

**Modelos  
de  
regressão multivariada**

**Fábio Esteves Nogueira**

DISSERTAÇÃO APRESENTADA  
AO  
INSTITUTO DE MATEMÁTICA E ESTATÍSTICA  
DA  
UNIVERSIDADE DE SÃO PAULO  
PARA  
OBTENÇÃO DO TÍTULO DE MESTRE  
EM  
CIÊNCIAS

Área de Concentração: **Estatística**  
Orientador: **Profa. Dra. Silvia Nagib Elian**

- São Paulo, fevereiro de 2007 -

**Modelos  
de  
regressão multivariada**

Este exemplar corresponde à redação final da dissertação de mestrado devidamente corrigida e defendida por Fábio Esteves Nogueira e aprovada pela comissão Julgadora.

São Paulo, maio de 2007.

Banca examinadora:

Profa. Dra. Silvia Nagib Elian (orientadora) – IME/USP

Profa. Dra. Carmen Diva Saldiva de André – IME/USP

Prof. Dr. Francisco José de Azevedo Cysneiros – UFPE

*“Impossível é apenas uma grande palavra usada por gente fraca que  
prefere viver no mundo como está  
em vez de usar o poder que tem para mudá-lo.  
Impossível não é um fato. É uma opinião.  
Impossível não é uma declaração. É um desafio.  
Impossível é hipotético. Impossível é temporário.  
Nada é Impossível”*

Mohamed Ali

*À Maria, minha mãe*

# Agradecimentos

---

Essa dissertação não é fruto do trabalho árduo de uma pessoa, na verdade é fruto do trabalho árduo de muitas pessoas. Agradeço primeiro aos meus amados pais que tanto se sacrificaram para que eu pudesse chegar até esse momento. Conteí com o apoio incondicional deles, sobretudo de minha mãe que (fora do combinado) foi morrer no céu poucos dias antes do término desse trabalho. Amo muito vocês.

Agradeço à minha orientadora, professora Silvia Nagib Elian, por ter acreditado e confiado em mim. Obrigado pela paciência, compreensão, motivação e pela solidariedade nos momentos de dificuldade. Não me arrependi em nenhum momento dessa escolha e tenho muito orgulho de ter sido orientado por você.

Ao meu grande amigo e companheiro de tantas batalhas, Ricardo Olivare de Magalhães, seu bom-humor, incentivo, empolgação e paciência foram fundamentais em diversos momentos. Tenho profunda admiração por você, muito obrigado. Não poderia esquecer de agradecer à Claudia Shigueko Suzuki (a Claudinha) por tudo o que sempre fez por mim, não teria chegado até aqui sem sua amizade. Você é mesmo muito especial.

À Paula Stefanoni Iwamizu, pelo carinho, pela atenção, pela amizade, por sua ajuda, por nunca ter se esquecido de mim e sobretudo pelo exemplo de caráter e integridade.

Devo agradecer aos professores do IME que me apresentaram esse mundo fantástico da estatística, que me fizeram evoluir como estudante e como profissional. Finalmente agradeço aos colegas João Ricardo Sato, Igor Quidim, Luis Gustavo do Amaral Vinha e Renata Aguemí, passamos juntos por muitos obstáculos. Valeu galera!

Por fim, agradeço a Deus.

# Resumo

---

Os modelos de Regressão Linear Multivariada apesar de serem pouco utilizados são muito úteis pois, dentre outras vantagens, permitem a construção de modelos considerando estruturas de correlação entre medidas tomadas na mesma ou em distintas unidades amostrais. Neste trabalho apresentamos os métodos de estimação dos parâmetros, medidas para análise de diagnóstico, procedimentos de seleção de variáveis e uma aplicação dessa técnica de modelagem em um conjunto de dados reais.

# Abstract

---

Multivariate Linear Regression Models are not frequently used although they are very useful. Working with this kind of model, it is possible to analyse correlated response variables jointly. In this dissertation, we dedicate initially to describe the inferencial methods in Multivariate Linear Regression models.

Further, we describe some measures of diagnostics and methods of variable selection in this model. Finally, some of the describe procedures are applied in a real data set.

# Sumário

<b>1</b>	<b>Introdução</b>	<b>3</b>
<b>2</b>	<b>Modelos de Regressão Multivariada</b>	<b>5</b>
2.1	Definição do Modelo . . . . .	5
2.2	Método de Estimação dos Parâmetros . . . . .	8
2.2.1	Estimador de Mínimos Quadrados de $\beta$ . . . . .	8
2.2.2	Estimação de $\beta$ quando $X_1 = X_2 = \dots = X_p = X^*$ e $\Omega = \Sigma \otimes I_n$ . . . . .	8
2.3	Estimação da Matriz de Variância e Covariância $\Sigma$ . . . . .	11
2.4	Teste de Hipóteses para os Parâmetros do Modelo de Regressão Multivariada . .	14
2.5	Predição através o modelo de Regressão Multivariada . . . . .	17
2.6	Estimação do vetor $\beta$ quando $X_1 \neq X_2 \neq \dots \neq X_p$ . . . . .	18
2.7	Ganho em Eficiência com a Estimação Conjunta . . . . .	21
2.8	Propriedades do Estimador em Dois Estágios . . . . .	22
<b>3</b>	<b>Análise de Diagnóstico</b>	<b>24</b>
3.1	Classes Gerais de Medidas de Influência . . . . .	24



3.2	Medidas da Classe $J_I^{tr}$ . . . . .	26
3.3	Medidas na Classe $J^{det}$ . . . . .	28
3.4	Alavanca e Componentes de Resíduos . . . . .	30
3.5	Exemplo 1 . . . . .	40
<b>4</b>	<b>Seleção de Variáveis</b>	<b>44</b>
4.1	Hipóteses para Seleção de Variáveis . . . . .	44
4.2	Testes Simultâneos . . . . .	46
4.3	Exemplo 2 . . . . .	50
4.4	Uso do Critério de Informação de Akaike . . . . .	52
<b>5</b>	<b>Aplicação</b>	<b>59</b>
<b>6</b>	<b>Conclusão</b>	<b>74</b>
<b>A</b>	<b>Programa de Cálculo do <math>AIC_c</math></b>	<b>76</b>
<b>B</b>	<b>Programas de Análise de Influência</b>	<b>78</b>
<b>C</b>	<b>Dados da Aplicação</b>	<b>84</b>
	<b>Referências Bibliográficas</b>	<b>87</b>

# Capítulo 1

## Introdução

As técnicas de Análise de Regressão Linear têm sido amplamente utilizadas tanto na literatura como em problemas práticos. Restringem-se predominantemente a modelos em que a variável resposta é única e está associada a um conjunto de variáveis preditoras, geralmente não aleatórias. Os livros de Análise de Regressão discutem detalhadamente todos os passos do ajuste e da posterior análise do modelo.

Muito menos, no entanto, tem sido apresentado quando se trata do ajuste de modelos de Regressão Linear Multivariada. Nessa situação, a variável resposta é multivariada, ou seja, é da forma  $\mathbf{Y} = (Y_1, \dots, Y_P)$ , sendo que as variáveis  $Y_1, \dots, Y_P$  são medidas na mesma unidade amostral e, como no caso univariado, associadas a  $P$  conjuntos de variáveis preditoras não aleatórias.

Ao contrário da Regressão Univariada, a análise de Regressão Multivariada é abordada apenas em capítulos de livros de Análise Multivariada de dados e sua utilização é limitada. O objetivo dessa dissertação é o de estudar com detalhes a técnica de Regressão Multivariada.

Apresentamos no Capítulo 2 uma análise dos procedimentos inferenciais relativos ao ajuste do modelo, à estimação de seus parâmetros e à construção dos testes de hipóteses e intervalos de confiança correspondentes.

No Capítulo 3 discutiremos medidas de diagnóstico associadas ao ajuste desse modelo. Tal assunto, exhaustivamente estudado no contexto de Regressão Univariada, é pouco conhecido em Regressão Multivariada. No Capítulo 4, estudaremos alguns processos de seleção de variáveis independentes. Encerramos o trabalho no Capítulo 5 com uma aplicação dos assuntos abordados nos capítulos anteriores.

# Capítulo 2

## Modelos de Regressão Multivariada

Neste capítulo serão apresentadas as definições do modelo de Regressão Multivariada e as técnicas de estimação dos parâmetros do modelo.

### 2.1 Definição do Modelo

Vamos supor o caso em que dispomos de  $p$  variáveis respostas

$$Y_1, \dots, Y_p.$$

A cada uma dessas  $p$  variáveis está associado um grupo de variáveis explicativas, então para a variável  $Y_1$  temos associado o grupo de variáveis explicativas  $X_{11}, \dots, X_{1K_1}$ . Procedendo dessa forma para as demais variáveis resposta,  $X_{p1}, \dots, X_{pK_p}$  é o conjunto de variáveis explicativas associadas à variável resposta  $Y_p$ .

Podemos definir, para cada uma das  $p$  regressões, um modelo linear geral (Graybill, 1976):

$$\mathbf{Y}_j = X_j \boldsymbol{\beta}_j + \boldsymbol{\varepsilon}_j, \quad j = 1, \dots, p.$$

Tomada uma amostra de  $n$  observações temos:

$\mathbf{Y}_j$  - vetor  $n \times 1$  contendo os valores da  $j$ -ésima variável resposta para os  $n$  elementos da amostra;

$X_j$  - matriz  $n \times (K_j + 1)$  de variáveis explicativas associada à  $j$ -ésima variável resposta;

$\boldsymbol{\beta}_j = [\beta_{0j}, \dots, \beta_{K_j}]'$  vetor de parâmetros  $(j = 1, \dots, p)$ .

Os elementos de  $\boldsymbol{\beta}_j$  podem ser estimados separadamente. No entanto, devido à correlação entre as  $p$  variáveis resposta, a análise conjunta se torna mais adequada.

Iniciaremos a análise conjunta, definindo os vetores e a matriz que irão compor o modelo. O vetor de variáveis resposta pode ser escrito da seguinte forma:

$$\mathbf{Y} = \begin{bmatrix} \mathbf{Y}_1 \\ \vdots \\ \mathbf{Y}_p \end{bmatrix} = \begin{bmatrix} Y_{11} \\ \vdots \\ Y_{1n} \\ \vdots \\ Y_{p1} \\ \vdots \\ Y_{pn} \end{bmatrix},$$

onde  $Y_{jk}$  é o  $k$ -ésimo valor da  $j$ -ésima variável resposta  $j = 1, \dots, p$  e  $k = 1, \dots, n$ .

A matriz de variáveis explicativas é dada por

$$X = \begin{bmatrix} X_1 & 0 & \dots & \dots & \dots & 0 \\ 0 & X_2 & \dots & \dots & \dots & 0 \\ 0 & 0 & X_3 & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & X_p \end{bmatrix}$$

em que  $X_1, \dots, X_p$  são as matrizes de planejamento associadas a cada uma das  $p$  regressões. A dimensão da matriz  $X$  é  $np \times (p + \sum_{i=1}^p K_i)$ .

O vetor de parâmetros pode ser escrito da seguinte forma:

$$\boldsymbol{\beta} = \begin{bmatrix} \boldsymbol{\beta}_1 \\ \vdots \\ \boldsymbol{\beta}_p \end{bmatrix} \text{ com } \boldsymbol{\beta}_j = \begin{bmatrix} \beta_{j0} \\ \vdots \\ \beta_{jK_j} \end{bmatrix} \quad j = 1, \dots, p$$

e o vetor de erros:

$$\boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_p \end{bmatrix} \text{ com } \boldsymbol{\varepsilon}_j = \begin{bmatrix} \varepsilon_{j1} \\ \vdots \\ \varepsilon_{jn} \end{bmatrix} \quad j = 1, \dots, p.$$

Por fim, a matriz de variância e covariância do vetor de erros é:

$$\text{Var}(\boldsymbol{\varepsilon}) = \begin{bmatrix} \text{Var}(\boldsymbol{\varepsilon}_1) & \dots & \dots & \dots & \text{Cov}(\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_p) \\ & \text{Var}(\boldsymbol{\varepsilon}_2) & \dots & \dots & \text{Cov}(\boldsymbol{\varepsilon}_2, \boldsymbol{\varepsilon}_p) \\ & & \ddots & \dots & \vdots \\ & & & \ddots & \vdots \\ & & & & \vdots \\ & & & & \text{Var}(\boldsymbol{\varepsilon}_p) \end{bmatrix} = \boldsymbol{\Omega}_{(np \times np)}.$$

Dessa forma, o modelo conjunto em notação matricial é dada por

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}. \quad (2.1)$$

## 2.2 Método de Estimação dos Parâmetros

Nesta seção vamos abordar a estimação do vetor  $\beta$  pelo método de mínimos quadrados.

### 2.2.1 Estimador de Mínimos Quadrados de $\beta$

Se  $\Omega$  é conhecido, o estimador de mínimos quadrados de  $\beta$  é:

$$\hat{\beta} = (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \mathbf{Y}. \quad (2.2)$$

Este estimador tem as seguintes propriedades:

- (i)  $\hat{\beta}$  é o melhor estimador linear não viesado de  $\beta$  se  $E(\varepsilon) = 0$  e  $Var(\varepsilon) = \Omega$ ;
- (ii)  $\hat{\beta}$  é não viesado uniformemente de variância mínima de  $\beta$  se  $\varepsilon \sim N_{np}(\mathbf{0}, \Omega)$ , além disso  $\hat{\beta}$  é o estimador de máxima verossimilhança de  $\beta$ .

O modelo de Regressão Multivariada descrito é bastante geral pois pode admitir:

- (i) Correlação entre medidas tomadas na mesma unidade amostral;
- (ii) Correlação entre medidas tomadas em unidades amostrais distintas.

### 2.2.2 Estimação de $\beta$ quando $X_1 = X_2 = \dots = X_p = X^*$ e $\Omega = \Sigma \otimes I_n$

Seja  $A \otimes B$  o produto de Kroneker entre as matrizes  $A$  e  $B$ . Quando as variáveis explicativas são as mesmas para as  $p$  regressões,  $Var(\varepsilon_j) = \sigma_j^2 I_n \quad \forall j = 1, \dots, p$  e  $Cov(\varepsilon_j, \varepsilon_l) = \sigma_{jl} I_n$  para

$j \neq l$ ,  $j, l = 1, 2, \dots, p$ , a matriz de planejamento é definida da seguinte maneira

$$X = \begin{bmatrix} X^* & 0 & \dots & \dots & \dots & 0 \\ 0 & X^* & \dots & \dots & \dots & 0 \\ 0 & 0 & X^* & \dots & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \dots & X^* \end{bmatrix} = I_p \otimes X^*$$

e  $\Omega = \Sigma \otimes I_n$  onde

$$\Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \dots & \dots & \sigma_{1p} \\ \sigma_{21} & \sigma_2^2 & \dots & \dots & \dots & \sigma_{2p} \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \sigma_{p1} & \sigma_{p2} & \dots & \dots & \dots & \sigma_p^2 \end{bmatrix}.$$

Observa-se que a dimensão de  $X^*$  é  $n \times K + 1$  pois  $K_1 = K_2 = \dots = K_p = K$ . Este padrão de matriz de covariância ocorre quando medidas tomadas em elementos amostrais distintos são não correlacionados e a covariância entre duas variáveis resposta medidas no mesmo elemento amostral é a mesma para todos os elementos.

Devido às propriedades do produto de Kroneker:

$$(A \otimes B)^{-1} = A^{-1} \otimes B^{-1};$$

$$(A \otimes B)(C \otimes D) = AC \otimes BD \quad \text{e}$$

$$(A \otimes B)' = A' \otimes B'$$

temos o seguinte resultado.

### Resultado:

A expressão 2.2 pode ser calculada como

$$\hat{\beta} = [I_p \otimes (X^{*'} X^*)^{-1} X^{*'}] \mathbf{Y}.$$



**Demonstração:**

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= (X' \Omega^{-1} X)^{-1} X' \Omega^{-1} \mathbf{Y} \\
&= [(I_p \otimes X^*)' (\Sigma \otimes I_n)^{-1} (I_p \otimes X^*)]^{-1} (I_p \otimes X^*)' (\Sigma \otimes I_n)^{-1} \mathbf{Y} \\
&= [(I_p \otimes X^*)' (\Sigma^{-1} \otimes I_n) (I_p \otimes X^*)]^{-1} (I_p \otimes X^*)' (\Sigma^{-1} \otimes I_n) \mathbf{Y} \\
&= [(\Sigma^{-1} \otimes X^{*'}) (I_p \otimes X^*)]^{-1} (\Sigma^{-1} \otimes X^{*'}) \mathbf{Y} \\
&= (\Sigma^{-1} \otimes X^{*'} X^*)^{-1} (\Sigma^{-1} \otimes X^{*'}) \mathbf{Y} \\
&= (\Sigma \otimes (X^{*'} X^*)^{-1}) (\Sigma^{-1} \otimes X^{*'}) \mathbf{Y} \\
\hat{\boldsymbol{\beta}} &= [I_p \otimes (X^{*'} X^*)^{-1} X^{*'}] \mathbf{Y}.
\end{aligned}$$

Nesse caso, como  $\hat{\boldsymbol{\beta}}$  é independente de  $\Sigma$ , podemos estimar  $\boldsymbol{\beta}$  sendo  $\Sigma$  conhecida ou não. Note que  $\hat{\boldsymbol{\beta}}$  coincide com o estimador de mínimos quadrados usual. Além disso,

$$\begin{aligned}
\hat{\boldsymbol{\beta}} &= [I_p \otimes (X^{*'} X^*)^{-1} X^{*'}] \mathbf{Y} \\
&= \begin{bmatrix} (X^{*'} X^*)^{-1} X^{*'} & 0 & \dots & 0 \\ 0 & (X^{*'} X^*)^{-1} X^{*'} & \dots & 0 \\ \vdots & \vdots & \vdots & \vdots \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & \dots & (X^{*'} X^*)^{-1} X^{*'} \end{bmatrix} \begin{bmatrix} \mathbf{Y}_1 \\ \mathbf{Y}_2 \\ \vdots \\ \vdots \\ \mathbf{Y}_p \end{bmatrix} \\
\hat{\boldsymbol{\beta}} &= \begin{bmatrix} (X^{*'} X^*)^{-1} X^{*'} \mathbf{Y}_1 \\ (X^{*'} X^*)^{-1} X^{*'} \mathbf{Y}_2 \\ \vdots \\ \vdots \\ (X^{*'} X^*)^{-1} X^{*'} \mathbf{Y}_p \end{bmatrix} = \begin{bmatrix} \hat{\boldsymbol{\beta}}_1 \\ \hat{\boldsymbol{\beta}}_2 \\ \vdots \\ \vdots \\ \hat{\boldsymbol{\beta}}_p \end{bmatrix}
\end{aligned}$$

com  $\hat{\boldsymbol{\beta}}_j = (X^{*'} X^*)^{-1} X^{*'} \mathbf{Y}_j \quad j=1, \dots, p$ .

Dessa forma, os estimadores de  $\boldsymbol{\beta}_j$ ,  $j = 1, \dots, p$ , determinados conjuntamente coincidem com os que seriam obtidos separadamente para cada variável resposta  $Y_j$ . Por esse motivo,

a situação é denominada de “Regressões Aparentemente não Correlacionadas”, denominação introduzida por Zellner (1962).

## 2.3 Estimação da Matriz de Variância e Covariância $\Sigma$

O caso mais comum no ajuste de modelos de Regressão Multivariada ocorre quando não conhecemos a matriz de variância e covariância  $\Sigma$ . Nesta seção iremos descrever o método de estimação de  $\Sigma$  apresentado em Johnson e Wichern (1998). Será aqui utilizada a notação introduzida pelos autores, em que o modelo (2.1) é escrito alternativamente como

$$Y_{(n \times p)} = X_{(n \times K+1)}^* \beta_{(K+1) \times p} + \varepsilon_{(n \times p)} \quad (2.3)$$

com

$$Y_{(n \times p)} = \begin{bmatrix} Y_{11} & \dots & Y_{1p} \\ \vdots & & \vdots \\ Y_{n1} & \dots & Y_{np} \end{bmatrix} = \left[ \mathbf{Y}_{(1)} \mid \dots \mid \mathbf{Y}_{(p)} \right]$$

$$\beta_{(K+1) \times p} = \begin{bmatrix} \beta_{01} & \dots & \beta_{0p} \\ \beta_{11} & \dots & \beta_{1p} \\ \vdots & & \vdots \\ \beta_{K1} & \dots & \beta_{Kp} \end{bmatrix} = \left[ \boldsymbol{\beta}_{(1)} \mid \dots \mid \boldsymbol{\beta}_{(p)} \right].$$

A matriz de planejamento será dada por:

$$X_{n \times (K+1)}^* = \begin{bmatrix} X_{10} & \dots & X_{1K} \\ \vdots & & \vdots \\ X_{n0} & \dots & X_{nK} \end{bmatrix}$$

e a matriz de erros é:

$$\varepsilon_{(n \times p)} = \begin{bmatrix} \varepsilon_{11} & \dots & \varepsilon_{1p} \\ \vdots & & \vdots \\ \varepsilon_{n1} & \dots & \varepsilon_{np} \end{bmatrix} = \left[ \varepsilon_{(1)} \mid \dots \mid \varepsilon_{(p)} \right].$$

A matriz  $Y$  pode ser escrita, alternativamente, como  $Y = [\mathbf{Y}_1, \dots, \mathbf{Y}_n]'$  sendo que os vetores  $\mathbf{Y}_j$  ( $j = 1, 2 \dots n$ ) correspondem às linhas de  $Y$  e são independentes,  $\mathbf{Y}_j' \sim N_p(\boldsymbol{\beta}' x_j, \Sigma)$ , com  $x_j$  sendo a  $j$ -ésima linha da matriz  $X^*$ .

Nestas condições:

$$\begin{aligned} (Y - X^* \boldsymbol{\beta})'(Y - X^* \boldsymbol{\beta}) &= (Y' - \boldsymbol{\beta}' X^{*'})'(Y' - \boldsymbol{\beta}' X^{*'})' \\ &= \begin{bmatrix} \mathbf{Y}_1 - \boldsymbol{\beta}' x_1 & \dots & \mathbf{Y}_n - \boldsymbol{\beta}' x_n \end{bmatrix} \begin{bmatrix} (\mathbf{Y}_1 - \boldsymbol{\beta}' x_1)' \\ \vdots \\ (\mathbf{Y}_n - \boldsymbol{\beta}' x_n)' \end{bmatrix} \\ &= \sum_{j=1}^n (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)(\mathbf{Y}_j - \boldsymbol{\beta}' x_j)' \end{aligned}$$

e

$$\sum_{j=1}^n (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)' \Sigma^{-1} (\mathbf{Y}_j - \boldsymbol{\beta}' x_j) = \sum_{j=1}^n \text{tr}[(\mathbf{Y}_j - \boldsymbol{\beta}' x_j)' \Sigma^{-1} (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)]$$

pois  $(\mathbf{Y}_j - \boldsymbol{\beta}' x_j)' \Sigma^{-1} (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)$  é uma matriz  $1 \times 1$ .

Por outro lado,

$$\begin{aligned} \sum_{j=1}^n \text{tr}[(\mathbf{Y}_j - \boldsymbol{\beta}' x_j)' \Sigma^{-1} (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)] &= \sum_{j=1}^n \text{tr}[\Sigma^{-1} (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)(\mathbf{Y}_j - \boldsymbol{\beta}' x_j)'] \\ &= \text{tr} \sum_{j=1}^n [\Sigma^{-1} (\mathbf{Y}_j - \boldsymbol{\beta}' x_j)(\mathbf{Y}_j - \boldsymbol{\beta}' x_j)'] \\ &= \text{tr}[\Sigma^{-1} (Y - X^* \boldsymbol{\beta})'(Y - X^* \boldsymbol{\beta})]. \end{aligned} \quad (2.4)$$

sendo que as duas primeiras igualdades devem-se às propriedades:

- $tr(AB) = tr(BA)$  sempre que  $AB$  e  $BA$  são produtos possíveis;
- $tr\left(\sum_{i=1}^m A_i\right) = \sum_{i=1}^m tr(A_i)$ ;

e finalmente, ao resultado

$$(Y - X^*\beta)'(Y - X^*\beta) = \sum_{j=1}^n (\mathbf{Y}_j - \beta'x_j)(\mathbf{Y}_j - \beta'x_j)' \text{ já demonstrado.}$$

Mas,

$$\begin{aligned} (\mathbf{Y} - X^*\beta)'(\mathbf{Y} - X^*\beta) &= [\mathbf{Y} - X^*\hat{\beta} + X^*(\hat{\beta} - \beta)]'[\mathbf{Y} - X^*\hat{\beta} + X^*(\hat{\beta} - \beta)] \\ &= (\mathbf{Y} - X^*\hat{\beta})'(\mathbf{Y} - X^*\hat{\beta}) + (\hat{\beta} - \beta)'X^{*'}X^*(\hat{\beta} - \beta) \\ &= \hat{\epsilon}'\hat{\epsilon} + (\hat{\beta} - \beta)'X^{*'}X^*(\hat{\beta} - \beta) \end{aligned} \quad (2.5)$$

onde  $\hat{\epsilon} = (\mathbf{Y} - X^*\hat{\beta})$ .

Então de (2.4) e (2.5) podemos escrever a função de verossimilhança do modelo de Regressão Linear Multivariada como:

$$\begin{aligned} L(\beta, \Sigma) &= \prod_{j=1}^p \frac{1}{(2\pi)^{n/2} |\Sigma|^{1/2}} \exp \left\{ -\frac{1}{2} (Y_j - \beta'X_j)' \Sigma^{-1} (Y_j - \beta'X_j) \right\} \\ &= C \frac{1}{|\Sigma|^{p/2}} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} (Y - X^*\beta)'(Y - X^*\beta)] \right\} \\ &= C \frac{1}{|\Sigma|^{p/2}} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} (\hat{\epsilon}'\hat{\epsilon} + (\hat{\beta} - \beta)'X^{*'}X^*(\hat{\beta} - \beta))] \right\} \\ L(\beta, \Sigma) &= C \frac{1}{|\Sigma|^{p/2}} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} (\hat{\epsilon}'\hat{\epsilon})] - \frac{1}{2} tr[X^*(\hat{\beta} - \beta)\Sigma^{-1}(\hat{\beta} - \beta)'X^{*'}] \right\} \end{aligned}$$

Nestas condições,

$$L(\hat{\beta}, \Sigma) = C \frac{1}{|\Sigma|^{p/2}} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} \hat{\epsilon}'\hat{\epsilon}] \right\}$$

em que  $C = \frac{1}{(2\pi)^{(pn)/2}}$ .

Como  $\Sigma$  é simétrica e positiva definida então, de acordo com Johnson e Wichern (1998, pg. 180),

$$\frac{1}{|\Sigma|^{p/2}} \exp \left\{ -\frac{1}{2} tr[\Sigma^{-1} \hat{\epsilon}'\hat{\epsilon}] \right\} \leq \left( \frac{1}{|\hat{\epsilon}'\hat{\epsilon}|} \right)^{p/2} \left( \frac{2n}{2} \right)^{(np)/2} \exp \left\{ -\frac{1}{2} np \right\}$$

logo

$$L(\hat{\beta}, \Sigma) = C \frac{1}{|\Sigma|^{p/2}} \exp \left\{ \frac{-1}{2} \text{tr}[\Sigma^{-1} \hat{\epsilon}' \hat{\epsilon}] \right\} \leq C \frac{1}{|\hat{\epsilon}' \hat{\epsilon}|^{p/2}} (n)^{\frac{np}{2}} \exp \left\{ \frac{-1}{2} np \right\} = L\left(\hat{\beta}, \frac{\hat{\epsilon}' \hat{\epsilon}}{n}\right)$$

Dessa forma o máximo de  $L(\hat{\beta}, \Sigma)$  será obtido quando

$$\Sigma = \frac{\hat{\epsilon}' \hat{\epsilon}}{n}$$

portanto o estimador de máxima verossimilhança de  $\Sigma$  é dado por  $\hat{\Sigma} = (\hat{\epsilon}' \hat{\epsilon})/n$ .

De acordo com Johnson e Wichern (1998), após o ajuste do modelo de Regressão Multivariada, uma análise de resíduos usual para cada modelo univariado deve ser feita. Adicionalmente, a normalidade do vetor de resíduos  $[\hat{\epsilon}_{j1}, \hat{\epsilon}_{j2}, \dots, \hat{\epsilon}_{jp}]$   $j = 1, 2, \dots, n$  deve ser avaliada através de técnicas de verificação de normalidade multivariada descritas na Seção 4.6 do referido texto.

## 2.4 Teste de Hipóteses para os Parâmetros do Modelo de Regressão Multivariada

Nesta seção apresentaremos o teste da razão de verossimilhanças para hipóteses relativas ao vetor de parâmetros  $\beta$ .

Utilizando novamente o modelo especificado em (2.3) quando  $X_1 = X_2 = \dots = X_p = X^*$  e:

$$Y_{(n \times p)} = \begin{bmatrix} Y_{11} & \dots & Y_{1p} \\ \vdots & & \vdots \\ Y_{n1} & \dots & Y_{np} \end{bmatrix} = \left[ Y_{(1)} \mid \dots \mid Y_{(p)} \right].$$

$$\beta_{(K+1) \times p} = \begin{bmatrix} \beta_{01} & \dots & \beta_{0p} \\ \beta_{11} & \dots & \beta_{1p} \\ \vdots & & \vdots \\ \beta_{K1} & \dots & \beta_{Kp} \end{bmatrix} = \left[ \beta_{(1)} \mid \dots \mid \beta_{(p)} \right].$$

A matriz de planejamento será dada por:

$$X_{n \times (K+1)}^* = \begin{bmatrix} X_{10} & \dots & X_{1K} \\ \vdots & & \vdots \\ X_{n0} & \dots & X_{nK} \end{bmatrix}.$$

e a matriz de erros é:

$$\varepsilon_{(n \times p)} = \begin{bmatrix} \varepsilon_{11} & \dots & \varepsilon_{1p} \\ \vdots & & \vdots \\ \varepsilon_{n1} & \dots & \varepsilon_{np} \end{bmatrix} = \left[ \varepsilon_{(1)} \mid \dots \mid \varepsilon_{(p)} \right].$$

Podemos particionar a matriz de parâmetros da seguinte maneira:

$$\beta = \begin{bmatrix} \beta_{(1)} \\ \text{-----} \\ \beta_{(2)} \end{bmatrix},$$

sendo que a dimensão de  $\beta_{(1)}$  é  $(q+1) \times p$  e a de  $\beta_{(2)}$  é  $(K-q) \times p$ .

De maneira análoga, a matriz  $X^*$  é particionada como

$$X^* = \left[ X_{(1)}^* \mid X_{(2)}^* \right]$$

de modo que dimensão de  $X_{(1)}^*$  é  $n \times (q+1)$  e a de  $X_{(2)}^*$  é  $n \times (K-q)$ .

Dessa forma, podemos escrever o modelo geral como:

$$E(Y) = X^* \beta = \begin{bmatrix} X_{(1)}^* & | & X_{(2)}^* \end{bmatrix} \begin{bmatrix} \beta_{(1)} \\ - - - - - \\ \beta_{(2)} \end{bmatrix} = X_{(1)}^* \beta_{(1)} + X_{(2)}^* \beta_{(2)}.$$

Vamos testar a hipótese:  $H_0 : \beta_{(2)} = 0$  e sob  $H_0$ , temos o modelo  $Y = X_{(1)}^* \beta_{(1)} + \varepsilon$ .

A estatística do teste da Razão de Verossimilhanças,  $\Lambda$ , pode ser expressa em termos das chamadas variâncias generalizadas:

$$\Lambda = \frac{\max_{\beta_{(1)} \Sigma} L(\beta_{(1)}, \Sigma)}{\max_{\beta \Sigma} L(\beta, \Sigma)} = \frac{L(\hat{\beta}_{(1)}, \hat{\Sigma}_{(1)})}{L(\hat{\beta}, \hat{\Sigma})} = \left( \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_{(1)}|} \right)^{\frac{n}{2}}.$$

Se a matriz de planejamento  $X^*$  for de posto completo  $(K+1)$ , e  $(K+1)+p \leq n$  e ainda, se os erros forem normalmente distribuídos, sob  $H_0$ ,  $n\hat{\Sigma}$  possui distribuição de Wishart com  $(n-K-1)$  graus de liberdade. Nas mesmas condições,  $n(\hat{\Sigma}_{(1)} - \hat{\Sigma})$  também possui distribuição de Wishart mas com  $K-q$  graus de liberdade. A matriz aleatória  $A_{(p \times p)}$  tem distribuição de Wishart com parâmetros  $n$  e  $W$  se sua função densidade de probabilidade é

$$f(A, W, n) = \begin{cases} \frac{|A|^{\frac{1}{2}(n-p-1)} \exp(-\frac{1}{2} \text{tr} AW^{-1})}{2^{\frac{np}{2}} \pi^{\frac{1}{4}p(p-1)} |W|^{\frac{1}{2}n} \prod_{i=1}^p \Gamma(\frac{n+1-i}{2})} & \text{para } A \text{ positiva definida,} \\ 0, & \text{caso contrário.} \end{cases}$$

O parâmetro  $n$  é usualmente denominado número de graus de liberdade da distribuição.

Verifica-se que o teste da razão de verossimilhanças é equivalente àquele que rejeita  $H_0$  para valores grandes de

$$-2 \ln \Lambda = -n \ln \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_{(1)}|} = -n \ln \frac{|n\hat{\Sigma}|}{|n\hat{\Sigma} + n(\hat{\Sigma}_{(1)} - \hat{\Sigma})|}.$$

Para  $n$  suficientemente grande, a estatística modificada

$$- \left[ n - K - 1 - \frac{1}{2}(p - K + q + 1) \right] \ln \frac{|\hat{\Sigma}|}{|\hat{\Sigma}_{(1)}|}$$

sob  $H_0$  tem distribuição aproximadamente qui-quadrado com  $p(K - q)$  graus de liberdade. Essa aproximação melhora se  $(n - K)$  e  $(n - p)$  também forem suficientemente grandes. Se não rejeitamos  $H_0$ , as variáveis  $X_{q+1}, X_{q+2}, \dots, X_K$  são descartadas do modelo.

## 2.5 Predição através o modelo de Regressão Multivariada

Nesta seção, vamos discutir o uso do modelo de Regressão Multivariada para realizar predições.

Suponha que o modelo especificado em (2.3) esteja adequadamente ajustado, então podemos empregá-lo para realizar predições.

Seja  $x_0^*$  o vetor contendo novos valores para as variáveis preditoras, para os quais desejamos fazer previsão.

Sabemos através do resultado 7.10 de Johnson e Wichern (1998) que:

$$\hat{\beta}' x_0^* \sim N_p(\beta' x_0^*, x_0^{*'} (X^{*'} X^*)^{-1} x_0^* \Sigma)$$

e

$n\hat{\Sigma} \sim W_{n-K-1}(\Sigma)$  onde  $W_{n-K-1}(\Sigma)$  representa a distribuição Wishart com  $n - K - 1$  graus de liberdade.

Como discutido na Seção 5.2 de Johnson e Wichern (1998), a variável aleatória

$$T = \left( \frac{\hat{\beta}' x_0^* - \beta' x_0^*}{\sqrt{x_0^{*'} (X^{*'} X^*)^{-1} x_0^*}} \right)' \left( \frac{n}{n - K - 1} \hat{\Sigma} \right)^{-1} \left( \frac{\hat{\beta}' x_0^* - \beta' x_0^*}{\sqrt{x_0^{*'} (X^{*'} X^*)^{-1} x_0^*}} \right)$$

tem distribuição  $T^2$  de Hotelling e fixando um nível de confiança de  $(1 - \alpha)$ , o elipsóide de confiança para  $\beta' x_0^*$  provém da desigualdade

$$(\beta' x_0^* - \hat{\beta}' x_0^*)' \left( \frac{n}{n - K - 1} \hat{\Sigma} \right)^{-1} (\beta' x_0^* - \hat{\beta}' x_0^*) \leq x_0^{*'} (X^{*'} X^*)^{-1} x_0^* \left[ \left( \frac{p(n - K - 1)}{n - K - p} \right) \right] F_{\alpha; p, n - K - p}$$



onde  $F_{\alpha;p,n-k-p}$  é o quantil de ordem  $1 - \alpha$  da distribuição F de Snedecor com  $p$  graus de liberdade no numerador e  $n - k - p$  graus de liberdade no denominador.

Os intervalos de confiança simultâneos para  $E(Y_i) = X^{*'}\beta_{(i)}$  são

$$x_0^{*'}\hat{\beta}_{(i)} \pm \sqrt{\left(\frac{p(n-k-1)}{n-k-p}\right) F_{\alpha;p,n-k-p}} \sqrt{x_0^{*'}(X^{*'}X^*)^{-1}x_0^* \left(\frac{n}{n-k-1}\hat{\sigma}_{ii}\right)} \quad i = 1, \dots, p \quad (2.6)$$

onde  $\hat{\beta}_{(i)}$  é a  $i$ -ésima coluna de  $\hat{\beta}$  e  $\hat{\sigma}_{ii}$  é o  $i$ -ésimo elemento da diagonal de  $\hat{\Sigma}$ .

Podemos ainda construir um elipsóide de predição para uma nova resposta  $Y_0 = \beta'x_0^* + \varepsilon$ . Então fixando um coeficiente de confiança  $(1 - \alpha)$  temos:

$$(Y_0 - \hat{\beta}'x_0^*)' \left(\frac{n}{n-k-1}\hat{\Sigma}\right)^{-1} (Y_0 - \hat{\beta}'x_0^*) \leq (1 + x_0^{*'}(X^{*'}X^*)^{-1}x_0^*) \left[\left(\frac{p(n-k-1)}{n-k-p}\right) F_{\alpha;p,n-k-p}\right]$$

com  $(Y_0 - \hat{\beta}'x_0^*) \sim N_p(0, (1 + x_0^{*'}(X^{*'}X^*)^{-1}x_0^*)\Sigma)$ .

Os intervalos simultâneos para a predição de respostas individuais  $Y_{0i}$  são:

$$x_0^{*'}\hat{\beta}_{(i)} \pm \sqrt{\left(\frac{p(n-k-1)}{n-k-p}\right) F_{\alpha;p,n-k-p}} \sqrt{1 + x_0^{*'}(X^{*'}X^*)^{-1}x_0^* \left(\frac{n}{n-k-1}\hat{\sigma}_{ii}\right)} \quad i = 1, \dots, p \quad (2.7)$$

onde  $\hat{\beta}_{(i)}$ ,  $\hat{\sigma}_{ii}$  e  $F_{\alpha;p,n-k-p}$  são as mesmas quantidades definidas em (2.6).

Comparando (2.6) com (2.7) percebemos que os intervalos para valores atuais das variáveis resposta são mais amplos que os correspondentes intervalos para os valores esperados, fato este que já ocorre no modelo de regressão univariada.

## 2.6 Estimação do vetor $\beta$ quando $X_1 \neq X_2 \neq \dots \neq X_p$

Uma vez estudado o caso em que, para cada uma das  $p$  regressões, a matriz de planejamento era a mesma, vamos considerar o caso em que cada regressão possui a sua própria matriz de planejamento.

Vamos supor que:

(i) A covariância entre  $Y_{mi}$  e  $Y_{hi}$  é a mesma para todas as unidades amostrais:

$$\begin{aligned} Cov(Y_{mi}, Y_{hi}) &= \sigma_{mh}^2 & i &= 1, \dots, n \\ m, h &= 1, \dots, p & (m \neq h) \end{aligned}$$

(ii) Medidas tomadas em unidades amostrais distintas são não correlacionadas:

$$\begin{aligned} Cov(Y_{mi}, Y_{hj}) &= 0 & i, j &= 1, \dots, n & (i \neq j) \\ m, h &= 1, \dots, p \end{aligned}$$

(iii)  $Var(Y_{mi}) = \sigma_m^2$   $i = 1, \dots, n$  e  $m = 1, \dots, p$

Como consequência de (i),(ii) e (iii) temos:

$$\begin{aligned} Var(\boldsymbol{\varepsilon}_m) &= \sigma_m^2 I_n & m &= 1, \dots, p \\ Cov(\boldsymbol{\varepsilon}_m, \boldsymbol{\varepsilon}_h) &= \sigma_{mh}^2 I_n & m, h &= 1, \dots, p & (m \neq h). \end{aligned}$$

Utilizando o modelo especificado em (2.1) temos

$$Var(\boldsymbol{\varepsilon}) = \begin{bmatrix} \sigma_1^2 I_n & \sigma_{12} I_n & \dots & \sigma_{1p} I_n \\ & \sigma_2^2 I_n & \dots & \sigma_{2p} I_n \\ & & \ddots & \vdots \\ & & & \sigma_p^2 I_n \end{bmatrix} = \Sigma \otimes I_n = \Omega$$

$$\text{onde } \Sigma = \begin{bmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1p} \\ & \sigma_2^2 & \dots & \sigma_{2p} \\ & & \ddots & \vdots \\ & & & \sigma_p^2 \end{bmatrix}.$$

Se  $\Sigma$  é conhecida, estima-se  $\beta$  da forma como especificada em (2.2). Além disso,  $Var(\hat{\beta}) = (X'\Omega^{-1}X)^{-1}$ , e como  $\Omega^{-1} = \Sigma^{-1} \otimes I_n$  então, após alguns cálculos obtemos

$$Var(\hat{\beta}) = \begin{bmatrix} \sigma_1^2 X_1' X_1 & \dots & \dots & \sigma_{1p} X_1' X_1 \\ & \sigma_2^2 X_2' X_2 & \dots & \sigma_{2p} X_2' X_2 \\ & & \ddots & \vdots \\ & & & \sigma_p^2 X_p' X_p \end{bmatrix}^{-1}.$$

Quando  $\Sigma$  é desconhecida estima-se  $\beta$  em dois estágios:

1. Estima-se  $\beta$  como no caso em que  $X_1 = \dots = X_p = X^*$ , ou seja, cada  $\hat{\beta}_j$  é estimado separadamente.
2. Com base nas estimativas obtidas no estágio anterior, estima-se  $\Sigma$  através de  $\hat{\Sigma}$  conforme descrito na Seção 2.3.

A partir daí estima-se novamente  $\beta$  usando (2.2) para  $\Omega = \hat{\Sigma}$ .

Para testar hipóteses do tipo

$$\begin{cases} H_0 : C\beta = \mathbf{m} \\ H_1 : C\beta \neq \mathbf{m}, \end{cases}$$

em que  $C$  e  $\mathbf{m}$  são, respectivamente, matrizes e vetores de constantes especificadas, utiliza-se o teste da razão de verossimilhanças, com estatística

$$W = \frac{(C\hat{\beta} - \mathbf{m})' [C(X'\hat{\Sigma}^{-1}X)^{-1}C'] (C\hat{\beta} - \mathbf{m})}{\frac{(\mathbf{Y}-X\hat{\beta})'(\mathbf{Y}-X\hat{\beta})}{n-p-1} p}.$$

Fixando um nível de significância  $\alpha$ , rejeita-se  $H_0$  se  $W \geq w$  onde  $w$  é o quantil de ordem  $(1 - \alpha)$  da distribuição F de Snedecor com  $p$  graus de liberdade no numerador e  $n - K - 1$  graus de liberdade no denominador.

## 2.7 Ganho em Eficiência com a Estimação Conjunta

Zellner(1962) apresenta um estudo sobre o ganho em eficiência na utilização do modelo de Regressão Multivariada em relação ao caso em que se ajusta  $p$  regressões separadas. Segundo o autor, quando estimamos conjuntamente os parâmetros das  $p$  regressões, obtemos um ganho de eficiência dos estimadores. Esse ganho ocorre essencialmente porque ao estimar conjuntamente os parâmetros, produzimos restrições nulas nos coeficientes das outras equações. Estas restrições podem ser observadas reescrevendo (2.1) como:

$$(\mathbf{Y}_1, \dots, \mathbf{Y}_p) = (X_1, \dots, X_p) \begin{bmatrix} \beta_1 & 0 & \dots & 0 \\ 0 & \beta_2 & \dots & 0 \\ \vdots & & \ddots & \vdots \\ 0 & \dots & \dots & \beta_p \end{bmatrix} + (\varepsilon_1, \dots, \varepsilon_p).$$

Vamos considerar o caso em que  $\sigma_{ii}^2 = \sigma^2$  e  $\sigma_{ij} = \sigma^2\rho$  para  $i \neq j$ , então  $\Sigma = \sigma^2[(1 - \rho)I + \rho\mathbf{E}\mathbf{E}']$  onde  $I$  é a matriz identidade de ordem  $p \times p$  e  $\mathbf{E}' = (1, \dots, 1)$  é um vetor de ordem  $p$ .

Então  $\Sigma^{-1} = \frac{1}{\sigma^2(1-\rho)}I - \frac{\alpha\rho\mathbf{E}\mathbf{E}'}{[1+(p-1)\rho]}$  onde  $\alpha = \frac{1}{\sigma^2(1-\rho)}$ . Dessa forma:

$$\begin{aligned} Var(\hat{\beta}) &= [X'(\Sigma^{-1} \otimes I)X]^{-1} \\ Var(\hat{\beta}) &= \begin{bmatrix} (\alpha - \gamma)X_1'X_1 & -\gamma X_1'X_2 & \dots & -\gamma X_1'X_p \\ -\gamma X_2'X_1 & (\alpha - \gamma)X_2'X_2 & \dots & -\gamma X_2'X_p \\ \vdots & \vdots & \ddots & \vdots \\ -\gamma X_p'X_1 & (\alpha - \gamma)X_p'X_2 & \dots & (\alpha - \gamma)X_p'X_p \end{bmatrix}^{-1} \end{aligned} \quad (2.8)$$

onde  $\gamma = \frac{\alpha\rho}{[1+(p-1)\rho]}$ .

Em particular, para o caso em que  $p = 2$ , a matriz de variância e covariância do estimador do vetor de parâmetros da primeira regressão é

$$Var(\hat{\beta}_1) = [(\alpha - \gamma)X_1'X_1 - \frac{-\gamma^2}{\alpha - \gamma}X_1'X_2(X_2'X_2)^{-1}X_2'X_1]^{-1}.$$

Zellner e Huang(1961) mostram que

$$|Var(\hat{\beta}_1)| = \frac{(1 - \rho^2)^{l_1}}{\prod_{i=1}^{l_1} (1 - \rho^2 r_i^2)} |\sigma^2(X_1'X_1)^{-1}| \quad (2.9)$$

onde  $l_1$  é o número de variáveis independentes da primeira equação, admitindo-se que  $l_1 \leq l_2$  e  $r_i$  é o  $i$ -ésimo coeficiente de correlação canônica (Johnson e Wichern (1998)) associada com o conjunto de variáveis  $X_1$  e  $X_2$ . Como  $0 \leq r_i^2 \leq 1$ , fica claro que:

$$|Var(\hat{\beta}_1)| \leq |\sigma^2(X_1'X_1)^{-1}|.$$

Note que  $|\sigma^2(X_1'X_1)^{-1}|$  é a variância generalizada do estimador do vetor de parâmetros da primeira equação quando estimamos os parâmetros separadamente. Define-se variância generalizada como sendo o determinante da matriz de variância e covariância. Logo estimar conjuntamente os parâmetros das equações de regressão é um processo mais eficiente.

Se  $r_i = 0$ , para todo  $i$ , a expressão (2.9) se reduz a

$$|Var(\hat{\beta}_1)| = |\sigma^2(X_1'X_1)^{-1}|.$$

Observa-se que  $r_i = 0$  quando  $X_1'X_2 = 0$ .

Além disso, como consequência de (2.8) se  $X_i'X_j = 0$  para  $i \neq j$  obtemos

$$Var(\hat{\beta}_1) = \left[ \frac{1 - \rho}{1 - \frac{\rho}{1 + \rho(p-1)}} \right] \sigma^2(X_1'X_1)^{-1}$$

Assim, se  $X_i'X_j = 0$   $i, j = 1, \dots, p$  ( $i \neq j$ ),  $Var(\hat{\beta}_1)$  se aproxima de  $(1 - \rho)\sigma^2(X_1'X_1)^{-1}$  para um número  $p$  muito grande de equações.

## 2.8 Propriedades do Estimador em Dois Estágios

Zellner (1962) apresenta um estudo das propriedades do estimador em dois estágios.

Seja

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

definido na Seção 2.6. O autor verifica que

$$\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{\mathbf{G}} + O_p(n^{-1})$$

com  $\hat{\boldsymbol{\beta}}_{\mathbf{G}}$  o estimador definido em (2.2) quando  $\Omega$  é conhecido.

Dizemos que a sequência de números reais  $\{a_n\}$   $n \geq 1$  é de ordem  $n^{-1}$  em probabilidade ( $a_n = O_p(n^{-1})$ ) se para todo número real  $\eta > 0$ , existir  $k = k(\eta)$  e  $n_0 = n_0(\eta)$  tal que

$$P\left(\left|\frac{a_n}{n^{-1}}\right| \geq k\right) \leq \eta, \quad \forall n > n_0,$$

ou seja, se a sequência  $\left\{\frac{a_n}{n^{-1}}\right\}_{n \geq 1}$  for limitada em probabilidade para todo  $n$  suficientemente grande.

Verifica-se adicionalmente que

$$\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \text{ e } \sqrt{n}(\hat{\boldsymbol{\beta}}_{\mathbf{G}} - \boldsymbol{\beta})$$

têm assintoticamente a mesma distribuição.

Dessa forma,  $\hat{\boldsymbol{\beta}}$  pode ser usado ao invés de  $\hat{\boldsymbol{\beta}}_{\mathbf{G}}$  sem muito prejuízo.

Neste capítulo, apresentamos a teoria básica associada ao Modelo de Regressão Multivariada. No próximo capítulo será abordado o problema de diagnóstico neste modelo.

# Capítulo 3

## Análise de Diagnóstico

Neste capítulo serão apresentadas medidas para a determinação de observações influentes, no modelo de Regressão Multivariada, com exemplos ilustrativos de algumas delas. Observações influentes são aquelas que individualmente, ou em conjunto com outras, exercem grande impacto nos valores de várias estimativas. Esta análise, amplamente discutida na literatura em modelos de regressão univariada, é menos freqüente no modelo multivariado.

Iniciaremos o estudo definindo algumas classes de medidas de influência, específicas para este modelo. Posteriormente, veremos como algumas medidas tradicionais na análise de regressão usual são adaptadas ao modelo multivariado.

### 3.1 Classes Gerais de Medidas de Influência

Barret e Ling (1992) consideram o modelo de regressão linear multivariada conforme apresentado em (2.3), ou seja,

$$Y_{n \times p} = X_{n \times (K+1)} \beta_{(K+1) \times p} + \varepsilon_{n \times p},$$

com  $E(\varepsilon) = 0$  e  $Cov(\varepsilon_i, \varepsilon_j) = \sigma_{ij}I$   $i, j = 1, 2, \dots, p$  é a matriz de covariância entre  $\varepsilon_i, \varepsilon_j$ ,

respectivamente  $i$ -ésima e  $j$ -ésima colunas de  $\varepsilon$ .

Dessa forma, as  $p$  medidas em cada unidade amostral têm matriz de covariância  $\Sigma$ , mas observações em unidades distintas são não correlacionadas.

O estimador de mínimos quadrados de  $\beta$ , que coincide com o de máxima verossimilhança sob normalidade, é dado por  $\hat{\beta} = (X'X)^{-1}X'Y$ . Define-se a matriz de resíduos  $E = Y - X\beta$  de modo que o estimador não viesado de  $\Sigma$  é  $\hat{\Sigma} = (E'E)/(n - K - 1)$ .

Todas as medidas consideradas aqui levam em conta a deleção de pontos. Em particular, será analisada a exclusão de  $m$  unidades amostrais pertencentes ao conjunto  $I \subset \{1, 2, \dots, n\}$ . Seja  $X_I$  a submatriz de  $X$  com  $m$  linhas e  $X_{(I)}$  a submatriz de  $X$  sem as  $m$  linhas que compõe  $X_I$  e  $Y_I, Y_{(I)}$  e  $\varepsilon_I$  similarmente definidas. O estimador de  $\beta$  com base nas observações restantes é dado por  $\hat{\beta}_{(I)} = (X'_{(I)}X_{(I)})^{-1}X'_{(I)}Y_{(I)}$ .

Barret e Ling (1992) definiram classes de medidas de influência que são funções de duas matrizes  $H_I$  e  $Q_I$ , definidas como se segue. Para o conjunto completo dos dados a matriz de projeção (ou matriz chapéu)  $H$  é dada por  $H = X(X'X)^{-1}X'$ . A submatriz  $H_I$  corresponde a

$$H_I = X_I(X'X)^{-1}X'_I. \quad (3.1)$$

A matriz  $Q$  é análoga a  $H$  com respeito à matriz de resíduos  $E$ , isto é,  $Q = E(E'E)^{-1}E'$  com sua correspondente submatriz

$$Q_I = E_I(E'E)^{-1}E'_I. \quad (3.2)$$

As classes gerais de medidas de influência associadas com o subconjunto de casos indexados por  $I$ , denotadas por  $J_I^{tr}, J_I^{det}$ , são definidas da seguinte forma:

$$J_I^{tr}(f; a, b) = f(n, K + 1, p, m)tr[H_I Q_I (I - H_I - Q_I)^a (I - H_I)^b] \quad (3.3)$$

e

$$J_I^{det}(f; a, b) = f(n, K + 1, p, m)det[(I - H_I - Q_I)^a (I - H_I)^b], \quad (3.4)$$



onde  $f$  é uma função das dimensões  $(n, K + 1, p)$  das matrizes presentes em (2.3) e das  $m$  linhas das submatrizes;  $I$  é matriz identidade de ordem  $m$  e  $a$  e  $b$  são valores inteiros associados a um particular membro de cada classe.

### 3.2 Medidas da Classe $J_I^{tr}$

Cook (1977, 1979) introduziu uma medida de influência baseada na comparação entre as estimativas de mínimos quadrados ordinários  $\hat{\beta}$  e  $\hat{\beta}_{(i)}$ , quando a  $i$ -ésima observação é retirada. Esta medida é definida como

$$d_i = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' [\widehat{Var}(\hat{\beta})]^{-1} (\hat{\beta}_{(i)} - \hat{\beta})}{K + 1} = \frac{(\hat{\beta}_{(i)} - \hat{\beta})' (X'X) (\hat{\beta}_{(i)} - \hat{\beta})}{(K + 1)\hat{\sigma}^2}$$

onde

$\hat{\beta}$  e  $\hat{\beta}_{(i)}$  são os estimadores de  $\beta$  com e sem a  $i$ -ésima observação respectivamente,

$\hat{\sigma}^2$  é o quadrado médio do resíduo e

$\beta$  é o vetor de parâmetros de regressão do modelo univariado.

Para utilizar a medida distância de Cook, âmbito da Regressão Multivariada, precisamos da seguinte definição:

Se  $A$  é uma matriz qualquer com  $q$  colunas,  $\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_q$ , de modo que  $A = [\mathbf{a}_1 \ \mathbf{a}_2 \ \dots \ \mathbf{a}_q]$ , o vetor  $vecA$  é definido como o vetor formado pelas colunas de  $A$ , ou seja

$$vecA = \begin{bmatrix} \mathbf{a}_1 \\ \vdots \\ \mathbf{a}_q \end{bmatrix}.$$

Dessa forma, a medida distância de Cook no âmbito do modelo de Regressão Multivariada, é dada por:

$$D_I = (vec\hat{\beta} - vec\hat{\beta}_{(I)})' [\hat{\Sigma}^{-1} \otimes (X'X)] (vec\hat{\beta} - vec\hat{\beta}_{(I)}) \left( \frac{1}{K + 1} \right), \quad (3.5)$$

Podemos assumir que (3.5) é um caso particular da medida geral

$$D_I(M, V) = (\text{vec}\hat{\beta} - \text{vec}\hat{\beta}_{(I)})' [\hat{V}^{-1} \otimes M] (\text{vec}\hat{\beta} - \text{vec}\hat{\beta}_{(I)}) \left( \frac{1}{K+1} \right) \quad (3.6)$$

onde  $V$  e  $M$  são matrizes positivas definidas. Se escolhermos  $V = (K+1)\hat{\Sigma}$  e  $M = X'X$ , a expressão (3.6) reduz-se a (3.5).

Quando tomamos  $V = (K+1)\hat{\Sigma}_{(I)}$  e  $M = X'X$  obtemos:

$$D_I = (\text{vec}\hat{\beta} - \text{vec}\hat{\beta}_{(I)})' [\hat{\Sigma}_{(I)}^{-1} \otimes (X'X)] (\text{vec}\hat{\beta} - \text{vec}\hat{\beta}_{(I)}) \left( \frac{1}{K+1} \right)$$

que é a versão multivariada da medida DFFITS. No caso univariado, na exclusão de uma única observação é tal que

$$(DFFITS)^2 = \frac{(\hat{\beta}_{(i)} - \hat{\beta})'(X'X)(\hat{\beta}_{(i)} - \hat{\beta})}{\hat{\sigma}_{(i)}^2}$$

em que

$\hat{\sigma}_{(i)}^2$  é o quadrado médio do resíduo do ajuste do modelo sem o ponto  $i$ .

Voltando à distância de Cook Multivariada definida em (3.5), é possível verificar que essa medida pertence a classe  $J_I^{tr}$ . Para isso, seguem os fatos:

Dadas as matrizes  $A_{m \times n}$ ,  $B_{p \times q}$ ,  $C_{p \times m}$  e  $D_{q \times n}$ , então

i)  $tr(AB) = (\text{vec}A)' \text{vec}B$  e  $\text{vec}(ABC) = (C' \otimes A) \text{vec}B$ .

Como consequência de (i) temos:

ii)  $tr[C'(BDA')] = (\text{vec}C)' \text{vec}(BDA') = (\text{vec}C')(A \otimes B) \text{vec}D$ .

Vale ainda a propriedade

iii)  $tr(AB) = tr(BA)$  se ambos os produtos  $AB$  e  $BA$  são possíveis.

Através de uma generalização direta da expressão (3.6.4) de Cook e Weisberg (1982, p.136), verifica-se que

iv)  $B - B_{(I)} = (X'X)^{-1}X_I(I - H_I)^{-1}E_I$

Como em  $D_I = (vec\hat{\beta} - vec\hat{\beta}_{(I)})' [\hat{\Sigma}_{(I)}^{-1} \otimes (X'X)] (vec\hat{\beta} - vec\hat{\beta}_{(I)}) \left(\frac{1}{K+1}\right)$ ,  $(vec\hat{\beta} - vec\hat{\beta}_{(I)}) = vec(\hat{\beta} - \hat{\beta}_{(I)})$ , então, devido a (iv) e aplicando-se os resultados (ii) e (iii), segue que

$$D_I[X'X, (K+1)\hat{\Sigma}] = tr\{(I - H_I)^{-1}X_I(X'X)^{-1}X'X(X'X)^{-1}X'_I(I - H_I)^{-1}E_I[(K+1)\hat{\Sigma}]^{-1}E'_I\}$$

$$D_I[X'X, (K+1)\hat{\Sigma}] = \left(\frac{n-K-1}{K+1}\right) tr[(I - H_I)^{-1}X_I(X'X)^{-1}X'_I(I - H_I)^{-1}E_I(E'E)^{-1}E'_I]$$

$$D_I[X'X, (K+1)\hat{\Sigma}] = \left(\frac{n-K-1}{K+1}\right) tr[(I - H_I)^{-1}H_I(I - H_I)^{-1}Q_I]$$

$$D_I[X'X, (K+1)\hat{\Sigma}] = \left(\frac{n-K-1}{K+1}\right) tr[(I - H_I)^{-2}H_IQ_I] = J_I^{tr} \left(\frac{n-K-1}{K+1}; 0, -2\right)$$

As duas últimas passagens são consequência do resultado (iii) e do fato que  $H_I(I - H_I)^{-1} = (I - H_I)^{-1}H_I$ .

Para a versão multivariada do DFFITS, após alguns cálculos, concluímos que

$$D_I[X'X, (K+1)\hat{\Sigma}_{(I)}] = J_I^{tr} \left(\frac{n-K-1-m}{K+1}; -1, -1\right).$$

### 3.3 Medidas na Classe $J^{det}$

Andrews e Pregibon (1978) propuseram uma medida de influência para regressão univariada que detecta pontos relativamente distantes do restante dos dados. Essa medida é dada por

$$AP_I = \frac{det(Z'_{(I)}Z_{(I)})}{det(Z'Z)} \quad (3.7)$$

onde  $Z = [X|Y]$  é a matriz  $X$  acrescida do vetor  $y$ . Dessa forma  $AP_I$  mede mudanças ocorridas no espaço gerado pelas colunas de  $Z$  sem distinguir se ela ocorre nas colunas de  $X$  ou de  $y$ . Essa medida pode ser generalizada simplesmente substituindo o vetor  $y$  pela matriz de variáveis resposta  $Y$  da regressão multivariada.

Agora,

$$Z'Z = \begin{bmatrix} X'X & X'Y \\ Y'X & Y'Y \end{bmatrix} = \begin{bmatrix} X' \\ Y' \end{bmatrix} \begin{bmatrix} X & Y \end{bmatrix}$$

Aplicando o lema

$$\det \begin{bmatrix} A & B \\ C & D \end{bmatrix} = \det(A)\det(D - CA^{-1}B), \quad \text{para } A \text{ e } D \text{ não singulares, temos:}$$

$$\det(Z'Z) = \det(X'X)\det(Y'Y - Y'X(X'X)^{-1}X'Y) = \det(X'X)\det(E'E) \quad (3.8)$$

pois,  $E'E = (Y - X(X'X)^{-1}X'Y)'(Y - X(X'X)^{-1}X'Y) = Y'Y - (Y'X(X'X)^{-1}X'Y)$  Similarmente

$$\det(Z'_{(I)}Z_{(I)}) = \det(X'_{(I)}X_{(I)})\det(E'_{(I)}E_{(I)}). \quad (3.9)$$

Para as matrizes  $X_I$ ,  $X_{(I)}$ ,  $H_I$ ,  $E$  e  $E_{(I)}$  definidas anteriormente valem os seguintes fatos

v)  $X_I(X'_{(I)}X_{(I)})^{-1}X'_I = H_I(I - H_I)^{-1}$ ,

vi)  $\det(X'_{(I)}X_{(I)}) = \det(X'X)\det(I - H_I)$ ,

vii)  $\det(E'_{(I)}E_{(I)}) = \det(E'E)\det[(I - H_I - Q_I)(I - H_I)^{-1}]$ .

Substituindo (3.7) e (3.8) em (3.6) e aplicando os resultados v) e vi) temos

$$AP_I = \det(I - H_I)\det[(I - H_I - Q_I)(I - H_I)^{-1}] = \det(I - H_I - Q_I) = J_{(1;1,0)}^{det}.$$

Belsley et al (1980,p.22 eq. 2.36) propuseram uma medida de influência baseada na matriz de covariância do estimador  $\hat{\beta}$ . Essa medida é dada por

$$RC_I = \frac{\det[\hat{\Sigma}_{(I)} \otimes (X'_{(I)}X_{(I)})^{-1}]}{\det[\hat{\Sigma} \otimes (X'X)^{-1}]}$$

Lembrando que  $\hat{\Sigma} = \frac{E'E}{n-K-1}$  e devido as propriedades de determinante do produto de Kronecker, essa medida pode ser expressa como

$$RC_I = \frac{\det[(n - K - m - 1)^{-1}(E'_{(I)}E_{(I)})]^p \det(X'X)^{K+1}}{\det[((n - k - 1)^{-1}E'E)^p \det(X'_{(I)}X_{(I)})^p]}$$

Após aplicarmos os resultados vi) e vii) obtemos

$$RC_I = \left[ \frac{n - K - 1}{n - K - m - 1} \right]^{(K+1)p} \det[(I - H_I - Q_I)^{K+1}(I - H_I)^{-(K+P+1)}], \text{ ou seja,}$$

$$RC_I = J_I^{det} \left[ \left( \frac{n - K - 1}{n - K - m - 1} \right)^{(K+1)p}; K + 1, -(p + K + 1) \right].$$

Podemos também medir a influência das observações nas matrizes de covariância dos valores ajustados,  $cov(X_I \text{vec} \hat{\beta})$ ,

$$RC_I = \frac{\det[cov(X_I \text{vec} \hat{\beta}_{(I)})]}{\det[cov(X_I \text{vec} \hat{\beta})]} = \frac{\det\{\hat{\Sigma}_{(I)} \otimes [(I - H_{(I)})^{-1} H_{(I)}]\}}{\det\{\hat{\Sigma}_{(I)} \otimes H_{(I)}\}}.$$

### 3.4 Alavanca e Componentes de Resíduos

Na seção anterior vimos a caracterização das medidas de influência em termos das submatrizes  $H_I$  e  $Q_I$ . A matriz  $H_I$  está usualmente relacionada com a detecção de pontos alavanca, que em geral, são pontos posicionados em regiões distantes do subespaço gerado pelas colunas na matriz  $X$ , e a matriz  $Q_I$  está relacionada com os componentes dos resíduos.

De modo geral, pontos alavanca estão posicionados em regiões distantes do subespaço gerado pelas colunas da matriz  $X$

Denotando a matriz alavanca por  $L_I$  e a matriz dos componentes dos resíduos por  $R_I$  obtemos a Tabela 3.1, que apresenta essa decomposição para as medidas estudadas nas seções anteriores.

Tabela 3.1: Decomposição das Medidas de Influência.

Medida	$L_I$	$R_I$
$D_I[X'X, (K+1)\hat{\Sigma}]$	$H_I(I - H_I)^{-1}$	$(I - H_I)^{-1/2} Q_I (I - H_I)^{-1/2}$
$D_I[X'X, (K+1)\hat{\Sigma}_{(I)}]$	$H_I(I - H_I)^{-1}$	$(I - H_I)^{-1/2} Q_I (I - H_I - Q_I)^{-1} (I - H_I)^{1/2}$
$RC^{-1} = \frac{1}{RC}$	$(I - H_I)^p$	$[(I - H_I - Q_I)^{-1} (I - H_I)]^{K+1}$

Na análise de influência de uma única observação, os elementos da diagonal principal das matrizes H e Q são essenciais.

Como agora estamos tratando da influência conjunta de m observações, no caso multivariado precisamos reduzir as informações das matrizes  $L_I$  e  $R_I$  em escalares. Examinando a caracterização dos pontos influêntes na classe  $J_i^{tr}$  em termos de um produto interno e aplicando um resultado elementar de geometria vetorial podemos escrever:

$$tr(L_I R_I) = [vec(L_I)]' vec(R_I) = \|vec(L_I)\| \|vec(R_I)\| \cos \theta_I$$

onde  $\theta_I$  é o ângulo entre  $vec(L_I)$  e  $vec(R_I)$ .

Como  $L_I$  e  $R_I$  são matrizes positivas semi-definidas, suas normas podem ser, alternativamente, expressas em termos de seus auto-valores. Por exemplo:

$$\|L_I\| = \left[ \sum_{j=1}^m \lambda_j^2(L_I) \right]^{1/2}$$

em que  $\lambda_j(L_I)$  representa o  $j$ -ésimo auto-valor de  $L_I$ . Denotando  $\|L_I\|$  e  $\|R_I\|$  por  $\mathcal{L}_I$  e  $\mathcal{R}_I$  respectivamente, nota-se que estas quantidades representam a contribuição da alavanca e dos resíduos na influência total.

Na realidade, a influência também depende de  $\cos \theta_I$ , ainda que a contribuição relativa de cada componente não seja afetada pelo ângulo  $\theta_I$ . Para analisar a contribuição efetiva dos subconjuntos sobre a influência total, podemos reescrever  $\mathcal{L}_I$  e  $\mathcal{R}_I$  da seguinte forma:  $\mathcal{L}_I^* = \mathcal{L}_I(\cos \theta_I)^{1/2}$  e  $\mathcal{R}_I^* = \mathcal{R}_I(\cos \theta_I)^{1/2}$  de modo que  $tr(L_I R_I) = \mathcal{L}_I^* \mathcal{R}_I^*$ . Por outro lado, as medidas  $J_I^{det}$  são da forma  $f(\cdot) det(L_I R_I)$ . Para reduzir as medidas desta classe a escalares basta calcular o determinante das matrizes  $L_I$  e  $R_I$ . Assim,  $\frac{1}{AP_I}$  tem componente de alavanca  $det[(I - H_I)^{-1}]$  e componente do resíduo  $det[(I - H_I - Q_I)^{-1}(I - H_I)]$ .

Para ilustrar a contribuição da alavanca e dos componentes do resíduo na influência total sugere-se a construção do gráfico de  $\log \mathcal{L}_I^*$  contra  $\log \mathcal{R}_I^*$ , pois, no caso da medida  $J_I^{tr}$ , como  $J_I^{tr}(f; a, b) = f(\cdot) tr(L_I R_I) = f(\cdot) \mathcal{L}_I^* \mathcal{R}_I^*$ , pontos com  $J_I^{tr}$  constante serão da forma  $\mathcal{L}_I^* \mathcal{R}_I^*$  constante.

Dessa forma, pontos de mesma influência serão tais que  $\mathcal{L}_{\mathcal{I}}^* \mathcal{R}_{\mathcal{I}}^* = C$ ,  $C$  constante, de modo que  $\log \mathcal{L}_{\mathcal{I}}^* + \log \mathcal{R}_{\mathcal{I}}^* = \log C$ . Portanto os contornos desse gráfico são retas com inclinação -1 e tais que a soma das coordenadas é o logaritmo da influência, como ilustra a Figura 3.1.

Afim de ilustrar a detecção de pontos alavanca ou influentes, os autores escolheram um particular conjunto de dados previamente estudado em Houssain e Naik (1989). A Regressão Multivariada foi usada para avaliar a qualidade dos dados de três testes de proficiência de aprendizado:

- Desempenho no teste gráfico de vocabulário por imagens ( $Y_1$ );
- Avaliação de um teste de desempenho do estudante ( $Y_2$ );
- Desempenho no teste de matrizes progressivas de Ravin ( $Y_3$ ).

As variáveis explicativas foram a soma do número de itens corretos (num total de vinte itens) em cinco tipos de tarefas de associação pareada:

- Nomeação ( $X_1$ );
- Figura ( $X_2$ );
- Nomeação de Figura ( $X_3$ );
- Nomeação de Ação ( $X_4$ );
- Sentença de Figura ( $X_5$ ).

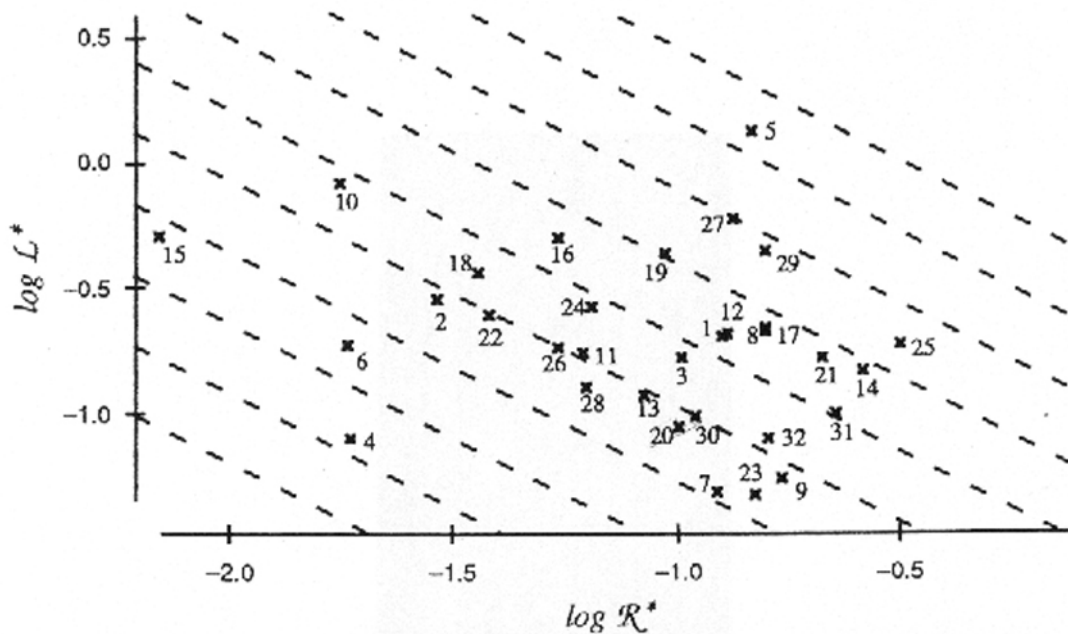
O objetivo do estudo era avaliar quão bem as variáveis explicativas conseguem prever as notas nos testes.

Para realizar a análise, foram selecionados 32 estudantes brancos de uma classe avançada de um colégio.

O modelo é  $Y_{32 \times 3} = X_{32 \times 6} \beta_{6 \times 3} + \varepsilon_{32 \times 3}$ , sendo que as linhas de  $\varepsilon$  são vetores aleatórios tri-dimensionais com distribuição  $N_3(0, \Sigma)$ .

Na Figura 3.1 foi construído o gráfico de  $\log \mathcal{L}_T^*$  contra  $\log \mathcal{R}_T^*$  para avaliar a influência de uma observação usando a distância de Cook Multivariada ( $D_I[X'X, (K+1)S]$ ). Os contornos do gráfico são espaçados de maneira que entre linhas adjacentes, da esquerda para a direita, existe o dobro de influência. Podemos notar que a observação {5} possui grande componente de resíduo e a maior alavanca e a observação {25} é a que possui maior componente do resíduo. Além disso, a observação {10} possui grande componente de resíduo mas não é uma observação alavanca.

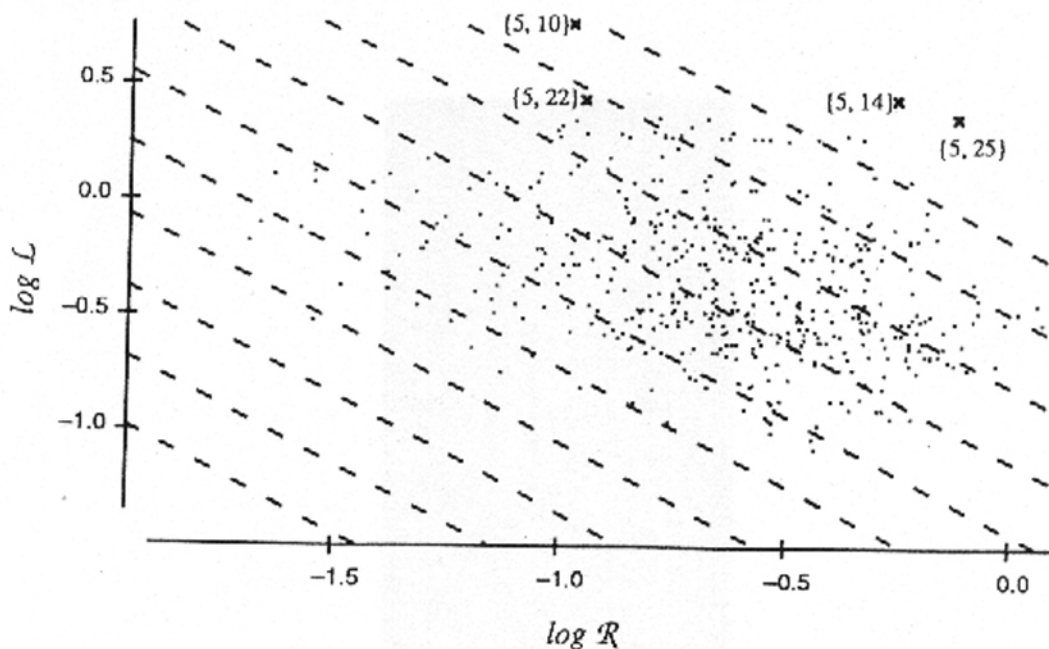
Figura 3.1: Gráfico de influência individual.



Os autores avaliam ainda os subconjuntos que são alavanca ou possuem grande componente de resíduo. A Figura 3.2 exibe o gráfico de  $\log \mathcal{L}_T$  contra  $\log \mathcal{R}_T$  para subconjuntos de tamanho  $m=2$ . Individualmente as observações {14} e {25} possuíam, relativamente, baixa influência total como ilustra a Figura 3.1, no entanto, em conjunto com a observação {5}, formaram os dois subconjuntos com maior influência usando a medida ( $D_I(X'X, pS)$ ).



Figura 3.2: Gráfico de influência para subconjuntos de tamanho 2.



Os gráficos sugeridos são muito úteis na ordenação da contribuição do componente de alavanca e do componente de resíduo de diferentes subconjuntos para uma dada medida de influência.

Na análise de diagnóstico, Houssain e Naik (1989) utilizaram o modelo (2.3)  $Y_{n \times p} = X_{n \times (K+1)}\beta_{(K+1) \times p} + \varepsilon_{n \times p}$  com  $vec(\varepsilon) \sim N(0, \Sigma \otimes I_n)$ .

Este modelo pode ser escrito alternativamente como

$$(vecY_1 \dots vecY_p) = (Xvec\beta_1 \dots Xvec\beta_p) + (vec\varepsilon_1 \dots vec\varepsilon_p)$$

onde  $vecY_1 \dots vecY_p$  representam as colunas da matriz Y e o restante da notação é definido de forma similar.

Os vetores  $vec\beta_1, \dots, vec\beta_p$  correspondem aos parâmetros de cada uma das p regressões e o estimador de cada  $vec\beta_j$  é dado por:

$$vec\hat{\beta}_j = (X'X)^{-1}X'vecY_j, \quad j = 1, 2, \dots, p.$$

Define-se ainda o vetor de resíduos de cada regressão ,  $vec\varepsilon_j = vecY_j - Xvec\hat{\beta}_j$ ,  $j = 1, 2, \dots, p$  de modo que se  $E = (vecE_1, vecE_2, \dots, vecE_p)$  e  $P = X(X'X)^{-1}X'$  então  $vec(E) \sim N(\mathbf{0}, \Sigma \otimes (I - P))$ .

Verifica-se que a matriz  $\hat{\beta} = (X'X)^{-1}X'Y$ , definida anteriormente, com a notação agora introduzida fica

$$\hat{\beta} = (vec\hat{\beta}_1, vec\hat{\beta}_2, \dots, vec\hat{\beta}_p)$$

e a matriz de resíduos  $E = Y - X\hat{\beta}$  pode ser escrita como:  $\begin{bmatrix} e'_1 \\ \vdots \\ e'_n \end{bmatrix}$ , sendo  $e'_1, e'_2, \dots, e'_n$  as linhas de  $e$ , cada uma de dimensão  $1 \times p$ .

A estimativa da matriz  $\Sigma$  é dada por

$$\hat{\Sigma} = \frac{E'E}{n - K - 1}.$$

Na análise de diagnóstico, os autores consideram inicialmente a retirada da  $i$ -ésima observação,  $i = 1, 2, \dots, n$ , que dá origem às seguintes quantidades

$$\hat{\Sigma}^{(i)} = \frac{E_{(i)}'E_{(i)}}{n - K - 2},$$

com

$$E_{(i)} = Y_{(i)} - X_{(i)}\beta_{(i)} = (\hat{\beta}_{1(i)} \dots \hat{\beta}_{p(i)})$$

onde  $\beta_{(i)} = (X_{(i)}'X_{(i)})^{-1}X_{(i)}'Y_{(i)}$

e

$$\hat{\beta}_{j(i)} = (X_{(i)}'X_{(i)})^{-1}X_{(i)}'Y_{j(i)}, \quad j = 1, 2, \dots, p$$

$$i = 1, 2, \dots, n$$

é o estimador de  $\beta_j$  calculado quando a  $i$ -ésima observação é retirada.

Os autores propõem métodos de detecção de pontos influentes que são extensões das medidas univariadas usuais. Dessa forma, na análise de influência, são definidas as quantidades

$$\tau_i^2 = \frac{1}{1 - p_i} e_i' \hat{\Sigma}^{-1} e_i \quad \text{e} \quad T_i^2 = \frac{1}{1 - p_i} e_i' \hat{\Sigma}_{(i)}^{-1} e_i, \quad i = 1, 2, \dots, n$$

onde  $e'_1, e'_2, \dots, e'_n$  são as linhas de  $E$ , cada uma de dimensão  $1 \times p$  e  $p_i$  é o  $i$ -ésimo elemento da diagonal principal da matriz  $P = X(X'X)^{-1}X'$ .

Observa-se que  $T_i^2$  possui distribuição  $T^2$  de Hotelling. Verifica-se ainda que  $\tau_i^2$  e  $T_i^2$  são as versões multivariadas, respectivamente, dos resíduos internamente e externamente “studentizados” (Cook e Weisberg, 1982). Estas estatísticas consideram os  $p$  resíduos, das  $p$  regressões, para cada indivíduo. Com base nessas quantidades, os autores definem as próximas medidas.

### **Distância de Cook Multivariada**

A influência da  $i$ -ésima observação sobre  $B$  é avaliada por:

$$C_i = \frac{1}{K + 1} \frac{p_i}{1 - p_i} \tau_i^2 \quad i = 1, 2, \dots, n.$$

Quanto maior o valor de  $C_i$ , mais influência possui a observação  $i$ . Sugere-se como ponto de corte valores baseados na distribuição F de Snedecor com  $K + 1$  e  $n - K - 1$  graus de liberdade no numerador e no denominador respectivamente.

### **Estatística de Welsch-Kuh**

Mede a influência da  $i$ -ésima observação sobre o valor predito  $\hat{y}_i$ , sendo definida por:

$$WK_i = \frac{p_i}{1 - p_i} T_i^2 \quad i = 1, 2, \dots, n$$

A  $i$ -ésima observação será influente sobre  $\hat{y}_i$  quando  $WK_i$  for maior que

$$\frac{K + 1}{n} \frac{p(n - K)}{n - K - p - 1} F_{(\alpha; p, n - K - p - 1)},$$

onde  $F_{(\alpha; p, n - K - p - 1)}$  é o quantil de ordem  $1 - \alpha$  da distribuição F de Snedecor com  $p$  graus de liberdade no numerador e  $n - K - p - 1$  graus de liberdade no denominador.

## Razão de Covariâncias

A razão de covariâncias mede a influência da  $i$ -ésima observação sobre a matriz de covariância de  $\hat{\beta}$  e pode ser mensurada por

$$\begin{aligned} \frac{\det[\text{cov}(\text{vec}(\beta_{(I)}))]}{\det[\text{cov}(\text{vec}(\beta))]} &= \frac{\det[\hat{\Sigma}_{(i)} \otimes (X'_{(i)}X_{(i)})^{-1}]}{\det[\hat{\Sigma} \otimes (X'X)^{-1}]} \\ &= \frac{[\det(X'_{(i)}X_{(i)})^{-1}]^p [\det(\hat{\Sigma}_{(i)})]^{K+1}}{[\det(X'X)^{-1}]^p [\det(\hat{\Sigma})]^{K+1}} = \left( \frac{1}{1-p_i} \right)^p \left\{ \frac{\det[\hat{\Sigma}_{(i)}]}{\det[\hat{\Sigma}]} \right\}^{K+1} \end{aligned}$$

Valores altos e baixos da razão de covariâncias são considerados significantes. Um ponto de corte para valores baixos da razão de covariância pode ser obtido usando o fato que

$$\frac{\det[\hat{\Sigma}_{(i)}]}{\det[\hat{\Sigma}]} = \left\{ 1 + \frac{T_i^2}{n-K-2} \right\}^{-1}$$

demonstrado em Rao (1973).

## Influência das linhas de $\beta$

A influência da  $i$ -ésima observação sobre a  $j$ -ésima linha de  $\beta$  pode ser medida pela estatística

$$\frac{T_i^2}{1-p_i} \frac{w_{ij}^2}{W_j'W_j} \quad \begin{array}{l} j = 1, 2, \dots, K+1 \\ i = 1, 2, \dots, n \end{array}$$

onde  $w_{ij}$  é o  $i$ -ésimo elemento do vetor  $W_j = (I - P_{[j]})X_j$ ,  $X_j$  é a  $j$ -ésima coluna de  $X$ ,  $P_{[j]} = X_{[j]}(X'_{[j]}X_{[j]})^{-1}X'_{[j]}$  e  $X_{[j]}$  é a matriz  $X$  retirada a coluna  $j$ .

A  $i$ -ésima observação pode ser considerada influente sobre a  $j$ -ésima linha de  $\beta$  se  $\frac{w_{ij}^2}{W_j'W_j} > \frac{1}{n}$ . É importante observar que a  $j$ -ésima linha de  $\beta$  contém as estimativas dos coeficientes de regressão da variável  $X_j$  para as  $p$  regressões consideradas.

Definidas essas medidas, Houssain e Naik (1989) sugerem que a estratégia de análise de diagnóstico multivariada seja similar à estratégia usada no caso univariado, analisando-se nessa ordem:

1. O gráfico de dispersão de  $Y_j$  contra  $X_l$ ;  $l = 1, 2, \dots, k$  e  $j = 1, 2, \dots, p$ ;
2. Os elementos da matriz P;
3. A alavanca parcial em  $\frac{w_{ij}^2}{W_j'W_j}$ ;
4. A influência individual das observações através de  $C_i$  e  $WK_i$ ;
5. A influência das observações sobre a matriz de covariância de B;
6. A influência de observações individuais sobre os coeficientes estimados através de  $\frac{T_i^2}{1-p_i} \frac{w_{ij}^2}{W_j'W_j}$ .

As medidas propostas por Houssain e Naik (1989) são extensões dos métodos utilizados nos modelos de regressão univariada. Até o momento, foram apresentadas as medidas tradicionais. Os autores discutem ainda medidas obtidas através de outras abordagens, conforme veremos a seguir.

### **Afastamento da Verossimilhança**

Se  $L(\theta)$  é o logaritmo da função de verossimilhança do modelo baseado nos dados completos, Cook e Weisberg (1982) e Cook (1987) definem o afastamento da verossimilhança  $LD_i$  como

$$LD_i(\theta) = 2[L(\hat{\theta}) - L(\hat{\theta}_{(i)})]$$

em que  $\hat{\theta}_{(i)}$  denota as estimativas de máxima verossimilhança do vetor de parâmetros  $\theta$  sem a  $i$ -ésima observação e  $\hat{\theta}$  a estimativa de máxima verossimilhança com base nos dados completos. Os autores sugerem comparar  $LD_i$  com os quantis da distribuição quiquadrado com  $K+1$  graus de liberdade, sendo que  $K+1$  é a dimensão de  $\theta$ .

Para o modelo definido em (2.3), o afastamento da verossimilhança para  $\beta$  sem considerar a matriz de covariância  $\Sigma$  é

$$LD_I(\beta/\Sigma) = 2[L(\hat{\beta}, \hat{\Sigma}) - L(\hat{\beta}_{(i)}, \hat{\Sigma}(\hat{\beta}_{(i)}))]$$

onde  $\hat{\Sigma}(\hat{\beta}_i)$  é a estimativa de  $\Sigma$  quando  $\beta$  é estimado por  $\hat{\beta}_i$ . Depois de algumas simplificações obtemos

$$LD_I(\beta/\Sigma) = n \log \left( 1 + \frac{1}{(n-K-1)} \frac{p_i}{(1-p_i)^2} e_i' \hat{\Sigma}^{-1} e_i \right) = n \log \left( 1 + \frac{K+1}{n-K-1} C_i \right),$$

onde  $C_i$  é a distância de Cook multivariada definida anteriormente. Então observamos que o afastamento da verossimilhança é determinado por  $C_i$ , sendo uma função crescente dessa quantidade. Similarmente verifica-se que a estatística  $WK_i$  pode ser obtida pela diferença de dois valores de  $LD_I$ .

A influência de uma observação na estimação conjunta de  $(\beta, \Sigma)$  é definida como

$$LD_i(\beta, \Sigma) = 2[L(\hat{\beta}, \hat{\Sigma}) - L(\hat{\beta}_i, \hat{\Sigma}_i)] = n \log \left( \frac{|\hat{\Sigma}_{(i)}|}{|\hat{\Sigma}|} \right) + \frac{T_i^2}{1-p_i} - p.$$

Se  $\Sigma$  é o único parâmetro de interesse, então o afastamento da verossimilhança fica

$$LD_i(\Sigma/\beta) = 2[L(\hat{\beta}, \hat{\Sigma}) - L(\hat{\beta}(\hat{\Sigma}_{(i)}), \hat{\Sigma}_{(i)})], \quad (3.10)$$

onde  $\hat{\beta}(\hat{\Sigma}_{(i)})$  é a estimativa de  $\beta$  quando  $\Sigma$  é estimado por  $\hat{\Sigma}_{(i)}$ .

Entretanto  $\hat{\beta}(\hat{\Sigma}_{(i)}) = \hat{\beta}$  então, após calculos obtém-se

$$LD_i(\Sigma/\beta) = n \log \left\{ \frac{|\hat{\Sigma}_{(i)}|}{|\hat{\Sigma}|} \right\} + n[tr(\hat{\Sigma}_{(i)}^{-1} \hat{\Sigma}) - p]. \quad (3.11)$$

Verifica-se ainda que a diferença entre (3.10) e (3.11) resulta em

$$LD(\beta, \Sigma) - LD(\Sigma/\beta) = \frac{n-1}{n-K-2} WK_{(i)}$$

onde  $WK_{(i)}$  é a estatística de Welsch-Kuh definida anteriormente.

### 3.5 Exemplo 1

Para ilustrar a utilidade das medidas abordadas, Houssain e Naik (1989) utilizaram dados retirados de Anderson (1984, pg.369). As variáveis analisadas foram: Peso da Semente ( $Y_1$ ), Peso da Palha ( $Y_2$ ) e Quantidade de fertilizante ( $X_1$ ) para uma amostra de oito observações,

Tabela 3.2: Dado utilizados.

$Y_1$ :	40	17	9	15	6	12	5	9
$Y_2$ :	53	19	10	29	13	27	19	30
$X_1$ :	24	11	5	12	7	14	11	18

Dessa forma o modelo é:

$$Y_{8 \times 2} = X_{8 \times 2} \beta_{2 \times 2} + \varepsilon_{8 \times 2}$$

com as linhas de  $\varepsilon$  independentemente distribuídas com distribuição  $N_2(0, \Sigma)$ . A primeira coluna de  $X$  é um vetor de uns e a segunda coluna contém os valores da variável  $X_1$ . As colunas da matriz  $Y$  correspondem aos valores de  $Y_1$  e  $Y_2$  e a matriz de parâmetros  $B$  é da forma  $B = \begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \end{bmatrix}$ .

As estatísticas: lâmbda de Wilks, traço de Pillai, traço de Hotelling-Lawey e a maior raiz de Roi (apresentadas em Anderson(1984)) foram utilizadas para testar a hipótese  $H_0 : \Gamma = \begin{bmatrix} \beta_{11} \\ \beta_{12} \end{bmatrix} = 0$  contra a hipótese alternativa  $H_1 : \Gamma \neq 0$ . Todas as estatísticas rejeitam a hipótese nula, levando-nos a concluir pela significância da variável  $X_1$  na explicação das variáveis resposta  $Y_1$  e  $Y_2$ .

Na Tabela 3.3 apresentamos as medidas de influência  $C_i$ ,  $WK_i$  e  $RC_i$ .

Tabela 3.3: Medidas de Influência.

Observação	$C_i$	$WK_i$	$RC_i$
1	0,923	7,825	3,720
2	0,406	0,818	0,376
3	0,548	0,450	2,009
4	0,402	0,598	0,284
5	0,470	0,006	1,788
6	0,404	0,048	1,189
7	0,406	0,123	1,045
8	0,457	1,700	0,492

Analisando a tabela, concluímos que a observação 1 é influente segundo todas as medidas consideradas, observamos que este elemento apresenta valor extremo em todas as variáveis consideradas.

Além disso seus valores de  $Y_1$  e  $Y_2$  estão muito acima do que seria esperado para seu valor de  $X_1$  conforme sugerem as Figuras 3.3 e 3.4, que são diagramas de dispersão entre  $Y_1$  e  $X_1$  e  $Y_2$  e  $X_1$ .



Figura 3.3: Diagrama de dispersão entre  $Y_1$  e  $X_1$ .

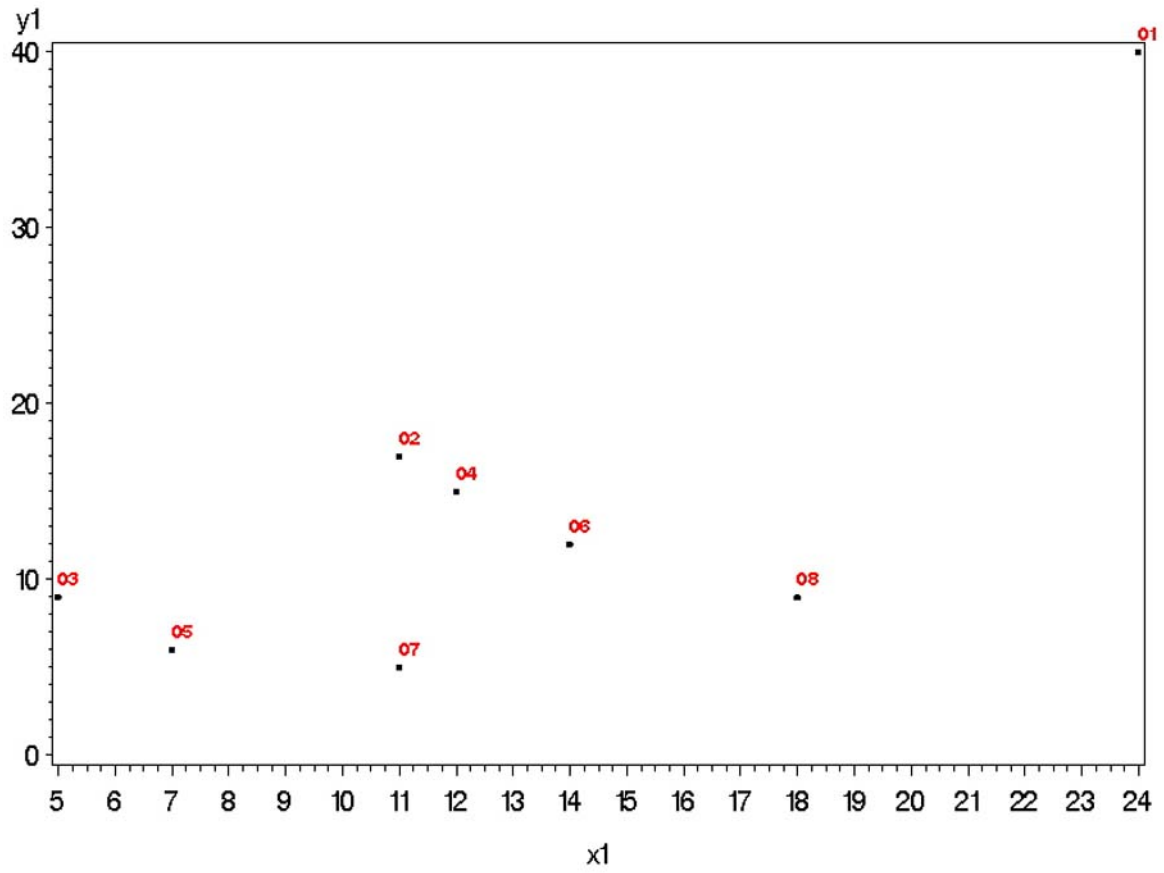
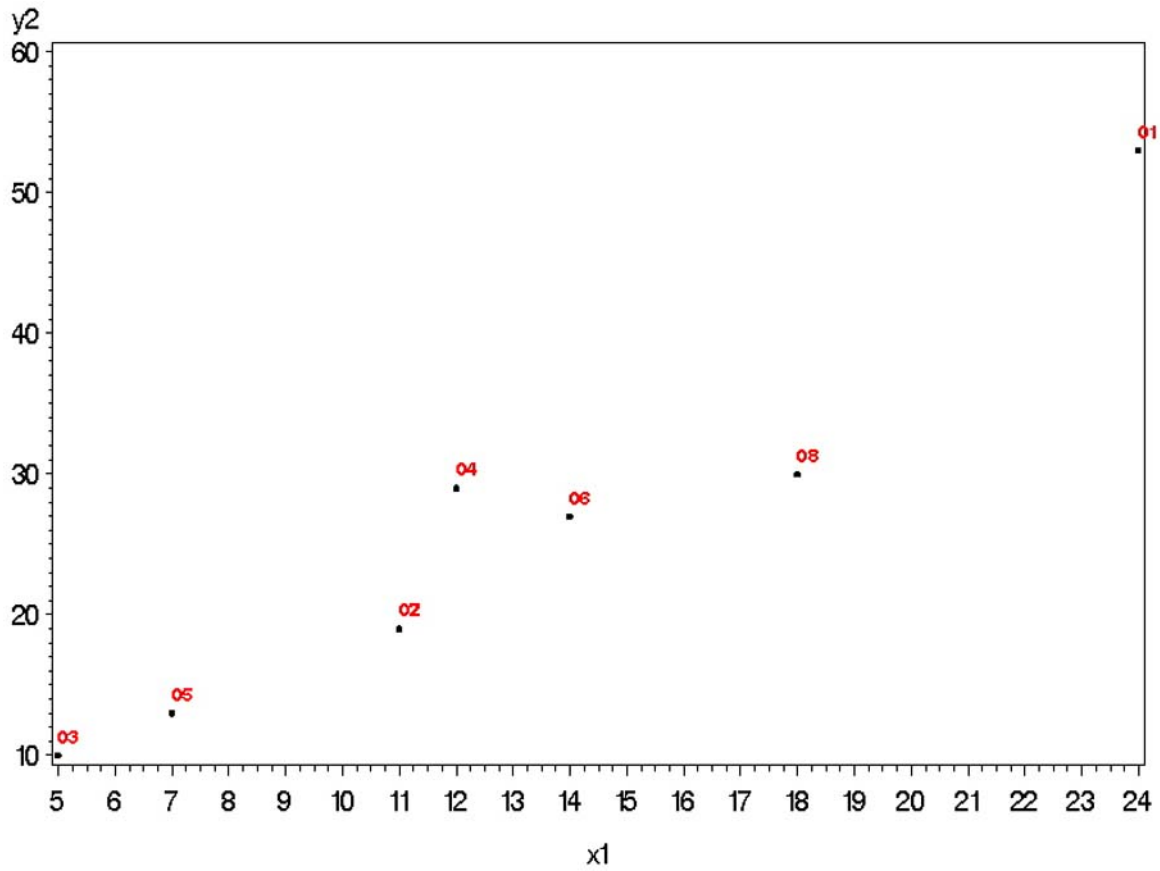


Figura 3.4: Diagrama de dispersão entre  $Y_2$  e  $X_1$ .



Neste capítulo, apresentamos várias classes de medidas de diagnóstico, a maioria delas extensões das medidas já existentes no contexto univariado. A análise de influência local, introduzida por Cook (1986), não foi considerada aqui e seria sugestão para um trabalho futuro.

# Capítulo 4

## Seleção de Variáveis

Neste capítulo iremos abordar alguns processos de seleção de variáveis independentes no modelo de Regressão Multivariada. Iniciaremos com o procedimento de testes simultâneos apresentado por McKay (1977) e posteriormente, discutiremos o Critério da Informação de Akaike (AIC), contemplado em Bedrick e Tsai (1994).

### 4.1 Hipóteses para Seleção de Variáveis

Suponha que o vetor aleatório  $\mathbf{X}$  tenha  $(m + p)$  componentes  $\mathbf{X}_1, \dots, \mathbf{X}_m, \mathbf{X}_{m+1}, \dots, \mathbf{X}_{m+p}$ , onde  $\mathbf{X}_g$  é um vetor composto por todas as observações da variável independente  $X_g$ . McKay (1977) supõe que  $\mathbf{X}$  tem distribuição normal  $(m + p)$ -variada com vetor de médias  $\boldsymbol{\mu}$  e matriz de covariância  $\Sigma$ . Consideremos  $u = \{1, \dots, m\}, s = \{m + 1, \dots, m + p\}$ , dessa forma podemos estabelecer as partições

$$\mathbf{X}' = (\mathbf{X}'_u, \mathbf{X}'_s), \quad \boldsymbol{\mu} = \begin{pmatrix} \boldsymbol{\mu}_u \\ \boldsymbol{\mu}_s \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \Sigma_{uu} & \Sigma_{us} \\ \Sigma_{su} & \Sigma_{ss} \end{pmatrix}.$$

Deseja-se avaliar a regressão de  $\mathbf{X}_u$  em  $\mathbf{X}_s$ . Esta regressão é a esperança condicional de

$\mathbf{X}_u$  dado  $\mathbf{X}_s$ , que na distribuição normal multivariada é da forma  $\beta_u + B_{us}\mathbf{X}_s$  onde

$$\beta_u = \mu_u - \Sigma_{us}\Sigma_{ss}^{-1}\mu_s \text{ e } B_{us} = \Sigma_{us}\Sigma_{ss}^{-1}.$$

A distribuição condicional de  $\mathbf{X}_u$  dado  $\mathbf{X}_s$  ( $\mathbf{X}_u|\mathbf{X}_s$ ) é normal m-variada com matriz de covariância  $\Sigma_{uu} - \Sigma_{us}\Sigma_{ss}^{-1}\Sigma_{su}$ .

A investigação das relações de regressão recaem naturalmente sobre a matriz  $B_{us}$  dos coeficientes da regressão. Se  $B_{us} = 0$  (ou, equivalentemente, se  $\Sigma_{us} = 0$ )  $\mathbf{X}_s$  não fornece nenhuma informação a respeito de  $\mathbf{X}_u$ . Porém se  $\Sigma_{us} \neq 0$  então  $\mathbf{X}_s$  fornece informação a respeito de  $\mathbf{X}_u$ . A qualidade das informações produzidas por  $\mathbf{X}_s$  pode ser melhorada determinando quais subconjuntos de  $\mathbf{X}_s$  são mais importantes na explicação de  $\mathbf{X}_u$ .

Para isso vamos considerar a partição  $\mathbf{X}_s' = (\mathbf{X}_f' \mathbf{X}_g')$  onde  $f \subset s$ ,  $g \subset s$  contendo  $p-r$  e  $r$  elementos respectivamente ( $1 \leq r < p$ ).

Intuitivamente,  $\mathbf{X}_f$  produz mais informação sobre  $\mathbf{X}_u$  do que  $\mathbf{X}_s$  se a matriz  $B_{ug,f}$ , que consiste das colunas de  $B_{us}$  associadas com  $\mathbf{X}_g$  é nula.

Todas as correlações entre  $\mathbf{X}_u$  e  $\mathbf{X}_s$  podem ser expressas em termos do quadrado das correlações canônicas

$$\rho_j^2 = c_j(\Sigma_{su}\Sigma_{uu}^{-1}\Sigma_{us}\Sigma_{ss}^{-1}),$$

em que  $c_j(A)$  denota a  $j$ -ésima maior raiz característica da matriz ( $A$ ),  $j = 1, \dots, s$  [ $s \leq \min(m, p)$ ].

As correlações de  $\rho_1^2, \dots, \rho_s^2$ , conjuntamente, medem a informação sobre as variações de  $\mathbf{X}_u$  vindas da relação de regressão com  $\mathbf{X}_s$ . Dessa forma  $\rho_j^{*2} = c_j(\Sigma_{fu}\Sigma_{uu}^{-1}\Sigma_{uf}\Sigma_{ff}^{-1})$   $j = 1, \dots, s^*$  [ $s^* \leq \min(m, p-r)$ ] mede a informação sobre as variações de  $\mathbf{X}_u$  vindas da relação de regressão com  $\mathbf{X}_f$ .

Se  $\rho_j^2 \geq \rho_j^{*2}$ ,  $j = 1, \dots, s^*$  ( $\leq s$ ) então  $\mathbf{X}_f$  não fornece mais informação que  $\mathbf{X}_s$ . Além disso,  $\rho_j^2 = \rho_j^{*2}$  se, e somente se,  $B_{ug,f} = 0$  (este fato está demonstrado no apêndice de McKay, 1977).

A menos que exista alguma informação, à priori, sugerindo que certos componentes de  $\mathbf{X}_s$  possuem uma importância especial, será necessário determinar se  $B_{ug.f} = 0$  para todas as possíveis partições  $(\mathbf{X}_f' \mathbf{X}_g')$  de  $\mathbf{X}_s'$ . Além disso, uma extensiva investigação da utilidade de  $\mathbf{X}_s$  em produzir informações sobre os componentes de  $\mathbf{X}_u$ , usualmente, irá requerer a análise da regressão do vetor  $\mathbf{X}_v$  (com  $q$  variáveis) sobre  $\mathbf{X}_s$ , onde  $v \subseteq u$  (e  $1 \leq q \leq m$ ) para uma ou todas possíveis escolhas de  $v$ .

Seja qual for a escolha, não há regressão de  $\mathbf{X}_v$  em  $\mathbf{X}_s$  se  $B_{vs} = 0$ , com as colunas de  $B_{vs}$  sendo constituídas pelas linhas de  $B_{us}$  correspondentes a  $\mathbf{X}_v$ . Se  $B_{vs} \neq 0$  então, pelos mesmos argumentos anteriores com  $v$  substituindo  $u$ , a importância relativa dos componentes de  $\mathbf{X}_s$  em prover informação sobre  $\mathbf{X}_v$  pode ser examinada determinando-se as partições  $(\mathbf{X}_f' \mathbf{X}_g')$  de  $\mathbf{X}_s'$  para as quais a matriz  $B_{vg.f}$ , formada pelas colunas de  $B_{vs}$  correspondentes a  $g$ , é igual à matriz nula.

Essa investigação requer testes simultâneos de um grande número de hipóteses  $w_{vs} : B_{vs} = 0$  especificada pela variação de  $v$  sobre todos os subconjuntos (não vazios) de  $u$ , e de um número de hipóteses  $w_{vg.f} : B_{vg.f} = 0$  especificadas variando  $v$  e pela variação de  $f$  (com  $s - f = g$ ) sobre todos os subconjuntos não vazios de  $s$ .

Apresentaremos a seguir uma breve descrição da construção de testes simultâneos, conforme introduzido em Morrison (1976).

## 4.2 Testes Simultâneos

Consideremos o modelo linear geral

$$X = A\xi + \varepsilon$$

onde  $X$  é formada por  $n$  vetores linha  $p$ -dimensionais,  $A$  é a matriz de planejamento e  $\xi$  matriz de parâmetros desconhecidos,  $q \times p$ .

A hipótese  $w_{vs}$  é um caso particular da hipótese linear geral:

$$\begin{cases} H_0 : C\xi M = 0 \\ H_1 : C\xi M \neq 0 \end{cases} \quad (4.1)$$

onde  $C$  é uma matriz  $g \times q$  referente às hipóteses dos elementos de uma dada coluna da matriz de parâmetros  $\xi$ . A matriz  $M$  possui dimensão  $p \times l$  e posto  $l \leq p$ . Podemos particionar as matrizes  $A$  e  $C$  em:  $A = [A_1 \ A_2]$  e  $C = [C_1 \ C_2]$  sendo  $A_1$  submatriz com  $r$  colunas e  $C_1$  submatriz de ordem  $g \times r$ .

A pós-multiplicação de  $C\xi$  por  $M$  permite a geração de hipóteses relativas a diferentes parâmetros de resposta.

A aplicação do teste da união-intersecção para a hipótese (4.1) permite-nos concluir que a hipótese multivariada é verdadeira se, e somente se, as hipóteses univariadas

$$H_0 : C\xi M a = 0 \quad (4.2)$$

são verdadeiras para todos os componentes não nulos do vetor  $a$ ,  $l$ -dimensional. A estatística de teste para qualquer uma das hipóteses univariadas é dada pela expressão

$$F(a) = \frac{(n-r)a' M' X' A_1 (A_1' A_1)^{-1} C_1' [C_1 (A_1' A_1)^{-1} C_1']^{-1} C_1 (A_1' A_1)^{-1} A_1' X M a}{ga' M' X' [I - A_1 (A_1' A_1)^{-1} A_1'] X M a}$$

Para um teste univariado de nível de significância  $\alpha$ , não rejeita-se a hipótese (4.2) se  $F(a) \leq F_{\alpha;g,n-r}$  onde  $F_{\alpha;g,n-r}$  é o percentil de ordem  $\alpha$  de uma distribuição F-Snedecor com  $g$  graus de liberdade no numerador e  $n-r$  graus de liberdade no denominador.

Não rejeita-se a hipótese multivariada (4.1) a um nível de significância  $\alpha$ , se  $\{F(a) \leq F_{\alpha;g,n-r}\}$  para todo vetor  $a \neq 0$ .

Esta região de não rejeição pode ser alternativamente definida por  $\max_a F(a) \leq F_{\alpha;g,n-r}$ .

Pode-se mostrar que  $F(a)$  é proporcional à maior raiz da equação

$$|H - \lambda E| = 0, \quad (4.3)$$

onde  $H = M' X' A_1 (A_1' A_1)^{-1} C_1' [C_1 (A_1' A_1)^{-1} C_1']^{-1} C_1 (A_1' A_1)^{-1} A_1' X M$  e  
 $E = M' X' [I - A_1 (A_1' A_1)^{-1} A_1'] X M$ .

Denotando a maior raiz de (4.3) por  $c_j$  onde  $j = \min(g, l)$ , não rejeita-se (4.1) a um nível de significância  $\alpha$  se

$$c_j \leq c(\alpha)$$

e rejeitamos caso contrário. Em que  $c(\alpha)$  é o quantil de ordem  $\alpha$  da distribuição da maior raiz característica quando (4.1) é verdadeira. Os quantis superiores dessa distribuição encontram-se tabelados (ver Morrison, 1976).

As raízes não nulas de  $|H - \lambda E| = 0$  são iguais as raízes características não nulas de  $HE^{-1}$ , e na prática é usualmente mais eficiente determinar as raízes características de  $HE^{-1}$ . É importante notar que  $HE^{-1}$  não necessariamente é simétrica, mas verifica-se que suas raízes características são números reais e não-negativos.

Como caso particular de (4.3) temos as hipóteses  $w_{us}$ ,  $w_{vs}$  e  $w_{vg,f}$  exercendo o papel de  $H_0$ , com matriz  $\xi = (B_u \ B_{us})$ . Para  $w_{vs}$  e  $w_{vg,f}$ , as respectivas matrizes  $H$  são  $S_{vs} S_{ss}^{-1} S_{sv}$  e  $S_{vs} S_{ss}^{-1} S_{sv} - S_{vf} S_{ff}^{-1} S_{fv}$ , onde  $S = \sum_{i=1}^n X_i X_i' - \{(\sum_{i=1}^n X_i)(\sum_{i=1}^n X_i')\} \frac{1}{n}$ .

Quando  $w_{vs}$  ou  $w_{vg,f}$  é verdadeira, a matriz  $H$  segue uma distribuição Wishart central com  $p$  ou  $r$  graus de liberdade. Tanto para  $w_{vs}$  quanto para  $w_{vg,f}$ , a matriz de erros é  $S_{vv} - S_{vs} S_{ss}^{-1} S_{sv}$  e segue distribuição Wishart com  $n - p - 1$  graus de liberdade.

Em ambos os casos, a matriz de hipóteses e de erros são independentes. Estas distribuições são condicionais para cada  $X_{si}$  fixado  $i = 1, \dots, n$ .

Para um nível de significância  $\alpha$  rejeita-se  $w_{vs}$  quando

$$c_1 \{S_{vs} S_{ss}^{-1} S_{sv} (S_{vv} - S_{vs} S_{ss}^{-1} S_{sv})^{-1}\} > \tau_{m,p,n-p-1}^\alpha$$

onde  $\tau_{m,p,n-p-1}^\alpha$  denota o quantil de ordem  $\alpha$  da distribuição da maior raiz característica. Para obter os quantis superior dessa distribuição consultando Morrison (1976) precisamos fixar  $\alpha$ ,

as entradas da tabela são  $s = \min(m, p)$ ,  $m$  e  $n - p - 1$ . Aqui  $H = S_{vs}S_{ss}^{-1}S_{sv}$  e  $E = S_{ss} - S_{vs}S_{ss}^{-1}S_{sv}$ . Nota-se que o valor crítico  $\tau_{m,p,n-p-1}^\alpha$  é usado em todos os testes.

A hipótese (4.3) também pode ser testada através das estatísticas

- traço de Hotelling-Lawley:  $T^2 = ntr(E^{-1}H)$ ;
- lambda de Wilks:  $\lambda = \det(E)/\det(E + H)$  e
- traço de Pillai:  $V = tr\{(E + H)^{-1}H\}$

entre outras.

No início da discussão, assumimos que  $\mathbf{X}$  tem distribuição normal  $(m+p)$ -variada. Esta suposição pode ser relaxada desde que a regressão de  $\mathbf{X}_u$  em  $\mathbf{X}_s$  seja linear nos parâmetros e a distribuição de  $\mathbf{X}_u|\mathbf{X}_s$  seja normal  $m$ -variada. De fato as variáveis em  $\mathbf{X}_s$  podem ser variáveis controladas; neste caso ainda faz sentido falarmos em correlação canônica se, como foi sugerido por Draper e Smith (1966), assumirmos que os valores de  $\mathbf{X}_s$  provém de uma população finita. Notemos ainda que as submatrizes de soma de quadrados e a matriz produto  $S$  presentes nas estatísticas envolvidas no procedimento de teste simultâneo podem, sem qualquer ajuste no procedimento, ser substituídas pelas submatrizes de correlação amostral, conforme descrito em Morrison (1976).



### 4.3 Exemplo 2

O procedimento proposto foi aplicado ao conjunto de dados de Waugh (1942), que examinou a relação entre cinco características de uma variedade de trigo canadense e quatro características da farinha obtida a partir dessa variedade. As características são:

Variáveis associadas à Farinha	Variáveis associadas ao Trigo
$X_1$ : Quantidade de trigo por barril de farinha	$X_5$ : Textura da semente
$X_2$ : Quantidade de cinzas	$X_6$ : Peso
$X_3$ : Quantidade de proteína crua	$X_7$ : Quantidade de sementes danificadas
$X_4$ : Índice de qualidade do glúten	$X_8$ : Material externo
	$X_9$ : Quantidade de proteína crua

As características da farinha estão representadas em  $\mathbf{X}_u$  e as características do trigo estão representadas em  $\mathbf{X}_s$ . Como o interesse é determinar quais características da farinha podem ser explicadas pelas características do trigo, o foco aqui está centrado nas regressões de subconjuntos de características da farinha,  $\mathbf{X}_u$ , no conjunto de características do trigo,  $\mathbf{X}_s$ .

Foi calculada a matriz de correlações amostrais das características  $X_1, \dots, X_9$  baseadas em  $n = 138$  observações, denotada por  $R$ , e  $R_{us}$  denota a submatriz de  $R$  consistindo de coeficientes de correlação amostral entre os componentes de  $\mathbf{X}_u$  e  $\mathbf{X}_s$ .

A hipótese de não existência de correlação entre o vetor  $\mathbf{X}_u$  e  $\mathbf{X}_s$  é rejeitada, com nível de significância  $\alpha = 0,10$ , se  $c_1\{HE^{-1}\}$  exceder o valor crítico  $\tau_{m,p,n-p-1}^\alpha = \tau_{4,5,132}^{0,10} \approx 0,15$  onde  $H = R_{us}R_{ss}^{-1}R_{su}$ ,  $E = R_{uu} - R_{us}^{-1}R_{su}$  e  $\tau_{4,5,132}^{0,10}$  é o quantil da distribuição da maior raiz característica com entradas  $s = \min\{4, 5\}$ ,  $m = 4$  e  $n - p - 1 = 132$ .

O autor chama de conjunto adequado a todo conjunto de características do trigo que não for rejeitado pelo procedimento de teste simultâneo para explicar um particular conjunto de características da farinha. Além disso, um conjunto de características do trigo é denomi-

nado conjunto mínimo adequado para um particular conjunto de características da farinha, se nenhum subconjunto dele é adequado.

Foram testadas todas as possíveis combinações entre as características do trigo e da farinha. Assim, para o conjunto de características  $X_1, X_2, X_4$  todos os conjuntos adequados são

(i)  $X_5, X_6, X_7, X_8$ ;

(ii)  $X_5, X_7, X_8, X_9$ ;

(iii)  $X_6, X_7, X_8, X_9$ ;

(iv)  $X_5, X_7, X_8$ .

De acordo com a definição, observamos que os conjuntos mínimos adequados são  $\{X_5, X_7, X_8\}$  e  $\{X_6, X_7, X_8, X_9\}$ .

O conjunto de características (iv) fornece tanta informação com respeito à  $X_1, X_2, X_4$  quanto os conjuntos apresentados em (i), (ii), (iii). Como possui o menor número de elementos, é o subconjunto que deve ser selecionado.

Uma alternativa aos testes simultâneos baseados na máxima raiz característica são os testes simultâneos baseados na razão de verossimilhanças. Aqui a estatística do teste da hipótese  $w_{vs}$  é

$$W_{vs} = \frac{(1 - U_{vs})}{U_{vs}} \quad \text{onde} \quad U_{vs} = \frac{|S_{vv} - S_{vs}S_{ss}^{-1}S_{sv}|}{|S_{vv}|}.$$

Para testar  $w_{vg.f}$ , a estatística é  $W_{vg.f} = \frac{(1 - U_{vg.f})}{U_{vg.f}}$  onde  $U_{vg.f} = |S_{vv} - S_{vs}S_{ss}^{-1}S_{sv}| / |S_{vv} - S_{vf}S_{ff}^{-1}S_{fv}|$ . Entretanto, Gabriel (1968, 1969) mostra que os testes simultâneos baseados na máxima raiz característica são mais poderosos.

## 4.4 Uso do Critério de Informação de Akaike

Quando se utiliza regressão como técnica de análise de dados, frequentemente determina-se vários modelos como candidatos a modelo final a ser adotado. Um método para escolher o modelo final, muito utilizado, é o Critério de Informação de Akaike (1973), (*AIC*), que foi concebido inicialmente para ser um estimador não viesado da esperança da Informação de Kullback-Leibler. Segundo esse critério, o modelo a ser adotado é aquele que possui o menor valor de *AIC*, sendo que essa medida é definida como

$$AIC = -2\log L(\hat{\beta}, \hat{\Sigma}) + 2q$$

onde  $L(\hat{\beta}, \hat{\Sigma})$  é a função de verossimilhança do modelo calculada em  $\beta = \hat{\beta}$  e  $\Sigma = \hat{\Sigma}$ , estimadores de máxima verossimilhança de  $\beta$  e  $\Sigma$  e  $q$  é o número de parâmetros do modelo.

Embora este critério seja muito utilizado, pode apresentar sérias deficiências. Para os modelos de regressão linear normal, Hurvich e Tsai (1989) mostraram que a medida *AIC* pode subestimar a esperança da Informação de Kullback-Leibler quando o tamanho da amostra é pequeno e quando o número de parâmetros do modelo é moderadamente elevado em comparação ao tamanho da amostra. Em seu artigo, Hurvich e Tsai (1989) apresentaram uma versão corrigida do *AIC*, denominada *AIC<sub>c</sub>* que é um estimador não viesado para a esperança da informação de Kullback-Leibler e tende a fornecer melhores escolhas de modelos para pequenas amostras. A extensão para os modelos de Regressão Multivariada foi apresentada por Bedrick e Tsai (1994).

Assim, seja o modelo dado em (2.3),  $Y_{n \times p} = X_{n \times (K+1)}\beta_{(K+1) \times p} + \varepsilon_{n \times p}$ , cuja função de verossimilhança sem os termos constantes é

$$L(\beta, \Sigma) = -0,5n \log |\Sigma| - 0,5tr\{(Y - X\beta)\Sigma^{-1}(Y - X\beta)'\},$$

o estimador de máxima verossimilhança de  $\beta$  é  $\hat{\beta} = (X'X)^{-1}X'Y$  e de  $\Sigma$  é  $\hat{\Sigma} = Y'(I - X(X'X)^{-1}X')Y/n$ . Sob este modelo  $vec(\hat{\beta}) \sim N_{(K+1)p}(vec(\beta), \Sigma \otimes (X'X)^{-1})$ . Além disso,  $\hat{\Sigma}$  e  $\hat{\beta}$  são independentes com  $n\hat{\Sigma} \sim Wishart_p(\Sigma, n - K - 1)$  e  $E(\hat{\Sigma}^{-1}) = d\Sigma^{-1}$ , onde  $d = n/\{n - (m + p + 1)\}$ , como demonstrado em Anderson (1984; pg. 270, 290).

Suponha que as variáveis resposta são da forma

$$Y = X_0\beta_0 + \varepsilon_0 \quad (4.4)$$

onde  $X_0$  tem dimensão  $n \times (K_0 + 1)$ ,  $\beta_0$  tem dimensão  $(K_0 + 1) \times p$ ,  $\varepsilon_0$  tem dimensão  $n \times p$  e as linhas de  $\varepsilon_0$  são independentes com média zero e matriz de variância e covariância  $\Sigma_0$ .

Uma medida muito útil para avaliar a discrepância entre o modelo verdadeiro (4.4) e o modelo ajustado (2.3) é a informação de Kullback-Leibler

$$\Delta(\beta, \Sigma) = E_0\{-2L(\beta, \Sigma)\} = n \log |\Sigma| + ntr(\Sigma^{-1}\Sigma_0) + tr\{(X_0\beta_0 - X\beta)\Sigma^{-1}(X_0\beta_0 - X\beta)'\},$$

onde  $E_0$  denota a esperança sob o modelo verdadeiro. Um critério razoável para julgar a qualidade do ajuste é  $E_0\{\Delta(\hat{\beta}, \hat{\Sigma})\}$ , onde  $\hat{\beta}, \hat{\Sigma}$  são os estimadores de máxima verossimilhança de  $\beta$  e  $\Sigma$  no modelo (2.3).

Dessa forma, dada uma coleção de modelos ajustados, o melhor modelo seria o que minimiza  $E_0\{\Delta(\hat{\beta}, \hat{\Sigma})\}$ .

O número de parâmetros desconhecidos do modelo ajustado é  $p(K + 1) + 0,5p(p + 1)$ . Então a medida *AIC* fica:

$$AIC = n(\log |\hat{\Sigma}| + p) + 2\{p(K + 1) + 0,5p(p + 1)\}$$

A constante  $np$  foi incluída para facilitar a comparação com a medida  $AIC_c$  definida a seguir. Esta modificação na definição usual do *AIC* não exerce impacto no critério do menor *AIC* porque esse ajuste independe do número de variáveis preditoras.

O *AIC* foi desenvolvido para ser um estimador aproximadamente não viesado para  $E_0\{\Delta(\hat{\beta}, \hat{\Sigma})\}$  quando o modelo ajustado inclui o modelo verdadeiro como um caso especial (Linhart e Zucchini, 1986, pg.245 citado em Bedrick e Tsai, 1994).

Sob esta suposição, as colunas de  $X$  podem ser rearranjadas de tal forma que  $X_0\beta_0 = X\beta^*$ , onde  $\beta^* = \begin{bmatrix} \beta_0 \\ \beta_1 \end{bmatrix}$  e  $\beta_1$  é uma matriz de zeros de ordem  $(K - K_0) \times p$ . Dessa forma,

excluí-se  $(K - K_0)$  variáveis independentes do modelo. Neste caso,

$$\Delta(\hat{\beta}, \hat{\Sigma}) = n \log |\hat{\Sigma}| + ntr(\hat{\Sigma}^{-1}\Sigma_0) + tr\{X(\beta^* - \hat{\beta})\hat{\Sigma}^{-1}(\beta^* - \hat{\beta})'X'\}.$$

Devido a propriedades de  $\hat{\beta}$  e  $\hat{\Sigma}$ , encontra-se  $E_0\{tr(\hat{\Sigma}^{-1}\Sigma_0)\} = dp$  e

$$\begin{aligned} E_0[tr\{X(\beta^* - \hat{\beta})\hat{\Sigma}^{-1}(\beta^* - \hat{\beta})'X'\}] &= E_0[tr\{\hat{\Sigma}^{-1}(\hat{\beta} - \beta^*)'(X'X)(\hat{\beta} - \beta^*)\}] \\ &= tr[E_0(\hat{\Sigma}^{-1})E_0\{(\hat{\beta} - \beta^*)'(X'X)(\hat{\beta} - \beta^*)\}] \\ &= d \ tr[\Sigma_0^{-1}E_0\{(\hat{\beta} - \beta^*)'(X'X)(\hat{\beta} - \beta^*)\}] \\ &= d \ E_0[vec(\hat{\beta} - \beta^*)'\{\Sigma_0^{-1} \otimes (X'X)\}vec(\hat{\beta} - \beta^*)] \\ &= dpK \end{aligned}$$

onde  $d = \frac{n}{n-(K+p+2)}$ , sendo que a última igualdade é consequência do fato que  $vec(\hat{\beta} - \beta^*)'\{\Sigma_0^{-1} \otimes (X'X)\}vec(\hat{\beta} - \beta^*)$  tem distribuição quiquadrado com  $pK$  graus de liberdade.

Portanto,  $E_0\{\Delta(\hat{\beta}, \hat{\Sigma})\} = E_0(n \log |\hat{\Sigma}|) + dpn + dp(K + 1)$  e assim

$$AIC_c = n \log |\hat{\Sigma}| + dp(n + k) = n(\log |\hat{\Sigma}| + p) + 2d\{p(K + 1) + 0,5p(p + 1)\}$$

é um estimador não viesado de

$$E_0\{\Delta(\hat{\beta}, \hat{\Sigma})\}.$$

É importante notar que a única diferença entre  $AIC$  e  $AIC_c$  é a inclusão do fator de escala  $d$  em  $AIC_c$ . Além disso, como  $d > 1$ ,  $AIC$  subestima  $E_0\{\Delta(\hat{\beta}, \hat{\Sigma})\}$ .

Bedrick e Tsai (1994) realizaram um estudo de simulação para avaliar nove critérios de seleção de modelos com resposta multivariada. Além das medidas  $AIC$  e  $AIC_c$  foram avaliadas as medidas

- $SIC = n \log |\hat{\Sigma}| + \{p(K + 1) + 0,5p(p + 1)\} \log(n)$  (Schwarz, 1978);
- $HQ = n \log |\hat{\Sigma}| + 2\{p(K + 1) + 0,5p(p + 1)\} \log(\log(n))$  (Hannan e Quinn, 1979);
- $ICOMP = 0,5(n + K + 1)\{p \log(tr(\hat{\Sigma})/p) - \log |\hat{\Sigma}|\} + 0,5p\{(K + 1) \log(tr(X'X)^{-1}/(K + 1)) - \log |(X'X)^{-1}|\}$  (Bozdogan, 1990);

- $PRESS = \sum_{i=1}^n \sum_{j=1}^p \hat{u}_{ij}^2 / (1 - h_i)$ , onde  $\hat{U} = [\hat{u}_{ij}] = Y - X\hat{\beta}$ , e  $h_i$  é o  $i$ -ésimo elemento da diagonal da matriz  $H = X(X'X)^{-1}X'$ , (Bedrick e Tsai, 1994).

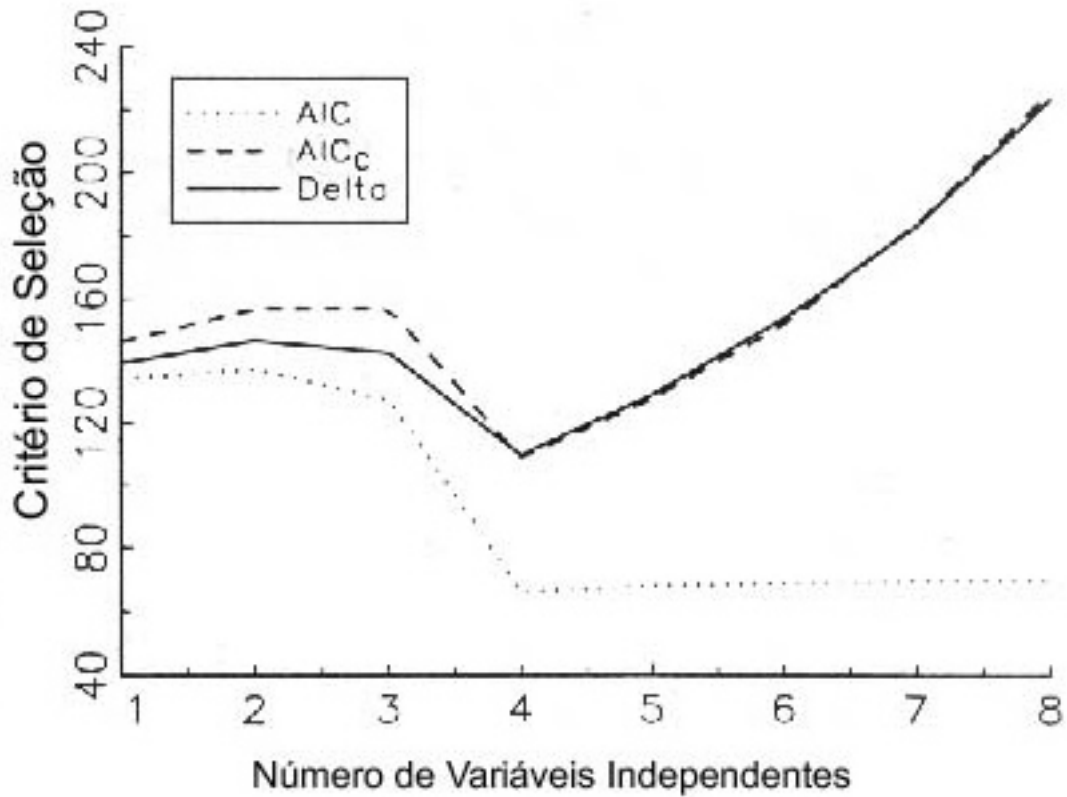
Os critérios a seguir foram avaliados com relação ao modelo completo com matriz de covariância  $\hat{\Sigma}_{max}$  e  $K_{max}$  variáveis independentes:

- $C_p^1 = |\hat{\Sigma}_{max}\hat{\Sigma}| - \{(n - k)/(n - K_{max})\}^p$  (Sparks, Coutsourides e Troskie, 1983);
- $C_p^2 = (n - K_{max})tr(\hat{\Sigma}_{max}^{-1}\hat{\Sigma}) + (K + 1 - n)p$  (Sparks, Coutsourides e Troskie, 1983);
- $FIC = n|\hat{\Sigma}| + |\hat{\Sigma}_{max}| \log |\hat{\Sigma}^{-1} \otimes (X'X)|$  (Bedrick e Tsai, 1994).

Quinhentas amostras de tamanho 20 e 35 foram geradas usando o modelo (4.4) com  $k_0 = 4$ , matriz de covariância  $\Sigma_0 = (1 - \rho)I_p + \rho J_p$ , onde  $J_p$  é uma matriz  $p \times p$  de uns, a matriz  $X$  possui 7 colunas sendo que as 4 primeiras são as colunas da matriz de planejamento do modelo verdadeiro  $X_0$ . A simulação considerou ainda cada combinação de  $\rho = 0, 3; 0, 6; 0, 8$  e  $p = 2; 4$ . O modelo selecionado é o que apresenta menor valor da medida.

A Tabela 4.1 apresenta o número de vezes que cada modelo foi selecionado em quatro casos particulares, resultantes das combinações dos valores de  $n$  e  $p$ . A Figura 4.1 mostra que a medida  $AIC_c$  possui um desempenho melhor que o  $AIC$ . Numa etapa seguinte foi construído um gráfico dos valores médios de  $AIC$ ,  $AIC_c$  e  $\Delta(\hat{\beta}, \hat{\Sigma})$  em função de  $K$ , para  $n = 20$ . Verificou-se que à medida que  $K$  se aproximava de  $K_0 = 4$ , os valores de  $\Delta(\hat{\beta}, \hat{\Sigma})$  decresciam rapidamente. Quando  $K > K_0$  os valores médios de  $AIC_c$ , que é não viesado para  $E\{\Delta(\hat{\beta}, \hat{\Sigma})\}$ , são praticamente iguais a  $\Delta(\hat{\beta}, \hat{\Sigma})$ , por outro lado, observou-se que  $AIC$  subestimava  $E\{\Delta(\hat{\beta}, \hat{\Sigma})\}$  drasticamente.

Figura 4.1: Desempenho dos Critérios de Seleção.



Quando o fator de correção  $d = \frac{n}{n-(K+p+2)}$  é muito maior que 1,  $AIC$  subestima substancialmente  $E\{\Delta(\hat{\beta}, \hat{\Sigma})\}$ . Isto ocorre quando o número de covariáveis  $K$  mais o número de respostas  $p$  são elevados em relação ao tamanho da amostra  $n$ . Com base nessa pesquisa empírica, Bedrick e Tsai (1994) sugerem usar  $AIC_c$  ao invés de  $AIC$  como critério de seleção de modelos. Constatou-se ainda um melhor desempenho de  $AIC_c$  com relação às demais medidas utilizadas.

Tabela 4.1: Número de vezes que o modelo de ordem  $K$  foi selecionado.

p	n	K	$AIC$	$AIC_c$	$SIC$	$HQ$	$PRESS$	$C_p^1$	$C_p^2$	$ICOMP$	$FIC$
2	20	1-3	0	0	0	0	0	0	0	0	0
		4	277	480	402	305	364	131	157	373	321
		5	70	19	48	65	63	72	68	43	65
		6	36	0	16	38	32	66	56	38	58
		7	48	1	18	40	23	86	81	20	29
		8	69	0	16	52	18	145	138	26	27
2	35	1-3	0	0	0	0	0	0	0	0	0
		4	346	457	473	411	365	163	177	405	426
		5	64	35	20	49	67	63	58	26	58
		6	36	7	6	18	31	67	67	29	11
		7	26	1	1	13	24	82	78	28	4
		8	28	0	0	9	13	125	120	12	1
4	20	1-3	0	0	0	0	0	0	0	0	0
		4	247	499	425	293	387	75	121	412	163
		5	60	1	41	55	63	43	58	40	77
		6	49	0	17	43	29	58	57	21	134
		7	46	0	7	37	14	96	78	17	80
		8	98	0	10	72	7	228	186	10	46
4	35	1-3	0	0	0	0	0	0	0	0	0
		4	379	489	489	447	411	124	164	378	432
		5	49	9	9	29	42	56	56	57	39
		6	37	1	1	14	19	58	61	26	21
		7	22	0	0	9	14	100	82	22	6
		8	13	1	1	1	14	162	137	17	2

Fonte: Bedrick e Tsai (1994)

No presente capítulo, apresentamos métodos clássicos de seleção de variáveis independentes no



modelo de Regressão Multivariada. Métodos de seleção de variáveis num contexto bayesiano, que não será abordado aqui, podem ser encontrados em Ibraim e Chen (1994).

# Capítulo 5

## Aplicação

Neste capítulo faremos uma aplicação da técnica de estimação apresentada no Capítulo 2, da análise de diagnóstico apresentada no Capítulo 3 e da seleção de modelos apresentada no Capítulo 4.

Os dados utilizados são um subconjunto dos dados provenientes do estudo O Impacto do Estresse Materno no Baixo Peso ao Nascimento (RAE-CEA-2000-23, Artes e Lobert, 2000). Contamos nessa aplicação com medidas antropométricas, socio-econômicas e variáveis que avaliam o nível de estresse de 50 gestantes do município de Jundiaí, no último trimestre de gestação. O estresse materno foi avaliado através de quatro variáveis distintas: Resultado do teste de Estado de ansiedade (EA), Resultado do teste de Traço de Ansiedade (TA), Resultado do Questionário Geral de Saúde (QGS) e Escala de Percepção de Estresse (EPE).

Como variáveis resposta utilizamos o peso da criança ao nascer (Peso), medido em gramas, e a idade gestacional do recém nascido (IG), medida em semanas. As variáveis explicativas ou independentes são: peso materno (kg), altura materna (m), idade (anos), renda *per capita*, resultado do teste de estado de ansiedade, resultado do teste de traço de ansiedade, resultado do questionário geral de saúde e escala de percepção de estresse. Os valores das variáveis encontram-se no Apêndice C.

Para essa aplicação, usamos o pacote estatístico SAS versão 8.0, que ajusta através do procedimento proc GLM o modelo dado em (2.3), isto é,  $Y_{n \times p} = X_{n \times (K+1)}\beta_{(K+1) \times p} + \varepsilon_{n \times p}$  supondo que  $vec(\varepsilon) \sim N(0, I_n \otimes \Sigma)$ . No caso  $n = 50$  e  $p = 2$ .

A matriz de parâmetros  $\beta$  é estimada por  $b = (X'X)^{-1}X'Y$  e a matriz de variância e covariância  $\Sigma$  é estimada por  $e = [(e'e)/(n-r)] = [(Y-Xb)'(Y-Xb)/(n-r)]$  onde  $r$  é o posto da matriz  $X$ .

O pacote estatístico SAS testa a hipótese linear geral multivariada  $H_0 : L\beta M = 0$  onde  $L$  é uma matriz referente às hipóteses dos elementos de uma dada coluna da matriz de parâmetros e  $M$  é uma matriz tal que a pós multiplicação de  $L\beta$  por ela permite a geração de hipóteses relativas a diferentes parâmetros de resposta.

Além disso, podemos obter desse pacote estatístico as estatísticas de teste multivariada sendo que algumas foram apresentadas no Capítulo 4:

- Lâmbda de Wilks =  $det(E)/det(E+H)$ ;
- Traço de Pilai =  $traço(H(H+E)^{-1})$ ;
- Traço de Hotelling =  $traço(E^{-1}H)$ ;
- Máxima Raiz de Roy =  $\lambda$  (maior auto-valor de  $E^{-1}H$ ).

onde

$$H = M'(Lb)'(L(X'X)^{-1}L'(Lb)M)$$

$$E = M'(Y'Y - b(X'X)b)M$$

que testam a significância conjunta das oito variáveis independentes, ou seja, testa  $H_0 : \Gamma =$

$$\begin{bmatrix} \beta_{01} & \beta_{02} \\ \beta_{11} & \beta_{12} \\ \vdots & \vdots \\ \beta_{81} & \beta_{82} \end{bmatrix} = 0.$$

Podemos obter ainda as medidas de diagnóstico: Distância de Cook, Razão de Co-variâncias, DFFITS, além da matriz Alavanca  $H = X(X'X)^{-1}X$ , os valores preditos pelos modelos, os resíduos ordinários (*observado – estimado*), os resíduos studentizados (com e sem a  $i$ -ésima observação), e os intervalos de confiança para os valores preditos e para os valores médios.

Para a variável resposta bivariada peso e idade gestacional do recém-nascido, ajustamos dois modelos, cada um considerando um conjunto diferente de variáveis explicativas referentes ao estresse materno. Devido à possível correlação entre as variáveis explicativas resultado do teste de estado de ansiedade e traço de ansiedade, estas variáveis foram colocadas em conjuntos distintos. Na Tabela 5.1 encontram-se as variáveis explicativas utilizadas no ajuste dos modelos e na Figura 5.1 apresentamos os diagramas de dispersão entre as variáveis resposta e explicativas.

Tabela 5.1: Variáveis Explicativas Utilizadas no Ajuste dos Modelos.

Modelo 1	Modelo 2
Peso Materno (PESOM)	Peso Materno (PESOM)
Altura Materna (ALTURAM)	Altura Materna (ALTURAM)
Idade Materna (IDADEM)	Idade Materna (IDADEM)
Renda <i>per capita</i>	Renda <i>per capita</i>
Estado de Ansiedade (EA)	Traço de Ansiedade (TA)
Questionário Geral de Saúde (QGS)	Escala de Percepção de Estresse (EPE)

Figura 5.1: Diagramas de Dispersão.

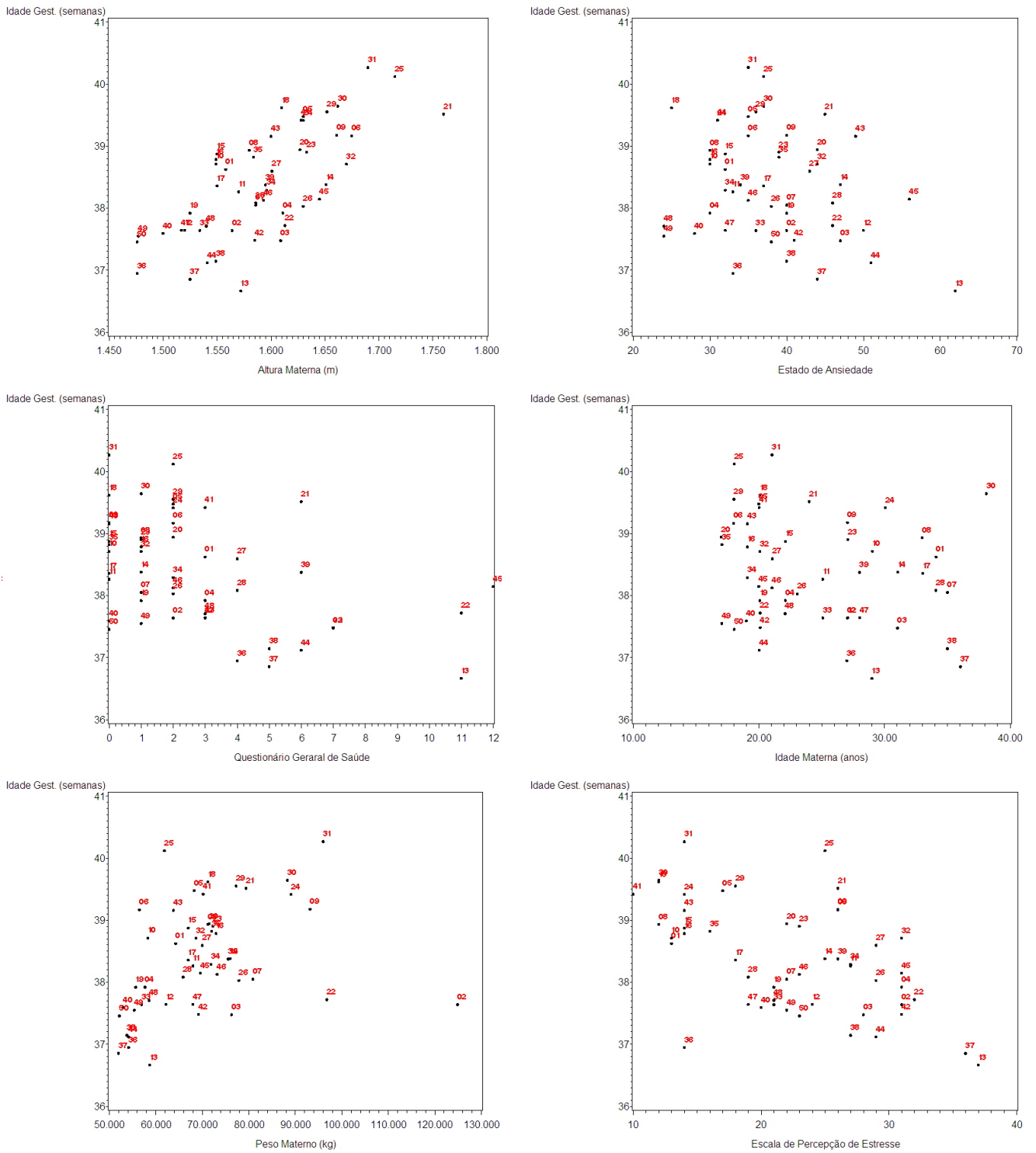


Figura 5.1: Diagramas de Dispersão.

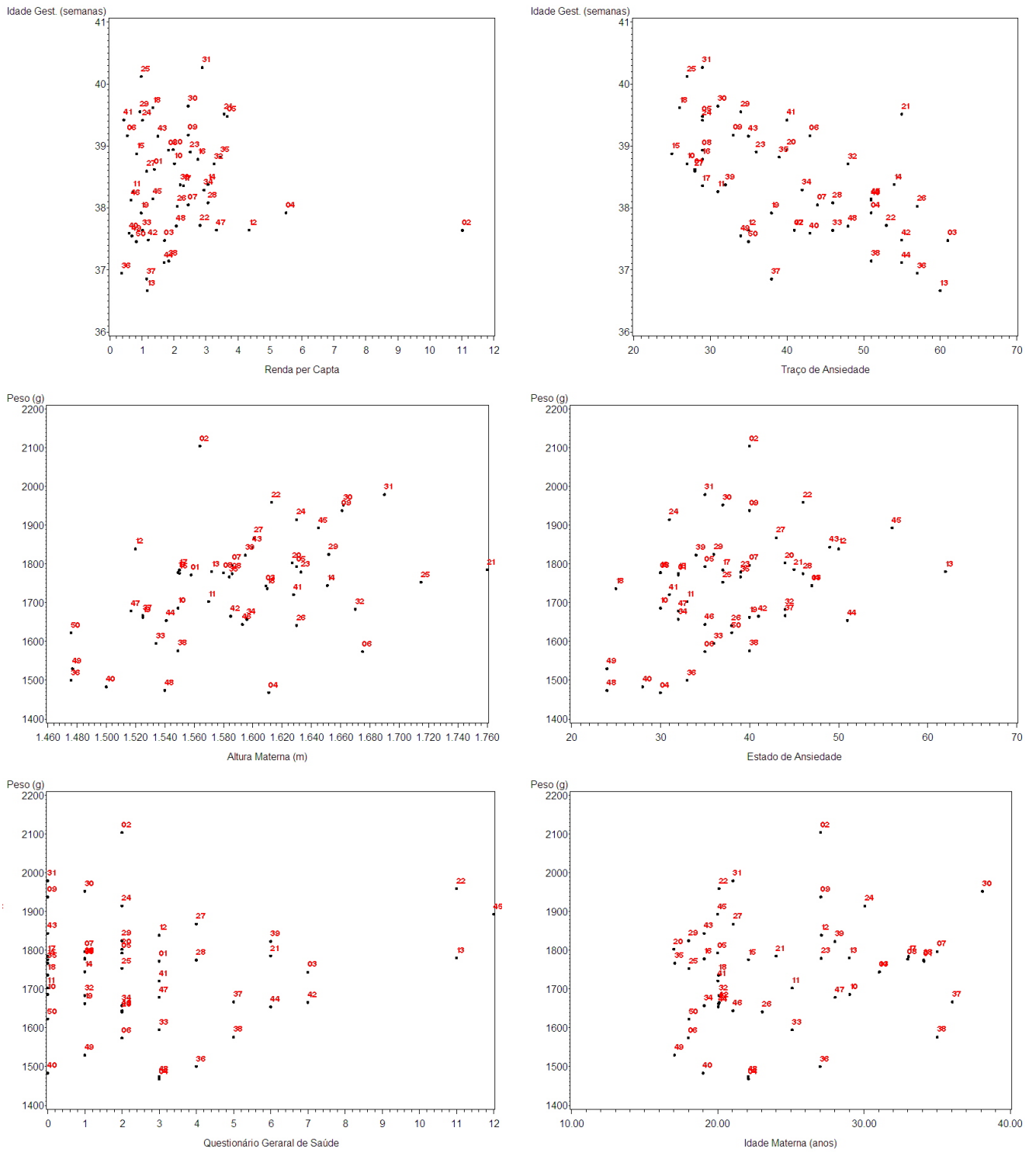
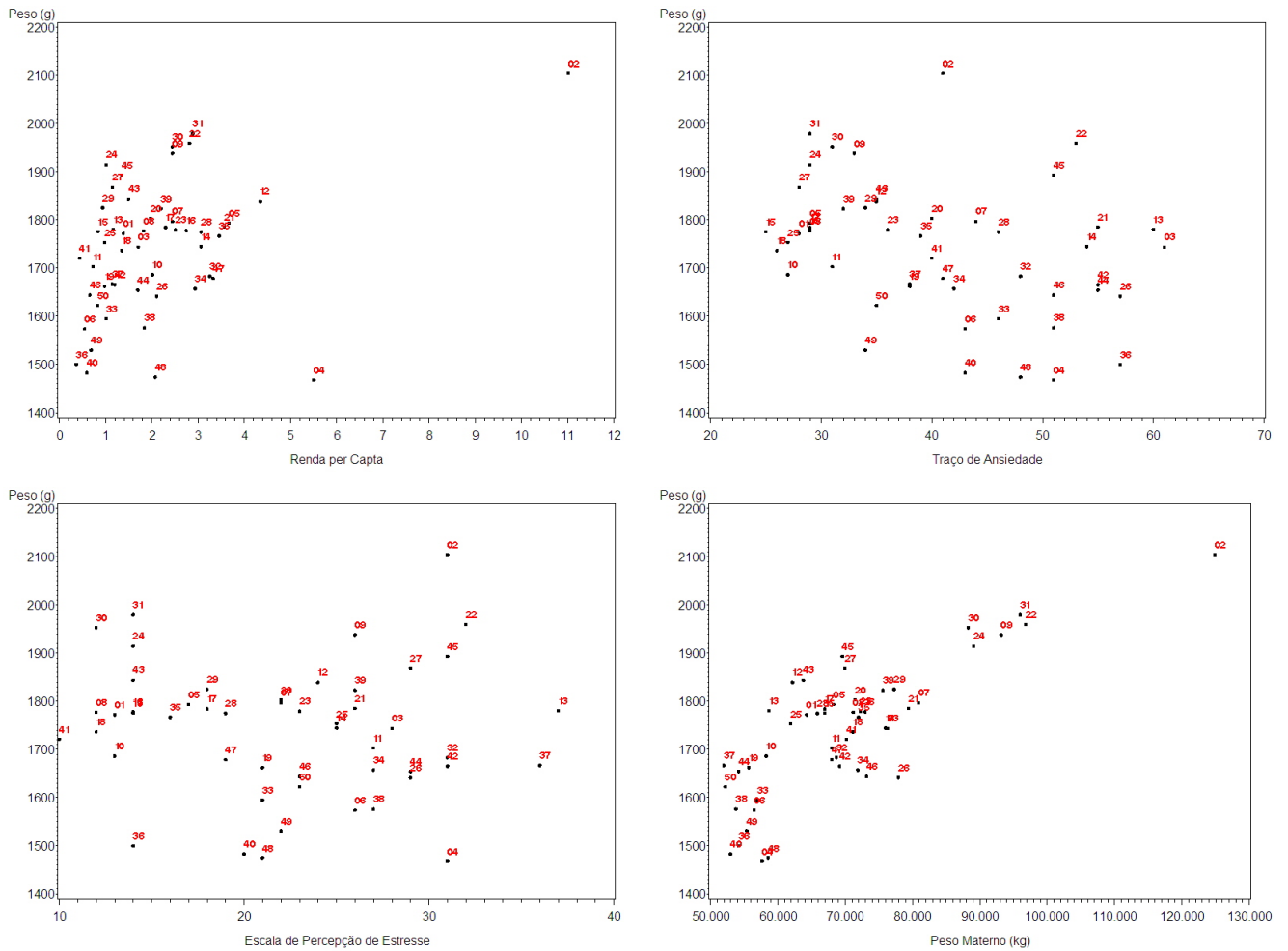


Figura 5.1: Diagramas de Dispersão.



Nas Tabelas 5.2 e 5.3 apresentamos as variáveis estatisticamente significantes, com nível de significância de 5%, para explicar o peso e a idade gestacional do recém-nascido nos modelos 1 e 2 respectivamente. Nessa análise, utilizamos os testes apresentados na Seção 2.4.

Tabela 5.2: Estimativas dos Parâmetros do Modelo 1

	Parâmetro	Estimativa	Erro Padrão	P-Valor
Peso	Intercepto	925,57	288,87	0,0026
	PESOM	8,53	1,05	<0,0001
	ALTURAM	-26,57	199,24	0,8945
	IDADEM	1,50	1,73	0,3925
	RENDA	-13,15	7,61	0,0913
	EA	7,16	1,64	<0,0001
	QGS	-7,18	4,29	0,1012
IG	Intercepto	22,61	1,62	<0,0001
	PESOM	0,01	0,01	0,1301
	ALTURAM	10,45	1,12	<0,0001
	IDADEM	-0,01	0,01	0,2877
	RENDA	-0,09	0,04	0,0347
	EA	-0,02	0,01	0,0447
	QGS	-0,12	0,02	<0,0001



Tabela 5.3: Estimativas dos Parâmetros do modelo 2.

	Parâmetro	Estimativa	Erro Padrão	P-Valor
Peso	Intercepto	894,50	302,69	0,0051
	PESOM	7,32	1,12	<0,0001
	ALTURAM	248,83	204,23	0,2297
	IDADEM	2,95	1,80	0,1088
	RENDA	-5,61	8,19	0,4974
	TA	-4,61	1,24	0,0006
	EPE	2,89	1,91	0,1381
IG	Intercepto	24,34	0,17	<0,0001
	PESOM	0,00	0,00	0,3382
	ALTURAM	10,65	0,12	<0,0001
	IDADEM	-0,02	0,00	<0,0001
	RENDA	0,01	0,00	0,0917
	TA	-0,03	0,00	<0,0001
	EPE	-0,05	0,00	<0,0001

No modelo 1 para a variável Peso, os coeficientes das variáveis explicativas estatisticamente diferentes de zero são: Peso Materno e Estado de Ansiedade. Já para a variável Idade Gestacional temos a Altura Materna, Renda *per capita*, Estado de Ansiedade e Questionário Geral de Saúde.

No modelo 2 para a variável Peso, os coeficientes estatisticamente diferentes de zero são: Peso Materno e Traço de Ansiedade. Já para a variável Idade Gestacional temos a Altura Materna, Idade Materna, Traço de Ansiedade e Escala de Percepção de Estresse.

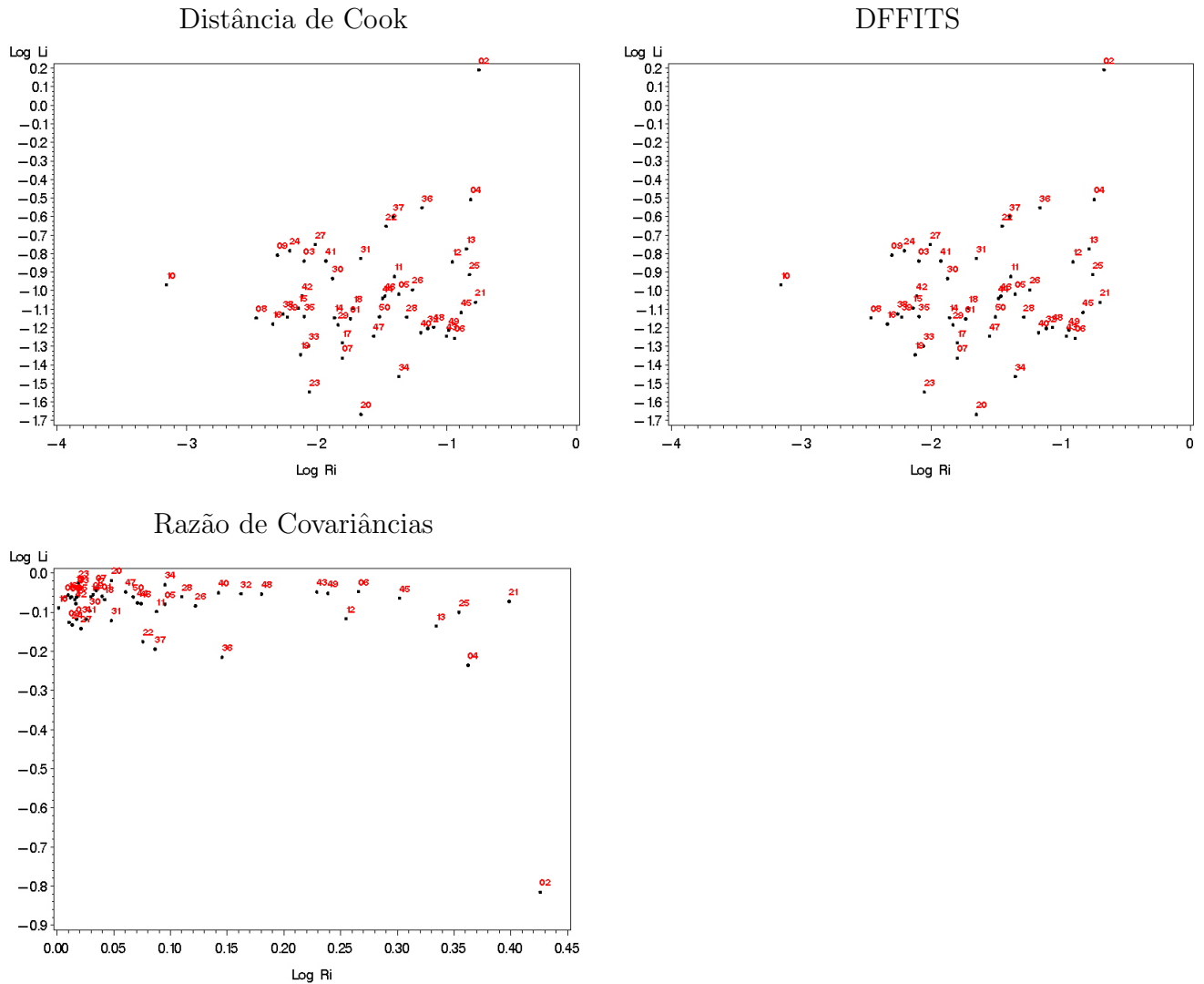
Para selecionar um dos dois modelos de Regressão Multivariada considerados, vamos utilizar o critério de Akaike com fator de correção sugerido por Bedrick e Tsai (1994),  $AIC_c$ . No modelo 1, calculamos  $AIC_c$  retirando a variável Idade Materna, que se mostrou não significativa em ambas as regressões, de modo que a matriz de planejamento seja a mesma em ambos os casos. Feito isso, obtivemos  $AIC_c = 288,68$ .

No modelo 2, a medida  $AIC_c$  foi calculada retirando-se a variável Renda *per capita* que se mostrou não significativa em ambas as regressões. Para o modelo obtido temos  $AIC_c = 170,33$ . Pelo critério de Akaike, o melhor modelo é o que apresenta menor  $AIC_c$ , então o modelo final selecionado será o modelo 2, sem a variável Renda *per capita*. No Apêndice A apresentamos a rotina que calculou a medida  $AIC_c$  para o modelo selecionado.

Afim de realizar a análise de diagnóstico apresentada no Capítulo 3, para o modelo selecionado, foram construídas rotinas em SAS utilizando o procedimento proc IML, avaliando a alavanca e os componentes dos resíduos através das medidas Distância de Cook, DFFITS e Razão de Covariâncias. No Apêndice B, apresentamos a rotina que calculou a medida distância de Cook para tal modelo.

A Figura 5.2 ilustra a influência de cada observação no modelo selecionado para as medidas distância de Cook, DFFITS e Razão de Covariâncias.

Figura 5.2: Influência individual.

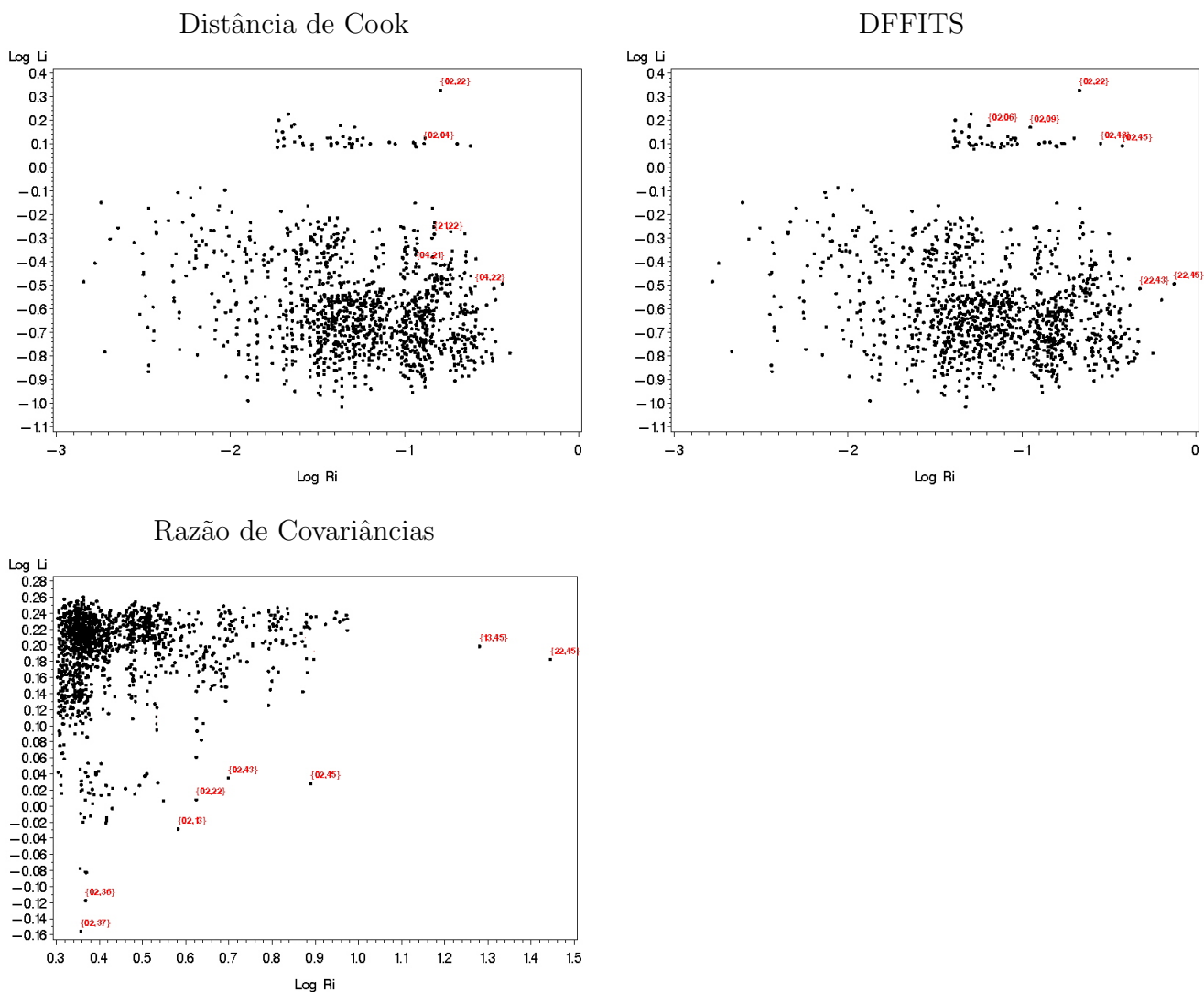


Concluimos através das medidas distância de Cook e DFFITS que a observação {02} possui a maior influência. Analisando esta observação, verificamos que refere-se à gestante com maior peso e que seu bebê é o mais pesado de todos. Tal observação já se mostrara discrepante nos gráficos de dispersão da Figura 5.1. Já a observação {10} possui o valor mais negativo para os componentes de resíduos. Essa observação refere-se à gestante cujo peso está entre os 25% menores, o Traço de Ansiedade está entre os 5% menores e a Escala de Percepção de Estresse está entre os 10% menores. Segundo a medida Razão de Covariâncias, além da observação {02} destacamos a observação {04}. Trata-se da gestante cujo peso está entre os 25% menores e o

seu bebê é o de menor peso. Destacamos também a observação {21} associada à gestante cujo peso está entre os 25% maiores e o peso de seu bebê está entre os 25% maiores.

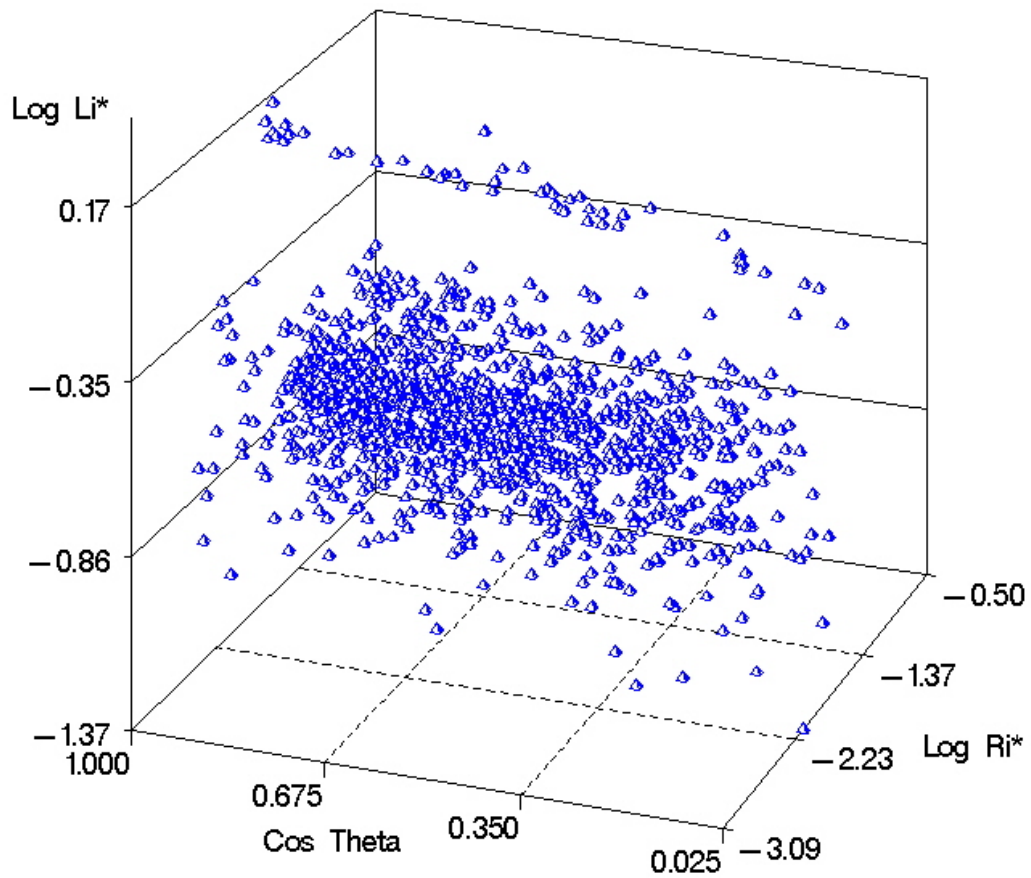
Quando analisamos a influência de subconjuntos de tamanho dois, para as três medidas consideradas, todos os que possuem como um dos elementos a observação {02} possuem maior influência. Na Figura 5.3 destacamos alguns casos.

Figura 5.3: Influência dos subconjuntos de tamanho dois.



Na Figura 5.7, construímos um gráfico tridimensional que considera o ângulo entre  $\log L_i$  e  $\log R_i$  para a medida distância de Cook.

Figura 5.7: Influência dos subconjuntos de tamanho dois.



Na Tabela 5.4 apresentamos os valores das medidas Distância de Cook, DFFITS e Razão de Covariâncias.

Tabela 5.4: Medidas de Diagnóstico.

n	Distância de Cook		DFFITS		Razão de Covariâncias	
	log Li	log Ri	log Li	log Ri	log Li	log Ri
1	-1,15	-1,74	-1,15	-1,73	-0,06	0,04
2	0,19	-0,75	0,19	-0,66	-0,82	0,43
3	-0,84	-2,1	-0,84	-2,09	-0,12	0,02
4	-0,51	-0,81	-0,51	-0,74	-0,24	0,36
5	-1,02	-1,37	-1,02	-1,35	-0,08	0,1
6	-1,26	-0,94	-1,26	-0,89	-0,05	0,27
7	-1,36	-1,8	-1,36	-1,8	-0,04	0,03
8	-1,15	-2,46	-1,15	-2,46	-0,06	0,01
9	-0,81	-2,3	-0,81	-2,3	-0,13	0,01
10	-0,97	-3,15	-0,97	-3,15	-0,09	0
11	-0,92	-1,4	-0,92	-1,38	-0,1	0,09
12	-0,84	-0,96	-0,84	-0,9	-0,12	0,26
13	-0,77	-0,85	-0,77	-0,78	-0,14	0,33
14	-1,14	-1,86	-1,14	-1,86	-0,06	0,03

Tabela 5.4: Medidas de Diagnóstico.

n	Distância de Cook		DFFITS		Razão de Covariâncias	
	log Li	log Ri	log Li	log Ri	log Li	log Ri
15	-1,09	-2,14	-1,09	-2,14	-0,07	0,02
16	-1,18	-2,34	-1,18	-2,34	-0,06	0,01
17	-1,28	-1,8	-1,28	-1,79	-0,04	0,03
18	-1,09	-1,72	-1,09	-1,71	-0,07	0,04
19	-1,34	-2,12	-1,34	-2,12	-0,04	0,02
20	-1,67	-1,66	-1,67	-1,65	-0,02	0,05
21	-1,06	-0,78	-1,06	-0,7	-0,07	0,4
22	-0,65	-1,46	-0,65	-1,45	-0,17	0,08
23	-1,55	-2,05	-1,55	-2,05	-0,02	0,02
24	-0,78	-2,21	-0,78	-2,2	-0,13	0,01
25	-0,91	-0,82	-0,91	-0,75	-0,1	0,35
26	-0,99	-1,26	-0,99	-1,24	-0,08	0,12
27	-0,75	-2,01	-0,75	-2,01	-0,14	0,02
28	-1,14	-1,31	-1,14	-1,28	-0,06	0,11
29	-1,18	-1,83	-1,18	-1,83	-0,06	0,03
30	-0,93	-1,88	-0,93	-1,87	-0,1	0,03
31	-0,82	-1,66	-0,82	-1,65	-0,12	0,05
32	-1,2	-1,14	-1,2	-1,11	-0,05	0,16
33	-1,3	-2,06	-1,3	-2,06	-0,04	0,02
34	-1,46	-1,37	-1,46	-1,35	-0,03	0,1
35	-1,14	-2,09	-1,14	-2,09	-0,06	0,02

Tabela 5.4: Medidas de Diagnóstico.

n	Distância de Cook		DFFITS		Razão de Covariâncias	
	log Li	log Ri	log Li	log Ri	log Li	log Ri
36	-0,55	-1,19	-0,55	-1,16	-0,21	0,15
37	-0,6	-1,41	-0,6	-1,39	-0,19	0,09
38	-1,12	-2,26	-1,12	-2,26	-0,06	0,01
39	-1,14	-2,23	-1,14	-2,22	-0,06	0,01
40	-1,23	-1,2	-1,23	-1,17	-0,05	0,14
41	-0,84	-1,93	-0,84	-1,92	-0,12	0,03
42	-1,03	-2,11	-1,03	-2,11	-0,08	0,02
43	-1,24	-1	-1,24	-0,95	-0,05	0,23
44	-1,04	-1,49	-1,04	-1,48	-0,08	0,07
45	-1,12	-0,89	-1,12	-0,83	-0,06	0,3
46	-1,03	-1,47	-1,03	-1,46	-0,08	0,07
47	-1,24	-1,56	-1,24	-1,55	-0,05	0,06
48	-1,2	-1,1	-1,2	-1,06	-0,05	0,18
49	-1,21	-0,98	-1,21	-0,93	-0,05	0,24
50	-1,14	-1,52	-1,14	-1,5	-0,06	0,07



# Capítulo 6

## Conclusão

Nesta dissertação foram estudados os modelos de Regressão Linear Multivariada. Inicialmente foi apresentado o método de estimação dos parâmetros por mínimos quadrados em duas situações: a primeira, em que para cada uma das  $p$  regressões a matriz de planejamento é a mesma e a segunda em que cada regressão possui a sua própria matriz de planejamento.

Uma das vantagens em se utilizar os modelos de Regressão Multivariada, em relação ao caso em que se ajusta as  $p$  regressões separadas, é que quando estimamos conjuntamente os parâmetros das  $p$  regressões, obtemos um ganho de eficiência dos estimadores.

Quando se utiliza a técnica de Regressão Multivariada, as principais medidas de análise de diagnóstico também podem ser utilizadas (com as devidas modificações) para a determinação de observações influentes, sendo que a análise de diagnóstico mais adequada deve levar em consideração conjuntos de observações e não apenas a análise de observações individuais.

Para essa técnica de modelagem também contamos com procedimentos de seleção de variáveis, sendo que um dos melhores procedimentos é o baseado na medida  $AIC_c$ , pois é um estimador não viesado para a esperança da medida de Informação de Kullback-Leibler e possui melhor desempenho que a medida  $AIC$  para situações em que o tamanho amostral é pequeno, além disso é facilmente implementado frente a outros métodos.

Para estudos futuros sugere-se a análise da influência local e métodos de seleção de variáveis baseados na teoria bayesiana.

# Apêndice A

## Programa de Cálculo do $AIC_c$

Encontra-se a seguir a rotina que construímos através do pacote estatístico SAS (1999), que calcula a medida  $AIC_c$  para o modelo selecionado.

```
proc iml;
use dados_finala;
namesy={'Peso' 'IG'};
namesx={'PESOM' 'ALTURAM' 'IDADEM' 'TA' 'EPE'};
namecod={'Codges'};
/* Armazenando as variáveis em matrizes*/
read all var{peso,ig} into y [colname=namesy];
read all var{PESOM ALTURAM IDADEM TA EPE} into m [colname=namesx];
read all var{Dcodges} into cod [colname=namecod];
/*contando o número de linhas da matriz m*/
l=nrow(m);
/*contando o número de linhas da matriz y*/
l1=nrow(m);
/*contando o número de colunas da matriz m*/
```

```

c=ncol(m);
/*contando o número de colunas da matriz y*/
c1=ncol(y);
/*adicionando uma coluna de uns na matriz m*/
x=j(1,1,1)||m[,1:c];
namescol='Intercepto'||namesx[,1:c];
mattrib x colname=namescol;
/* estimando beta */
beta=inv(t(x)*x)*t(x)*y;
/* resíduo ordinário */
e=y-x*beta;
/*contando o número de colunas de x*/
c2=ncol(x);
/*Calculando o posto da matrix X*/
rank=round(trace(ginv(X)*X));
/*Estimando a matriz de variância e covariância */
sigma_hat=(t(e)*e)/(l1-rank);
/* Calculando o AICc*/
d=l1/(l1-(c2+c1+1));
AIC_c=l1*log10(det(sigma_hat))+d*c1*(l1+c2);
start aic;
/*Imprime o AICc na tela */
print AIC_c;
finish;
run aic;
quit;

```

# Apêndice B

## Programas de Análise de Influência

Encontra-se a seguir a rotina que construímos através do pacote estatístico SAS (1999), que calcula a influência das observações individuais e posteriormente, a rotina que calcula a influência das observações para subconjuntos de tamanho dois, utilizando a medida Distância de Cook. Essa medida é calculada para o modelo selecionado.

```
proc iml;
  use dados_finala;
  namesy={'Peso' 'IG'};
  namesx={'PESOM' 'ALTURAM' 'IDADEM' 'TA' 'EPE'};
  namecod={'Codges'};
  /* Armazenando as variáveis em matrizes*/
  read all var{peso,ig} into y [colname=namesy];
  read all var{PESOM ALTURAM IDADEM TA EPE} into m [colname=namesx];
  read all var{Dcodges} into cod [colname=namecod];
  /*contando o número de linhas da matriz m*/
  l=nrow(m);
  /*contando o número de colunas da matriz m*/
```

```

c=ncol(m);
/*contando o número de colunas da matriz y*/
c1=ncol(y);
/*adicionando uma coluna de uns na matriz m*/
x=j(1,1,1)||m[,1:c];
namescol='Intercepto'||namesx[,1:c];
mattrib x colname=namescol;
/* estimando beta */
beta=inv(t(x)*x)*t(x)*y;
/* resíduo ordinário */
e=y-x*beta;
/*Matriz Identidade*/
ident=I(2);
/*contando o número de colunas de x*/
c2=ncol(x);
/*contando o número de colunas de e*/
c3=ncol(e);
inversax=inv(t(x)*x);
inversae=inv(t(e)*e);
start influencia;
i=1;
do while(i <= l);
xi=x[i,1:c2];
yi=y[i,1:c1];
ei=e[i,1:c3];
hi=xi*inversax*t(xi);
qi=ei*inversae*t(ei);
/* Componente de Alavanca */

```

```

Li=hi/(1-hi);
/* Componente dos Resíduos*/
Ri=(1-hi)**-0.5*qi*(1-hi)**-0.5;
/* Traço das Matrizes Li e Ri */
trLi=trace(Li);
trRi=trace(Ri);
/* Armazenando o número da gestante */
codges=cod[i,1];
/* Gerando o Arquivo de Saída */
name_res='LI' || 'RI' || 'Cod';
influence=trLi||trRi||codges//influence;
mattrib influence [colname=name_res];
i=i+1;
end;
create saida3 from influence [colname=name_res];
append from influence;
finish;
run influencia;
quit;

proc iml;
use dados_finala;
namesy={'Peso' 'IG'};
namesx={'PESOM' 'ALTURAM' 'IDADEM' 'TA' 'EPE'};
namecod={'Codges'};
/* Armazenando as variáveis em matrizes*/
read all var{peso,ig} into y [colname=namesy];
read all var{PESOM ALTURAM IDADEM TA EPE} into m [colname=namesx];

```

```

read all var{Dcodges} into cod [colname=namecod];
/*contando o número de linhas da matriz m*/
l=nrow(m);
/*contando o número de linhas da matriz y*/
l1=nrow(m);
/*contando o número de colunas da matriz m*/
c=ncol(m);
/*contando o número de colunas da matriz y*/
c1=ncol(y);
/*adicionando uma coluna de uns na matriz m*/
x=j(1,1,1)||m[,1:c];
namescol='Intercepto'||namesx[,1:c];
mattrib x colname=namescol;
/* estimando beta */
beta=inv(t(x)*x)*t(x)*y;
/* resíduo ordinário */
e=y-x*beta;
/*Matriz Identidade*/
ident=I(2);
/*contando o número de colunas de x*/
c2=ncol(x);
/*contando o número de colunas de e*/
c3=ncol(e);
inversax=inv(t(x)*x);
inversae=inv(t(e)*e);
/*Calculando o posto da matrix X*/
rank=round(trace(ginv(X)*X));
/*Estimando a matriz de variância e covariância */

```



```

sigma_hat=(t(e)*e)/(l1-rank);
start influencia;
i=1;
do while(i < l);
do j = i+1 to l;
xi=x[i,1:c2]/x[j,1:c2];
yi=y[i,1:c1]/y[j,1:c1];
ei=e[i,1:c3]/e[j,1:c3];
hi=xi*inversax*t(xi);
qi=ei*inversae*t(ei);
/* Componente de Alavanca */
Li=hi*inv(ident-hi);
/* Componente dos Resíduos*/
Ri=(ident-hi)**-0.5*qi*(ident-hi)**-0.5;
/* Calculando Norma de Li e Ri*/
norma_Li=sum(t(EIGVAL(Li))*EIGVAL(Li)**0.5);
norma_Ri=sum(t(EIGVAL(Ri))*EIGVAL(Ri)**0.5);
/* Transformando em vetor */
vecLi=shape(Li,4);
vecRi=shape(Ri,4);
/* Ângulo entre os vetores */
cos_theta=sum(vecLi#vecRi)/(norma_Li*norma_Ri);
LI_star=norma_Li*(cos_theta)**0.5;
RI_star=norma_Ri*(cos_theta)**0.5;
/* Traço das Matrizes Li e Ri */
trLi=trace(Li);
trRi=trace(Ri);
codges=cod[i,1]||cod[j,1];

```

```
name_res='LI' || 'RI' || 'LI_st' || 'RI_st' || 'COS' || 'Par1' || 'Par2';
influence=trLi || trRi || LI_Star || RI_star || cos_theta || codges // influence;
mattrib influence [colname=name_res];
end;
i=i+1;
end;
create saida from influence [colname=name_res];
append from influence;
finish;
run influencia;
quit;
```

# Apêndice C

## Dados da Aplicação

Tabela B.1: Dados usados na Aplicação.

n	PESOM	ALTURAM	IDADEM	RENDA	EA	TA	QGS	EPE	PESO	IG
1	64,3	1,558	34,11	1,39	32	28	3	13	1772,38	38,62
2	124,9	1,564	27,04	11,02	40	41	2	31	2104,66	37,64
3	76,3	1,609	31,04	1,71	47	61	7	28	1744,21	37,48
4	57,7	1,611	22,11	5,51	30	51	3	31	1468,35	37,92
5	68,3	1,63	20	3,67	35	29	2	17	1793,47	39,48
6	56,5	1,675	18	0,55	35	43	2	26	1574,18	39,17
7	80,9	1,586	35,02	2,45	40	44	1	22	1797,3	38,05
8	71,2	1,58	33	1,83	30	29	1	12	1777,46	38,93
9	93,2	1,661	27,06	2,45	40	33	0	26	1938,45	39,18
10	58,3	1,549	29,04	2,02	30	27	0	13	1686,58	38,72
11	68	1,57	25,1	0,73	33	31	0	27	1703,41	38,27
12	62,2	1,52	27,1	4,35	50	35	3	24	1839,51	37,64
13	58,7	1,572	29	1,17	62	60	11	37	1780,59	36,67
14	76	1,651	31,07	3,07	47	54	1	25	1744,97	38,38

Tabela B.1: Dados usados na Aplicação.

n	PESOM	ALTURAM	IDADEM	RENDA	EA	TA	QGS	EPE	PESO	IG
15	67	1,55	22,11	0,84	32	25	0	14	1776,23	38,88
16	73	1,549	19,08	2,75	30	29	1	14	1778,37	38,79
17	67	1,55	33,06	2,3	37	29	0	18	1784,75	38,36
18	71,2	1,61	20,07	1,35	25	26	0	12	1736,74	39,62
19	55,7	1,525	20,07	0,98	40	38	1	21	1662,7	37,92
20	71,5	1,627	17	1,98	44	40	2	22	1803,47	38,94
21	79,4	1,76	24	3,57	45	55	6	26	1785,77	39,52
22	96,8	1,613	20,1	2,82	46	53	11	32	1959,84	37,72
23	72,3	1,633	27,09	2,51	39	36	1	23	1779,93	38,91
24	89,1	1,63	30,06	1,02	31	29	2	14	1915,11	39,42
25	61,9	1,715	18,05	0,98	37	27	2	25	1753,54	40,12
26	77,9	1,63	23,06	2,11	38	57	2	29	1641,62	38,03
27	70	1,601	21,07	1,15	43	28	4	29	1868,22	38,6
28	65,9	1,586	34,09	3,07	46	46	4	19	1775,37	38,09
29	77,3	1,652	18,02	0,94	36	34	2	18	1825,41	39,55
30	88,3	1,662	38,11	2,45	37	31	1	12	1952,95	39,65
31	96	1,69	21,05	2,89	35	29	0	14	1979,76	40,27
32	68,7	1,67	20,09	3,26	44	48	1	31	1683,71	38,71
33	57	1,534	25,1	1,02	36	46	3	21	1595,33	37,64
34	71,9	1,596	19,08	2,94	32	42	2	27	1657,66	38,29
35	72	1,584	17,07	3,46	39	39	0	16	1767,24	38,82

Tabela B.1: Dados usados na Aplicação.

n	PESOM	ALTURAM	IDADEM	RENDA	EA	TA	QGS	EPE	PESO	IG
36	54,2	1,476	27,01	0,37	33	57	4	14	1500,55	36,95
37	52	1,525	36,05	1,15	44	38	5	36	1667,28	36,86
38	53,8	1,549	35,02	1,84	40	51	5	27	1576,59	37,15
39	75,6	1,595	28,02	2,2	34	32	6	26	1823,37	38,38
40	53	1,5	19,01	0,6	28	43	0	20	1483,24	37,59
41	70,2	1,628	20,02	0,44	31	40	3	10	1721,44	39,42
42	69,2	1,585	20,11	1,2	41	55	7	31	1665,85	37,48
43	63,8	1,6	19,08	1,5	49	35	0	14	1844,14	39,16
44	54,2	1,541	20,04	1,7	51	55	6	29	1654,65	37,12
45	69,6	1,645	20	1,35	56	51	12	31	1893,78	38,15
46	73,2	1,593	21,05	0,66	35	51	2	23	1644,35	38,13
47	68	1,517	28,04	3,33	32	41	3	19	1678,96	37,64
48	58,6	1,54	22,09	2,08	24	48	3	21	1474,05	37,71
49	55,4	1,477	17,05	0,69	24	34	1	22	1529,92	37,55
50	52,2	1,476	18,04	0,83	38	35	0	23	1623,05	37,46

# Referências Bibliográficas

- [1] Anderson, T. W. (1984). *An Introduction to Multivariate Statistical Analysis*. New York: Willey.
- [2] Andrews, D. F. and Pregbon, D. (1978). “Finding the Outliers that Matter” *Journal of the Royal Statistical Society, Series B*, 40, 85-93.
- [3] Artes, R. e Lobert, H. J. (2000). *O Impacto do Estresse Materno no Baixo Peso ao Nascimento*, São Paulo, IME-USP. (RAE-CEA-2000-23).
- [4] Barret, E. B. and Ling, F. R. (1992). “General Classes of Influence Measures for Multivariate Regression.” *Journal of the American Statistical Association*, 87, 184-191.
- [5] Bedrick, E. J. and Tsai, C. (1994). “Model Selection for Multivariate Regression in Small Samples.” *Biometrics*, 50, 226-231.
- [6] Belsley, D. A., Kuh, E. and Welsch, R. E. (1980). *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*. New York: Willey.
- [7] Bozdogan, H. (1990). “On the information-based measure of covariance complexity and its application to the evaluation of multivariate linear models.” *Communications in Statistics - Theory and Methods*, 19, 221-278.
- [8] Cook, R. D. (1977). “Detection of Influential Observation in Linear Regression.” *Technometrics*, 19, 15-18.

- [9] Cook, R. D. (1979). "Influential Observation in Linear Regression." *Journal of the American Statistical Association*, 74, 169-174.
- [10] Cook, R. D. (1986). "Assessment of Local Influence." *Journal of the Royal Statistical Society, Series B*, 48, 133-169.
- [11] Cook, R. D. (1987). "Influence Assessment." *Journal of Applied Statistics*, 14, 117-131.
- [12] Cook, R. D. and Weisberg, S. (1982). *Residuals and Influence in Regression*. London: Chapman and Hall.
- [13] Drapper, N. W. and Smith, H. (1966). *Applied Regression Analysis*. New York: Wiley.
- [14] Gabriel, K. R. (1968). "Simultaneous Test Procedures in Multivariate Analysis of Variance." *Biometrics*, 55, 489-504.
- [15] Gabriel, K. R. (1969). "Simultaneous Test Procedures - Some Theory of Multiple Comparisons." *Ann. Math. Statist.*, 40, 224-250.
- [16] Graybill, F. A. (1976). *Theory and Application of Linear Model*. North Scituate Massachusetts: Duxbury Press.
- [17] Hannan, E. and Quinn, B. G. (1979). "The determination of the order of autoregression" *Journal of the Royal Statistical Society, Series B*, 41, 190-195.
- [18] Houssain, A. and Naik, D. I. (1989). "Detection of Influential Observations in Multivariate Regression." *Journal of Applied Statistics*, 16, 25-37.
- [19] Hurvich, C. M. and Tsai, C. L. (1989). "Regression and Times Series Model Selection in Small Samples." *Biometrika*, 76, 297-307.
- [20] Ibraim, J. G. and Chen, M. H. (1997). "Predictive Variable Selection for the Multivariate Linear Model." *Biometrics*, 53, 465-478.

- [21] Johnson, R. A. and Wichern, D. W. (1998). *Applied Multivariate Statistical Analysis*, fifth edition. New York: Prentice Hall.
- [22] Linhart, H. and Zucchini, W. (1986). *Model Selection*. New York: Willey.
- [23] McKay, R. J. (1977). "Variable Selection in Multivariate Regression: An Application of Simultaneous Test Procedures" *Journal of the Royal Statistical Society, Series B*, 39, 381-386.
- [24] Morrison, D. F. (1976). *Multivariate Statistical Methods*, second edition. New York: McGraw-Hill.
- [25] Rao, C. R. (1973). *Linear Statistical Inference and its Applications*. New York: John Wiley.
- [26] SAS Institute Inc., SAS OnlineDoc®, Version 8, Cary, NC: SAS Institute Inc., 1999.
- [27] Schwarz, G. (1978). "Estimating the Dimension of a Model." *Annals of Statistics*, 6, 461-464.
- [28] Sparks, R. S., Coutsourides, D. and Troskie, L. (1983). "The multivariate  $C_p$ ." *Communications in Statistics - Theory and Methods*, 12, 1775-1793.
- [29] Waugh, F. V. (1942). "Regression between sets of variates." *Econometrica*, 10, 290-310.
- [30] Zellner, A. (1962). "An Efficient Method of Estimation Seemingly Unrelated Regressions and Test for Aggregation Bias." *Journal of the American Statistical Association*, 57, 348-368.