

ESTIMAÇÃO DE β NO MODELO DE
REGRESSÃO LINEAR MÚLTIPLA
NA PRESENÇA DE MULTICOLINEARIDADE

Maria da Conceição Farias Freitas

Dissertação apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo

para a obtenção do grau de mestre
em
Estatística

Área de concentração : Análise de Regressão

ORIENTADOR : *Prof Dr. CLOVIS DE ARAUJO PERES*

Este trabalho foi parcialmente financiado pelo CNPq .

São Paulo, setembro de 1988.

ERRATAS

NOTAÇÕES.

Onde se lê “ $\omega_i, i=1, \dots, p$ ”, leia-se “ $\omega_i, i=0, \dots, p.$ ”

ÍNDICE .

Onde se lê “1.4.4-Indicador relcionado...” , leia-se “1.4.4-Indicador relacionado...”
Onde se lê “ 3.1 - ... Utlizando-se ... ” leia-se “ ... Utilizando-se .. ”.

CAPÍTULO - 1 -

Pag. - 3 - :

- onde se lê “ variáveis ”, leia-se “ variáveis ” .
- onde se lê “ funções ”, leia-se “ funções ”.
- onde se lê “ características ”, leia-se “ características ” .

Pag. - 4 - :

- onde se lê “ economicas ”, leia-se “ econômicas ”.

Pag. - 5 - :

- onde se lê “ pesença ”, leia-se “ presença ” .
- onde se lê :

$$X_{ij} \rightarrow \left\{ \frac{(n-1)}{n} \right\}^{\frac{1}{2}}, j = 1e2.$$

leia-se :

$$X_{1j} \rightarrow \left\{ \frac{(n-1)}{n} \right\}^{\frac{1}{2}}, j = 1e2.$$

expressão que se encontra em Gunst e Mason - 1985 - [3] .

Pag. - 6 - :

- onde se lê “ necessariamente ”, leia-se “ necessariamente ” .

Pag. - 10 - :

- onde se lê :

$$\sum_{i=1}^n X_{ij} \phi(Y - i - X_{i1} \beta_1 - X_{i2} \beta_2) = 0$$

leia-se :

$$\sum_{i=1}^n X_{ij} \phi(Y_i - X_{i1} \beta_1 - X_{i2} \beta_2) = 0$$

Pag. - 13 - :

- onde se lê “ discriminados ” , leia-se “ discriminados ” .
- onde se lê “ nescessário ” , leia-se “ necessário ” .

Pag. - 14 - :

- onde se lê “ XtX ” , leia-se “ X^tX ” .
- onde se lê “ portano ” , leia-se “ portanto ” .

CAPÍTULO - 2 -

Pag. - 18 - :

- Deve-se eliminar a frase : “ pode-se ilustrar facilmente o caso em que $p=2$ ” .

Pag. - 20 - :

- onde se lê “ $(vies)^2(\hat{\beta}_{CP})$ ” , leia - se “ $(vies)^2(\hat{\beta}_{CP})$ ” .
- onde se lê “ monotonica ” , leia-se “ monotônica ” .

Pag. - 22 - :

- onde se lê “ empiricos ” , leia - se “ empíricos ” .

Pag. - 26 - :

- onde se lê “ máximos ” , leia-se “ máximo ” .

Pag. - 27 - :

- onde se lê “ pré fix-ada ” , leia-se “ pre fixa-da ” .

Pag. - 30 - :

- onde se lê “ Gráficos - 4 ” , leia-se “ Gráfico - 4 ” . par

Pag. - 31 - :

- onde se lê “ $Y = SQE$ ” , leia-se “ $Y = \frac{SQE}{2}$ ” .
- onde se lê “ $b_0 = 0$ ” , leia-se $b_0 = \frac{Y^tY}{2}$ ” .

Desta forma obtem-se :

$$Y = b_0 + b^t x + \left(\frac{1}{2}\right) x^t B x = \frac{SQE}{2}$$

conforme foi proposto no início do texto. Da forma em que esta originalmente escrito o que coincide são as equações :

$$\frac{\partial Y}{\partial \beta} = 0$$

e

$$\frac{\partial SQE}{\partial \beta} = 0$$

Pag. - 34 - :

- onde se lê “ diiferença ” , leia-se “ diferença ” .
- onde se lê “ $(X^T X)^{-1}$ ” , leia-se “ $(X^t X)^{-1}$ ” .

Pag. - 35 - :

- onde se lê “ $tr(\sum \hat{\beta}_k)$ ” , leia-se “ $tr(\sum \hat{\beta}_k)$ ” .
- onde se lê “ gráfico 5 ” , leia-se “ gráfico 4 ” .
- onde se lê “ $\hat{\beta}$ pode ser escrito como uma transformação linear de $\hat{\beta}_{MQ}$... ” , leia-se “ $\hat{\beta}_k$ pode ser escrito ... ” .

Pag. - 40 - :

- onde se lê “ $\hat{\beta}_{LR}$ ” , leia-se “ $\hat{\beta}_{RL}$ ” .

Pag. - 42 - :

- onde se lê :

$$f(a) = \eta^2 \sum_{j=0}^p a_j^2 \omega_j - 2\mu_0 \left(\sum_{j=0}^p a_j \Upsilon_{0j} - 1 \right)$$

leia-se :

$$f(a) = \eta^2 \sum_{j=0}^p a_j^2 \omega_j - 2\mu_0 \left(\sum_{j=0}^p a_j \Upsilon_{0j} - 1 \right)$$

Pag. - 44 - :

- onde se lê “ $EQM(\hat{\beta}_{RL})$ ” , leia-se “ $EQM(\hat{\beta}_{RL})$ ” .
- onde se lê “ $\hat{\beta}_{LR}$ ” , leia-se “ $\hat{\beta}_{RL}$ ” .

Pag. - 50 - :

- no gráfico , onde se lê “ $\hat{\beta}_\xi$ ” , leia-se “ $\hat{\beta}_\delta$, ou $\hat{\beta}_\rho$ ” .

Pag. - 51 - :

- onde se lê “ \hat{beta}_{MQ} ” , leia-se “ $\hat{\beta}_{MQ}$ ” .

CAPÍTULO - 3 -

Pag. - 60 - :

- onde se lê “monotonica” , leia-se “monotônica” .

Pag. - 61 - :

- onde se lê “ $EQM(\hat{\beta}_k) < \hat{\beta}_{MQ}$ ” , leia-se “ $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{MQ})$ ” .

Pag. - 63 - :

- onde se lê :

$$a_j = \Upsilon_{0j} \omega^{-1} \left(\sum_{j=0}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1}$$

leia-se :

$$a_j = \Upsilon_{0j} \omega_j^{-1} \left(\sum_{j=0}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1}$$

Pag. - 72 - :

- onde se lê “Mayer e Willke - 1973 - [36]” , leia-se “Mayer e Willke - 1973 - [15]” .

Pag. - 79 - :

- onde se lê “ $EQMG\hat{\beta}_\rho$ ” , leia-se “ $EQMG(\hat{\beta}_\rho)$ ” .

Pag. - 80 - :

- onde se lê “ $EQMG\hat{\beta}_k$ ” , leia-se “ $EQMG(\hat{\beta}_k)$ ” .

Pag. - 82 - :

- onde se lê :

“ Tem-se 15 comparações possíveis entre os 6 estimadores estudados :

$\hat{\beta}_{MQ}$, $\hat{\beta}_{CP}$, $\hat{\beta}_{RL}$, $\hat{\beta}_{IT}$ e $\hat{\beta}_\rho$ ”

leia-se :

“ Tem -se ... estudados :

$\hat{\beta}_{MQ}$, $\hat{\beta}_{CP}$, $\hat{\beta}_{RL}$, $\hat{\beta}_{IT}$, $\hat{\beta}_\rho$ e $\hat{\beta}_k$ ”.

CAPÍTULO - 4 -

Pag. - 92 - :

- onde se lê “ $var(\beta_i)$ ” , leia-se “ $var(\hat{\beta}_i)$ ” .

Pag. - 94 - :

- onde se lê “ $var(\hat{\beta})$ ” , leia-se “ $\hat{var}(\hat{\beta})$ ” .

- onde se lê “ $var(\beta)$ ” , leia-se “ $var(\hat{\beta})$ ” .

Pag. - 95 - :

- onde se lê “ $\hat{\beta}^t \beta$ ” , leia-se “ $\beta^t \beta$ ” .

Pag. - 104 - :

- onde se lê “ Tabela - 13 - Resultadso ... ” , leia-se “ Tabela - 13 - Resultados ... ” .

APÊNDICE - 2 -

Pag. - 109 - :

- onde se lê :

$$k_1 = \phi^2 + \sum_{i=2}^n X_{i1}^2 \dots e \dots k_2 + c^2 \phi^2 + \sum_{i=2}^n X_{i2}^2$$

leia-se :

$$k_1 = \phi^2 + \sum_{i=2}^n X_{i1}^2 \dots e \dots k_2 = c^2 \phi^2 + \sum_{i=2}^n X_{i2}^2$$

APÊNDICE - 4 -

Pag. - 112 - :

- onde se lê “ defir ” , leia-se “ definir ” .

APÊNDICE - 6 -

Pag. - 114 - :

- onde se lê “ Dados utilizados para a análise do capítulo 4 .” , leia-se “ Tabela - 1 -
Dados utilizados para a análise do capítulo 4 .”æ

AGRADECIMENTOS

Este trabalho contou com a DECISIVA colaboração de :

FUNAD - Fundo de Pesquisa do Instituto de Administração ,
ALOISIO PINTO ALVES .

UPD - Unidade de Processamento de Dados - FEA - USP ,
LUIZ ÁLVARO LEÃO GIL
E demais funcionarios.

SIANG WUN SONG

WILTON DE OLIVEIRA BUSSAB

CARLOS EDGARD HARLE

E em especial , desejo agradecer :

ao incentivo constante de MINHA FAMÍLIA :
OS PAIS : LUIZ ALVES DE FREITAS e JOANA ANITA DE FARIAS FREITAS.
OS IRMÃOS : JOSÉ , JOÃO E LUIZ.
E R. YURI TANDEL.

a paciente ORIENTAÇÃO de *CLOVIS DE ARAUJO PERES* , cuja dedicação e
experiencia contribuíram para enfrentar as fases mais difíceis.

MUITO OBRIGADO.

setembro - 1988 -

“ Tudo quanto te vier a mão para fazer , faze-o conforme as tuas forças , porque na sepultura , para onde tu vais , não há obra , nem indústria , nem ciência , nem sabedoria alguma . ”

“Confia no Senhor de todo o teu coração e não te estribes no teu próprio entendimento.”

Pois

“ O temor do Senhor é o principio da sabedoria. ”

Eclesiastes 9:10
Provérbios 3:5
Provérbios 9:10

ESTIMAÇÃO DE β NO MODELO DE REGRESSÃO LINEAR MÚLTIPLA

NA PRESENÇA DE MULTICOLINEARIDADE

RESUMO

O objetivo principal deste trabalho é comparar métodos alternativos ao de mínimos quadrados, para se estimar β , no modelo de regressão linear múltipla na presença de multicolinearidade.

- O capítulo - 1 - descreve a multicolinearidade, enfatizando sua caracterização como um problema na estimação dos parâmetros.

- O capítulo - 2 - apresenta os cinco estimadores viesados mais conhecidos na literatura, apresentando suas vantagens sobre o método de mínimos quadrados, diante do problema em questão. São eles:

- *MÉTODO DAS COMPONENTES PRINCIPAIS*, cuja primeira publicação com o objetivo de resolver o problema de multicolinearidade, é atribuído a M. G. Kendall em 1957.

- *ESTIMADORES EM CRISTAS*, conhecido como estimador RIDGE, atribuído a A. E. Hoerl e R. W. Kennard, fruto de seu trabalho conjunto em 1970.

- *MÉTODO DAS RAIZES LATENTES*, atribuído a D. Hawkins, em 1973, também desenvolvido independentemente por R. F. Gunst e R. L. Mason em 1974.

- *ESTIMADORES CONTRAÍDOS*, cuja primeira publicação atribui-se a W. James e C. Stein em 1961.

- *MÉTODO INTERATIVO*, atribuído a G. Trenkler, publicado em 1978.

- O capítulo - 3 - refere-se a comparação entre estes estimadores e o de mínimos quadrados e entre si, utilizando-se dois critérios:

- O EQM - ERRO QUADRÁTICO MÉDIO.

- O EQMG - ERRO QUADRÁTICO MÉDIO GENERALIZADO.

- O capítulo - 4 - consiste na análise de um conjunto de dados, com o intuito de ilustrar os resultados pesquisados nos capítulos anteriores. As estimativas obtidas para estes cinco métodos foram calculadas através de processamento em computador, Burroughs 6900, utilizando-se a linguagem Fortran, e algumas rotinas já prontas do IMSL.

Os originais do texto deste trabalho foram editados e impressos em microcomputador, utilizando-se a rotina T_EX: Donald E. Knuth - 1984 - [48].

NOTAÇÕES

X : Denota a matriz X padronizada no modelo $Y = X\beta + \epsilon$, de maneira a obter-se:
 $E(X_i) = 0$ e $\text{var}(X_i) = n - 1$.

Z : $= \Gamma X$ com $Z^t Z = \Delta$ onde :
 Γ é a matriz cujas colunas são os AUTOVETORES normalizados de $X^t X$.
 Δ é a matriz diagonal, cujos elementos são os AUTOVALORES de $X^t X$.

$\sum \hat{\beta}$: Denota a matriz de Variância e Covariância do estimador $\hat{\beta}$.

$\lambda_i, i=1...p$: Denota AUTOVALORES de $X^t X$ em ordem decrescente:
 $\lambda_1 > \lambda_2 > \dots > \lambda_p$.

Γ : Denota a matriz cuja coluna- i , Γ_i é o i -ésimo AUTOVETOR NORMALIZADO de $X^t X$, portanto associado a λ_i .

Δ : Denota a matriz diagonal, cujos elementos são os autovalores de $X^t X$ -
 $\lambda_i, i=1, \dots, n$.

$\omega_i, i=1...p$: Denota AUTOVALORES de $A^t A$ onde $A = (YX)$ e $\omega_0 < \omega_1 < \dots < \omega_p$

Ω : Denota a matriz diagonal, cujos elementos são os autovalores de $(A^t A)$ com $A = (YX)$.

Υ : Matriz cujas colunas são os autovetores de $A^t A$ com $A = (YX)$

EQM ($\hat{\beta}$) - ERRO QUADRÁTICO MÉDIO :

$$E(\hat{\beta} - \beta)^t(\hat{\beta} - \beta) = \sigma^2 \text{tr}(X^t X) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}$$

EQMG ($\hat{\beta}$) - ERRO QUADRÁTICO MÉDIO GENERALIZADO :

$$E(\hat{\beta} - \beta)^t T (\hat{\beta} - \beta)$$

onde T - é uma matriz simétrica NND .

$VAR(\hat{\beta}) =$ Variância ($\hat{\beta}$) .

$Corr(\hat{\beta}) =$ Matriz de Correlação de $\hat{\beta}$.

Tr : Abreviação para traço de uma matriz.

A^t : Denota a Matriz Transposta de A .

ÍNDICE

CAPÍTULO - 1 - A MULTICOLINEARIDADE - DEFINIÇÃO , FONTES , EFEITOS E DETECÇÃO .

1.1 - O CONCEITO DE MULTICOLINEARIDADE .	1
1.2 - FONTES DE MULTICOLINEARIDADE .	2
1.2.1 - Multicolinearidade induzida pela especificação do modelo.	2
1.2.2 - Multicolinearidade induzida por propriedades inerentes a população.	3
1.2.3 - Multicolinearidade induzida por deficiências na amostragem .	3
1.2.4 - Multicolinearidade induzida por modelos mal definidos .	4
1.2.5 - Multicolinearidade induzida pela presença de outliers.	4
1.3 - EFEITOS DA PRESENÇA DE MULTICOLINEARIDADE .	6
1.3.1 - Efeito da multicolinearidade sobre a estimação de β .	8
1.4 - DETECÇÃO E MEDIDAS DO GRAU DE MULTICOLINEARIDADE .	11
1.4.1 - A matriz $X^t X$.	11
1.4.2 - V.I.F. - Fator de Inflação da Variância .	11
1.4.3 - Razão entre os comprimentos de intervalos de confiança de β_j .	13
1.4.4 - Indicador relacionado ao tamanho da amostra .	13
1.4.5 - Autovalores e Autovetores de $X^t X$.	14
1.4.6 - O número da condição espectral .	14
1.4.7 - Valores singulares de $X^t X$ na decomposição de $var(\hat{\beta}_j)$ - a matriz π .	15
1.4.8 - O determinante de $X^t X$.	16

CAPÍTULO - 2 - CINCO MÉTODOS MAIS UTILIZADOS PARA A SOLUÇÃO DO PROBLEMA DE MULTICOLINEARIDADE.

2.1 - O MÉTODO DAS COMPONENTES PRINCIPAIS .	18
2.1.1 - Descrição formal e algumas propriedades de $\hat{\beta}_{CP}$.	19
2.1.2 - Métodos de seleção das componentes principais .	21
2.2 - REGRESSÃO EM CRISTAS .	23
2.2.1 - Desenvolvimento original do método de análise de cristas .	23
2.2.2 - Extensão da análise de cristas como método de estimação de β na regressão linear .	31
2.2.3 - Algumas propriedades do estimador $\hat{\beta}_k$.	34

2.3 - O MÉTODO DAS RAÍZES LATENTES .	36
2.3.1 - Descrição do método das raízes latentes.	37
2.3.2 - Interpretação geométrica do método das raízes latentes.	38
2.3.3 - Obtenção de β_{RL} .	40
2.3.4 - Algumas propriedades de $\hat{\beta}_{RL}$.	44
2.4 - OS ESTIMADORES CONTRAÍDOS .	46
2.4.1 - Apresentação do estimador contraído.	47
2.4.2 - Algumas propriedades estatísticas de alguns estimadores contraídos .	49
2.4.3 - Comentários e problemas não solucionados .	52
2.5 - O ESTIMADOR INTERATIVO .	53
2.5.1 - Definição e algumas propriedades estatísticas .	53
CAPÍTULO - 3 - COMPARAÇÕES ENTRE OS ESTIMADORES .	56
3.1 - COMPARAÇÕES ENTRE OS ESTIMADORES UTILIZANDO-SE O EQM -ERRO QUADRÁTICO MÉDIO .	57
3.1.1 - Componentes principais x Mínimos quadrados .	58
3.1.2 - Regressão em cristas x Mínimos quadrados .	60
3.1.3 - Raízes latentes x Mínimos quadrados .	62
3.1.4 - Estimador contraído x Mínimos quadrados .	65
3.1.5 - Estimador interativo x Mínimos quadrados .	67
3.1.6 - Componentes principais x Estimador em cristas .	68
3.2 - COMPARAÇÕES ENTRE OS ESTIMADORES UTILIZANDO-SE O EQMG - ERRO QUADRÁTICO MÉDIO GENERALIZADO .	70
3.2.1 - Componentes Principais x Mínimos quadrados .	73
3.2.2 - Regressão em cristas x Mínimos quadrados .	73
3.2.3 - Estimador contraído x Mínimos quadrados .	75
3.2.4 - Estimador interativo x Mínimos quadrados .	76
3.2.5 - Estimador contraído x Estimador em cristas .	79
3.2.6 - Estimador em cristas x Componentes principais .	80
3.3 - RESUMO GERAL DAS COMPARAÇÕES OBTIDAS E ALGUNS COMENTÁRIOS .	82

CAPÍTULO - 4 - ANÁLISE DE DADOS.	84
4.1 - INTRODUÇÃO .	84
4.2 - DETECÇÃO E MEDIDAS DO GRAU DE MULTICOLINEARIDADE .	86
4.3 - COMPARAÇÕES ENTRE AS ESTIMATIVAS OBTIDAS .	92
4.3.1 - Descrição das estimativas e comparações preliminares .	92
4.4 - COMPARAÇÕES ENTRE OS ESTIMADORES BASEANDO-SE NO EQM .	96
4.5 - COMPARAÇÕES ENTRE OS ESTIMADORES BASEANDO-SE NO EQMG .	101
5 - CONCLUSÃO .	106

APÊNDICES

APÊNDICE - 1 - A multicolinearidade é um problema?	108
APÊNDICE - 2 - Multicolinearidade induzida por outlier .	109
APÊNDICE - 3 - Decomposição espectral de uma matriz A $n \times p$.	110
APÊNDICE - 4 - Inversa generalizada de uma matriz .	111
APÊNDICE - 5 - Matrizes N.N.D. - Não Negativas Definida.	113
APÊNDICE - 6 - Dados utilizados para análise no capítulo 4 .	114
REFERÊNCIAS BIBLIOGRAFICAS .	115

CAPÍTULO - 1 -

MULTICOLINEARIDADE

DEFINIÇÃO - FONTES - EFEITOS E DETECÇÃO

1.1 - O CONCEITO DE MULTICOLINEARIDADE

Estimadores de mínimos quadrados dos parâmetros de regressão linear múltipla têm sido usados frequentemente nos últimos 50 anos com o intuito de ajustar-se funções a um conjunto de dados

O modelo de regressão linear múltipla é descrito por:

$$Y = X\beta + \epsilon$$

onde

Y: é um vetor $n \times 1$, chamado de variável dependente ou variável resposta.

X: é uma matriz $n \times p$, de n observações de p variáveis chamadas de variáveis preditoras, não estocásticas.

β : é um vetor $p \times 1$ de parâmetros, a ser estimado.

ϵ : é um vetor $n \times 1$ de erros ou resíduos com $\epsilon_i \sim N(0, \sigma^2)$.

No estudo da estimação do parâmetro β , merece relevância o comportamento das variáveis preditoras. Quando algumas ou todas as variáveis são perfeitamente colineares, diz-se estar diante da presença de *multicolinearidade perfeita*, situação em que não se pode calcular os estimadores de mínimos quadrados dos parâmetros porque X^tX é singular. Como na prática isto não ocorre, o termo multicolinearidade também é utilizado quando as variáveis são correlacionadas ou seja X^tX é *quase* singular, nesta situação pode-se calcular o estimador de mínimos quadrados de β , contudo algumas propriedades estatísticas desejáveis, são prejudicadas.

R. Gunst-1983-[2], define multicolinearidade utilizando o conceito de quase dependência entre as variáveis preditoras, ou a quase singularidade de X^tX , da seguinte forma:

Diz-se existir *multicolinearidade*, se para algum $\eta \geq 0$ existir um vetor c , $c = (c_0, c_1, \dots, c_p)$, de elementos não todos nulos, tal que:

$$\sum_{j=0}^p c_j x_j = \delta \text{ com } \|\delta\| \leq \eta \|c\|$$

Os itens seguintes apresentam as fontes, os efeitos e as maneiras de detectar-se a multicolinearidade.

1.2 - FONTES DA MULTICOLINEARIDADE

É importante compreender as diferenças entre as várias causas da presença da multicolinearidade, uma vez que a interpretação do modelo resultante está extremamente ligada a elas.

A multicolinearidade pode surgir por várias razões:

- A especificação do modelo.
- Propriedades inerentes a população.
- Deficiências no método de amostragem.
- Modelos mal definidos, ou seja, o número de observações é menor que o número de variáveis.
- Devido a presença de outliers.

Para cada uma destas causas R. Gunst - 1983 - [2], apresenta exemplos, alguns deles são discutidos a seguir.

1.2.1 - MULTICOLINEARIDADE INDUZIDA PELA ESPECIFICAÇÃO DO MODELO.

Em alguns casos a multicolinearidade é introduzida no modelo através da definição das variáveis preditoras. Um exemplo clássico é dado pelo modelo definido de maneira a obedecer a seguinte restrição:

$$\sum_{j=1}^p x_{ij} = 1, i = 1 \dots n.$$

Neste caso, teremos multicolinearidade perfeita nos modelos com o termo constante.

Um outro exemplo é obtido através do modelo onde a variável resposta pode ser representada corretamente por uma função polinomial de uma única variável preditora X :

$$Y = \sum_{i=1}^p X^{k_i} + \epsilon$$

A correlação entre duas potências de X , X^{k_1} e X^{k_2} por exemplo, pode se tornar muito próximo de "1". (Bradley e Srivastava - 1979 - [49]).

Exemplos mais comuns são os casos em que muitas variáveis preditoras são colocadas no modelo, devido a incerteza do pesquisador com relação aquelas que realmente influenciam a variável resposta.

Alguns autores defendem a idéia de que a reespecificação do modelo tende a suavizar a presença de multicolinearidade, em contrapartida, outros afirmam que ao reespecificar-se o

modelo, através de transformações, os efeitos da multicolinearidade podem ser transferidos das variáveis originais para as variáveis transformadas. Por exemplo:

- Suponha que x_j e x_k são duas funções de tempo, monotônicas e colineares. Suponha ainda que a transformação $w_i = x_j/x_k$, suavize a presença de multicolinearidade. No caso em que este quociente for quase constante ao longo do tempo, aparecerá uma relação quase linear entre w_i e o termo constante α , do modelo descrito por $y = \alpha + \beta x$.

1.2.2 - MULTICOLINEARIDADE INDUZIDA POR PROPRIEDADES INERENTES A POPULAÇÃO.

Neste caso a multicolinearidade sempre ocorrerá independentemente do modelo especificado ou do processo de amostragem utilizado.

Exemplo:

Foram coletados dados de 33 mulheres negras candidatas ao departamento de polícia, com o intuito de construir uma equação para explicar altura em função de outras nove variáveis, entre as quais:

- X: o comprimento da parte superior do braço.
- Y: o comprimento do antebraço.
- W: o comprimento da parte superior da perna.
- K: o comprimento da parte inferior da perna.
- A: é o quociente Y/X
- B: é o quociente K/W .

Devido a características da população estudada, o campo de variação das quatro primeiras variáveis é muito pequeno, tornando-as muito correlacionadas com os seus logarítimos.

Neste caso a multicolinearidade pode ser representada pelas equações :

$$\ln(Y) - \ln(X) = \ln(A)$$

$$\ln(K) - \ln(W) = \ln(B)$$

1.2.3 - MULTICOLINEARIDADE INDUZIDA POR DEFICIÊNCIAS NO PROCESSO DE AMOSTRAGEM.

A multicolinearidade que ocorre por causa de deficiências no plano amostral é artificial, devido ao fato de os dados coletados serem específicos daquele subconjunto do espaço amostral com aquelas características.

Um exemplo interessante é apresentado por R. Gunst - 1983 - [2], extraído de dados coletados na França durante o período de 1949 a 1959, referentes a explicação da variável importação através das variáveis produção, estoques e consumo domésticos. Verificou-se uma forte dependência entre a produção e o consumo. Como em 1960 a França entrou para o Mercado Comum Europeu, atribuiu-se a este fato as peculiaridades econômicas da época.

1.2.4 - MULTICOLINEARIDADE INDUZIDA POR MODELOS MAL DEFINIDOS.

É o caso em que há mais parâmetros a serem estimados que observações, pois nesta situação $X^t X$ não é de posto completo, isto é fácil ver. Considere:

n : o número de observações.

p : o número de parâmetros a serem estimados.

Sabe-se que :

$\text{posto}(X) = r \leq \min(n, p)$ com $(r \leq n)$ e que

$\text{posto}(X^t X) = r$

A multicolinearidade desaparece ao se coletar mais dados.

1.2.5 - MULTICOLINEARIDADE INDUZIDA PELA PRESENÇA DE OUTLIERS.

Este tipo de multicolinearidade só ocorre quando dois ou mais valores das variáveis preditoras assumem valores muito grandes em relação aos demais.

Como no caso de deficiências na amostragem, a multicolinearidade induzida por outliers é artificial, ou seja, eliminando-se os outliers, solucionam-se os possíveis problemas.

Diagramas de dispersão dos pontos são indicados e quanto mais próximos os outliers uns dos outros, mais provável é que a multicolinearidade seja induzida por estas observações.

Considere o exemplo em que se tem duas variáveis preditoras, seja:

x_i^* , a i -ésima linha de X , considerando-se X sem o termo constante.

Considere ainda :

$$x_i^* = 1(x_{i1}, x_{i2}), i = 2 \dots n, x_{i1}, x_{i2} > 0.$$

$$x_1^* = k(x_{i1}, x_{i2}), \text{ para algum } (i).$$

Quando $k \rightarrow \infty$, a primeira observação, x_1^* , torna-se arbitrariamente grande em relação as demais.

Considere ainda a matriz $H = X(X'X)^{-1}X'$. Os elementos de sua diagonal h_i , são indicadores da presença de outliers, quando estão próximos de um.

Mostra-se que no exemplo acima, $h_1 \rightarrow 1$ e ainda que:

$$X_{ij} \rightarrow \left\{ \frac{(n-1)}{n} \right\}^{\frac{1}{2}}, j = 1 \text{ e } 2.$$

$$X_{ij} \rightarrow \{n(n-1)\}^{\frac{-1}{2}}, i \neq 1 \text{ e } j = 1, 2.$$

Conclue-se que para k suficientemente grande, $X_1 - X_2$ aproxima-se arbitrariamente de zero o que consiste exatamente na descrição de multicolinearidade apresentada no início deste capítulo.

1.3 - EFEITOS DA PRESENÇA DE MULTICOLINEARIDADE

Sérios efeitos sobre os estimadores de mínimos quadrados são encontrados. A relação abaixo ajuda-nos a avaliar a necessidade de um estudo detalhado sobre a multicolinearidade:

- Alterações nas estimativas dos parâmetros. E conseqüentemente nas previsões.
- Aumento das estimativas das variancias e covariancias dos estimadores, tendo como consequencia o aumento na amplitude dos intervalos de confiança.
- O poder de alguns testes são afetados assim como o parâmetro de não centralidade de t-Student para o teste: $H_0 : \beta_j = 0$ contra $H_a : \beta_j \neq 0$.

A multicolinearidade é descrita através de combinações lineares das variáveis preditoras. Vamos denotar a j-ésima coluna da matriz X por X_j . Diremos que existe *multicolinearidade perfeita* no modelo, no caso em que os regressores são padronizados, se existir $a_i, i=1...p$, não todos nulos tal que:

$$\sum_{j=1}^p a_j X_j = 0$$

Quando os regressores não estão padronizados ou simplesmente centrados a relação acima fica substituída por:

$$\sum_{j=1}^p a_j X_j = m \tag{1}$$

onde "m" é um vetor de constantes não necessariamente iguais a zero, (Montgomery e Peck - 1982 -[1]).

No caso em que a expressão (1), acima, é exata o posto de $(X^t X) < p$ implicando na não existencia de $(X^t X)^{-1}$.

Se (1) é "quase verdade" para algum subconjunto de colunas de X, existirá uma "quase dependencia" linear em X. Como vimos também neste caso diz-se existir a multicolinearidade. Podemos então dizer que *todo conjunto de dados possui multicolinearidade com uma diferença de intensidade.*

Para melhor visualizarmos os efeitos da presença de multicolinearidade convém escrevermos X , utilizando sua decomposição espectral, ou decomposição em valores singulares, [apêndice 3]:

$$X = H\Delta^{1/2}\Gamma, \text{ onde:}$$

H : é uma matriz $n \times p$, cujas colunas são os *autovetores* normalizados de XX^t , portanto $H^tH = I$.

Γ : é uma matriz $p \times p$, cujas colunas são os *autovalores* normalizados de X^tX , portanto $\Gamma^t\Gamma = I$.

$\Delta^{1/2}$: é uma matriz diagonal $p \times p$, cujos elementos são as raízes quadradas dos autovalores de X^tX , $\sqrt{\lambda_i}$, com $\lambda_1 > \lambda_2 > \dots > \lambda_p$.

Utilizando a expressão acima obtém-se:

$$(X^tX)^{-1} = \sum_{j=1}^p \frac{1}{\lambda_j} \Gamma_j \Gamma_j^t \quad (2)$$

$$\hat{\beta}_{MQ} = (X^tX)^{-1} X^tY = \sum_{i=1}^p \frac{1}{\lambda_i} \Gamma_i \Gamma_i^t X^tY = \sum_{i=1}^p \frac{\Gamma_i}{\lambda_i} c_i \quad (3)$$

com $c_i = \Gamma_i^t X^tY$, um escalar.

$$\sum_{\hat{\beta}_{MQ}} = \sigma^2 (X^tX)^{-1} = \text{diag } \sigma^2 \left\{ \sum_{i=1}^p \frac{1}{\lambda_i} \Gamma_i \Gamma_i^t \right\} \quad (4)$$

Gunst e Mason - 1980 - [8].

1.3.1 - EFEITO DA MULTICOLINEARIDADE SOBRE A ESTIMAÇÃO DE β .

SOBRE OS ELEMENTOS DE $\hat{\beta}_{MQ}$.

É interessante notar que se X^tX for quase singular, um ou mais de seus autovalores serão próximos de zero. Conseqüentemente isto trará uma instabilidade na expressão (3), uma vez que os autovalores fazem parte do denominador. Esta instabilidade é refletida por uma superestimativa dos parâmetros e por eventuais inversões de sinal nos elementos de $\hat{\beta}_{MQ}$.

SOBRE A MATRIZ $\sum_{\beta(MQ)}$.

A variancia e a covariancia dos estimadores de β também ficam inflacionadas. Isto é fácil ver através do denominador da expressão (4) , o mesmo ocorre com o erro quadrático médio : - $EQM(\hat{\beta}_{MQ})$, dado por :

$$EQM(\hat{\beta}_{MQ}) = E(\hat{\beta}_{MQ} - \beta)^t(\hat{\beta}_{MQ} - \beta) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}$$

SOBRE FUNÇÕES LINEARES DOS PARÂMETROS .

Nas situações em que o interesse é estimar funções paramétricas do tipo $a_t\beta$, onde a_t é um vetor de constantes, mostra-se que na presença de multicolinearidade, o estimador de mínimos quadrados de $a_t\beta$, terá mínima variância quando $a_t = \Gamma_1^t$, onde Γ_1^t , é o autovetor de X^tX associado ao maior autovalor. (Gunst e Mason - 1980 - [8])

SOBRE ALGUMAS PROPRIEDADES ESTATÍSTICAS.

Devido ao fato de as variancias e as covariancias serem afetadas pela multicolinearidade, os intervalos de confiança assim como os testes de hipóteses sobre os parâmetros também o são.

A fim de verificarmos isto, vamos expressar de forma conveniente o comprimento do intervalo de confiança para β :

$$\frac{2t_{n-p-1}\hat{\sigma}}{\sqrt{1-R_j^2}}$$

R_j : é o coeficiente de determinação resultante da regressão múltipla de X_j sobre os demais $(p-1) X_k$.

$\hat{\sigma}^2$: refere-se ao estimador de mínimos quadrados de σ^2 .

Quanto mais forte for a multicolinearidade , mais próximo de 1 é R_j^2 e maior é o intervalo de confiança.

Com relação ao teste de hipótese, verificaremos que o poder do teste fica reduzido. Isto ocorre porque o parâmetro de não centralidade μ , da estatística utilizada, diminui na presença de multicolinearidade uma vez que μ , é dado por

$$\mu = \frac{\beta_j \sqrt{1-R_j^2}}{2\sigma}$$

Quanto maior R_j menor é o valor de μ

SOBRE AS PREVISÕES.

Os valores estimados de Y, ficam também bastante prejudicados sobre a presença de multicolinearidade, com exceção dos casos em que os NOVOS valores de X escolhidos seguem a mesma estrutura de multicolinearidade que os valores de X utilizados no ajuste do modelo.

Para verificarmos esta afirmação denotaremos:

X_i^* : Os novos valores de X .

\hat{Y}^* : O valor da previsão.

Obtendo-se :

$$\hat{Y}^* = X_i^{*t} \beta$$

$$\sum \hat{Y}^* = \sigma^2 \left\{ \frac{1}{n} + \sum_{j=1}^p \lambda_j^{-1} (X_i^{*t} \Gamma_j)^2 \right\}$$

onde $\sum \hat{Y}^*$ denota a matriz de var e covar de \hat{Y}^* .

A expressão acima mostra que se tem um aumento significativo em $\sum \hat{Y}^*$, devido aos valores pequenos de λ_i no denominador.

Este aumento só desaparece, se $(X^{*t} \Gamma_j)$ for próximo de zero .Ou seja se X^* seguir a estrutura de multicolinearidade de X.Pois $X^t \Gamma_j = \lambda_j \simeq 0$,considerando-se $\lambda_j \simeq 0$

Outros efeitos , típicos de situações mais específicas, são apresentados na literatura. Alguns estimadores robustos , tipo M, por exemplo, são afetados pela multicolinearidade induzida pela presença de outliers.Para o caso em que $p=2$, num modelo sem intercepto, este estimador é dado pela solução de um sistema do tipo:

$$\sum_{i=1}^n X_{ij} \phi(Y - i - X_{i1} \beta_1 - X_{i2} \beta_2) = 0$$

para $j=1,2$ e $\phi(\cdot)$ uma função convexa, simétrica e limitada com $\phi(0) = 0$. Considerando-se os outliers expressos como $X_{11} = \Phi$ e $X_{12} = k\Phi$, mostra-se que a solução é tal que:

$$\hat{\beta}_1 + k\hat{\beta}_2 = Y_i / \Phi \rightarrow 0$$

Ou seja , independentemente dos valores reais dos parâmetros o limite acima sempre é zero .(Gunst - 1983 - [2]) .

Pelo que se discutiu até o presente momento, torna-se evidente que o problema da presença de multicolinearidade é *um problema de GRAU* e não necessariamente uma questão de presença ou ausência.Sendo assim algumas medidas de intensidade são fáceis de serem obtidas e estão descritas no item a seguir.

1.4 - DETECÇÃO E MEDIDAS DO GRAU DE MULTICOLINEARIDADE.

As medidas abaixo refletem o grau de multicolinearidade, portanto servem como instrumentos de detecção, algumas delas fornecendo ainda informações sobre quais os regressores envolvidos.

1.4.1 - A MATRIZ $X^t X$

Uma investigação simples e bastante esclarecedora é obtida através do exame dos elementos de $X^t X$. Como X está padronizada, $X^t X$ é a matriz de CORRELAÇÃO de X , $X^t X = Corr(X)$, um elemento desta matriz próximo a 1, indica a existência de forte relação linear entre duas variáveis preditoras.

É interessante notar que se 3 ou mais regressores estiverem relacionados linearmente, na mesma equação, não significa necessariamente que as correlações duas a duas serão próximas de 1 e portanto este tipo de multicolinearidade não é detectada pelas correlações simples.

1.4.2 - V I F - FATOR DE INFLAÇÃO DA VARIANCIA.

Vimos que a matriz de var e covar de $\hat{\beta}_{MQ}$, é dada por:

$$\sum \hat{\beta}_{MQ} = \sigma^2 (X^t X)^{-1}$$

Podemos reescrever a variancia de $\hat{\beta}_{MQ(j)}$, substituindo os elementos da diagonal de $(X^t X)^{-1}$ por $\frac{1}{1-R_j^2}$:

$$var(\hat{\beta}_{MQ(j)}) = \sigma^2 \frac{1}{1-R_j^2}$$

Lembrando que: R_j^2 : é o coeficiente de determinação obtido da regressão múltipla de X_j sobre os demais X_k e que na presença de multicolinearidade: $R_j^2 \simeq 1$ e $\frac{1}{1-R_j^2} \rightarrow \infty$ podemos dizer que o fator $\frac{1}{1-R_j^2}$, é um fator de inflação de variancia devido a multicolinearidade, denotado por:

$$VIF_j = \frac{1}{1-R_j^2}$$

que pode ser reescrito utilizando-se a expressão 4 do item 1.3 como

$$VIF_j = \sum_{i=1}^p \frac{\Gamma_{ji}^2}{\lambda_i}$$

A conclusão imediata é que quanto maior o VIF_j mais forte a multicolinearidade, causada pela presença de X_j .

1.4.3 - RAZÃO ENTRE OS COMPRIMENTOS DE INTERVALOS DE CONFIANÇA DE β_j .

Uma outra medida do grau de multicolinearidade é dada pelo quociente dos comprimentos dos intervalos de confiança discriminados abaixo:

L^* :O comprimento do intervalo de confiança de β_j , para o modelo onde X é ortogonal.Isto é , sem multicolinearidade.

L :O comprimento do intervalo de confiança de β_j , para o modelo em que há presença de multicolinearidade.

Vimos que $L = \frac{2t_{n-p-1}\hat{\sigma}}{\sqrt{1-R_j^2}}$.Quando X é ortogonal, $R_j^2 = 0$ e portanto $L^* = 2t_{n-p-1}\hat{\sigma}$, então:

$$\frac{L}{L^*} = \frac{1}{\sqrt{1-R_j^2}} = \sqrt{VIF_j}$$

Verificamos que quanto mais acentuada a multicolinearidade, maior é o VIF_j , portanto maior o intervalo de confiança para β_j .

1.4.4 - INDICADOR RELACIONADO AO TAMANHO DA AMOSTRA

Uma outra medida do efeito da multicolinearidade é o aumento necessário no tamanho da amostra para que o intervalo de confiança para β_j ,no modelo com multicolinearidade tenha o mesmo comprimento que este mesmo intervalo no modelo ortogonal, (Willan e Watts - 1978 - [50]) .

O tamanho ideal para se obter esta igualdade é n_j^* , solução de :

$$\frac{n_j^*}{f_{n_j^*-p-1}} = \frac{n(1-R_j^2)}{f_{n-p-1}}$$

onde $f_{n-p-1} = t_{n-p-1}^2$, tem distribuição f-Fisher com n-p-1 graus de liberdade.

Quanto maior é a multicolinearidade que envolve X_j , menor é a precisão na estimação de β_j e portanto , maior é o tamanho de amostra necessário para se atingir a precisão desejada.

1.4.5 - AUTOVALORES E AUTOVETORES DE X^tX .

O exame dos autovalores e autovetores de X^tX , descrevem eficazmente a estrutura de multicolinearidade no modelo de regressão linear.

Seja Γ_i , um autovetor de X^tX correspondente ao autovalor λ_i . Da relação :

$$X\Gamma_i = \lambda_i\Gamma_i$$

que é equivalente a definição de autovetor de X^tX , verifica-se que quando λ_i , é próximo de zero, os elementos de Γ_i mostram a estrutura de multicolinearidade existente em X .

1.4.6 - O NÚMERO DA CONDIÇÃO ESPECTRAL

O número da condição espectral é definido por:

$$K = \left\{ \frac{\max(\lambda_i)}{\min(\lambda_i)} \right\}^{1/2}$$

Na ausencia de multicolinearidade, isto é , X^tX ortogonal, ou $X^tX = \theta I_d$, os autovalores de X^tX são iguais entre si e portanto $k=1$. Isto acontece mesmo para o caso em que $\theta \rightarrow 0$, situação em que os autovalores são muito pequenos, dando uma falsa aparência de multicolinearidade.

Isto ilustra porque alguns autores se referem a esta medida como sendo a mais estavel medida do grau de multicolinearidade . Quanto maior é o valor de " k " , mais forte é a multicolinearidade, alguns autores consideram 100 como um valor indicador de forte multicolinearidade.

(Vinod e Ullah - 1981 - [4]) .

1.4.7 - VALORES SINGULARES DE X^tX
 NA DECOMPOSIÇÃO DE $VAR(\hat{\beta}_j)$ - A MATRIZ Π

Utilizando a decomposição em valores singulares de X^tX , verificou-se que:

$$(X^tX)^{-1} = \sum_{i=1}^p \frac{\Gamma_i \Gamma_i^t}{\lambda_i}$$

portanto :

$$\begin{aligned} var(\hat{\beta}_j) &= \sigma^2 VIF_j = \sigma^2 \left(\sum_{i=1}^p \frac{\Gamma_{ji}^2}{\lambda_i} \right) = \\ &\sigma^2 \left(\frac{\Gamma_{j1}^2}{\lambda_1} + \frac{\Gamma_{j2}^2}{\lambda_2} + \dots + \frac{\Gamma_{jp}^2}{\lambda_p} \right) \end{aligned}$$

Dividindo ambos os membros da equação acima por $\sigma^2 VIF_j$, obtem-se:

$$1 = \frac{\Gamma_{j1}^2}{\lambda_1 VIF_j} + \frac{\Gamma_{j2}^2}{\lambda_2 VIF_j} + \dots + \frac{\Gamma_{jp}^2}{\lambda_p VIF_j}$$

Portanto cada parcela desta soma reflete a contribuição percentual de cada autovalor na variancia de $\hat{\beta}_j$. Quanto menor for λ_i , em relação aos demais, maior é a sua contribuição.

Estes elementos são apresentados na literatura por meio de uma matriz, a matriz Π , cujos elementos são :

$$\pi_{ij} = \frac{\Gamma_{ji}^2}{\lambda_i VIF_j}$$

O capítulo de análise de dados explora mais detalhadamente as informações que podem ser extraídas da matriz Π .

1.4.8 - O DETERMINANTE DE $X'X$.

Quando X esta padronizada , $0 \leq | X'X | \leq 1$, e $| X'X | = 1$ se X for ortogonal e $| X'X | \simeq 0$ quando existe forte dependencia linear entre os regressores. Portanto o determinante de $X'X$ pode ser usado como indicador de multicolinearidade. Verifica-se também que a razão entre os volumes das regiões de confiança para β obtidas na presença e na ausencia de multicolinearidade é dada por :

$$\sqrt{| X'X |}$$

que se interpreta como uma medida da perda da estimativa, devido a presença de multicolinearidade.

Uma vez caracterizado o problema de multicolinearidade, analisando-se os aspectos mais fundamentais para a compreensão do modelo resultante, resta-nos ainda algumas duvidas:

Que técnica para estimar-se os parâmetros é mais resistente aos seus efeitos?

É uma das questões que o capítulo seguinte tenta solucionar.

CAPÍTULO - 2 -

CINCO MÉTODOS MAIS UTILIZADOS PARA A SOLUÇÃO DO PROBLEMA DE MULTICOLINEARIDADE

Até o presente momento estudou-se as causas , os efeitos da presença de multicolinearidade e alguns métodos de detecção, enfatizando-se os problemas que surgem ao utilizar-se o estimador de mínimos quadrados.

Numa situação em que existe uma função paramétrica interpretável que solucione os problemas do pesquisador, a multicolinearidade pode ser eliminada através de uma transformação de variáveis que ortogonalize o modelo. Óbviamente este não é o contexto analisado neste trabalho, esta transformação nos será útil para outros propósitos e esta apresentada no apêndice 1.

Alguns métodos alternativos são propostos na literatura, escolheu-se os mais conhecidos para objeto de estudo e comparação:

- Método das *COMPONENTES PRINCIPAIS*
- Método da *REGRESSÃO EM CRISTAS*
- Método das *RAÍZES LATENTES*
- Método dos *ESTIMADORES CONTRAÍDOS*
- Método *INTERATIVO*

Este capítulo descreve estas técnicas apresentando um breve histórico de sua utilização NO CONTEXTO DE ANÁLISE DE REGRESSÃO , como solução para os problemas induzidos pela multicolinearidade.

Apresentam-se também as propriedades conhecidas que justificam a utilização destes estimadores, como alternativa ao de mínimos quadrados.

2.1 - O MÉTODO DAS COMPONENTES PRINCIPAIS

A técnica de componentes principais é bem geral, aplicável a qualquer conjunto de dados multivariados, representado por uma amostra de tamanho n de um vetor aleatório: (X_1, X_2, \dots, X_p) .

O método determina um novo vetor, a partir de X_1, X_2, \dots, X_p , que vamos denotar por $Z = (Z_1, Z_2, \dots, Z_p)$, com as seguintes características:

$$Z_i = \sum_{j=1}^p a_{ij} X_j = a_i X \quad (1)$$

$$\text{var}(Z_1) \geq \text{var}(Z_2) \geq \dots \geq \text{var}(Z_p) \quad (2)$$

$$\sum_{i=1}^p \text{var}(Z_i) = \sum_{i=1}^p \text{var}(X_i) \quad (3)$$

$$\text{Cov}(Z_i Z_j) = 0 \text{ para } i \neq j \quad (4)$$

A transformação acima é linear e representa uma rotação do antigo sistema de eixos, (X_1, X_2, \dots, X_p) para o novo sistema (Z_1, Z_2, \dots, Z_p) , onde a variabilidade está mais evidenciada nas primeiras componentes, (expressão 2) acima. Pode-se ilustrar facilmente o caso em que $p=2$: Esta técnica pode ser aplicada no contexto de análise de regressão onde o vetor (X_1, X_2, \dots, X_p) representa as p variáveis preditoras do modelo $Y = X\beta + \epsilon$. Os primeiros estudos nesta direção são atribuídos a Kendall - 1957 - [43], que desenvolveu os principais resultados hoje conhecidos, por exemplo, a utilidade deste método na estimação de β na presença de multicolinearidade.

Muitos autores contribuíram para a formalização algébrica das idéias apresentadas por Kendall, destacando-se Massy - 1965 - [44] e Yoel Haitovsky - 1966 - [45], Marsh - 1982 - [27], oferece um resumo das publicações em ordem cronológica.

Os autores mais recentes têm contribuído no sentido de simplificar a apresentação matemática de $\hat{\beta}_{CP}$, expressando-o de forma conveniente facilitando, por exemplo, comparações com outros estimadores.

2.1.1 - DESCRIÇÃO FORMAL E ALGUMAS PROPRIEDADES DE $\hat{\beta}_{CP}$.

A transformação linear que ortogonaliza o modelo é dada por :

$$Z = \Gamma X \quad (1)$$

e é tal que :

$$Z^t X = \Delta$$

com:

Δ : Matriz diagonal cujos elementos : $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p$ são os autovalores de $X^t X$.

Γ : Matriz cujas colunas são os autovetores de $X^t X$, com $\Gamma^t \Gamma = I_d$.

As colunas de $Z = (Z_1, Z_2, \dots, Z_p)$ são chamadas de *COMPONENTES PRINCIPAIS*. A variancia de cada componente Z_j , é dada por :

$$var(Z_j) = \sum_{i=1}^n Z_{ij}^2 = \lambda_j$$

(Draper e Smith - 1981 - [30]) .

Na presença de multicolinearidade, um ou mais autovalores de $X^t X$, suponha que “ s ” autovalores, são próximos de zero: $\lambda_{p-s+1}, \dots, \lambda_p \simeq 0$. Pela relação (1) , Z_j pode ser aproximado por uma constante, identificando-se a combinação linear ,ou “ s ” combinações lineares, responsável pela multicolinearidade através dos elementos de Γ_j : $j=p-s+1, \dots, p$.

Os demais Z_j , associados a $\lambda_1, \dots, \lambda_{p-s}$, explicam *quase a totalidade* da variação original já que :

$$\sum_{j=1}^{p-s} var(Z_j) = \sum_{j=1}^{p-s} var(Z_j) \simeq \sum_{j=1}^p var(Z_j) = p$$

$\hat{\beta}_{CP}$, o estimador de β no modelo $Y = X\beta + \epsilon$, obtido através da técnica de COMPONENTES PRINCIPAIS é dado por :

$$\hat{\beta}_{CP} = \sum_{j=1}^{p-s} \lambda_j^{-1} \Gamma_j^t X^t Y \Gamma_j$$

Existem outras maneiras de obter-se $\hat{\beta}_{CP}$, entre elas é usual utilizar-se a noção de inversa generalizada de uma matriz, (Marquardt - 1970 - [20]), obtendo-se :

$$\hat{\beta}_{CP} = (Z^t Z)^+ Z^t Y$$

onde ,

$(Z^t Z)^+$ é a inversa generalizada de $Z^t Z$, descrita no apêndice 4.

As expressões para esperança, variância e erro quadrático médio de $\hat{\beta}_{CP}$ são mais facilmente obtidas utilizando-se esta formalização. Gunst e Mason - 1977 -[16], apresentam expressões simplificadas para estas estatísticas:

$$E(\hat{\beta}_{CP}) = \beta - \sum_{j=p-s}^p (\Gamma_j^t \beta) \Gamma_j$$

$$\sum \hat{\beta}_{CP} = \sigma^2 \sum_{j=1}^{p-s} \lambda_j^{-1} \Gamma_j \Gamma_j^t$$

$$\begin{aligned} EQM(\hat{\beta}_{CP}) &= Tr(\sum \hat{\beta}_{CP}) + (vies)^2(\hat{\beta}_{CP} = \\ &= \sigma^2 \sum_{j=1}^{p-s} \lambda_j^{-1} + \sum_{j=p-s+1}^p (\Gamma_j^p \beta)^2 \end{aligned}$$

Marquardt - 1970-[20], verificou que o viés de $\hat{\beta}_{CP}$ é uma função monotônica decrescente do posto de $X^t X$ e descreve $\hat{\beta}_{CP}$ como uma transformação linear de $\hat{\beta}_{MQ}$, (apêndice-4).

Podemos inferir pelos resultados obtidos até agora que selecionando-se as componentes obter-se-á maior precisão e melhores estimadores, pois os termos omitidos são exatamente os responsáveis pelos efeitos da presença de multicolinearidade.

Não existe um procedimento universal para se realizar esta seleção, o item seguinte apresenta tres métodos bastante discutidos na literatura.

2.1.2 - MÉTODOS DE SELEÇÃO DAS COMPONENTES PRINCIPAIS.

Os critérios para selecionar-se as componentes principais no modelo de regressão, geralmente tem como base dois indicadores:

- O tamanho dos autovalores de $X^t X$ e
- Testes estatísticos da significancia de cada componente do vetor de parâmetros.

CRITÉRIO - 1 -

Baseando-se somente no tamanho dos autovalores , alguns autores sugerem a eliminação das componentes associadas aos autovalores , menores que um determinado número , mais usualmente menores que “UM”.(Coxe - 1982 - [27]) .

CRITÉRIO - 2 -

Outros autores decidem eliminar as componentes até que uma certa porcentagem, mínima, da variancia total seja atingida. Por exemplo se esta porcentagem é 75 por cento e “ s ” componentes foram eliminadas, significa que:

$$\frac{\sum_{j=1}^{p-s} var(Z_j)}{\sum_{j=1}^p var(Z_j)} = \frac{\sum_{j=1}^{p-s} \lambda_j}{\sum_{j=1}^p \lambda_j} \geq 0.75$$

CRITÉRIO - 3-

Os testes estatísticos referem-se a significância de cada componente α_j , do vetor de parâmetros α , do modelo $Y = Z\alpha + \epsilon$, proveniente da transformação $Z = X\Gamma$ e $\alpha = \Gamma^t\beta$, (apêndice 1). A estatística para o teste $H_0 : \alpha_j = 0$ contra $H_a : \alpha_j \neq 0$, t_j , tem distribuição t-Student e é dada por :

$$t_j = \frac{\lambda_j^{1/2} \hat{\alpha}_j}{\sqrt{\hat{\sigma}_\alpha^2}}$$

com parâmetro de não centralidade dado por : $\mu_j = \frac{\lambda_j^{1/2} \alpha_j}{\sigma}$

onde:

$\hat{\alpha}_j$: é o estimador de mínimos quadrados de α_j .

e

$\hat{\sigma}_\alpha^2$: é o estimador da variancia de $\hat{\alpha}_j$, $var(\hat{\alpha}) = (Z^t Z)^{-1} \sigma^2$.

(Gunst e Mason - 1985 - [29])

Apesar de serem tres métodos bem aceitos na literatura, Jolliffe-1982-[28], chama a atenção para o fato de ao eliminar-se as componentes, reduzindo-se as superestimativas nas variancias, pode haver um aumento não controlado dos vieses dos estimadores, havendo portanto uma compensação no cálculo do erro quadrático médio , já que $EQM = VARIANCIA + VIES^2$.

Em resposta a este fato, Gunst e Mason - 1985 - [29] , oferece um estudo onde conclue que somente para os casos em que a multicolinearidade é muito forte, os resultados dos métodos descritos acima não são satisfatórios, ocorrendo por exemplo, uma redução acentuada do poder do teste t-Student no caso de testar-se a significância de cada componente.

Estudos empiricos teem mostrado que se os menores autovalores de $X^t X$ forem próximos de "0.1", há pouco comprometimento dos resultados de seleção discutidos acima.

Uma outra crítica usual é o fato destes critérios negligenciarem o valor preditivo dos termos excluidos. Gunst e Mason - 1977 - [16], comentam outros métodos que levam em consideração este valor preditivo, onde a eliminação NÃO tem como base, necessariamente o fato da componente estar associada a um autovalor muito pequeno, podendo as vezes, uma componente com esta característica, permanecer no modelo.æ

2.2 - REGRESSÃO EM CRISTAS

A nomenclatura, regressão “ ridge ” , traduzida neste trabalho como regressão em cristas, (Jorge Oishi - 1983 - [42]), tem sua origem na interpretação inicial do método criado por A. E. Hoerl em 1959. O item 2.2.1 descreve a técnica de análise de cristas na forma em que foi desenvolvida inicialmente. O item 2.2.2 apresenta sua extensão para a estimação dos parâmetros na regressão linear e o item 2.2.3 , descreve algumas propriedades deste estimador .

2.2.1 - DESENVOLVIMENTO ORIGINAL DO MÉTODO DE ANÁLISE DE CRISTAS

Para melhor compreensão apresentaremos inicialmente a interpretação geométrica desta técnica para posteriormente formalizarmos a sua descrição e os principais resultados de interesse neste trabalho.

A ORIGEM DA ANÁLISE DE CRISTAS.

Arthur E. Hoel em 1959, durante sua ligação profissional com o grupo Du Pont, (E. I. du Pont de Nemours e Company) , estava envolvido com problemas de otimização de processos industriais. A solução matemática destes problemas consistia na maximização ou minimização de funções do tipo:

$$(1) \quad Y = b_0 + \sum_{i=1}^p b_i x_i + \sum_{i=1}^{p-1} \sum_{j=i+1}^p b_{ij} x_i x_j + \sum_{i=1}^p b_{ii} x_i^2$$

onde,

- x_i : representa o efeito isolado de cada variável.
- $x_i x_j$: representa o efeito de interação.
- x_i^2 : representa o efeito quadrático de cada variável.

Este mesmo modelo em forma matricial é dado por:

$$(2) \quad Y = b_0 + b^t X_i + (1/2)X_i^t B X \quad \text{onde,}$$

- b_{px1} : é o vetor de coeficientes lineares.
- X_i : é um vetor $px1$ que contem os valores das variáveis independentes,(no contexto de regressão sera a i -ésima linha de X) .
- b_0 : representa o termo constante.
- B_{pp} : é uma matriz simétrica cujos valores da diagonal são duas vezes os termos quadráticos e os valores abaixo da diagonal são os termos de interação.

Quando as variáveis estão padronizadas, de maneira que suas medias sejam iguais a zero e variancias iguais a um, a região que contem os pontos observados pode ser interpretada como uma superfície centrada na origem.

O método consiste em determinar os pontos de máximo e mínimo, chamados de picos e vales , da superfície Y , restritos a região descrita por $\sum_{i=1}^p x_i^2 = r^2$, com r fixo.

Não só os pontos são determinados como também os caminhos mais curtos a partir da origem que levam a estes pontos. Estes caminhos são chamados de cristas, originando o título do método.

Uma analogia bastante ilustrativa é a descrita por Jorge Oishi - 1983 - [42]:

-“ Suponha que Y representa uma superfície num espaço Euclidiano de dimensão 3 e que voce esta de olhos vendados , em cima da origem $(0,0)$ e deseja sair do lugar e localizar o ponto mais baixo,(alto) ,da superfície. O procedimento que voce adota é estendendo uma das pernas examinar a superfície ao redor numa circunferencia cujo raio r é a projeção do comprimento da sua perna.Voce identifica e marca o ponto mais baixo, (alto) ,da superfície naquela circunferencia.

Suponha ainda que voce possa variar a vontade o comprimento da sua perna e repetindo o procedimento, voce determina todos os pontos de mínimo, (máximo) , a partir da origem.

Unindo todos estes pontos, sobre as diversas circunferencias, voce obtem um caminho que certamente passa pelo mínimo,(máximo) ,absoluto .

Este caminho recebe o nome de CRISTA , (ridge) , da função.”

Desta forma pode-se determinar os pontos críticos da superfície em função de sua distancia ao centro $(0,0)$.Isto é equivalente a determinar os pontos críticos da função Y , restritos a uma circunferencia de raio r , com centro na origem.

Passaremos a descrever formalmente o método, ressaltando a interpretação gráfica, de maneira a ter-se no plano bidimensional uma descrição ampla do comportamento de uma superfície , mesmo para o caso de mais de duas variáveis independentes.

- DESCRIÇÃO ALGÉBRICA DO MÉTODO DE CRISTAS -

Para facilitar o tratamento algébrico , consideraremos um modelo com tres variáveis independentes. A equação (1) do item 2.2.1 , para $p=3$, é dada por:

$$Y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_3 + b_{11} x_1^2 + b_{22} x_2^2 + b_{33} x_3^2 + b_{12} x_1 x_2 + b_{13} x_1 x_3 + b_{23} x_2 x_3 \quad (3)$$

com a restrição:

$$x_1^2 + x_2^2 + x_3^2 = r^2 \quad \text{ou} \quad x_3 = (+)(-)\sqrt{r^2 - x_1^2 - x_2^2} \quad (4)$$

O passo inicial é determinarmos os pontos críticos , candidatos a máximos ou mínimos da função .

Substituindo-se (4) em (3) e derivando Y em relação a x_1 e x_2 , obtem-se:

$$\frac{\partial Y}{\partial x_1} = \left\{ \frac{x_1(b_3 + b_{13}x_1 + b_{23}x_2)}{\sqrt{r^2 - x_1^2 - x_2^2}} + b_{13}\sqrt{r^2 - x_1^2 - x_2^2} + (b_1 + 2b_{11}x_1 - 2b_{33}x_1 + b_{12}x_2) \right\}$$

$$\frac{\partial Y}{\partial x_2} = \left\{ \frac{x_2(b_3 + b_{13}x_1 + b_{23}x_2)}{\sqrt{r^2 - x_1^2 - x_2^2}} + b_{23}\sqrt{r^2 - x_1^2 - x_2^2} + (b_2 + 2b_{22}x_2 - 2b_{33}x_2 + b_{12}x_1) \right\}$$

Igualando-se a zero , tem-se :

$$x_1 \left[2b_{11} - 2b_{33} - \frac{(b_3 + b_{13}x_1 + b_{23}x_2)}{x_3} \right] + b_{12}x_2 + b_{13}x_3 = -b_1$$

$$x_2 \left[2b_{22} - 2b_{33} - \frac{(b_3 + b_{13}x_1 + b_{23}x_2)}{x_3} \right] + b_{12}x_1 + b_{23}x_3 = -b_2$$

A resolução deste sistema é simplificada introduzindo-se o seguinte parâmetro:

$$\lambda = \frac{b_3 + b_{13}x_1 + b_{23}x_2}{x_3}$$

Obtem-se então o seguinte sistema :

$$b_{13}x_1 + b_{23}x_2 - \lambda x_3 = -b_3$$

$$[2b_{11} - 2b_{33} - \lambda]x_1 + b_{12}x_2 + b_{13}x_3 = -b_1 \quad (5)$$

$$b_{12}x_1 + [(2b_{22} - 2b_{33}) - \lambda]x_2 + b_{23}x_3 = -b_2$$

Fixando-se λ os valores de x_1, x_2, x_3 podem ser determinados e representam as coordenadas de um ponto na esfera $\sum_{i=1}^p x_i^2 = r^2$.

Vamos analisar para que valores de λ o sistema (5) tem solução.

Para tal considere a matriz cujos elementos são os coeficientes de x_1, x_2, x_3 :

$$(B - \lambda I) = \begin{pmatrix} 2(b_{11} - b_{33}) - \lambda & b_{12} & b_{13} \\ b_{12} & 2(b_{22} - b_{33}) - \lambda & b_{23} \\ b_{13} & b_{23} & -\lambda \end{pmatrix}$$

O sistema(5) , não tem solução se o determinante da matriz $(B - \lambda I)$ for igual a zero. Isto acontece quando λ for igual a um dos autovalores de B . Caso contrario, o sistema tem solução que é dada pelos pontos críticos da função:

$$Y + b_0 + b^t x + (1/2)x^t B x$$

sujeito a restrição $x^t x = r^2$.

E esta solução é dada por:

$$x_s = -(B - \lambda I)^{-1} b , \quad (6)$$

onde λ NÃO é um autovalor de B .

Pode-se classificar estes pontos x_s , no sentido de determinar se são pontos de sela, de máximos ou de mínimo, para tal, verifica-se a posição de λ em relação aos autovalores de B .

Considere $\lambda_1 > \lambda_2 > \dots > \lambda_p$ os autovalores de B. Prova-se que :

- Se $\lambda > \max(\lambda_1, \lambda_2, \dots, \lambda_p)$, a solução do sistema (5) representa um ponto de *máximo absoluto* na esfera definida pelo raio r , que por sua vez é função de λ .

- Se $\lambda < \min(\lambda_1, \lambda_2, \dots, \lambda_p)$ a solução do sistema representa um ponto de *mínimo absoluto* na esfera definida pelo raio r .

- Se $\lambda_j < \lambda < \lambda_{j+1}$, os pontos encontrados são de *máximo ou mínimo locais* .

É interessante notar ainda que se existem autovalores positivos e negativos, a superfície possui um *ponto de sela* .

Além disto Y cresce ou decresce na direção de um ponto de máximo ou de mínimo somente uma vez dentro da esfera de raio r , ou seja, a crista é monotônica no conjunto de números reais, com no máximo uma exceção ,(R. W. Hoerl - 1987 - [46]) .

A interpretação original do problema é um pouco diferente do que a apresentada acima. Originalmente selecionar-se-ia o raio r da circunferencia sobre a qual se deseja estudar e determinar-se-ia os pontos de máximos e mínimos dentro desta região pré fixada. Isto exigiria um grande número de cálculos, existindo ainda a possibilidade de obter-se raízes múltiplas, ou ainda o número de variáveis independentes ser muito grande.

A introdução do parâmetro λ , simplifica bastante os cálculos e em contra partida r não é mais escolhido e sim determinado em função de λ .

É interessante notar que o método de maximização , (ou de minimização) , descrito é o próprio método de LAGRANGE, onde λ é o multiplicador . Verificaremos no item 2.2.2 que λ determina $(\hat{\beta}, r, \hat{Y})$, no contexto de análise de regressão.(R. W. Hoerl - 1985 - [31]).

RECURSOS GRÁFICOS DO MÉTODO EM CRISTAS.

A riqueza gráfica deste método esta analisada detalhadamente nas referencias A. E. Hoerl - 1959 - [40] e R. W. Hoerl - 1985 - [31]. Neste trabalho faremos uma pequena ilustração somente daqueles que serão utilizados posteriormente no contexto de estimação dos parâmetros do modelo de regressão linear .

GRÁFICO - 1 -

É o gráfico de Y por r , onde se representa os valores de máximo ou mínimo de Y , para cada valor de r . Em outras palavras é a evolução da crista dentro da região delimitada pela circunferencia cujo raio é r .

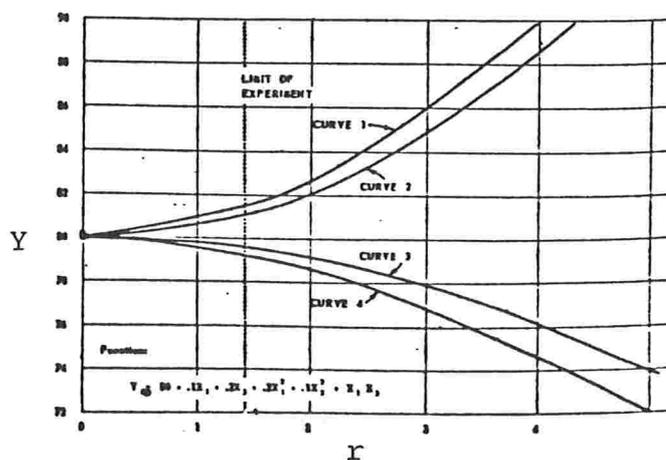


GRÁFICO - 2 -

É um gráfico de curvas de nível, portanto cada curva representa um valor fixo de Y . Neste gráfico podemos desenhar as cristas, que são os caminhos obtidos unindo-se os pontos de máximo ou de mínimo de acordo com a variação de r .

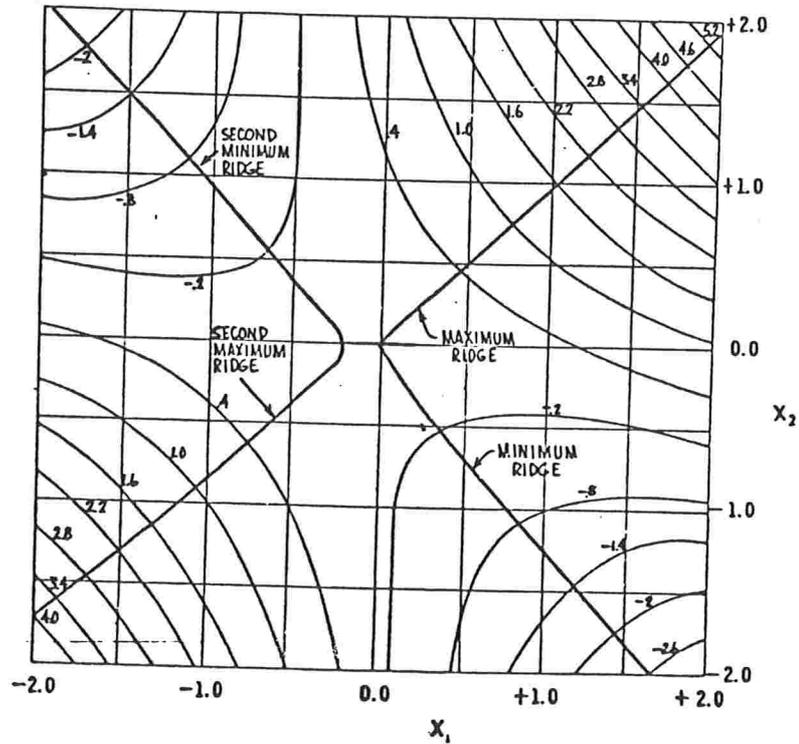
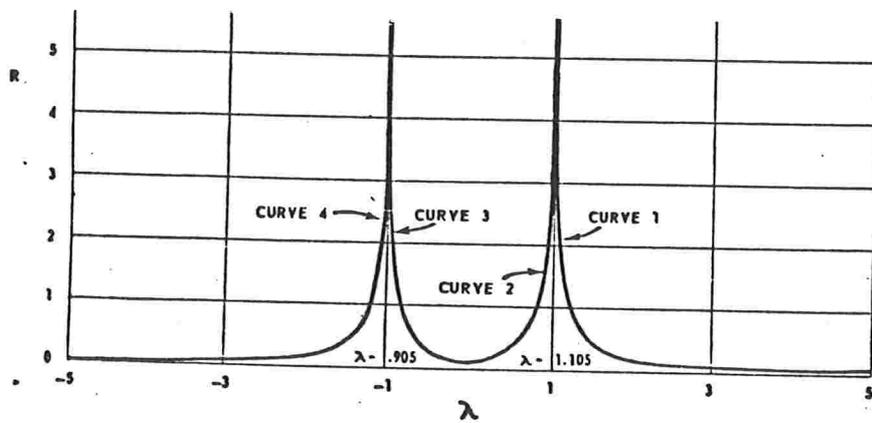


GRÁFICO - 3 -

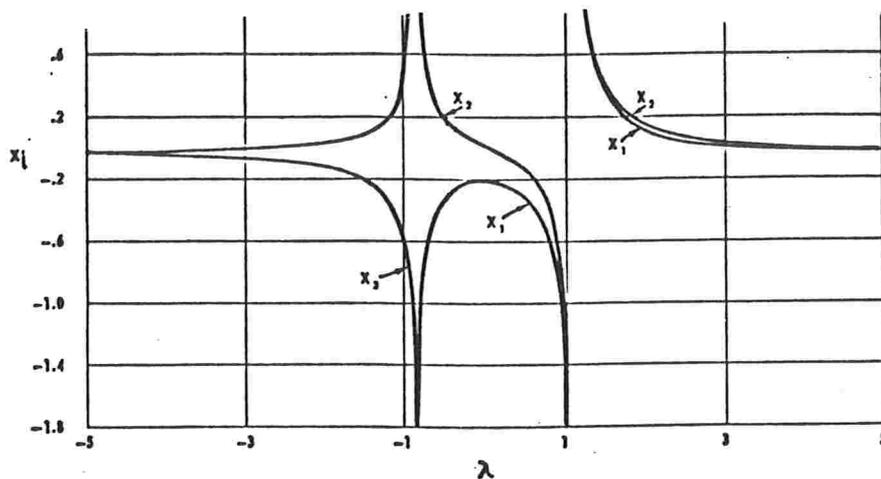
É o gráfico de r por λ , ou seja, fornece a variação de λ de acordo com a distancia, r , a origem.



GRÁFICOS - 4

O gráfico 4 permite detectar a solução mais próxima da origem para a qual os coeficientes começam a estabilizar-se.

Estes coeficientes são as coordenadas dos pontos de máximo ou de mínimo encontrados variando-se r ou λ .



Durante o tempo em que A. E. Hoerl desenvolvia o método de análise em cristas, uma outra questão chamava sua atenção: Os problemas da estimação dos parâmetros na regressão linear utilizando-se a técnica de mínimos quadrados, quando as variáveis preditoras são correlacionadas.

Hoerl notou que os parâmetros numa regressão linear podiam ser estimados pelo método de cristas, obtendo-se algumas vantagens sobre o método de mínimos quadrados. Desta forma surgiu a regressão em cristas, cujo desenvolvimento original está descrito na seção seguinte.

2.2.2 - EXTENSÃO DA ANÁLISE DE CRISTAS COMO MÉTODO DE ESTIMAÇÃO DE β NA REGRESSÃO LINEAR

A. E. Hoerl notou que a soma de quadrados residual numa regressão linear múltipla, denotada por “SQE”, pode ser escrita da mesma forma que as funções que estavam sendo estudadas através da análise de cristas:

$$Y = b_0 + b^t x + (1/2)x^t B x$$

equação 1, item 2.2.1

com :

$$Y = SQE$$

$$B = X^t X$$

$$b_{px1} = -X^t Y$$

$$x_{px1} = \beta$$

$$b_0 = 0$$

ou seja, o SQE é dado por:

$$SQE = (Y - X\beta)^t (Y - X\beta) = Y^t Y - 2\beta^t X^t Y + \beta^t (X^t X) \beta$$

Esta ligação permitiu o desenvolvimento do estimador em cristas obtido através da minimização do SQE sujeito a restrição $\beta^t \beta = r^2$ ou seja $\sum_{i=1}^p \beta_i^2 = r^2$.

O ponto de mínimo encontrado, x_s , na expressão 6, da seção anterior, é o estimador de β procurado.

Seguindo o mesmo procedimento já descrito, utiliza-se o método de Lagrange para minimizar-se F:

$$F = (Y - X\beta)^t (Y - X\beta) - \lambda(\beta^t \beta - r^2)$$

cuja derivada é dada por:

$$\frac{\partial F}{\partial \beta} = -2X^t Y + 2X^t X \beta - 2\lambda \beta$$

onde λ ,é o multiplicador de Lagrange.

Segundo os resultados descritos no item 2.2.1, para determinarmos o ponto de mínimo de F ,devemos tomar λ menor que o menor autovalor de B , $B = X^t X = Corr(X)$. Neste caso , B é não negativa definida , todos os autovalores são maiores que zero:

$0 \leq \lambda_0 \leq \dots \leq \lambda_1$.Tem-se então 3 escolhas :

- $\lambda = 0$.
- $0 \leq \lambda \leq \lambda_p$.
- $\lambda \leq 0 \leq \lambda_p$.

- Se $\lambda = 0$, obtem-se a solução :

$$x = -(B - \lambda I)^{-1} b = (X^t X)^{-1} X^t Y$$

que é a solução de mínimos quadrados, sendo o ponto de mínimo absoluto de F , sem restrição.

- Se tomarmos $0 < \lambda < \lambda_1$ obteremos uma solução mais longe da origem que a solução de mínimos quadrados.

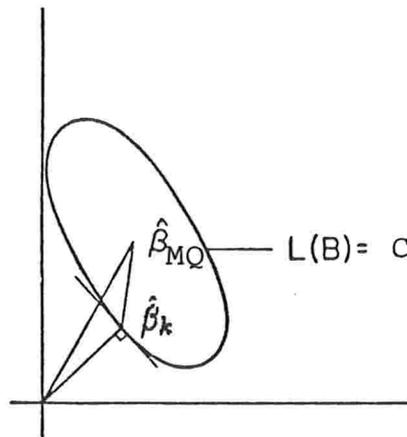
Então escolhe-se $\lambda < 0$ que possibilita a determinação de um ponto de mínimo de F dentro da circunferencia $\beta^t \beta = r^2$ que é mais próximo a origem que $\hat{\beta}_{MQ}$.

Com base nas considerações acima define-se como ESTIMADOR EM CRISTAS, (ridge), a solução dada por:

$$x = -(B - \lambda I)^{-1} b = (X^t X - \lambda I)^{-1} X^t Y =$$
$$\hat{\beta}_k = (X^t X + kI)^{-1} X^t Y$$

onde $k = -\lambda > 0$.

Na figura abaixo , representamos o estimador em cristas e o estimador de mínimos quadrados para $p=2$:



A elipse $L(\hat{B}) = (Y - X\hat{\beta})^t(Y - X\hat{\beta}) = C, C > 0$ é centrada em $\hat{\beta}_{MQ}$, e é o lugar geométrico dos pontos no espaço paramétrico onde a soma de quadrados residual é constante igual a C e maior que o valor mínimo absoluto obtido pela solução de mínimos quadrados.

A circunferencia é tangente a elipse no ponto dado pela solução do método de cristas, de forma que esta solução tem menor norma euclidiana, portanto é mais próximo da origem que a solução de mínimos quadrados.

De modo geral , a solução baseada no método de cristas é a que tem menor comprimento entre todos os vetores que representem soluções com a mesma soma de quadrados residual.

A utilização deste método no contexto de regressão linear com o objetivo de estimação dos parâmetros concretizou-se com uma publicação , em 1970, fruto do trabalho conjunto de A. E. Hoerl e seu assistente R. W. Kennard -[19]. Nesta publicação os autores apresentaram detalhadamente a utilidade e a importancia deste estimador e sua interpretação gráfica .Destaca-se o TRAÇO DA CRISTA , (gráfico 4 da seção 2.2) ,gráfico de “ k ” por $\hat{\beta}_k$ onde se pode determinar o valor de “k” a partir do qual os elementos de $\hat{\beta}_k$ começam a estabilizar-se.

Surgiram muitas contribuições e outras interpretações para esta técnica, contudo a filosofia original facilita a compreensão de sua aplicabilidade no contexto análise de regressão . O item seguinte descreve algumas propriedades de $\hat{\beta}_k$, entre elas, sua utilização na presença de multicolinearidade.

2.2.3 - ALGUMAS PROPRIEDADES DO ESTIMADOR $\hat{\beta}_k$.

Verifica-se que a principal diferença em relação a $\hat{\beta}_{MQ}$ é o acréscimo na diagonal de $X^t X$ de uma constante , k :

$$\hat{\beta}_k = (X^t X + kI)^{-1} X^t Y$$

Este acréscimo , elimina a instabilidade computacional no cálculo de $(X^t X)^{-1}$ no caso em que X é quase singular, ou seja na presença de multicolinearidade .Convém lembrar também que nesta situação um ou mais autovalores de $X^t X$ são muito próximos de zero, já os autovalores de $(X^t X + kI)$ são iguais a $\lambda_i + k$, $i=1, \dots, p$.

Existe uma outra formulação mais abrangente chamada forma canônica do estimador em cristas,(Hoerl - Kennard - 1970 -[19] e Lawless -1976 - [33]) :

$$\hat{\beta}_k + (X^t X + K)^{-1} X^t Y$$

onde :

- X esta padronizada, portanto as variáveis independentes possuem médias iguais a zero e variancias iguais a um.

- K é uma matriz diagonal, cujos elementos $K_i > 0$ são diferentes entre si .

Da mesma forma que expressamos $\hat{\beta}_{MQ}$ e $\hat{\beta}_{CP}$ como combinação linear dos autovetores de $X^t X$, o mesmo pode ser feito com $\hat{\beta}_k$, (Gunst e Mason - 1977 - [16]) :

$$\hat{\beta}_k = \sum_{i=1}^p (\lambda_j + k)^{-1} [\Gamma_j^t X^t Y] \Gamma_j = \sum_{i=1}^p c_j \Gamma_j$$

onde:

$$\frac{\Gamma_j^t X^t Y}{(\lambda_j + k)} = c_j \text{ é um escalar .}$$

Para k fixo,conhece-se algumas propriedades deste estimador. Na prática é usual utilizar-se as estimativas de mínimos quadrados de β como ponto de partida para estimar-se k , tornando-o uma variável aleatória .

Deduz-se, para k fixo, expressões para esperança , variancia e erro quadrático médio:

$$E(\hat{\beta}_k) = \beta - k \sum_{j=1}^p (\lambda_j + k)^{-1} (\Gamma_j^t \beta) \Gamma_j$$

$$\sum_{\hat{\beta}_k} = \sigma^2 \sum_{j=1}^p \lambda_j (\lambda_j + k)^{-2} \Gamma_j \Gamma_j^t$$

Para calcularmos o EQM vamos determinar a soma das variancias de $\hat{\beta}_{k(i)}$ e o viés de $\hat{\beta}_k$.

Hoerl e Kennard - 1970 - [19] , mostraram que :

- O $\text{tr}(\sum \hat{\beta}_k)$ é uma função *monotônica decrescente em k*, dada por:

$$\text{tr}(\sum_{\hat{\beta}_k}) = \sigma^2 \sum_{j=1}^p \lambda_j (\lambda_j + k)^{-2}$$

e que também o viés $\hat{\beta}_k$ é uma função *monotônica crescente em k* , cujo quadrado é dado por :

$$\text{viés}^2(\hat{\beta}_k) = k^2 \sum_{j=1}^p (\lambda_j + k)^{-2} (\Gamma_j^t \beta)^2$$

Obtendo-se :

$$EQM(\hat{\beta}_k) = \sigma^2 \sum_{j=1}^p \lambda_j (\lambda_j + k)^{-2} + k^2 \sum_{j=1}^p (\lambda_j + k)^{-2} (\Gamma_j^t \beta)^2$$

Desta forma o $EQM(\hat{\beta}_k)$, primeiro decresce com k , atinge um valor mínimo e depois cresce, (Gunst e Mason - 1977 -[16]) , o que indica a existencia de um valor ótimo.

Este valor é tradicionalmente determinado utilizando-se o gráfico 5 da seção 2.2 chamado o traço da crista, que é o gráfico de $\hat{\beta}_{k(i)} x k, k > 0$. Verifica-se o comportamento de $\hat{\beta}_{k(i)}$, na medida em que k varia, sendo de interesse o valor de k para o qual os $\hat{\beta}_{k(i)}$ começam a estabilizar-se.

A maior controvérsia deste método se refere exatamente a escolha de k, uma vez que as propriedades descritas em Hoerl e Kennard -1970 - [19], só são válidas para k fixo.

Ainda se desconhece uma solução melhor, no entanto estudos empíricos, como simulações, tem mostrado a superioridade desta técnica com relação a de mínimos quadrados, mesmo para o caso de estimar-se k baseando-se em outras estimativas iniciais dos parâmetros.

É necessário pesquisar-se mais profundamente a estimação de " k " , sendo de utilidade saber-se que $\hat{\beta}$ pode ser escrito como uma transformação linear de $\hat{\beta}_{MQ}$:

$$\hat{\beta}_k = (X^t X + kI)^{-1} X^t Y \quad \text{com } X^t Y = (X^t X) \hat{\beta}_{MQ} \quad \text{e portanto :}$$

$$\hat{\beta}_k = (X^t X + kI)^{-1} (X^t X) \hat{\beta}_{MQ} =$$

$$\hat{\beta}_k = M_k \hat{\beta}_{MQ} \quad \text{com } M_k = (X^t X + kI)^{-1} (X^t X) \quad \text{uma matriz } p \times p .$$

2.3 - O MÉTODO DAS RAÍZES LATENTES

Este método é recente na literatura, (1973-1974), tendo a vantagem de possuir uma interpretação geométrica atraente e um enfoque inovador ao dar atenção para as informações contidas na matriz :

$$(Y \ X)^t (Y \ X)$$

ao invés de :

$$(X^t X)$$

como é habitual no contexto de regressão linear.

De forma análoga a técnica de componentes principais aplicada a regressão linear, após a detecção das prováveis combinações lineares existentes entre as variáveis preditoras, constrói-se o estimador $\hat{\beta}_{LR}$, eliminando-se os termos responsáveis pela multicolinearidade, tendo como objeto central de análise a matriz $(Y \ X)^t (Y \ X)$, referida na literatura frequentemente como a *matriz X aumentada pela coluna Y*.

Apresentaremos um breve histórico do desenvolvimento deste método, sua interpretação geométrica e a formalização matemática do estimador, finalizando com uma descrição das propriedades estatísticas conhecidas até o presente momento.

O método das raízes latentes foi desenvolvido independentemente por D. M. Hawkins em 1973 - [11] e Webster - Gunst e Mason em 1974 - [10], esta última formalização é a adotada neste trabalho, por sua simplicidade e frequência na literatura.

As publicações subsequentes mais importantes são dos mesmos autores, destacando-se White - Gunst - 1979 - [26], onde se encontra algumas propriedades assintóticas de $\hat{\beta}_{RL}$ sob determinadas condições sobre os parâmetros β e σ^2 .

Há também contribuições de outros autores, por exemplo, Sharma - James - 1981 - [12], consegue uma expressão matemática única para $\hat{\beta}_{RL}, \hat{\beta}_{MQ}, \hat{\beta}_k$, diferenciando-os por um fator de ponderação, facilitando as comparações entre os métodos.

Na ausência de um estudo analítico mais completo das propriedades estatísticas, métodos empíricos, como simulações, tem guiado os defensores desta técnica como alternativa a de mínimos quadrados na presença de multicolinearidade.

2.3.1 - DESCRIÇÃO DO MÉTODO DAS RAÍZES LATENTES

Da mesma forma que os autovalores e autovetores de (X^tX) são elementos importantes na detecção, caracterização e eliminação da multicolinearidade, os autovalores e autovetores de (A^tA) , $A = (YX)$, permitem-nos:

- Identificar a singularidade em X .
- Determinar se a singularidade encontrada em X , tem importância na relação de dependência entre Y e X .
- Obter um novo estimador, $\hat{\beta}_{RL}$, que pode ser interpretado como o estimador de mínimos quadrados retirando-se os termos responsáveis pela multicolinearidade.

Uma diferença importante entre o estimador $\hat{\beta}_{RL}$ e o estimador $\hat{\beta}_{CP}$ consiste no fato de através dos autovalores e autovetores de (A^tA) poder-se detectar se algumas das combinações lineares encontradas em A tem importância na determinação de Y , o que não é possível na técnica de componentes principais, com isto define-se os conceitos:

- SINGULARIDADE PREDITIVA: que se refere a singularidade existente em (YX) devido exclusivamente a especificação do modelo, que descreve Y como uma combinação linear de X : $Y = X\beta + \epsilon$

- SINGULARIDADE NÃO PREDITIVA: que se refere aquela que é proveniente de combinações lineares em X , sem valor preditivo, ou seja, as que produzem muito pouca ou quase nenhuma influência em Y .

O atributo maior de $\hat{\beta}_{RL}$, fica por conta da eliminação dos termos que dizem respeito somente a singularidade não preditiva, no intuito de não prejudicar a estimação de Y .

A seção seguinte mostra a interpretação geométrica do método, facilitando a compreensão e identificação dos termos referentes a singularidade preditiva e não preditiva.

2.3.2 - INTERPRETAÇÃO GEOMÉTRICA DO MÉTODO DAS RAIZES LATENTES

Os autovetores de $(A^t A)$, representados pelas colunas de Υ , definem um conjunto de eixos ortogonais: $(\Upsilon_0, \Upsilon_1, \dots, \Upsilon_p)$, que gera um espaço de dimensão $(p+1)$, representado pela base ortonormal, (Z_0, Z_1, \dots, Z_p) .

Este mesmo espaço pode ser definido pelos eixos ortogonais $(Y, X_1, X_2, \dots, X_p)$ e uma outra base ortonormal: $e = (e_0, e_1, \dots, e_p)$.

Os eixos Z_j podem ser representados por combinações lineares de elementos da base e , da seguinte maneira:

$$Z_j = \sum_{r=0}^p \Upsilon_{rj} e_r$$

ou seja,

$$Z_{ij} = \Upsilon_{0j} Y_i + \sum_{i=1}^p X_{ir} \Upsilon_{rj}$$

O primeiro elemento de Υ_j , que é Υ_{0j} , representa o cosseno entre os eixos Y e Z_j , os outros elementos, Υ_{ij} , $i=1, \dots, p$, representam os cossenos entre X_i e Z_j , $i=1, \dots, p$. Resumindo:

$$\Upsilon_j = \begin{pmatrix} \Upsilon_{0j} \\ \Upsilon_{1j} \\ \vdots \\ \vdots \\ \Upsilon_{(p+1)j} \end{pmatrix} = \begin{pmatrix} \cos Z_j \hat{Y} \\ \cos Z_j \hat{X}_1 \\ \vdots \\ \vdots \\ \cos Z_j \hat{X}_p \end{pmatrix}$$

Para definirmos $\hat{\beta}_{RL}$, precisamos primeiro identificar a singularidade não preditiva, separando-a daquela de carácter preditivo. Para tal, denotaremos os autovalores de $A^t A$ por ω_j , $j=0, 1, \dots, p$, com $\omega_0 \leq \omega_1 \leq \dots \leq \omega_p$. Verificaremos na seção seguinte que:

$$\omega_j = \sum_{i=1}^n \left\{ \Upsilon_{0j} Y_i + \sum_{r=1}^p X_{ir} \Upsilon_{rj} \right\}^2$$

Suponha que “ r ” autovalores de $A^t A$ são próximos de zero, então “r-s” entre estes autovalores, são próximos de zero devido exclusivamente a relação $Y = X\beta + \epsilon$ com $r - s > 0$, ou seja, existe pelo menos um autovalor de $A^t A$ próximo de zero, já que Y é descrito como uma função linear de X. Os “ s ” autovalores restantes, $s \geq 0$ dizem respeito as relações de dependência linear que envolvem somente elementos de X. A distinção entre estes dois tipos de *proximidade a zero* é facilmente descrita verificando-se que na expressão de ω_j , acima, tem-se duas situações :

- $\omega_j \simeq 0$ e $\Upsilon_{0j} \neq 0$: neste caso a obtém-se da expressão para ω_j :

$$\Upsilon_{0j} Y_i + \sum_{i=r}^p X_{ir} \Upsilon_{rj} \simeq 0$$

Identificando-se uma combinação linear de *CARACTER PREDITIVO*.

- $\omega_j \simeq 0$ e $\Upsilon_{0j} \simeq 0$: neste caso obtém-se :

$$\sum_{r=1}^p X_{ir} \Upsilon_{rj} \simeq 0$$

Identificando-se, portanto, uma combinação linear que envolve somente os X_j , logo, de *CARACTER NÃO PREDITIVO*. É interessante notar que, Υ_{0j} , é o cosseno entre os eixos Y e Υ_j , ou seja Υ_j é quase ortogonal a Y.

A seção seguinte descreve formalmente $\hat{\beta}_{RL}$, mostrando que ele pode ser interpretado como uma modificação do estimador de mínimos quadrados devido a eliminação dos termos referentes a singularidade não preditiva.

2.3.3 - OBTENÇÃO DE $\hat{\beta}_{LR}$

Para facilidade de compreensão, vamos dividir em duas partes o caminho para obter $\hat{\beta}_{LR}$. Na primeira parte deduziremos $\hat{\beta}_{MQ}$ a partir dos autovetores de $A^t A$. Na segunda parte identificaremos as parcelas relacionadas a multicolinearidade não preditiva e eliminando-as obteremos $\hat{\beta}_{LR}$.

PARTE - 1 - Obtenção de $\hat{\beta}_{MQ}$ a partir dos autovetores de $A^t A$.

Ja vimos que o estimador de mínimos quadrados, pode ser descrito como uma combinação linear dos autovetores de $X^t X$. De forma análoga, podemos expressá-lo como combinação linear de “*parte dos elementos*” dos autovetores de $A^t A$.

Considere a transformação linear de A descrita por $A\Upsilon$, então :

$$(A\Upsilon)^t(A\Upsilon) = \Omega$$

onde $\Omega_{(p+1) \times (p+1)}$, é a matriz diagonal cujos elementos são ω_i . Podemos escrever:

$$(A\Upsilon_j)^t(A\Upsilon_j) = \omega_j \tag{1}$$

$$= \Upsilon_j^t(A^t A)\Upsilon_j = \omega_j$$

$$= \sum_{i=1}^n (Y_i \Upsilon_{0j} + \sum_{r=1}^p x_{ir} \Upsilon_{rj})^2 = \omega_j$$

fazendo $\omega_j = 0$, tem-se :

$$\sum_{i=1}^n (Y_i \Upsilon_{0j} + \sum_{r=1}^p x_{ir} \Upsilon_{rj})^2 = 0$$

então ,

$$Y_i \Upsilon_{0j} + \sum_{r=1}^p x_{ir} \Upsilon_{rj} = 0$$

A equação acima, pode ser reescrita substituindo-se Y_i pela variável Y_i^* , definida pela relação:

$$Y_i = \frac{Y_i^* - \bar{Y}^*}{\sqrt{\sum_{i=1}^n (Y_i^* - \bar{Y}^*)^2}}$$

obtendo-se:

$$\hat{Y}_i^* = \bar{Y}^* - \eta \Upsilon_{0j}^{-1} \sum_{r=1}^p X_{ir} \Upsilon_{rj} \quad (2)$$

$$\text{com } \eta = \sqrt{\sum_{i=1}^n (Y_i^* - \bar{Y}^*)^2}$$

ou, em notação matricial:

$$Y_i^* = \bar{Y}^* - \eta \Upsilon_{0j}^{-1} X \Upsilon_j^0 \quad j=0,1,\dots,p. \quad (3)$$

onde $\Upsilon_j^{0t} = (\Upsilon_{1j}, \Upsilon_{2j}, \dots, \Upsilon_{pj})$, ou seja, é Υ_j sem o primeiro elemento: Υ_{0j} .

Se $\Upsilon_{0j} \neq 0$ $j=0,1,\dots,p$, obtem-se em (3), $(p+1)$ equações de regressão do tipo $Y = a + bX$ com:

$$a = \bar{Y}^* \quad , j=0,1,\dots,p.$$

$$b = -\eta \Upsilon_{0j}^{-1} \Upsilon_j^0 \quad , j=0,1,\dots,p.$$

Nenhuma das $(p+1)$ equações acima, pode ser considerada isoladamente no sentido de estimar-se Y_i^* com precisão, é razoável portanto considerar-se como estimador de Y_i^* , uma combinação linear dos diversos \hat{Y}_i^* :

$$\hat{Y}^* = \sum_{i=1}^p c_i \hat{Y}_i^*$$

tomando-se $c_i = a_j \Upsilon_{0j}$ de forma que $\sum_{j=0}^p a_j \Upsilon_{0j} = 1$, obtem-se:

$$\hat{Y}^* = \sum_{j=0}^p a_j \Upsilon_{0j} \hat{Y}_j^* = \sum_{j=0}^p a_j \Upsilon_{0j} [\bar{Y}^* - \eta \Upsilon_{0j}^{-1} X \Upsilon_j^0]$$

$$\hat{Y}^* = \bar{Y}^* - \eta X \left(\sum_{j=0}^p a_j \Upsilon_j^0 \right)$$

onde:

$$b = -\eta \sum_{j=0}^p a_j \Upsilon_j^0 \quad (4)$$

O SQE, soma de quadrados residual, obtida para \hat{Y}^* , definido acima é dada por:

$$SQE = (Y - \hat{Y}^*)^t (Y - \hat{Y}^*) = \eta^2 a^t \Omega a = \eta^2 \sum_{j=0}^p a_j^2 \omega_j$$

Utiliza-se o método de Lagrange para minimizar-se o SQE sujeito a restrição

$$\sum_{j=0}^p a_j \Upsilon_{0j} = 1$$

Obtendo-se :

$$f(a) = \eta^2 \sum_{j=0}^p a_j^2 \omega_j - 2\mu_0 \left(\sum_{j=0}^p a_j \Upsilon_{0j} - 1 \right) \quad (5)$$

onde :

$$-2\mu_0 \quad , \text{ é o multiplicador de Lagrange, com } \mu_0 = \eta^2 \left(\sum_{j=0}^p \Upsilon_{0j}^2 \omega_j^{-1} \right)^{-1}$$

A solução é dada por:

$$a_j = \eta^{-2} \omega_j^{-1} \Upsilon_{0j} \mu_0, j = 0, 1, \dots, p$$

Substituindo-se esta solução em (4) , obtem-se:

$$b = -\eta \sum_{j=0}^p a_j \Upsilon_j^0 = -\eta \left(\sum_{j=0}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1} \sum_{j=0}^p \Upsilon_{0j} \omega_j^{-1} \Upsilon_j^0 = \hat{\beta}_{MQ}$$

Mostrou-se que $\hat{\beta}_{MQ}$ pode ser escrito como combinação linear de Υ_j^0 ,que é Υ_j sem o elemento Υ_{0j} ,j=0,1,...,p.

PARTE - 2 - Obtenção de $\hat{\beta}_{RL}$

Já verificamos que a inovação deste estimador é eliminar somente a singularidade não preditiva , retendo a singularidade com valor preditivo . Suponha que os autovetores $\Upsilon_0, \Upsilon_1, \dots, \Upsilon_{s-1}$ correspondam a situação que descreve a singularidade não preditiva, portanto $\omega_j \simeq 0$ e $\Upsilon_{0j} \simeq 0$, para $j=0,1,\dots,s-1$.O procedimento para minimizarmos o SQE levando este fato em consideração é o mesmo descrito na parte - 1 - ,sendo que na expressão (5) , considera-se em f(a) , $a_0 = a_1 = a_2 = \dots = a_{s-1} = 0$, sujeita a mesma restrição.

Obtem-se como solução :

$$a_j = \Upsilon_{0j} \omega_j^{-1} \left(\sum_{j=s}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1}$$

para $j=s, (s+1), \dots, p$ e portanto $\hat{\beta}_{RL}$, é dado por:

$$\hat{\beta}_{RL} = -\eta \sum_{j=s}^p a_j \Upsilon_j^0 = -\eta \left(\sum_{j=s}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1} \sum_{j=s}^p \Upsilon_{0j} \omega_j^{-1} \Upsilon_j^0$$

2.3.4 - ALGUMAS PROPRIEDADES DE $\hat{\beta}_{RL}$.

Além das propriedades descritas nas seções anteriores , que em parte justificam o uso deste estimador como alternativa para o método de mínimos quadrados, pouco se conhece sobre $\hat{\beta}_{RL}$, basta dizer que nada se sabe a respeito de sua distribuição.

Webster - Gunst e Mason - 1976 - [21] , obtiveram algumas conclusões do comportamento deste método através de simulações. Alguns resultados são apresentados a seguir.

- 1. - O menor autovalor de $A^t A$ é muito próximo do menor autovalor de $X^t X$: $\omega_0 \simeq \lambda_p$.

- 2. - Se $\omega_0 \simeq 0$ e $|\Upsilon_{00}| \simeq 0$, tem-se $\Upsilon_0^0 \simeq \Gamma_p$, ou seja existe uma proximidade entre os autovetores que correspondem aos menores autovalores, tanto de $X^t X$ como de $A^t A$. Convém lembrar que $\Upsilon_0^0 = \Upsilon_0$ sem o primeiro elemento Υ_{00} .

Este resultado é bastante intuitivo já que tanto Υ_j^0 quanto Γ_p descrevem a multicolinearidade existente em X :

$$\sum_{i=1}^p X_i \Gamma_{i1} \simeq 0 \quad \text{item 1.4.5 .}$$

$$\sum_{i=1}^p X_i \Upsilon_j^0 \simeq 0 \quad \text{item 2.3.3 .}$$

Verificou-se que a esperança e o erro quadrático médio de $\hat{\beta}_{CP}$, são boas aproximações para $E(\hat{\beta}_{RL})$ e $EQM(\hat{\beta}_{RL})$:

$$E(\hat{\beta}_{RL}) \simeq E(\hat{\beta}_{CP}) = \beta - \sum_{j=1}^{s-1} (\Gamma_j^t \beta) \Gamma_j$$

$$EQM(\hat{\beta}_{RL}) \simeq EQM(\hat{\beta}_{CP}) = \sigma^2 \sum_{j=s}^p \lambda_j^{-1} + \sum_{j=1}^{s-1} (\Gamma_j^t \beta)^2$$

Sharma - James - 1981 - [12] , descreve os estimadores $\hat{\beta}_k, \hat{\beta}_{MQ}, \hat{\beta}_{LR}$, através de uma expressão algébrica única, diferenciando-os por um fator de ponderação, apresentando-os como combinações lineares de Υ_j^0 . Esta expressão é dada por :

$$- \left\{ \frac{\sum_{j=0}^p p_j \Upsilon_{0j} \omega_j^{-1} \Upsilon_j^0}{\sum_{j=0}^p p_j \Upsilon_{0j}^2 \omega_j^{-1}} \right\}$$

onde p_j é o fator de ponderação, cujos valores particulares, definem os estimadores citados acima da seguinte forma:

$$\hat{\beta}_k \quad \text{para } p_j = \frac{1}{\omega_j + k}$$

$$\hat{\beta}_{MQ} \quad \text{para } p_j = -\eta$$

$$\hat{\beta}_{RL} \quad \text{para } \begin{cases} p_j = 0, & \text{se } \omega_j = 0 \\ p_j = 1, & \text{se } \omega_j \neq 0 \end{cases}$$

Isto facilita a comparação entre os estimadores, contudo as limitações ainda existentes só serão eliminadas mediante a ampliação do conhecimento sobre o comportamento de $\hat{\beta}_{RL}$.

2.4 - OS ESTIMADORES CONTRAÍDOS

A terminologia *estimador contraído*, é mais recente na literatura do que a própria idéia de se estreitar o tamanho do estimador, via um fator multiplicativo do tipo:

$$\hat{\beta}_\rho = \rho \hat{\beta}_{MQ}$$

com $\rho > 0$.

J. W. Tukey em 1960, estudando a regressão polinomial, propôs multiplicar o estimador de mínimos quadrados do parâmetro relacionado ao termo quadrático, por uma constante ρ , com $\rho \in (0, 1)$.

Este novo estimador, sendo menor que $\hat{\beta}_{MQ}$, tem menor erro quadrático médio, quando β , o verdadeiro valor do parâmetro, é pequeno; e tem maior viés, portanto maior erro quadrático médio, quando β é grande.

Embora Tukey tenha proporcionado esta idéia inicial, a contribuição maior é atribuída a C. Stein, mediante seus estudos iniciados em 1956, referentes a estimação da média de um vetor com distribuição normal, p-variada, que permitiu a introdução dos estimadores contraídos na literatura.

A primeira publicação data de 1961, sendo fruto de seu trabalho conjunto com W. James. Já neste primeiro trabalho, apresentam propriedades estatísticas relevantes, por exemplo a existência de condições sobre o modelo, que permitem a estes estimadores possuírem menor *erro quadrático médio generalizado** que o estimador de mínimos quadrados.

Estes resultados foram ampliados, transformando o estimador de James - Stein, num caso particular de uma classe mais geral de estimadores desenvolvida por Baranchik em 1964, 1967 e 1970.

A ligação destes resultados, relacionando-os com o contexto de análise de regressão deve-se a Stanley - Sclove - 1968 - [34].

Uma apresentação resumida e simples dos estimadores mais conhecidos pode ser encontrada em Mayer e Willke - 1973 - [15], de onde se pesquisou grande parte dos resultados apresentados neste trabalho.

* Oportunamente apresentar-se-á o erro quadrático médio generalizado : EQMG, que por enquanto pode ser entendido como uma ponderação dos elementos do erro quadrático médio usual.

2.4.1 - APRESENTAÇÃO DO ESTIMADOR CONTRAÍDO

Todos os estimadores estudados até agora, $\hat{\beta}_{CP}, \hat{\beta}_k, \hat{\beta}_{RL}$, fazem parte de uma classe mais geral de estimadores, a classe das transformações lineares de $\hat{\beta}_{MQ}$.

Seja C esta classe.

Então se $b(A) \in C$, com A uma matriz fixa, podemos escrever:

$$b(A) = A\hat{\beta}_{MQ}$$

e

$$\begin{aligned} SQE(b(A)) &= (Y - Xb(A))^t(Y - Xb(A)) = \\ &= (Y - X\hat{\beta}_{MQ})^t(Y - X\hat{\beta}_{MQ}) + \hat{\beta}_{MQ}^t(A - I)^t(A - I)\hat{\beta}_{MQ} \\ &= SQE(\hat{\beta}_{MQ}) + L(A) \end{aligned}$$

com $L(A) = \hat{\beta}_{MQ}^t(A - I)^t(A - I)\hat{\beta}_{MQ}$

Definimos uma subclasse de C , $C(\tau)$, de modo que se $b(A) \in C(\tau)$, $L(A) = \tau$, ou seja,

$$SQE(b(A)) = SQE(\hat{\beta}_{MQ}) + \tau$$

Hoerl e Kennard - 1970 - [19], mostraram que o estimador de $\hat{\beta}_k$, cristas, possui mínima norma euclidiana em $C(\tau)$.

Mayer e Willke - 1973 - [15], mostraram que se considerarmos uma definição adequada de norma, o estimador contraído passa a ser o estimador de norma mínima em $C(\tau)$.

RESULTADO - 1 -

Se $A = \rho I$ para algum $\rho \in (0, 1)$ e $b(A) \in C(\tau)$ então:

$$n^*b(\rho I) = \min_{C(\tau)} n^*b(A)$$

onde:

- $n^*(b(A))$, "norma dependente dos dados" é definida por:

$$n^*b(A) = b^t(A)(X^tX)b(A)$$

Ou seja, o estimador $\rho\hat{\beta}_{MQ}$ é de mínima norma, $n^*(b(A))$ em $C(\tau)$.

Com isto vamos definir uma nova classe de estimadores:

Considere uma classe de estimadores definida por:

$$\hat{\beta}_\rho = \rho \hat{\beta}_{MQ}$$

com $\rho \in (0, \infty)$

O estimador $\hat{\beta}_\rho$ é chamado de *Estimador Contraído*, e ρ é o *fator de contração*.

- Se ρ é um escalar fixo, então $\hat{\beta}_\rho$ é *DETERMINISTICAMENTE CONTRAÍDO*.

- Se ρ é uma função de $\hat{\beta}_{MQ}^t \hat{\beta}_{MQ}$, ou seja $\rho = f(\hat{\beta}_{MQ}^t \hat{\beta}_{MQ})$, então $\hat{\beta}_\rho$ é dito ser *ESTOCÁSTICAMENTE CONTRAÍDO*.

Da mesma forma que o estimador $\hat{\beta}_k$ possui um parâmetro k para ser previamente fixado, antes de se determinar o valor do estimador, também em $\hat{\beta}_\rho$, devemos pré determinar ρ .

O método proposto por Hoerl e Kennard - 1970 - [19], para estimar k em $\hat{\beta}_k$ é examinar o traço da crista. Este método não serve para a estimação de ρ , pois $\hat{\beta}_\rho$ cresce linearmente com ρ . Para alguns casos como veremos posteriormente, podemos usar um método análogo.

As estimativas de k e ρ oferecem problemas semelhantes, pois geralmente baseiam-se em $\hat{\beta}_{MQ}$, o que os torna variáveis aleatórias, sendo que as expressões para os momentos só são conhecidas para k e ρ fixos.

A seção seguinte apresenta algumas propriedades conhecidas para tres destes estimadores, os mais discutidos na literatura.

2.4.2 - ALGUMAS PROPRIEDADES ESTADÍSTICAS DE ALGUNS ESTIMADORES CONTRAÍDOS

O ESTIMADOR $\hat{\beta}_\rho = \rho\hat{\beta}_{MQ}$, para ρ fixo.

Embora não se saiba claramente como se estimar ρ sem a ajuda inicial da estimativa de mínimos quadrados de β , é fácil verificar que para ρ fixo, tem-se :

$$- E(\hat{\beta}_\rho) = \rho\beta$$

$$- \text{Cov}(\hat{\beta}_\rho) = \rho^2\sigma^2(X^tX)^{-1}$$

$$- \text{viés}(\hat{\beta}_\rho) = (1 - \rho)^2\beta^t\beta$$

$$- \text{EQM}(\hat{\beta}_\rho) = \rho^2\sigma^2 \text{tr} (X^tX)^{-1} + (1 - \rho)^2(\beta^t\beta)$$

Alem de ser um estimador bastante simples, $\hat{\beta}_\rho$, possui menor EQM do que o estimador $\hat{\beta}_{MQ}$. Outras vantagens deste estimador serão apresentadas no capítulo 3.

O ESTIMADOR $\hat{\beta}_\delta = f(\delta, \hat{\beta}_{MQ})\hat{\beta}_{MQ}$

O resultado abaixo define um estimador estocasticamente contraído, que possui menor $\text{tr}(\text{Cov}(\hat{\beta}))$, dentro da classe $C(\tau)$, definida no item anterior.

RESULTADO - 2 -

Considere ainda a classe $C(\tau)$, dos estimadores do tipo $b(A) = A\hat{\beta}_{MQ}$ com :
 $SQE(b(A)) = SQE(\hat{\beta}_{MQ}) + \tau$, prova-se que:

Se $A = A^* = \delta\hat{\beta}_{MQ}\hat{\beta}_{MQ}^t(I + \delta\hat{\beta}_{MQ}\hat{\beta}_{MQ}^t)^{-1}$, para algum $\delta \in (0, \infty)$ tem-se :

$$tr(\sum b(A^*)) = \min_{C(\tau)} tr(\sum b(A))$$

e ainda que :

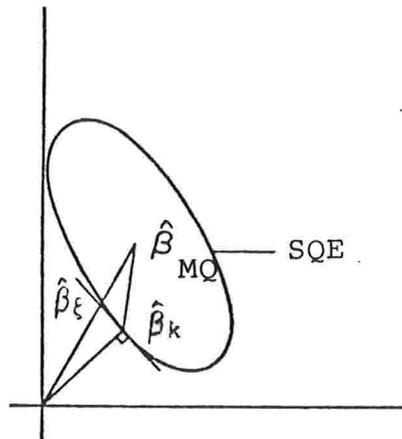
$$\hat{\beta}_\delta = \delta\hat{\beta}_{MQ}\hat{\beta}_{MQ}^t(I + \delta\hat{\beta}_{MQ}\hat{\beta}_{MQ}^t)^{-1}\hat{\beta}_{MQ}$$

é um estimador estocásticamente contraído .

A diferença entre os estimadores deterministicamente e estocásticamente contraídos é mínima na prática pois:

- Se $\hat{\beta}_\rho$ e $\hat{\beta}_\delta$ pertencem a $C(\tau)$, para τ fixo, então $\hat{\beta}_\rho = \hat{\beta}_\delta$.
- Se ρ for fixado após observar-se $\hat{\beta}_{MQ}$, então ele se torna estocásticamente contraído.

O esquema abaixo posiciona $\hat{\beta}_{MQ}, \hat{\beta}_k, \hat{\beta}_\rho$ ou $\hat{\beta}_\delta$, na elipse descrita por :
 $SQE(b(A)) = SQE(\hat{\beta}_{MQ}) + \tau$.



A elipse é o lugar geométrico dos pontos que tem soma de quadrados residual constante.

O estimador $\hat{\beta}_{MQ}$ está no centro, o estimador $\hat{\beta}_k$, esta tangente a elipse no ponto mais próximo da origem e o estimador $\hat{\beta}_\rho$ ou $\hat{\beta}_\delta$, é a intersecção entre a elipse e a linha que liga o centro, $\hat{\beta}_{MQ}$, a origem $(0, 0)$.

O ESTIMADOR DE JAMES E STEIN : $\hat{\beta}_\xi$.

Existe ainda um outro estimador estocásticamente contraído, $\hat{\beta}_\xi$, frequentemente utilizado na literatura, que possui EQMG, erro quadrático médio generalizado, mínimo em $C(\tau)$.

Neste momento vamos somente apresentar $\hat{\beta}_\xi$, os resultados comparativos com outros estimadores será detalhadamente explorado no capítulo 3.

Considere $b(A) = A\hat{\beta}_{MQ}$ com $A = A^* = [1 + \xi s^2 (\hat{\beta}_{MQ}^t \hat{\beta}_{MQ})^{-1}]$, onde :

$$s^2 = Y^t Y - \hat{\beta}_{MQ}^t (X^t X) \hat{\beta}_{MQ} = SQE(\hat{\beta}_{MQ})$$

$$\hat{\beta}_\xi = A^* \hat{\beta}_{MQ} = [1 + \xi s^2 (\hat{\beta}_{MQ}^t \hat{\beta}_{MQ})^{-1}] \hat{\beta}_{MQ}$$

Estes tres estimadores são os mais simples encontrados na literatura entre os estimadores contraídos. Existem muitos aspectos ainda não solucionados ou que precisam ser revistos. A seção seguinte apresenta alguns comentários neste sentido.

2.4.3 - COMENTARIOS E PROBLEMAS NÃO SOLUCIONADOS.

Embora os resultados anteriores sejam de relevancia, existe uma grande barreira na utilização destes estimadores devido ao desconhecimento de suas distribuições e mesmo de seus momentos.

Não há procedimentos ótimos indicados na literatura para se estimar ρ , δ ou ξ . Um procedimento análogo ao descrito por Hoerl e Kennard -1970 - [19], o traço da crista, tem sido satisfatório somente para o caso da estimação de δ , escolhe-se δ para o qual os componentes de $\hat{\beta}_\delta$ comecem a estabilizar, para tal utiliza-se o gráfico de δ por $\hat{\beta}_{\delta(i)}$.

Mais esforços são necessários, mesmo, porque não há como garantir que na presença de multicolinearidade o estimador contraído terá menor erro quadrático médio que o estimador de mínimos quadrados, contudo, alguns autores tem obtido resultados favoráveis neste sentido, através de simulações.

Existem outras formas de se apresentar o estimador contraído, bem como outros resultados referentes a estimadores mais particulares. Veja por exemplo Farebrother - 1978-[35] que apresenta uma outra classe descrita por:

$$\hat{\beta}_* = \hat{\beta}_{MQ} + A^t(b - \hat{\beta}_{MQ})$$

onde $A_{p \times p}$ e $b_{p \times 1}$ são matrizes arbitrárias.

É possível que o melhor caminho a seguir seja uma caracterização mais completa dos estimadores contraídos mais simples, possibilitando uma interpretação adequada e com isto maior aplicabilidade.

2.5 - ESTIMADOR INTERATIVO

Este estimador foi desenvolvido mais recentemente, tendo surgido na literatura em 1978 com a publicação de Gotz Trenkler: AN ITERATION ESTIMATOR FOR THE LINEAR MODEL - Compstat -1978 WIEN: Physica-Verlag , 125-131.

Vamos denotá-lo por $\hat{\beta}_{IT}$. O seu comportamento é análogo a $\hat{\beta}_k$, no sentido que ambos, tanto o estimador em cristas como o interativo convergem para $\hat{\beta}_{MQ}$.

O resultado desta convergencia no caso do estimador $\hat{\beta}_{IT}$ deve-se ao trabalho de Ben - Israel e Charnes em 1963: CONTRIBUTIONS TO THE THEORY OF GENERALIZED INVERSES - J.Soc.Indust. Appl.Math. 11, 667-699.

Gotz Trenkler , em 1980 - [36] e em 1984 - [24], estudou a comparação deste com outros estimadores viesados mais comuns na literatura , de onde extraiu-se os resultados apresentados neste trabalho.

2.5.1 - DEFINIÇÃO E ALGUMAS PROPRIEDADES ESTATÍSTICAS

Considere o modelo usual de regressão , $Y = X\beta + \epsilon$, sob as hipóteses usuais.

O estimador interativo é dado por:

$$\hat{\beta}_{m,\alpha} = \alpha \sum_{i=0}^m (I - \alpha X^t X)^i X^t Y$$

com:

$$0 < \alpha < \frac{1}{\lambda_1}$$

para $m = 0, 1, \dots$ e $\lambda_1 > \lambda_2 > \dots > \lambda_p$, autovalores de $X^t X$.

Vamos denotá-lo por $\hat{\beta}_{IT}$, exceto em situações onde se faz necessário lembrar que este estimador depende de dois parâmetros, “m ” e α .

Uma outra forma de escreve-lo é :

$$\hat{\beta}_{m,\alpha} = X_{m,\alpha} Y$$

com $X_{m,\alpha} = \alpha \sum_{i=0}^m (I - \alpha X^t X)^i X^t$.

A vantagem desta última formalização é que a matriz $X_{m,\alpha}$ pode ser calculada por um processo iterativo , onde :

$$X_{0,\alpha} = \alpha X^t$$

$$X_{(m+1),\alpha} = (I - \alpha X^t X) X_{m,\alpha} + \alpha X^t$$

Esta sequencia , definida por $X_{m,\alpha}$, converge quando $m \rightarrow \infty$ para $(X^t X)^{-1} X^t$, mesmo quando X é singular:

$$\lim_{m \rightarrow \infty} X_{m,\alpha} = (X^t X)^{-1} X^t$$

- O SEGUNDO MOMENTO E O VIÉS DE $\hat{\beta}_{IT}$.

Gotz Trenkler - 1980 - [36] , apresenta as expressões para o viés e para a matriz de var covar de $\hat{\beta}_{IT}$:

$$\sum(\hat{\beta}_{IT}) = (X_{m,\alpha} X - I) \beta \beta^t (X_{m,\alpha} X - I) + \sigma^2 X_{m,\alpha} X^t_{m,\alpha}$$

e o viés é dado por:

$$\Gamma^t \Lambda_{m,\alpha} \Gamma \beta \beta^t \Gamma^t \Lambda_{m,\alpha} \Gamma$$

onde: Γ $p \times p$ uma matriz cujas colunas são os autovetores normalizados de $X^t X$.

$\Lambda_{m,\alpha}$ matriz diagonal cujos elementos são dados por $-(1 - \alpha \lambda_j)^{m+1}$, $j=1,2,\dots,p$.

SEMELHANÇA COM O ESTIMADOR EM CRISTAS - $\hat{\beta}_k$.

Existe uma semelhança caracterizada unicamente pela convergencia ao estimador de mínimos quadrados, pois também para $\hat{\beta}_k$ tem-se:

$$\lim_{k \rightarrow \infty} (X^t X + kI)^{-1} X^t = (X^t X)^{-1} X^t$$

O que, de maneira análoga a $X_{m,\alpha}$, também ocorre quando X é singular.

ALGUMAS VANTAGENS DE $\hat{\beta}_{IT}$.

Apesar de ser um estimador menos conhecido que os já apresentados $\hat{\beta}_{IT}$ oferece algumas vantagens:

- Não é necessário nenhum cálculo de inversão de matrizes para obter-se as estimativas.
- A escolha de $\alpha \in (0, \frac{1}{tr(X^t X)})$ é suficiente para se ter a convergência de $\hat{\beta}_{IT}$, mesmo que X seja de posto não completo, ou seja, mesmo na presença de multicolinearidade.
- Pode-se examinar o “traço da interação” , que é análogo ao traço da crista, para determinar-se o valor ótimo de “ m ”. É o gráfico de $\hat{\beta}_{IT}$ por $SQE(\hat{\beta}_{IT})$ para valores crescentes de m ,onde se escolhe um valor de “ m ” para o qual os elementos de $\hat{\beta}_{IT}$ comecem a estabilizar-se, portanto os resultados obtidos para a escolha de k na regressão em cristas, são aplicáveis para a escolha de m em $\hat{\beta}_{IT}$.

O capítulo de comparações entre os estimadores , capítulo 3, contem outras informações que auxiliam na compreensão de se ter em $\hat{\beta}_{IT}$ uma alternativa para $\hat{\beta}_{MQ}$ na presença de multicolinearidade, baseando esta conclusão no cálculo do erro quadrático médio generalizado ,onde apresentaremos as condições sobre as quais se tem $EQMG(\hat{\beta}_{IT}) < EQMG(\hat{\beta}_{MQ})$.

COMPARAÇÕES ENTRE OS ESTIMADORES

Geralmente as comparações entre os estimadores viesados são feitas utilizando-se o erro quadrático médio - EQM:

$$EQM(\hat{\beta}) = E(\hat{\beta} - \beta)^t(\hat{\beta} - \beta) = \sum_{I=1}^p [var(\hat{\beta}_i) + vies^2(\hat{\beta}_i)]$$

Considera-se melhor o estimador que apresentar menor EQM .

No caso dos cinco estimadores estudados no capítulo anterior, somente algumas comparações são possíveis , utilizando-se o EQM, dependendo ainda de determinadas condições sobre os parâmetros β e σ^2 .

Em 1974 , C. M. Theobald - [37] , apresentou um teorema que envolve comparações entre estimadores viesados , baseando-se no erro quadrático médio generalizado, o **EQMG** , que será definido oportunamente .

Com este novo parâmetro ,é possível maior número de comparações, além de reafirmar aquelas obtidas utilizando-se o EQM .

Este capítulo esta dividido em duas partes :

- No item 3.1 , apresenta-se todas as comparações possíveis utilizando-se o EQM .
- No item 3.2, apresenta-se as comparações possíveis utilizando-se o EQMG.

Em ambos os itens , tem-se as comparações de cada estimador viesado com o estimador de mínimos quadrados, em seguida aquelas referentes aos estimadores viesados entre si :

- *Método das Componentes Principais: $\hat{\beta}_{CP}$* .
- *Método da regressão em Cristas : $\hat{\beta}_k$*
- *Método das raízes Latentes : $\hat{\beta}_{RL}$*
- *Estimador Contraído : $\hat{\beta}_\rho, \hat{\beta}_\delta, \hat{\beta}_\xi$*
- *Método Interativo : $\hat{\beta}_{IT}$* .

3.1 - COMPARAÇÕES ENTRE OS ESTIMADORES UTILIZANDO-SE O ERRO QUADRÁTICO MÉDIO - EQM

Entre os estimadores viesados estudados , não se conhece expressões para o EQM de $\hat{\beta}_k$, cristas, para k aleatório, $\hat{\beta}_\delta$, qualquer estimador contraído com parâmetro estocástico e $\hat{\beta}_{RL}$, o estimador obtido pelo método das raízes latentes.

Para contornar esta situação , utiliza-se , quando possível , aproximações empíricas para $\hat{\beta}_{RL}$ verificou-se no item 2.3.4 que:

$$EQM(\hat{\beta}_{RL}) \simeq EQM(\hat{\beta}_{CP})$$

E no caso de $\hat{\beta}_k$ e $\hat{\beta}_\delta$, com k e δ aleatórios , é procedimento usual , utilizar-se a expressão do EQM , obtida quando os parâmetros são fixos.

Desta forma somente algumas comparações são possíveis e estão descritas neste item.

3.1.1 - COMPONENTES PRINCIPAIS x MÍNIMOS QUADRADOS

$$\hat{\beta}_{CP} \text{ x } \hat{\beta}_{MQ}$$

RESULTADO - 1 -

Uma condição suficiente para que $EQM(\hat{\beta}_{CP}) < EQM(\hat{\beta}_{MQ})$, é dada por :

$$\sum_{j=(p-s+1)}^p \frac{1}{\lambda_j} > \frac{1}{\sigma^2}(\beta^t \beta)$$

onde $\lambda_1 > \lambda_2 > \dots > \lambda_p$, λ_i , autovalores de $X^t X$ com $\lambda_{(p-s+1)} \simeq \dots \simeq \lambda_p \simeq 0$.

(D. W. Maquardt - 1970 - [20])

Prova:

Vimos no item 2.1.1 que ,

$$EQM(\hat{\beta}_{CP}) = \sigma^2 \sum_{j=1}^{p-s} \frac{1}{\lambda_j} + \sum_{j=(p-s+1)}^p (\Gamma_j^t \beta)^2$$

e na expressão 4 do item 1.3.1

$$EQM(\hat{\beta}_{MQ}) = \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}$$

A condição é obtida diretamente da resolução da equação $EQM(\hat{\beta}_{CP}) < EQM(\hat{\beta}_{MQ})$:

$$\sigma^2 \sum_{j=1}^{p-s} \frac{1}{\lambda_j} + \sum_{j=p-s+1}^p (\Gamma_j^t \beta)^2 < \sigma^2 \sum_{j=1}^p \frac{1}{\lambda_j}$$

$$\sigma^2 \sum_{j=p-s+1}^p \lambda_j^{-1} > \sum_{j=p-s+1}^p (\Gamma_j^t \beta)^2 \tag{1}$$

Esta última expressão é a condição que precisa ser satisfeita para considerar-se o estimador $\hat{\beta}_{CP}$ melhor que $\hat{\beta}_{MQ}$.

Na prática, a verificação desta condição é simplificada, sendo substituída por uma condição mais forte:

Seja $\Gamma^t \beta = \xi$ então $\xi^t \xi = \beta^t \beta$ e portanto:

$\beta^t \beta \geq \sum_{j=p-s+1}^p \xi_j^2$, onde $\xi_j = \Gamma_j^t \beta$ que em (1) resulta em :

$$\sigma^2 \sum_{j=p-s+1}^p \frac{1}{\lambda_j} > \beta^t \beta$$

c. q. d.

3.1.2 - REGRESSÃO EM CRISTAS x MÍNIMOS QUADRADOS

$$\hat{\beta}_k \text{ x } \hat{\beta}_{MQ}$$

A existencia de um valor ótimo de k , de maneira a atingir-se um valor mínimo para o $EQM(\hat{\beta}_k)$ já foi discutida no item 2.2.3, onde vimos que $EQM(\hat{\beta}_k)$ primeiro decresce com k , atinge um mínimo e depois cresce.

O resultado seguinte, obtido por Hoerl e Kennard - 1970 -[19], apresenta este valor ótimo para k fixo.

RESULTADO - 2 -

Sempre existe $k > 0$ tal que $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{MQ})$.

Prova:

Vimos no item 2.2.3 que :

$$EQM(\hat{\beta}_k) = \varphi_1(k) + \varphi_2(k) = \sigma^2 \sum_{j=1}^p \lambda_j (\lambda_j + k)^{-2} + k^2 \sum_{j=1}^p (\lambda_j + k)^{-2} (\Gamma_j^t \beta)^2$$

Com :

$\varphi_1(k) = tr \sum(\hat{\beta}_k)$: uma função contínua monotônica decrescente em k .

$\varphi_2(k) = vies^2(\hat{\beta}_k)$: uma função contínua monotônica crescente em k .

e φ_1' e φ_2' não positiva e não negativa, respectivamente.

Além disto , vimos que $\hat{\beta}_k = \hat{\beta}_{MQ}$ para $k = 0$, item 2.2.2 ,então para provarmos que existe k tal que $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{MQ})$, basta provarmos que na vizinhança do zero , $EQM(\hat{\beta}_k)$ decresce . Ou seja desejamos determinar $k_0 > 0$, de maneira que para $k < k_0$ tenha-se :

$$\frac{\partial EQM(\hat{\beta}_k)}{\partial k} < 0$$

$$\frac{\partial EQM(\hat{\beta}_k)}{\partial k} = -2\sigma^2 \sum_{i=1}^p [\lambda_i(\lambda_i + k)] + 2k \frac{\sum_{i=1}^p [\lambda_i(\Gamma_i^t \beta)^2]}{(\lambda_i + k)^3}$$

Mostra-se que se :

$$k < \frac{\sigma^2}{\max_j (\Gamma_j^t \beta)^2} \quad \text{tem-se} \quad \frac{\partial EQM(\hat{\beta}_k)}{\partial k} < 0 \quad (2)$$

De forma análoga ao que foi visto para $\hat{\beta}_{CP}$, verificou-se que a existencia de um valor ótimo de k , depende do tamanho de $\beta^t \beta$.

Pela expressão (2) , acima, se $\beta^t \beta$ for muito grande, parece impossível tomar-se $k \neq 0$ tal que $EQM(\hat{\beta}_k) < \hat{\beta}_{MQ}$ para $0 < k < k_0$. Na prática isto não ocorre, tanto é que o gráfico da crista sempre tem permitido a determinação de valores satisfatórios.

Existe uma outra formulação do resultado -2, que é muito encontrada na literatura:

RESULTADO - 2' -

Se $\beta^t \beta$ for limitada , então existe $k > 0$, fixo tal que $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{MQ})$.
(Marquardt - 1970 - [20]) .

Este resultado é mais conhecido e é um caso particular do resultado - 2 - , acima demonstrado.

3.1.3 - RAÍZES LATENTES x MÍNIMOS QUADRADOS

$$\hat{\beta}_{RL} \text{ x } \hat{\beta}_{MQ}$$

Conforme se verificou no item 2.3.4 , não se conhece expressões para a esperança, variância ou erro quadrático médio do estimador $\hat{\beta}_{RL}$. No entanto estudos empíricos tem fornecido parâmetros de comparação com outros estimadores, obtendo-se resultados muito satisfatórios do ponto de vista prático.

Através de simulações realizadas em situações típicas, que ocorrem na prática, Webster - Gunst e Mason - 1974 -[10] ,observaram que se obtém uma redução bastante significativa do $EQM(\hat{\beta}_{RL})$ em relação ao $EQM(\hat{\beta}_{MQ})$,verificando ainda ser bastante razoável a aproximação de $EQM(\hat{\beta}_{RL})$ por $EQM(\hat{\beta}_{CP})$.

Resultados um pouco mais abrangentes foram obtidos por Gunst e Mason -1976 - [21] onde as simulações realizadas , revelaram que as estimativas dos elementos de β , que *não são* relacionadas a multicolinearidade , são praticamente idênticas em ambos os processos, mínimos quadrados e raízes latentes e ainda verificaram existir uma única situação onde o estimador de mínimos quadrados é melhor que o estimador das raízes latentes. Para descrevermos esta situação vamos supor que o autovetor associado ao menor autovalor de $X^t X$, Γ_p , é tal que alguns de seus elementos são muito pequenos, próximos de zero, obtendo-se como aproximação da expressão $\Gamma_p X \simeq 0$ a expressão:

$$\Gamma_{1p}X_1 + \Gamma_{2p}X_2 + \Gamma_{5p}X_5 \simeq 0$$

para o caso em que somente Γ_{1p} , Γ_{2p} , Γ_{5p} , forem consideradas diferentes de zero. Conclui-se então que X_1 , X_2 , X_5 , são as variáveis mais importantes na descrição da multicolinearidade e ainda que os efeitos sobre a estimação de β_1 , β_2 e β_5 são mais intensos. Vamos considerar estes componentes de β e Γ_p , separados , compondo os vetores β^* e Γ_p^* :

β^* : é um subconjunto de elementos de β , composto somente por aqueles envolvidos com a multicolinearidade, neste exemplo $\beta^* = (\beta_1, \beta_2, \beta_5)$.

Γ_p^* : um subconjunto dos elementos de Γ_p , que são aqueles de maior peso na descrição da multicolinearidade , no exemplo acima , $\Gamma_p^* = (\Gamma_{1p}, \Gamma_{2p}, \Gamma_{5p})$.

O resultado - 3 apresenta a condição em que se tem $\hat{\beta}_{MQ}$ mais preciso que $\hat{\beta}_{RL}$ em termos dos vetores Γ_p^* e β^* :

RESULTADO - 3 -

Se $\beta^* = c\Gamma^*$ com $|c|$ grande, ou seja β^* , for paralelo a Γ_p^* então :

$$EQM(\hat{\beta}_{MQ}) < EQM(\hat{\beta}_{RL})$$

Convém fazermos alguns comentários que tornem evidente o resultado acima.

Estudos empíricos tem mostrado que existe algumas relações entre os autovetores de X^tX , Γ_j , e os autovetores de A^tA sem o primeiro elemento, Υ_j^0 . Entre elas verificou-se que:

$$\Upsilon_j^0 \perp \Upsilon_i^0, \quad i \neq j, \quad \text{onde } \Upsilon_j^0 = (\Upsilon_{1j} \dots \Upsilon_{jp})^t = \Upsilon_j \text{ sem o termo } \Upsilon_{0j}. \quad (3).$$

$$\Upsilon_0^0 \simeq \Gamma_p \quad (4).$$

$$\Upsilon_{j \neq 0}^0 \perp \Gamma_p \quad (5).$$

A partir das relações acima, vamos mostrar que $\hat{\beta}_{MQ}$ é dominado essencialmente por Γ_p e que $\hat{\beta}_{RL}$ é dominado por vetores que são ortogonais a Γ_p . A partir deste resultado não é difícil acreditar que se β^* for paralelo a Γ_p^* , $\hat{\beta}_{MQ}$, é mais preciso que $\hat{\beta}_{RL}$, uma vez que os componentes de β que não estão envolvidos com a multicolinearidade, possuem estimativas semelhantes utilizando-se ambos os estimadores.

Suponha que somente um autovalor de X^tX e um de A^tA seja próximo de zero, $\lambda_p \simeq 0$ e $\omega_0 \simeq 0$, respectivamente*, então:

$$\hat{\beta}_{RL} = -\eta \sum_{j=1}^p a_j \Upsilon_j^0$$

onde:

$$\eta = \sqrt{\sum_{i=1}^n (Y_i^* - \bar{Y}^*)^2}$$

$$\text{e } a_j = \Upsilon_{0j} \omega_j^{-1} \left(\sum_{j=1}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1}$$

(item 2.3.3)

$$\hat{\beta}_{MQ} = -\eta \sum_{j=0}^p a_j \Upsilon_j^0$$

onde :

$$a_j = \Upsilon_{0j} \omega_j^{-1} \left(\sum_{j=0}^p \frac{\Upsilon_{0j}^2}{\omega_j} \right)^{-1} \quad (\text{ item 2.3.4 }).$$

* Convém lembrar que $\omega_0 < \dots < \omega_p$ e que $\lambda_1 > \dots > \lambda_p$

O coeficiente do termo Υ_0^0 em $\hat{\beta}_{MQ}$, é dado por ω_0^{-1} , que é muito grande pois estamos supondo $\omega_0 \simeq 0$.

Desta forma:

$\hat{\beta}_{MQ}$ é dominado por Υ_0^0 , que por sua vez é quase paralelo a Γ_p , (expressão 4), portanto $\hat{\beta}_{MQ}$ é dominado por Γ_p .

$\hat{\beta}_{RL}$ é dominado por vetores que são essencialmente ortogonais a Γ_p . (expressão 5).

Conclue-se então que se β^* for paralelo a Γ_p^* , é de se esperar que o estimador de mínimos quadrados seja mais preciso que $\hat{\beta}_{RL}$, pois nenhuma combinação linear de vetores ortogonais a Γ_p resultará numa boa estimativa, ou num bom estimador de β . O cálculo do EQM dos dois estimadores confirma a superioridade de $\hat{\beta}_{MQ}$, nesta situação:

$$EQM(\hat{\beta}_{RL}) \simeq EQM(\hat{\beta}_{CP}) = \sigma^2 \sum_{j=2}^p \lambda_j^{-1} + (\Gamma_p^t \beta)^2$$

onde $(\Gamma_p^t \beta)^2$ é o viés de $\hat{\beta}_{RL}$.

$$EQM(\hat{\beta}_{MQ}) = \sigma^2 \sum_{j=1}^p \lambda_j^{-1}$$

O viés de $\hat{\beta}_{RL}$ torna-se muito grande se β for paralelo ou quase paralelo a Γ_p , ultrapassando a redução na variancia, obtida pela omissão do termo λ_1 no denominador.

3.1.4 - ESTIMADOR CONTRAÍDO x MÍNIMOS QUADRADOS

$$(\hat{\beta}_\rho \text{ e } \hat{\beta}_\xi) \text{ x } \hat{\beta}_{MQ}$$

Os resultados apresentados a seguir complementam a descrição da utilização dos estimadores contraídos como método alternativo de estimação de β na presença de multicolinearidade.

RESULTADO - 4 -

Para todo β , existe ρ fixo, $\in (0, 1)$, tal que :

$$EQM(\hat{\beta}_\rho) < EQM(\hat{\beta}_{MQ})$$

(Mayier e Willke - 1973 - [15]).

Prova:

Vimos no item 2.4.2 que:

$$EQM(\hat{\beta}_\rho) = \rho^2 \sigma^2 \text{tr}(X^t X)^{-1} + (1 - \rho)^2 \beta^t \beta$$

A resolução da equação :

$EQM(\hat{\beta}_\rho) - EQM(\hat{\beta}_{MQ}) < 0$, leva a condição:

$$\begin{aligned} \rho^2 \sigma^2 (X^t X)^{-1} + (1 - \rho)^2 \beta^t \beta - \sigma^2 \text{tr}(X^t X)^{-1} < 0 &\iff \\ \iff \rho > \frac{\beta^t \beta - \sigma^2 \text{tr}(X^t X)^{-1}}{\beta^t \beta + \sigma^2 \text{tr}(X^t X)^{-1}} \end{aligned}$$

Como :

$$\beta^t \beta - \sigma^2 \text{tr}(X^t X)^{-1} < \beta^t \beta + \sigma^2 \text{tr}(X^t X)^{-1}$$

conclue-se que existe $\rho < 1$ tal que $EQM(\hat{\beta}_\rho) < EQM(\hat{\beta}_{MQ})$.

c.q.d.

Existe um outro resultado que merece destaque na descrição dos estimadores contrai-
dos:

RESULTADO - 5 -

Seja :

$$\hat{\beta}_\xi = [1 + \xi s^2 (\hat{\beta}_{MQ}^t \hat{\beta}_{MQ})^{-1}] \hat{\beta}_{MQ}$$

com :

$$s^2 = SQE(\hat{\beta}_{MQ})$$

Se $p \geq 3$ e $0 < \xi < 2(p-2)/(n-p+2)$ e $X^t X = I$
então $EQM(\hat{\beta}_\xi) < EQM(\hat{\beta}_{MQ})$.

Este resultado é atribuído a James-Stein em 1961 e está descrito em Sclove -1968-
[34].

O fato do resultado acima só ser aplicável quando $X^t X = I$, não nos permite justificar sua utilização na presença de multicolinearidade. Entretanto esta exigência é uma particularização de um resultado mais abrangente, envolvendo o EQMG, que será discutido no item 3.2.

A importância deste resultado reside no fato de este ser o *único estimador viesado*, com parâmetro estocástico que se pode garantir uma redução no EQM em comparação ao EQM de $\hat{\beta}_{MQ}$.

3.1.5 - ESTIMADOR INTERATIVO x MÍNIMOS QUADRADOS

$$\hat{\beta}_{IT} \text{ x } \hat{\beta}_{MQ}$$

No item 2.5.1 , apresentou-se as expressões para $\sum \hat{\beta}_{IT}$ e para o viés ,extraídas dos estudos de Gotz Trenkler -1980 - [36] , contudo não se encontra na literatura referencia ou algum resultado que descreva a comparação do EQM de $\hat{\beta}_{IT}$ com o EQM de $\hat{\beta}_{MQ}$.

Gotz Trenkler refere-se a estudos ainda não concluídos neste sentido, prometendo uma futura publicação de seus resultados.

O item 3.2 refere-se as comparações entre os estimadores viesados baseando-se no EQMG . Com este outro parâmetro de comparação é possível obter-se alguns resultados que serão comentados oportunamente.

Os resultados apresentados até agora servem como justificativa para a utilização destes estimadores como alternativas a $\hat{\beta}_{MQ}$. Entretanto novamente surge a questão :

Mas qual a melhor alternativa?

Com o intuito de respondermos esta pergunta é necessário efetuar-se comparações entre estas técnicas.

A única comparação possível ,através do EQM , é a de $\hat{\beta}_{CP}$ com $\hat{\beta}_k$, para k fixo, assunto que passaremos a discutir.

3.1.6 - COMPONENTES PRINCIPAIS x ESTIMADOR EM CRISTAS

$$\hat{\beta}_{CP} \text{ x } \hat{\beta}_k$$

Baye e Parker - 1984 - [22] , apresenta os estimadores $\hat{\beta}_k, \hat{\beta}_{CP}, \hat{\beta}_{MQ}$ como casos particulares de uma classe mais geral de estimadores , chamada classe ' $r - k$ ' , dada por :

$$b_r(k) = \Gamma_r(\Gamma_r^t X^t X \Gamma_r + k I_r)^{-1} \Gamma_r^t X^t Y$$

com :

$$r = \text{posto}(X^t X) \leq p$$

$$k \geq 0$$

Γ_r : é uma matriz $p \times r$, obtida eliminando-se $(p - r)$ colunas de Γ , que são os $(p - r)$ autovetores de $X^t X$ associados aos $(p - r)$ autovalores próximos de zero.

Esta classe satisfaz as seguintes propriedades:

$$E(b_r(k) - \beta) = (\Gamma_r(\Delta_r + k I_r)^{-1} \Delta_r \Gamma_r^t) - I_p \beta$$

$$\sum(b_r(k)) = \sigma^2 \Gamma_r(\Delta_r + k I_r)^{-1} \Delta_r (\Delta_r + k I_r)^{-1} \Gamma_r^t$$

$$EQM(b_r(k)) = \sum_{i=1}^r [(\delta_i - 1)^2 \alpha_i^2 + \sigma^2 \frac{\delta_i^2}{\lambda_i}] + \sum_{i=r+1}^p \alpha_i^2$$

com:

$$\alpha_i = \Gamma^t \beta \text{ e } \delta_i = \frac{\lambda_i}{\lambda_i + k}$$

Δ : a matriz diagonal cujos elementos são os autovalores de $X^t X$.

Verifica-se que :

$$b_r(0) = \hat{\beta}_{CP}$$

$$b_p(k) = \hat{\beta}_k$$

$$b_p(0) = \hat{\beta}_{MQ}$$

Com os estudos de Baye e Parker - 1984 - [22] , o resultado 2 do item 3.1.2 , que permite a comparação de $\hat{\beta}_k$ com $\hat{\beta}_{CP}$, tornou-se um caso particular de um resultado mais geral, que vamos apresentar a seguir.

RESULTADO - 6 -

Se $\sigma^2 > 0$, então existe $k > 0$ tal que :

$$EQM(b_r(k)) < EQM(b_r(0))$$

para todo $0 < r \leq p$.

Prova:

É suficiente mostrar que $EQM(b_r(k))$, como função de k , é decrescente na vizinhança de zero. Para tal deriva-se $EQM(b_r(k))$ e verifica-se que esta derivada é menor que zero para $k = 0$:

$$\frac{\partial(b_r(k))}{\partial k} = -2\sigma^2 \sum_{i=1}^r \lambda_i^{-2} < 0$$

Determina-se o valor de “ k ” , para o qual $EQM(b_r(k))$ atinge o mínimo, seja k_{ot} este valor ,($k_{ot} = k$ - ótimo) :

$$k_{ot} = \sigma^2 \frac{\sum_{i=1}^r \delta_i^3 \lambda_i^2}{\sum_{i=1}^r \delta_i^3 \alpha_i^2 \lambda_i^2}$$

Como o denominador da expressão para k_{ot} é uma média ponderada dos α_i , pode-se escrever:

$$k_{ot} \geq \frac{\sigma^2}{\max_i \alpha_i^2} > \frac{\sigma^2}{\alpha^t \alpha} = \frac{\sigma^2}{\beta^t \beta}$$

ou seja, k -ótimo > 0 , de onde se conclue que $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{CP})$, para algum $k > 0$,fixo.

Verificou-se que entre os estimadores estudados , de forma direta ou indireta, o aumento da precisão de um sobre o outro esta relacionado a condições sobre os parâmetros β e σ^2 , ou em alguns casos , nem se conhece a expressão para o EQM , a não ser por aproximações empíricas, como no caso de $\hat{\beta}_{RL}$.

O EQMG , já citado algumas vezes neste trabalho, permite algumas comparações que não puderam ser efetuadas através do EQM.O item 3.2, a seguir,irá redescrever as comparações já obtidas e acrescentar algumas outras baseando-se no EQMG .

3.2 - COMPARAÇÕES ENTRE OS ESTIMADORES UTILIZANDO-SE O EQMG

A argumentação original dos autores Hoerl e Kennard - 1970 - [19] , que leva a conclusão da superioridade de $\hat{\beta}_k$ sobre $\hat{\beta}_{MQ}$ na presença de multicolinearidade, baseando-se no EQM , foi duramente criticada por J. A. Nelder tendo por base o trabalho de Lindley e Smith - 1972 -(JRSS-B vol. 34 , pág. 1 a 18) , onde por métodos Bayesianos foram mais convincentes ao apresentarem a mesma conclusão.

As críticas de Nelder , motivaram a C. M. Theobald - 1974 - [37] , a estudar uma outra forma de comparação entre $\hat{\beta}_k$ e $\hat{\beta}_{MQ}$, que fortalecesse a idéia original de Hoerl e Kennard .

Theobald estudou uma medida já utilizada em 1966 no contexto de planejamento de experimentos, por Karlin e Studden :

O EQMG - erro quadrático médio generalizado :

$$EQMG(\hat{\beta}) = E(\hat{\beta} - \beta)^t T (\hat{\beta} - \beta)$$

onde T é uma matriz simétrica , não negativa definida, (matrizes nnd - ver apêndice 5) .

Embora o objetivo central de Theobald - 1974 - [37] fosse a comparação de $\hat{\beta}_k$ com $\hat{\beta}_{MQ}$, de maneira a relacionar os erros quadráticos médios envolvidos na expressão $E(\hat{\beta} - \beta)^t T (\hat{\beta} - \beta)$, com os segundos momentos: $E(\hat{\beta} - \beta)(\hat{\beta} - \beta)^t$ de cada estimador, o resultado obtido foi mais geral, possibilitando a introdução de um novo parâmetro de comparação entre os estimadores viesados.

TEOREMA - 1 - TEOREMA DE THEOBALD

Seja $\sum b_j$ a matriz de segundo momento , (var e covar) , de b_j , $j= 1,2$:

$$\sum b_j = E(b_j - \beta)(b_j - \beta)^t$$

e

$$EQMG(b_j) = E(b_j - \beta)^t T (b_j - \beta)$$

Então :

$\sum b_1 - \sum b_2$, é não negativa definida , (nnd) , $\iff EQMG(b_2) \leq EQMG(b_1)$, para todo T nnd .

Theobald - 1974 - [37] .

Prova :

É fácil ver que :

$$EQMG(b_j) = tr\{T \sum(b_j)\}$$

$$EQMG(b_1) - EQMG(b_2) = tr\{T[\sum(b_1) - \sum(b_2)]\}$$

Sendo assim , basta mostrar que toda matriz simétrica N é nnd $\iff trTN \geq 0$, para todo T nnd. Para tal é conveniente representarmos N , segundo sua decomposição em valores singulares,(apêndice - 3) :

$$N = \sum_{i=1}^p \mu_i q_i q_i^t$$

onde :

$\mu_1, \mu_2, \dots, \mu_p$ são os autovalores de N.

q_1, q_2, \dots, q_p são os respectivos autovetores.

(\implies) - Supondo que N é nnd $\implies \mu_i \geq 0 \implies tr\{TN\} = tr\{\sum_{i=1}^p \mu_i q_i T q_i^t\} \geq 0$.

(\impliedby) - Supondo que $tr\{TN\} \geq 0$ tem-se $\mu_i \geq 0 \implies N$ é nnd.

c.q.d.

É comum encontrar-se na literatura uma outra versão deste mesmo teorema, onde se substitue a hipótese de $\sum b_1 - \sum b_2$ ser “nnd” por “pd” : positiva definida ou “psd” : positiva semi definida :

$$\sum b_1 - \sum b_2 , \text{ é pd } \iff EQMG(b_2) < EQMG(b_1) \text{ para todo T pd .}$$

$$\sum b_1 - \sum b_2 , \text{ é psd } \iff EQMG(b_2) \leq EQMG(b_1) \text{ para todo T psd .}$$

Gunst e Mason - 1977 - [16] .

Alguns autores verificaram que se pode descrever o comportamento de $\sum b_1 - \sum b_2$, no sentido de ser ou não ser nnd (ou psd) , através de condições sobre os parâmetros β e σ^2 . Vamos denotar estes valores dos parâmetros através do conjunto $\Omega(\beta, \sigma^2)$.

Se $\Omega(\beta, \sigma^2) = \emptyset$, o teorema de Theobald não é aplicavel para a comparação dos estimadores em questão. Caso contrario existe determinadas condições sobre os parâmetros que permite a aplicação do teorema.

Estes resultados serão apresentados a seguir , para os estimadores em estudo e estão provados em:

- Gunst e Mason - 1976 - [18] e 1977 - [16] .
- J. Michael Price -1982 -[38] .
- G. Trenkler - 1980 - [36] .

A única exceção refere-se as comparações que envolvem os estimadores contraídos , para os quais utiliza-se outros resultados extraídos de Mayer e Willke - 1973 - [36] , que serão oportunamente enunciados .

3.2.1 - COMPONENTES PRINCIPAIS x MÍNIMOS QUADRADOS

$$\hat{\beta}_{CP} \text{ x } \hat{\beta}_{MQ}$$

RESULTADO - 1 -

Uma condição necessária e suficiente para que $\sum \hat{\beta}_{MQ} - \sum \hat{\beta}_{CP}$ seja pd é dada por :

$$\frac{\sum_{p-r}^p \lambda_j (\Gamma_j^t \beta)^2}{\sigma^2} < 1$$

onde $\lambda_p \simeq \dots \simeq \lambda_{p-r} \simeq 0$.

Podendo-se neste caso aplicar o teorema de Theobald, concluindo-se que:

$$EQMG(\hat{\beta}_{CP}) < EQMG(\hat{\beta}_{MQ})$$

(Gunst e Mason - 1976 - [18] .)

3.2.2. - ESTIMADOR EM CRISTAS x MÍNIMOS QUADRADOS

$$\hat{\beta}_k \text{ x } \hat{\beta}_{MQ}$$

Para se enunciar esta comparação é conveniente descrever o estimador em cristas da seguinte forma:

$$\hat{\beta}_k = (X^t X + \Gamma K \Gamma^t)^{-1} X^t Y$$

onde K é uma matriz $p \times p$, diagonal , cujos elementos K_i , são não negativos com pelo menos um elemento diferente de zero.

RESULTADO - 2 -

Uma condição necessária e suficiente para que $\sum \hat{\beta}_{MQ} - \sum \hat{\beta}_k$, seja psd, é dada por:

$$\frac{Z^t A^+ Z}{\sigma^2} < 1$$

onde :

$$Z = (\Delta + K)^{-1} K \Gamma^t \beta$$

$$\Delta = \text{diag}(\lambda_i)$$

e

$$A^+ = \text{INVG}(\Delta^{-1} - (\Delta + K)^{-1} \Delta)$$

com : INVG - A Inversa Generalizada, (ver apêndice 4) .

e Δ é a matriz diagonal dos autovalores de $X^t X$.

E aplicando-se o Teorema de Theobald conclue-se que:

$$EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_{MQ})$$

Uma maneira simplificada de descrevermos esta condição quando k é um escalar é dada por Gunst e Mason - 1976 - [18] :

$$0 < k < \frac{2\sigma^2}{\beta^t \beta}$$

3.2.3 - ESTIMADOR CONTRAÍDO x MÍNIMOS QUADRADOS

$$\hat{\beta}_\rho \text{ x } \hat{\beta}_{MQ}$$

O estimador $\hat{\beta}_\rho$ com $\rho \in (0, 1)$, pode ser escrito como um estimador em cristas, (J. M. Price - 1982 - [38]), utilizando-se a forma descrita no item anterior, 3.2.2 .

Para $K = \frac{(1-\alpha)}{\alpha} \Delta$ com $\alpha > 1$ tem-se :

$$\begin{aligned} \hat{\beta}_\rho &= (X^t X + \Gamma K \Gamma^t)^{-1} X^t X \hat{\beta}_{MQ} \\ &= \Gamma(\Delta + K)^{-1} \Delta \Gamma^t \hat{\beta}_{MQ} \end{aligned}$$

$$\begin{aligned} \hat{\beta}_\rho &= \Gamma\left(\Delta + \frac{(1-\alpha)\Delta}{\alpha}\right)^{-1} \Delta \Gamma^t \hat{\beta}_{MQ} \\ &= \rho \hat{\beta}_{MQ} \end{aligned}$$

onde $\rho = \Gamma\left(\Delta + \frac{(1-\alpha)\Delta}{\alpha}\right)^{-1} \Delta \Gamma^t$ com $\rho \in (0, 1)$

Desta forma os resultados obtidos para $\hat{\beta}_k$ são válidos para $\hat{\beta}_\rho$ e o resultado - 3 , descrito abaixo, nada mais é que um caso particular do resultado 2, do item 3.2.2 :

RESULTADO - 3 -

Se $p \geq 2$, então $\Omega(\beta, \sigma^2) = \emptyset$

Se $p = 1$, $\sum \hat{\beta}_{MQ} - \sum \hat{\beta}_\rho$, é psd se:

$$\frac{(1-\rho)^2(1-\rho^2)^{-1} \beta^t X^t X \beta}{\sigma^2} < 1$$

E aplicando-se o teorema Theobald tem-se:

$$EQMG(\hat{\beta}_\rho) < EQMG(\hat{\beta}_{MQ})$$

$$\hat{\beta}_\xi \text{ x } \hat{\beta}_{MQ}$$

O resultado abaixo , apresentado por Sclove - 1968 - [34] , é atribuído a James Stein - 1961.

RESULTADO - 4 -

Seja :

$$\hat{\beta}_\xi = [1 + \xi s^2 (\hat{\beta}_{MQ}^t \hat{\beta}_{MQ})^{-1}] \hat{\beta}_{MQ}$$

com $s^2 = SQE(\hat{\beta}_{MQ})$

Se $p \geq 3$ e $0 < \xi < \frac{2(p-2)}{(n-p+2)}$, então :

$$EQMG(\hat{\beta}_\xi) < EQMG(\hat{\beta}_{MQ})$$

e ainda para : $\xi = \frac{2(p-2)}{(n-p+2)}$ tem-se :

$$EQMG(\hat{\beta}_\xi) = \min_{\xi} EQMG(\hat{\beta}_\xi)$$

Outros estimadores contraídos são propostos na literatura , no entanto não são usados na prática por serem expressões complicadas e de difícil interpretação , além de não se conhecer suas distribuições , nem sequer os seus momentos .

3.2.4 - ESTIMADOR INTERATIVO x MÍNIMOS QUADRADOS

$$\hat{\beta}_{IT} \text{ x } \hat{\beta}_{MQ}$$

Para efetuar-se esta comparação é necessário escrever-se os estimadores da seguinte forma :

$$\hat{\beta}_j = C_j Y, j = 1, 2.$$

cujas matrizes de segundo momentos são dadas por :

$$\sum \hat{\beta}_j = (C_j X - I) \beta \beta^t (C_j X - I)^t + \sigma^2 C_j C_j^t$$

onde C_j é uma matriz real .

Trenkler - 1980 - [36] ,apresenta uma condição para que $\sum \hat{\beta}_1 - \sum \hat{\beta}_2$ seja pd :

RESULTADO - 5 -

Se $C_1 C_1^t - C_2 C_2^t$ for pd e :

$$\beta^t (C_2 X - I)^t (C_1 C_1^t - C_2 C_2^t)^{-1} (C_2 X - I) \beta < \sigma^2$$

então $\sum \hat{\beta}_1 - \sum \hat{\beta}_2$, é pd .

O procedimento é usar o resultado acima para $\hat{\beta}_1 = \hat{\beta}_{MQ}$ e $\hat{\beta}_2 = \hat{\beta}_{IT}$, e obter-se $\sum \hat{\beta}_{MQ} - \sum \hat{\beta}_{IT}$ pd e então utilizar-se o teorema de Theobald para a verificação final de $EQMG(\hat{\beta}_{IT}) < EQMG(\hat{\beta}_{MQ})$.

Para tal usaremos a decomposição em valores singulares de $X^t X$, apresentada no item 1.3 :

$$X = H \Delta^{1/2} \Gamma$$

onde :

- H : é a matriz cujas colunas são os autovetores normalizados de XX^t .
- Δ : é a matriz diagonal cujos elementos são os autovalores de $X^t X$
- Γ : é a matriz cujas colunas são os autovetores normalizados de $X^t X$.

$\hat{\beta}_{MQ}$ e $\hat{\beta}_{IT}$, podem ser escritos como :

$$\hat{\beta}_{MQ} = \Gamma \Delta^{-1} \Delta^{1/2} H^t Y \rightarrow C_{MQ} = \Gamma \Delta^{-1} \Delta^{1/2} H^t$$

$$\hat{\beta}_{IT} = \Gamma [\Delta_{m,\alpha}; 0] H^t Y \rightarrow C_{IT} = \Gamma [\Delta_{m,\alpha}; 0] H^t$$

onde :

$$\Delta_{m,\alpha} : \text{ é a matriz diagonal - } \text{diag} \left\{ \frac{1 - (1 - \alpha \lambda_j)^{m+1}}{\sqrt{\lambda_j}} \right\}_{j=1}^p$$

Desta forma mostra-se que :

$$(C_{\hat{\beta}_{MQ}} C_{\hat{\beta}_{MQ}}^t - C_{\hat{\beta}_{IT}} C_{\hat{\beta}_{IT}}^t) = \Gamma (\Delta^{-1} - \Delta_{m,\alpha}^2) \Gamma$$

é pd e utilizando-se o resultado - 5 - , tem-se que $\sum \hat{\beta}_{MQ} - \sum \hat{\beta}_{IT}$, é pd se :

$$\beta^t \Gamma \Delta_{m,\alpha}^* (\Delta^{-1} - \Delta_{m,\alpha}^2)^{-1} \Delta_{m,\alpha}^* \Gamma^t \beta < \sigma^2$$

com :

$$\Delta_{m,\alpha}^* = \text{diag} \{ (1 - \alpha \lambda_j)^{m+1} \}$$

ou podemos obter a expressão :

$$\beta^t \Gamma \text{diag} \left\{ \frac{\lambda_j \lambda_j(m, \alpha)}{\alpha - \lambda_j(M, \alpha)} \right\}_{j=1}^p \Gamma^t \beta < \sigma^2$$

com $\lambda_j(m, \alpha) = (1 - \alpha \lambda_j)^{m+1}$.

Sob esta condição, aplicando-se o Teorema de Theobald, conclue-se que:

$$EQMG(\hat{\beta}_{IT}) < EQMG(\hat{\beta}_{MQ})$$

É interessante notar ainda que :

$$\lim_{m \rightarrow \infty} \frac{\lambda_j \lambda_j(m, \alpha)}{\alpha - \lambda_j(m, \alpha)} \rightarrow 0$$

Isto significa que :

$$\exists m_0 \forall m \geq m_0 \rightarrow \sum_{\hat{\beta}_{MQ}} - \sum_{\hat{\beta}_{IT}} , \text{ é pd .}$$

Este valor de m_0 depende de β e σ^2 , que são desconhecidos .

Supondo que :

$$\beta^t \beta < k \sigma^2$$

para $k > 0$,então m_0 , é o menor valor que satisfaz:

$$\lambda_{min}(m + 1, \alpha) = (1 - \alpha \lambda_{min})^{m+1} < (k \lambda_{max})^{-1}$$

O resultado - 5 - também é utilizado para mostrar que não se obtém conclusão na comparação de $\hat{\beta}_{IT}$ com $\hat{\beta}_\rho$ nem com $\hat{\beta}_k$.Em ambos os casos :

- para "m" fixo , determina-se valores dos parâmetros, k e ρ , de maneira a obter-se tanto $EQMG(\hat{\beta}_{IT}) < EQMG(\hat{\beta}_k)$ como $EQMG(\hat{\beta}_{IT}) < EQMG(\hat{\beta}_\rho)$, para valores distintos de k e ρ .

- No entanto se primeiro fixarmos k ou ρ ; obter-se-á " m "tal que a situação se inverte , ou seja , teremos $EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_{IT})$, ou $EQMG(\hat{\beta}_\rho) < EQMG(\hat{\beta}_{IT})$, para valores distintos de " m " .

Com relação a comparação de $\hat{\beta}_{CP}$ com $\hat{\beta}_{IT}$ segundo este critério , tem-se :

$C_{\hat{\beta}_{IT}} C_{\hat{\beta}_{IT}}^t - C_{\hat{\beta}_{CP}} C_{\hat{\beta}_{CP}}^t$, não pd para quaisquer valores de β e σ^2 . Além disto, não se conhece na literatura uma condição que garanta a verificação das hipóteses do Teorema de Theobald.

J. M. Price - 1982 - [38] ,também apresenta condições sobre os parâmetros que possibilitam a comparação ,entre si, de dois estimadores em cristas, dois estimadores contraídos, ou dois obtidos pelo método das componentes principais.

Os resultados obtidos que dizem respeito a comparação de $\hat{\beta}_{k_1}$ com $\hat{\beta}_{k_2}$, possibilitam a comparação de $\hat{\beta}_\rho$ com $\hat{\beta}_k$, uma vez que $\hat{\beta}_\rho$ pode ser escrito como um estimador em cristas ,(item 3.2.3) .

3.2.5 - ESTIMADOR CONTRAÍDO x ESTIMADOR EM CRISTAS

$$\hat{\beta}_\rho \text{ x } \hat{\beta}_k$$

Da mesma forma que os resultados obtidos para a comparação de $\hat{\beta}_{IT}$ com $\hat{\beta}_k$, pode-se determinar valores dos parâmetros para os quais tem-se tanto $EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_\rho)$ como o inverso . O resultado - 6 - apresenta estas duas situações .

RESULTADO - 6 -

Seja :

$\hat{\beta}_\rho = \rho \hat{\beta}_{MQ}$ com ρ fixo e $\hat{\beta}_k = (X^t X + \Gamma K \Gamma^t)^{-1} X^t Y$ com K uma matriz diagonal de elementos K_i , não todos nulos , então :

a - Existem valores de β e σ^2 para os quais $\sum_{\hat{\beta}_\rho} - \sum_{\hat{\beta}_k}$,é psd se :

$$\left(\frac{1-\rho}{\rho} \right) \lambda_i > K_i$$

para no MÁXIMO um valor de “ i ” .

b - Existem valores de β e σ^2 para os quais $\sum_{\hat{\beta}_k} - \sum_{\hat{\beta}_\rho}$ é psd se :

$$\left(\frac{1-\rho}{\rho} \right) \lambda_i < K_i$$

para no MÁXIMO um valor de “ i ” .

Utilizando-se o Teorema de Theobald em “ a” , conclue-se que

$$EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_\rho)$$

e em “ b” , conclue-se que :

$$EQMG(\hat{\beta}_p) < EQMG\hat{\beta}_k)$$

A contribuição maior apresentada por J. M. Price - 1982 - [38] , esta no fato de explorar com mais detalhamento a comparação de $\hat{\beta}_{CP}$ com $\hat{\beta}_k$, cujo resultado esta apresentado a seguir.

3.2.6 - ESTIMADOR EM CRISTAS x COMPONENTES PRINCIPAIS

$$\hat{\beta}_k \text{ x } \hat{\beta}_{CP}$$

É possível comparar-se estes dois estimadores, desde que se retire uma só componente no cálculo de $\hat{\beta}_{CP}$, ou se retire $(p - 1)$, de um total de p .

RESULTADO - 7 -

a - $\Omega(\hat{\beta}_{CP}, \hat{\beta}_k) \neq 0$, se $\hat{\beta}_{CP}$ for calculado retirando-se *somente 1 componente* de um total de p . Neste caso , existem valores de β e de σ^2 de forma que :

$\sum_{\hat{\beta}_{CP}} - \sum_{\hat{\beta}_k}$ seja pd , concluindo-se através do Teorema de Theobald que :

$$EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_{CP})$$

b - $\Omega(\hat{\beta}_{CP}, \hat{\beta}_k) \neq 0$, se $\hat{\beta}_{CP}$ for calculado retirando-se $(p - 1)$ *componentes* de um total de p . Neste caso , existem valores de β e de σ^2 de forma que :

$\sum_{\hat{\beta}_k} - \sum_{\hat{\beta}_{CP}}$ seja pd , concluindo-se através do Teorema de Theobald que :

$$EQMG(\hat{\beta}_{CP}) < EQMG(\hat{\beta}_k)$$

Estes resultados contradizem o comentário de Gunst e Mason - 1976 - [18] , onde estes autores afirmam que nada se pode dizer com relação a esta comparação via o EQMG.

Outros trabalhos com o intuito de elaborar comparações entre estimadores viesados foram realizados, alguns utilizando simulações e outros baseando-se na distância euclidiana entre os estimadores. D. Trenkler e G. Trenkler - 1984 - [24] , ressaltam que na presença de multicolinearidade ,para uma escolha adequada de “ m” , $\hat{\beta}_{IT}$, tende a ser próximo de $\hat{\beta}_{CP}$, do ponto de vista da distância euclidiana.O mesmo ocorre com $\hat{\beta}_k$, para uma escolha adequada de k .Isto significa que estes estimadores são bons substitutos para $\hat{\beta}_{CP}$. O mesmo não ocorre para $\hat{\beta}_\rho$, o que faz com que a regressão em cristas não seja considerada simplesmente como um processo de contrair o estimador $\hat{\beta}_{MQ}$, como afirmam alguns autores.

3.3 - RESUMO GERAL DAS COMPARAÇÕES OBTIDAS E ALGUNS COMENTÁRIOS.

Tem-se 15 comparações possíveis entre os 6 estimadores estudados:

$$\hat{\beta}_{MQ}, \hat{\beta}_{CP}, \hat{\beta}_{RL}, \hat{\beta}_{IT} \text{ e } \hat{\beta}_{\rho}$$

onde $\hat{\beta}_{\rho}$ representa a classe dos estimadores contraídos: $\hat{\beta}_{\rho}, \hat{\beta}_{\delta}, \hat{\beta}_{\xi}$.

Entre as comparações discutidas nos itens 3.1 e 3.2 obteve-se:

7 resultados conclusivos em relação a escolha do melhor estimador foram obtidos através do EQMG, 5 destes resultados foram confirmados utilizando-se o EQM e os outros 2, referentes a comparação de $\hat{\beta}_{IT}$ x $\hat{\beta}_{MQ}$ e $\hat{\beta}_{\rho}$ x $\hat{\beta}_k$ só podem ser comparados através do EQMG.

2 outras comparações não são conclusivas, ou seja não se pode determinar qual o melhor estimador entre $\hat{\beta}_{IT}$ x $\hat{\beta}_{\rho}$ e $\hat{\beta}_{IT}$ x $\hat{\beta}_k$.

5 comparações não são possíveis devido a ausência de propriedades estatísticas, são as referentes a $\hat{\beta}_{RL}$ com $\hat{\beta}_{CP}, \hat{\beta}_{\rho}, \hat{\beta}_k$ e $\hat{\beta}_{IT}$ e $\hat{\beta}_{IT}$ com $\hat{\beta}_{CP}$.

1 comparação não é comentada na literatura: $\hat{\beta}_{CP}$ com $\hat{\beta}_{\rho}$. No entanto, vimos no item 3.2.3, que $\hat{\beta}_{\rho}$ pode ser escrito como um estimador em cristas, sendo então válido o resultado - 7 do item 3.2.6 onde se compara $\hat{\beta}_k$ com $\hat{\beta}_{CP}$.

O quadro a seguir resume os resultados descritos.

RESUMO GERAL DAS COMPARAÇÕES.

Métodos	Melhor Estimador	Condição para < EQM	Condição para < EQMG
MQ x CP	$\hat{\beta}_{CP}$	$\sum_{j=(p-r)}^p \frac{1}{\lambda_j} > \frac{1}{\sigma^2}(\beta^t \beta)$ (item 3.1.1)	$\sum_{p-r}^p \frac{\lambda_j (\Gamma_j^t \beta)^2}{\sigma^2} < 1$ (item 3.2.1)
MQ x k	$\hat{\beta}_k$	$k < \frac{\sigma^2}{\max_j (\Gamma_j^t \beta)^2}$ (item 3.1.2)	$0 < k < \frac{2\sigma^2}{\beta^t \beta}$ (item 3.2.2)
MQ x RL	$\hat{\beta}_{MQ}$	$\beta^* = c\Gamma^*$ (item 3.1.3)	não é possível
MQ x ρ, ξ	$\hat{\beta}_\rho$	$\rho > \frac{\beta^t \beta - \sigma^2 \text{tr}(X^t X)^{-1}}{\beta^t \beta + \sigma^2 \text{tr}(X^t X)^{-1}}$ (item 3.1.4)	$p = 1$ e $\frac{(1-\rho)^2(1-\rho^2)^{-1} \beta^t X^t X \beta}{\sigma^2} < 1$ (item 3.2.3)
	$\hat{\beta}_\xi$	$p \geq 3$ e $0 < \xi < \frac{2(p-2)}{(n-p+2)}$ e $X^t X = I$ (item 3.1.4)	$p \geq 3$ e $0 < \xi < \frac{2(p-2)}{(n-p+2)}$ (item 3.2.3)
MQ x IT	$\hat{\beta}_{IT}$	não é possível	$C_1 C_1^t - C_2 C_2^t$ for pd e $\beta^t (C_2 X - I)^t (C_1 C_1^t - C_2 C_2^t)^{-1} (C_2 X - I) \beta < \sigma^2$ (item 3.2.4)
CP x k	$\hat{\beta}_k$	$k \geq \frac{\sigma^2}{\beta^t \beta}$ (item 3.1.6)	$\Omega(\beta, \sigma^2) \neq \emptyset$ se em $\hat{\beta}_{CP}$ retira-se uma compon.(item 3.2.6)
	$\hat{\beta}_{CP}$	não é possível	$\Omega(\beta, \sigma^2) \neq \emptyset$ se em $\hat{\beta}_{CP}$ retira-se $(p-1)$ compon.(item 3.2.6)
ρ e k	$\hat{\beta}_k$	não é possível	$(\frac{1-\rho}{\rho}) \lambda_i > k$ para no max. 1 valor de i .
	$\hat{\beta}_\rho$	não é possível	$(\frac{1-\rho}{\rho}) \lambda_i < k$ para no max. 1 valor de i . (item 3.2.5)

CAPÍTULO - 4 -

ANÁLISE DE DADOS

4.1 - INTRODUÇÃO

O conjunto de dados que será analisado, foi extraído de Bussab - 1984 - [47] e refere-se ao estudo da distribuição geográfica dos médicos no estado de São Paulo, dividido em subregiões administrativas.

Os dados foram obtidos do cadastro da DATEC - Distribuidora , Gráfica e Mala Direta Ltda. , que é uma empresa especializada em fornecer para as grandes indústrias farmacêuticas, listas atualizadas de nomes e endereços de médicos, para visitas de propagandistas. Segundo a própria empresa, cre-se que cerca de 10 por cento dos médicos em atividade, não fazem parte de seu cadastro.

Os dados referem-se a situação em 1980 . Algumas entre as variáveis pesquisadas foram transformadas, descrevemos a seguir cada variável e o tratamento estatístico inicial realizado por Bussab - 1984 - [47] , que originou os dados que vamos utilizar:

- **LOGPOP** : Log do número de habitantes por subregião administrativa. Fonte: Seade .

- **POPURB** : índice de urbanização, que é o quociente da população urbana pela população total .Fonte : Censo - 1970 .

- **LEITOS** : Número de leitos por 1000 habitantes, utilizado como indicador da infraestrutura hospitalar. Fonte : Coordenadoria de Assistência Hospitalar - Secretaria da Saúde.

- **MAIS50** : porcentagem da população acima de 50 anos, que é o quociente entre a população acima de 50 anos e o total de habitantes. Fonte : Censo-1970 .

- **INDESC** : Índice de escolaridade , medido através do quociente entre os concluintes do segundo grau e os concluintes do primeiro grau. Fonte : Secretaria da Educação - 1979 .

- **CONDHAB** : Índice de habitabilidade ou condições de habitação, foi medido baseando-se nas seguintes variáveis:

- . Número de prédios com iluminação elétrica.
- . Número de prédios com ligação de água.

. Número de prédios com ligação de esgoto.

Transformou-se estas variáveis em proporção em relação ao número de prédios existentes na sub-região. Construiu-se o índice, utilizando-se análise fatorial, adotando-se o primeiro fator como índice, cuja explicação corresponde a 87 por cento da variabilidade total e cada comunalidade esta acima de 86 por cento. Os escores fatoriais correspondem aos índices de habitabilidade ou condições de habitação , em cada subregião, sendo os valores positivos os indicadores de boas condições. Fonte: Censo-1970 .

- **BENSDUR** : Índice de bens duráveis, ou de desenvolvimento da região. Utilizou-se também análise fatorial , com um único fator, a partir da proporção de domicílios duráveis em relação ao total de domicílios e o número de automóveis por 1000 habitantes. Os escores positivos indicam alta proporção de bens duráveis. Fonte: Censo-1970 e boletim do DNER - 1978 .

- **CONDECO** : Índice de níveis de renda ou condições econômicas. Utilizou-se medidas indiretas para avaliar-se as condições econômicas, como a arrecadação do ICM e total de depósitos bancários per capita, formando-se também um índice correspondente ao primeiro score fatorial, com uma explicação de 75 por cento da variação total. Fonte : Seade - 1978 .

- **MÉDICOS** : Variável dependente , que representa o número de médicos por 1000 habitantes.

Devido a grande variabilidade de algumas medidas descritas acima e para que as suposições do modelo fossem satisfeitas, algumas transformações matemáticas foram necessárias. Os dados em Bussab - 1984 - [47] , apresentados na tabela - 1 - , apêndice - 6 - , já encontram-se transformados segundo a descrição a seguir:

- X_1 ou LOGPOP : Log da população .
- X_2 ou POPURB : Proporção da população urbana.
- X_3 ou LEITOS : Leitos por 1000 habitantes , dividido por 4.5.
- X_4 ou MAIS 50 : $\text{Arcoseno} \sqrt{+50 \text{anos/pop} - 1970}$.
- X_5 ou INDESC : $\text{Arcoseno} \sqrt{2.\text{grau}/1.\text{grau}}$.
- X_6 ou CONDHAB : Escore fatorial para as condições habitacionais.
- X_7 ou BENSDUR : Escore fatorial para bens duráveis.
- X_8 ou CONDECO : Escore fatorial para condições econômicas.
- Y ou MÉDICOS : Número de médicos por 1000 habitantes .

4.2 - DETECÇÃO E MEDIDAS DO GRAU DE MULTICOLINEARIDADE

Conforme estudamos no capítulo um , existem alguns indicadores simples da presença de multicolinearidade, entre eles os elementos de $X^t X$, que são as correlações duas a duas das variáveis independentes ,apresentadas na tabela - 2 - .

TABELA - 2 - A MATRIZ $X^t X$

	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8
X_1								
X_2	0,526							
X_3	-0,021	0,209						
X_4	0,109	0,477	0,408					
X_5	0,151	0,460	0,470	0,366				
X_6	0,370	0,636	0,462	0,707	0,574			
X_7	0,350	0,713	0,373	0,619	0,609	0,810		
X_8	0,536	0,715	0,321	0,482	0,397	0,664	0,778	

Obteve-se 5 correlações acima de 70 por cento, envolvendo as variáveis:

- POPURB com BENSDUR : X_2 e $X_7 = 0,713$.
- POPURB com CONDECO : X_2 e $X_8 = 0,715$.
- MAIS50 com CONDHAB : X_4 e $X_6 = 0,707$.
- BENSDUR com CONDECO : X_7 e $X_8 = 0,778$.
- CONDHAB com BENSDUR : X_6 e $X_7 = 0,810$.

Estas variáveis muito provavelmente estão envolvidas na descrição da multicolinearidade, contudo , pode ser que 3 ou mais variáveis estejam relacionadas linearmente, sem que a correlação duas a duas seja próxima de 1.Neste caso , a detecção é feita por outros meios, por exemplo, através do VIF .

O V. I. F. - FATOR DE INFLAÇÃO DA VARIANCIA

Esta é uma medida muito usada tanto na detecção como na avaliação do grau de multicolinearidade .

$$VIF(X_j) = \frac{1}{1 - R_j^2}$$

onde R_j^2 : é o coeficiente de determinação obtido na regressão de X_j sobre os demais X_k .

A tabela -3-,mostra os resultados obtidos para cada uma das variáveis X_1, X_2, \dots, X_8 .

TABELA - 3 - V.I.F.(X_j) e R_j^2

X_j	V.I.F.(X_j)	R_j^2
X_1 - LOGPOP	1,8037	0,4456
X_2 - POPURB	2,7284	0,6335
X_3 - LEITOS	1,5670	0,3618
X_4 - MAIS50	2,2738	0,5602
X_5 - INDESC	1,9712	0,4927
X_6 - CONDHAB	4,2671	0,7656
X_7 - BENSUR	5,2555	0,8097
X_8 - CONDECO	3,5920	0,7216

Não existe na literatura uma regra que estabeleça um número a partir do qual o V.I.F. seja considerado grande .

É interessante portanto calcular-se R_j^2 a partir de cada V.I.F.(X_j) para se obter um melhor meio de avaliação .

Entre os valores de R_j^2 observados, obteve-se tres acima de 70 por cento, para CONDHAB , BENSUR e CONDECO, reafirmando o envolvimento destas variáveis na descrição da multicolinearidade.

O maior valor de R_j^2 observado foi $R_7^2 = 0,8097$, concluindo-se que deva existir uma combinação linear das outras variáveis ou parte delas , que resultem em X_7 : BENSUR .

Uma outra maneira eficaz de detectar a presença de multicolinearidade é a análise dos autovalores de $X^t X$, cujos resultados estão na tabela - 4 - .

TABELA - 4 - AUTOVALORES DE $X^t X$

λ_1	=	4,4863
λ_2	=	1,2884
λ_3	=	0,6789
λ_4	=	0,5512
λ_5	=	0,3782
λ_6	=	0,2941
λ_7	=	0,1998
λ_8	=	0,1231

Autovalores próximos de zero são indicadores de forte relação linear entre os regressores. Usualmente considera-se $\lambda_p < 0.1$, para se caracterizar uma multicolinearidade muito forte. Nesta análise , $\lambda_8 = 0.1231$, indicando a existencia de multicolinearidade num grau não muito elevado .

A tabela - 5 - mostra os autovetores associados a autovalores menores que 1 . A análise desta tabela nos levará a identificar quais as variáveis envolvidas na descrição da multicolinearidade.

TABELA - 5 - AUTOVETORES DE X^tX ASSOCIADOS A AUTOVALORES < 1

X_j	Γ_8 $\lambda_8 = 0,1231$	Γ_7 $\lambda_7 = 0,1998$	Γ_6 $\lambda_6 = 0,2941$	Γ_5 $\lambda_5 = 0,3782$	Γ_4 $\lambda_4 = 0,5512$	Γ_3 $\lambda_3 = 0,6789$
X_1	-0,1990	-0,1183	-0,0571	-0,5173	+0,3736	+0,2773
X_2	+0,0447	+0,2314	+0,7955	+0,2301	-0,1787	+0,0000
X_3	-0,1356	+0,0806	+0,1797	+0,1130	+0,6762	+0,3344
X_4	-0,0625	-0,4574	+0,1815	-0,3703	+0,0521	-0,6587
X_5	+0,1412	-0,3280	-0,0104	-0,2542	-0,5264	+0,5885
X_6	+0,4364	+0,6474	-0,3030	-0,2938	-0,0077	+0,5885
X_7	-0,7364	+0,1534	-0,3412	+0,2647	-0,2374	-0,0760
X_8	+0,4283	-0,4082	-0,3007	+0,5555	+0,1871	-0,0264

Γ_8 esta associado ao menor autovalor ,se considerarmos $\lambda_8 \simeq 0$,podemos escrever:

$$X\Gamma_8 \simeq 0$$

(item 1.4.5) ,ou seja,

$$-0,1990X_1 + 0,0447X_2 - 0,1356X_3 - 0,0625X_4 + \\ +0,1412X_5 + 0,4364X_6 - 0,7364X_7 + 0,4283X_8 \simeq 0$$

Os 3 maiores coeficientes ,da equação acima , são aqueles relacionados a :

X_6 : CONDHAB .

X_7 : BENS DUR .

X_8 : CONDECO .

Isto é uma indicação que exista uma mesma combinação linear que relacione estas 3 variáveis.Convém ainda lembrar que para estas 3 encontrou-se os maiores valores do V.I.F.

Através da tabela - 5 - , nota-se que há uma tendencia de sempre os coeficientes das 3 variáveis citadas acima serem os maiores.Também merecem destaque:

X_2 : POPURB ,cujo coeficiente obtido através de Γ_6 ,é 0,7955 .

X_3 : LEITOS ,cujo coeficiente obtido através de Γ_4 ,é 0,6772 .

X_4 : MAIS50 ,cujo coeficiente obtido através de Γ_3 ,é -0,6587 .

É fácil ver que a existencia de autovalores próximos de zero é a maior evidencia da presença de multicolinearidade.Como vimos no item 1.4.6 o número da condição espectral

$$k = \sqrt{\frac{\max_i \lambda_i}{\min_i \lambda_i}} = 6,0369$$

Verificou-se portanto que o maior autovalor é aproximadamente 36 vezes o menor autovalor, ($k^2 = 36,44$).

O determinante de $X^t X$, que pode ser calculado através do produto dos autovalores, é pequeno se tivermos algum $\lambda_j \simeq 0$. Neste caso:

$$|X^t X| = 0,0059$$

Além de um meio de detecção, o $|X^t X|$ também mede o grau de multicolinearidade, pois $\sqrt{|X^t X|}$, é a razão entre os volumes das regiões de confiança para β no modelo ortogonal e no modelo em que $X^t X \neq I$.

$$\frac{\text{volume}(\text{modelo} - X^t X \neq I)}{\text{volume}(\text{modelo} - X^t X = I)} = \frac{1}{\sqrt{0,0059}} = 13,02$$

Conclui-se que a região de confiança conjunta para β , no modelo original, portanto com multicolinearidade, tem volume 13 vezes maior que o volume da região obtida para o modelo ortogonal.

Analogamente a razão entre o tamanho do intervalo de confiança para β_j isoladamente, nos 2 modelos é medida através de $\sqrt{VIF_j}$. Neste caso o maior valor relaciona-se a β_7 :

$$\sqrt{VIF_7} = \sqrt{5,2555} = 2,30$$

ou seja, o intervalo de confiança de β_7 no modelo com multicolinearidade é 2,30 vezes maior que o mesmo intervalo no modelo ortogonal.

Para finalizar apresentaremos a matriz Π , descrita no item 1.4.7, onde verificaremos a influencia de cada λ_i , no cálculo da variancia de β_j . O elemento π_{ij} representa a proporção da variancia de $\hat{\beta}_{MQ(j)}$, que é atribuída a influencia de λ_i .

TABELA - 6 - A MATRIZ Π

	1	2	3	4	5	6	7	8
1	0,0066	0,0121	0,0084	0,0113	0,0115	0,0094	0,0079	0,0098
2	0,1747	0,0224	0,1495	0,0222	0,0324	0,0015	0,0000	0,0125
3	0,0628	0,0000	0,1051	0,2811	0,2588	0,0085	0,0016	0,0003
4	0,1403	0,0212	0,5294	0,0022	0,2550	0,0000	0,0195	0,0177
5	0,3923	0,0513	0,0215	0,1595	0,0867	0,0535	0,0352	0,2272
6	0,0061	0,7887	0,0700	0,0493	0,0002	0,0732	0,0753	0,0856
7	0,0388	0,0983	0,0208	0,4605	0,2732	0,4915	0,0224	0,2322
8	0,1783	0,0059	0,0953	0,0140	0,0822	0,3625	0,8381	0,4148

- Os elementos da j-ésima coluna de π , indicam a proporção de variancia de β_j , que é atribuída a cada autovalor λ_i , $i=1,2,\dots,p$.

Logo :

$$\sum_{i=1}^{p=8} \pi_{ij} = 1$$

- Por outro lado os elementos da i -ésima linha de π , indicam a contribuição do autovalor λ_i em cada variancia de cada um dos estimadores $\hat{\beta}_j$, $j=1,2,\dots,8$.

Por exemplo:

A linha 8 de π , π_{8j} , $j=1,\dots,p$, apresenta a influencia de λ_8 , o menor autovalor de $X^t X$, nas variancias de $\hat{\beta}_j$ $j=1,2,\dots,8$. O maior valor apontado é :

$$\pi_{87} = 0,8381$$

interpretando-se que 83,81 por cento da variancia de $\hat{\beta}_7$ deve-se a influencia de λ_8 , na parcela $\frac{\Gamma_8^t \Gamma_8}{\lambda_8}$.

Outros valores merecem destaque :

78,87 por cento da variancia de $\hat{\beta}_2$, deve-se a influencia de λ_6 , pois $\pi_{62} = 0,7887$.

52,94 por cento da variancia de $\hat{\beta}_3$, deve-se a influencia de λ_4 , pois $\pi_{43} = 0,5294$.

49,15 por cento da variancia de $\hat{\beta}_6$, deve-se a influencia de λ_7 , pois $\pi_{76} = 0,4915$.

CONCLUSÃO SOBRE A CARACTERIZAÇÃO DA MULTICOLINEARIDADE

NESTE CONJUNTO DE DADOS

As variáveis X_6 - CONDHAB, X_7 - BENSUR e X_8 - CONDECO são as mais importantes na descrição da multicolinearidade. Provavelmente, existe uma única relação linear que envolve estas três variáveis, que pode ser aproximada pelos maiores coeficientes de Γ_8 :

$$0,4364X_6 - 0,7364X_7 + 0,4283X_8 \simeq 0$$

Entretanto não se pode negligenciar outras influências menores como a possível relação existente entre as variáveis:

- MAIS50 e CONDHAB, cujo coeficiente de correlação linear obtido foi de 0,707.
- POPURB e CONDECO, com coeficiente de correlação linear igual a 0,715.

De maneira geral, verificou-se que existe a multicolinearidade neste conjunto de dados num grau não muito elevado, porém o suficiente para afetar a estimação de β . Estes efeitos ainda aparecerão ao se comparar os resultados obtidos pelo método de mínimos quadrados com aqueles obtidos pelos demais métodos estudados no capítulo 2, assunto que passaremos a discutir.

4.3 - COMPARAÇÕES ENTRE AS ESTIMATIVAS OBTIDAS

4.3.1 - DESCRIÇÃO DAS ESTIMATIVAS E COMPARAÇÕES PRELIMINARES

Uma das maneiras imediatas de se verificar o desempenho dos métodos estudados, é compará-los com o resultado obtido através da técnica de mínimos quadrados que sabemos super-estimar os parâmetros e suas variancias, quando $X^t X$ é quase singular.

A tabela - 7 - apresenta os valores obtidos na estimação de β e $var(\beta_i)$, $i=1, \dots, 8$.

TABELA - 7 - ESTIMATIVAS DE β E RESPECTIVAS VARIANCIAS

β	MIN QUAD		COMP PRINC CRISTAS				R LATENT		CONTR		INTERA
	$\hat{\beta}$	$var \hat{\beta}_i$	$\hat{\beta}$	$var \hat{\beta}_i$	$\hat{\beta}$	$var \hat{\beta}_i$	$\hat{\beta}$	$var \hat{\beta}_i$	$\hat{\beta}$	$var \hat{\beta}_i$	$\hat{\beta}$
$\hat{\beta}_1$	0,6991	0,0727	0,4169	0,0569	0,4753	0,0314	0,2109	0,0597	0,6439	0,5943	0,4430
$\hat{\beta}_2$	0,3414	0,1100	0,5420	0,0985	0,3293	0,0374	0,3335	0,1093	0,3143	0,2901	0,7051
$\hat{\beta}_3$	0,7849	0,0631	0,7007	0,0558	0,5623	0,0313	0,3790	0,0571	0,7230	0,6674	0,0284
$\hat{\beta}_4$	0,1269	0,0916	-,2426	0,0481	0,0324	0,0343	-,0595	0,0903	0,1179	0,1088	0,1672
$\hat{\beta}_5$	0,4824	0,0794	0,4072	0,0512	0,4203	0,0333	0,5176	0,0729	0,4435	0,4094	-,3240
$\hat{\beta}_6$	-,8390	0,1720	0,0392	0,0252	-,2487	0,0374	-1,8242	0,1097	-,7728	-,7134	-,1423
$\hat{\beta}_7$	0,7853	0,2118	0,1358	0,0296	0,4036	0,0359	-,4584	0,0343	0,7228	0,6673	-,1676
$\hat{\beta}_8$	0,2185	0,1448	0,3838	0,0512	0,3227	0,0365	0,7313	0,0848	0,2012	0,1857	0,2605

PROCEDIMENTOS UTILIZADOS :

$$\hat{\beta}_{CP}$$

Para se calcular $\hat{\beta}_{CP}$, retirou-se 2 componentes, permanecendo 95,96 por cento da variação total.

$$\hat{\beta}_k$$

Para se calcular $\hat{\beta}_k$, determinou-se $k = 0,22$, através do gráfico da crista. Utilizamos na construção do gráfico, valores de k a partir de 0,01 até 0,35 com intervalos de 0,01.

$\hat{\beta}_{RL}$

Retirou-se uma componente, para procedermos esta eliminação, baseou-se em :

$$|Y_{00}| < 0,1$$

e

$$\omega_0 < 0,3$$

Com isto permaneceu 98,82 por cento da variação total. Utilizou-se também, a aproximação:

$$var(\hat{\beta}_{RL}) \simeq var(\hat{\beta}_{CP})$$

sendo que retirou-se somente uma componente para o cálculo de $var(\hat{\beta}_{CP})$.

$\hat{\beta}_\xi$ e $\hat{\beta}_\rho$

a - Determinou-se ξ em $\hat{\beta}_\xi$ de maneira a satisfazer as hipóteses do resultado 4, da seção 3.2 do cap. 3, garantindo desta forma que :

$$EQMG(\hat{\beta}_\xi) < EQMG(\hat{\beta}_{MQ})$$

Calculou-se ξ como sendo o ponto médio do intervalo $(0, 2(p-2)/(n-p+2))$ obtendo-se:

$$\xi = 0,143 \text{ para } n=48 \text{ e } p=8.$$

$$\hat{\beta}_\xi = (1 - \xi s^2 (\hat{\beta}_{MQ}^t \hat{\beta}_{MQ}))^{-1} \hat{\beta}_{MQ} = 0,9208 \hat{\beta}_{MQ}$$

com $s^2 = SQE_{MQ} = 1,5726$, obtido no modelo com intercepto e X padronizada.

b - Determinou-se ρ em $\hat{\beta}_\rho$, supondo que uma redução de 15 por cento em $\hat{\beta}_{MQ}$ fosse suficiente para eliminar os efeitos da multicolinearidade. Este valor foi determinado subjetivamente, baseando-se no grau não muito elevado de multicolinearidade. Desta forma:

$$\hat{\beta}_\rho = 0,85 \hat{\beta}_{MQ}$$

$\hat{\beta}_{IT}$

Para o cálculo de $\hat{\beta}_{IT}$ utilizou-se $\alpha = 0.0625$, que é o ponto médio do intervalo $(0, \frac{1}{tr(X^t X)})$ e $m=20$. A escolha de m baseou-se num procedimento análogo ao da análise do traço da crista, isto é, calculou-se $\hat{\beta}_{IT}$ para diversos valores de m , e escolheu-se aquele em que os elementos de $\hat{\beta}_{IT}$ começaram a estabilizar-se.

As inversões de sinais obtidas entre algumas estimativas de um mesmo $\beta_{(i)}$, também são um sintoma da presença de multicolinearidade, como vimos no item 1.3.2.

Verifica-se também uma redução significativa nos diversos valores de $\hat{\beta}$ e $var \hat{\beta}$ se comparados aos valores obtidos pelo método de mínimos quadrados. Expressou-se esta redução em termos percentuais da seguinte maneira :

$$\frac{|\hat{\beta}_{MQ}| - |\hat{\beta}_i|}{|\hat{\beta}_{MQ}|} \times 100$$

e

$$\frac{var \hat{\beta}_{MQ} - var \hat{\beta}_i}{var \hat{\beta}_{MQ}} \times 100$$

cujos valores estão descritos na tabela - 8 - .

TABELA - 8 -PORCENTAGEM DE REDUÇÃO NAS ESTIMATIVAS DE β e $var \beta$:

β	Redução na VAR			Redução em $\hat{\beta}$					
	$\hat{\beta}_{CP}$	$\hat{\beta}_k$	$\hat{\beta}_{RL}$	$\hat{\beta}_{CP}$	$\hat{\beta}_k$	$\hat{\beta}_{RL}$	$\hat{\beta}_\xi$	$\hat{\beta}_\rho$	$\hat{\beta}_{IT}$
β_1	22	57	19	40	32	70	8	15	37
β_2	10	66	0,6	-59	4	2	8	15	-107
β_3	12	50	10	11	28	52	8	15	96
β_4	47	63	1	-91	74	53	8	15	-32
β_5	36	58	8	16	13	-7	8	15	33
β_6	85	78	36	95	70	-117	8	15	83
β_7	86	83	83	83	49	42	8	15	79
β_8	65	75	41	-76	-48	-235	8	15	-19

Os estimadores em cristas e os contraídos, $\hat{\beta}_k, \hat{\beta}_\rho, \hat{\beta}_\xi$, apresentaram os resultados mais satisfatórios do ponto de vista de redução nas estimativas de β e $var \beta$. Em alguns casos houve um aumento do valor calculado, representado na tabela acima pelos sinais negativos.

É interessante notar que as maiores porcentagens de redução são referentes a $\hat{\beta}_6, \hat{\beta}_7$ e $\hat{\beta}_8$, que são os coeficientes de X_6, X_7, X_8 , COND HAB, BENS DUR E CONDECO, detectadas no item anterior como as principais variáveis responsáveis pela existência de multicolinearidade.

Comparações entre os estimadores, baseando-se no EQM e no EQMG , dependem de condições sobre os parâmetros, mais especificamente , $\beta^t\beta$ e σ^2 .

Utilizaremos como estimador de σ^2 , o QME , quadrado médio residual , obtido no método de mínimos quadrados:

$$\hat{\sigma}^2 = QME_{MQ} = 0,0403$$

Discutiremos regiões para $\beta^t\beta$, de maneira que obtenhamos os resultados desejados. Convém no entanto apresentarmos as diversas estimativas de $\beta^t\beta$, obtidas através dos 5 métodos estudados.

TABELA - 9 - ESTIMATIVAS DE $\beta^t\beta$

Método	Valor estimado de $\hat{\beta}^t\beta$
MIN QUAD	2,8388
COMP PRINC	1,3505
CRISTAS	1,0530
R LATENT	4,6434
CONTRAIDO	2,4069
INTERAT	0,9434

As estimativas obtidas são muito diferentes entre si, contudo pode-se inferir que $\beta^t\beta$ não é muito grande, podemos aceitar que $1 < \beta^t\beta < 5$. Esta informação é suficiente para avaliarmos os resultados subsequentes.

4.4 - COMPARAÇÕES ENTRE OS ESTIMADORES BASEANDO-SE NO EQM

Todos os resultados utilizados nesta descrição estão apresentados no capítulo 3 .

$$\hat{\beta}_{CP} \text{ x } \hat{\beta}_{MQ}$$

Para afirmarmos que $EQM(\hat{\beta}_{CP}) < EQM(\hat{\beta}_{MQ})$ devemos verificar se :

$$\sum_{j=7}^8 1/\lambda_j > \frac{\beta^t \beta}{\sigma^2}$$

(Resultado 1, item 3.1.1)

Este resultado pode ser escrito como :

$$\beta^t \beta < \sigma^2 \left\{ \frac{1}{\lambda_7} + \frac{1}{\lambda_8} \right\}$$

$$\beta^t \beta < 0,0403 \left\{ \frac{1}{0,1231} + \frac{1}{0,1998} \right\}$$

$$\beta^t \beta < 0,5291$$

Tendo em vista as estimativas de $\beta^t \beta$ apresentadas na tabela - 9 - , parece-nos que o estimador obtido pelo método de componentes principais não se apresenta melhor que o de mínimos quadrados para este conjunto de dados.

Se calculássemos $\hat{\beta}_{CP}$ retirando-se 5 componentes , obter-se-ia 81 por cento da variação total e a condição acima seria dada por :

$$\sum_{j=4}^8 \frac{1}{\lambda_j} > \frac{\beta^t \beta}{\sigma^2}$$

equivalente a :

$$\beta^t \beta < 0,8100$$

Mesmo neste caso seria difícil concordar que o resultado é favorável a $\hat{\beta}_{CP}$

$$\hat{\beta}_k \text{ x } \hat{\beta}_{MQ}$$

Vimos que se obtem $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{MQ})$ se :

$$k < \frac{\sigma^2}{\max_j(\Gamma_j^t \beta)^2}$$

(Resultado 2 , item 3.1.2)

No nosso exemplo utilizou-se $k=0,22$, para calcular-se $\hat{\beta}_k$.Supondo que foi uma boa escolha para k , podemos escrever:

$$\max_j(\Gamma_j^t \beta)^2 < \frac{\sigma^2}{0,22} \rightarrow \max_j(\Gamma_j^t \beta)^2 < 0,1832$$

As estimativas obtidas para $\max_j(\Gamma_j^t \beta)^2$, utilizando-se $\hat{\beta}_{MQ}$ e $\hat{\beta}_{CP}$ são :

$$\max_j(\Gamma_j^t \hat{\beta}_{MQ})^2 = 1,0601$$

$$\max_j(\Gamma_j^t \hat{\beta}_{CP})^2 = 0,5625$$

O que nos leva a crer que nesta situação , $EQM(\hat{\beta}_{MQ}) < EQM(\hat{\beta}_k)$

Supondo agora que $\max_j(\Gamma_j^t \beta)^2 \simeq 0,5600$, qual será o valor de k adequado de maneira a obter-se $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{MQ})$?

$$k < \frac{\sigma^2}{0,56} \rightarrow k < 0,0720$$

Calculou-se $\hat{\beta}_{k=0,07}$,o estimador obtido esta na tabela 10 :

TABELA - 10 - $\hat{\beta}_k$ para $k=0,07$.

	$\hat{\beta}_1$	$\hat{\beta}_2$	$\hat{\beta}_3$	$\hat{\beta}_4$	$\hat{\beta}_5$	$\hat{\beta}_6$	$\hat{\beta}_7$	$\hat{\beta}_8$
$\hat{\beta}_k$	0,5874	0,3445	0,6845	0,0653	0,4648	-0,5310	0,5605	0,2950
$\hat{\beta}_{MQ}$	0,6991	0,3414	0,7849	0,1269	0,4824	0,8390	0,7853	0,2185

$$\hat{\beta}_\rho \text{ x } \hat{\beta}_{MQ}$$

Para termos $EQM(\hat{\beta}_\rho) < EQM(\hat{\beta}_{MQ})$, devemos ter:

$$\rho > \left\{ \frac{\beta^t \beta - tr(\sigma^2(X^t X)^{-1})}{\beta^t \beta + tr(\sigma^2(X^t X)^{-1})} \right\}$$

(Resultado 4 , item 3.1.4)

Como $tr(\sigma^2(X^t X)^{-1}) = \sum_{i=1}^8 var(\hat{\beta}_{MQ(i)}) = 0,9454$ obtem-se :

$$\rho > \left\{ \frac{\beta^t \beta - 0,9454}{\beta^t \beta + 0,9454} \right\}$$

Fixando-se $\beta^t \beta = 1,5$, o que é razoavel segundo as estimativas obtidas na tabela - 9 - , tem-se :

$$\rho > \frac{1,5 - 0,9454}{1,5 + 0,9454} \Rightarrow \rho > 0,2268$$

O que nos leva a concluir que $\hat{\beta}_\rho$ é um estimador mais preciso que $\hat{\beta}_{MQ}$, para este conjunto de dados .

$$\hat{\beta}_{RL} \times \hat{\beta}_{MQ}$$

Vimos no item 3.1.3, que $\hat{\beta}_{MQ}$, é essencialmente dominado pelo autovetor associado ao menor autovalor de $X'X$, Γ_p , e que $\hat{\beta}_{RL}$, é essencialmente dominado por vetores que são ortogonais a Γ_p .

Descreveu-se também a única situação em que se tem $EQM(\hat{\beta}_{MQ}) < EQM(\hat{\beta}_{RL})$:

$$\beta^* = c\Gamma_p^*$$

com "c" = um escalar e β^* e Γ_p^* , definidos no item 3.1.3.

Neste conjunto de dados a redução esperada nas estimativas obtidas utilizando-se $\hat{\beta}_{RL}$, em relação as superestimativas de $\hat{\beta}_{MQ}$, não foram muito satisfatórias. Uma possível justificativa é estarmos diante da situação descrita acima, $\beta^* = c\Gamma_p^*$, o que também não foi possível detectar-se.

Verificou-se ainda que $\Upsilon_0^0 \simeq \Gamma_p$, resultado 4, item 3.1.3, podemos dizer que:

$$\Upsilon_0^0 \simeq -1\Gamma_p.$$

A tabela 11, apresenta os elementos dos dois autovetores, para que o leitor possa fazer a sua comparação:

TABELA 11 - Υ_0^0 e Γ_p

Υ_0^0	Γ_p
0,2615	-0,1992
0,0047	0,0447
0,2176	-0,1356
0,0998	-0,0625
-0,0189	0,1419
-0,5282	0,4364
0,6606	-0,7364
-0,2748	0,4283

$$\hat{\beta}_{CP} \times \hat{\beta}_k$$

Para termos $EQM(\hat{\beta}_k) < EQM(\hat{\beta}_{CP})$ devemos ter:

$$k \geq \frac{\sigma^2}{\beta^t \beta}$$

(Resultado 6 , item 3.1.6)

Nesta análise , considerando-se $k=0,22$, determinado através do traço da crista,obtem-se:

$$\beta^t \beta \geq \frac{\sigma^2}{0,22} \Rightarrow \beta^t \beta \geq 0,1832$$

Podemos inferir que $\hat{\beta}_k$, é melhor que $\hat{\beta}_{CP}$, já que as estimativas para $\beta^t \beta$, obtidas na tabela 9 , não contradizem a condição .

Entretanto calculou-se qual seria o valor de k adequado para obter-se a mesma conclusão acima se tivéssemos $\beta^t \beta < 5$, obteve-se :

$$k \geq 0,00806$$

O que reafirma a conclusão anterior já que se utilizou $k = 0,22$.

A tabela 12 , resume os resultados obtidos até o presente momento, portanto somente baseando-se no EQM :

TABELA - 12 - RESULTADOS DAS COMPARAÇÕES ENTRE OS ESTIMADORES UTILIZANDO-SE O EQM

ESTIMADORES	MELHOR	CONDIÇÃO
$\hat{\beta}_{CP} \times \hat{\beta}_{MQ}$	$\hat{\beta}_{CP}$	Se $\beta^t \beta < 0,5291$
$\hat{\beta}_k \times \hat{\beta}_{MQ}$	$\hat{\beta}_k$	Se $k < 0,07$ e $\max_j(\Gamma_j^t; \beta)^2 = 0,56$ Se $k = 0,22$ e $\max_j(\Gamma_j^t; \beta)^2 < 0,1832$
$\hat{\beta}_\rho \times \hat{\beta}_{MQ}$	$\hat{\beta}_\rho$	Se $\rho = 0,2268$ e $\beta^t \beta > 1,5$ ou se $\rho > 0,2268$ e $\beta^t \beta = 1,5$
$\hat{\beta}_{CP} \times \hat{\beta}_k$	$\hat{\beta}_k$	Para $k=0,22$ se $\beta^t \beta > 0,1832$ Para $\beta^t \beta < 5$ se $k \geq 0,0806$

4.5 - COMPARAÇÕES ENTRE OS ESTIMADORES ATRAVÉS DO EQMG

Vimos no capítulo 3 que é possível determinar-se condições para σ^2 e $\beta^t\beta$, de maneira que alguns dos estimadores viesados estudados tenham menor EQMG, erro quadrático médio generalizado, que o estimador de mínimos quadrados.

Mesmo utilizando-se este outro parâmetro de comparação, continua-se dependendo de condições sobre β e σ^2 .

Em alguns casos esta comparação não é possível, por exemplo, comparar-se $\hat{\beta}_{CP}$ com $\hat{\beta}_k$ se o número de componentes retiradas em $\hat{\beta}_{CP}$ for diferente de 1 ou de $(p-1)$, (resultado 7 - item 3.2.6).

Contudo é conveniente ilustrarmos os resultados obtidos.

$$\hat{\beta}_{CP} \times \hat{\beta}_{MQ}$$

Para se obter $EQMG(\hat{\beta}_{CP}) < EQMG(\hat{\beta}_{MQ})$, deve-se verificar se:

$$\sum_{j=p-r}^p \frac{\lambda_j(\Gamma_j^t\beta)^2}{\sigma^2} < 1$$

onde $\lambda_{p-r} \simeq \dots \simeq \lambda_p \simeq 0$.
(resultado 1, item 3.2.1)

obteve-se ;

$$\sum_{j=7}^8 \lambda_j(\Gamma_j^t\beta)^2 < 0,0403$$

Neste caso fica difícil isolar-se β ou $\beta^t\beta$, deixando a condição na forma acima.

$$\hat{\beta}_{MQ} \text{ x } \hat{\beta}_k$$

Para obter-se $EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_{MQ})$, basta verificar se:

$$0 < k < \frac{2\sigma^2}{\beta^t\beta}$$

resultado -2 - , item 3.2.2

Supondo que $\beta^t\beta < 5$ obtem-se :

$$0 < k < \frac{2(0,0403)}{5} \Rightarrow 0 < k < 0,0161$$

No inicio desta análise fixou-se $k=0,22$,substituindo-se este valor na condição acima , obtem-se $\hat{\beta}_k$ com menor erro quadrático médio que $\hat{\beta}_{MQ}$ se :

$$\beta^t\beta < 0,3664$$

$$\hat{\beta}_\rho \text{ x } \hat{\beta}_{MQ}$$

Da mesma forma que na comparação de $\hat{\beta}_{CP}$ com $\hat{\beta}_{MQ}$, não foi possível isolar-se $\beta^t\beta$,neste caso verificou-se que se:

$$\beta^t X^t X \beta < \frac{\sigma^2}{(1-\rho)^2(1-\rho^2)^{-1}} = 0,50$$

tem -se $EQMG(\hat{\beta}_\rho) < EQMG(\hat{\beta}_{MQ})$ para $\rho = 0,85$. (Resultado 3 , item 3.2.3) .

$$\hat{\beta}_\xi \text{ x } \hat{\beta}_{MQ}$$

Nesta análise β_ξ já foi calculado segundo as exigencias do resultado 4 do item 3.2.3 , de maneira a garantir-se que $EQMG(\hat{\beta}_\xi) < EQMG(\hat{\beta}_{MQ})$.

$$\hat{\beta}_{IT} \times \hat{\beta}_{MQ}$$

Para obtermos $EQMG(\hat{\beta}_{IT}) < EQMG(\hat{\beta}_{MQ})$, devemos determinar o menor “ m ” que satisfaça a equação :

$$(1 - \alpha\lambda_p)^{m+1} < \frac{1}{k\lambda_1}$$

sob a hipótese de que $\beta^t\beta < k\sigma^2$.

(Resultado 5 , item 3.2.4)

Utilizou-se:

- $\beta^t\beta < 5$

- $\alpha = 0,0625$ que é o ponto médio do intervalo $(0, \frac{1}{tr(X^tX)})$

- $\lambda_p = 0,1231$

- $\lambda_1 = 4,4863$

obtendo-se a equação :

$$(1 - (0,00625)(0,1231))^{m+1} < 0,0018$$

E determinou-se $m = 818$.

Com estes parâmetros $\alpha = 0,0625$ e $m = 818$, calculou-se $\hat{\beta}_{IT}$:

$$\hat{\beta}_{IT} = \begin{pmatrix} 0,4106 \\ 1,7265 \\ 0,3048 \\ 0,5064 \\ -0,5134 \\ -0,5817 \\ -0,9597 \\ 0,3116 \end{pmatrix}$$

O estimador $\hat{\beta}_{IT}$ converge para $\hat{\beta}_{MQ}$, (item 2.5.2), o que não foi possível ser detectado nesta análise para este conjunto de dados. Calculou-se $\hat{\beta}_{IT}$ para valores de m variando de 1 até 1200, verificando-se sua distancia euclidiana em relação ao estimador de mínimos quadrados, obtendo-se para os valores mais altos de “m” uma distancia que oscilava em torno de 2,42.

COMPARAÇÕES ENTRE OS ESTIMADORES VIESADOS

$$\hat{\beta}_\rho \quad \times \quad \hat{\beta}_k$$

O resultado 6 no item 3.2.5 , afirma que :

a - $EQMG(\hat{\beta}_k) < EQMG(\hat{\beta}_\rho)$ se $\left(\frac{1-\rho}{\rho}\right) \lambda_i > k_i$ para no máximo um valor de i .

b - $EQMG(\hat{\beta}_\rho) < EQMG(\hat{\beta}_k)$ se $\left(\frac{1-\rho}{\rho}\right) \lambda_i < k_i$ para no máximo um valor de i .

Existem 2 valores de “ i ” , para os quais a condição (a) esta satisfeita. E 6 para a condição (b) ,os resultados estão descritos na tabela 13 . Conclue-se portanto que este critério não pode ser usado neste conjunto de dados .

TABELA - 13 - RESULTADSO USADOS NA COMPARAÇÃO DE $\hat{\beta}_\rho$ COM $\hat{\beta}_k$

λ_i	$\left(\frac{1-\rho}{\rho}\right) \lambda_i$
4,4863	0,7917
1,2884	0,2274
0,6789	0,1198
0,5512	0,0973
0,3782	0,0667
0,2944	0,0520
0,1998	0,0353
0,1231	0,0217

$$\hat{\beta}_{CP} \text{ x } \hat{\beta}_k$$

Vimos no resultado 7 do item 3.2.6 que existem alguns valores dos parâmetros para os quais ora um estimador ora outro pode ser considerado melhor. Contudo não está explícita qual a exigência sobre β e σ^2 para se obter o resultado desejado.

RESUMO DOS RESULTADOS OBTIDOS ATRAVÉS DO EQMG

A tabela 14, resume os resultados obtidos para a comparação dos estimadores através do EQMG.

TABELA - 14 - COMPARAÇÃO ENTRE OS ESTIMADORES ATRAVÉS DO EQMG

COMPARAÇÃO	MELHOR	CONDIÇÃO
$\hat{\beta}_{MQ} \text{ x } \hat{\beta}_{CP}$	$\hat{\beta}_{CP}$	Se $\sum_{j=7}^8 \lambda_j (\Gamma_j^t \beta)^2 < 0,0403$
$\hat{\beta}_{MQ} \text{ x } \hat{\beta}_k$	$\hat{\beta}_k$	Para $k = 0,22$ se $\beta^t \beta < 0,3664$ Para $\beta^t \beta < 5$ se $0 < k < 0,0161$
$\hat{\beta}_{MQ} \text{ x } \hat{\beta}_\rho$	$\hat{\beta}_\rho$	se $\beta^t X^t X \beta < 0,50$ para $\rho = 0,85$
$\hat{\beta}_{MQ} \text{ x } \hat{\beta}_\xi$	$\hat{\beta}_\xi$	se $0 < \xi < 2(p-2)(n-p+2)^{-1}$
$\hat{\beta}_{IT} \text{ x } \hat{\beta}_{CP}$	$\hat{\beta}_{CP}$	Para $\alpha = 0,0625$ e $m = 818$ se $\beta^t \beta < 5$

5 - CONCLUSÃO

Vimos que a utilização dos 6 métodos estudados para a estimação dos parâmetros num modelo de regressão linear múltipla, são provenientes de interpretações diversas em relação ao tratamento dos efeitos da multicolinearidade. Alguns destes métodos são específicos para regressão, enquanto que outros, por exemplo a análise de Cristas e o método de componentes principais são técnicas mais gerais e podem ser utilizadas em outros contextos.

Embora provenientes de concepções diferentes, constatou-se que existe uma expressão geral através da qual se descreve os diferentes estimadores estudados, $\hat{\beta}_\rho, \hat{\beta}_{CP}, \hat{\beta}_{MQ}$ e $\hat{\beta}_k$. Esta expressão é dada por :

$$\hat{\beta} = \sum_{j=1}^p a_j c_j \Gamma_j$$

onde:

$c_j = \Gamma_j^t X^t Y$, é um escalar .

a_j é um fator de ponderação que diferencia $\hat{\beta}_\rho, \hat{\beta}_{CP}, \hat{\beta}_{MQ}$ e $\hat{\beta}_k$.

Γ_j , $j=1, \dots, p$, são os autovetores de $X^t X$.

(Gunst e Mason - 1977 - [16])

Os 4 estimadores são obtidos da seguinte maneira:

$\hat{\beta}_{MQ}$ para $a_j = \lambda_j^{-1}$, $j=1, \dots, p$.

$\hat{\beta}_{CP}$ para $a_j = \begin{cases} 0, & \text{para } j < r + 1, \\ \lambda_j^{-1}, & \text{para } j=r+1, \dots, p \end{cases}$

$\hat{\beta}_k$ para $a_j = (\lambda_j + k)^{-1}$, $j=1, \dots, p$

$\hat{\beta}_\rho$ para $a_j = \frac{\rho}{\lambda_j}$, $0 \leq \rho < 1$, $j=1, \dots, p$

Ou seja, estes estimadores podem ser expressos como combinação linear dos autovetores de $X^t X$.

De forma análoga $\hat{\beta}_{RL}$ pode ser escrito como combinação linear dos autovetores de $(YX)^t(YX)$ retirando-se seu primeiro elemento :

$$\hat{\beta}_{RL} = \sum_{j=s}^p f_j \Upsilon_j^0$$

onde :

$\Upsilon_j^0 = \Upsilon_j$, autovetor de $(YX)^t(YX)$, sem o primeiro elemento Υ_{0j}

$$f_j = -\frac{\eta \Upsilon_{0j} \lambda_j^{-1}}{\sum_{j=1}^p \Upsilon_{0j}^2 \lambda_j^{-1}} \quad , \quad \eta = \sum_{i=1}^n (Y_i - \bar{Y})^2 .$$

(item 2.3.3)

E ainda vimos que :

$$\hat{\beta}_{MQ} = \sum_{j=1}^p f_j \Upsilon_j^0$$

onde :

f_j e Υ_j^0 são os mesmos definidos acima .

(item 2.3.3)

Com relação ao conjunto de dados analisado , embora não tenha sido completamente satisfatório no sentido de se obter uma multicolinearidade de grau muito forte, foi de extrema utilidade para ilustrar os resultados apresentados bem como verificar a facilidade de cálculo das diversas estimativas devido a similaridade das expressões acima.

Toda a análise de dados foi obtida utilizando-se a linguagem FORTRAN , com o auxílio de rotinas do IMSL, processadas no computador Burroughs 6900, do CCE - Centro de Computação Eletrônica - USP .

Outro aspecto que merece destaque na pesquisa que envolveu este trabalho , é a redescoberta da interpretação original do método de Análise de Cristas,(item 2.2.1 e 2.2.2, capítulo 2) , que permitiu verificarmos que a constante k em $\hat{\beta}_k = (X^t X + kI)^{-1} X^t Y$, não consiste num simples acréscimo na diagonal de $X^t X$ no intuito de reduzir-se os efeitos da singularidade.

Embora quase nada pode ser efetivamente feito utilizando-se $\hat{\beta}_{RL}$, sua inclusão neste trabalho ficou por conta de sua rica interpretação geométrica,(item 2.3.2,capítulo 2), que entre outras coisas possibilitou a separação da multicolinearidade preditiva e multicolinearidade não preditiva , (definidas no item 2.3.1) .

A ausência de propriedades estatísticas que envolva $\hat{\beta}_{RL}$ é um convite para a elaboração de futuras pesquisas.

Finalizo com o desejo de compartilhar com o leitor um sentimento de satisfação e gratidão diante da rica experiência que o desenvolver deste trabalho me proporcionou , tanto no âmbito da pesquisa científica, bem como no relacionamento que pude desenvolver com todos aqueles que prontamente se envolveram no desenvolvimento e conclusão desta dissertação.

A MULTICOLINEARIDADE É UM PROBLEMA ?

A multicolinearidade na maioria das vezes é um problema, excetuando-se os casos em que exista uma transformação linear dos parâmetros, que seja interpretável e de interesse para o pesquisador.

Além disto esta transformação deveria ser exatamente aquela que ortogonaliza a matriz X :

$$Z = X\Gamma$$

onde :

- Γ : é a matriz cujas colunas são os autovetores normalizados de X^tX .
- Desta forma tem-se $\Gamma^t\Gamma = I$ e conseqüentemente:

$$Z^tZ = (X\Gamma)(X\Gamma)^t = \Gamma^tX^tX\Gamma = \Delta$$

com :

- Δ : é a matriz diagonal cujos elementos são os autovalores de X^tX .

Assim , as colunas de Z formam um novo conjunto de variáveis preditoras , ortogonais, definindo o novo modelo:

$$Y = \alpha Z + \epsilon$$

com :

- $\alpha = \Gamma^t\beta$.

Esta transformação linear é bastante explorada no item 2.1 , no capítulo 2, como um artifício para melhor entendermos as implicações provenientes da presença da multicolinearidade.

MULTICOLINEARIDADE INDUZIDA POR OUTLIER

Para ilustrar como um outlier pode induzir a multicolinearidade, vamos considerar no modelo $Y = X\beta + \epsilon$, descrito através de 2 variáveis preditoras, sendo X descrita por :

$$\begin{pmatrix} 1 & X_{11} & X_{12} \\ 1 & X_{21} & X_{22} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & X_{n1} & X_{n2} \end{pmatrix}$$

Suponha que X_{11} e X_{12} são arbitrariamente grandes:

$X_{11} = \phi$ e $X_{12} = c\phi$, com $c > 0$. Normalizando-se as colunas de X, obtém-se :

$$\begin{pmatrix} 1 & \frac{\phi}{k_1} & \frac{c\phi}{k_2} \\ 1 & \frac{X_{21}}{k_1} & \frac{X_{22}}{k_2} \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot \\ 1 & \frac{X_{n1}}{k_1} & \frac{X_{n2}}{k_2} \end{pmatrix}$$

com :

$$k_1 = \phi^2 + \sum_{i=2}^n X_{i1}^2 \dots e \dots k_2 + c^2 \phi^2 + \sum_{i=2}^n X_{i2}^2$$

Desta forma tem-se para $i=2, \dots, n$:

$$\lim_{\phi \rightarrow \infty} X_{1j} \rightarrow 1 \dots e \dots \lim_{\phi \rightarrow \infty} X_{ij} \rightarrow 0$$

Concluindo-se que para $\eta > 0$, arbitrariamente pequeno, existe ϕ^* tal que :

Para todo $\phi > \phi^*$ tem-se :

$$\begin{aligned} X_1 - X_2 &= \delta \\ \|\delta\| &\leq \sqrt{2}\eta \end{aligned}$$

Que é umas das formas de se descrever a multicolinearidade, apresentada no capítulo 1, item 1.1.

APÊNDICE - 3 -

A DECOMPOSIÇÃO ESPECTRAL DE UMA MATRIZ $A_{n \times p}$ OU DECOMPOSIÇÃO EM VALORES SINGULARES

Uma matriz arbitrária A , de dimensão $n \times p$, pode ser escrita da seguinte maneira:

$$A = USV^t$$

onde :

- U , matriz $n \times r$ e V , matriz $p \times r$ são ortogonais.
- S , matriz $r \times r$, é diagonal de elementos não negativos.

No modelo de regressão linear, a decomposição de X , é muito conveniente pois ao expressarmos os estimadores de mínimos quadrados é fácil ver os termos inflacionados pela presença de autovalores próximos de zero. Se $n \geq p$ e $r \leq p$ obtem-se :

$$X = HD\Gamma^t$$

onde :

- H : é $n \times p$, tal que suas colunas são os autovetores normalizados de XX^t , associados aos p autovalores diferentes de zero. Portanto $H^tH = I$. Convém ressaltar que estes p autovalores são os mesmos de X^tX .

- Γ : é $p \times p$, tal que suas colunas são os autovetores normalizados de X^tX , associados aos p autovalores diferentes de zero : $\Gamma^t\Gamma = I$.

- D : é uma matriz diagonal $p \times p$, cujos elementos são as raízes quadradas dos autovalores de X^tX , diferentes de zero, chamados de *VALORES SINGULARES DE X*.

Em alguns casos denota-se D por $\Delta^{1/2}$, onde Δ , é a matriz diagonal dos autovalores de X^tX .

(J. Mandel - 1982 - [9])

INVERSA GENERALIZADA DE UMA MATRIZ

Considere :

- A : uma matriz pxp , de posto r , $r \leq p$.
- Z : tal que $Z^t Z = A$.
- Γ : matriz pxp , cujas colunas são os autovetores normalizados de Z .
- Δ : matriz pxp , diagonal, cujos elementos são os autovalores de Z .

- Se A tem posto completo , $r = p$, então :

$$A^{-1} = \Gamma \Delta^{-1} \Gamma^t$$

onde A^{-1} , é a inversa , usual de A .

- Se A tem posto r , $r < p$, $(p - r)$ elementos de Δ são zero.

- Se A é quase singular , $p-r$ elementos de Δ são próximos de zero, e pode-se subdividir Δ e Γ em duas partes:

$$\Gamma = (\Gamma_r : \Gamma_{p-r})$$

com Γ_r : de dimensão pxr , e Γ_{p-r} : de dimensão $px(p - r)$.

$$\Delta = \begin{pmatrix} \Delta_r & 0 \\ 0 & \Delta_{p-r} \end{pmatrix}$$

com Δ_r : de dimensão rxr e Δ_{p-r} : de dimensão $(p - r)x(p - r)$.

A INVERSA GENERALIZADA DE A , é dada por :

$$A_r^+ = \Gamma_r \Delta_r^{-1} \Gamma_r^t$$

A INVERSA GENERALIZADA DE X^tX

$$(X^tX)^{-1} = \Gamma\Delta^{-1}\Gamma^t = \sum_{j=1}^p \lambda_j^{-1}\Gamma_j\Gamma_j^t$$

se (X^tX) tem posto completo .

Se X^tX , é quase singular , não podemos calcular $(X^tX)^{-1}$. Para definir sua inversa generalizada , considere as matrizes :

$\Delta^* = \Delta$,fazendo-se iguais a zero ,os elementos que são próximos de zero.

Γ_{p-r} , é a matriz Γ , excluindo-se as colunas que são os autovetores associados aos r autovalores próximos de zero.

Tem-se :

$$(X^tX)_r^+ = \Gamma_{p-r}\Delta_{p-r}^{-1}\Gamma_{p-r}^t = \sum_{j=1}^{p-r} \lambda_j^{-1}\Gamma_j\Gamma_j^t$$

A INVERSA GENERALIZADA E O ESTIMADOR $\hat{\beta}_{CP}$

Muitos autores utilizam a noção de inversa generalizada para a apresentação de $\hat{\beta}_{CP}$

$$\hat{\beta}_{CP} = (X^tX)_r^+ X^tY = \Gamma_{p-r}\Delta_{p-r}^{-1}\Gamma_{p-r}^t X^tY$$

Esta também é uma forma conveniente para verificar-se que $\hat{\beta}_{CP}$, é uma transformação linear de $\hat{\beta}_{MQ}$, basta substituir X^tY por $X^tX\hat{\beta}_{MQ}$:

$$\hat{\beta}_{CP} = W_{p-r}\hat{\beta}_{MQ}$$

com $W_{p-r} = \Gamma_{p-r}\Delta_{p-r}^{-1}\Gamma_{p-r}^t X^tX$.

MATRIZES N.N.D. - NÃO NEGATIVAS DEFINIDA

- Uma forma quadrática é dita POSITIVA DEFINIDA se :

$X^tAX > 0$ para todo $X \neq 0$. A matriz A associada a forma quadrática é tal que $A^t = A$, e é dita ser uma matriz POSITIVA DEFINIDA-PD.

- Uma forma quadrática é dita ser POSITIVA SEMI DEFINIDA se :

$X^tAX \geq 0$ para todo X e $X^tAX = 0$ para ALGUM $X \neq 0$, e a matriz A é dita ser POSITIVA SEMI DEFINIDA - PSD.

As duas classes de matrizes , PD e PSD , são ditas NÃO NEGATIVAS DEFINIDA - NND .

Outras definições aparecem na literatura, neste trabalho , utilizou-se a de Shayle R. Searle - 1982 - [39] , pág. 77 . No entanto é universal o fato de ter-se A uma matriz simétrica, pois as formas quadráticas sempre podem ser expressas por matrizes simétricas.

APÊNDICE - 6 -

DADOS UTILIZADOS PARA A ANÁLISE DO CAPÍTULO - 4 -

ESTADO DE SÃO PAULO
LISTAGEM DAS VARIÁVEIS TRANSFORMADAS

Região	Subregião	OBS	X1	X2	X3	X4	X5	X6	X7	X8	Y
II Litoral	Santos	1	5,97329	0,945199	0,92093	0,373940	0,453513	0,2859	0,2658	2,5599	1,17402
	São Sebastião	2	4,81588	0,794992	0,71985	0,355651	0,283832	-1,6347	-1,1193	-1,1872	0,63759
	Vale do Ribeira	3	5,14537	0,345003	0,75688	0,334921	0,333158	-1,6518	-2,4179	-1,2474	0,38639
III Vale do Paraíba	S.J. dos Campos	4	5,60729	0,762694	0,79482	0,325479	0,379587	0,3342	0,8647	1,5998	1,07449
	Teubetê	5	5,41086	0,712457	0,92064	0,358205	0,412334	-0,3561	-0,0917	-0,0815	1,13376
	Guaratiningetá	6	5,48550	0,712428	0,99106	0,344607	0,379559	0,6425	-0,6526	-0,7787	0,59507
IV Sorocaba	Sorocaba	7	5,74461	0,691815	0,87544	0,375207	0,395591	0,7929	0,5397	0,5507	0,77602
	Tatuf	8	4,98237	0,653123	0,76603	0,397039	0,440749	0,4099	0,4185	-0,0133	0,61444
	Itapetininga	9	5,05459	0,538896	0,78781	0,365139	0,420709	0,0651	-1,6188	-1,4499	0,42330
	Capão Bonito	10	4,96966	0,259479	0,27643	0,322967	0,284444	-1,6415	-2,1960	-0,7752	0,18230
	Itapeva	11	5,25652	0,309839	0,43209	0,330207	0,320463	-0,7727	-1,9308	-1,3347	0,17727
	Avaré	12	5,05289	0,511531	0,70040	0,348027	0,413731	-0,0980	-0,9494	-1,0656	0,50464
	Botucatu	13	5,03708	0,602917	1,35072	0,386167	0,490083	0,6960	0,2819	-0,5315	0,69980
	Campinas	14	6,07406	0,763218	0,91237	0,363712	0,406555	1,0049	1,3315	2,2335	1,26653
	Piracicaba	15	5,40071	0,760140	0,74368	0,372735	0,439950	1,2010	1,7166	1,6109	1,00159
	Limeira	16	5,48017	0,765988	0,73851	0,360892	0,449126	1,0493	1,0602	1,2800	0,65870
	Rio Claro	17	5,07491	0,773148	0,69757	0,403983	0,473980	1,4140	2,3994	0,0532	0,63316
	V Campinas	S.J. Boa Vista	18	5,09417	0,607291	1,25539	0,374895	0,422278	1,0631	0,8365	0,3647
Casa Branca		19	5,10083	0,560922	1,23323	0,362597	0,407752	1,0512	0,5161	0,3590	0,77596
Jundiaí		20	5,51785	0,795442	0,65622	0,371278	0,412590	0,7992	1,2598	1,7155	0,91336
Bragança Paulista		21	5,20014	0,510709	0,77373	0,387872	0,377114	0,4827	0,2377	-0,5326	0,49830
Ribeirão Preto		22	5,68496	0,757719	0,87121	0,361377	0,502307	1,0766	1,0427	0,8060	1,87760
VI Ribeirão Preto	Francis	23	5,26294	0,712227	0,61375	0,336056	0,440347	-0,1987	0,2111	-0,0463	0,80237
	Ituverava	24	4,92094	0,567902	0,81044	0,322676	0,485903	-0,4986	-0,6777	-0,7724	0,53385
	S.J. Barra	25	4,89372	0,651326	0,60740	0,330450	0,393715	0,3990	0,1022	0,5142	0,54486
	Barretos	26	5,01439	0,712953	0,92011	0,348735	0,426610	-0,1059	-0,0488	-0,0385	0,92671
	Jaboticabal	27	5,30194	0,665592	1,41370	0,375932	0,421155	0,5243	0,6299	0,5043	0,63966
	Araraquara	28	5,33474	0,668654	1,15048	0,377866	0,435410	1,0746	0,9440	0,5145	0,95308
	São Carlos	29	5,13043	0,772669	0,90185	0,386696	0,470101	1,1955	1,0875	1,2543	0,83684
	Bauru	30	5,46496	0,688043	0,93471	0,364440	0,439547	0,7152	0,6682	0,2150	0,86335
VII Bauru	Lins	31	4,95088	0,624661	1,48206	0,253326	0,515568	0,2764	0,3591	-0,0225	1,05257
	Jaú	32	5,06554	0,689831	2,17848	0,385742	0,454770	1,4677	0,7899	1,7158	1,06631
VIII S.J. do Rio Preto	S.J. Rio Preto	33	5,57793	0,506489	1,43711	0,355253	0,454051	0,3824	0,4259	0,1823	1,27548
	Catanduva	34	5,18092	0,562007	0,79848	0,366318	0,390165	-0,0101	0,4790	0,5993	0,81093
	Votuporanga	35	4,98698	0,521079	0,72818	0,324248	0,423168	-1,2273	-0,9791	-0,8095	0,51522
	Fernandópolis	36	4,96738	0,445583	0,59170	0,313566	0,378899	-1,6989	-1,1418	-0,8219	0,53900
	Jales	37	5,10946	0,417738	0,65286	0,307996	0,362699	-2,8561	-1,1406	-1,0232	0,54405
IX Aracatuba	Aracatuba	38	5,52704	0,615285	0,56456	0,333633	0,429622	-0,3966	-0,2596	-0,3529	0,74284
	Andradina	39	5,39862	0,526942	0,58220	0,292692	0,412066	-1,7852	-1,3321	-1,5743	0,31551
X Presidente Prudente	Pres. Prudente	40	5,41296	0,589682	0,93338	0,377420	0,420600	-0,6894	-0,2333	-0,3620	0,80372
	Pres. Venceslau	41	5,21348	0,461398	0,49477	0,290381	0,358717	-1,0493	-0,6059	-1,3033	0,36700
	Draçena	42	4,90301	0,496654	0,97518	0,321264	0,414885	-0,9150	-0,3119	-0,2728	0,73764
	Adamantina	43	4,85387	0,469170	1,13244	0,327457	0,433057	-0,9098	-0,0296	-0,0299	0,71400
	Oswaldo Cruz	44	4,62425	0,407755	0,37130	0,387775	0,387752	-0,6380	-0,0811	-0,4035	0,66513
XI Marília	Marília	45	5,32682	0,578153	1,50880	0,345161	0,418756	0,3995	0,1227	-0,1752	1,19678
	Assis	46	5,19185	0,520477	0,80435	0,342815	0,432716	0,2145	-0,2373	-0,3706	0,94507
	Ourinhos	47	5,27859	0,540221	0,93602	0,343727	0,409107	0,5595	-0,3250	-0,9421	0,47380
	Tupã	48	4,89313	0,580204	1,05447	0,346761	0,402544	-0,4425	-0,1903	-0,4153	0,92088

X1 - log de População
X2 - proporção de população urbana
X3 - leitos por 1 000 habitantes / 4,5
X4 - arco seno $\sqrt{\frac{\text{população com mais de 50 anos}}{\text{pop. de 1970}}}$

X5 - arco seno $\sqrt{\frac{\text{matrículas no 2º grau}}{\text{matrículas no 1º grau}}}$
X6 - escore fatorial para as condições habitacionais
X7 - escore fatorial para os bens duráveis
X8 - escore fatorial para condição econômica
Y - nº de médicos por 1 000 habitantes

REFERÊNCIAS BIBLIOGRÁFICAS

- [1] - MONTGOMERY, D. C. E PECK, E. A. *Introduction to linear regression analysis*. New York , John Wiley, 1982. 504p. (Wiley Series in Probability and Mathematical Statistics)
- [2] - GUNST, R. F. Regression analysis with multicolinear predictor variables: definition, detection and effects. *Comm. Statist. A.: Theory Methods*, 12(19) : 2217-2260, 1983 .
- [3] - MASON, R. L. E GUNST R. F. Outlier-induced collinearities. *Technometrics*, 27 (4) : 401-407, 1985.
- [4] - VINOD, H. D. E ULLAH, A. *Recent advances in regression methods*. New York, Marcel Dekker, c1981. 361p. (Statistics Textbooks and Monographs, 41)
- [5] - LAWSON, C. L. E HANSON, R. J. Basic linear algebra including projections. in: *e Solving least squares problems*. Englewood cliffs, Prentice-Hall, c1974. p.233-239.
- [6] - HOFFMAN, K. E KUNZE, R. A. *Algebra linear* . 2.ed. São Paulo, Livros Técnicos e Científicos, 1979. 514p.
- [7] - MANSFIELD, E. R. E HELMS, B. P. detecting multicollinearity. *Amer. Statist.*, 36(3 part 1) :158-160, 1982.
- [8] - MASON, R. L. E GUNST, R. F. *Regression analysis and its application: a data oriented approach*. New York, Marcel Dekker, c1980. 402p. (Statistics Textbooks and Monographs, 34)
- [9] - MANDEL, J. Use of the singular value decomposition in regression analysis. *Amer. Statist.*, 36 (1) : 15-24, 1982.
- [10] - WEBSTER, J. T.; GUNST, R. F.;MASON, R. L. Latent root regression analysis. *Technometrics*, 16 (4) :513-522, 1974.
- [11] - HAWKINS, D. M. On the investigation of alternative regressions by principal component analysis. *J. Roy. Statist. Soc. Ser. C.*, 22 (3) :275-286, 1973.

[12] - SHARMA, S. E JAMES, W. L. Latent root regression: an alternative procedure for estimating parameters in the presence of multicollinearity. *J. Mark. Res.*, 18 : 154-161, 1981.

[13] - HAWKINS, D. M. Relations between ridge regression and eigenanalysis of the augmented correlation matrix. *technometrics*, 17 (4) : 477-480, 1975.

[14] - MASON, R. L. Latent root regression: a biased regression methodology for use with collinear predictor variabls. *Comm. Statist. A.: Theory Methods*, 15 (9) :2651-2678, 1986.

[15] - MAYER, L. S. E WILLKE, T. A. On biased estimation in linear models. *Technometrics*, 15 (3) : 497-508, 1973.

[16] - GUNST, R. F. E MASON, R. L. Biased estimation in regression: an evaluation using mean squared error. *J. Amer. Statist. Assoc.*, 72 (359) :616-628, 1977.

[17] - HOCKING, R. R.; SPEED, F. M.; LYNN, M. J. A class of biased estimators in linear regression. *Technometrics*, 18(4) :425-437, 1976.

[18] - GUNST, R. F. E MASON, R. L. Generalized mean squared error properties of regression estimators. *Comm. Statist. A.: Theory Methods*, (15) : 1501-1508, 1976.

[19] - HOERL, A. E. E KENNARD, R. W. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics*, 12 (1) : 55-67, 1970.

[20] - MARQUARDT, D. W. Generalized inverses, ridge regression, biased linear estimation, and nonlinear estimation. *Technometrics*, 12 (3) : 591-612, 1970.

[21] - GUNST, R. F.;WEBSTER, J. T.;MASON, R. L. A comparison of least squares and latent root regression estimators. *Technometrics*, 18 (1) :75-83, 1976.

[22] - BAYE, M. R. E PARKER, D. F. Combining ridge and principal component regression: a money demand illustration. *Comm. Statist. A.: Theory Methods*, 13(2) ; 197-205, 1984.

[23] - HSUAN, F. C. Ridge regression from principal component point of view. *Comm. Statist. A.: Theory Methods*, 10(19) :1981-1995, 1981.

[24] - TRENKLER, D. E TRENKLER, G. On the Euclidian distance between biased estimators. *Comm. Statist. A.: Theory Methods*, 13 (3) :273-284, 1984.

- [25] - LOWERRE, J. M. On the mean square error of parameter estimates for some biased estimators. *Technometrics*, 16 (3) :461-464, 1974.
- [26] - WHITE, J. W. E GUNST, R. F. Latent root regression: large sample analysis. *Technometrics*, 21(4):481-488, 1979.
- [27a] - COXE, K. Selection rules for principal components regression, comparison with latent root regression. In:MEETING OF THE AMERICAN STATISTICAL ASSOCIATION, Cincinnati, 1982. *Proceedings of the Business and Economic Statistics Section*. Washington, ASA, s.d. p.222-227.
- [27b] - MARSH, L. C. On linking ridge regression and principal components regression. In: MEETING OF THE AMERICAN STATISTICAL ASSOCIATION, Cincinnati, 1982. *Proceedings of the Business and Economic Statistics Section*. Washington, ASA, s.d. p.228-233.
- [28] - JOLLIFFE, I. T. A note on the use of principal components in regression. *J. Roy. Statist. Soc. Ser. C.* , 31(3) :300-303, 1982.
- [29] - MASON, R. L. E GUNST, R. F. Selecting principal components in regression. *Statistics and Probability Letters*, 3(6) :299-301, 1982.
- [30] - DRAPER, N. R. E SMITH, H. *Applied regression analysis*. 2.ed. New York, John Wiley, c1981. 709p. (Wiley Series in Probability and Mathematical Statistics)
- [31] - HOERL, R. W. Ridge analysis 25 years later. *Amer. Statist.*, 39 (3) : 186-192, 1985.
- [32] - McDONALD, G. C. Some algebraic properties of ridge coefficients. *J. Roy. Statist. soc. Ser. B.*, 42(1) :31-34, 1980.
- [33] - LAWLESS, J. F. Mean squared error properties of generalized ridge estimators. *J. Amer. Statist. Assoc.*, 76(374) : 462-466, 1981.
- [34] - SCLOVE, S. L. Improved estimators for coefficients in linear regression. *J. Amer. statist. Assoc.*, 63(322) :596-606, 1968.
- [35] - FAREBROTHER, R. W. A class of shrinkage estimators. *J. Roy. Statist. Soc. Ser. B.*, 40(1) :47-49, 1978.

[36] - TRENKLER, G. Generalized mean squared error comparisons of biased regression estimators. *Comm. Statist. A.:Theory Methods*, 9(12) :1247-1259, 1980.

[37] - THEOBALD, C. M. Generalizations of mean square error applied to ridge regression. *J. Roy. Statist. Soc. Ser. B.*, 36(2) : 103-106, 1974.

[38] - PRICE, J. M. Comparisons among regression estimators under the generalized mean square error criterion. *Comm. Statist. A.: Theory Methods*, 11 (17) :1965-1984, 1982.

[39] - SEARLE, S. R. *Matrix algebra useful for statistics*. New York, John Wiley, c1982. 438p. (Wiley Series in Probability and Mathematical Statistics. Applied Probability and Statistics)

[40] - HOERL, A. E. Optimum solution of many variables equations. *Chem. Engrg. Progress*, 55 (2) :69-78, 1959.

[41] - HOERL, A. E. Application of ridge analysis to regression problems. *Chem. Engrg. Progress*, 58(3) :54-59, 1962.

[42] - OISHI, J. *Regressão sobre cristas*. São Paulo, 1983. 100p. Dissertação (Mestrado) - IME-USP.

[43] - KENDALL, M. C. *A course in multivariate analysis*. New York, Hafner, 1961. 185p. (Griffin's Statistical Monographs and Courses, 2)

[44] - MASSY, W. F. Principal components regression in exploratory statistical research. *J. Amer. Statist. Assoc.*, 60 (309) :234-256, 1965.

[45] - HAITOVSKY, Y. A note on regression on principal components. *Amer. statist.*, 20(4) :28-29, 1986.

[46] - HOERL, R. W. The application of ridge techniques to mixture data: ridge analysis. *Technometrics*, 29(2) :161-172, 1987.

[47] - BUSSAB, W. O. Os médicos de São Paulo, onde estão? *Estudos FUNDAP: Residência Médica*: 53-65, 1984.

[48] - KNUTH, D. E. *The texbook*. Reading, Addison-Wesley, c1984. 483p.

[49] - BRADLEY, A. R. E SRIVASTAVA, S. S. Correlation in polynomial regression. *Amer. Statist*, 33(1) : 11-14, 1979.

[50] - WILLAN, R. A. E WATTS, G. D. Meaningful multicollinearity measures.
Technometrics, 20(4) : 407-412, 1978.