

# MLG

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

23 de agosto de 2017

# Hipóteses do Modelo de Regressão Linear

(1) Linearidade

$$E(Y_i) = \beta_0 + \beta_1 X_{i1} + \beta_2 X_{i2} + \dots + \beta_k X_{ik}.$$

(2) Heterocidade da Variância

$$\text{Var}(\varepsilon_i) = \sigma^2$$

(3) Não correlação

$$\text{Cov}(\varepsilon_i, \varepsilon_j) = 0 \text{ for all } i \neq j.$$

(4) Normalidade

$$\varepsilon_i \sim N(0, \sigma^2) \text{ or } \varepsilon = (\varepsilon_1, \dots, \varepsilon_n)' \sim N_n(0, \sigma^2 I).$$

- Resíduos

O resíduo  $e_i$  é a diferença entre o valor observado  $Y_i$  e o valor ajustado  $\hat{Y}_i$ :

$$e_i = Y_i - \hat{Y}_i.$$

O resíduo pode ser pensado como o erro observado, em distinção ao erro verdadeiro  $\varepsilon_i$  do modelo de regressão:

$$\varepsilon_i = Y_i - E[Y_i].$$

- Propriedades do Resíduo

- Média

A média dos  $n$  resíduos  $e_i$ 's é

$$\bar{e} = \frac{1}{n} \sum_{i=1}^n e_i = 0.$$

- Propriedades do Resíduo

- Variância

A variância estimada dos  $n$  resíduos  $e_i$ 's é

$$S_e^2 = \frac{1}{n-p} \sum_{i=1}^n (e_i - \bar{e})^2 = \frac{1}{n-p} \sum_{i=1}^n e_i^2 = \text{MSE}.$$

- Note que  $\hat{\mathbf{Y}} = \mathbf{X}\beta = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} = \mathbf{H}\mathbf{Y}$  e  $\text{Var}(\hat{\mathbf{Y}}) = \sigma^2\mathbf{H}$ . Portanto temos,

$$\text{Var}(\hat{Y}_i) = \sigma^2 h_{ii} \text{ e } \text{Cov}(\hat{Y}_i, \hat{Y}_j) = \sigma^2 h_{ij} \text{ para } i \neq j.$$

- Uma nota interessante: Porque  $\hat{\mathbf{Y}} = \mathbf{H}\mathbf{Y}$ , temos

$$\hat{Y}_i = h_{i1} Y_1 + h_{i2} Y_2 + \cdots + h_{in} Y_n,$$

equivalente

$$\hat{Y}_i = b_0 + b_1 X_{i1} + b_2 X_{i2} + \cdots + b_k X_{ik}.$$

- Como sabemos os resíduos,  $e_i$ 's, não são independentes e podem ser visto como um resultado das restrições que os resíduos estão sujeitos:

$$\sum_{i=1}^n e_i = 0 \text{ (pelo intercepto),}$$

$$\sum_{i=1}^n X_{ij}e_i = 0 \text{ (pelo } j\text{-ésimo covariável } X_j \text{ para } j = 1, 2, \dots, k).$$

Uma consequência imediata das  $k + 1$  restrições é

$$\sum_{i=1}^n e_i \hat{Y}_i = 0.$$

como no modelo linear simples.

- O desvio padrão estimados do resíduo  $e_i$  é dado por

$$se(e_i) = \sqrt{\text{MSE}(1 - h_{ii})}.$$

# Características dos Modelos que podem ser estudadas nos Resíduos

- (1) A função de regressão não é linear.
- (2) Os erros não possuem variância constante.
- (3) Os erros não são independentes.
- (4) Os modelo ajusta a maioria dos dados ao menos de alguns outliers.
- (5) Os erros não possuem distribuição normal.

# Diagnósticos obtidos com os Resíduos

- Gráfico dos resíduos contra os preditores ou covariáveis (linearidade).
- Gráfico dos resíduos contra os valores ajustados (constância na variância e linearidade).
- Gráfico dos resíduos contra o tempo (independência).
- Box plot dos resíduos (outlier).
- Normal probability plot dos resíduos (normalidade).

- Como as inferências da análise dependem fortemente da suposição de normalidade, testes de normalidade devem ser usados
  - Um Q-Qplot para verificar normalidade
  - Testes padrão de normalidade, e.x. **Shapiro-Wilk Test**, **Anderson-Darling Test**, e **Kolmogorov-Smirnoff test**.
- Se a normalidade é violada, uma possível solução é a transformação da resposta.
- Gráficos Básicos dos Resíduos
  - (i) Um gráfico de matriz entre o  $Y_i$  e cada preditora  $X_{ij}$  é útil para verificar linearidade.
  - (ii) Um gráfico do resíduo  $e_i$  versus  $\hat{Y}_i$  e um gráfico do resíduo  $e_i$  versus cada  $X_{ij}$  podem também ser utilizado para verificar hipótese linearidade.

- Lembrando, temos que  $e_i \sim N(0, \sigma^2(1 - h_{ii}))$

Assim

$$\frac{e_i}{\sigma\sqrt{1 - h_{ii}}} \sim N(0, 1)$$

- Resíduo Interno Studentizado
  - O *resíduo interno studentizado* é definido como

$$r_i = \frac{e_i}{se(e_i)} = \frac{e_i}{\hat{\sigma}\sqrt{1 - h_{ii}}} = \frac{e_i}{\sqrt{MSE(1 - h_{ii})}}$$

- Os  $r_i$  também são correlacionados uns com os outros. Em particular,

$$\text{Cov}(r_i, r_j) = \frac{-h_{ij}}{\sqrt{(1 - h_{ii})(1 - h_{jj})}}$$

- Como  $e_i$  e MSE são dependentes,  $r_i$  não possui distribuição- $t$ .
- Sob a suposição de linearidade e normalidade, temos (*Cook and Weisberg, 1982*)

$$\frac{r_i^2}{n-p} \sim \text{beta}\left(\frac{1}{2}, \frac{n-p}{2}\right).$$

Assim,

$$|r_i| \leq n-p.$$

- Um gráfico de  $r_i$  versus  $\hat{Y}_i$  ainda é útil para detectar discrepâncias das hipóteses do modelo. Espera-se ver um padrão aleatório e  $|r_i| \leq 3$ .

# Resíduos Deletados Padronizados (Resíduos Externo Studentizado)

- Para remover a dependência entre  $e_i$  e MSE, omitimos a  $i$ -ésima observação no cálculo do MSE.
- A idéia por trás é, para cada  $i = 1, 2, \dots, n$ , ajuste o modelo de regressão com todos os dados removendo a  $i$ -ésima observação, e deixe  $MSE_{(i)}$  ser o MSE resultante.
- Assim,  $e_i$  é independente de  $MSE_{(i)}$ .
- Seja

$$t_i = \frac{e_i}{\sqrt{MSE_{(i)}(1 - h_{ii})}},$$

chamado de *resíduo deletado padronizado*.

- Podemos mostrar que  $t_i$  possui distribuição  $t_{n-1-p}$ , i.e.,

$$t_i \sim t_{n-1-p}.$$

- Matematicamente, temos

$$(n-p)\text{MSE} = (n-p-1)\text{MSE}_{(i)} + \frac{e_i^2}{1-h_{ii}}$$

e a relação

$$t_i = e_i \left[ \frac{n-p-1}{\text{SSE}(1-h_{ii}) - e_i^2} \right]^{1/2}.$$

Dessa forma, na prática, para calcular  $t_i$  não é necessário re-ajustar o MLE.

- Um gráfico do *resíduo deletado padronizado* versus  $\hat{Y}_i$  ou um box plot dos *resíduos deletados padronizados* é muito efetivo para detectar outliers.
- As observações no extremos das cauda da distribuição  $t_{n-p-1}$  são possíveis outliers. Uma regra é dizer que a  $i$ -ésima observação é um possível outlier se

$$|t_i| \geq 2.$$

- Pode se mostrar que  $h_{ii}$  é uma medida de distância entre os valores de  $X$  da  $i$ -ésima observação e a média dos valores de  $X$  para todos os  $n$  casos.
- Um grande valor de  $h_{ii}$  indica que  $i$ -ésimo caso é distante do centro das observações de  $X$ , ou seja,  $X_i$  é um outlier em relação a  $X$
- Os elementos das diagonais  $h_{ii}$  é chamado de ponto de *alavanca* (em termos dos valores de  $X$ ) para o  $i$ -ésimo caso.

- Regra básica: Se

$$h_{ii} \geq \frac{2p}{n},$$

onde  $p = k + 1$ , o  $i$ -ésimo caso é considerado um outlier em relação aos valores de  $X$

- Outra regra sugerida:  
valores de  $h_{ii}$  excedendo 0.5 indica forte ponto de alavanca, enquanto valores entre 0.2 e 0.5 indicam ponto de alavanca moderado.

- Lembre-se que

$$\hat{Y}_i = h_{ii} Y_i + \sum_{j \neq i} h_{ij} Y_j,$$

portanto, maior  $h_{ii}$ , mais importante o valor de  $Y_i$  é para determinar  $\hat{Y}_i$ .

- Outra observação é que, maior  $h_{ii}$ , menor a variância do resíduo  $e_i$ . ( $\text{Var}(e_i) = \sigma^2(1 - h_{ii})$ .)
- Quando  $h_{ii} = 1$ , podemos mostrar que  $h_{ij} = 0$  for  $j \neq i$ . Então, nesse caso,  $\hat{Y}_i = Y_i$ , indicando  $e_i = 0$ , e  $\text{Var}(e_i) = 0$ .

**Questão 1:** Casos considerados pontos de alavanca fortes pode ser um outlier em termos das observações  $Y$ ?

**Questão 2:** Podemos pensar num exemplo gráfico no qual um ponto de alavanca forte é um outlier em termos da observação  $Y$ , respectivamente?

- Observação Influente

Após identificar observações que são outliers com respeito aos valores de  $Y$  e/ou valores de  $X$ , o próximo passo é determinar se essas observações são ou não pontos influentes.

Consideramos uma observação influente se a exclusão dessa observação causa uma grande mudança no ajuste da função regressão.

- Definição

$$\text{DFFITS}_i = \frac{\hat{Y}_i - \hat{Y}_{i(i)}}{\sqrt{\text{MSE}_{(i)} h_{ii}}},$$

onde  $\hat{Y}_{i(i)}$  denota o valor predito para  $i$ -ésima observação obtida ao deletar a  $i$ -ésima observação para ajustar o modelo de regressão e  $\text{MSE}_{(i)}$  é o MSE obtido ao deletar a  $i$ -ésima observação do modelo de regressão.

- Calculo

$$\text{DFFITS}_i = t_i \sqrt{\frac{h_{ii}}{1 - h_{ii}}},$$

onde  $t_i$  é o resíduo deletado padronizado.

- Regra básica:

Se  $|\text{DFFITS}_i| \geq 1$  para bancos pequenos ou médios e  $|\text{DFFITS}_i| \geq 2\sqrt{\frac{p}{n}}$  ( $p = k + 1$ ) para bancos grandes, a  $i$ -ésima observação é considerada influente.

- Formulação Matemática

Seja

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

Também seja

$$\hat{\beta}_{(i)} = ((\mathbf{X}_{(i)})'\mathbf{X}_{(i)})^{-1}(\mathbf{X}_{(i)})'\mathbf{Y}_{(i)},$$

onde  $\mathbf{X}_{(i)}$  é  $\mathbf{X}$  com a  $i$ -ésima linha deletada, e  $\mathbf{Y}_{(i)}$  é  $\mathbf{Y}$  com o  $i$ -ésimo elemento deletado.

O objetivo é comparar  $\hat{\beta}$  com  $\hat{\beta}_{(i)}$  para determinar se a  $i$ -ésima observação é influente.

- Cook (1972) propôs que a influência da  $i$ -ésima observação deve ser medida em função da distância quadrática

$$D_i = (\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})'(\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)})/(\rho\text{MSE}),$$

onde  $\hat{\mathbf{Y}} = \mathbf{X}\hat{\boldsymbol{\beta}}$ ,  $\hat{\mathbf{Y}}_{(i)} = \mathbf{X}\hat{\boldsymbol{\beta}}_{(i)}$ ,  $\rho = \dim(\boldsymbol{\beta}) = k + 1$ , e o MSE é obtido pela regressão com todas observações.

- Como  $\hat{\mathbf{Y}} - \hat{\mathbf{Y}}_{(i)} = \mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})$ , também pode-se escrever

$$D_i = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})'\mathbf{X}'\mathbf{X}(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}}_{(i)})/(\rho\text{MSE}).$$

- Cálculo do  $D_i$

Pode-se mostrar que

$$D_i = \left( \frac{e_i}{\sqrt{\text{MSE}(1 - h_{ii})}} \right)^2 \left( \frac{h_{ii}}{1 - h_{ii}} \right) \cdot \frac{1}{p} = \frac{r_i^2}{p} \cdot \frac{h_{ii}}{1 - h_{ii}},$$

onde  $r_i$  é o resíduo interno studentizado.

- Magnitude de  $D_i$

Observações com grandes valores de  $D_i$  são aqueles nos quais exclusão irão resultar em grandes mudanças na análise.

- Regra básica

Se  $D_i \geq 0.8$ , a  $i$ -ésima observação será considerada influente.

- Uma medida de influência para a  $i$ -ésima observação em cada coeficiente de regressão  $\hat{\beta}_j$  ( $j = 0, 1, 2, \dots, k$ ) é dado pela diferença nas estimativas entre o coeficiente de regressão  $\hat{\beta}_k$  baseado em todas as  $n$  observações e o coeficiente de regressão obtido quando a  $i$ -ésima observação é removida, denotado por  $\hat{\beta}_{j(i)}$ .
- A medida DFBETAS é definida como

$$\text{DFBETAS}_{j(i)} = \frac{\hat{\beta}_j - \hat{\beta}_{j(i)}}{\sqrt{\text{MSE}_{(i)} c_{jj}}},$$

onde  $c_{jj}$  é o  $j$ -ésimo elemento diagonal da matriz  $(\mathbf{X}'\mathbf{X})^{-1}$ .

- Note que  $\text{Var}(\hat{\beta}) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}$  e, portanto, a variância de  $\hat{\beta}_j$  é

$$\text{Var}(\hat{\beta}_j) = \sigma^2 c_{jj}.$$

- Apesar da fórmula não ser tão simples como a apresentada para os  $DFFITs_i$ , os  $DFBETAS_{j(i)}$  também podem ser calculados através dos resultados obtidos na regressão original (Belsley, D.A., Kuh, E., and Welsch, R.E. (1980)).
- Regra básica  
A  $i$ -ésima é considerada ser influene se

$$|DFBETAS_{j(i)}| \geq \begin{cases} 1 & \text{para bancos pequenos ou médios} \\ \frac{2}{\sqrt{n}} & \text{pra bancos grandes.} \end{cases}$$

# Influência na matriz de Covariância – COVRATIO

- A estatística COVRATIO mede a mudança no determinante na matriz de covariância das estimativas deletando a  $i$ -ésima observação

$$\text{COVRATIO} = \frac{\det\left(\text{MSE}_{(i)}((\mathbf{X}(i))'\mathbf{X}(i))^{-1}\right)}{\det\left(\text{MSE}(\mathbf{X}'\mathbf{X})^{-1}\right)},$$

onde  $\mathbf{X}(i)$  é  $\mathbf{X}$  com a  $i$ -ésima linha deletada, e  $\text{MSE}_{(i)}$  é o MSE obtido quando a  $i$ -ésima observação é omitida para ajustar o modelo de regressão.

- Cálculo

$$\text{COVRATIO} = \left[ \left( \frac{n-p-1}{n-p} + \frac{t_i}{n-p} \right)^k (1 - h_{ii}) \right]^{-1}.$$

- O ponto de corte  
Belsley, Kuh, and Welsch sugerem que observações com

$$|\text{COVRATIO} - 1| \geq \frac{3p}{n},$$

onde  $p = k + 1$  e  $n$  é o numero de observações utilizadas para ajustar o modelo, são passíveis de investigação.

- Referência  
Belsley, D.A., Kuh, E., and Welsch, R.E. (1980), Regression Diagnostics, New York: John Wiley & Sons, Inc.

# Remédios para Corrigir o Problema de Pontos Influentes

- (1) Fazer transformações.
- (2) Mudar a distribuição do erro.
- (3) Aumentar o modelo adicionando covariáveis, termos de interação e termos quadráticos.
- (4) Adicionar observações para aumentar o poder do ajuste.

Pontos influentes comumente são causados por outliers nas covariáveis, e portanto transformações nas covariáveis e/ou resposta comumente corrigem esse problema.

- O que é colinearidade?
  - Suponha que temos apenas dois preditores  $X_1$  e  $X_2$ . Então  $X_1$  e  $X_2$  são ditos colineares se existe constantes  $c_0$ ,  $c_1$  e  $c_2$  tal queremos

$$c_1 X_1 + c_2 X_2 = c_0$$

e aproximadamente colineares se

$$c_1 X_1 + c_2 X_2 \approx c_0$$

- Essa idéia pode ser estendido no caso que temos  $X_1, \dots, X_k$  preditores. Logo,  $X_1, \dots, X_k$  são ditos colineares se temos

$$c_1 X_1 + \dots + c_k X_k = c_0$$

e aproximadamente

$$c_1 X_1 + \dots + c_k X_k \approx c_0$$

- Seja o modelo linear  $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$ . Então o modelo é dito colinear ou multicolinear se pelo menos uma coluna de  $\mathbf{X}$  é combinação linear das outras.
- Matematicamente temos que a colinearidade ocorre se

$$\det(\mathbf{X}'\mathbf{X}) = 0$$

- Na prática tratamos o problema de multicolinearidade quando

$$\det(\mathbf{X}'\mathbf{X}) \approx 0$$

# Qual o problema de multicolinearidade?

- Vamos pensar no modelo simples com 2 covariáveis,  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon$  com  $E(\varepsilon) = 0$  e  $\text{Var}(\varepsilon) = \sigma^2$ .
- Seja  $r_{12}$  a correlação amostral de Pearson  $X_1$  and  $X_2$ , e seja  $S_{X_j}^2$  a variância amostral de  $X_j$ . Portanto, se  $\hat{\beta}_j$  é o estimador de mínimos quadrados de  $\beta_j$ . Pode se mostrar que

$$\text{Var}(\hat{\beta}_j) = \sigma^2 \left( \frac{1}{1 - r_{12}^2} \right) \left( \frac{1}{(n-1)S_{X_j}^2} \right)$$

for  $j = 1, 2$ . Assim,

$$\text{Var}(\hat{\beta}_j) \rightarrow \infty \text{ if } r_{12}^2 \rightarrow 1,$$

e  $\text{Var}(\hat{\beta}_j)$  é minimizada quando  $r_{12} = 0$ .

- No caso geral, considere  $Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k + \varepsilon$  e  $k > 2$ . Seja  $\hat{\beta}_j$  o estimador de mínimos quadrados de  $\beta_j$ . Então

$$\text{Var}(\hat{\beta}_j) = \sigma^2 \left( \frac{1}{1 - R_j^2} \right) \left( \frac{1}{(n-1)S_{X_j}^2} \right)$$

onde  $R_j^2$  é o coeficiente de determinação múltipla  $R^2$  quando  $X_j$  é regressado nos outros  $k - 1$   $X$ 's para  $j = 1, 2, \dots, k$ .

- Logo, quando  $X_j$  é aproximadamente linearmente dependente dos outros  $X$ 's,  $R_j^2$  é perto de 1. Nesse caso,  $\text{Var}(\hat{\beta}_j)$  fica extremamente grande.

- Lembre-se que o estimador de mínimos quadrados é da forma

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$$

com variância

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

Assim, a multicolinearidade gera um matriz  $\mathbf{X}'\mathbf{X}$  aproximadamente singular, o que pode causar estimativas poucos precisas de  $\hat{\beta}$  e uma grande variância em  $\hat{\beta}$ .

- Logo, um teste  $t$  para testar

$$H_0 : \beta_j = 0$$

pode ser enganatório.

- Se alguns dos coeficientes forem estimados com o sinal trocado, baseando em algum conhecimento prévio
- Preditores conhecidos de serem importantes (através de conhecimento a priori) possuem uma estatística  $t$  pequena
- Remoção de uma linha ou coluna da matriz  $\mathbf{X}$  matrix produz uma mudança surpreendente no modelo ajustado
- A correlação entre pares de preditores é surpreendentemente grande

# Fator de Inflação da Variância

- O fator de inflação da variância (VIF) para o  $j$ -ésimo preditor  $X_j$  é definido como

$$\text{VIF}_j = \frac{1}{1 - R_j^2},$$

onde  $R_j^2$  = é o coeficiente de determinação múltipla  $R^2$  quando  $X_j$  é regressado nos outros  $k - 1$   $X$ 's, para  $j = 1, 2, \dots, k$ .

- Medindo colinearidade:

VIF  $\geq 10$  é uma indicação de multicolinearidade severa

VIF  $> 5$  deve ser investigado.

- Centrar e/ou escalonar e/ou padronizar empiricamente os preditores, ou seja,

Seja  $X_{1j}, X_{2j}, \dots, X_{nj}$  os valores observados de  $X_j$  para cada indivíduo. Defina  $\bar{X}_j$  and  $s_{X_j}$  como a média amostral e o desvio padrão amostral dos  $X_{ij}$ 's.

Centrar:  $X_{ij} - \bar{X}_j$ ;

Escalonar:  $X_{ij}/s_{X_j}$ ;

Padronizar:  $(X_{ij} - \bar{X}_j)/s_{X_j}$ .

Essa técnica é particularmente útil para tirar a correlação entre os preditores e o intercepto, e também é muito útil em modelos que contém interação de alta ordem (interações maiores do que quadrática).

- Remover um ou vários preditores (normalmente aquelas que possuem alto valores de VIF) do modelo
- Combinar alguns preditores lineares.
- Quebrar o padrão de multicolinearidade adicionando novas observações, i.e., adicionando mais dados
- Usando algum tipo de regressão penalizada:
  - ridge regression
  - lasso regression

- Iremos apresentar teste de hipótese para funções lineares de  $\beta$

$$H : C'\beta = d$$

é chamado de hipótese linear geral, na qual  $C_{p \times s}$  é uma matriz de posto  $s$  e  $d = (d_1, \dots, d_s)'$  é um vetor de constantes.

- Essas restrições também podem ser vistas como:

$$H : c_1'\beta = d_1, \dots, c_s'\beta = d_s$$

- O  $r(C)$  deve ser  $s$  para que nenhum teste  $H$  seja redundante, ou seja, possa ser escrito como combinação linear de outros.

- Um hipótese é dita testável se  $C'\beta$  é escrito em termos de funções estimáveis de  $\beta$ .
- Logo, para  $C'\beta$  ser válido temos que  $C$  deve satisfazer a condição  $T'X = C'$  ou  $C'W = C'$ .
- Sob  $H_0$  devemos obter a solução do modelo restrito

$$\mathbf{X}'\mathbf{X}\tilde{\beta}_H + C\lambda = \mathbf{X}'y, \quad C'\tilde{\beta}_H = d$$

que é dada por

$$\tilde{\beta}_H = \tilde{\beta} - GC(C'GC)^{-1}(C\tilde{\beta} - d)$$

- Suponha que um experimento tenha resposta  $Y$  e dois preditores  $X_1$  e  $X_2$ , obtendo um banco de dados  $\{(Y_i, X_{1i}, X_{2i}), i = 1, 2, \dots, n\}$ . O cientista calcula que, apesar de ambos  $X_1$  e  $X_2$  afetarem  $Y$ , o verdadeiro preditor de interesse é na verdade a diferença  $X_1 - X_2$ .
- Se ambos  $X_1$  e  $X_2$  são necessários, inicialmente ajusta-se o modelo

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \varepsilon,$$

- mas, caso sua suspeita esteja correta, o modelo deve ser

$$Y = \beta_0 + \beta(X_1 - X_2) + \varepsilon$$

Como podemos checar?

- Matematicamente, a pergunta verdadeiramente é: “Pode ser verdade que  $\beta_1 = -\beta_2 = \beta$ ?”
- Logo, gostaríamos de testar a hipótese nula  $H_0 : \beta_1 + \beta_2 = 0$  versus a hipótese alternativa  $H_1 : \beta_1 + \beta_2 \neq 0$ . No formato de matriz, teremos que  $H_0$  pode ser escrito como

$$(0 \quad 1 \quad -1) \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = 0.$$

- Seja o modelo:  $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2$ .
- Se queremos testar:  $H_0 : \beta_1 = \beta_2 = 0$  (duas funções lineares independentes).  
(Por “independente”, queremos dizer linearmente independente, ou seja, uma restrição não é combinação linear da outra.)
- Matematicamente, podemos escrever  $H_0$  na forma matricial como

$$\begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = 0.$$

- Seja:  $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_5 X_5$ .

$$H_0 : \beta_3 = 0,$$

$$\beta_4 = 0,$$

$$\beta_5 = 0 \quad (3 \text{ funções lineares independentes}).$$

Nesse caso estamos testando de algum  $\beta_j = 0$ .

- Seja:  $E(Y) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_k X_k$ .

$$H_0 : \beta_1 - \beta_2 = 0,$$

$$\beta_2 - \beta_3 = 0,$$

$$\vdots$$

$$\beta_{k-1} - \beta_k = 0 \quad (k - 1 \text{ funções lineares independentes}).$$

Veja que nesse caso estamos testando se  $H_0$ :

$$\beta_1 = \beta_2 = \dots = \beta_k = \beta.$$

# Testando a hipótese linear geral $\mathbb{C}\beta = \mathbf{0}$

- Suponha que o modelo em consideração seja

$$E(\mathbf{Y}) = \mathbf{X}\beta,$$

onde  $\mathbf{Y}$  é  $(n \times 1)$ ,  $\mathbf{X}$  é  $(n \times p)$  com  $p = k + 1$ , e  $\beta$  é  $(p \times 1)$ . Se  $\mathbf{X}'\mathbf{X}$  é não singular, i.e.,  $(\mathbf{X}'\mathbf{X})^{-1}$  existe, temos

$$\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}.$$

- E a soma quadrática dos erros é

$$\text{SSE} = \mathbf{Y}'\mathbf{Y} - \hat{\beta}'\mathbf{X}'\mathbf{Y}$$

com  $(n - p)$  graus de liberdade.

- Suponha que queremos testar a hipótese linear:

$$H_0 : C\beta = \mathbf{0},$$

onde  $C$  gere  $q$  condições independentes nos parâmetros  $\beta_0, \beta_1, \beta_2, \dots, \beta_k$ , onde  $k = p - 1$ .

- Como agora temos  $q$  restrições vindas de  $C$ . Podemos ajustar o modelo reduzido da forma:

$$E(\mathbf{Y}) = \mathbf{Z}\alpha,$$

onde é o modelo reduzido e  $\alpha$  é o vetor de parâmetros a ser estimado,  $\mathbf{Z}$  é  $n \times (p - q)$ , e  $\alpha$  é  $(p - q) \times 1$ .

- O estimado de  $\alpha$  é dado por

$$\hat{\alpha} = (\mathbf{Z}'\mathbf{Z})^{-1}\mathbf{Z}'\mathbf{Y},$$

dado a existência de  $(\mathbf{Z}'\mathbf{Z})^{-1}$ .

- Agora a nova soma de quadrado dos erros é

$$\text{SSW} = \mathbf{Y}'\mathbf{Y} - \hat{\alpha}'\mathbf{Z}'\mathbf{Y},$$

com  $(n - p + q)$  graus de liberdade.

A diferença entre  $\text{SSW} - \text{SSE}$  é chamada de soma dos quadrados da hipótese:  $H_0: C\beta = \mathbf{0}$  e possui  $(n - p + q) - (n - p) = q$  graus de liberdade.

- Sob a suposição usual de normalidade pode se mostrar que a razão

$$F = \left( \frac{\text{SSW} - \text{SSE}}{q} \right) / \left( \frac{\text{SSE}}{n - p} \right) \sim F_{q, n-p},$$

se  $H_0$  é verdadeira.

- Essa estatística- $F$  pode então ser usada para testar se  $H_0$  é verdadeira.

- Matematicamente, não é necessário ajustar ambos os modelos. Pode se mostrar que

$$\mathbf{Q} = \text{SSW} - \text{SSE} = (C\hat{\beta})'(C(\mathbf{X}'\mathbf{X})^{-1}C')^{-1}(C\hat{\beta}).$$

- No caso geral, quando  $H_0: C\beta = d$ , temos

$$\mathbf{Q} = \text{SSW} - \text{SSE} = (C\hat{\beta} - d)'(C(\mathbf{X}'\mathbf{X})^{-1}C')^{-1}(C\hat{\beta} - d)$$

- Logo,

$$F = \left( \frac{\mathbf{Q}}{q} \right) / \left( \frac{\text{SSE}}{n-p} \right) \sim F_{q, n-p},$$

se  $H_0$  é verdade.