = c|A|$
  9. Seja B obtido trocando de posição uma coluna ou linha de A, então $|B| = -|A|$
  10. Seja $A_{m\times n}$ e $B_{n\times m}$, então $|I_m + AB| = |I_n + BA|$

Propriedades Matricial

- **Inverso de uma Matriz**: Seja $A_{n\times n}$. Se existe $B_{n\times n}$ tal que $AB = I_n$ (e $BA = I_n$), então B é chamado de inverso de A e denotamos por A^{-1}.
- Propriedades do inverso:
  1. A^{-1} é unico
  2. $(AB)^{-1} = B^{-1}A^{-1}$
  3. $(cA)^{-1} = (Ac)^{-1} = \frac{A^{-1}}{c}$
  4. Se $|A| \neq 0$, então A' e A^{-1} são não singulares e $(A')^{-1} = (A^{-1})'$
  5. $(A+BCD)^{-1} = A^{-1} - A^{-1}B(C^{-1} + DA^{-1}B)^{-1}DA^{-1}$ (Sherman-Morisson-Woodbury Theorem)
  6. $|A^{-1}| = |A|^{-1}$

- **Espaço das Colunas de uma Matriz**: Seja $A_{m\times n}$ na qual as colunas m-dimensionais são $\boldsymbol{a}_1, \ldots, \boldsymbol{a}_n$. O espaço vetorial gerado pelas n colunas de A é chamado de espaço das colunas de A ($\mathcal{C}(A)$). A dimensão do espaço das colunas de A é o número de colunas linarmente independetes.
- **Espaço Nulo de uma Matriz**: O espaço nulo, $\mathcal{N}(A)$, de uma matriz $A_{m\times n}$ consiste de todos os vetores n-dimensionais $\boldsymbol{x}$ tal que $A\boldsymbol{x} = 0$, ou seja,

$$\mathcal{N}(A) = \{\boldsymbol{x} \in \Re^n \text{ tal que } A\boldsymbol{x} = 0\}$$

Propriedades Matricial

- Propriedades do espaço de colunas:
  1. $dim[\mathcal{C}(A)] = n - dim[\mathcal{N}(A)]$
  2. $\mathcal{N}(A) = \{\mathcal{C}(A)\}^{\perp}$
  3. $\mathcal{C}(A'A) = \mathcal{C}(A')$
  4. Para qualquer A e B, $\mathcal{C}(AB) \subseteq \mathcal{C}(A)$

Propriedades Matricial

- **Rank de uma Matriz**: Seja $A_{m\times n}$. Dizemos que a matriz A possui rank completo de linha se $r(A) = m$ (somente se $m \leq n$), e possui rank completo de coluna se $r(A) = n$ (somente se $n \leq m$). Dizemos que uma matriz tem rank r ($r(A) = r$) se o rank das colunas é igual ao rank das linhas iguais a r.
- Propriedades de rank:
  1. $A_{m\times n}$ tem rank r se a maior sub-matriz não singular de A possui tamanho r.
  2. Para $A_{m\times n}$, $r(A) \leq \min(m,n)$
  3. $r(A+B) \leq r(A) + r(B)$
  4. $r(AB) \leq \min\{r(A), r(B)\}$
  5. Para matrizes não singulares A, B e uma matriz qualquer C
  $$r(C) = r(AC) = r(CB) = r(ACB)$$
  6. $r(A) = r(A') = r(A'A) = r(AA')$
  7. $r(A,b) \geq r(A)$, isso é, adicionar uma coluna em a nunca reduz o seu rank

Propriedades Matricial

- Seja as seguintes notações:
  - $A_{n\times n}$: matriz de constantes
  - $X_{n\times 1}$: vetor de variáveis (ou parâmetros)
  - $a_{n\times 1}$: vetor de constantes .
- Derivadas de Matrizes:
  - $\frac{\partial a^T X}{\partial X} = a$
  - $\frac{\partial X^T X}{\partial X} = 2X$
  - $\frac{\partial a^T Ax}{\partial x} = A'a$
  - $\frac{\partial x^T Ax}{\partial x} = Ax + A^T x$

- Seja $X_{n\times 1}$ um vetor de variáveis aleatórias e $\mathbf{A}_{n\times n}$ uma matriz simétrica. A $E(\mathbf{X}) = \mu$ e a $\text{Var}(\mathbf{X}) = \Sigma = (\sigma_{ij})$. Temos que:

$$E(\mathbf{X}'\mathbf{A}\mathbf{X}) = tr(\mathbf{A}\Sigma) + \mu'\mathbf{A}\mu$$

- Prova

$$\begin{aligned}
E(\mathbf{X}'\mathbf{A}\mathbf{X}) &= tr(E[\mathbf{X}'\mathbf{A}\mathbf{X}]) \\
&= E[tr(\mathbf{X}'\mathbf{A}\mathbf{X})] = E[tr(\mathbf{A}\mathbf{X}\mathbf{X}')] \\
&= tr[E(\mathbf{A}\mathbf{X}\mathbf{X}')] = tr[\mathbf{A}E(\mathbf{X}\mathbf{X}')] \\
&= tr[\mathbf{A}(\text{Var}(\mathbf{X}) + \mu\mu')] = tr(\mathbf{A}\Sigma) + tr(\mathbf{A}\mu\mu') \\
&= tr(\mathbf{A}\Sigma) + tr(\mu'\mathbf{A}\mu) \\
&= tr(\mathbf{A}\Sigma) + \mu'\mathbf{A}\mu
\end{aligned}$$

- Assim, como na regressão linear simples, uma regressão linear múltipla supoem que função de regressão $E(Y|X)$ é linearmente dependente dos preditores $X_1, \ldots, X_p$.
- Regressões lineares são simples e comumente fornecem um descrição adequada e interpretável de como as variáveis exploratórias afetam as resposta.

- Suponha que temos $X' = (X_1, \ldots, X_p)$ e queremos prever uma resposta Y. Uma regressão linear é da formado

$$Y = \beta_0 + \sum_{i=1}^{p} X_i \beta_i$$

onde β_i não são conhecidos.

- As variáveis X_i podem vir de diferentes fontes
  - inputs quantitativos (valores reais, medidas)
  - tranformação de inputs quantitatvos ($\log, \sqrt{}$,etc)
  - expansão de base ($X_2 = X_1^2$, $X_3 = X_1^3$)
  - inputs qualitativos ("dummy" e.x. genero, classes)
  - interação ($X_3 = X_1.X_2$)

- Normalmente tem-se um banco de observações $(x_1, y_1), \ldots, (x_n, y_n)$, no qual o objetivo é estimar os parâmetros β.
- Ao extender a noção de regressão simples para regressão múltipla o método de mínimos quadrados continua sendo utilizado, no qual as estimativas de β são escolhidas derviando o *SEQ* em relação a $\beta_0, \cdots, \beta_k$

$$
\begin{aligned}
SEQ(\beta) &= \sum_{i=1}^{n} (Y_i - f(X_i, \beta))^2 \\
&= \sum_{i=1}^{n} (Y_i - \beta_0 - \sum_{j=1}^{k} X_{ij}\beta_j)^2.
\end{aligned}
$$

Intuição do métodos de minímos quadrados

Elements of Statistical Learning (2nd Ed.) ©Hastie, Tibshirani & Friedman 2009 Chap 3

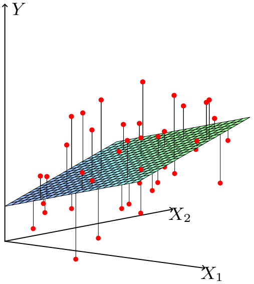


FIGURE 3.1. *Linear least squares fitting with $X \in \mathbb{R}^2$. We seek the linear function of X that minimizes the sum of squared residuals from Y.*

- Podemos representar tanto a regressão simples, assim como a regressão múltipla na forma matricial. Seja $p = k + 1$ and

$$\underset{n\times 1}{\mathbf{Y}} = \begin{pmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{pmatrix}, \; \underset{n\times p}{\mathbf{X}} = \begin{pmatrix} 1 & X_{11} & \cdots & X_{1k} \\ 1 & X_{21} & \cdots & X_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ 1 & X_{n1} & \cdots & X_{nk} \end{pmatrix}, \; \underset{p\times 1}{\beta} = \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_k \end{pmatrix}, \; \underset{n\times 1}{\varepsilon} = \begin{pmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{pmatrix}.$$

  Aqui, $\underset{n\times p}{\mathbf{X}}$ é chamado de matriz de desenho. Dessa forma o modelo de regressão pode escrito como:

$$\underset{n\times 1}{\mathbf{Y}} = \underset{n\times p}{\mathbf{X}} \underset{p\times 1}{\beta} + \underset{n\times 1}{\varepsilon},$$

- Nesse caso, seja **X** uma matriz $n \times p$, onde a primeira coluna de **X** é de 1's.
- Defina **Y** como um vetor $v \times 1$ de respostas. Logo

$$SEQ(\beta) = (\mathbf{Y} - \mathbf{X}\beta)'(\mathbf{Y} - \mathbf{X}\beta)$$

(mostrar no quadro equivalência)

- Para minimizar *SEQ* devemos derivar em relação a β e igualar a 0

$$\frac{\partial SEQ(\beta)}{\partial \beta} = -2\mathbf{X}^\top(y - \mathbf{X}\beta) \equiv 0 \text{ (Verificar)}$$

- Seja $\hat{\beta} = \begin{pmatrix} \hat{\beta}_0 \\ \hat{\beta}_1 \end{pmatrix}$. As equações normais são da forma

$$\mathbf{X}'\mathbf{X}\hat{\beta} = \mathbf{X}'\mathbf{Y}.$$

- Assumindo que **X** tem posto completo e resolvendo as equações normais temos que
$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'y$$
- No caso da RLS podemos visualizar as matrizes

$$\mathbf{X}'\mathbf{Y} = \begin{pmatrix} \sum_{i=1}^{n} Y_i \\ \sum_{i=1}^{n} X_i Y_i \end{pmatrix}, \quad \mathbf{X}'\mathbf{X} = \begin{pmatrix} n & \sum_{i=1}^{n} X_i \\ \sum_{i=1}^{n} X_i & \sum_{i=1}^{n} X_i^2 \end{pmatrix}$$

e

$$(\mathbf{X}'\mathbf{X})^{-1} = \frac{1}{nS_{XX}} \begin{pmatrix} \sum_{i=1}^{n} X_i^2 & -n\overline{X} \\ -n\overline{X} & n \end{pmatrix} = \begin{pmatrix} \frac{1}{n} + \frac{\overline{X}^2}{S_{XX}} & -\frac{\overline{X}}{S_{XX}} \\ -\frac{\overline{X}}{S_{XX}} & \frac{1}{S_{XX}} \end{pmatrix}.$$

- Assim temos o que o valor estimado pela regressção é da forma

$$\hat{y} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}^\top\mathbf{X})^{-1}\mathbf{X}^\top y \tag{4}$$

$$\underset{n\times 1}{\hat{\mathbf{Y}}} = \begin{pmatrix} \hat{Y}_1 \\ \hat{Y}_2 \\ \vdots \\ \hat{Y}_n \end{pmatrix} = \mathbf{X}\hat{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \underset{n\times n}{\mathbf{H}}\,\underset{n\times 1}{\mathbf{Y}},$$

onde

$$\underset{n\times n}{\mathbf{H}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$$

é chamada de **Hat Matrix** ou **Matriz de projeção**.

- Propriedades da matriz de projeção:
  - $\mathbf{H}$ é simétrico, ou seja, $\mathbf{H}' = \mathbf{H}$
  - $\mathbf{H}$ é idepotente, ou seja, $\mathbf{H}^2 = \mathbf{H}$
  - $tr(\mathbf{H}) = p$ onde p é o posto de $\mathbf{X}$
  - $\mathbf{H}$ é a matriz de projeção no plano gerado pelas colunas de $\mathbf{X}$, ou seja, $\mathbf{HX} = \mathbf{X}$
  - Quais são os autovalores de $\mathbf{H}$?

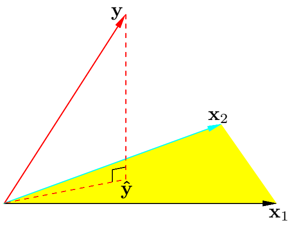


FIGURE 3.2. *The N-dimensional geometry of least squares regression with two predictors. The outcome vector* $\mathbf{y}$ *is orthogonally projected onto the hyperplane spanned by the input vectors* $\mathbf{x}_1$ *and* $\mathbf{x}_2$*. The projection* $\hat{\mathbf{y}}$ *represents the vector of the least squares predictions*

- Quando $\mathbf{X}$ não possui posto completo, então $\hat{\beta}$ não pode ser estimado de forma única
- Porém $\hat{y} = \mathbf{X}\hat{\beta}$ é único, pois $\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ unicamente define a projeção de y no espaço gerado por $\mathbf{X}$
- Portanto apesar das infinitas soluções para β temos que $\hat{y}$ continua sendo estimado de forma única

- Até o momento não foi necessária nenhuma hipótese sobre o modelo.
- Para inferência vamos supor que

$$\begin{aligned} Y &= \mathbf{X}\beta + \varepsilon \\ \varepsilon &\sim N(0, \sigma^2 \mathbf{I}) \end{aligned}$$

- Pois

$$\underset{n\times n}{\mathrm{Var}(\mathbf{Y})} = \underset{n\times n}{\mathrm{Var}(\varepsilon)} = \begin{pmatrix} \mathrm{Var}(\varepsilon_1) & \mathrm{Cov}(\varepsilon_1, \varepsilon_2) & \cdots & \mathrm{Cov}(\varepsilon_1, \varepsilon_n) \\ \mathrm{Cov}(\varepsilon_2, \varepsilon_1) & \mathrm{Var}(\varepsilon_2) & \cdots & \mathrm{Cov}(\varepsilon_2, \varepsilon_n) \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{Cov}(\varepsilon_n, \varepsilon_1) & \mathrm{Cov}(\varepsilon_n, \varepsilon_2) & \cdots & \mathrm{Var}(\varepsilon_n) \end{pmatrix} = \sigma^2 \underset{n\times n}{\mathbf{I}},$$

onde $\mathbf{I}$ é a matriz de identidade $n \times n$.

- Dessa forma temos que

$$\text{Var}(\hat{\beta}) = (\mathbf{X}'\mathbf{X})^{-1}\sigma^2$$

- Sob a hipótese de normalidade é simples mostrar que

$$\hat{\beta} \sim N(\beta, (\mathbf{X}^\top\mathbf{X})^{-1}\sigma^2)$$

- Resíduos
  O vetor de resíduos é dado por

$$\underset{n\times 1}{e} = \begin{pmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{pmatrix} = \mathbf{Y} - \hat{\mathbf{Y}} = \mathbf{Y} - \mathbf{HY} = (\mathbf{I} - \mathbf{H})\mathbf{Y}.$$

  É fácil mostrar que a matriz $(\mathbf{I} - \mathbf{H})$ também é *simétrica* e *idepotente*.

- Estamando a variância do erro
  A soma quadrática dos erros é dada por

$$\text{SSE} = e'e = \sum_{i=1}^{n} e_i^2.$$

Dessa forma o estimador apropriado para σ^2 é

$$\hat{\sigma}^2 = \text{MSE} = \frac{1}{n-p}\text{SSE} = \frac{1}{n-p}\sum_{i=1}^{n} e_i^2,$$

- Pela normalidade podemos mostrar

$$(N-p)\hat{\sigma}^2 \sim \sigma^2\chi^2_{n-p} \text{ (verifique)}$$

- Matriz de Covariância dos Resíduos

$$\begin{aligned} & \mathrm{Var}(e) = \mathrm{Cov}((\mathbf{I}-\mathbf{H})\mathbf{Y},(\mathbf{I}-\mathbf{H})\mathbf{Y}) \\ = \quad & (\mathbf{I}-\mathbf{H})\mathrm{Var}(\mathbf{Y})(\mathbf{I}-\mathbf{H})' = \sigma^2(\mathbf{I}-\mathbf{H}). \end{aligned}$$

- Logo, a matriz estimada de covariância para os resíduos é dada por

$$\widehat{\mathrm{Var}}(e) = \mathrm{MSE}(\mathbf{I}-\mathbf{H}).$$

Inverso Generalizado, G-inverso

- Antes de apresentar a definição do g-inverso apresentamos os resultados:
- Seja $\mathbf{A}$ e $\mathbf{B}$ matrizes $m \times n$. Seja $\mathbf{C}_{p\times m}$ e $\mathbf{D}_{n\times p}$.
  1. se $\mathbf{CA} = \mathbf{CB}$ então $\mathbf{A} = \mathbf{B}$.
  2. se $\mathbf{AD} = \mathbf{BD}$ então $\mathbf{A} = \mathbf{B}$.
  3. se $\mathbf{CAD} = \mathbf{CBD}$ então $\mathbf{A} = \mathbf{B}$.
  4. $\mathbf{AB} = \mathbf{AC}$ se e somente se $\mathbf{A}'\mathbf{AB} = \mathbf{A}'\mathbf{AC}$
  5. $\mathbf{EA}' = \mathbf{FA}'$ se e somente se $\mathbf{EA}'\mathbf{A} = \mathbf{FA}'\mathbf{A}$

Inverso Generalizado, G-inverso

- **G-inverso**: Um inverso generalizado (g-inverso) de uma matriz $m \times n$ A é uma matriz $\mathbf{G}_{n\times m}$ se ela satisfaz a seguinte relação

$$\mathbf{AGA} = \mathbf{A}$$

- Denotamos o g-inverso por $\mathbf{A}^{-}$
- Teorema: A matriz $\mathbf{G}$ g-inverso de uma matriz real $\mathbf{A}$ sempre existe. E $\mathbf{G} = \mathbf{A}^{-1}$ se $\mathbf{A}$ é não singular.

Inverso Generalizado, G-inverso

- Seja $\mathbf{A}_{m \times n}$ de posto r. Então
  - $\mathbf{A}^{-}\mathbf{A}$ e $\mathbf{A}\mathbf{A}^{-}$ são idepotente
  - $(\mathbf{I} - \mathbf{A}^{-}\mathbf{A})$ e $(\mathbf{I} - \mathbf{A}\mathbf{A}^{-})$ são idepotente
- Seja $\mathbf{G}$ g-inverso de $\mathbf{A}'\mathbf{A}$, então
  - $\mathbf{G}'$ é g-inverso de $\mathbf{A}'\mathbf{A}$
  - $\mathbf{G}\mathbf{A}'$ é g-inverso $\mathbf{A}$, tal que $\mathbf{A}\mathbf{G}\mathbf{A}'\mathbf{A} = \mathbf{A}$
  - $\mathbf{A}\mathbf{G}\mathbf{A}'$ é invariante a escolha de $\mathbf{G}$, ou seja,

$$\mathbf{A}\mathbf{G}_1\mathbf{A}' = \mathbf{A}\mathbf{G}_2\mathbf{A}'$$

  - $\mathbf{A}\mathbf{G}\mathbf{A}'$ é simetrica

- Como comentamos, a não ser que $r(\mathbf{X}) = p$, $\tilde{\beta}$ não é único.
- Apesar de no modelo de posto completo podermos estimar qualquer função de β devemos nos restringir apenas algumas funções de β quando $r(\mathbf{X}) < p$.
- Essas funções são chamdas de funções estimáveis
- **Definição:** Uma função linear paramétrica $\mathbf{c}'\beta$ é chamada estimável de β se existe um vetor n-dimensional $\mathbf{t} = (t_1, \ldots, t_n)'$ tal que a esperança da combinação linear $\mathbf{t}'\mathbf{y} = \mathbf{c}'\beta$, isto é,

$$E(\mathbf{t}'\mathbf{y}) = \mathbf{c}'\beta$$

- Ou seja, $\mathbf{c}'\beta$ se existe um função linear de $\mathbf{y}$ tal que o valor esperado é igual à $\mathbf{c}'\beta$.

Funções Estimáveis

- Se $r(\mathbf{X}) = p$ então qualquer função linear de β é estimável.
- Esse não é o caso para modelos no qual $r(\mathbf{X}) < p$. Nesses casos, devemos verificar que a função de β é estimável.
- Para garantir que uma função $\mathbf{c}'\beta$ devemos verificar:
  - Uma função $\mathbf{c}'\beta$ é estimável se e somente se $\mathbf{c}' = \mathbf{t}'\mathbf{X}$ para um vetor $\mathbf{t}$
  - Uma função $\mathbf{c}'\beta$ é estimável se e somente se $\mathbf{c}' = \mathbf{c}'\mathbf{W}$ onde $\mathbf{W} = \mathbf{G}\mathbf{X}'\mathbf{X}$.

Funções Estimáveis

- Desses resultados é possíve mostrar que:
  - O valor esperado de qualquer observação é estimável.
  - Qualquer combinação linear de funções estimáveis é estimável.
  - Dado uma função estimável $\mathbf{c}'\beta$, a quantidade $\mathbf{c}'\tilde{\beta}$ é invariante a escolha de $\tilde{\beta}$.

O Teorema de Gauss-Markov

- Seja $\mathbf{c}'\beta$ uma função estimável de β e seja $\tilde{\beta}$ uma solução qualquer para as equações normais.
- O Teorma de Gauss-Markov afirma $\mathbf{c}'\tilde{\beta}$ é o melhor estimador não viesado de $\mathbf{c}'\beta$ com variância $\text{Var}(\mathbf{c}'\tilde{\beta}) = \sigma^2\mathbf{c}'\mathbf{G}\mathbf{c}$.
- Se $\tilde{\beta} = \hat{\beta}$ então $\text{Var}(\mathbf{c}'\hat{\beta}) = \sigma^2\mathbf{c}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{c}$.