REGRESSÃO SOBRE CRISTAS

JORGE OISHI

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: ESTATÍSTICA

ORIENTADOR: *Prof. Dr. Wilton de Oliveira Bussab*

- SÃO PAULO, JULHO DE 1983 -

À

*Ignes, Jorginho,
Marcos e aos meus
pais.*

AGRADECIMENTOS

- Ao Prof. Dr. Wilton de Oliveira Bussab, pela orienta
ção segura e paciente que sempre nos proporcionou.
- Ao Prof. Dr. Abraham Laredo Sicsū, pelos preciosos co
mentários que muito nos auxiliaram.
- Ao Prof. Dr. José Ferreira de Carvalho, pela motiva
ção e incentivos incansáveis.
- Aos amigos do Departamento de Computação e Estatística
da UFSCar pelo apoio de todas as horas.
- Aos CNPq, Finep e Embrapa, pelo suporte financeiro
parcial a este trabalho.
- À Ana Sigoli Fernandes Matheus pela confecção dos inū
meros gráficos.
- À Maria Fatima L. Dallantonia, pelo rāpido e excelen
te trabalho de datilografia.

ÍNDICE

CAPÍTULO I	- INTRODUÇÃO	1
CAPÍTULO II	- ANÁLISE DE CRISTAS	6
2.1	- Introdução	6
2.2	- O Método	8
2.3	- Modelo Geral	15
2.4	- Comentários	16
CAPÍTULO III	- REGRESSÃO SOBRE CRISTAS	18
3.1	- Introdução	18
3.2	- O Método dos Mínimos Quadrados	19
3.3	- Multicolinearidade	20
3.4	- O Estimador Através das Cristas	23
3.5	- Interpretação Geométrica	26
3.6	- Modos Alternativos para a Determinação de $\underline{B}(k)$	27
3.7	- Propriedades de $\underline{B}(k)$	28
3.8	- O Erro Quadrático Médio de $\underline{B}(k)$: EQM ($\underline{B}(k)$)	31
CAPÍTULO IV	- CRITÉRIOS PARA SELEÇÃO DE k	37
4.1	- Introdução	37
4.2	- Critério Baseado no Teorema da Existência	38

4.3 - $\underline{B}(k)$ Através da Estimação de $\underline{\beta}'\underline{\beta}$	39
4.4 - Critério Através do Enfoque Bayesiano	40
4.5 - Critério Baseado no Traço da Crista	42
4.6 - Exemplos	49
 CAPÍTULO V - SELEÇÃO DE VARIÁVEIS	66
5.1 - Introdução	66
5.2 - Seleção de Variáveis	67
5.3 - Exemplos	69
5.4 - Comentários	74
 CAPÍTULO VI - CONCLUSÃO	77
6.1 - Sugestões para a Obtenção de k	79
6.2 - Sugestões para a Seleção de Variáveis	81
6.3 - Pontos que Necessitam de Mais Pesquisas ...	81
6.4 - Derivações e Trabalhos Relacionados	82
 APÊNDICE	86
 REFERÊNCIAS	94

C A P Í T U L O I

INTRODUÇÃO

A Análise de Regressão Linear, tornou-se sem dúvida, em uma das mais utilizadas técnicas estatísticas. Seu emprego tem-se efetivado em quase todos os campos do conhecimento científico, principalmente em função de seu baixo custo computacional, aplicabilidade em um número muito grande de problemas, e um embasamento teórico muito bem fundamentado.

Entretanto, em algumas áreas em que os experimentos não são devidamente planejados, surgem alguns problemas, onde esta técnica pode não ser a mais adequada.

A multicolinearidade, devido a um forte relacionamento entre as variáveis independentes, é um desses problemas.

Nesses casos, a solução de mínimos quadrados, técnica padrão na resolução de problemas de Regressão Linear, pode apresentar sérios problemas tanto no campo estatístico como computacional.

Estatisticamente, o principal problema é a pouca precisão com que as estimativas dos coeficientes são obtidas, isto é, a multicolinearidade pode originar uma variância muito grande, sendo praticamente inútil para propósitos práticos. Do ponto de vista computacional a precisão dos resultados está diretamente relacionada com o Número de Condição, que é uma das medidas da multicolinearidade. Quanto maior for este número, menor será o número de dígitos significativos da solução.

A multicolinearidade é portanto, um sério contratempo para o analista de dados e não é difícil de ser encontrada em algumas áreas, como por exemplo: Econometria, onde quase sempre um forte relacionamento entre as variáveis independentes surge naturalmente.

As formas de detecção e tratamento da multicolinearidade ainda não estão totalmente desenvolvidas, existindo ainda alguma controvérsia sobre a melhor forma de solucionar os problemas originados pela sua presença. Entretanto, algumas técnicas já são razoavelmente aceitas nos tratamentos, co

mo por exemplo, a estimação viesada, das quais Regressão sobre Cristas (Ridge Regression) é uma delas. (Para maiores detalhes, consultar Besley e outros (1980)).

Este trabalho tem por objetivo apresentar a técnica de Regressão sobre Cristas, suas propriedades e sua utilização, cujo enfoque tem recebido ultimamente, por parte de vários autores, uma grande atenção. Ver por exemplo: Chatterjee e Price (1977), Besley e outros (1980), Gunst e Mason (1980), Draper e Smith (1981), etc.

Vamos mostrar que esta técnica pode ser uma opção bastante interessante nos casos onde a multicolinearidade é alta. Ela apresenta o estimador $\underline{B}(k) = (X'X + kI)^{-1}X'\underline{y}$ que é tendencioso mas possui algumas propriedades ótimas com relação ao Erro Quadrático Médio (EQM).

Uma característica importante é a possibilidade de representação gráfica das estimativas dos coeficientes obtidos segundo esta técnica, que permite numa simples inspeção visual, detectar possíveis instabilidades nos sistemas de equações normais quando sujeitos a pequenas perturbações dos dados.

Esse dispositivo gráfico, que recebeu o nome de Traço da Crista, pode vir a ser uma ótima ferramenta para o analista de dados na detecção e análise da multicolinearidade prejudicial à solução de Mínimos Quadrados.

Além disso, vamos mostrar que através da sua utilização, é possível selecionar um subconjunto de variáveis

que seja mais estável do que o conjunto completo. Partindo do fato que uma interpretação da multicolinearidade é a de que algumas variáveis não conseguem adicionar mais informações do que aquela já existente através das outras, cujo relacionamento é alto, vamos mostrar que através da utilização do Traço da Crista poderemos selecionar um subconjunto de variáveis cujo grau de multicolinearidade é mais baixo, ou de outro modo, eliminar aquelas variáveis que já não conseguem acrescentar mais informações, e só conseguem aumentar a instabilidade das soluções dos sistemas de equações normais.

Esta técnica foi introduzida por Arthur E. Hoerl em 1959, com o nome de Análise de Cristas (Ridge Analysis), na tentativa de oferecer uma técnica mais refinada para os problemas de otimização industrial.

Logo após, em 1962, Hoerl lançou um trabalho onde pela primeira vez utilizava a Análise de Cristas em problemas de Regressão Linear e posteriormente em 1970, num trabalho conjunto com R.W. Kennard, introduziram definitivamente a Regressão sobre Cristas.

Dessa forma iniciaremos no Capítulo II, com a Análise de Cristas, mostrando as raízes dessa técnica, enquanto que no Capítulo III, vamos mostrar o seu desenvolvimento em Regressão Linear, bem como demonstraremos algumas propriedades interessantes, em termos de EQM.

Como $B(k)$ depende do parâmetro k , veremos no Capí

tulo IV, alguns critérios para a escolha de um valor ótimo para k , onde entretanto, a ênfase maior será dada ao Traço da Crista, com algumas aplicações vistas no final do mesmo.

No Capítulo V, veremos como poderemos utilizar o Traço da Crista em Seleção de Variáveis, e alguns exemplos serão dados. No Capítulo VI, algumas recomendações e algumas indicações de outros desenvolvimentos que não foram considerados aqui serão expostos.

Algumas tabelas resultantes de dois trabalhos de simulação comparando os critérios expostos no Capítulo IV, são dados no Apêndice.

C A P Í T U L O I I

ANÁLISE DE CRISTAS

2.1 - INTRODUÇÃO

Este método foi introduzido por Hoerl (1959), na tentativa de desenvolver uma técnica alternativa de otimiza
ção a ser aplicada em problemas industriais.

Os métodos tradicionais de otimização de até en
tão, eram de difícil manipulação numérica e quase impraticá
veis quando o número de variáveis envolvidas era muito gran
de. Havia a necessidade de métodos mais fáceis de se traba
lhar e interpretar.

Hoerl desenvolveu então este método que chamou de

Análise de Cristas (Ridge Analysis), que consiste em analisar as várias cristas (Ridges) a partir da origem, na busca de uma solução alternativa.

Essas cristas representam os caminhos constituídos dos pontos de máximo ou de mínimo da função a ser otimizada, quando certas restrições são impostas.

No nosso caso estaremos interessados na resolução de sistemas lineares, onde se verifica o problema de mal condicionamento, tendo como consequência a instabilidade da solução com relação aos dados apresentados.

Vamos estudar o sistema:

$$A\tilde{x} = \tilde{b}, \quad (2.1.1)$$

onde:

A : matriz conhecida de ordem $n \times p$, de posto p ,

$\tilde{b}$: vetor $n \times 1$ conhecido.

A solução obtida através do método dos Mínimos Quadrados é dada por:

$$\hat{\tilde{x}} = (A'A)^{-1}A'\tilde{b} \quad (2.1.2)$$

que é conseguida minimizando a soma de quadrados residual:

$$S = (\tilde{b} - A\tilde{x})'(\tilde{b} - A\tilde{x})$$

Essa solução no entanto às vezes não é satisfató

ria nos sistemas mal condicionados, isto é, nos sistemas onde pequenas perturbações nos coeficientes acarretam grandes mudanças na resposta.

A proposta da Análise de Cristas, é a de servir de opção nestes casos, além de introduzir um dispositivo gráfico que permite através de uma simples visualização, a verificação dessa instabilidade.

2.2 - O MÉTODO

Vamos apresentar este método através de um exemplo bem simples.

Seja o sistema linear:

$$\begin{aligned} 10x+9y &= 19 \\ 9x+8y &= 17 \end{aligned} \tag{2.2.1}$$

cujá solução exata é: $x=y = 1$.

Alterando ligeiramente um dos coeficientes na primeira equação de (2.2.1):

$$\begin{aligned} 10,1x+9y &= 19 \\ 9x+8y &= 17 \end{aligned} \tag{2.2.2}$$

a solução passa a ser: $x = 5,0$ e $y = -3,5$.

Portanto, uma perturbação de 1% em um dos coeficientes acarreta uma grande mudança na solução exata do problema. Essa é uma das características de sistemas mal condi

cionados. Pequenos erros de arredondamento, por exemplo, poderão nos levar a soluções bastante afastadas da verdadeira.

Vamos mostrar um processo que pode ser usado para a verificação dessa instabilidade.

Seja S a soma de quadrados residual dada por:

$$\begin{aligned} S &= (19-10x-9y)^2 + (17-9x-8y)^2 \\ &= 181x^2 + 145y^2 - 686x - 614y + 324xy + 650 \end{aligned}$$

Minimizando S , obtemos a solução de Mínimos Quadrados: $x=y = 1$.

Vamos a partir de agora, raciocinar do seguinte modo:

Suponhamos que S representa uma superfície na qual estamos de olhos vendados, em cima da origem $(0,0)$ e desejamos saber sem sair do lugar, onde se localiza o ponto mais baixo (alto) da superfície.

Um dos procedimentos que podemos adotar é: estendendo uma das pernas e examinando a superfície ao nosso redor num círculo cujo raio é o comprimento da mesma, podemos anotar o ponto mais baixo (alto) da superfície naquele círculo.

Se pudessemos variar à nossa vontade o tamanho da perna, poderíamos determinar todos os pontos de mínimo (máximo) da superfície, a partir da origem.

Unindo todos esses pontos de mínimos (máximos) so

bre os círculos, vamos obter um caminho sobre a superfície que certamente vai passar pelo mínimo (máximo) absoluto. Esse caminho recebeu o nome de CRISTA (Ridge) da função.

Dessa forma podemos determinar os pontos críticos da superfície em função de sua distância ao centro $(0,0)$. Isto é equivalente a determinar os pontos críticos da função S restritos a círculos de raio R com centro na origem.

Formalmente, no nosso exemplo, devemos minimizar S sujeito a restrição:

$$x^2 + y^2 = R^2 \quad (2.2.3)$$

Pelo método de Lagrange para máximos e mínimos condicionados temos:

$$F = S + k(x^2 + y^2 - R^2),$$

onde k : multiplicador de Lagrange.

Derivando F com relação a x e a y e igualando a zero, obtemos o seguinte sistema:

$$(181+k)x + 162y = 343 \quad (2.2.4)$$

$$162x + (145+k)y = 307$$

que tem como incógnitas, x , y e k .

Portanto, para acharmos uma solução para o sistema acima, precisamos conhecer o valor de k ou R . Se não temos conhecimento a priori desses parâmetros temos um novo proble

ma: como determinar k ou R .

Infelizmente esta técnica não foi suficientemente desenvolvida para fornecer um critério ótimo para a determinação de k . Algumas sugestões subjetivas podem ser encontradas em Hoerl (1962).

Porém independentemente desse fato, podemos atribuir valores para k ou R , e conseqüentemente encontrar os correspondentes x e y através do sistema (2.2.4). Dessa forma se fizermos k ou R variar num intervalo, x e y também vão variar em um intervalo e poderemos então colocá-los num gráfico bidimensional, com k ou R no eixo horizontal e x e y no eixo vertical. Esse gráfico representará as coordenadas da Crista mínima da função S , ou seja, para cada k ou R , (x, y) representará as coordenadas de um ponto mínimo relativo. Esse gráfico recebeu o nome de TRAÇO da Crista.

Do ponto de vista computacional é extremamente mais fácil atribuir valores a k e acharmos x e y do que atribuir valores a R . Observe que para resolver o sistema através de R precisamos resolver uma equação envolvendo uma raiz quadrada.

Por exemplo: se $k=0$, então (2.2.4) fica:

$$181x + 162y = 343 \quad (2.2.5)$$

$$162x + 145y = 307$$

de onde obtemos facilmente: $x=y=1$, e substituindo em (2.2.3)

temos: $R = \sqrt{2}$.

Outra vantagem de utilizar k ao invés de R é que o mínimo absoluto de S é encontrado em $k=0$, ou seja, fazendo $k=0$ encontramos o ponto de mínimo absoluto da função, ($x=y=1$) sobre o círculo cujo raio é $R=\sqrt{2}$. Esta propriedade é válida em qualquer sistema, como veremos em 2.3.

A relação existente entre k e R , pode ser encontrada, depois de algumas manipulações algébricas:

$$R = \frac{\sqrt{(2+1300k+211898k^2)}}{(1+326k+k^2)} \quad (2.2.6)$$

onde $R=0$ se $k=+\infty$, e $R=+\infty$ se $k=-\lambda_2$: menor autovalor da matriz dos coeficientes do sistema (2.2.5).

Isto é decorrência do fato que se $k=-\lambda_2$, o sistema (2.2.4) não tem solução única pois a matriz dos coeficientes torna-se singular, como é fácil observar:

$$\begin{vmatrix} 181+k & 162 \\ 162 & 145+k \end{vmatrix} = \begin{vmatrix} 181-\lambda_2 & 162 \\ 162 & 145-\lambda_2 \end{vmatrix} = 0,$$

fazendo x e y tender para $\pm\infty$, e consequentemente $R=+\infty$.

Neste problema os autovalores em questão são:

$$\lambda_1 = 325.99693 \quad \text{e} \quad \lambda_2 = 0.00307$$

logo, a medida que $k \rightarrow -0.00307$, $R \rightarrow +\infty$.

Dessa forma fazendo k ou R variar, vamos achar a

Crista de S . O gráfico abaixo nos mostra a variação de x e y contra R . Como S também é função de R , podemos colocá-lo no mesmo gráfico adotando uma escala diferente. No caso por conveniência ao invés de S , colocamos $\sqrt{S}$. Nesse gráfico k está variando de $+\infty$ a $-\lambda_2$.

GRÁFICO 2.2.1

Pelo comportamento dos gráficos de x e y podemos ver que são bastante influenciados pelos valores de $R > \sqrt{2}$, ou $k < 0$, enquanto que para esses mesmos valores, o resíduo não se altera muito. É possível então perceber a existência de soluções bastante afastadas da exata, que no entanto não apresentam erros elevados. Por exemplo:

Se $k = -0,0030264$ ou $R \approx 6$, temos

$$x = -2,81 \quad \text{e} \quad y = 5,29$$

acarretando um erro de $S = 0,261$, ou seja, a "solução" $(-2,81; 5,29)$ embora muito distante da exata $(1,1)$, não representa um erro muito elevado. Substituindo na equação original (2.2.1), obtemos:

$$10x+9y = 10(-2,81)+9(5,29) = 18,7$$

$$9x+8y = 9(-2,81)+8(5,29) = 17,3$$

o que implica em uma diferença de 1,6% na primeira equação e 1,7% na segunda, enquanto x tem uma variação de 381% e y 429%.

Para valores de $k \geq 0$, ou $R \leq \sqrt{2}$, as curvas são bastante estáveis, enquanto o resíduo cresce rapidamente. Isto quer dizer que a solução de Mínimos Quadrados, apesar do mal condicionamento do sistema é a solução que reúne as melhores condições de estabilidade. Em outras palavras, a solução de Mínimos Quadrados $(1,1)$ está numa região estável do gráfico dos coeficientes, e como ela apresenta o menor resíduo (no caso é zero), ela é a melhor solução do sistema.

Podemos então concluir que este gráfico pode ser extremamente útil em diagnosticar a instabilidade de um sistema de equações.

Nos próximos capítulos veremos como aproveitar este potencial para tentarmos achar uma solução mais estável do

que a solução dos Mínimos Quadrados em problemas de Regressão Linear.

2.3 - MODELO GERAL

No sistema geral (2.1.1):

$$A\tilde{x} = \tilde{b}$$

o sistema (2.2.4) é obtido minimizando:

$$S = (\tilde{b} - A\tilde{x})'(\tilde{b} - A\tilde{x})$$

sujeito a restrição:

$$\tilde{x}'\tilde{x} = R^2$$

Achando o mínimo condicionado através de Lagrange:

$$F = S + k(\tilde{x}'\tilde{x} - R^2)$$

derivando F com relação a $\tilde{x}$ e igualando a zero, obtemos:

$$\hat{\tilde{x}}(k) = (A'A + kI)^{-1}A'\tilde{b} \quad (2.3.1)$$

para algum k , que é a solução obtida através do Método das Cristas.

Se $k=0$, $\hat{\tilde{x}}(k) = (A'A)^{-1}A'\tilde{b}$: solução dos Mínimos Quadrados do sistema (2.1.1).

Observe que isto é equivalente a acrescentar a

constante k à diagonal da matriz $A'A$ antes de invertê-la. Dessa forma é possível melhorar a estabilidade do sistema.

O Traço da Crista é a representação gráfica dos coeficientes $\hat{x}_i(k)$'s, $i=1,2,\dots,p$, conjuntamente com o gráfico dos resíduos, contra k . Um fato importante é que este gráfico pode ser construído com qualquer número de variáveis. Como veremos no Capítulo V, isso será de grande valia na aplicação deste método em seleção de variáveis.

2.4 - COMENTÁRIOS

O problema de solucionar o sistema de equações:

$$A\tilde{x} = \tilde{b}, \quad (2.4.1)$$

onde A é não singular, porém mal condicionada foi também abordado pelo matemático Hadamard no começo do século e depois por Tychonov, sob o nome de Problemas Mal Postos (Improperly Posed). Ver Lavrentiev (1967).

Mais especificamente, tratou-se da minimização de:

$$R(x) = ||A\tilde{x} - \tilde{b}||^2 + \psi(x),$$

onde $||.||$: norma qualquer e $\psi(x)$ é uma função cuidadosamente escolhida a fim de assegurar a estabilidade do sistema.

A justificativa para este procedimento baseia-se no seguinte raciocínio:

Admite-se o fato de que o problema geral de solu

cionar (2.4.1) onde A é mal condicionada, é impossível. Não se pode garantir precisão.

Ao invés de uma única solução, temos para (2.4.1) uma classe de soluções, ou seja, para cada $\varepsilon > 0$, podemos encontrar o conjunto dos vetores $\hat{x}_{\varepsilon}$ tal que:

$$||A\hat{x}_{\varepsilon} - b|| < \varepsilon,$$

onde $||.||$ é alguma norma convenientemente escolhida.

Devido ao mal condicionamento de A não temos possibilidade de escolher o vetor que desejamos neste classe, sem alguma informação adicional sobre x . A função $\psi(x)$ representa o uso dessa informação adicional.

A tese defendida é a de que em qualquer investigação analítica ou científica sempre existe mais informação a respeito de x do que aquela contida em (2.4.1).

A solução $\hat{x}(k)$ do Método das Cristas é uma particularização deste, onde:

$$\psi(x) = k(||x||^2 - R^2),$$

com $||.||$: norma Euclidiana.

C A P Í T U L O I I I

REGRESSÃO SOBRE CRISTAS

3.1 - INTRODUÇÃO

Na resolução de problemas de Regressão Linear, o método geralmente empregado é o de Mínimos Quadrados.

Porém existem alguns casos especiais, que dificul tam bastante a análise. O mal condicionamento da matriz $X'X$, devido a multicolinearidade é um desses problemas. Isto faz com que os resultados obtidos através desse método, às vezes não sejam estáveis.

Um dos objetivos da Análise de Cristas, visto no capítulo anterior era exatamente a de tentar solucionar esse

problema através do acréscimo de uma constante k aos elementos da diagonal da matriz dos coeficientes.

Neste capítulo vamos mostrar como esta técnica pode ser aplicada a problemas de regressão.

3.2 - O MÉTODO DOS MÍNIMOS QUADRADOS

Seja o modelo linear geral:

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad (3.2.1)$$

onde:

X : matriz de regressão $n \times p$, $n > p$ e posto $(X) = p$.

$\underline{Y}$: vetor de respostas $n \times 1$,

$\underline{\beta}$: vetor de parâmetros $p \times 1$,

$\underline{\varepsilon}$: vetor de erros $n \times 1$, com as suposições usuais: $E(\underline{\varepsilon}) = \underline{0}$
e $Var(\underline{\varepsilon}) = \sigma^2 I_n$

O estimador de Mínimos Quadrados de $\underline{\beta}$ é obtido minimizando a soma de quadrados residual:

$$SQR = (\underline{Y} - X\underline{\beta})'(\underline{Y} - X\underline{\beta}),$$

ou seja,

$$\underline{b} = (X'X)^{-1}X'\underline{Y} \quad (3.2.2)$$

que é não tendencioso e de variância mínima uniformemente.

Em nosso estudo vamos considerar por conveniência as variáveis na sua forma padronizada de tal forma que a matriz $X'X$, forneça as correlações entre as variáveis independentes e o vetor $X'Y$ as correlações entre as variáveis independentes e a variável dependente.

Essa forma de tratamento é recomendada por alguns autores quando há indícios de multicolinearidade, ver Marquardt e Snee (1975). No nosso caso tentaremos justificar essa decisão mais tarde.

3.3 - MULTICOLINEARIDADE

Um dos maiores problemas encontrado em regressão é o da multicolinearidade, que é devido ao forte relacionamento entre as variáveis independentes, isto é, quando existem algumas variáveis independentes com alto coeficiente de correlação múltipla.

Devido a esse fato alguns pesquisadores utilizam como medida de multicolinearidade, os elementos diagonais da inversa da matriz $X'X$, que pode ser escrito como:

$$(X'X)^{-1}_{ii} = (1 - R_i^2)^{-1}, \quad i=1, \dots, p$$

denominado VIF: fator de inflação da variância, onde R_i^2 é o quadrado do coeficiente de correlação múltipla da i -ésima variável independente sobre as $p-1$ variáveis restantes. Se a i -ésima estiver fortemente relacionada com outras, R_i^2 está

próximo de 1 e conseqüentemente, $(X'X)^{-1}_{ii}$ será bastante grande.

Outro fato decorrente da multicolinearidade está na magnitude dos autovalores de $X'X$. Se $\lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p > 0$, são esses autovalores, quanto mais mal condicionada for $X'X$, mais λ_p estará próximo de zero. Entretanto a medida mais conhecida para medir mal condicionamento de uma matriz de um sistema é o Número de Condição que neste caso, é definido por:

$$\kappa(X) = \sqrt{\frac{\lambda_1}{\lambda_p}} \geq 1 \quad (3.3.1)$$

que vale exatamente 1 quando a matriz $X'X$ for ortogonal ($X'X = I$).

Segundo Besley e outros (1980), temos uma fraca dependência se $\kappa(X)$ estiver entre 5 e 10, enquanto que uma dependência forte está associada a um número de condição maior que 30.

Vamos mostrar agora um dos problemas na estimação de parâmetros quando a matriz $X'X$ é mal condicionada.

Seja L : distância Euclidiana entre $\underline{b}$ e $\underline{\beta}$, ou seja:

$$L^2 = (\underline{b} - \underline{\beta})'(\underline{b} - \underline{\beta})$$

Logo:

$$\begin{aligned} E(L^2) &= E(\underline{b} - \underline{\beta})'(\underline{b} - \underline{\beta}) = E(\underline{b}'\underline{b}) - \underline{\beta}'\underline{\beta} \\ &= \sigma^2 \text{tr}(X'X)^{-1} + \underline{\beta}'\underline{\beta} - \underline{\beta}'\underline{\beta} \end{aligned}$$

$$\begin{aligned}
 &= \sigma^2 \text{tr}(X'X)^{-1} \\
 &= \sigma^2 \sum \frac{1}{\lambda_i} > \frac{\sigma^2}{\lambda_p}
 \end{aligned} \tag{3.3.2}$$

ou de outra forma:

$$E[\underline{b}'\underline{b}] = \underline{\beta}'\underline{\beta} + \sigma^2 \sum \frac{1}{\lambda_i} > \underline{\beta}'\underline{\beta} + \frac{\sigma^2}{\lambda_p} \tag{3.3.3}$$

e sob a hipótese de normalidade de ε , temos:

$$\text{Var}(L^2) = 2\sigma^4 \text{tr}(X'X)^{-2} = 2\sigma^4 \sum \left(\frac{1}{\lambda_i}\right)^2 > \frac{2\sigma^4}{\lambda_p^2} \tag{3.3.4}$$

Se a matriz $X'X$ for mal condicionada, teremos λ_p pequeno, e isso acarretará num limite inferior bem grande para as expressões acima.

Uma outra forma de verificar a instabilidade de $\underline{b}$ é através do Número de Condição.

Pode-se mostrar, ver Besley e outros (1980) pág. 177, que a sensibilidade da variância de $\underline{b}$, com relação a perturbações nos dados, tem como limite superior $2\kappa(X)$, isto é, uma perturbação de 1% em qualquer elemento de X , pode resultar em uma modificação de até $2\kappa(X)\%$ na variância de $\underline{b}$.

Mostramos dessa forma que o estimador $\underline{b}$ pode apresentar problemas de instabilidade quando a multicolinearidade é muito alta. Mas como $\underline{b}$ é o estimador não tendencioso com a menor variância, parece-nos lógico a procura de outros estimadores na classe dos estimadores tendenciosos. Partimos en

tão do princípio de sacrificarmos um pouco da não tendencisidade pela melhor precisão, procurando um ponto de equilíbrio entre ambas.

3.4 - O ESTIMADOR ATRAVÉS DAS CRISTAS

Seja então o modelo geral dado em (3.2.1):

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$$

e suponhamos $\underline{B}$ um estimador qualquer de $\underline{\beta}$. Vamos denotar por:

$$\Phi = (\underline{Y} - X\underline{B})'(\underline{Y} - X\underline{B})$$

a soma de quadrados residual com relação a $\underline{B}$.

Aplicando o método proposto na Análise de Cristas na função Φ , podemos verificar a existência de soluções distintas de $\underline{b}$, que talvez sejam mais estáveis a pequenas perturbações nos dados do que $\underline{b}$.

Essas soluções são encontradas minimizando Φ e determinando os pontos de mínimos sobre esferas *n*-dimensionais de raio R e centro na origem. Esses pontos descrevem um caminho na superfície da soma de quadrados dos resíduos Φ , chamado de Crista Mínima.

Formalmente, isso é equivalente a minimizar ϕ sujeito à restrição:

$$\underline{B}'\underline{B} = R^2$$

e pelo Método de Lagrange aplicado a função:

$$F = (\underline{Y} - \underline{X}\underline{B})'(\underline{Y} - \underline{X}\underline{B}) + k(\underline{B}'\underline{B} - R^2)$$

onde k : multiplicador de Lagrange, obtemos:

$$\frac{\partial F}{\partial \underline{B}} = -2\underline{X}'\underline{Y} + 2\underline{X}'\underline{X}\underline{B} + 2k\underline{B} = 0$$

$$(\underline{X}'\underline{X} + k\underline{I})\underline{B} = \underline{X}'\underline{Y}$$

e portanto:

$$\underline{B}(k) = (\underline{X}'\underline{X} + k\underline{I})^{-1}\underline{X}'\underline{Y} \quad (3.4.2)$$

é o estimador de $\underline{\beta}$, através do Método das Cristas.

Se $k=0$, $\underline{B}(k)=\underline{b}$, isto é, o estimador de Mínimos Quadrados também pertence a essa classe de estimadores.

Observem que este processo é equivalente a acrescentar uma perturbação k na diagonal de $\underline{X}'\underline{X}$ antes de invertê-la, como já foi comentado no ítem 2.3 do capítulo anterior. Neste caso a matriz $\underline{X}'\underline{X}$ está na forma de correlação e portanto, se aumentarmos a sua diagonal, é certo que para quantidades positivas de k teremos soluções mais estáveis do ponto de vista numérico.

Além disso, pode ser mostrado facilmente que os autovalores da matriz $(\underline{X}'\underline{X} + k\underline{I})$ são $\lambda_i + k$, $i=1, \dots, p$.

Logo, o número de condição do novo sistema, se $k > 0$, é:

$$\kappa(X(k)) = \sqrt{\frac{\lambda_1 + k}{\lambda_p + k}} < \sqrt{\frac{\lambda_1}{\lambda_p}}$$

ou seja, para $k > 0$, o número de condição do sistema (3.4.2) é sensivelmente menor do que o do sistema de equações normais do sistema original, principalmente no caso de λ_p muito pequeno.

Os valores de k podem ser tanto positivos quanto negativos. Entretanto, a medida que temos k negativo se aproximando de λ_p , tanto o número de condição como os coeficientes de $\underline{B}(k)$ tendem a aumentar rapidamente em valor absoluto, pois como já vimos, em $k = -\lambda_p$, a matriz $(X'X + kI)$ é singular devido a:

$$|X'X + kI| = |X'X - \lambda_p I| = 0.$$

Dessa forma é possível perceber a instabilidade da solução de Mínimos Quadrados quando a matriz $X'X$ é mal condicionada, ou seja, se o menor autovalor estiver bem próximo de zero, para os valores de k na vizinhança de zero, $\underline{B}(k)$ poderá ser bastante instável, e conseqüentemente como em $k=0$, $\underline{B}(k) = \underline{b}$, $\underline{b}$ poderá também ser instável.

Como o nosso objetivo é a procura de soluções mais estáveis vamos portanto, estudar apenas as soluções determinadas por valores positivos de k .

3.5 - INTERPRETAÇÃO GEOMÉTRICA

Na Figura 3.5.1 abaixo, damos uma interpretação geométrica da Regressão sobre Cristas, para o caso de um exemplo hipotético envolvendo apenas dois parâmetros β_1 e β_2 .

GRÁFICO 3.5.1

O ponto $\underline{b}$ no centro da elipse é a solução de mínimos quadrados. Em $\underline{b}$, a soma de quadrados residual Φ , é mínima. A pequena elipse é o lugar geométrico dos pontos no espaço paramétrico onde a soma de quadrados residual é constante para algum valor maior que o valor mínimo, isto é, todos os pontos nesta pequena elipse fornecem o mesmo valor para a soma de quadrados residual. O círculo com centro na origem é tangente a pequena elipse no ponto $\underline{B}(k)$. Note que $\underline{B}(k)$ é o vetor de menor comprimento (norma Euclidiana) nesta elipse.

De modo geral, a solução baseada no método das cristas, é a que tem o menor comprimento dentre todos os vetores que apresentam a mesma soma de quadrados residual.

3.6 - MODOS ALTERNATIVOS PARA A DETERMINAÇÃO DE $\underline{B}(k)$

Alguns autores mostraram que $\underline{B}(k)$ pode ser obtido através de vários caminhos.

Mayer e Wilke (1973), mostraram que $\underline{B}(k)$ é uma transformação linear de $\underline{b}$, isto é, existe uma matriz A tal que $\underline{B}(k)$ pode ser escrito como:

$$\underline{B}(k) = A\underline{b} \quad (3.6.1)$$

onde a matriz A é dada por:

$$A = (k(X'X)^{-1} + I)^{-1}. \quad (3.6.2)$$

Marquardt (1970), mostrou que $\underline{B}(k)$ pode ser obtido, aumentando a matriz X com algumas linhas convenientemente escolhidas, e depois determinando o estimador de Mínimos Quadrados do sistema aumentado.

Lindley e Smith (1972), utilizando o enfoque Bayesiano, mostraram que $\underline{B}(k)$ é a média da distribuição a posteriori, se a distribuição a priori de β é normal $N(\underline{0}, \sigma_*^2 I)$, onde k é dado pelo quociente de σ^2 por σ_*^2 .

3.7 - PROPRIEDADES DE $\underline{B}(k)$

Apresentaremos a seguir algumas propriedades de $\underline{B}(k)$ que achamos interessantes. As demonstrações não foram feitas de modo rigoroso, mas apenas, na maioria delas esboçadas.

PROPRIEDADE 3.7.1 - $\underline{B}(k)$, para $k > 0$ é um estimador tendencioso de $\underline{\beta}$:

DEMONSTRAÇÃO:

$$\underline{B}(k) = (X'X + kI)^{-1} X' \underline{y} = (k(X'X)^{-1} + I)^{-1} \underline{b}$$

logo:

$$E(\underline{B}(k)) = (k(X'X)^{-1} + I)^{-1} \underline{\beta} \quad \text{e}$$

$$E(\underline{B}(k)) \neq \underline{\beta} \quad \text{se} \quad k > 0 \quad //$$

PROPRIEDADE 3.7.2 - A variância de $\underline{B}(k)$ é:

$$\text{Var}(\underline{B}(k)) = \sigma^2 P D P',$$

onde

$$D = \text{diag}\left(\frac{\lambda_i}{(\lambda_i + k)^2}\right)$$

e P : transformação ortogonal tal que $P \Lambda P' = X'X$ e $\Lambda = \text{diag}(\lambda_i)$, matriz cuja diagonal são os autovalores de $X'X$.

DEMONSTRAÇÃO:

Sendo $X'X$ simétrica, positiva definida, existe P :
ortogonal tal que:

$$X'X = P\Lambda P' \quad \text{e} \quad \Lambda: \text{matriz dos autovalores.}$$

Portanto:

$$\begin{aligned} \text{Var}(\underline{\tilde{B}}(k)) &= \text{Var}(k(X'X)^{-1} + I)^{-1} \underline{\tilde{b}}) \\ &= \sigma^2 (k(X'X)^{-1} + I)^{-1} (X'X + kI)^{-1} \\ &= \sigma^2 (kP\Lambda^{-1}P' + PP')^{-1} (kPP' + P\Lambda P')^{-1} \\ &= \sigma^2 P(k\Lambda^{-1} + I)^{-1} P' P(kI + \Lambda)^{-1} P' \\ &= \sigma^2 P(kI + \Lambda)^{-2} P' \\ &= \sigma^2 PDP' \end{aligned} \quad //$$

COROLÁRIO:

A variância de cada $B_i(k)$ é dada por:

$$\text{Var}(B_i(k)) = \sigma^2 \frac{\lambda_i}{(k + \lambda_i)^2}, \quad i=1, 2, \dots, p$$

que é decrescente com k .

Esta propriedade mostra que a variabilidade de $B_i(k)$ é menor do que a de b_i , e essa variabilidade diminui a medida que k aumenta, ou λ_p diminui.

PROPRIEDADE 3.7.3 - A variância Total de $\underline{B}(k)$, isto é, o traço de $Var(\underline{B}(k))$ é monótona decrescente com k .

DEMONSTRAÇÃO:

Imediata a partir de:

$$VT(\underline{B}(k)) = \sum Var(B_i(k)) \quad //$$

PROPRIEDADE 3.7.4 - O quadrado do viés de $\underline{B}(k)$, $\Delta^2(\underline{B}(k))$, é monótona crescente com k , e tem por limite superior $\underline{\beta}'\underline{\beta}$.

DEMONSTRAÇÃO:

$$\begin{aligned} \Delta^2(\underline{B}(k)) &= (E(\underline{B}(k)) - \underline{\beta})' (E(\underline{B}(k)) - \underline{\beta}) \\ &= ((k(X'X)^{-1} + I)^{-1} \underline{\beta} - \underline{\beta})' ((k(X'X)^{-1} + I)^{-1} \underline{\beta} - \underline{\beta}) \\ &= \underline{\beta}' P((k\Lambda^{-1} + I)^{-1} - I)' ((k\Lambda^{-1} + I)^{-1} - I) P' \underline{\beta} \\ &= \underline{\beta}' P((k\Lambda^{-1} + I)^{-1} - I)^2 P' \underline{\beta} \\ &= \underline{\alpha}' ((k\Lambda^{-1} + I)^{-1} - I)^2 \underline{\alpha}, \quad \text{onde } \underline{\alpha} = P' \underline{\beta} \\ &= \sum \frac{\alpha_i^2 k^2}{(k + \lambda_i)^2} \\ &= \sum \frac{\alpha_i^2}{(\frac{\lambda_i}{k} + 1)^2}, \quad \text{para } k > 0 \end{aligned}$$

donde podemos concluir que $\Delta^2(\underline{B}(k))$ é monótona crescente com k , e

$$\lim_{k \rightarrow \infty} \Delta^2(\underline{B}(k)) = \Sigma \alpha_i^2 = \underline{\alpha}' \underline{\alpha} = \underline{\beta}' P P' \underline{\beta} = \underline{\beta}' \underline{\beta}. \quad //$$

3.8 - O ERRO QUADRÁTICO MÉDIO DE $\underline{B}(k)$: EQM ($\underline{B}(k)$)

Seja:

$$L^2(k) = (\underline{B}(k) - \underline{\beta})' (\underline{B}(k) - \underline{\beta}),$$

e

$$E(L^2(k)) = E((\underline{B}(k) - \underline{\beta})' (\underline{B}(k) - \underline{\beta})) \quad (3.8.1)$$

o Erro Quadrático Médio de $\underline{B}(k)$.

Fazendo:

$$\underline{\beta} = (k(X'X)^{-1} + I)^{-1} \underline{b} - (k(X'X)^{-1} + I)^{-1} \underline{\beta} + \underline{\beta}$$

$$\underline{Z} = (k(X'X)^{-1} + I)^{-1}$$

e substituindo em (3.8.1) temos, após certa manipulação algébrica:

$$\begin{aligned} E(L^2(k)) &= E((\underline{b} - \underline{\beta})' \underline{Z}' \underline{Z} (\underline{b} - \underline{\beta})) + (\underline{Z} \underline{\beta} - \underline{\beta})' (\underline{Z} \underline{\beta} - \underline{\beta}) \\ &= \sigma^2 \text{tr}(X'X)^{-1} \underline{Z}' \underline{Z} + \underline{\beta}' (\underline{Z} - I)' (\underline{Z} - I) \underline{\beta} \\ &= \sigma^2 (\text{tr}(X'X + kI)^{-1} - k \text{tr}(X'X + kI)^{-2}) + \\ &\quad + k^2 \underline{\beta}' (X'X + kI)^{-2} \underline{\beta} \end{aligned}$$

$$= \sigma^2 \sum \frac{\lambda_i}{(\lambda_i + k)^2} + \sum \frac{k^2 \alpha_i^2}{(\lambda_i + k)^2} \quad (3.8.2)$$

$$= VT(\underline{B}(k)) + \Delta^2(\underline{B}(k)) \quad (3.8.3)$$

ou seja, o $EQM(\underline{B}(k))$ pode também ser decomposto na soma da Variância Total mais o quadrado do Viês.

Observando então as propriedades (3.7.3) e (3.7.4), notamos que enquanto a Variância Total diminui, o quadrado do Viês aumenta com k , o que faz com que o $EQM(\underline{B}(k))$ tenha algumas propriedades interessantes, como veremos a seguir.

PROPRIEDADE 3.8.1 - A inclinação da Variância Total na vizinhança de $k=0$ é negativa e aumenta em valor absoluto quando $\lambda_p \rightarrow 0$.

DEMONSTRAÇÃO:

$$VT(\underline{B}(k)) = \sigma^2 \sum \frac{\lambda_i}{(\lambda_i + k)^2}$$

$$\frac{d}{dk} VT(\underline{B}(k)) = -2\sigma^2 \sum \frac{\lambda_i}{(\lambda_i + k)^3}$$

de onde:

$$\lim_{k \rightarrow 0} \frac{d}{dk} VT(\underline{B}(k)) = -2\sigma^2 \sum \frac{1}{\lambda_i^2}$$

e

$$\lim_{\substack{k \rightarrow 0 \\ \lambda \rightarrow 0 \\ p}} \frac{d}{dk} VT(\underline{B}(k)) = -\infty$$

//

Esta propriedade mostra que podemos ter grande re
dução na Variância Total para pequenos valores de k e λ_p .

Por outro lado, é preciso que o viés nas mesmas
condições não seja muito grande. É o que veremos na proposi
ção seguinte.

PROPOSIÇÃO 3.8.2 - A taxa de crescimento de $\Delta^2(\underline{B}(k))$ na vizi
nhança de zero é quase nula.

DEMONSTRAÇÃO:

$$\Delta^2(\underline{B}(k)) = \Sigma \frac{k^2 \alpha_i^2}{(\lambda_i + k)^2}$$

$$\frac{d}{dk} \Delta^2(\underline{B}(k)) = 2 \Sigma \frac{k \lambda_i \alpha_i^2}{(\lambda_i + k)^3}$$

$$\lim_{k \rightarrow 0} \frac{d}{dk} \Delta^2(\underline{B}(k)) = 0$$

//

Pelas propriedades acima, observamos que devem
existir valores de k que ofereçam boas propriedades com rela
ção ao $EQM(\underline{B}(k))$.

O teorema seguinte vai nos garantir essa existên
cia.

TEOREMA 3.8.1 - (EXISTÊNCIA)

Se $\beta' \beta < \infty$, então existe $k > 0$ tal que:

$$EQM(\underline{B}(k)) < EQM(\underline{b}) = \sigma^2 \sum \frac{1}{\lambda_i}$$

DEMONSTRAÇÃO:

De (3.8.2) temos:

$$EQM(\underline{B}(k)) = E(L^2(k)) = VT(\underline{B}(k)) + \Delta^2(\underline{B}(k))$$

$$= \sigma^2 \sum \frac{\lambda_i}{(\lambda_i + k)^2} + \sum \frac{k^2 \sigma_i^2}{(\lambda_i + k)^2}$$

$$\frac{d}{dk} EQM(\underline{B}(k)) = -2\sigma^2 \sum \frac{\lambda_i}{(\lambda_i + k)^3} + 2k \sum \frac{\lambda_i}{(\lambda_i + k)^3}$$

Logo:

$$\lim_{k \rightarrow 0} \frac{d}{dk} EQM(\underline{B}(k)) = -2\sigma^2 \sum \left(\frac{1}{\lambda}\right)^2 < 0$$

que indica a inclinação negativa de $EQM(\underline{B}(k))$ nas vizinhanças de $k=0$.

Logo, isto vai implicar na existência de valores $k > 0$ que torna $EQM(\underline{B}(k)) < EQM(\underline{B}(0)) = EQM(\underline{b})$.

Alguns desses valores são encontrados, resolvendo a inequação:

$$\frac{d}{dk} EQM(\underline{B}(k)) \leq 0$$

de onde obtemos:

$$k \leq \frac{p\sigma^2}{\beta'\beta} \quad //$$

Infelizmente, esses valores de k são delimitados por uma função dos parâmetros σ^2 e β , sendo portanto difícil a sua utilização na prática.

Este teorema evidencia também o fato de que, quanto mais pequeno for λ_p , maior pode ser a vantagem obtida com a solução $B(k)$ com relação a b .

Porém essa vantagem somente pode ser aproveitada a partir do momento em que conhecemos o valor do parâmetro k . Entretanto, esse valor de k , cujo ganho é máximo, é obtido fazendo:

$$\frac{d}{dk} EQM(B(k)) = 0$$

de onde obtemos:

$$k_{ot} = \frac{p\sigma^2}{\beta'\beta}$$

ou seja, o valor de k que minimiza EQM é função dos parâmetros σ^2 e β .

Esse fato tem-se constituído no maior motivo para as críticas ao método. O problema de determinar um valor para k , que satisfaça as condições ótimas acima tem movimentado os pesquisadores da área e vários critérios foram propostos.

No capítulo seguinte vamos enunciar alguns desses critérios existentes, onde daremos mais ênfase ao critério de nominado de Traço de Crista.

C A P Í T U L O I V

CRITÉRIOS PARA SELEÇÃO DE k

4.1 - INTRODUÇÃO

O teorema 3.7.1 garante a existência de valores k onde o Erro Quadrático Médio de $\underline{B}(k)$ é menor do que o Erro Quadrático Médio de $\underline{b}$. Isto nos permite concluir que se o critério para a escolha do melhor estimador for o que possuir o menor EQM , para os valores de k mencionados acima, $\underline{B}(k)$ é melhor do que $\underline{b}$.

Entretanto, esses valores dependem dos parâmetros σ^2 e β , o que torna difícil a sua obtenção na prática.

Para a escolha de um valor para k , vários crité

rios foram propostos, com a maioria deles obtendo k como função dos estimadores $\hat{\sigma}^2$ e $\underline{b}$, onde $\hat{\sigma}^2$ é o estimador usual de σ^2 .

Infelizmente os k 's encontrados através desses critérios são variáveis aleatórias, o que faz com que a partir de obtido $\hat{k}_0$ não sabemos de fato se $\underline{B}(\hat{k}_0)$ possui a propriedade descrita no teorema da existência, isto é, se $EQM(\underline{B}(\hat{k}_0))$ é realmente menor do que $EQM(\underline{b})$.

Isto fez com que vários trabalhos de comparações utilizando simulação fossem efetuados, visando a procura de um critério ótimo, ver por exemplo, Dempster e outros (1975), Mc Donald e Galarneau (1975), Hoerl, Kennard e Baldwin (1975), Wichern e Churchill (1978), etc.

Apresentaremos a seguir alguns dos critérios dentre os que achamos possuir alguma evidência empírica, tendo por base o que foi exposto até agora.

4.2 - CRITÉRIO BASEADO NO TEOREMA DA EXISTÊNCIA

Vimos através do Teorema da Existência que o valor de k que minimiza o Erro Quadrático Médio de $\underline{B}(k)$ é dado por:

$$k_{\text{ót}} = \frac{p\sigma^2}{\underline{\beta}'\underline{\beta}}$$

Logo, um valor para k pode ser obtido fazendo:

$$\hat{k} = \frac{p\hat{\sigma}^2}{\underline{\underline{b}}'\underline{\underline{b}}} \quad (4.2.1)$$

onde $\hat{\sigma}^2$: estimador não tendencioso de σ^2 .

Este critério foi proposto por Hoerl, Kennard e Baldwin (1975), onde apresentou bons resultados em um trabalho de simulação, quando comparado em termos de *EQM*, com o estimador obtido através dos Mínimos Quadrados. Entretanto, em outros trabalhos comparativos, não apresentou o mesmo rendimento sendo considerado como um critério que deve ser melhor estudado antes de ser aplicado. Ver Wichern e Churchill (1978).

O principal problema desse critério é o fato de que às vezes $\underline{\underline{b}}'\underline{\underline{b}}$ pode ser muito grande devido a multicolinearidade, fazendo com que $\hat{k}$ seja muito pequeno, e consequentemente $\underline{\underline{B}}(\hat{k})$ próximo de $\underline{\underline{b}}$.

Este fato pode ter originado o seguinte critério:

4.3 - $\underline{\underline{B}}(k)$ ATRAVÉS DA ESTIMAÇÃO DE $\underline{\underline{\beta}}'\underline{\underline{\beta}}$

Este critério tem por base, o fato que:

$$E(\underline{\underline{b}}'\underline{\underline{b}}) = \underline{\underline{\beta}}'\underline{\underline{\beta}} + \sigma^2 \sum \frac{1}{\lambda_i}$$

como foi visto em 3.2, isto é, $\underline{\underline{b}}'\underline{\underline{b}}$ superestima $\underline{\underline{\beta}}'\underline{\underline{\beta}}$ gerando problemas na estimação de k através de 4.2.

Isto fez com que McDonald e Galarneau (1975), introduzissem o seguinte critério:

Seja

$$Q = \underline{b}' \underline{b} - \hat{\sigma}^2 \sum \frac{1}{\lambda_i}$$

isto é, Q é um estimador não tendencioso de $\beta' \beta$.

Escolha então k tal que:

$$\underline{B}'(k) \underline{B}(k) = Q, \quad \text{se } Q > 0, \quad (4.3.1)$$

caso contrário, faça $\underline{B}(k) = \underline{b}$.

Se através dos dados encontrarmos evidências que $\underline{b}$ superestima β , ou seja, $\underline{b}' \underline{b}$ é muito maior do que $\beta' \beta$, $\underline{B}(k)$ satisfazendo (4.3.1) pode ser um bom estimador de β .

Bons resultados através desse critério foram obtidos nos trabalhos de McDonald e Galarneau (1975) e Wichern e Churchill (1978), nos casos onde a multicolinearidade é alta.

4.4 - CRITÉRIO ATRAVÉS DO ENFOQUE BAYESIANO

Um outro método para a obtenção do estimador $\underline{B}(k)$ é através do Enfoque Bayesiano. Isto é, se a distribuição a priori de β é normal $N(\underline{0}, \sigma_*^2 I)$, então pode-se demonstrar (ver Lindley e Smith (1972)) que o estimador $\underline{B}(k)$ é a média da distribuição a posteriori, onde:

$$k = \frac{\sigma^2}{\sigma_*^2}. \quad (4.4.1)$$

Tendo em vista esse fato, Lawless e Wang (1976),

propuseram o seguinte critério para a escolha de k :

$$\hat{k} = \frac{p\hat{\sigma}^2}{\sum \lambda_i \hat{\alpha}_i^2} \quad (4.4.2)$$

onde $\hat{\alpha} = P'b$ e P : matriz dos autovetores de $X'X$.

Lawless e Wang mostraram que esse estimador é mais eficiente do que b , em termos de *EQM*.

No entanto, Wichern e Churchill (1978) comparam esse e outros critérios, inclusive 4.2 e 4.3, com o de Mínimos Quadrados, e não o recomendam sem antes de mais estudos serem desenvolvidos.

No Apêndice, veremos alguns resultados de alguns trabalhos de simulação realizados, comparando alguns desses critérios aqui enunciados.

Todos os critérios vistos até agora, fornecem um único valor para k . Existem no entanto, outros que procuram encontrar k através do exame do Traço da Crista, já mencionada anteriormente. Esses critérios geralmente fornecem mais de um valor para k . Na verdade trata-se de encontrar intervalos onde supostos estimadores $B(k)$'s mais estáveis podem se encontrar.

Existem portanto dois propósitos distintos mas não conflitantes nesses critérios. Os critérios acima que buscam uma solução em termos de *EQM*, e o que vamos enunciar a seguir, que procura uma solução mais estável.

Vamos enunciar aqui apenas o critério sugerido por

Hoerl e Kennard (1970a).

4.5 - CRITÉRIO BASEADO NO TRAÇO DA CRISTA

O Traço da Crista é o gráfico dos coeficientes $B_i(k)$, $i=1,2,\dots,p$, contra k .

Para cada $B_i(k)$ teremos uma curva que inicia em b_i , quando $k=0$ e tende a zero, quando k tende a $+\infty$.

Este processo, como já foi dito anteriormente, tem a vantagem de mostrar num mesmo gráfico, a variação de todos os coeficientes $B_i(k)$'s simultaneamente, independentemente do número de variáveis envolvidas no sistema.

Às vezes costuma-se colocar no mesmo gráfico, com escala diferente, a soma de quadrados dos resíduos com relação a $\underline{B}(k)$, que é dada por:

$$\Phi(k) = (\underline{Y} - \underline{X}\underline{B}(k))'(\underline{Y} - \underline{X}\underline{B}(k)) \quad (4.5.1)$$

$$= \underline{Y}'\underline{Y} - \underline{B}'(k)\underline{X}'\underline{Y} - \underline{Y}'\underline{X}\underline{B}(k) + \underline{B}'(k)\underline{X}'\underline{X}\underline{B}(k)$$

Entretanto a maioria utiliza um gráfico separado para evitar confusão entre os coeficientes e o resíduo.

O fato de trabalharmos com as variáveis na forma padronizada, evita que tenhamos uma diferença às vezes brutal na escala entre os coeficientes, que dessa forma podem ser diretamente comparados.

Através desse gráfico é possível então "perceber"

a sensibilidade dos coeficientes, com relação a pequenas perturbações na diagonal da matriz $X'X$. Assim, se para pequenas perturbações, os coeficientes sofrerem grandes mudanças sem que a soma de quadrados residual se altere muito, teremos problemas de instabilidade no sistema.

Já vimos também que nos casos onde a multicolinearidade é alta, teremos problemas de mal condicionamento na matriz $X'X$, gerando com isso a instabilidade na solução dos Mínimos Quadrados.

Essa instabilidade pode então ser verificada no gráfico dos coeficientes $B_i(k)$'s com pequenas variações de k .

Nos casos onde a multicolinearidade não é alta, os coeficientes se comportam geralmente da seguinte forma:

- i) As curvas dos $B_i(k)$'s decrescem suavemente com k , sem grandes variações, enquanto que o resíduo cresce sensivelmente.
- ii) Alguns $B_i(k)$'s podem apresentar uma variação mais acentuada para pequenos valores de k , estabilizando-se logo, enquanto que o resíduo também cresce rapidamente.

Com isto queremos dizer que em problemas relativamente estáveis, a solução muda pouco com pequenas perturbações nos dados, e embora em alguns casos a solução possa mudar bastante, o resíduo também aumenta sensivelmente, o que

não acontece com problemas mal condicionados onde podemos ter soluções bastante diferentes uma da outra, sem que as somas de quadrados residual difiram muito entre si, e da soma de quadrados residual mínima.

Cabe ressaltar, que estes aspectos foram observados de um número grande de problemas analisados, e não analiticamente. Podemos apresentar uma análise mais rigorosa apenas no caso extremo. Vamos mostrar o que acontece com os coeficientes no caso em que o sistema é completamente ortogonal, isto é, quando $X'X=I$. É evidente que se queremos achar uma solução que seja estável, devemos levar em consideração este aspecto da questão.

Neste caso,

$$\begin{aligned}\underline{\underline{B}}(k) &= (X'X+kI)^{-1}X'\underline{\underline{y}} = (X'X+kI)^{-1}X'X\underline{\underline{b}} \\ &= (I+kI)^{-1}\underline{\underline{b}} = \frac{\underline{\underline{b}}}{(1+k)}\end{aligned}$$

que depende de k através da função:

$$\frac{1}{(1+k)}$$

que é bem comportada e decresce suavemente a zero quando k cresce para $+\infty$. O gráfico seguinte nos dá uma idéia do comportamento dos coeficientes em função de k .

GRÁFICO 4.5.1

Entretanto é de se esperar, que nos problemas com moderada multicolinearidade a variação seja bem mais acentuada, sem que isso acarrete em problemas na estimação de β , através de $\hat{b}$.

Com relação a soma de quadrados residual temos o seguinte:

$$\begin{aligned}
 \Phi(k) &= (\underline{Y} - X\underline{B}(k))'(\underline{Y} - X\underline{B}(k)) \\
 &= (\underline{Y} - X(X'X + kI)^{-1}X'\underline{Y})'(\underline{Y} - X(X'X + kI)^{-1}X'\underline{Y}) \\
 &= (\underline{Y} - X(I + kI)^{-1}X'\underline{Y})'(\underline{Y} - X(I + kI)^{-1}X'\underline{Y}) \\
 &= (\underline{Y} - \frac{XX'\underline{Y}}{(1+k)})'(\underline{Y} - \frac{XX'\underline{Y}}{(1+k)})
 \end{aligned}$$

$$= \underline{\underline{Y}}' \underline{\underline{Y}} - 2 \frac{\underline{\underline{b}}' \underline{\underline{b}}}{(1+k)} + \frac{\underline{\underline{b}}' \underline{\underline{b}}}{(1+k)^2}$$

$$= \underline{\underline{Y}}' \underline{\underline{Y}} - \frac{1+2k}{(1+k)^2} \underline{\underline{b}}' \underline{\underline{b}}$$

Portanto o acréscimo na soma de quadrados residual em um sistema ortogonal é determinado pela função:

$$g(k) = \frac{1+2k}{(1+k)^2}$$

Uma forma um pouco mais fácil de interpretar é da da por:

$$\begin{aligned} \Phi(k) - \Phi(0) &= \underline{\underline{Y}}' \underline{\underline{Y}} - g(k) \underline{\underline{b}}' \underline{\underline{b}} - (\underline{\underline{Y}}' \underline{\underline{Y}} - \underline{\underline{b}}' \underline{\underline{b}}) \\ &= (1-g(k)) \underline{\underline{b}}' \underline{\underline{b}} \\ &= \left(\frac{k}{1+k} \right)^2 \underline{\underline{b}}' \underline{\underline{b}} \end{aligned}$$

isto é, a diferença na soma de quadrados dos resíduos é determinada pela expressão:

$$f(k) = \left(\frac{k}{1+k} \right)^2$$

que tende a 1 quando k tende ao $+\infty$.

GRÁFICO 4.5.2

Observando o comportamento dessa função podemos ver que quando k está próximo de zero a inclinação de $f(k)$ é suave, aumentando gradativamente a medida que k aumenta. Entretanto essa curva tem um ponto de inflexão em $k=0.5$, o que acarreta uma diminuição dessa tendência depois desse ponto.

A diferença máxima observada é $\tilde{b}'\tilde{b}$, o que implica em que a soma de quadrados residual máxima é:

$$\Phi(k) = \tilde{Y}'\tilde{Y}$$

e como estamos tratando com as variáveis na forma padronizada, o máximo que $\Phi(k)$ pode atingir é 1.

Com base nesses fatos, Hoerl e Kennard (1970b), sugeriram um critério para a escolha de k .

Segundo esses autores, um bom valor pode ser obtido, olhando para o Traço da Crista e escolhendo o k onde:

- Os coeficientes já se estabilizaram e terão o comportamento semelhante ao de um sistema ortogonal.
- Os coeficientes com sinais aparentemente incorretos em $k=0$, terão mudado para o sinal apropriado.
- Os resíduos não serão muito elevados, com relação ao resíduo obtido em $k=0$.

Com essas indicações, os autores afirmam não ser difícil conseguir bons valores para k , através da análise do Traço, para a grande maioria dos problemas. Segundo eles é necessário apenas um pouco de prática para achar um valor para k que resulte em uma solução mais estável.

Entretanto, um aspecto importante deve ser ressaltado. Nem sempre um sistema mal condicionado vai resultar numa solução de Mínimos Quadrados com problemas sérios de instabilidade.

Consequentemente, nem sempre, $\underline{B}(k)$ obtido através do Traço deverá diferir de $\underline{b}$. Às vezes, a melhor solução através do Traço está exatamente em $k=0$, ou seja, quando $\underline{B}(k)=\underline{b}$, embora a multicolinearidade seja forte.

Nesse sentido o critério do Traço da Crista tem a

sensibilidade de detectar situações onde embora a multicolinearidade seja alta, o estimador de Mínimos Quadrados ainda é o melhor.

É preciso contudo, tomar muito cuidado com a escala adotada para k . Um gráfico extremamente variável numa escala pode parecer bastante comportado em outra. Para evitar esse problema é conveniente às vezes utilizar a porcentagem da variação como medida.

Como a matriz $X'X$ está na forma de correlação, e sua diagonal é constituída apenas de 1's, iremos fazer k variar de 0 a 0,8, ou seja, aumentamos a diagonal de $X'X$ em no máximo 80%, que acreditamos ser mais do que o suficiente para detectar instabilidades em qualquer problema de regressão.

A seguir vamos ver alguns exemplos.

4.6 - EXEMPLOS

EXEMPLO 4.6.1 - ESTUDO DA PERFORMANCE DOS SUPERVISORES

Este exemplo é analisado no livro do Chatterjee e Price (1977), pág. 59-64, e refere-se a um levantamento sobre funcionários de escritório de uma grande organização financeira incluindo questões relacionadas com a satisfação dos empregados com seus supervisores. O nosso interesse para com este exemplo é devido ao fato de que ele não apresenta nenhum problema de multicolinearidade como poderá ser constatado através do Número de Condição. Veremos entretanto, como a nossa

técnica funciona nestes casos.

Foram medidas seis variáveis relacionadas com questões que tratam do relacionamento entre os empregados e seus supervisores e os dados estão na tabela abaixo:

TABELA 4.6.1 - DADOS NÃO PADRONIZADOS

X_1	X_2	X_3	X_4	X_5	X_6	Y
51.	30.	39.	61.	92.	45.	43.
64.	51.	54.	63.	73.	47.	63.
70.	68.	69.	76.	86.	48.	71.
63.	45.	47.	54.	84.	35.	61.
78.	56.	66.	71.	83.	47.	81.
55.	49.	44.	54.	49.	34.	43.
67.	42.	56.	66.	68.	35.	58.
75.	50.	55.	70.	66.	41.	71.
82.	72.	67.	71.	83.	31.	72.
61.	45.	47.	62.	80.	41.	67.
53.	53.	58.	58.	67.	34.	64.
60.	47.	39.	59.	74.	41.	67.
62.	57.	42.	55.	63.	25.	69.
83.	83.	45.	59.	77.	35.	68.
77.	54.	72.	79.	77.	46.	77.
90.	50.	72.	60.	54.	36.	81.
85.	64.	69.	79.	79.	63.	74.
60.	65.	75.	55.	80.	60.	65.
70.	46.	57.	75.	85.	46.	65.
58.	68.	54.	64.	78.	52.	50.
40.	33.	34.	43.	64.	33.	50.
61.	52.	62.	66.	80.	41.	64.
66.	52.	50.	63.	80.	37.	53.
37.	42.	58.	50.	57.	49.	40.
54.	42.	48.	66.	75.	33.	63.
77.	66.	63.	88.	76.	72.	66.
75.	58.	74.	80.	78.	49.	78.
57.	44.	45.	51.	83.	38.	48.
85.	71.	71.	77.	74.	55.	85.
82.	39.	59.	64.	78.	39.	82.

A matriz de correlação entre as variáveis independentes é dada por:

MATRIZ DE CORRELAÇÃO

$$\begin{pmatrix} 1.00000 & 0.55829 & 0.59674 & 0.66920 & 0.18771 & 0.22458 \\ 0.55829 & 1.0000 & 0.49333 & 0.44548 & 0.14723 & 0.34329 \\ 0.59674 & 0.49333 & 1.00000 & 0.64031 & 0.11596 & 0.53162 \\ 0.66920 & 0.44548 & 0.64031 & 1.00000 & 0.37688 & 0.57419 \\ 0.18771 & 0.14723 & 0.11596 & 0.37688 & 1.00000 & 0.28334 \\ 0.22458 & 0.34329 & 0.53162 & 0.57419 & 0.28334 & 1.00000 \end{pmatrix}$$

e o vetor de autovalores é:

$$(3.16922 \quad 1.00635 \quad 0.76291 \quad 0.55252 \quad 0.31725 \quad 0.19176)$$

Uma rápida análise da matriz de correlação revela que não temos correlações altas, bem como não temos autovalores muito pequenos que causem problemas. O número de condição espectral definido por:

$$\kappa(X) = \sqrt{\frac{\lambda_1}{\lambda_p}},$$

onde os λ_i 's são respectivamente maior e o menor autovalor da matriz de correlação acima, fica:

$$\kappa(X) = \sqrt{\frac{3.16922}{0.19176}} = 4.06533,$$

que não é grande o bastante para causar problemas.

A equação estimada através do método dos Mínimos Quadrados é:

$$\hat{Y} = 10.787 + 0.613X_1 - 0.073X_2 + 0.320X_3 + 0.081X_4 + 0.038X_5 - 0.217X_6$$

(0.93) (3.81) (-0.54) (1.91) (0.37) (0.25) (-1.22)

onde os números entre parêntesis são os correspondentes aos valores de t .

Embora este seja um problema onde esta técnica não deveria apresentar bons resultados, vamos assim mesmo aplicar os critérios vistos até agora, principalmente o Traço da Crista, e analisar a estabilidade do sistema.

O Traço e os coeficientes são dados na Tabela 4.6.2.

Numa simples inspeção visual dos Gráficos 4.6.1 e 4.6.2, observamos os seguintes aspectos:

- i) O coeficiente de X_1 é o que tem maior influência para toda a variação de k , e tem uma variação um pouco mais acentuada do que os demais.
- ii) O coeficiente da variável X_4 é quase nulo em $k=0$, entretanto ele cresce inicialmente com k para depois estabilizar para k maior que 0,3. Isto pode significar que esta variável deve estar sendo mascarada por outra.
- iii) Os coeficientes de X_2 e X_5 são quase nulos, isto é, não são significantes para nenhum valor de k .

iv) Os coeficientes de X_3 e X_6 são estáveis, não sofrendo com as variações de k .

v) Outro ponto interessante está no gráfico dos resíduos. Observem que para k entre 0 e 0,1 a soma de quadrados residual se manteve praticamente constante. Isto quer dizer que qualquer solução $\underline{B}(k)$ com k nesse intervalo será tão boa quanto a solução de Mínimos Quadrados em termos de precisão.

TABELA 4.6.2

k 18	$B1(k)$ 18	$B2(k)$ 18	$B3(k)$ 18	$B4(k)$ 18	$B5(k)$ 18	$B6(<)$ 18	SQR 18
0.0	0.6707	-0.0734	0.3088	0.0698	0.0311	-0.1834	0.2673
0.01	0.6577	-0.0671	0.3056	0.0772	0.0298	-0.1830	0.2674
0.02	0.6456	-0.0613	0.3024	0.0839	0.0286	-0.1823	0.2677
0.03	0.6343	-0.0559	0.2993	0.0898	0.0275	-0.1813	0.2681
0.04	0.6237	-0.0507	0.2964	0.0951	0.0265	-0.1801	0.2686
0.05	0.6136	-0.0459	0.2934	0.0999	0.0256	-0.1787	0.2693
0.06	0.6041	-0.0413	0.2906	0.1042	0.0248	-0.1771	0.2700
0.07	0.5951	-0.0370	0.2879	0.1081	0.0240	-0.1755	0.2708
0.08	0.5865	-0.0329	0.2852	0.1116	0.0233	-0.1738	0.2717
0.09	0.5783	-0.0291	0.2827	0.1148	0.0227	-0.1720	0.2727
0.10	0.5705	-0.0254	0.2801	0.1178	0.0221	-0.1701	0.2737
0.20	0.5072	0.0031	0.2585	0.1360	0.0184	-0.1505	0.2857
0.30	0.4611	0.0218	0.2415	0.1434	0.0168	-0.1319	0.2995
0.40	0.4251	0.0347	0.2277	0.1461	0.0162	-0.1156	0.3138
0.50	0.3958	0.0438	0.2161	0.1463	0.0161	-0.1016	0.3279
0.60	0.3712	0.0504	0.2062	0.1452	0.0162	-0.0895	0.3416
0.70	0.3502	0.0552	0.1974	0.1434	0.0164	-0.0791	0.3549
0.80	0.3319	0.0588	0.1897	0.1413	0.0167	-0.0702	0.3676

GRÁFICO 4.6.1 - TRAÇO DA CRISTA

GRÁFICO 4.6.2 - RESÍDUO

De modo geral, os coeficientes $B_i(k)$'s são estáveis, revelando que o estimador de Mínimos Quadrados não apresenta problemas de estabilidade, o que reflete as condições já percebidas anteriormente. Neste caso o estimador $\underline{b}$ é o melhor estimador de β .

Os critérios 4.2, 4.3 e 4.4 apresentam aproximadamente o mesmo valor para k , ou seja:

$$\hat{k} = 0,1$$

cujos $B_i(\hat{k})$'s são comparados com os b_i 's, como podemos ver na Tabela 4.6.3.

TABELA 4.6.3

i	1	2	3	4	5	6
b_i 's	0,6707	-0,0734	0,3089	0,0698	0,0312	-0,1835
B_i 's	0,5706	-0,0254	0,2802	0,1178	0,0222	-0,1702
Var. Perc.	14,92%	65,40%	9,29%	68,77%	28,85%	7,25%

onde a variação percentual é dada por: $1 - \frac{B_i}{b_i}$.

A soma de quadrados residual aumentou de 0,2674 para 0,2737, ou seja, 3,36%.

Observem que as maiores diferenças nos coeficientes individuais estão nos casos em que são quase nulos, ou

seja, são os que podem ser considerados não significantes, en quanto que para os não nulos a diferença é pequena.

EXEMPLO 4.6.2

Este exemplo foi tirado do livro do Draper e Smith (1981), pág. 352-360.

Os dados abaixo representam as variáveis:

Y : É a quantidade de água utilizada (galões).

X_1 : Temperatura média mensal ($^{\circ}F$).

X_2 : Quantidade produzida (M libras).

X_3 : Número de máquinas operando por dia no mês.

X_4 : Número de pessoas na lista de pagamento mensal.

X_5 : Um número aleatório de dois dígitos.

TABELA 4.6.4 - DADOS ORIGINAIS NÃO PADRONIZADOS

X_1	X_2	X_3	X_4	X_5	Y
55.8	7107	21	129	52	3067
65.2	6373	22	141	68	2828
70.9	6796	22	153	29	2891
77.4	9208	20	166	23	2994
79.3	14792	25	193	40	3082
81.0	14564	23	189	14	3898
71.9	11964	20	175	96	3502
63.9	13526	23	186	94	3060
54.5	12656	20	190	54	3211
39.5	14119	20	187	37	3286
44.5	16691	22	195	42	3542
43.6	14571	19	206	22	3125
56.0	13619	22	198	28	3022
64.7	14575	22	192	7	2922
73.0	14556	21	191	42	3950
78.9	18573	21	200	33	4488
79.4	15618	22	200	92	3295

A matriz de correlação é:

$$\begin{pmatrix} 1,0000 & -0,0241 & 0,4376 & -0,0821 & 0,1076 \\ -0,0241 & 1,0000 & 0,1057 & 0,9185 & -0,1115 \\ 0,4376 & 0,1057 & 1,0000 & 0,0319 & 0,0377 \\ -0,0821 & 0,9185 & 0,0319 & 1,0000 & -0,1590 \\ 0,1076 & -0,1115 & 0,0377 & -0,1590 & 1,0000 \end{pmatrix}$$

Existindo correlação alta apenas entre X_2 e X_4 , cujo coeficiente é de: 0,9185.

Os autovalores da matriz acima são:

$$1,96866 \quad 1,46294 \quad 0,94196 \quad 0,54920 \quad 0,07724$$

Apesar de $\lambda_5=0,07724$ ser relativamente pequeno, o Número de Condição Espectral, é:

$$\kappa(X) = \sqrt{\frac{1,96866}{0,07724}} = 5,$$

que não é alto, e portanto não deveria afetar a análise através de $\underline{b}$.

Entretanto, verificando o Traço da Crista, achamos algumas coisas interessantes.

O Traço da Crista, bem como a tabela com os coeficientes $B_i(k)$'s são dados na Tabela 4.6.5. e Gráficos 4.6.3 e 4.6.4.

TABELA 4.6.5

k 18	$B1(k)$ 18	$B2(k)$ 18	$B3(k)$ 18	$B4(k)$ 18	$B5(k)$ 18	SQR 18
0.0	0.4308	1.7110	-0.4193	-1.1375	-0.0846	0.2227
0.01	0.4252	1.5456	-0.4008	-0.9755	-0.0766	0.2269
0.02	0.4195	1.4139	-0.3850	-0.8470	-0.0702	0.2364
0.03	0.4138	1.3064	-0.3712	-0.7428	-0.0650	0.2481
0.04	0.4081	1.2169	-0.3589	-0.6567	-0.0607	0.2607
0.05	0.4025	1.1413	-0.3479	-0.5843	-0.0570	0.2732
0.06	0.3970	1.0765	-0.3378	-0.5227	-0.0538	0.2855
0.07	0.3915	1.0203	-0.3285	-0.4697	-0.0511	0.2973
0.08	0.3862	0.9710	-0.3200	-0.4236	-0.0487	0.3084
0.09	0.3809	0.9275	-0.3119	-0.3832	-0.0466	0.3190
0.10	0.3758	0.8888	-0.3044	-0.3475	-0.0447	0.3290
0.20	0.3309	0.6501	-0.2476	-0.1377	-0.0333	0.4036
0.30	0.2955	0.5322	-0.2094	-0.0454	-0.0278	0.4514
0.40	0.2669	0.4598	-0.1812	0.0040	-0.0246	0.4863
0.50	0.2435	0.4098	-0.1594	0.0334	-0.0225	0.5140
0.60	0.2240	0.3725	-0.1420	0.0519	-0.0210	0.5370
0.70	0.2074	0.3432	-0.1278	0.0640	-0.0199	0.5568
0.80	0.1931	0.3193	-0.1160	0.0721	-0.0189	0.5743

A variação dos Gráficos 4.6.3 e 4.6.4 agora é bem mais acentuada, revelando um pouco mais de instabilidade. Por exemplo, o comportamento dos coeficientes das variáveis X_2 e X_4 é bastante variável para pequenos valores de k , ambos decrescendo em valor absoluto na mesma proporção, apenas com o sinal trocado. Esse tipo de comportamento geralmente é originado por uma alta correlação positiva

GRÁFICO 4.6.3 - TRAÇO DA CRISTA

GRÁFICO 4.6.4 - RESÍDUO

entre elas. Como já foi mencionado, $r_{2,4}=0,9185$.

Entretanto eles se estabilizam após $k=0,1$, enquanto o resíduo aumenta rapidamente.

No caso do coeficiente da variável X_4 temos um caso em que há mudança de sinal no coeficiente. Ele é o maior coeficiente com sinal negativo, dirige-se rapidamente a zero e passa a ser positivo para k entre 0,3 e 0,4. Essa instabilidade vai sugerir, como veremos no próximo capítulo, que ele pode ser eliminado da equação, e que com a sua eliminação os coeficientes das demais variáveis melhorarão de comportamento.

Os coeficientes das variáveis X_1 e X_3 se mantêm relativamente estável enquanto que $B_5(k)$ é nulo para qualquer valor de k .

Segundo o critério do Traço, escolheríamos $k=0,08$, onde os coeficientes já estão razoavelmente estáveis, e a soma de quadrados residual não se elevou demais.

Com esse valor de k , as estimativas dos coeficientes serão:

TABELA 4.6.6

i	1	2	3	4	5
b_i	0,431	1,711	-0,419	-1,138	-0,085
$B_i(k)$	0,386	0,971	-0,320	-0,423	-0,049
$v.p.\%$	(10,44%)	(43,25%)	(23,63%)	(62,74%)	(42,35%)

Os números entre parêntesis são as variações percentuais de B_i com relação a b_i .

Em termos dessa variação temos que até $k=0,08$, enquanto que a partir daí o crescimento torna-se um pouco mais lento.

A escolha do valor de k nesse caso, recai então sobre ele, visto que, o critério de melhor desempenho apresenta um valor de k muito próximo de zero, ou seja, $k=0,002$. Neste ponto entretanto a variação ainda é muito grande para uma estimativa ser estável.

EXEMPLO 4.6.3

Este exemplo foi retirado do livro do Draper e Smith, capítulo 6, onde foi utilizado como exemplo para aplicação das diversas técnicas de seleção de variáveis. Os dados podem ser encontrados no Apêndice B do livro.

Nosso interesse nesse exemplo está no fato de que apresenta agora duas correlações altas ao invés de uma como no caso anterior, além de apresentar forte multicolinearidade como pode ser constatado.

A matriz de correlação é dada por:

$$\begin{pmatrix} 1,00000 & 0,22858 & -0,82413 & -0,24544 \\ 0,22858 & 1,00000 & -0,13924 & -0,97296 \\ -0,82413 & -0,13924 & 1,00000 & 0,02954 \\ -0,24544 & -0,97296 & 0,02954 & 1,00000 \end{pmatrix}$$

e os autovalores são:

(2,23571 1,57607 0,18661 0,00162)

apresentando um número de Condição de 37,1 que pode ser considerado alto.

Vamos mostrar o comportamento do Traço, e tentar achar a solução que apresenta melhor estabilidade, bem como aquelas que supostamente fazem com que $\tilde{B}(k)$ tenha EQM menor do que $\tilde{b}$.

O Traço da Crista é mostrado a seguir.

TABELA 4.6.7

k 20	$B1(k)$ 20	$B2(k)$ 20	$B3(k)$ 20	$B4(k)$ 20	SQR 20
0.0	0.6065	0.5278	0.0434	-0.1602	0.0176
0.01	0.5142	0.3166	-0.0549	-0.3815	0.0178
0.02	0.4974	0.3032	-0.0694	-0.3941	0.0180
0.03	0.4853	0.2996	-0.0788	-0.3963	0.0183
0.04	0.4751	0.2986	-0.0863	-0.3957	0.0186
0.05	0.4659	0.2984	-0.0927	-0.3941	0.0190
0.06	0.4576	0.2985	-0.0984	-0.3920	0.0195
0.07	0.4499	0.2989	-0.1034	-0.3897	0.0200
0.08	0.4428	0.2992	-0.1079	-0.3872	0.0205
0.09	0.4362	0.2995	-0.1120	-0.3848	0.0211
0.10	0.4299	0.2998	-0.1157	-0.3824	0.0217
0.20	0.3827	0.2991	-0.1391	-0.3601	0.0289
0.30	0.3510	0.2943	-0.1494	-0.3418	0.0378
0.40	0.3272	0.2876	-0.1537	-0.3260	0.0480
0.50	0.3082	0.2802	-0.1549	-0.3122	0.0591
0.60	0.2922	0.2726	-0.1543	-0.2998	0.0710
0.70	0.2784	0.2650	-0.1527	-0.2885	0.0834
0.80	0.2663	0.2575	-0.1505	-0.2782	0.0961

GRÁFICO 4.6.5 - TRAÇO DA CRISTA

GRÁFICO 4.6.6 - RESÍDUO

Alguns comentários a respeito dos gráficos que são bastante significativos:

- i) Observem o comportamento dos $B_z(k)$'s para k entre 0 e 0,01. O decréscimo em todos os coeficientes são notáveis, enquanto que o acréscimo na soma de quadrados residual é mínimo. A partir de $k > 0,01$ existe a estabilidade para todos os coeficientes. Por exemplo, um aumento de 1% na diagonal de $X'X$ acarreta uma variação de 66,7% em $B_2(k)$, enquanto que o acréscimo no resíduo é de 1,4%.
- ii) Para k entre 0 e 0,01 $B_3(k)$ muda de sinal e começa a crescer em valor absoluto.
- iii) Vejam a coincidência na variação dos coeficientes entre X_1 e X_3 e X_2 e X_4 . Esse tipo de comportamento, onde as variações são as mesmas, apenas deslocadas na escala, geralmente acontecem quando temos altas correlações negativas.

Com o exposto acima não fica difícil dizer que um bom valor para k deve estar entre 0,01 e 0,02.

Escolhendo $k=0,01$, temos boas razões para acreditar que $B(0,01)$ é bastante estável.

Os critérios 4.2, 4.3 e 4.4, apresentam os seguintes valores para k :

Critério 4.2 - $\hat{k} = 0,01$

Critério 4.3 - $\hat{k} = 0$, pois $Q < 0$

Critério 4.4 - $\hat{k} = 0,008$.

Neste caso, os k 's estão razoavelmente próximos um do outro, favorecendo então a escolha de $k=0,01$.

O estimador através do método das Cristas é dado na tabela abaixo e comparado com o de mínimos quadrados.

TABELA 4.6.8

i	1	2	3	4	SQR
b_i	0,6065	0,5279	0,0434	-0,1602	0,0176
$B_i(0,01)$	0,5143	0,3166	-0,0549	-0,3816	0,0179
V.P.%	15,20%	40,03%	226,50%	138,20%	1,14%

Embora neste caso a variação brutal seja no coeficiente pequeno B_3 , as variações também são grandes nos coeficientes significativos, o que significa que os mínimos quadrados realmente fornecem uma solução bastante instável.

C A P Í T U L O V

SELEÇÃO DE VARIÁVEIS

5.1 - INTRODUÇÃO

Uma aplicação bastante interessante desta técnica, através da utilização do Traço da Crista, é em Seleção de Variáveis.

Isto foi inicialmente sugerido por Hoerl e Kennard (1970b), que introduziram algumas regras para eliminar as variáveis que supostamente causam a instabilidade do sistema.

Porém a repercussão desse processo, parece não ter encontrado eco nos pesquisadores da área. Muito pouco foi escrito a respeito desde então. Acreditamos que esse pouco in

teresse seja devido às dificuldades de comparações entre os vários métodos.

No próximo ítem vamos enunciar os passos sugeridos por Hoerl e Kennard, bem como algumas modificações propostas para tentar aperfeiçoar este critério de seleção.

5.2 - SELEÇÃO DE VARIÁVEIS

Não faremos aqui nenhum comentário a respeito das diversas técnicas de selecionar variáveis, pois esse assunto já é bem conhecido, e um extenso tratado sobre o mesmo pode ser encontrado em Hocking (1976), Chatterjee e Price (1977) e Draper e Smith (1981).

Como já vimos, a técnica de Cristas pode detectar instabilidades nos sistemas através do Traço. Consequentemente, pode-se tentar utilizar este critério a fim de selecionar um subconjunto menor de variáveis que seja mais estável do que o conjunto completo.

Hoerl e Kennard (1970b), sugeriram os seguintes passos, utilizando o Traço da Crista, para eliminar variáveis:

- a) Elimine as variáveis que possuem coeficientes estáveis, porém com baixo poder de predição, isto é, aqueles coeficientes estáveis, porém pequenos. Isto pode ser feito facilmente devido ao fato das variáveis

veis estarem na forma padronizada e os coeficientes são portanto, diretamente comparáveis.

- b) Elimine as variáveis que apresentam coeficientes muito instáveis que dirigem-se rapidamente a zero.
- c) Elimine uma ou mais variáveis com coeficientes instáveis.

O subconjunto restante terá o comportamento de um sistema aproximadamente ortogonal.

A verificação dessa ortogonalidade é feita através do gráfico onde são colocados $\sum b_i^2(k)$ e $\sum \frac{b_i^2}{(1+k)^2}$ contra k . Se os gráficos estiverem quase coincidentes para-se o processo.

A equação final será obtida do modelo completo, utilizando o critério do Traço para achar k , e colocando os coeficientes das variáveis eliminadas iguais a zero.

Vamos no entanto, utilizar um critério um pouco diferente.

Sugerimos eliminar:

- As variáveis que não tem poder de predição, isto é, as que tem os coeficientes pequenos,
- Passo a passo, as variáveis cujos coeficientes são instáveis e dirigem-se a zero rapidamente, começando com a variável mais relacionada com as outras, através de alguma medida como por exemplo, o coeficiente

de correlação múltipla R_i^2 .

Agindo iterativamente dessa forma, após alguns passos teremos um comportamento semelhante ao de um sistema relativamente estável, e a equação final será estimada através do método dos mínimos quadrados, do subconjunto restante.

Diferimos portanto, do critério proposto por Hoerl e Kennard em dois pontos: eles sugerem a eliminação simultânea no ítem (b), enquanto sugerimos a eliminação passo a passo, pois embora o trabalho computacional seja maior, podemos observar melhor o comportamento de cada variável eliminada do modelo. O segundo ponto é que eles estimam a equação final através do método das Cristas do modelo completo, enquanto que estimamos a equação final através dos mínimos quadrados do subconjunto resultante. Baseamo-nos no fato de que se esse subconjunto resultante for aproximadamente ortogonal, o melhor estimador é sem dúvida, o de mínimos quadrados.

Vamos a seguir ver algumas aplicações.

5.3 - EXEMPLOS

EXEMPLO 5.3.1 - (Continuação do Exemplo 4.6.1)

Vamos aplicar a nossa técnica neste exemplo a título de curiosidade, pois evidentemente, este não é o exemplo para o qual esta técnica foi desenvolvida.

Segundo o critério, a primeira variável a ser eli

minada do modelo é a X_5 , que tem o coeficiente quase nulo em todos os k 's examinados.

A segunda variável eliminada é X_2 , pois ela dirige-se rapidamente a zero, e neste caso ela é única com esse comportamento.

Achando novamente o Traço sem essas duas variáveis, teremos os Gráficos 5.3.1 e 5.3.2 a seguir.

Observem que a eliminação das duas variáveis em na da mudou o comportamento das demais. Todas mantiveram o mesmo comportamento, como já era esperado em virtude da quase ortogonalidade. Se desejarmos ficar com apenas três variáveis poderemos eliminar tranquilamente a variável X_4 , pois seu coeficiente é o mais baixo e não mudou com a eliminação das outras variáveis.

A solução final em termos de coeficientes estandarizados é:

$$b_1 = 0,6356, \quad b_3 = 0,2891, \quad b_4 = 0,0907, \quad b_6 = -0,1935$$

com uma soma de quadrados residual de 0,2715.

EXEMPLO 5.3.2 - (Continuação do Exemplo 4.6.2)

Neste problema já existe um pouco mais de instabilidade, sendo portanto um pouco mais apropriado à aplicação de nosso critério.

Evidentemente a primeira variável a ser eliminada

GRÁFICO 5.3.1 - TRAÇO DA CRISTA

GRÁFICO 5.3.2 - RESÍDUO

é a X_5 , cujo coeficiente é quase nulo em quase toda a extensão examinada. Esse comportamento era esperado, devido a natureza da variável em questão. Um critério que não a eliminasse, evidentemente não seria um bom critério.

Existe apenas uma variável instável que dirige-se a zero, que é X_4 . Nesse caso não há também outra alternativa.

Como já dissemos também essa variável é altamente correlacionada com a variável X_2 e sua presença deve estar influenciando fortemente a outra.

Os Gráficos 5.3.3 e 5.3.4 com as variáveis eliminadas são dados a seguir.

Vejam o comportamento do Traço sem as duas variáveis. Toda a instabilidade da variável X_2 desapareceu e seu coeficiente agora é bastante estável, enquanto que o comportamento das demais se manteve constante.

A soma de quadrados residual teve uma elevação significativa, o que também era esperado devido ao comportamento do gráfico dos resíduos em (4.6.2).

Podemos agora utilizar o critério dos Mínimos Quadrados para encontrar a solução do problema.

A equação final, é então estimada pelos coeficientes na sua forma padronizada:

$$b_1 = 0,4607, \quad b_2 = 0,6802, \quad b_3 = -0,3624$$

e a soma de quadrados residual é 0,4071.

GRÁFICO 5.3.3 - TRAÇO DA CRISTA

GRÁFICO 5.3.4 - RESÍDUO

EXEMPLO 5.3.3 - (Continuação do Exemplo 4.6.3)

Draper e Smith utilizaram este exemplo para aplicar as técnicas de seleção de variáveis. Vamos aplicar também o nosso critério e ver a equação resultante.

Segundo o critério, a variável eliminada inicialmente é a X_3 pois ela é a única que dirige-se a zero.

Assim, sem essa variável, obtemos os Gráficos 5.3.5 e 5.3.6.

Observem que com a eliminação de X_3 houve uma substancial redução na variabilidade dos coeficientes, embora ainda tenha sobrado um pouco de variabilidade em X_2 e X_4 , pela correlação alta existente entre elas, não deve ser mais problemas estimá-las.

Os coeficientes estimados serão:

$$b_1 = 0,5677, \quad b_2 = 0,4304, \quad b_4 = -0,2632$$

e a soma de quadrados dos resíduos é: $SQR = 0,0176$.

5.4 - COMENTÁRIOS

Mostramos nos itens anteriores, uma forma de selecionar variáveis através da Técnica de Cristas.

Evidentemente não podemos concluir que o subconjunto resultante representa o "melhor" subconjunto das variáveis inicialmente no modelo, mas podemos apenas afirmar que trata-se de um subconjunto mais estável, segundo o critério de esta

GRÁFICO 5.3.5 - TRAÇO DA CRISTA

GRÁFICO 5.3.6 - RESÍDUO

bilidade do Traço.

Nos casos onde o objetivo é a estabilidade, acreditamos que isto pode proporcionar bons resultados embora a multicolinearidade seja alta.

Infelizmente, como já dissemos anteriormente, este enfoque da técnica não recebeu a devida atenção dos pesquisadores, e conseqüentemente os resultados não estão ainda bem desenvolvidos. É necessário portanto, mais trabalho nesse sentido a fim de verificar o potencial real desta técnica nessa área. Neste sentido, a nossa sugestão representa apenas os passos iniciais de um trabalho que estamos desenvolvendo e que posteriormente pretendemos divulgar.

CAPÍTULO VI

CONCLUSÃO

Mostramos portanto a técnica de Regressão sobre Cristas, e sua aplicação em alguns problemas, através dos critérios enunciados para a determinação de k .

Esses critérios são apenas alguns dos citados na literatura. O nosso interesse por eles, é que além de serem intuitivos, fáceis de aplicar, eles foram comparados em alguns trabalhos realizados.

Devido a falta de uma teoria mais bem elaborada, esses critérios foram comparados através de experimentos simu

lados. Vários trabalhos de simulação foram realizados, como por exemplo, Dempster e outros (1977), McDonald e Galarneau (1975), Hoerl, Kennard e Baldwin (1975), Wichern e Churchill (1978), além de muitos outros, cuja relação pode ser encontrada na bibliografia.

Destes trabalhos, algumas conclusões podem ser tiradas:

- Não existe um critério que seja uniformemente melhor. Dependendo do grau de multicolinearidade, direção do vetor de parâmetros $\underline{\beta}$, e da razão sinal-ruído $\underline{\beta}'\underline{\beta}/\sigma^2$, um critério pode ser melhor do que o outro e vice-versa.
- Nos casos de alta multicolinearidade $\underline{B}(k)$ é melhor do que $\underline{b}$ em termos de EQM, para quase todos os critérios.
- Nos casos de baixa multicolinearidade, os mínimos quadrados são melhores do que $\underline{B}(k)$, em todas as ocasiões.
- Dentre os Critérios 4.2., 4.3 e 4.4, o melhor deles é sem dúvida o Critério 4.3, embora com algumas restrições. O número de vezes que coincide com o de mínimos quadrados, pode ser muito grande.
- O critério ótimo para determinar k ainda não foi encontrado. Mais trabalhos serão necessários, pois o

potencial desta técnica ainda não foi plenamente alcançado.

Entretanto, apesar das restrições, podemos obter bons resultados, em problemas com alta multicolinearidade, utilizando os recursos dessa técnica, através do Traço da Crista, aliado ao emprego dos critérios enunciados.

Através do Traço podemos visualizar a instabilidade produzida pela multicolinearidade na maioria dos problemas. Um gráfico altamente variável, é sinônimo de instabilidade na solução dos mínimos quadrados.

6.1 - SUGESTÕES PARA A OBTENÇÃO DE k

Após estes fatos, podemos fazer algumas sugestões para a obtenção de $B(k)$.

Sugerimos a utilização dos mínimos quadrados nos casos, onde embora haja alguns indícios de moderada multicolinearidade, o Traço se mostra relativamente estável.

Nos casos onde o Traço se mostrar instável, sugerimos a aplicação da Técnica de Cristas, mesmo onde aparentemente a multicolinearidade não é muito alta. Os limites para essa instabilidade são obtidos apenas empiricamente, digamos que se uma perturbação de 1% na diagonal de $X'X$ acarretar numa diferença superior a 10% nos coeficientes, com um acréscimo não superior a 5% nos resíduos, teremos problemas de instabilidade.

Reafirmamos que estes limites são apenas algumas sugestões iniciais e que mais trabalhos estão sendo feitos no sentido de aprimorá-los.

Para a utilização dos critérios, nos casos necessários, sugerimos a aplicação de todos eles para determinar k .

Se eles fornecerem k 's próximos um do outro, é um sinal de que o sistema não é tão mal comportado, e poderemos tomar qualquer um para nosso estimador, pois evidentemente todos serão quase iguais.

O caso mais provável de acontecer no entanto, é haver bastante diferença entre os k 's. Neste caso, se eles estiverem na região estável do Traço tome o menor deles, assim estaremos aliando a estabilidade fornecida pelo Traço com uma provável diminuição do EQM fornecido pelos critérios.

Se nenhum deles estiver na região de estabilidade do Traço determinada pelos limites acima, devemos nos orientar pelo objetivo do trabalho, por exemplo.

Se queremos previsão, ou seja, estabilidade devemos escolher k através do Traço; se estimação precisa de β , utilizaremos o Critério 4.3, embora o Traço não deva nunca ser ignorado.

Observamos entretanto, que as sugestões acima, não são de forma alguma infalíveis, representam apenas mais uma tentativa empírica de obtenção de um bom estimador para o parâmetro, baseado em algumas experiências realizadas.

6.2 - SUGESTÕES PARA A SELEÇÃO DE VARIÁVEIS

Quanto a seleção de variáveis, esta técnica pode ser de utilidade na determinação de um subconjunto de variáveis que seja mais estável do que o conjunto todo. Nestes problemas devemos utilizar apenas o Traço da Crista como indicador de estabilidade, não levando em conta os limites acima, mas o número desejado de variáveis que deve permanecer no modelo. Se no critério de seleção empregado, o subconjunto resultante não for o desejado, ou se ainda persistir alguma instabilidade, elimina-se uma ou mais variáveis com os coeficientes mais instáveis.

Finalmente, recomendamos que o Traço da Crista seja obtido mesmo nos problemas mais simples, em qualquer ocasião, pois o trabalho computacional é bastante rápido e fácil com a ajuda dos "pacotes" computacionais, tais como o MINITAB, BMDP e SAS, que já trazem subprogramas apropriados para a sua obtenção.

Em nossos estudos utilizamos os pacotes Statistical Analysis System (SAS), implantado no Instituto de Pesquisas Energéticas (IPEN) em São Paulo e o MINITAB da Universidade Federal de São Carlos.

6.3 - PONTOS QUE NECESSITAM DE MAIS PESQUISAS

Como pontos que necessitam de mais trabalhos, existem vários que poderemos citar:

- A procura de um critério ótimo para a escolha de k ,
- O desenvolvimento da teoria que permita fazer inferências mais apropriadas.
- O relacionamento com outras técnicas como por exemplo, diagnósticos. Nesse caso, podemos dizer que em algumas aplicações notamos que a medida que k aumenta, diminui muito o número de pontos influentes, detectados pelos critérios diagnósticos, tais como, diagonal da matriz de projeção H (matriz chapéu) DFFITS, etc. Um exemplo disto, pode ser encontrado em Besley e outros (1980), pág. 219-229. Talvez um estudo mais detalhado neste sentido possa trazer mais alguns resultados entre pontos influentes e multicolinearidade, e talvez quem sabe, um outro critério para a escolha de k , como por exemplo, use o k que elimine os pontos influentes.

6.4 - DERIVAÇÕES E TRABALHOS RELACIONADOS

Deixamos de abordar de forma proposital, algumas derivações relacionadas com esta técnica, que agora merecem alguns comentários. Como veremos, $B(k)$ pode ser considerado como caso particular de um número muito grande de técnicas que visam principalmente a multicolinearidade.

Não levamos em consideração essas técnicas, pois

o nosso objetivo era o de tentar inicialmente apresentar $\underline{B}(k)$ e mostrar o seu comportamento frente aos problemas com alta multicolinearidade. Entretanto, não resta dúvida de que são algumas alternativas atraentes para se continuar pesquisando.

Uma forma que deixamos sem a devida abordagem é a forma generalizada de $\underline{B}(k)$.

Hoerl e Kennard (1970a), introduziram também esta forma de estimar $\underline{\beta}$, que é dada por:

$$\hat{\underline{\alpha}}_* = (\Lambda + K)^{-1} \underline{X}' \underline{Y}$$

onde:

Λ : Matriz diagonal formada pelos autovalores de $\underline{X}'\underline{X}$,

$$\underline{X} = \underline{X}P',$$

$$K = \text{diag}(k_i), \quad \text{e}$$

$\hat{\underline{\alpha}}_*$ Estimador generalizado de $\underline{\alpha} = P\underline{\beta}$.

Apesar da afirmação dos autores de que este estimador apresenta as mesmas propriedades que $\underline{B}(k)$, ele não possui algumas características importantes como a facilidade de obtenção e nem o equivalente ao Traço da Crista. Um critério para a escolha dos k_i 's é sugerido pelos autores, e uma propriedade de convergência desse critério é mostrada em Hemmerle (1975).

Uma forma da qual $\underline{B}(k)$ é caso particular é a dos chamados estimadores "Shrunken", isto é, encolhidos, que apre

sentam a tendência de diminuir em direção a zero. Estes estimadores, foram estudados por Mayer e Willke (1973). Neste caso, $\underline{B}(k)$ pode ser escrito como $\underline{B}(k) = A\underline{b}$, onde $A = (I + k(X'X)^{-1})^{-1}$.

Se $A = \lambda I$, teremos, por exemplo, um estimador do tipo Stein.

Porém, a forma mais bem aceita de $\underline{B}(k)$, é através da teoria Bayesiana, quando pode ser mostrado que $\underline{B}(k)$ é a média da distribuição a posteriori, sob a hipótese a priori de que: $\underline{\beta} \sim N(\underline{0}, \sigma_*^2 I)$. Entretanto, como foi mostrado em alguns trabalhos de simulação, este critério nem sempre apresenta bons resultados. Porém aqui, o parâmetro k tem uma interpretação mais objetiva, pois $k = \sigma^2 / \sigma_*^2$. Se o conhecimento a priori for bem preciso, poderemos apresentar $\underline{B}(k)$ também preciso.

Um outro caso no qual $\underline{B}(k)$ é caso particular, foi abordado por Theil e Goldeberger (1961) e Theil (1971), ver Besley e outros (1980), no que chamaram de Estimação Mista (Mixed), que é dada por:

$$\underline{b}_{EM} = (X'X/\sigma^2 + R'V^{-1}R)^{-1}(X'\underline{y}/\sigma^2 + R'V^{-1}\underline{c})$$

onde:

$$\underline{c} = R\underline{\beta} + \underline{\xi}, \quad \text{com } E(\underline{\xi}) = \underline{0} \quad \text{e} \quad \text{Var}(\underline{\xi}) = V$$

com

R : Matriz de constantes conhecidas.

$$\text{Se } R'R = I, \quad V = \lambda^2 I \quad \text{e} \quad \underline{c} = \underline{0}, \quad \underline{b}_{EM} = \underline{B}(k), \quad \text{com } k = \sigma^2 / \lambda^2.$$

Temos portanto, que $B(k)$ é caso particular de um número grande de técnicas que visam particularmente o trato da multicolinearidade. Não é preciso dizer que só estes requerimentos já são motivos suficientes para apresentá-lo como um forte candidato, na busca de soluções alternativas para o problema da multicolinearidade.

APÊNDICE

Vamos mostrar aqui, dois trabalhos de simulação, comparando os critérios enunciados no Capítulo IV, ou seja, os critérios:

$$4.2 - \hat{k} = \frac{p\hat{\sigma}^2}{\underline{\underline{b}}'\underline{\underline{b}}}$$

4.3 - Escolha k tal que $\underline{\underline{B}}'(k)\underline{\underline{B}}(k)=Q$, se $Q>0$, ou faça $\underline{\underline{B}}(k)=\underline{\underline{b}}$, se $Q\leq 0$.

$$4.4 - k = \frac{p\hat{\sigma}^2}{\sum \lambda_i \hat{\alpha}_i^2}, \text{ onde } \hat{\underline{\underline{\alpha}}} = P'\underline{\underline{b}}$$

Embora os trabalhos sejam mais completos, mostraremos apenas as comparações relacionadas com os critérios acima.

O primeiro trabalho é o de McDonald e Galarneau (1975), que faz uma comparação de duas versões do Critério 4.3. O nosso interesse por esse trabalho é que além de comparar um dos critérios propostos aqui, ele apresenta uma regra ótima que permite nesse caso, em que o vetor de parâmetros é conhecido determinar o valor de k que apresenta o menor EQM .

O segundo trabalho é o desenvolvido por Wichern e Churchill (1978), que faz uma comparação entre os três critérios acima, e também com o de Mínimos Quadrados.

Faremos apenas um breve resumo, e apresentaremos as tabelas construídas a partir dos resultados simulados. Os interessados em maiores detalhes, deverão consultar os trabalhos originais.

Apesar dos trabalhos diferirem entre si no número de variáveis envolvidas, número de critérios e objetivos, as simulações apresentam as mesmas características básicas, que são as seguintes:

São supostos apenas dois valores para o parâmetro β , em ambos os trabalhos:

$$\beta = \beta_M \quad \beta = \beta_m, \quad \text{onde}$$

β_M e β_m são respectivamente os autovetores normalizados correspondentes ao maior autovalor de $X'X$.

Essas direções foram as escolhidas, porque representam as que oferecem a vantagem máxima e mínima, respectivamente

mente, de $\underline{B}(k)$ sobre $\underline{b}$, como foi demonstrado em Nehouse e Oman (1971).

Outros trabalhos utilizando outras direções além dessas foram realizados, ver por exemplo, Gunst e Mason (1977).

Foram supostos também alguns valores para a razão sinal-ruído $\beta'\beta/\sigma^2$, e finalmente algumas suposições sobre as correlações entre as variáveis independentes, foram feitas.

A comparação foi feita através da razão entre o EQM de $\underline{B}(k)$ com o EQM de $\underline{b}$, que vamos denotar por R_t , onde t é o tipo de critério utilizado:

$$R_t = \frac{EQM(\underline{B}(k))}{EQM(\underline{b})}(t).$$

No trabalho de McDonald e Galarneau, temos apenas $t=1, 2$, onde $t=1$, representa o Critério 4.3, enquanto que $t=2$, é o critério otimal.

Em Wichern e Churchill, os Critérios 4.2, 4.3 e 4.4 são respectivamente, $t=1, 2$ e 3 .

Mostraremos na Tabela A1, as variações de R_t em função de σ , enquanto que na Tabela A2, temos as variações de R_t em função das correlações, referentes ao trabalho de McDonald e Galarneau.

Observando as tabelas, podemos concluir que nos casos em que $\underline{\beta}=\underline{\beta}_M$, o estimador obtido através de 4.3, tem R_t sempre menor do que 1, enquanto que em $\underline{\beta}=\underline{\beta}_m$, ele é menor do que 1, somente nos casos de σ e α maiores.

O fato mais importante, entretanto está, no critério ótimo R_2 . Observem que em todos os casos o ganho foi bastante acentuado com relação aos mínimos quadrados. Isto é forte motivação para que se continue a busca de critérios para a escolha de k ótimo.

Quanto ao Critério 4.3, como em aplicações evidentemente não sabemos a direção de β , é mais importante, ver o seu desempenho em função de α e σ . Neste aspecto, podemos notar que quando α é 0,99, temos um R_1 menor que 1 para quaisquer combinações dos outros parâmetros, o que significa que nestes casos ele é um bom critério.

No trabalho de Wichern e Churchill, cujos resultados são mostrados na Tabela A3, o mal condicionamento é medido através da razão:

$$\lambda = \frac{\lambda_1}{\lambda_5}$$

Olhando a tabela, algumas conclusões imediatas podem ser obtidas:

No caso de $\beta = \beta_M$ todos os critérios tem desempenho melhor do que o método de mínimos quadrados, inclusive quando λ é pequeno, só empatando no caso de $\rho = \beta' \beta / \sigma^2 = 10.000$, e $\lambda = 9$.

Se $\beta = \beta_m$, existe alguma alternância, como por exemplo, o Critério 4.4 é bastante ruim nos casos de λ alto e $\rho = 10.000$, melhorando um pouco quando ρ diminui, mas mesmo assim, é o pior de todos os três, enquanto que o Critério 4.3,

tem o melhor desempenho de todos os três, só perdendo para os mínimos quadrados nos casos de baixa multicolinearidade e $\beta = \beta_m$.

O Critério 4.2 vem em segundo entre os critérios e na maioria dos casos tem bom desempenho.

Em resumo, desses dois trabalhos a conclusão final é a de que:

- Dentre os três critérios, o que apresentou melhor resultado é o 4.3.
- Em segundo lugar vem o Critério 4.2, enquanto que o 4.4 só deve ser aplicado nos casos em que temos alta multicolinearidade e alguma informação a priori que a direção de β , é próxima de β_M , ou que $\rho = \beta' \beta / \sigma^2$ é pequeno.
- O critério ótimo ainda não foi encontrado, mais trabalhos são necessários nesse sentido.

TABELA A.1

TABELA A.2

TABELA A.3

$\lambda = 378$						$\lambda = 580$					
$\beta = \beta_M$						$\beta = \beta_m$					
$R_t \backslash \rho$	10.000	100	4	1	.04	$R_t \backslash \rho$	10.000	100	4	1	.04
1	.9976	.8163	.3848	.3457	.2754	1	1.0014	1.0544	.5058	.3465	.2792
2	.8435	.3960	.4275	.5674	.6205	2	1.0000	1.0256	.7013	.5801	.6259
3	.9995	.9519	.4211	.1316	.0071	3	14.	9.3830	.4260	.1225	.0091

$\lambda = 41$						$\lambda = 21$					
$\beta = \beta_M$						$\beta = \beta_m$					
$R_t \backslash \rho$	10.000	100	4	1	.04	$R_t \backslash \rho$	10.000	100	4	1	.04
1	.9972	.7920	.2730	.2237	.2685	1	1.0085	1.4410	.6268	.3153	.2129
2	.8510	.4361	.4538	.5745	.7312	2	1.0000	1.0000	.7882	.6646	.6375
3	.9993	.9243	.2896	.0822	.0202	3	121.	.17.	.7113	.1733	.0204

$\lambda = 9$						$\lambda = 21$					
$\beta = \beta_M$						$\beta = \beta_m$					
$R_t \backslash \rho$	10.000	100	4	1	.04	$R_t \backslash \rho$	10.000	100	4	1	.04
1	.9997	.9731	.6722	.4399	.3032	1	1.0004	1.0299	.9973	.6939	.3652
2	1.0000	.8407	.4833	.3627	.5220	2	1.0000	1.0089	.9460	.8779	.6490
3	.9999	.9935	.8677	.6084	.1102	3	1.0280	3.0404	2.0350	.7109	.1156

$\lambda = 9$						$\lambda = 21$					
$\beta = \beta_M$						$\beta = \beta_m$					
$R_t \backslash \rho$	10.000	100	4	1	.04	$R_t \backslash \rho$	10.000	100	4	1	.04
1	.9998	.9839	.7346	.5148	.3430	1	.9995	1.0312	1.3608	.8910	.3458
2	1.0000	.9499	.6096	.4916	.6672	2	1.0000	1.0078	1.0809	.8972	.6630
3	.9998	.9937	.8608	.6245	.1926	3	1.0092	2.4401	3.1704	1.1881	.1739

$\lambda = 9$						$\lambda = 21$					
$\beta = \beta_M$						$\beta = \beta_m$					
$R_t \backslash \rho$	10.000	100	4	1	.04	$R_t \backslash \rho$	10.000	100	4	1	.04
1	1.0000	.9936	.8814	.6988	.4348	1	1.0000	1.0095	1.1807	1.1753	.4559
2	1.0000	.9870	.8336	.6665	.6578	2	1.0000	1.0027	1.0214	1.0838	.7161
3	1.0000	.9974	.9421	.8081	.3777	3	1.0011	1.1160	2.1370	1.5628	.3636

REFERÊNCIAS

- ALLEN, D.M. (1974) - "The Relationship Between Variable Selection and Data Augmentation and a Method for Prediction", *Technometrics*, 16, 125-127.
- ASKIN, R.G., and MONTGOMERY, D.C. (1980) - Augmented Robust Estimators, *Technometrics*, 22, 333-341.
- BEATON, A.E., RUBIN, D.B., and BARONE, J.L. (1976) - "The Acceptability of Regression Solutions: Another Look Computational Accuracy", *J. Am. Statist. Assoc.*, 71, 158-168.
- BELINFANTE, A., and COXE, K. (1982) - "Multicollinearity and Ridge Regression", Paper presented at the annual meeting of the American Statistical Association, 17 de agosto de 1982.
- BELLMAN, R. (1979) - *Introduction to Matrix Analysis*, TATA McGraw-Hill Publishing Company.
- BESLEY, D.A., KUH, E., and WELCH, R.E. (1980) - *Regression Diagnostics: Identifying Influential Data and Sources of Collinearity*, John Wiley & Sons.
- BROWN, P.J. (1977) - "Centering and Scaling in Ridge Regression", *Technometrics*, 19, 35-36.
- BROWN, P.J., and ZIDEK, J.V. (1980) - "Adaptive Multivariate

Ridge Regression", *The Annals of Statistics*, 8, 64-74.

CAUSEY, B.D. (1980) - "A Frequentist Analysis of a Class of Ridge Regression Estimators", *J. Am. Statist. Assoc.*, 75, 736-738.

CHATTERJEE, S., and PRICE, B. (1977) - *Regression Analysis by Example*, John Wiley & Sons.

DEMPSTER, A.P., SCHATZOFF, M., and WERMUTH, N. (1977) - "A Simulation Study of Alternatives to Ordinary Least Squares", *J. Am. Statist. Assoc.*, 72, 77-106.

DRAPER, N.R., and VAN NOSTRAND, R.C. (1978) - "Ridge Regression - is it Worthwhile?", University of Wisconsin Statistics Department Technical Report n° 501.

DRAPER, N.R., and VAN NOSTRAND, R.C. (1979) - "Ridge Regression and James-Stein Estimation: Review and Comments", *Technometrics*, 21, 451-466.

DRAPER, N.R., and SMITH, H. (1981) - *Applied Regression Analysis*, John Wiley & Sons, Second Edition.

ERICKSON, G.M. (1981) - "Using Ridge Regression to Estimate Directly Lagged Effects in Marketing", *J. Am. Statist. Assoc.* 76, 766-773.

FAREBROTHER, R.W. (1975) - "The Minimum Mean Square Error Linear Estimator and Ridge Regression", *Technometrics*, 17,

127-128.

GOLUB, G.H., HEATH, M., and WAHBA, G. (1979) - "Generalized Cross-Validation as a Method for Choosing a Good Ridge Parameter", *Technometrics*, 21, 215-223.

GUILKEY, D.K., and MURPHY, J.L. (1975) - "Directed Ridge Regression, Techniques in Cases of Multicollinearity", *J. Am. Statist. Assoc.*, 70, 769-775.

GUNST, R.F., and MASON, R.L. (1977) - "Biased Estimation in Regression: An Evaluation using Mean Square Error", *J. Amer. Statist. Assoc.*, 72, 616-628.

GUNST, R.F., and MASON, R.L. (1980) - *Regression Analysis and Its Applications: A Data Oriented Approach*, Marcel-Dekker.

HAWKINS, D.M. (1975) - "Relations Between Ridge Regression and Eigenanalysis of the Augmented Correlations Matrix", *Technometrics*, 17, 477-480.

HEMMERLE, W.J. (1975) - "An Explicit Solution for Generalized Ridge Regression", *Technometrics*, 17, 309-314.

HOCKING, R.R. (1976) - "The Analysis and Selection of Variables in Linear Regression", *Biometrics*, 32, 1-49.

HOCKING, R.R., SPEED, F.M., and LYNN, M.J. (1976) - "A Class of Biased Estimators in Linear Regression", *Technometrics*, 18, 425-437.

HOERL, A.E. (1959) - "Optimum Solution of Many Variables Equations", *Chem. Eng. Prog.*

HOERL, A.E. (1962) - "Application of Ridge Analysis to Regression Problems", *Chem. Eng. Prog.*, 60, 54-59.

HOERL, A.E. and KENNARD, R.W. (1970a) - "Ridge Regression: Biased Estimation for Nonorthogonal Problems", *Technometrics*, 12, 55-68.

HOERL, A.E., and KENNARD, R.W. (1970b) - "Ridge Regression: Applications to Nonorthogonal Problems", *Technometrics*, 12, 69-82.

HOERL, A.E., KENNARD, R.W., and BALDWIN, K.F. (1975) - "Ridge Regression: Some Simulations", *Comm. in Statist.*, 4, 105123.

HOERL, A.E., and KENNARD, R.W. (1976) - "Ridge Regression: Iterative Estimation of the Biasing Parameters", *Comm. in Statist.*, A5, 77-88.

JAMES, W., and STEIN, C. (1961) - "Estimation with Quadratic Loss", *Proceedings of the Fourth Berkeley Symposium, Vol. I*, 361-379.

KHURI, A.I., and MYERS, R.H. (1979) - "Modified Ridge Analysis", *Technometrics*, 21, 467-473.

- LAVRENTIEV, M.M. (1967) - *Some Improperly Posed Problems of Mathematical Physics*, Springer-Verlag, Berlin.
- LAWLESS, J.F., and WANG, P. (1976) - "A Simulation Study of Ridge and Others Regression Estimators", *Comm.in Statist.*, A5, 307-323.
- LEAMER, E.E. (1981) - "Coordinate-Free Ridge Regression Bounds", *J. Am. Statist. Assoc.*, 76, 842-849.
- LINDLEY, D.V., and SMITH, A.F.M. (1972) - "Bayes Estimates for the Linear Models", *J. Roy. Statist. Soc. B*, 34, 1-41.
- MARQUARDT, D.W. (1970) - "Generalized Inverses, Ridge Regression, Biased Linear Estimation, and Nonlinear Estimation", *Technometrics*, 12, 591-612.
- MARQUARDT, D.W., and SNEE, R.D. (1975) - "Ridge Regression in Prattice", *The American Statistician*, 29, 3-20.
- MAYER, L.S., and WILLKE, T.A. (1973) - "On Biased Estimation in Linear Models", *Technometrics*, 15, 497-508.
- McDONALD, G.C., and GALARNEAU, D.I. (1975) - "A Monte Carlo Evaluation of Some Ridge Type Estimators", *J. Am. Statist. Assoc.*, 70, 407-416.
- McDONALD, G.C. (1980) - "Some Algebraic Properties of Ridge Coefficients", *J. Roy. Statist. Soc.*, B, 42, 31-34.

- MULLET, G.M. (1976) - "Why Regression Coefficients Have the Wrong Sign", *J. of Quality Technology*, 8, 121-126.
- NEWHOUSE, J.P., and OMAN, S.D. (1971) - "An Evaluation of Ridge Estimators", *Rand Corporation*, R-716-PR.
- OBENCHAIN, R.L. (1975) - "Ridge Analysis Following a Preliminary Tests of the Shrunk Hypothesis", *Technometrics*, 17, 431-445.
- OBENCHAIN, R.L. (1977) - "Classical F-Tests and Confidence Regions for Ridge Regression", *Technometrics*, 19, 429-439.
- OBENCHAIN, R.L. (1978) - "Good and Optimal Ridge Estimators", *The Annals of Statistics*, 6, 1111-1121.
- OMAN, S.D. (1981) - "A Confidence Bound Approach to Choosing the Biasing Parameter in Ridge Regression", *J. Am. Statist. Assoc.*, 76, 452-461.
- PRICE, B. (1977) - "Ridge Regression: Application to Nonexperimental Data", *Psychological Bulletin*, 84, 759-766.
- SMITH, G., and CAMPBELL, F. (1980) - "A Critique of Some Ridge Regression Methods", *J. Am. Statist. Assoc.*, 75, 74-103.
- SWINDEL, B.F. (1981) - "Geometry of Ridge Regression Illustrated", *The American Statistician*, 35, 12-15.
- VINOD, H.D. (1976) - "Application of New Ridge Regression Me

thods to a Study of Bell System Scale Economies", *J. Am. Statist. Assoc.*, 71, 835-841.

WICHERN, D.W., and CHURCHILL, G.A. (1978) - "A Comparison of Ridge Estimators", *Technometrics*, 20, 301-311.