



**LUZIA APARECIDA DA COSTA**

**NOVO ESTIMADOR DE CUMEEIRA DE  
RAO COM APLICAÇÃO EM SELEÇÃO  
GENÔMICA**

**LAVRAS-MG**

**2015**

**LUZIA APARECIDA DA COSTA**

**NOVO ESTIMADOR DE CUMEEIRA DE RAO COM APLICAÇÃO EM  
SELEÇÃO GENÔMICA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutora.

Orientador  
Dr. Lucas Monteiro Chaves

**LAVRAS-MG  
2015**

**Ficha Catalográfica Elaborada pelo Sistema de Geração de Ficha  
Catalográfica da Biblioteca Universitária da UFLA, com dados  
informados pelo(a) próprio(a) autor(a).**

Costa, Luzia Aparecida da.

Novo estimador de cumeieira de Rao com aplicação em seleção  
genômica /

Luzia Aparecida da Costa. – Lavras : UFLA, 2014.

126 p. : il.

Tese (Doutorado) - Universidade Federal de Lavras, 2015.

Orientador: Lucas Monteiro Chaves.

Bibliografia.

1. Análise Multivariada. 2. Estimadores Viesados. 3.  
Multicolinearidade. 4. Ridge Regression. I. Universidade Federal  
de Lavras. II. Título.

**LUZIA APARECIDA DA COSTA**

**NOVO ESTIMADOR DE CUMEEIRA DE RAO COM APLICAÇÃO EM  
SELEÇÃO GENÔMICA**

Tese apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-Graduação em Estatística e Experimentação Agropecuária, área de concentração em Estatística e Experimentação Agropecuária, para a obtenção do título de Doutora.

APROVADA em 03 de Junho de 2015.

Dr. Daniel Furtado Ferreira	UFLA
Dr. Devanil Jaques	UFLA
Dr. Fabyano Fonseca e Silva	UFV
Dr. Marcos Santos de Oliveira	UFSJ

Dr. Lucas Monteiro Chaves  
Orientador

**LAVRAS-MG  
2015**

A Deus que me abriu os caminhos e  
aos meus pais, Ângelo (*in memoriam*) e Aparecida,  
pelo grande amor dispensado.

**DEDICO**

## AGRADECIMENTOS

A Deus, por ter trilhado comigo todo o tempo, me amparando, me ajudando e me fortificando nos momentos mais difíceis.

Aos meus pais, Ângelo (*in memoriam*) e Aparecida pelo incentivo aos estudos desde cedo.

Aos meus tios Jurandir e Fátima que são como pais para mim.

A toda minha família pelos incentivos: avó, tios e tias, primos e primas.

À Universidade Federal de Lavras, pela oportunidade de aprimoramento acadêmico e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

Ao professor Lucas Monteiro Chaves pela orientação, apoio e dedicação nestes quatro anos. E mais que isso, pela amizade.

Aos membros da banca examinadora, pelas críticas e sugestões a este trabalho.

Aos professores do Departamento de Ciências Exatas, pelos conhecimentos transmitidos.

Aos funcionários do Departamento de Ciências Exatas, pela prontidão em ajudar.

Aos professores Paulo Sávio Lopes e Simone Eliza Facioni Guimaraes, responsáveis pela granja de melhoramento de suínos do DZO-UFV e Laboratório de Biotecnologia Animal da UFV, respectivamente, pela concessão dos dados utilizados nas aplicações deste trabalho.

Aos colegas do curso de Pós-graduação em Estatística e Experimentação Agropecuária, pelo convívio.

Aos amigos e amigas que tornaram essa jornada mais leve, dentre eles: Fernanda, Renata, Richeli, Márcia, Danielle, Mariana, Big, "Intoju", Rafaeli, Ivone

Lima, Ivone Martins, Andressa, Thauana, Serjão e Celina.

Ao grande amor da minha vida.

A todos que, direta ou indiretamente, contribuíram para a realização deste trabalho, muito obrigada!

**“O sucesso é a habilidade de se ir de fracasso em fracasso sem perder o entusiasmo.”**

Desconhecido



## RESUMO

Um novo estimador de cumeieira é obtido a partir de uma abordagem geométrica da teoria geral dos estimadores de cumeieira. Obtido a partir da intersecção de dois elipsoides ortogonais e possuindo uma expressão algébrica simples este estimador é bastante natural. Exemplos são utilizados que sugerem que as curvas de cumeieira deste estimador se estabilizam mais rapidamente que a do estimador de cumeieira usual. Este fato pode facilitar a estimação do valor ótimo para o parâmetro, uma área de pesquisa central na teoria. Foram obtidas expressões para a variância total e o erro quadrático médio. O desempenho do estimador proposto foi estudado pela aplicação a um conjunto de dados já analisados na literatura pela utilização do estimador de cumeieira usual. Uma análise dos dados de seleção genômica relativos à resposta pH da carne suína aos 45 minutos e 24 horas após o abate. Em todas as aplicações os resultados obtidos pelo estimador proposto são comparados com os resultados obtidos pelo estimador de cumeieira usual e pelo estimador de Mayer e Wilke (1973).

Palavras-chave: Análise Multivariada. Estimadores Viesados. Multicolinearidade. Ridge Regression.

## ABSTRACT

A new ridge estimator is obtained from a geometric approach of the general theory of the ridge estimators. Obtained from the intersection of two orthogonal ellipsoids and having a simple algebraic expression this estimator is quite natural. Examples are used to suggest that curves ridge from this estimator stabilize faster than the usual ridge estimators. This fact may facilitate the estimation of great value for the parameter, a central area of research in theory. Expressions were obtained for the total variance and the mean quadratic error. The performance of the proposed estimator was studied by applying a set of data already analyzed in the literature by using the usual ridge estimator. Data analysis of genomic selection on the pH response of pork at 45 minutes and 24 hours after slaughter. In all applications the results obtained by the proposed estimator are compared with the results obtained by the usual ridge and Mayer and Wilke (1973) estimator.

Keywords: Multivariate Analysis. Biased estimators. Multicollinearity. Ridge Regression.

## LISTA DE FIGURAS

Figura 1	Transformação linear de $\mathbb{R}^n$ em $\mathbb{R}^p$ . . . . .	19
Figura 2	Funcional linear de $\mathbb{R}^n$ em $\mathbb{R}$ . . . . .	20
Figura 3	Interpretação geométrica do produto interno . . . . .	21
Figura 4	Projeção ortogonal de um vetor aleatório . . . . .	24
Figura 5	Eixos principais do elipsoide . . . . .	31
Figura 6	Centro do elipsoide . . . . .	32
Figura 7	Elipsóides com pré-imagem de esferas . . . . .	33
Figura 8	Geometria do modelo linear . . . . .	36
Figura 9	Método geométrico dos quadrados mínimos . . . . .	39
Figura 10	As imagens e os espaços nulos de $A$ e de $A'$ . . . . .	41
Figura 11	Geometria das imagens e os espaços nulos de $A$ e de $A'$ . . . . .	42
Figura 12	Estimativas com a mesma soma de quadrados dos erros . . . . .	50
Figura 13	Triângulo retângulo fundamental . . . . .	52
Figura 14	Relação de um estimador com o estimador de quadrados mínimos . . . . .	53
Figura 15	Ajuste com a mesma soma de quadrados de resíduos . . . . .	54
Figura 16	Estimativas com a mesma soma de quadrados de resíduos . . . . .	55
Figura 17	Elipsoide centrado no estimador de quadrados mínimos . . . . .	57
Figura 18	Traço de cumeeira - caso ortogonal . . . . .	61
Figura 19	Traço de cumeeira . . . . .	62
Figura 20	Mudança para parâmetros canônicos . . . . .	63
Figura 21	Comportamento do EQM. Fonte: Hoerl e Kennard (1970a) . . . . .	79
Figura 22	Reparametrização linear . . . . .	89
Figura 23	Elipsoide centrado em $\hat{\beta}$ . . . . .	90
Figura 24	Interpretação geométrica do estimador de Mayer & Wilke . . . . .	91
Figura 25	Estudo de curvas de cumeeira . . . . .	93
Figura 26	Região de utilização do estimador de cumeeira de Rao . . . . .	94
Figura 27	Os estimadores de Hoerl & Kennard e de Mayer & Wilke . . . . .	95
Figura 28	Região de utilização do estimador de cumeeira de Rao . . . . .	96
Figura 29	Curvas traço de cumeeira do estimador de Hoerl & Kennard . . . . .	102
Figura 30	Curvas traço de cumeeira do estimador de Mayer & Wilke . . . . .	103
Figura 31	Curvas traço de cumeeira do estimador proposto . . . . .	104
Figura 32	Curvas traço de cumeeira do estimador de Hoerl & Kennard . . . . .	107
Figura 33	Curvas traço de cumeeira do estimador proposto . . . . .	109
Figura 34	Curvas traço de cumeeira do estimador de Mayer & Wilke . . . . .	110
Figura 35	Curvas de cumeeira resposta $pH_{45}$ estimador Hoerl & Kennard . . . . .	115
Figura 36	Curvas de cumeeira resposta $pH_{45}$ estimador Proposto . . . . .	116
Figura 37	Curvas de cumeeira resposta $pH_u$ estimador Hoerl & Kennard . . . . .	117

Figura 38	Curvas de cumeeira resposta $pH_u$ estimador Proposto . . . . .	118
Figura 39	Curvas de cumeeira resposta $pH_u$ fatores relevantes estimador Proposto . . . . .	119
Figura 40	Curvas de cumeeira resposta $pH_u$ fatores relevantes estimador Hoerl & Kennard . . . . .	120
Figura 41	Curvas de cumeeira resposta $pH_{45}$ fatores relevantes estimador Proposto . . . . .	121
Figura 42	Curvas de cumeeira resposta $pH_{45}$ fatores relevantes estimador Hoerl & Kennard . . . . .	122

## LISTA DE TABELAS

Tabela 1	Erro quadrático médio . . . . .	105
Tabela 2	Identificação dos SNPs reportados como sendo os mais relevantes para os fenótipos $pH_{(45)}$ (45 min após o abate) e $pH_u$ (24 horas após o abate) . . . . .	114

## SUMÁRIO

<b>1</b>	<b>INTRODUÇÃO</b>	15
<b>2</b>	<b>REFERÊNCIAL TEÓRICO</b>	18
<b>2.1</b>	<b>Vetores Aleatórios</b>	18
<b>2.2</b>	<b>Formas Quadráticas</b>	26
<b>2.3</b>	<b>Elipsoides</b>	29
<b>2.3.1</b>	<b>Regressão Linear Múltipla</b>	34
<b>2.4</b>	<b>Erro quadrático médio de estimadores</b>	36
<b>2.4.1</b>	<b>Estimador de quadrados mínimos</b>	38
<b>2.5</b>	<b>Regressão linear múltipla com coeficientes normalizados</b>	42
<b>2.5.1</b>	<b>Algumas propriedades dos estimadores de quadrados mínimos</b>	45
<b>2.6</b>	<b>Regressão de cumeeira - uma estimação viesada</b>	49
<b>2.7</b>	<b>Uma interpretação Bayesiana dos estimadores de cumeeira</b>	58
<b>2.8</b>	<b>Exemplos de estimadores de cumeeira</b>	60
<b>2.9</b>	<b>Propriedades dos estimadores de cumeeira</b>	66
<b>2.10</b>	<b>A procura pelo valor ótimo de k</b>	80
<b>2.11</b>	<b>Outros métodos de obtenção de estimadores de cumeeira</b>	84
<b>3</b>	<b>RESULTADOS E DISCUSSÃO</b>	87
<b>3.1</b>	<b>O estimador de Mayer &amp; Wilke</b>	87
<b>3.2</b>	<b>Uma proposta para um novo estimador de cumeeira de Rao</b>	91
<b>3.3</b>	<b>Erro quadrático médio do estimador proposto</b>	98
<b>3.4</b>	<b>Um exemplo por simulação computacional</b>	100
<b>3.5</b>	<b>Desempenho do estimador proposto aplicado a um conjunto de dados reais</b>	105
<b>3.6</b>	<b>Uma aplicação a dados de seleção genômica</b>	112
<b>4</b>	<b>CONCLUSÕES</b>	124
	<b>REFERÊNCIAS</b>	125

## 1 INTRODUÇÃO

O método de estimação, baseado em minimizar a soma de quadrados de resíduos, apesar de não apresentar viés e ter a menor variância possível pode não ser adequado quando em presença de quasecolinearidade. Uma maneira de se contornar esse problema é permitir a ocorrência de viés no processo de estimação. Hoerl e Kennard (1970a) propuseram um método, denominado regressão de cumeeira (ridge), que se mostrou adequado nesta situação, isto é, análise de dados em que as covariáveis são altamente correlacionadas. Alguns trabalhos tem apresentado resultados satisfatórios quando se utiliza a regressão de cumeeira (ridge), dentre eles Cule e Lorio (2012, 2013). Algebricamente o método é extremamente simples, sendo uma modificação linear dependente de um parâmetro real, denominado parâmetro de cumeeira do estimador de quadrados mínimos. Uma das vantagens deste método é que se pode plotar, para cada parâmetro estimado, uma curva em função do parâmetro de cumeeira, denominada curva traço de cumeeira que é um excelente método gráfico para a escolha do valor do parâmetro de cumeeira mais adequado. O método de regressão de cumeeira é um caso particular de se obter estimadores de encolhimento (*shrinkage*). As estimativas de quadrados mínimos são encolhidas em direção ao vetor nulo. Desta forma com uma escolha adequada do parâmetro de cumeeira tem-se um balanceamento entre variância e viés de forma que o erro quadrático médio do estimador de cumeeira se torna muitas vezes substancialmente menor que o erro quadrático médio do estimador de quadrados mínimos. Em razão de sua simplicidade algébrica, é comum que o usuário utilize este estimador sem se preocupar da forma com que ele foi obtido. Esta é uma atitude que pode levar a erros na compreensão da análise da regressão de cumeeira. O método de regressão de cumeeira é uma modificação

do método de quadrados mínimos penalizado, de fato é o método de quadrados mínimos que pode ser descrito geometricamente. Esta é a ênfase neste trabalho, explicitar os aspectos geométricos da teoria observados os dados e fixado um valor para o parâmetro de cumeeira. Fica determinado no espaço paramétrico um elipsoide, definido pelo mau delineamento, centrado na estimativa de quadrados mínimos. A estimativa de cumeeira é obtida tomando-se sobre este elipsoide o vetor mais próximo da origem. Posteriormente foi definido um estimador de cumeeira, denominado neste trabalho de estimador de cumeeira de Rao, que generaliza o anteriormente descrito. Escolhendo-se uma matriz definida positiva  $G$ , o estimador de cumeeira de Rao é obtido tomando-se o vetor definido pela intersecção do elipsoide centrado na estimativa de quadrados mínimos e outro elipsoide na origem definido pela matriz  $G$ . A contribuição deste trabalho é sugerir uma matriz  $G$  que define um elipsoide que possui os eixos principais centrados na estimativa de quadrados mínimos, mas com o tamanho destes eixos invertidos. Por exemplo, o eixo principal de maior tamanho de um elipsoide corresponde ao mesmo eixo só que com o menor tamanho. Pode-se afirmar apenas para efeitos visuais que os dois elipsoides são ortogonais. A estrutura do trabalho segue primeiramente um estudo bastante detalhado do estimador de cumeeira. As propriedades do estimador de cumeeira de Rao são abordadas de maneira mais rápida pois se tem um artigo do autor versando sobre este tópico. Outra questão que não foi abordada aqui é a de se obter estimadores para o valor do parâmetro de cumeeira que minimiza o erro quadrático médio do correspondente estimador. Este assunto por si só seria suficiente para uma tese dada a sua complexidade. Finalmente o estimador proposto é aplicado em três situações: um exemplo da literatura em que é estudado, entre outros, o comportamento do estimador de cumeeira, em um conjunto de dados reais também já analisados pelo uso do estimador de cumeeira e finalmente uma



aplicação a um conjunto de dados de seleção genômica analisados anteriormente pela técnica PLS (*partial least squares*). Deve-se observar que as aplicações não visavam um estudo computacional exaustivo das propriedades do estimador proposto, mas apenas mostrar que o estimador proposto apresenta boas propriedades e que merece um estudo mais apropriado.

## 2 REFERÊNCIAL TEÓRICO

### 2.1 Vetores Aleatórios

Quando se quer medir vários aspectos aleatórios de um fenômeno, ou mesmo fazer medidas repetidas de algum deles, tem-se a teoria das variáveis aleatórias multidimensionais,  $\mathbf{Y} = (Y_1, \dots, Y_n)$ , em que cada coordenada é uma variável aleatória unidimensional. O exemplo usual é quando se tem uma amostra aleatória simples  $\mathbf{Y} = (Y_1, \dots, Y_n)$ . Neste caso os  $Y_i$  são medidas independentes de uma mesma grandeza populacional. Outra situação é quando se tem uma variável multidimensional propriamente dita, isto é, medidas de várias características de um fenômeno aleatório. Pode-se também ter uma amostra de tais variáveis. Portanto, as coordenadas podem estar relacionadas a valores de características diferentes ou valores obtidos por repetição.

Os valores que  $\mathbf{Y} = (Y_1, \dots, Y_n)$  assume podem ser vistos como vetores no espaço  $\mathbb{R}^n$  e, portanto, podem também ser apropriadamente denominados de vetores aleatórios. Neste trabalho as variáveis aleatórias multidimensionais são vistas como vetores aleatórios e são estudadas suas projeções em subespaços vetoriais, utilizando-se uma abordagem inteiramente geométrica. Nos textos de estatística multivariada, o enfoque é muito mais algébrico do que geométrico. Um dos objetivos com este texto é justamente o de explicitar as vantagens da abordagem geométrica em razão de sua concisão e generalidade. Os conceitos e notação referentes à álgebra linear são os usuais nos livros textos. Quando uma base do espaço vetorial for escolhida será usada a notação  $\mathbb{R}^n$ . Caso o espaço vetorial seja abstrato, isto é, sem uma escolha particular de base, será utilizada a notação  $V$ . Uma observação essencial sobre a notação para que se evitem dúvidas é a de uma vez fixada bases no domínio e no contradomínio, uma transformação linear pode ser

representada por uma matriz, e não será feita distinção entre estes dois conceitos, transformações lineares e matrizes, estas serão tratadas de forma indistinta. Uma situação comum em estatística é a transformação linear de dados, como mudanças de escalas, por exemplo, ou o procedimento de se trabalhar com a média amostral. Tais procedimentos, em termos de álgebra linear, consistem em aplicar transformações lineares, fazer projeções ortogonais, calcular ângulos, etc, em vetores. Da mesma forma todos estes procedimentos podem ser aplicados em vetores aleatórios.

As transformações lineares podem ser representadas como na Figura 1.

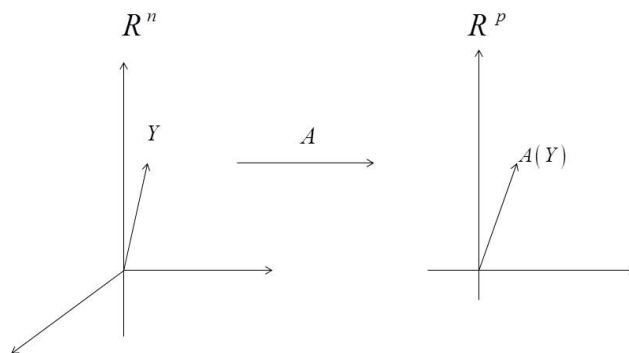


Figura 1 Transformação linear de  $\mathbb{R}^n$  em  $\mathbb{R}^p$

Um caso particular é o de funcionais lineares em que a transformação linear é definida por uma matriz  $n \times 1$ .

Aplicando esta transformação linear em um vetor aleatório  $\mathbf{Y}$  tem-se (Figura2):

$$A(\mathbf{Y}) = x_1 Y_1 + x_2 Y_2 + \dots + x_n Y_n$$

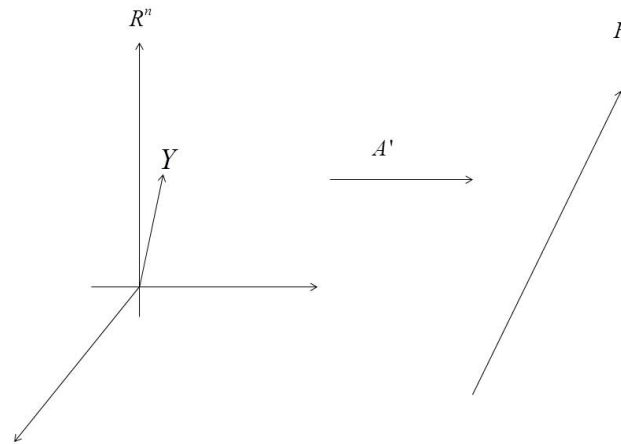


Figura 2 Funcional linear de  $\mathbb{R}^n$  em  $\mathbb{R}$

Considerando  $A$  uma matriz coluna e  $A'$  sua transposta, então, em termos de produto de matrizes, temos que o funcional linear pode ser expresso por:

$$\begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}$$

O produto interno do vetor aleatório com um vetor  $\mathbf{x} = (x_1, \dots, x_n)$ , é dado por  $\mathbf{x}'\mathbf{Y} = x_1Y_1 + \dots + x_nY_n$  ou pelas notações equivalentes  $\mathbf{x} \cdot \mathbf{Y}$  e  $\langle \mathbf{x}, \mathbf{Y} \rangle$ . Dessa forma o funcional linear pode ser expresso por  $A(\mathbf{Y}) = \mathbf{x}'\mathbf{Y}$ . Em termos de matrizes se  $\mathbf{x}$  é considerado uma matriz  $n \times 1$ , ou seja,

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

então  $\mathbf{x}'\mathbf{Y}$  pode ser expresso como o produto de matrizes :

$$\mathbf{x}'\mathbf{Y} = \begin{bmatrix} x_1 & x_2 & \cdots & x_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} .$$

As duas notações serão usadas indistintamente. A interpretação geométrica do produto interno descrita pela Figura 3 é  $\mathbf{x}'\mathbf{Y} = \|\mathbf{x}\| \cdot \|\mathbf{Y}\| \cdot \cos(\theta)$ , portanto  $\mathbf{x}'\mathbf{Y}$  significa o número  $\|\mathbf{x}\|$  vezes o tamanho do vetor aleatório obtido pela projeção do vetor  $\mathbf{Y}$  no subespaço unidimensional gerado pelo vetor  $\mathbf{x}$ .

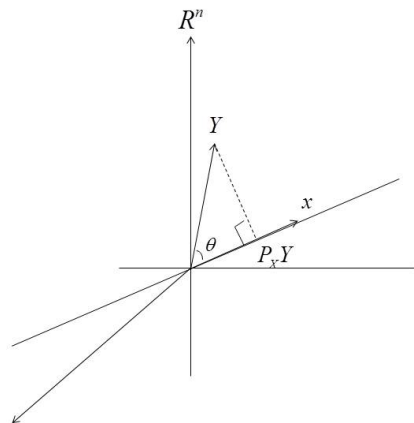


Figura 3 Interpretação geométrica do produto interno

A esperança de um vetor aleatório é o vetor dado pelas esperanças de cada componente  $E(\mathbf{Y}) = \boldsymbol{\beta} = [E(Y_1), \dots, E(Y_n)]'$ .

Outra característica fundamental de um vetor aleatório é sua matriz de variâncias e covariâncias também denominada matriz de dispersão, que será denotada  $cov(\mathbf{Y})$ .

Em termos matriciais a definição é  $cov(\mathbf{Y}) = E((\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))')$ . As relações entre o vetor esperança, a matriz de variâncias e covariâncias e transformações lineares são dadas por:

**Proposição 1.**  $E(\mathbf{x}'\mathbf{Y}) = \mathbf{x}'E(\mathbf{Y}) = \mathbf{x}'\boldsymbol{\beta}$

*Demonstração.*

$$\begin{aligned}
 E(\mathbf{x}'\mathbf{Y}) &= E(x_1Y_1 + x_2Y_2 + \dots + x_nY_n) \\
 &= x_1E(Y_1) + x_2E(Y_2) + \dots + x_nE(Y_n) \\
 &= \begin{bmatrix} x_1 & x_2 & \dots & x_n \end{bmatrix} \begin{bmatrix} E(Y_1) \\ E(Y_2) \\ \vdots \\ E(Y_n) \end{bmatrix} \\
 &= \mathbf{x}'E(\mathbf{Y}) \\
 &= \mathbf{x}'\boldsymbol{\beta}
 \end{aligned}$$

□

**Proposição 2.**  $cov(\mathbf{x}'\mathbf{Y}) = \mathbf{x}'cov(\mathbf{Y})\mathbf{x}$ . No caso particular em que  $cov(\mathbf{Y}) = \sigma^2I$  tem-se  $cov(\mathbf{x}'\mathbf{Y}) = \sigma^2\mathbf{x}'\mathbf{x} = \sigma^2\|\mathbf{x}\|^2$ .

*Demonstração.*

$$\begin{aligned}
 cov(\mathbf{x}'\mathbf{Y}) &= E((\mathbf{x}'\mathbf{Y} - E(\mathbf{x}'\mathbf{Y}))^2) \\
 &= E((\mathbf{x}'\mathbf{Y} - \mathbf{x}'E(\mathbf{Y}))'(\mathbf{x}'\mathbf{Y} - \mathbf{x}'E(\mathbf{Y}))) \\
 &= E((\mathbf{x}'(\mathbf{Y} - E(\mathbf{Y})))'(\mathbf{x}'(\mathbf{Y} - E(\mathbf{Y})))) \\
 &= E(\mathbf{x}'(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))'\mathbf{x}) \\
 &= \mathbf{x}'E((\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))')\mathbf{x} \\
 &= \mathbf{x}'cov(\mathbf{Y})\mathbf{x}
 \end{aligned}$$

□

**Proposição 3.** *Dados dois vetores  $\mathbf{x}$  e  $\mathbf{z}$ , a covariância entre  $\mathbf{x}'\mathbf{Y}$  e  $\mathbf{z}'\mathbf{Y}$  é dada por  $cov(\mathbf{x}'\mathbf{Y}, \mathbf{z}'\mathbf{Y}) = \mathbf{x}'cov(\mathbf{Y})\mathbf{z}$ .*

*Demonstração.*

$$\begin{aligned}
 cov(\mathbf{x}'\mathbf{Y}, \mathbf{z}'\mathbf{Y}) &= E((\mathbf{x}'\mathbf{Y} - E(\mathbf{x}'\mathbf{Y}))(\mathbf{z}'\mathbf{Y} - E(\mathbf{z}'\mathbf{Y}))) \\
 &= E((\mathbf{x}'\mathbf{Y} - \mathbf{x}'E(\mathbf{Y}))(\mathbf{z}'\mathbf{Y} - \mathbf{z}'E(\mathbf{Y}))) \\
 &= E(\mathbf{x}'(\mathbf{Y} - E(\mathbf{Y}))\mathbf{z}'(\mathbf{Y} - E(\mathbf{Y}))) \\
 &= E(\mathbf{x}'(\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))'\mathbf{z}) \\
 &= \mathbf{x}'E((\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))')\mathbf{z} \\
 &= \mathbf{x}'cov(\mathbf{Y})\mathbf{z}
 \end{aligned}$$

□

Note que se  $cov(\mathbf{Y}) = \sigma^2 I$  então  $cov(\mathbf{x}'\mathbf{Y}, \mathbf{z}'\mathbf{Y}) = \sigma^2 \mathbf{x}'\mathbf{z}$  e portanto se  $\mathbf{x}$  e  $\mathbf{z}$  são ortogonais, as variáveis  $\mathbf{x}'\mathbf{Y}$  e  $\mathbf{z}'\mathbf{Y}$  são não correlacionadas.

Alguns teoremas importantes envolvendo projetores ortogonais de um vetor aleatório  $\mathbf{Y}$  serão apresentados a seguir. Deve ser observado que sendo  $\mathbf{Y}$  um vetor aleatório em um espaço vetorial,  $V$  a sua projeção ortogonal em um subespaço  $W$  de  $V$ , denotada por  $P_W(\mathbf{Y})$ , também é um vetor aleatório, que depende obviamente de  $\mathbf{Y}$  (Figura 4). O vetor  $P_W(\mathbf{Y})$  pode ser expresso em relação a uma base ortogonal  $\mathbf{u}_1, \dots, \mathbf{u}_k$  de  $W$  por:

$$P_W(\mathbf{Y}) = \left( \frac{\mathbf{Y} \cdot \mathbf{u}_1}{\mathbf{u}_1 \cdot \mathbf{u}_1} \right) \mathbf{u}_1 + \left( \frac{\mathbf{Y} \cdot \mathbf{u}_2}{\mathbf{u}_2 \cdot \mathbf{u}_2} \right) \mathbf{u}_2 + \dots + \left( \frac{\mathbf{Y} \cdot \mathbf{u}_k}{\mathbf{u}_k \cdot \mathbf{u}_k} \right) \mathbf{u}_k$$

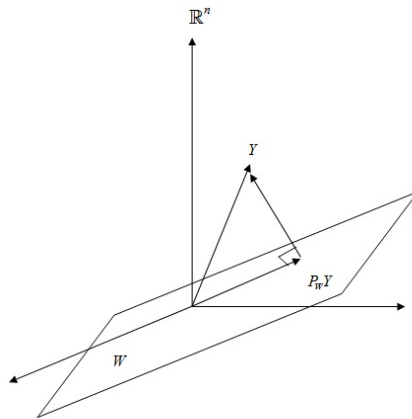


Figura 4 Projeção ortogonal de um vetor aleatório

Uma vez definida uma base no espaço vetorial  $V$  este pode ser naturalmente identificado no espaço  $\mathbb{R}^n$ . Um fato importante sobre projetores ortogonais é sua caracterização como matriz. Uma matriz quadrada tal que  $A^2 = A$  é chamada matriz de projeção ou projetor. Tem-se que um projetor  $A$  restrito à imagem de  $A$ ,  $Im(A)$ , é a identidade, ou seja,  $A(A\mathbf{x}) = A^2\mathbf{x} = A\mathbf{x}, \forall \mathbf{x} \in \mathbb{R}^n$ .  $I - A$  também é um projetor, pois  $(I - A)^2 = I - 2A + A^2 = I - A$ .

Como,



$$\begin{aligned}
A((I-A)\mathbf{x}) &= A(\mathbf{x}-A\mathbf{x}) = A\mathbf{x}-A^2\mathbf{x} = A\mathbf{x}-A\mathbf{x} = 0 \Rightarrow \text{Im}(I-A) \subset \text{ker}(A) \\
\mathbf{x} \in \text{ker}(A) &\Rightarrow (I-A)\mathbf{x} = \mathbf{x}-A\mathbf{x} = \mathbf{x}-0 \Rightarrow \mathbf{x} \in \text{Im}(I-A), \text{ segue então que} \\
&\text{Im}(I-A) = \text{ker}(A) \text{ e } \text{Im}(A) = \text{ker}(I-A).
\end{aligned}$$

Uma matriz de projeção  $A$  é dita um projetor ortogonal se  $A\mathbf{v} - \mathbf{v}$  é perpendicular ao subespaço  $\text{Im}(A)$  para todo  $\mathbf{v} \in \mathbb{R}^n$ .

**Proposição 4.** *Uma matriz de projeção é simétrica se, e somente se, é um projetor ortogonal*

$$\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle \Leftrightarrow \langle A\mathbf{v} - \mathbf{v}, A\mathbf{w} \rangle = 0$$

*Demonstração.* ( $\Rightarrow$ ) Se  $A$  é simétrica

$$\begin{aligned}
\langle \mathbf{v} - A\mathbf{v}, A\mathbf{w} \rangle &= \langle \mathbf{v}, A\mathbf{w} \rangle - \langle A\mathbf{v}, A\mathbf{w} \rangle \\
&= \langle \mathbf{v}, A\mathbf{w} \rangle - \langle \mathbf{v}, A^2\mathbf{w} \rangle \\
&= \langle \mathbf{v}, A\mathbf{w} \rangle - \langle \mathbf{v}, A\mathbf{w} \rangle \\
&= 0; \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n
\end{aligned}$$

( $\Leftarrow$ ) Se  $A$  é um projetor ortogonal,  $A\mathbf{v} - \mathbf{v}$  é perpendicular a  $A\mathbf{w}$ , isto é:  $\langle A\mathbf{v} - \mathbf{v}, A\mathbf{w} \rangle = 0 \Rightarrow \langle A\mathbf{v}, A\mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle$ , da mesma forma  $A\mathbf{w} - \mathbf{w}$  é perpendicular a  $A\mathbf{v}$ , isto é  $\langle A\mathbf{w} - \mathbf{w}, A\mathbf{v} \rangle = 0 \Rightarrow \langle A\mathbf{w}, A\mathbf{v} \rangle = \langle \mathbf{w}, A\mathbf{v} \rangle$ , portanto  $\langle A\mathbf{v}, \mathbf{w} \rangle = \langle \mathbf{v}, A\mathbf{w} \rangle; \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^n$ .

□

**Teorema 1.** *Considere que  $E(\mathbf{Y}) = \boldsymbol{\beta}$ ,  $\text{cov}(\mathbf{Y}) = \sigma^2 I$  e que  $W$  é um subespaço  $d$ -dimensional de  $V$ . Então:*

- $E(P_W(\mathbf{Y})) = P_W(E(\mathbf{Y})) = P_W(\boldsymbol{\beta})$

- $E(\|P_W(\mathbf{Y})\|^2) = \|P_W(\boldsymbol{\beta})\|^2 + d\sigma^2$

*Demonstração.* Seja  $\mathbf{w}_1, \dots, \mathbf{w}_d$  uma base ortonormal de  $W$ . Sem perda de generalidade tem-se:

$$P_W(\mathbf{Y}) = (\mathbf{w}_1 \cdot \mathbf{Y})\mathbf{w}_1 + \dots + (\mathbf{w}_d \cdot \mathbf{Y})\mathbf{w}_d$$

$$\|P_W(\mathbf{Y})\|^2 = (\mathbf{w}_1 \cdot \mathbf{Y})^2 + \dots + (\mathbf{w}_d \cdot \mathbf{Y})^2$$

Logo:

$$\begin{aligned} E(\|P_W(\mathbf{Y})\|^2) &= E((\mathbf{w}_1 \cdot \mathbf{Y})^2) + \dots + E((\mathbf{w}_d \cdot \mathbf{Y})^2) \\ &= \text{var}(\mathbf{w}_1 \cdot \mathbf{Y}) + (E(\mathbf{w}_1 \cdot \mathbf{Y}))^2 + \dots + \text{var}(\mathbf{w}_d \cdot \mathbf{Y}) + (E(\mathbf{w}_d \cdot \mathbf{Y}))^2 \\ &= \|\mathbf{w}_1\|^2 \sigma^2 + (\mathbf{w}_1 \cdot \boldsymbol{\beta})^2 + \dots + \|\mathbf{w}_d\|^2 \sigma^2 + (\mathbf{w}_d \cdot \boldsymbol{\beta})^2 \\ &= d\sigma^2 + \|P_W(\boldsymbol{\beta})\|^2 \end{aligned}$$

□

## 2.2 Formas Quadráticas

Formas quadráticas ocupam um lugar central em vários ramos da matemática, incluindo teoria dos números, álgebra linear, teoria dos grupos, geometria diferencial, e desempenha papel fundamental em estatística. Neste trabalho será necessário o uso da esperança e variância de formas quadráticas.

**Proposição 5.** Se  $E(\mathbf{Y}) = \boldsymbol{\beta}$  então  $E(\mathbf{Y}'A\mathbf{Y}) = \boldsymbol{\beta}'A\boldsymbol{\beta} + \text{tr}(A\text{cov}(\mathbf{Y}))$ .

*Demonstração.*

$$\begin{aligned}
 \mathbf{Y}'\mathbf{A}\mathbf{Y} &= \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} \\
 &= \begin{bmatrix} Y_1 & Y_2 & \cdots & Y_n \end{bmatrix} \begin{bmatrix} a_{11}Y_1 + a_{12}Y_2 + \cdots + a_{1n}Y_n \\ a_{21}Y_1 + a_{22}Y_2 + \cdots + a_{2n}Y_n \\ \vdots \\ a_{n1}Y_1 + a_{n2}Y_2 + \cdots + a_{nn}Y_n \end{bmatrix} \\
 &= Y_1(a_{11}Y_1 + \cdots + a_{1n}Y_n) + Y_2(a_{21}Y_1 + \cdots + a_{2n}Y_n) + \cdots + \\
 &+ Y_n(a_{n1}Y_1 + \cdots + a_{nn}Y_n) \\
 &= a_{11}Y_1Y_1 + a_{12}Y_1Y_2 + \cdots + a_{nn}Y_nY_n \\
 &= \sum_{j=1}^n \sum_{i=1}^n a_{ij}Y_iY_j
 \end{aligned}$$

Com isso, tem-se que:

$$\begin{aligned}
 E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) &= E\left(\sum_{j=1}^n \sum_{i=1}^n a_{ij}Y_iY_j\right) \\
 &= \sum_{j=1}^n \sum_{i=1}^n a_{ij}E(Y_iY_j)
 \end{aligned}$$

Definindo

$$\begin{aligned}
E((Y_i - \beta_i)(Y_j - \beta_j)) &= \sigma_{ij} \\
E(Y_i Y_j) - \beta_i E(Y_j) - \beta_j E(Y_i) + \beta_i \beta_j &= \sigma_{ij} \\
E(Y_i Y_j) - \beta_i \beta_j - \beta_j \beta_i + \beta_i \beta_j &= \sigma_{ij} \\
E(Y_i Y_j) &= \sigma_{ij} + \beta_i \beta_j
\end{aligned}$$

$$\begin{aligned}
&= E\left(\sum_{j=1}^n \sum_{i=1}^n a_{ij} (\sigma_{ij} + \beta_i \beta_j)\right) \\
&= \underbrace{\sum_{j=1}^n \sum_{i=1}^n a_{ij} \sigma_{ij}}_{\text{tr}(A \text{cov}(\mathbf{Y}))} + \underbrace{\sum_{j=1}^n \sum_{i=1}^n a_{ij} \beta_i \beta_j}_{\boldsymbol{\beta}' A \boldsymbol{\beta}}
\end{aligned}$$

Pois

$$\begin{aligned}
\text{cov}(\mathbf{Y}) &= E((\mathbf{Y} - E(\mathbf{Y}))(\mathbf{Y} - E(\mathbf{Y}))') \\
&= E((\mathbf{Y} - X\boldsymbol{\beta})(\mathbf{Y} - X\boldsymbol{\beta})') \\
&= E\left(\begin{pmatrix} Y_1 - \beta_1 \\ Y_2 - \beta_2 \\ \vdots \\ Y_n - \beta_n \end{pmatrix} \begin{pmatrix} Y - \beta_1 & Y - \beta_2 & \cdots & Y - \beta_n \end{pmatrix}\right) \\
&= (E((Y_i - \beta_i)(Y_j - \beta_j)))_{ij} \\
&= (\sigma_{ij})_{ij} \quad \forall i, j = 1, 2, \dots, n
\end{aligned}$$

Observe que  $\sum_{j=1}^n \sum_{i=1}^n a_{ij} \sigma_{ij} = \text{tr}(A \text{cov}(\mathbf{Y}))$ , pois:

$$\begin{bmatrix} a_{11} & a_{12} & \cdots & a_{1n} \\ a_{21} & a_{22} & \cdots & a_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ a_{n1} & a_{n2} & \cdots & a_{nn} \end{bmatrix} \begin{bmatrix} \sigma_{11} & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{21} & \sigma_{22} & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{n1} & \sigma_{n2} & \cdots & \sigma_{nn} \end{bmatrix} = \\
\begin{bmatrix} a_{11}\sigma_{11} + \cdots + a_{1n}\sigma_{n1} & \cdots & a_{11}\sigma_{1n} + \cdots + a_{1n}\sigma_{nn} \\ \vdots & \ddots & \vdots \\ \vdots & \ddots & \vdots \\ a_{n1}\sigma_{11} + \cdots + a_{nn}\sigma_{nn} & \cdots & a_{n1}\sigma_{1n} + \cdots + a_{nn}\sigma_{nn} \end{bmatrix}$$

Logo,  $E(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = \boldsymbol{\beta}'\mathbf{A}\boldsymbol{\beta} + tr(\mathbf{A}cov(\mathbf{Y}))$ . □

**Proposição 6.** Se  $\mathbf{Y}$  é variável aleatória de uma normal multivariada com  $cov(\mathbf{Y}) = \boldsymbol{\Sigma}$ ,  $N(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ , então  $var(\mathbf{Y}'\mathbf{A}\mathbf{Y}) = 2tr[(\mathbf{A}\boldsymbol{\Sigma})^2] + 4\boldsymbol{\beta}'\mathbf{A}\boldsymbol{\Sigma}\boldsymbol{\beta}$

*Demonstração.* (RENCHEER, 2008, p. 109) □

### 2.3 Elipsoides

Para a compreensão das propriedades dos estimadores *ridge* e de outros estimadores de encolhimento é útil uma abordagem geométrica dos elipsoides. Se  $A_{n \times n}$  é simétrica positiva definida, as curvas de nível da função  $g : \mathbb{R}^n \rightarrow \mathbb{R}$ ,  $g(\mathbf{x}) = \mathbf{x}'\mathbf{A}\mathbf{x} = \sum_{i,j} a_{ij}x_ix_j$ , isto é, a equação  $\mathbf{x}'\mathbf{A}\mathbf{x} = c$  define um elipsoide. Com uma mudança de coordenadas os elipsoides podem ser expressos de forma mais simples. Como  $A$  é simétrica existe uma matriz  $P_{n \times n}$  ortogonal, tal que:

$$P'AP = \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix}$$

Se  $\mathbf{y} = P'\mathbf{x}$ ,  $\mathbf{x} = P\mathbf{y}$  nas novas coordenadas, então:

$$\begin{aligned}c &= \mathbf{x}'A\mathbf{x} \\&= (P\mathbf{y})'A(P\mathbf{y}) \\&= \mathbf{y}'P'AP\mathbf{y} \\&= \mathbf{y}' \begin{bmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{bmatrix} \mathbf{y} \\&= \sum \lambda_i y_i^2\end{aligned}$$

Portanto,

$$\frac{y_1^2}{\frac{c}{\lambda_1}} + \dots + \frac{y_n^2}{\frac{c}{\lambda_n}} = 1$$

Segue daí a descrição do elipsoide em termos dos eixos principais como descrito na Figura 5:

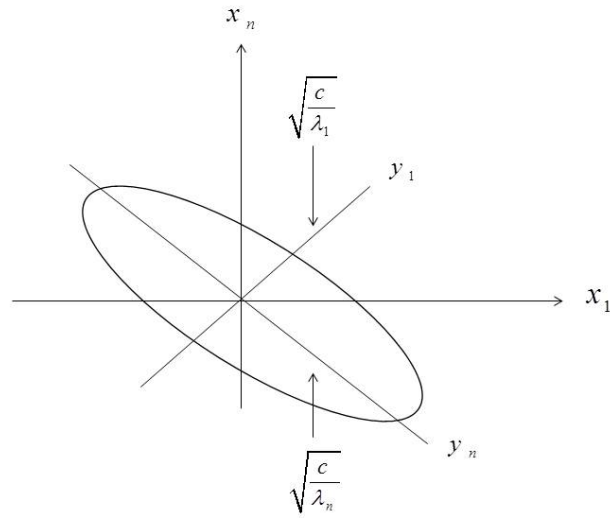


Figura 5 Eixos principais do elipsoide

De uma maneira mais geral fazendo-se um deslocamento temos os elipsoides  $(\mathbf{x} - \mathbf{x}_0)'A(\mathbf{x} - \mathbf{x}_0) = c$ , como descrito na descrito na Figura 6:

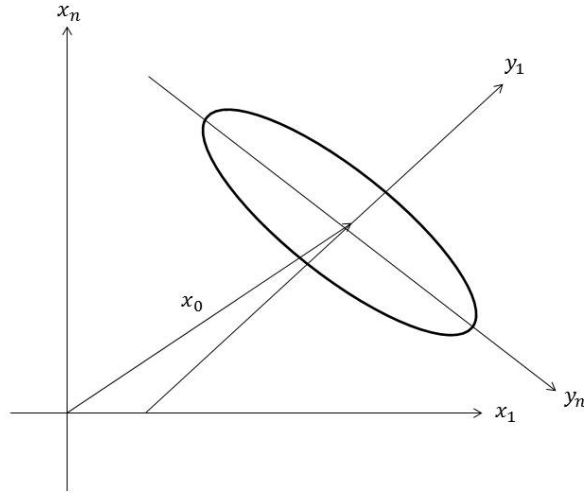


Figura 6 Centro do elipsoide

Quando se tem uma transformação linear de posto completo  $A : \mathbb{R}^n \rightarrow \mathbb{R}^n$ ,  $\mathbf{x} \rightarrow A(\mathbf{x}) = \mathbf{y}$ , os elipsoides descrevem a pré-imagem de esferas. Se  $S = \{\|\mathbf{y} - \mathbf{y}_0\| = r, \mathbf{y} \in \mathbb{R}^n\}$ , se  $\mathbf{y} = A\mathbf{x}$  então:

$$\begin{aligned}
 r^2 &= \|\mathbf{y} - \mathbf{y}_0\|^2 \\
 &= (\mathbf{y} - \mathbf{y}_0)' (\mathbf{y} - \mathbf{y}_0) \\
 &= (A(\mathbf{x}) - A(\mathbf{x}_0))' (A(\mathbf{x}) - A(\mathbf{x}_0)) \\
 &= (\mathbf{x} - \mathbf{x}_0)' A' A (\mathbf{x} - \mathbf{x}_0)
 \end{aligned}$$

Quando  $A : \mathbb{R}^p \rightarrow \mathbb{R}^n$ ,  $A(\boldsymbol{\beta}) = \mathbf{y}$  é de posto completo e  $n > p$  ainda se tem um resultado semelhante. Considere a esfera  $\|\mathbf{y} - \mathbf{y}_0\| = r$  contida na  $Im A$ .



Então:

$$\begin{aligned}
 r^2 &= \|\mathbf{y} - \mathbf{y}_0\|^2 \\
 &= (\mathbf{y} - \mathbf{y}_0)' (\mathbf{y} - \mathbf{y}_0) \\
 &= (A(\boldsymbol{\beta}) - A(\boldsymbol{\beta}_0))' (A(\boldsymbol{\beta}) - A(\boldsymbol{\beta}_0)) \\
 &= (\boldsymbol{\beta} - \boldsymbol{\beta}_0)' A' A (\boldsymbol{\beta} - \boldsymbol{\beta}_0).
 \end{aligned}$$

Esta equação define um elipsoide em  $\mathbb{R}^p$  centrado em  $\boldsymbol{\beta}_0$  como descrito na Figura 7

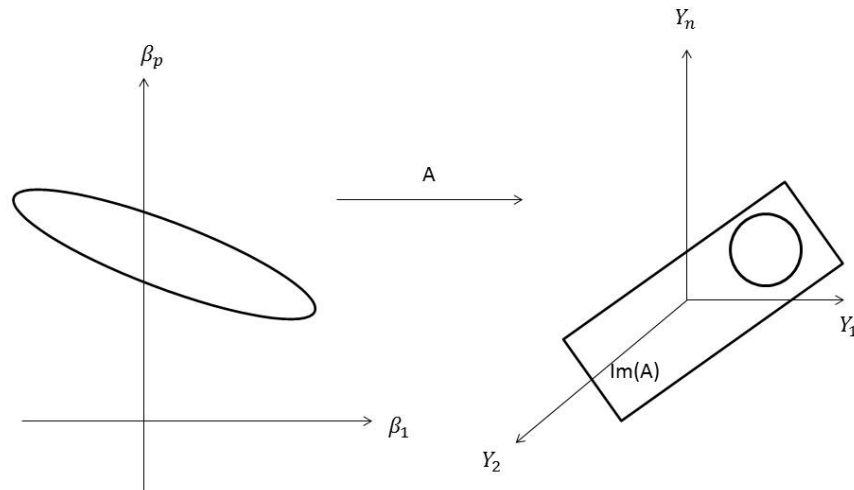


Figura 7 Elipsoides com pré-imagem de esferas

Note que o inverso também é verdadeiro. A imagem de certos elipsoides são esferas. Na situação anterior o elipsoide  $(\boldsymbol{\beta} - \boldsymbol{\beta}_0)' A' A (\boldsymbol{\beta} - \boldsymbol{\beta}_0) = r^2$  tem como imagem a esfera  $\|A\boldsymbol{\beta} - A\boldsymbol{\beta}_0\| = r$  no subespaço  $Im A$ .

### 2.3.1 Regressão Linear Múltipla

Esta seção baseia-se em Rencher (2008). Na regressão múltipla procura-se prever uma variável resposta (ou seja, dependente)  $\mathbf{Y}$ , assumindo que ela tem uma relação linear com diversas variáveis independentes  $x_1, x_2, \dots, x_p$ .

O modelo de regressão linear múltipla pode ser expresso como:

$$\mathbf{Y} = \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p + \varepsilon. \quad (1)$$

Uma vez que os  $\beta$ 's podem ser estimados pelo método dos mínimos quadrados e para estimá-los, é utilizada uma amostra com  $n$  observações de  $\mathbf{Y}$  associados com os  $x$ 's. Tem-se, portanto para a  $i$ -ésima observação:

$$Y_i = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_i. \quad (2)$$

Observe que se  $x_{i1} = 1 \forall i = 1, \dots, n$  o coeficiente  $\beta_1$  é denominado intercepto. É comum na literatura que o intercepto seja denominado por  $\beta_0$ . No intuito de simplificar a variação dos contadores, denotar o intercepto por  $\beta_1$  implica que o espaço paramétrico tenha dimensão  $p$  e não dimensão  $p + 1$  caso se optasse pela notação utilizando  $\beta_0$ . As pressuposições do modelo são:

1.  $E(\varepsilon_i) = 0, \forall i = 1, 2, \dots, n$ , isto é,  $E(Y_i) = \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip}$ .
2.  $\text{var}(\varepsilon_i) = \sigma^2$ , isto é,  $\text{var}(Y_i) = \sigma^2$ .
3.  $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ , isto é,  $\text{cov}(Y_i, Y_j) = 0 \forall i \neq j$ .

De acordo com a suposição 1 tem-se que todos os  $x$ 's estão incluídos no modelo de forma linear; a suposição 2 estabelece que a variância de  $Y$  é constante e não depende dos  $\mathbf{x}$ 's e da suposição 3 os  $Y$ 's não são correlacionados entre si, o que geralmente acontece em amostras aleatórias.

A identidade (2) pode ser reescrita em sua forma matricial:

$$\begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix} = \begin{bmatrix} x_{11} & x_{12} & \cdots & x_{1p} \\ x_{21} & x_{22} & \cdots & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n1} & x_{n2} & \cdots & x_{np} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

ou

$$\mathbf{Y}_{n \times 1} = \mathbf{X}_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1} \quad (3)$$

e então as três suposições matricialmente são dadas por:

1.  $E(\boldsymbol{\varepsilon}) = 0$ , isto é,  $E(\mathbf{Y}) = \mathbf{X}\boldsymbol{\beta}$ .
2.  $cov(\boldsymbol{\varepsilon})_{n \times n} = \sigma^2 I$ , isto é,  $cov(\mathbf{Y}) = \sigma^2 I$ .

Assumiremos que  $n > p$  e  $posto(X) = p$ . Se  $n < p$  ou se existir alguma relação linear entre os vetores coluna da matriz  $X$  não terá posto (coluna) completo. Se os valores  $x_{ij}$  da matriz  $X$  forem planejados (escolhidos) pelo pesquisador, a matriz  $X$  é denominada de matriz do delineamento.

Neste trabalho foi adotada uma abordagem geométrica na qual foram utilizados de forma indistinta os conceitos de matriz e transformação linear. Quando fixadas bases no domínio e no contra-domínio, uma transformação linear é representada por uma matriz. Tal identificação não gerará dúvidas, pois as bases escolhidas estarão bem determinadas. Neste sentido temos a representação geométrica do modelo linear na Figura 8 em que  $Im(X)$  é o subespaço vetorial definido pela imagem da transformação linear  $X$ :

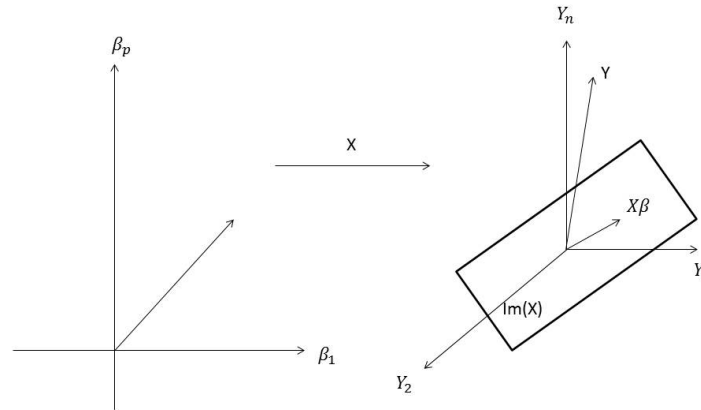


Figura 8 Geometria do modelo linear

#### 2.4 Erro quadrático médio de estimadores

Dado um estimador  $\hat{\beta}$  de um vetor de parâmetros  $\beta$ , seu erro quadrático é definido por

$$L_1^2 = (\hat{\beta} - \beta)'(\hat{\beta} - \beta). \quad (4)$$

O erro quadrático médio é definido como  $EQM(\hat{\beta}) = E(L_1^2)$ . O erro quadrático médio de um estimador pode ser decomposto em termos de variância e do quadrado do viés.

##### Proposição 7.

$$\begin{aligned} EQM(\hat{\beta}) &= tr(cov(\hat{\beta})) + \|E(\hat{\beta}) - \beta\|^2 \\ &= \sum_{i=1}^p var(\hat{\beta}_i) + \|E(\hat{\beta}) - \beta\|^2 \end{aligned}$$

*Demonstração.*

$$\begin{aligned}
EQM(\hat{\beta}) &= E((\hat{\beta} - \beta)'(\hat{\beta} - \beta)) \\
&= E((\hat{\beta} - E(\hat{\beta}) + E(\hat{\beta}) - \beta)'(\hat{\beta} - E(\hat{\beta}) + E(\hat{\beta}) - \beta)) \\
&= E((\hat{\beta} - E(\hat{\beta}))'(\hat{\beta} - E(\hat{\beta})) + (\hat{\beta} - E(\hat{\beta}))'(E(\hat{\beta}) - \beta) \\
&\quad + (E(\hat{\beta}) - \beta)'(\hat{\beta} - E(\hat{\beta})) + (E(\hat{\beta}) - \beta)'(E(\hat{\beta}) - \beta)) \\
&= E((\hat{\beta} - E(\hat{\beta}))'(\hat{\beta} - E(\hat{\beta}))) + (E(\hat{\beta}) - \beta)'(E(\hat{\beta}) - \beta) \\
&\quad + (E(\hat{\beta}) - \beta)'E(\hat{\beta} - E(\hat{\beta})) + (E(\hat{\beta}) - \beta)'(E(\hat{\beta}) - \beta) \\
&= E(\hat{\beta} - E(\hat{\beta}))'(\hat{\beta} - E(\hat{\beta})) + (E(\hat{\beta}) - \beta)'(E(\hat{\beta}) - \beta) \\
&= E(\text{tr}((\hat{\beta} - E(\hat{\beta}))(\hat{\beta} - E(\hat{\beta}))) + \|E(\hat{\beta}) - \beta\|^2 \\
&= \text{tr}(\text{cov}(\hat{\beta})) + \|E(\hat{\beta}) - \beta\|^2 \\
&= \sum_{i=1}^p \text{var}(\hat{\beta}_i) + \sum_{i=1}^p (\text{viés}(\beta_i))^2
\end{aligned}$$

□

OBS.: A proposição 7 também pode ser demonstrada utilizando-se a proposição 5. Veja que mesmo tendo-se um estimador não viesado  $\hat{\beta}$  do vetor  $\beta$ , a norma ao quadrado deste estimador é um estimador viesado da norma ao quadrado do vetor de parâmetros, de fato determina uma superestimação.

**Proposição 8.** Se  $\hat{\beta}$  é um estimador não viesado de  $\beta$ , então  $E(\|\hat{\beta}\|^2) = EQM(\hat{\beta}) + \|\beta\|^2$ .

*Demonstração.*

$$\begin{aligned}
 E(\|\hat{\beta}\|^2) &= E\left(\sum_{i=1}^p \hat{\beta}_i^2\right) \\
 &= \sum_{i=1}^p E(\hat{\beta}_i^2) \\
 &= \sum_{i=1}^p \left(\text{var}(\hat{\beta}_i) + (E(\hat{\beta}_i))^2\right) \\
 &= \sum_{i=1}^p \text{var}(\hat{\beta}_i) + \sum_{i=1}^p (E(\hat{\beta}_i))^2 \\
 &= \sum_{i=1}^p \text{var}(\hat{\beta}_i) + \sum_{i=1}^p \beta_i^2 \\
 &= \sum_{i=1}^p \text{var}(\hat{\beta}_i) + \|\beta\|^2
 \end{aligned}$$

□

OBS.: A proposição 8 também pode ser demonstrada utilizando-se a proposição 5.

#### 2.4.1 Estimador de quadrados mínimos

Será discutida aqui a abordagem de quadrados mínimos para estimar os  $\beta$ 's apresentados no modelo (1) e nenhuma suposição sobre a distribuição de  $Y$  será necessária na obtenção destes estimadores. Logo, procura-se os estimadores de  $\beta_1, \beta_2, \dots, \beta_p$  que minimizam a soma de quadrados dos desvios dos  $n$  valores observados de  $\mathbf{Y}$  em relação aos preditos, ou seja, busca-se  $\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p$  tais que  $\hat{y} = \hat{\beta}_1 x_1 + \hat{\beta}_2 x_2 + \dots + \hat{\beta}_p x_p$  que minimizem

$$\begin{aligned} \sum_{i=1}^n \hat{\varepsilon}_i^2 &= \sum_{i=1}^n (y_i - \hat{y}_i)^2 \\ &= \sum_{i=1}^n (y_i - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} - \dots - \hat{\beta}_p x_{ip})^2 \end{aligned} \quad (5)$$

Vetorialmente deve-se obter o vetor definido pela projeção ortogonal do vetor  $\mathbf{y}$  observado em  $Im(X)$ ,  $\hat{\mathbf{y}} = P_{Im(X)}(\mathbf{y})$ , uma vez que a projeção ortogonal minimiza a distância do vetor  $\mathbf{y}$  ao subespaço de  $Im(X)$ . (Figura 9)

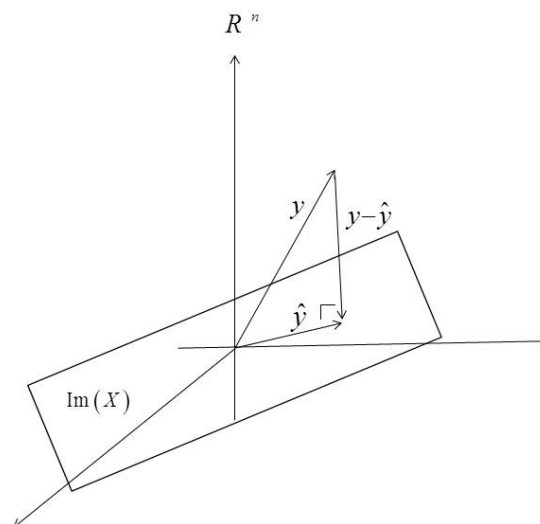


Figura 9 Método geométrico dos quadrados mínimos

A obtenção dos estimadores de quadrados mínimos como sendo aquela que minimiza a distância representada na Figura 9 pode ser demonstrada geometricamente com o seguinte teorema:

**Teorema 2.** *Se  $\mathbf{Y} = X\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ , em que  $X$  é  $n \times p$  de posto  $p < n$ , então o valor de  $\hat{\boldsymbol{\beta}} = (\hat{\beta}_1, \hat{\beta}_2, \dots, \hat{\beta}_p)'$  que minimiza (5) é igual à*

$$\hat{\boldsymbol{\beta}} = (X'X)^{-1}X'\mathbf{Y} \quad (6)$$

*Demonstração.* Seja  $Im(X) \subset \mathbb{R}^n$  o espaço coluna ou espaço imagem de  $X$ , isto é, se  $\mathbf{b} \in \mathbb{R}^p$  então  $X\mathbf{b} \in Im(X)$ ;  $ker(X) \subset \mathbb{R}^p$  o espaço nulo (kernel) de  $X$ , ou seja, se  $\mathbf{b} \in ker(X)$  então  $X\mathbf{b} = 0$ . Utilizando o resultado fundamental da álgebra linear: Se  $A_{n \times p}$  é a matriz de uma transformação linear de  $\mathbb{R}^p$  em  $\mathbb{R}^n$ , então o espaço nulo e  $A$ ,  $ker(A)$ , é ortogonal ao espaço imagem de  $A'$ ,  $Im(A')$ . Tem-se também que os espaços  $Im(A)$  e  $ker(A')$  são ortogonais (Figura 10).

No nosso caso específico, como a transformação  $X$  é de posto coluna completo o espaço nulo de  $X$  restringe-se à origem de  $\mathbb{R}^p$  e sua imagem tem dimensão  $p$ . Decorre também da imagem de  $X'$ , a qual posto  $p$  e, portanto, é o próprio  $\mathbb{R}^p$ .

Considere um vetor de observações  $\mathbf{Y} = (Y_1, Y_2, \dots, Y_n)'$ . Em razão dos erros experimentais a probabilidade  $P(\mathbf{Y} \in Im(X)) = 0$ . Como queremos minimizar  $L(\mathbf{b}) = \|\mathbf{Y} - X\mathbf{b}\|^2$  e o vetor  $X\mathbf{b}$  está na  $Im(X)$ , segue que  $L(\mathbf{b})$  é mínimo quando  $X\mathbf{b}$  é a projeção ortogonal de  $\mathbf{Y}$  em  $Im(X)$ , com notação  $P_{Im(X)}(\mathbf{Y})$ , e conseqüentemente, existe um único vetor de parâmetros  $\hat{\boldsymbol{\beta}}$  tal que  $X\hat{\boldsymbol{\beta}} = P_{Im(X)}(\mathbf{Y})$  (Equações normais na forma geométrica). Considere agora o vetor de parâmetros  $\mathbf{b} = X'P_{Im(X)}(\mathbf{Y})$  (Figura 11).

Observe que, como  $\mathbf{Y} - P_{Im(X)}(\mathbf{Y})$  é ortogonal a qualquer vetor na ima-



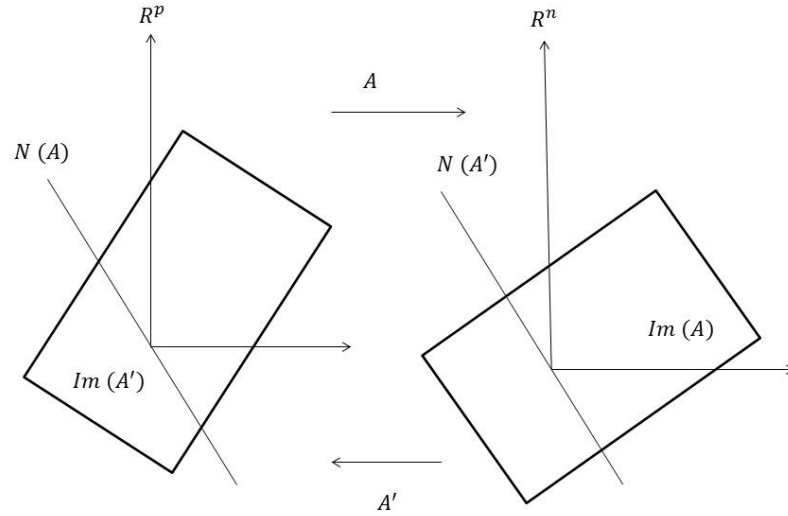


Figura 10 As imagens e os espaços nulos de  $A$  e de  $A'$

gem de  $X$ ,  $\mathbf{Y} - P_{Im(X)}\mathbf{Y}$  está no espaço nulo de  $X'$ . Segue que  $X'(\mathbf{Y} - P_{Im(X)}\mathbf{Y}) = 0$  e, portanto, podemos escrever:

$$\mathbf{b} = X'P_{Im(X)}\mathbf{Y} = X'P_{Im(X)}\mathbf{Y} + X'(\mathbf{Y} - P_{Im(X)}\mathbf{Y}) = X'\mathbf{Y} \quad (7)$$

Como  $X\hat{\boldsymbol{\beta}} = P_{Im(X)}\mathbf{Y}$ , podemos também escrever:

$$\mathbf{b} = X'X\hat{\boldsymbol{\beta}}. \quad (8)$$

Considerando (7) e (8):

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

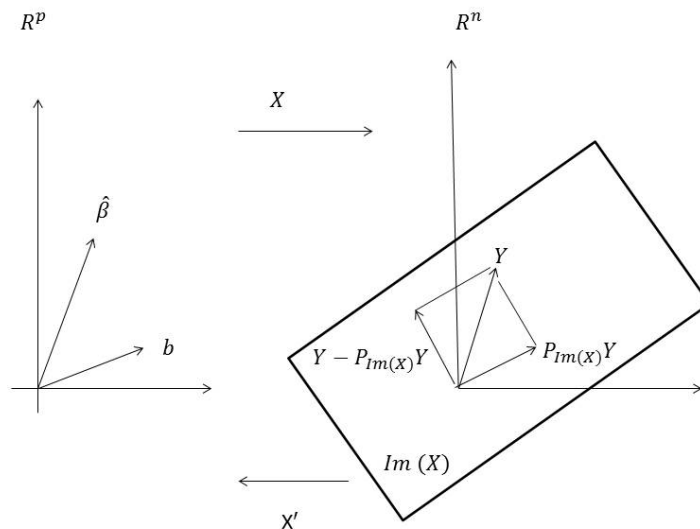


Figura 11 Geometria das imagens e os espaços nulos de  $A$  e de  $A'$

□

## 2.5 Regressão linear múltipla com coeficientes normalizados

Quando se tem um modelo linear  $Y = X\beta + \varepsilon$ , a matriz  $X$  dos valores das covariáveis é, em algumas situações, experimental totalmente determinada pelo pesquisador. No entanto, em outras não se tem um controle total sobre seus valores. Mas, em ambas as situações, para a análise do modelo, a matriz  $X$  pode ser alterada para que se tenha uma expressão adequada que ajude na interpretação do modelo. A primeira modificação na matriz  $X$  é que esta pode ser centralizada, isto é, pode ser modificada para que cada uma de suas colunas tenha soma zero. A transformação é  $X_c = X - \frac{1}{n}JX = \left(I - \frac{1}{n}J\right)X$ , em que  $J$  é uma matriz com todas as entradas iguais a 1, neste caso o modelo é denominado modelo na forma

centralizada (RENCHEER, 2008, p.154).

Suponhamos então, que  $X = (x_{ij})$ ,  $i = 1, 2, \dots, n$  e  $j = 1, 2, \dots, p$  esteja na forma centralizada, isto é,  $\sum_{i=1}^p x_{ij} = 0$ ,  $\forall j$ . Neste caso as entradas da matriz  $X'X = (a_{ij})$ ;  $i, j = 1, 2, \dots, n$  são da forma  $a_{ij} = \sum_{s=1}^n x_{si}x_{sj}$ . Se as covariáveis  $X_i$  e  $X_j$  são consideradas variáveis aleatórias, a  $i$ -ésima coluna e a  $j$ -ésima coluna de  $X$  podem ser consideradas amostras de  $X_i$  e  $X_j$ , respectivamente. Logo  $a_{ij} = \sum_{s=1}^n x_{si}x_{sj}$  é a covariância amostral entre  $X_i$  e  $X_j$  e é pelo método dos momentos um estimador de  $cov(X_i, X_j)$ . Observe que  $a_{ii} = \sum x_{si}^2$  é a variância amostral de  $X_i$ . Desta forma a matriz  $X'X$  pode ser descrita como uma matriz de variâncias-covariâncias amostrais.

A ideia agora é modificar  $X$  para que se tenha no lugar de covariâncias; correlação. Para isto é necessário uma reparametrização do modelo. Considere a matriz diagonal  $D = (d_{ij})$ ,  $d_{ii} = \frac{1}{\sqrt{a_{ii}}}$ ,  $d_{ij} = 0$ ,  $i \neq j$ .

O modelo reparametrizado é:

$$\mathbf{Y} = XD\boldsymbol{\theta} + \boldsymbol{\varepsilon} = \tilde{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$$

$$\tilde{X}'\tilde{X} = (XD)'(XD) = D'(X'X)D = (c_{ij}).$$

É fácil ver que  $c_{ij} = \frac{a_{ij}}{\sqrt{a_{ii}}\sqrt{a_{jj}}}$  e portanto o coeficiente de correlação amostral entre  $X_i$  e  $X_j$ . Para o modelo  $\tilde{X}\boldsymbol{\theta} + \boldsymbol{\varepsilon}$  o estimador de quadrados mínimos

é:

$$\begin{aligned}
 \hat{\boldsymbol{\theta}} &= (\tilde{X}'\tilde{X})^{-1}\tilde{X}'\mathbf{Y} \\
 &= ((XD)'(XD))^{-1}(XD)'\mathbf{Y} \\
 &= (D'X'XD)^{-1}D'X'\mathbf{Y} \\
 &= D^{-1}(X'X)^{-1}DD^{-1}X'\mathbf{Y} \\
 &= D^{-1}(X'X)^{-1}X'\mathbf{Y} \\
 &= D^{-1}\hat{\boldsymbol{\beta}}
 \end{aligned}$$

Logo a equação de predição fica da forma:

$$\hat{y} = \hat{\boldsymbol{\theta}}'\tilde{X} = \hat{\theta}_1\tilde{x}_1 + \dots + \hat{\theta}_p\tilde{x}_p.$$

Observe que as covariáveis  $\tilde{x}_i$  são relativas à matriz  $\tilde{X}$ , isto é, não são as unidades em que o experimento foi feito. A relação entre as covariáveis  $x_i$  e  $\tilde{x}_i$  seguem da relação  $\tilde{X} = XD$  o que implica  $\tilde{x}_i = \frac{x_i}{\sqrt{a_{ii}}}$ . Como  $\hat{\theta}_i = \sqrt{a_{ii}}\hat{\beta}_i$

$$\begin{aligned}
 \hat{y} &= \hat{\boldsymbol{\theta}}'\tilde{X} = \hat{\theta}_1\tilde{x}_1 + \dots + \hat{\theta}_p\tilde{x}_p \\
 &= \sqrt{a_{11}}\hat{\beta}_1 \frac{x_1}{\sqrt{a_{11}}} + \dots + \sqrt{a_{pp}}\hat{\beta}_p \frac{x_p}{\sqrt{a_{pp}}} \\
 &= \hat{\beta}_1 x_1 + \dots + \hat{\beta}_p x_p
 \end{aligned}$$

Portanto, a equação de predição é invariável em relação a esta reparametrização. Em razão deste fato é usual se trabalhar com a matriz  $X$  tal que  $X'X$ , esteja na forma de correlação, pois desta forma temos uma boa descrição amostral da correlação entre as covariáveis, o que nos permite uma melhor compreensão de como a variável resposta é afetada pelas covariáveis.

### 2.5.1 Algumas propriedades dos estimadores de quadrados mínimos

Nesta seção serão abordadas algumas propriedades dos estimadores de quadrados mínimos que serão úteis para demonstrações futuras.

**Propriedade 1.** *O estimador de quadrados mínimos é não viesado.*

*Demonstração.*

$$\begin{aligned}
 E(\hat{\beta}) &= E((X'X)^{-1}X'\mathbf{Y}) \\
 &= (X'X)^{-1}X'E(\mathbf{Y}) \\
 &= \underbrace{(X'X)^{-1}X'X}_{I}\beta \\
 &= \beta
 \end{aligned}$$

□

**Propriedade 2.** *Se  $\text{cov}(\mathbf{Y}) = \sigma^2 I$ , a matriz de covariâncias de  $\hat{\beta}$  é  $\sigma^2(X'X)^{-1}$ .*

*Demonstração.*

$$\begin{aligned}
 \text{cov}(\hat{\beta}) &= \text{cov}((X'X)^{-1}X'\mathbf{Y}) \\
 &= (X'X)^{-1}X'\text{cov}(\mathbf{Y})((X'X)^{-1}X')' \\
 &= (X'X)^{-1}X'\sigma^2 I X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}X'X(X'X)^{-1} \\
 &= \sigma^2(X'X)^{-1}
 \end{aligned}$$

□

**Teorema 3.** *(Teorema de Gauss-Markov): Se  $E(\mathbf{Y}) = X\beta$  e  $\text{cov}(\mathbf{Y}) = \sigma^2 I$ , os estimadores de quadrados mínimos  $\hat{\beta}_j$ , para  $j = 1, 2, \dots, p$ , têm variância mínima dentre*

*todos os estimadores lineares não viesados.*

*Demonstração.* (RENCHEER, 2008, p. 146) □

A matriz  $X'X$  é simétrica e positiva definida. Portanto, admite  $p$  autovalores  $\lambda_{max} = \lambda_1 \geq \lambda_2 \geq \dots \geq \lambda_p = \lambda_{min} > 0$ .

**Proposição 9.** *Para o estimador de quadrados mínimos tem-se*

$$EQM(\hat{\beta}) = \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \quad (9)$$

$$\text{Se } \mathbf{Y} \sim N_n(X\boldsymbol{\beta}, \sigma^2 I) : \text{var}(L_1^2) = 2\sigma^4 \sum_1^p \frac{1}{\lambda_i^2} \quad (10)$$

Considerando  $L_1^2 = (\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})'(\hat{\boldsymbol{\beta}} - \hat{\boldsymbol{\beta}})$

*Demonstração.*

$$\begin{aligned}
 EQM(\hat{\beta}) &= E((\hat{\beta} - \beta)'(\hat{\beta} - \beta)) \\
 &= E(\text{tr}((\hat{\beta} - \beta)'(\hat{\beta} - \beta))) \\
 &= E(\text{tr}(\hat{\beta} - \beta)(\hat{\beta} - \beta)') \\
 &= \text{tr} E((\hat{\beta} - \beta)(\hat{\beta} - \beta)') \\
 &= \text{tr} \text{cov}(\hat{\beta}) \\
 &= \text{tr} (\sigma^2 (X'X)^{-1}) \\
 &= \sigma^2 \text{tr} (X'X)^{-1} \\
 &= \sigma^2 \text{tr} (PP'X'X)^{-1} \\
 &= \sigma^2 \text{tr} (P'(X'X)^{-1}P) \\
 &= \sigma^2 \text{tr} \begin{pmatrix} \frac{1}{\lambda_1} & & \\ & \ddots & \\ & & \frac{1}{\lambda_p} \end{pmatrix} \\
 &= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i}
 \end{aligned}$$

em que  $P_{p \times p}$  é uma matriz ortogonal que diagonaliza  $X'X$ .

$$\begin{aligned}
\text{var}(L_1^2) &= \text{var}(((X'X)^{-1}X'\mathbf{Y} - \boldsymbol{\beta})'((X'X)^{-1}X'\mathbf{Y} - \boldsymbol{\beta})) \\
&= \text{var}(((X'X)^{-1}X'\mathbf{Y} - (X'X)^{-1}(X'X)\boldsymbol{\beta})'((X'X)^{-1}X'\mathbf{Y}) \\
&\quad - (X'X)^{-1}(X'X)\boldsymbol{\beta}) \\
&= \text{var}([X'\mathbf{Y} - (X'X)\boldsymbol{\beta}]'(X'X)^{-1}[(X'X)^{-1}(X'\mathbf{Y} - (X'X)\boldsymbol{\beta})]) \\
&= \text{var}([X'\mathbf{Y} - (X'X)\boldsymbol{\beta}](X'X)^{-2}[(X'\mathbf{Y} - (X'X)\boldsymbol{\beta})]) \\
&= \text{var}([X'(\mathbf{Y} - X\boldsymbol{\beta})](X'X)^{-2}[(X'(\mathbf{Y} - X\boldsymbol{\beta}))]) \\
&= \text{var}((\mathbf{Y} - X\boldsymbol{\beta})X(X'X)^{-2}X'(\mathbf{Y} - X\boldsymbol{\beta}))
\end{aligned}$$

Substituindo  $\mathbf{Y}$  por  $\mathbf{Y} - X\boldsymbol{\beta}$ ,  $A = X(X'X)^{-2}X'$  e  $\Sigma = \sigma^2 I$  na proposição (6) e observando que  $E(\mathbf{Y} - X\hat{\boldsymbol{\beta}}) = 0$

$$\begin{aligned}
\text{var}(L_1^2) &= 2\text{tr}(\sigma^2(X(X'X)^{-2})X')^2 \\
&= 2\sigma^4\text{tr}[X(X'X)^{-2}X'X(X'X)^{-2}X'] \\
&= 2\sigma^4\text{tr}[X(X'X)^{-2}X'X^{-1}X'] \\
&= 2\sigma^4\text{tr}(X(X'X)^{-3}X') \\
&= 2\sigma^4\text{tr}[(X'X)^{-3}X'X] \\
&= 2\sigma^4\text{tr}[(X'X)^{-2}] \\
&= 2\sigma^4 \sum_{i=1}^p \frac{1}{\lambda_i^2}
\end{aligned}$$

□

Portanto, observa-se que os autovalores da matriz  $X'X$  são extremamente relevantes, pois os autovalores afetam o erro quadrático médio das estimativas. Portanto, autovalores pequenos implicam em variabilidade alta para o estimador



de quadrados mínimos, o que pode implicar que tais estimadores não sejam adequados.

## 2.6 Regressão de cumeeira - uma estimação viesada

Em problemas de regressão linear com um número grande de covariáveis é comum que algumas dessas variáveis, ou combinações lineares entre elas, sejam altamente relacionadas. Tal fato implicará que alguns dos vetores colunas da matriz  $X$  ou combinações lineares destas colunas, estão próximos sob um ponto de vista geométrico. Tal fato implica que o determinante da matriz quadrada  $X'X$  é próximo de zero. Como o determinante é produto dos autovalores pode-se afirmar que pelo menos um deles é próximo de zero. Neste caso, como já observado anteriormente a variabilidade total do estimador de quadrados mínimos é alta e pode inviabilizar o seu uso. Como o estimador de quadrados mínimos é o melhor entre os não viesados uma possibilidade para se contornar este problema é a utilização de estimadores viesados.

Uma estratégia para obter estimadores viesados é utilizar a ideia de encolhimento. Isto é, obter estimadores pelo método dos quadrados mínimos que diminuam o tamanho das estimativas.

Para que o estimador obtido por encolhimento possua propriedades interessantes é razoável que ele seja obtido a partir de um problema de minimização. Esta é a filosofia para se obter os denominados estimadores de cumeeira (*ridge*) (HOERL; KENNARD, 1970a).

A ideia para a definição dos estimadores de cumeeira e outros estimadores de encolhimento é utilizar elipsoides centrados na estimativa de quadrados mínimos  $\hat{\beta}$ , descrito na Figura 12:

Seja  $B$  outro estimador qualquer do vetor de parâmetros  $\beta$ . A soma de

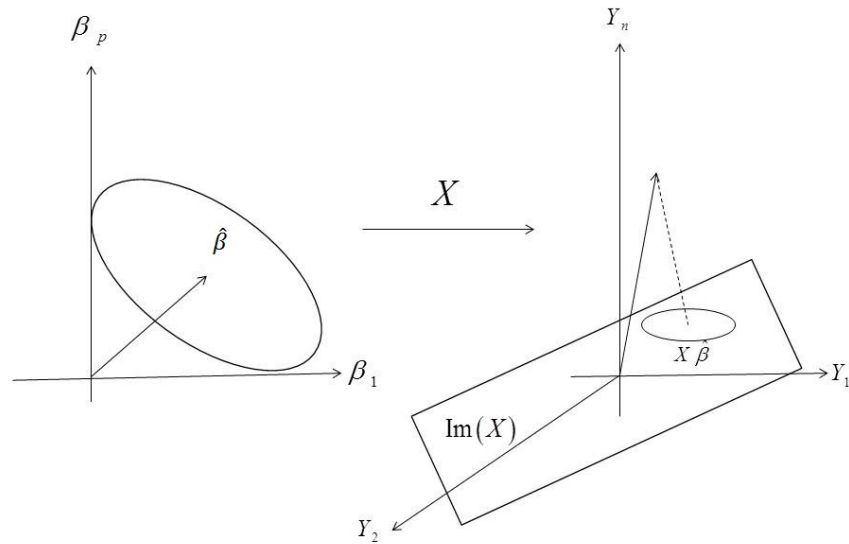


Figura 12 Estimativas com a mesma soma de quadrados dos erros

quadrados dos resíduos deste estimador é dada por:

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

Somando e subtraindo  $X\hat{\beta}$  tem-se:

$$\begin{aligned}
 \phi &= (\mathbf{Y} - XB)'(\mathbf{Y} - XB) \\
 &= (\mathbf{Y} - X\hat{\beta} + X\hat{\beta} - XB)'(\mathbf{Y} - X\hat{\beta} + X\hat{\beta} - XB) \\
 &= (\mathbf{Y} - X\hat{\beta})'(\mathbf{Y} - X\hat{\beta}) + 2(X\hat{\beta} - XB)'(\mathbf{Y} - X\hat{\beta}) \\
 &\quad + (X\hat{\beta} - XB)'(X\hat{\beta} - XB) \\
 &= (\mathbf{Y} - X\hat{\beta})'(\mathbf{Y} - X\hat{\beta}) + (X\hat{\beta} - XB)'(X\hat{\beta} - XB) \\
 &= (\mathbf{Y} - X\hat{\beta})'(\mathbf{Y} - X\hat{\beta}) + (\hat{\beta} - B)'X'X(\hat{\beta} - B) \\
 &= \|\mathbf{Y} - X\hat{\beta}\|^2 + (\hat{\beta} - B)'X'X(\hat{\beta} - B)
 \end{aligned}$$

Tem-se que  $(X\hat{\beta} - XB)'(\mathbf{Y} - X\hat{\beta}) = 0$  pois  $(X\hat{\beta} - XB) = X(\hat{\beta} - B)$  é um vetor em  $Im(X)$  e  $(\mathbf{Y} - X\hat{\beta})$  é perpendicular a  $Im(X)$ . Este fato é geometricamente descrito pela Figura 13:

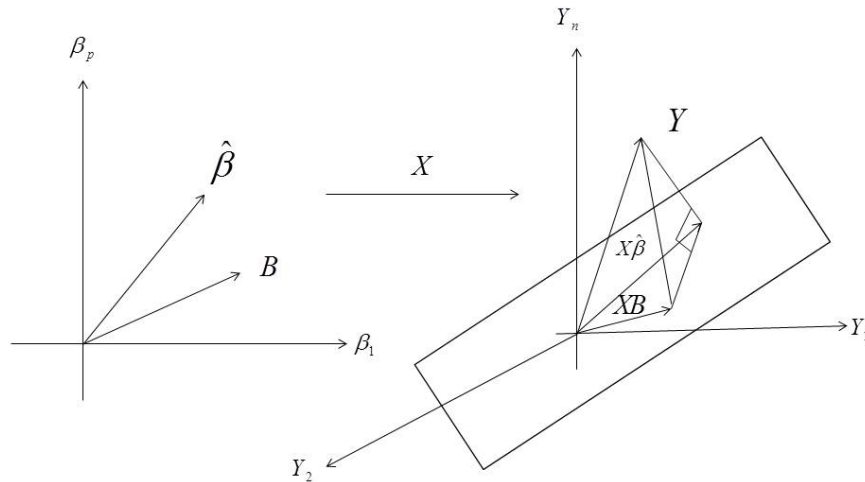


Figura 13 Triângulo retângulo fundamental

Portanto, a soma de quadrados dos resíduos do estimador  $B$ , isto é, o quadrado da distância do vetor de dados  $Y$  e sua estimativa  $XB$  pode ser decomposto em duas partes: o quadrado da distância de  $Y$  à sua estimativa de quadrados mínimos,  $\|Y - X\hat{\beta}\|^2$ , que será denominada  $\phi_{min}$  e o quadrado da distância de  $XB$  a  $X\hat{\beta}$ , isto é,  $\|X\hat{\beta} - XB\|^2 = (X\hat{\beta} - XB)'(X\hat{\beta} - XB) = (\hat{\beta} - B)'X'X(\hat{\beta} - B)$ . Observe que  $\phi_{min}$  não depende do estimador  $B$  escolhido, pois  $\phi = \phi_{min} + \|X\hat{\beta} - XB\|^2$ .

Como o que se quer é obter o estimador  $B$ , que tenha um erro quadrático menor que o erro quadrático do estimador de quadrados mínimos, suponha  $Y$  fixo e também  $\phi = cte$ . Isto é, estamos supondo que a soma de quadrados dos resíduos é constante.

Com  $Y$  fixo, fica também fixo  $X\hat{\beta}$ . Como  $\|Y - XB\|^2 = \|Y - X\hat{\beta}\|^2 +$

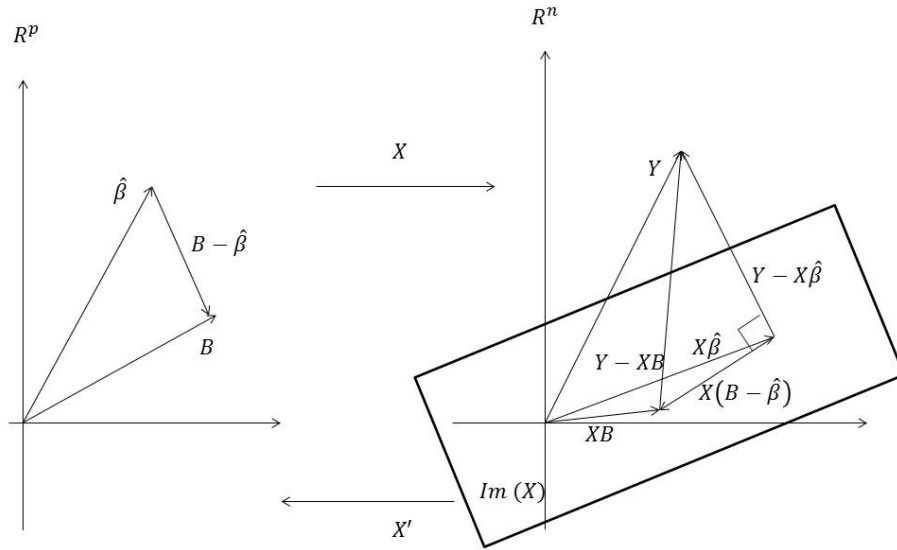


Figura 14 Relação de um estimador com o estimador de quadrados mínimos

$\|X\hat{\beta} - XB\|^2$  e  $\|Y - XB\|^2$  está fixado,  $\|X\hat{\beta} - XB\|^2$  também fica fixo em um valor que será denominado  $\phi_0$ . Desta forma o estimador  $B$  pode assumir valores no espaço paramétrico tais que  $XB$  está em uma esfera em  $Im(X)$  centrada em  $X\hat{\beta}$  e raio  $\sqrt{\phi_0}$ .

$$\begin{aligned}
 \phi_0 &= \|X\hat{\beta} - XB\|^2 \\
 &= (X\hat{\beta} - XB)'(X\hat{\beta} - XB) \\
 &= (X(\hat{\beta} - B))'X(\hat{\beta} - B) \\
 &= (\hat{\beta} - B)'X'X(\hat{\beta} - B)
 \end{aligned}$$

Da teoria das formas quadráticas e elipsoides temos que como o vetor  $XB$  está contido em uma esfera centrada em  $X\hat{\beta}$  e raio  $\sqrt{\phi_0}$ , os vetores  $B$  estão contidos no elipsoide  $(\hat{\beta} - B)'X'X(\hat{\beta} - B) = \phi_0$  no espaço de parâmetros  $R^p$  (Figura 16).

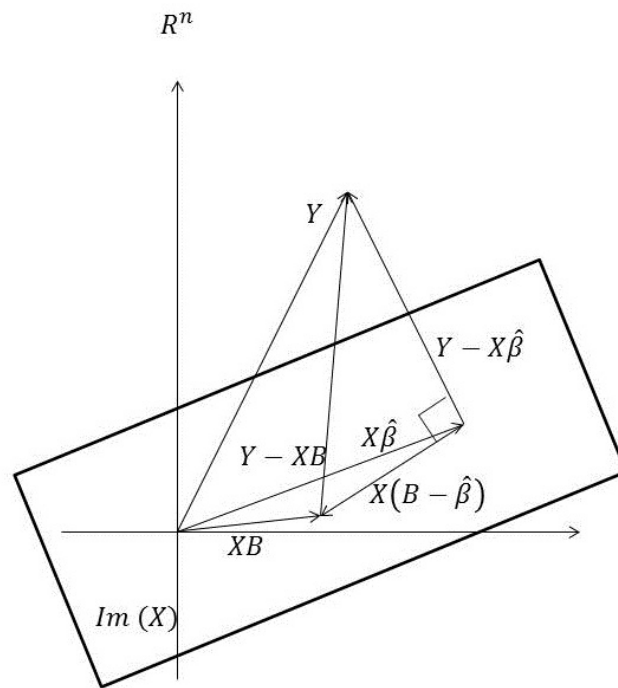


Figura 15 Ajuste com a mesma soma de quadrados de resíduos

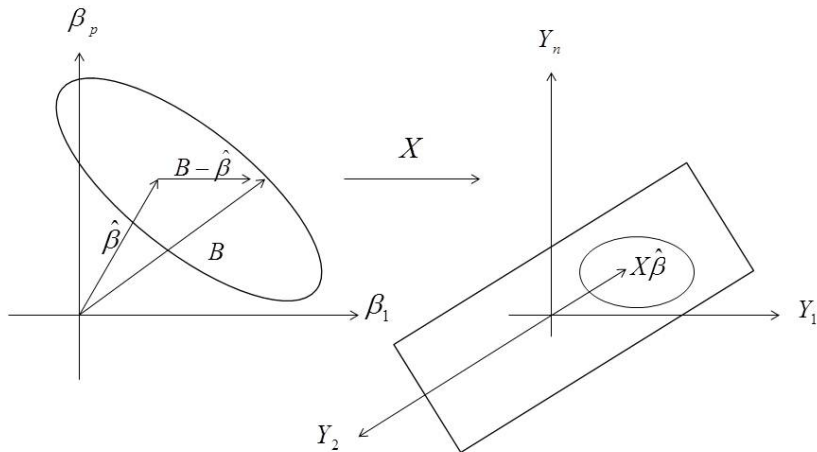


Figura 16 Estimativas com a mesma soma de quadrados de resíduos

Pode-se agora formular o problema variacional. Como o vetor de dados  $\mathbf{Y}$  está fixo, queremos os vetores das estimativas de  $B$  sobre uma elipse tal que o tamanho de  $B$  seja mínimo. Temos então um problema de minimização com restrição, isto é, se quer minimizar  $B'B$  sujeito à restrição

$$(\hat{\beta} - B)'(X'X)(\hat{\beta} - B) = \phi_0.$$

A lagrangiana é dada por:

$$H = B'B + \lambda[(\hat{\beta} - B)'(X'X)(\hat{\beta} - B) - \phi_0].$$

Logo:

$$\frac{\partial H}{\partial B} = 2B + 2\lambda[(X'X)B - (X'X)\hat{\beta}] = 0.$$

$$\begin{aligned}
B + \lambda(X'X)B &= \lambda(X'X)\hat{\beta} \\
(I + \lambda(X'X))B &= \lambda(X'X)\hat{\beta} \\
B &= \lambda(I + \lambda(X'X))^{-1}(X'X)\hat{\beta} \\
&= \left(\frac{1}{\lambda}I + (X'X)\right)^{-1} (X'X)\hat{\beta} \\
&= \left(\frac{1}{\lambda}I + (X'X)\right)^{-1} (X'X)\hat{\beta} \\
&= \left(\frac{1}{\lambda}I + (X'X)\right)^{-1} (X'X)(X'X^{-1}X'Y) \\
&= \left(\frac{1}{\lambda}I + (X'X)\right)^{-1} X'Y
\end{aligned}$$

fazendo  $k = \frac{1}{\lambda}$  tem-se finalmente  $B = [(X'X) + kI]^{-1}X'Y$ . Este estimador será denotado por  $\hat{\beta}(k)$  e é denominado o estimador de cuneeira com parâmetro  $k$  ou estimador de cuneeira de Hoerl & Kennard. A soma do número  $k$  na diagonal de  $X'X$  é intuitivamente um encolhimento, uma vez que o valor de  $k$  muito grande fará a matriz  $(X'X + kI)^{-1}$  ficar próxima à matriz nula. A curva  $\hat{\beta}(k)$  em  $R^p$  obtida quando se varia  $k$ , (observe que o vetor de dados  $Y$  está fixo) é denominada de traço de cuneeira (ridge trace).  $\hat{\beta}(0) = \hat{\beta}$  e se está interessado no comportamento de  $\hat{\beta}(k)$  para  $k \geq 0$ .

Os estimadores de cuneeira admitem também uma interessante interpretação em termos de verossimilhança para o caso  $Y \sim N_n(X\beta, \sigma^2I)$ . A função de verossimilhança é dada por:

$$L(\beta) = \frac{1}{v/2\pi\sigma^2} \exp\left(-\frac{1}{2\sigma^2}(\mathbf{Y} - X\beta)'(\mathbf{Y} - X\beta)\right)$$

Ou seja



$$L(\beta) = \frac{1}{\sqrt{2\pi\sigma^2}^n} \exp\left(-\frac{1}{2\sigma^2} \left[ (\mathbf{Y} - X\hat{\beta})'(\mathbf{Y} - X\hat{\beta}) + (\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) \right]\right)$$

Segue que sobre o elipsoide  $(\beta - \hat{\beta})'X'X(\beta - \hat{\beta}) = \phi_0$ , a verossimilhança é constante, uma vez que  $(\mathbf{Y} - X\hat{\beta})$  só depende dos dados. Uma vez fixado o valor  $\phi_0$ , isto equivale a se fixar um valor de  $k$ . O estimador de cumeira  $\hat{\beta}(k)$  é obtido então por um procedimento conservador. Entre todos os vetores  $\beta$  com a mesma verossimilhança é escolhido aquele de menor comprimento como descrito na Figura 17

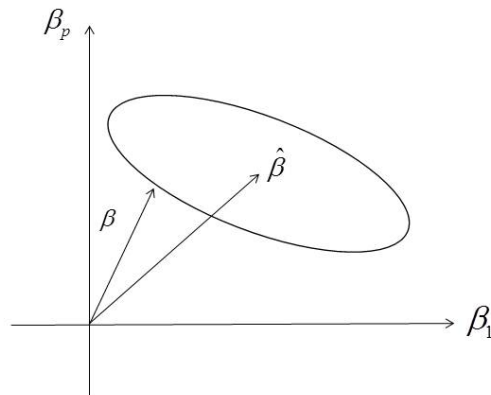


Figura 17 Elipsoide centrado no estimador de quadrados mínimos

## 2.7 Uma interpretação Bayesiana dos estimadores de cumeieira

O estimador de cumeieira pode ser obtido através de uma abordagem Bayesiana. Considere a regressão linear:

$$\mathbf{Y}_{n \times 1} = X_{n \times p} \boldsymbol{\beta}_{p \times 1} + \boldsymbol{\varepsilon}_{n \times 1}$$

$$\boldsymbol{\varepsilon}_{n \times 1} \sim N_n(0, \sigma^2 I_{n \times n}).$$

Considere  $\sigma^2$  conhecido e uma distribuição a priori para o vetor de parâmetros  $\boldsymbol{\beta}_{p \times 1}$  dada por uma normal  $p$ -variada:

$$\boldsymbol{\beta}_{p \times 1} | \sigma_\lambda^2 \sim N_p\left(\mathbf{0}, \frac{\sigma^2}{\lambda} I\right)$$

A distribuição a posteriori de  $\boldsymbol{\beta}$  dado  $\mathbf{Y}$  e  $X$  é:

$$\begin{aligned} f_{\boldsymbol{\beta}}(\boldsymbol{\beta}, \sigma^2 | \mathbf{Y}, X, \lambda) &\propto f_{\mathbf{Y}}(\mathbf{Y} | X, \boldsymbol{\beta}, \sigma^2) f_{\boldsymbol{\beta}}(\boldsymbol{\beta} | \sigma^2, \lambda) \\ &\propto \sigma^{-n} e^{\left[-\frac{1}{2\sigma^2} (\mathbf{Y}-X\boldsymbol{\beta})' (\mathbf{Y}-X\boldsymbol{\beta})\right]} \sigma^{-p} e^{\left[-\frac{\lambda}{2\sigma^2} \boldsymbol{\beta}' \boldsymbol{\beta}\right]} \\ &\propto \sigma^{-(n+p)} e^{\left[-\frac{1}{2\sigma^2} [(\mathbf{Y}-X\boldsymbol{\beta})' (\mathbf{Y}-X\boldsymbol{\beta}) - \lambda \boldsymbol{\beta}' \boldsymbol{\beta}]\right]} \\ &\propto \sigma^{-(n+p)} e^{\left[-\frac{1}{2\sigma^2} [\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'X\boldsymbol{\beta} - \boldsymbol{\beta}'X'\mathbf{Y} + \boldsymbol{\beta}'X'X\boldsymbol{\beta} + \lambda \boldsymbol{\beta}' \boldsymbol{\beta}]\right]} \\ &\propto \sigma^{-(n+p)} e^{\left[-\frac{1}{2\sigma^2} [\mathbf{Y}'\mathbf{Y} - \mathbf{Y}'X\boldsymbol{\beta} - \boldsymbol{\beta}'X'\mathbf{Y} + \boldsymbol{\beta}'[X'X\boldsymbol{\beta} + \lambda I]\boldsymbol{\beta}]\right]} \end{aligned}$$

Para explicitar a distribuição posteriori de  $\boldsymbol{\beta}$  é necessário completar os quadrados na expressão:

$$\begin{aligned}
(Y - X\beta)'(Y - X\beta) + \lambda\beta'\beta &= Y'Y - Y'X\beta - \beta'X'Y + \beta'X'X\beta + \lambda\beta'\beta \\
&= Y'Y - 2Y'X\beta + \beta'(X'X + \lambda I)\beta
\end{aligned}$$

ou seja, queremos determinar  $W$  tal que:

$$\begin{aligned}
(\beta - W)'(X'X + \lambda I)(\beta - W) &= Y'Y - 2Y'X\beta + \beta'(X'X + \lambda I)\beta \\
&= \beta'(X'X + \lambda I)\beta - 2W'(X'X + \lambda I)\beta \\
&\quad + W'(X'X + \lambda I)W \tag{11}
\end{aligned}$$

Então:  $W = (X'X + \lambda I)^{-1}X'Y$ . Observe que :  $Y'X\beta = W'(X'X + \lambda I)\beta$ .

$$\begin{aligned}
(Y - X\beta)'(Y - X\beta) + \lambda\beta'\beta &= Y'Y + \beta'(X'X + \lambda I)\beta - 2W'(X'X + \lambda I)\beta \\
&= Y'Y + \beta'(X'X + \lambda I)\beta - 2W'(X'X + \lambda I)\beta \\
&\quad + W'(X'X + \lambda I)W - W'(X'X + \lambda I)W \\
&= (\beta - W)'(X'X + \lambda I)(\beta - W) + Y'Y \\
&\quad - W'(X'X + \lambda I)W \tag{12}
\end{aligned}$$

logo,  $f_{\beta}(\beta, \sigma^2 | \mathbf{Y}, X) \propto \exp((\beta - W)'(X'X + \lambda I)(\beta - W) + Y'Y - W'(X'X + \lambda I)W) \propto ((\beta - W)'(X'X + \lambda I)(\beta - W))$ .

Portanto, a posteriori é uma normal com média  $W = (X'X + \lambda I)^{-1}X'Y$ .

Segue então, que o estimador de Bayes do vetor  $\beta$  é justamente o estimador de cumeieira  $\hat{\beta}(\lambda) = (X'X + \lambda I)^{-1}X'Y$ .

## 2.8 Exemplos de estimadores de cumeieira

O exemplo mais simples para a aplicação do estimador de cumeieira é quando a matriz de delineamento  $X$  possui suas colunas ortonormais, isto é, as colunas possuem norma unitária e são ortogonais entre si. Neste caso tem-se  $X'X = I$  e o estimador de cumeieira fica da forma:

$$\begin{aligned}\hat{\beta}(k) &= (I + kX'X)^{-1}X'\mathbf{Y} \\ &= (I + kI)^{-1}X'\mathbf{Y} \\ &= \left(\frac{1}{1+k}\right)X'\mathbf{Y}\end{aligned}$$

A função traço de cumeieira são retas que tendem para a origem quando  $k$  cresce (Figura 18).

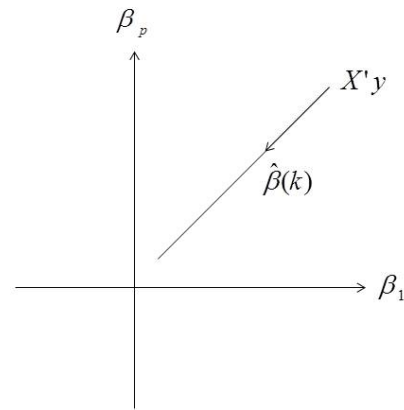


Figura 18 Traço de cumeeira - caso ortogonal

Um caso mais geral é quando  $X'X$  é uma matriz diagonal, ou seja,

$$X'X = \Lambda = \begin{pmatrix} \lambda_1 & 0 & \cdots & 0 \\ 0 & \lambda_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \lambda_p \end{pmatrix}$$

Com isso tem-se que:

$$\begin{aligned}
\hat{\beta}(k) &= (X'X + kI)^{-1}X'\mathbf{Y} \\
&= (\Lambda + kI)^{-1}X'\mathbf{Y} \\
&= \begin{pmatrix} \frac{1}{\lambda_1 + k} & 0 & \cdots & 0 \\ 0 & \frac{1}{\lambda_2 + k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\lambda_p + k} \end{pmatrix} X'\mathbf{Y}
\end{aligned}$$

Neste caso como função de  $k$  as estimativas não seguem mais uma reta, mas tendem a zero quando  $k$  cresce (Figura 19).

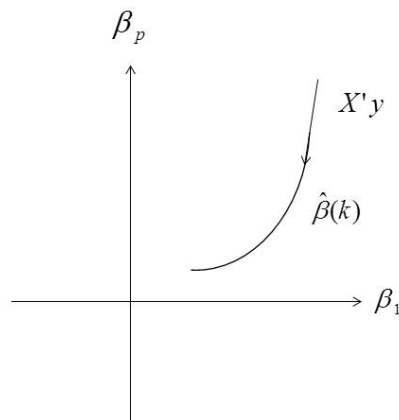


Figura 19 Traço de cumeeira

De fato esta curva é o comportamento genérico das estimativas de cumeira em função de  $k$  a menos de uma transformação linear ortogonal, como exposto a seguir.

Como  $X'X$  é uma matriz simétrica, existe uma matriz ortogonal  $P_{p \times p}$ , tal que  $P'(X'X)P = \Lambda$ , em que  $\Lambda$  é diagonal com entradas dadas pelos autovalores  $\lambda_i$ ,  $i = 1, 2, \dots, p$  de  $X'X$ . Utilizando  $P$  será definida uma mudança de parâmetros,  $\beta = P\alpha$ :

$$\mathbf{Y} = X\beta + \varepsilon = XP\alpha + \varepsilon = X^*\alpha + \varepsilon$$

com  $X^* = XP$ . A mudança nos parâmetros pode ser descrita geometricamente pela Figura 20:

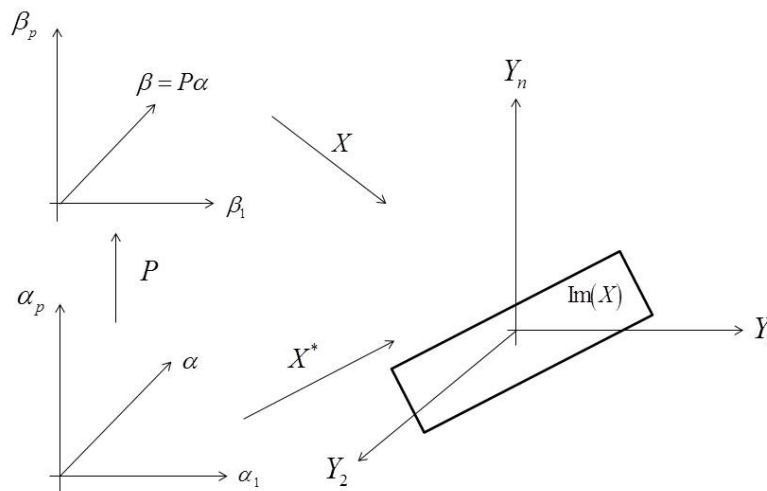


Figura 20 Mudança para parâmetros canônicos

Em relação aos novos parâmetros  $E(\mathbf{Y}) = XP\alpha$ . O estimador de qua-

drados mínimos é:

$$\begin{aligned}
 \hat{\boldsymbol{\alpha}} &= ((XP)'(XP))^{-1}(XP)'\mathbf{Y} \\
 &= (P'X'XP)^{-1}P'X'\mathbf{Y} \\
 &= P(X'X)^{-1}(P')^{-1}P'X'\mathbf{Y} \\
 &= P'(X'X)^{-1}X'\mathbf{Y} \\
 &= P'\hat{\boldsymbol{\beta}}
 \end{aligned}$$

$$\begin{aligned}
 cov(\hat{\boldsymbol{\alpha}}) &= cov(P'\hat{\boldsymbol{\beta}}) \\
 &= P'cov(\hat{\boldsymbol{\beta}})(P^{-1})' \\
 &= P'\sigma^2(X'X)^{-1}(P^{-1})' \\
 &= \sigma^2P^{-1}(X'X)^{-1}(P^{-1})' \\
 &= \sigma^2(P'X'XP)^{-1} \\
 &= \sigma^2\Lambda^{-1}
 \end{aligned}$$



O estimador de cumeieira é:

$$\begin{aligned}
 \hat{\alpha}(k) &= (\Lambda + kI)^{-1}P'X'\mathbf{Y} \\
 &= (PX'XP' + kI)^{-1}P'X'\mathbf{Y} \\
 &= (PX'XP' + kPP')^{-1}P'X'\mathbf{Y} \\
 &= (P(X'X + kI)P')^{-1}P'X'\mathbf{Y} \\
 &= P'(X'X + kI)^{-1}PP'X'\mathbf{Y} \\
 &= P'(X'X + kI)^{-1}X'\mathbf{Y} \\
 &= P'\hat{\beta}(k)
 \end{aligned}$$

E, portanto, as curvas das estimativas  $\hat{\beta}(k)$  e  $\hat{\alpha}(k)$  estão relacionadas por uma transformação ortogonal e são, portanto essencialmente iguais. Os parâmetros  $\alpha$  são denominados parâmetros canônicos.

O caso em que  $X'X = \Lambda$  é uma matriz diagonal sugere uma definição um pouco mais geral do estimador de cumeieira. No lugar de se somar  $kI$  à matriz  $X'X$ , soma-se uma matriz diagonal da forma:

$$K = \begin{pmatrix} k_1 & 0 & \cdots & 0 \\ 0 & k_2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \cdots & k_p \end{pmatrix}$$

Obtendo-se o estimador  $\hat{\beta}(k_1, k_2, \dots, k_p) = (X'X + K)^{-1}X'Y$ , denominado estimador de cumeieira generalizado. Para o caso  $X'X = \Lambda$ , o estimador fica

da forma:

$$\begin{aligned}
 \hat{\beta}(k_1, k_2, \dots, k_p) &= (X'X + K)^{-1} X'Y \\
 &= (\Lambda + K)^{-1} X'Y \\
 &= \begin{pmatrix} \frac{1}{\lambda_1 + k_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\lambda_2 + k_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{1}{\lambda_p + k_p} \end{pmatrix} X'Y
 \end{aligned}$$

Um inconveniente desta definição mais geral é que não se tem mais as curvas de cumeeira.

## 2.9 Propriedades dos estimadores de cumeeira

**Proposição 10.** *O estimador de cumeeira está relacionado de forma linear com o estimador de quadrados mínimos.*

*Demonstração.*

$$\begin{aligned}
 \hat{\beta}(k) &= (X'X + kI)^{-1} X'Y \\
 &= ((X'X)(I + k(X'X)^{-1})^{-1})^{-1} X'Y \\
 &= (I + k(X'X)^{-1})^{-1} (X'X)^{-1} X'Y \\
 &= (I + k(X'X)^{-1})^{-1} \hat{\beta}
 \end{aligned} \tag{13}$$

□

Portanto, se  $Z(k) = (I + k(X'X)^{-1})^{-1}$  então  $\hat{\beta}(k) = Z(k)\hat{\beta}$ . Considere  $W(k) = (X'X + kI)^{-1}$  e  $Z(k) = (I + k(X'X)^{-1})^{-1}$ .

Afirmação: os autovalores de  $W(k)$  são  $\frac{1}{\lambda_i + k}$  e os de  $Z(k)$  são  $\frac{\lambda_i}{\lambda_i + k}$ .

Os autovalores da matriz  $X'X$  satisfazem  $X'X\mathbf{v}_i = \lambda_i\mathbf{v}_i$ .

$$\begin{aligned} W^{-1}(k) &= (X'X + kI)\mathbf{v}_i \\ &= X'X\mathbf{v}_i + k\mathbf{v}_i \\ &= \lambda_i\mathbf{v}_i + k\mathbf{v}_i \\ &= (\lambda_i + k)\mathbf{v}_i \end{aligned}$$

E, portanto os autovalores de  $W(k)$  são  $\frac{1}{\lambda_i + k}$ .

Analogamente, se  $Z(k) = (I + k(X'X)^{-1})^{-1}$ , tem-se para  $Z^{-1}(k)$  os seguintes autovalores:

$$\begin{aligned} Z^{-1}(k)\mathbf{v}_i &= (I + k(X'X)^{-1})\mathbf{v}_i \\ &= \mathbf{v}_i + k(X'X)^{-1}\mathbf{v}_i \\ &= \mathbf{v}_i + k\left(\frac{1}{\lambda_i}\mathbf{v}_i\right) \\ &= \mathbf{v}_i + \frac{k}{\lambda_i}\mathbf{v}_i \\ &= \left(1 + \frac{k}{\lambda_i}\right)\mathbf{v}_i \\ &= \left(\frac{\lambda_i + k}{\lambda_i}\right)\mathbf{v}_i \end{aligned}$$

e, portanto os autovalores da matriz  $Z(k)$  são  $\frac{\lambda_i}{\lambda_i + k}$ . Ainda é possível observar

uma relação entre as matrizes  $W(k)$  e  $Z(k)$ . Segue:

$$\begin{aligned}
 W(k) &= (X'X + kI)^{-1} \\
 &= (X'X + k(X'X)^{-1}(X'X))^{-1} \\
 &= (I + k(X'X)^{-1})^{-1}(X'X)^{-1} \\
 &= Z(k)(X'X)^{-1}
 \end{aligned}$$

E, portanto  $W(k)(X'X) = Z(k)$ . Note que

$$\begin{aligned}
 Z(k) + kW(k) &= W(k)(X'X) + kW(k) \\
 &= W(k)(X'X + kI) \\
 &= I
 \end{aligned}$$

E, portanto,  $(I - kW(k)) = Z(k)$

A mesma matriz ortogonal  $P$  que diagonaliza  $X'X$  também diagonaliza  $Z(k)$  e  $W(k)$ .

$$\begin{aligned}
 P'Z(k)P &= P'(I + k(X'X)^{-1})^{-1}P \\
 &= (P'(I + k(X'X)^{-1})P)^{-1} \\
 &= (I + kP'(X'X)^{-1}P)^{-1} \\
 &= (I + k(P'(X'X)^{-1}P)^{-1})^{-1} \\
 &= (I + k\Lambda^{-1})^{-1} \\
 &= D\left(\frac{\lambda_1}{\lambda_1 + k}, \dots, \frac{\lambda_p}{\lambda_p + k}\right)
 \end{aligned}$$

A matriz  $D \left( \frac{\lambda_1}{\lambda_1 + k}, \dots, \frac{\lambda_p}{\lambda_p + k} \right)$  será denotada por  $D(k)$ .

**Proposição 11.** *O estimador de cumeeira é um estimador de encolhimento, isto é,  $\|\hat{\beta}(k)\| \leq \|\hat{\beta}\|$ .*

*Demonstração.* Seja  $P$  tal que utilizando as relações:

$$P'Z(k)P = D(k)$$

$$Z(k) = PD(k)P'$$

$$\hat{\beta}(k) = Z(k)\hat{\beta}$$

$$\begin{aligned} (\hat{\beta}(k))'(\hat{\beta}(k)) &= \hat{\beta}'Z'(k)Z(k)\hat{\beta} \\ &= \hat{\beta}'P'D(k)PP'D(k)P\hat{\beta} \\ &= \hat{\beta}'P'(D(k))^2P\hat{\beta} \\ &= (P\hat{\beta})'(D(k))^2(P\hat{\beta}) \end{aligned}$$

Fazendo  $\xi_i = \frac{\lambda_i}{\lambda_i + k}$  tem-se para  $k > 0$ :

$$\begin{aligned}
\|\hat{\beta}(k)\|^2 &= \begin{pmatrix} v_1 & v_2 & \cdots & v_p \end{pmatrix} \begin{pmatrix} \xi_1^2 & 0 & \cdots & 0 \\ 0 & \xi_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \xi_p^2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_p \end{pmatrix} \\
&= \begin{pmatrix} v_1 \xi_1^2 & v_2 \xi_2^2 & \cdots & v_p \xi_p^2 \end{pmatrix} \begin{pmatrix} v_1 \\ v_2 \\ \vdots \\ v_p \end{pmatrix} \\
&= v_1 \xi_1^2 v_1 + v_2 \xi_2^2 v_2 + \cdots + v_p \xi_p^2 v_p \\
&= v_1^2 \xi_1^2 + v_2^2 \xi_2^2 + \cdots + v_p^2 \xi_p^2 \\
&\leq v_1^2 \xi_{max}^2 + v_2^2 \xi_{max}^2 + \cdots + v_p^2 \xi_{max}^2 \\
&= \xi_{max}^2 (v_1^2 + \cdots + v_p^2) \\
&= \xi_{max}^2 \|P\hat{\beta}\|^2 \\
&= \xi_{max}^2 \|\hat{\beta}\|^2 \\
&\leq \|\hat{\beta}\|^2
\end{aligned}$$

□

**Proposição 12.** A soma de quadrados dos resíduos de  $\hat{\beta}(k)$ :

$$\phi(k) = (\mathbf{Y} - X\hat{\beta}(k))'(\mathbf{Y} - X\hat{\beta}(k)) \quad (14)$$

pode ser reescrita da seguinte forma:

$$\phi(k) = \mathbf{Y}'\mathbf{Y} - (\hat{\boldsymbol{\beta}}(k))'X'\mathbf{Y} - k(\hat{\boldsymbol{\beta}}(k))'(\hat{\boldsymbol{\beta}}(k)). \quad (15)$$

*Demonstração.* Observe que:  $\hat{\boldsymbol{\beta}}(k) = (X'X + kI)^{-1}X'\mathbf{Y} \Rightarrow (X'X + kI)\hat{\boldsymbol{\beta}}(k) = X'\mathbf{Y} \Rightarrow X'\mathbf{Y} - X'X\hat{\boldsymbol{\beta}}(k) = (X'X + kI)\hat{\boldsymbol{\beta}}(k) - X'X\hat{\boldsymbol{\beta}}(k) = k\hat{\boldsymbol{\beta}}(k)$

$$\begin{aligned} \phi(k) &= (\mathbf{Y} - X\hat{\boldsymbol{\beta}}(k))'(\mathbf{Y} - X\hat{\boldsymbol{\beta}}(k)) \\ &= \mathbf{Y}'\mathbf{Y} - (\hat{\boldsymbol{\beta}}(k))'X'\mathbf{Y} - \mathbf{Y}'X\hat{\boldsymbol{\beta}}(k) + (\hat{\boldsymbol{\beta}}(k))'X'X(\hat{\boldsymbol{\beta}}(k)) \\ &= \mathbf{Y}'\mathbf{Y} - (\hat{\boldsymbol{\beta}}(k))'X'\mathbf{Y} - [\mathbf{Y}'X - (\hat{\boldsymbol{\beta}}(k))'X'X](\hat{\boldsymbol{\beta}}(k)) \\ &= \mathbf{Y}'\mathbf{Y} - (\hat{\boldsymbol{\beta}}(k))'X'\mathbf{Y} - k(\hat{\boldsymbol{\beta}}(k))'(\hat{\boldsymbol{\beta}}(k)) \end{aligned}$$

□

**Proposição 13.** *A matriz de variâncias e covariâncias de  $\hat{\boldsymbol{\beta}}(k)$  é dada por:*

$$\begin{aligned} \text{cov}(\hat{\boldsymbol{\beta}}(k)) &= \text{cov}(Z(k)\hat{\boldsymbol{\beta}}) \\ &= Z(k)\text{cov}(\hat{\boldsymbol{\beta}})Z(k)' \\ &= Z(k)\sigma^2(X'X)^{-1}Z(k)' \\ &= \sigma^2 Z(k)(X'X)^{-1}Z(k)' \end{aligned}$$

Portanto a soma das variâncias dos estimadores  $\hat{\beta}_i(k)$  é  $\text{tr}(\sigma^2 Z(k)(X'X)^{-1}Z(k)').$

O erro quadrático médio de  $\hat{\beta}(k)$  é:

$$\begin{aligned}
E[L_1^2(k)] &= E((\hat{\beta}(k) - \beta)'(\hat{\beta}(k) - \beta)) \\
&= E((Z(k)\hat{\beta} - \beta)'(Z(k)\hat{\beta} - \beta)) \\
&= E((Z(k)\hat{\beta} - Z(k)\beta + Z(k)\beta - \beta)'(Z(k)\hat{\beta} - Z(k)\beta + Z(k)\beta - \beta)) \\
&= E(((Z(k)\hat{\beta} - Z(k)\beta) + (Z(k)\beta - \beta))'((Z(k)\hat{\beta} - Z(k)\beta) + (Z(k)\beta - \beta))) \\
&= E(((Z(k)\hat{\beta} - Z(k)\beta)'(Z(k)\hat{\beta} - Z(k)\beta) + (Z(k)\beta - \beta)'(Z(k)\beta - \beta)) + \\
&\quad + \underbrace{((Z(k)\hat{\beta} - Z(k)\beta)'(Z(k)\beta - \beta))}_I + \underbrace{(Z(k)\beta - \beta)'(Z(k)\hat{\beta} - Z(k)\beta)}_{II})
\end{aligned}$$

$I$  e  $II$  é nulo, pois:

$$\begin{aligned}
E((Z(k)\hat{\beta} - Z(k)\beta)'(Z(k)\beta - \beta)) &= E((Z(k)\hat{\beta} - Z(k)\beta)')E((Z(k)\beta - \beta)) \\
&= E((Z(k)\hat{\beta} - Z(k)\beta)')E(Z(k)\beta - \beta) \\
&= E(Z(k)(\hat{\beta} - \beta)')E(Z(k)\beta - \beta) \\
&= Z(k)E(\hat{\beta} - \beta)(Z(k)\beta - \beta) \\
&= Z(k)(E[\hat{\beta}] - \beta)(Z(k)\beta - \beta) \\
&= Z(k)(\beta - \beta)(Z(k)\beta - \beta) \\
&= Z(k)(0)E(Z(k)\beta - \beta) \\
&= 0
\end{aligned}$$

A interpretação para a parte  $II$  é análoga. Retomando o cálculo de  $E[L_1^2(k)]$ , tem-se que:

$$\begin{aligned}
E((Z(k)\hat{\beta} - Z(k)\beta)'(Z(k)\hat{\beta} - Z(k)\beta) + (Z(k)\beta - \beta)'(Z(k)\beta - \beta)) &= \\
E((\hat{\beta} - \beta)'Z(k)'Z(k)(\hat{\beta} - \beta)) + \beta'(Z(k) - I)'(Z(k) - I)\beta
\end{aligned}$$



Observe que se tem aqui a seguinte forma quadrática:

$$Y'AY = (\hat{\beta} - \beta)'Z(k)'Z(k)(\hat{\beta} - \beta)$$

Portanto, como visto anteriormente na seção 2.2, pode-se pela proposição 5 concluir que:

$$\begin{aligned} E[(\hat{\beta} - \beta)'Z(k)'Z(k)(\hat{\beta} - \beta)] &= tr(\sigma^2(X'X)^{-1}Z(k)'Z(k)) \\ &= \sigma^2 tr(X'X)^{-1}(Z(k)'Z(k)) \end{aligned}$$

Observando que:

$$Z(k) = I - kW(k) \Rightarrow Z(k) - I = -kW(k)$$

$$\begin{aligned} \beta'(Z(k) - I)'(Z(k) - I)\beta &= \beta'[-k(X'X + kI)^{-1}][-k(X'X + kI)^{-1}]\beta \\ &= k^2\beta'(X'X + kI)^{-2}\beta \end{aligned}$$

e

$$\begin{aligned} Z(k)'Z(k) &= [I - k(X'X + kI)^{-1}]^2 \\ &= I^2 - 2kI(X'X + kI)^{-1} + k^2(X'X + kI)^{-2} \end{aligned}$$

Logo,

$$\begin{aligned} E[L_1^2(k)] &= \sigma^2 tr[(X'X)^{-1}[I - 2k(X'X + kI)^{-1} + k^2(X'X + kI)^{-2}] + \\ &+ k^2\beta'(X'X + kI)^{-2}\beta] \end{aligned}$$

$X'X = PDP'$ , em que  $D = diagonal(\lambda_1, \lambda_2, \dots, \lambda_p)$  e  $P$  é uma matriz

ortogonal. Sendo assim,  $(X'X)^m = PD^mP'$ ,  $P'X'XP = D$ ,  $P'P = PP' = I$ ,  
 $tr(P'X'XP) = tr(D) = \sum \lambda_i$

Observe que a matriz  $(I - 2k(X'X + kI))$  também é simétrica, e:

$$\begin{aligned}
 P'(I - 2k(X'X + kI)^{-1})P &= P'IP - 2kP'(X'X + kI)^{-1}P \\
 &= P'IP - 2kP'(X'X)P - 2k^2P'IP \\
 &= P'P - 2kD - 2k^2P'P \\
 &= I - 2kD - 2k^2I \\
 &= D_1
 \end{aligned}$$

Analogamente, temos:

$$\begin{aligned}
 P'k^{-2}(X'X + kI)^2P &= k^{-2}P'(X'X + kI)^2P \\
 &= k^{-2}P'((X'X)^2 + 2(X'X)kI + k^2I^2)P \\
 &= k^{-2}P'((X'X)^2 + 2k(X'X) + k^2I)P \\
 &= k^{-2}(P'(X'X)^2P + 2kP'(X'X)P + k^2P'IP) \\
 &= k^{-2}[D^2 + 2kD + k^2I] \\
 &= D_2
 \end{aligned}$$

o que demonstra que a matriz  $P$  diagonaliza as matrizes  $A = (I - 2k(X'X + kI))$

e  $B = k^{-2}(X'X + kI)^2$  e, portanto,

$$\begin{aligned}
 \text{tr}[(X'X)^{-1}(A + B)] &= \text{tr}[(X'X)^{-1} - (X'X)^{-1}2k(X'X + kI)^{-1} + \\
 &\quad + (X'X)^{-1}k^{-2}(X'X + kI)^2] \\
 &= \text{tr}[P'(X'X)^{-1}P] + \\
 &\quad + \text{tr}[P'(X'X)^{-1}PP'2k(X'X + kI)^{-1}P] + \\
 &\quad + \text{tr}[P'(X'X)^{-1}PP'k^{-2}(X'X + kI)^2P] \\
 &= \sum \frac{1}{\lambda_i} - 2k \sum \frac{1}{\lambda_i} \frac{1}{\lambda_i + k} + k^2 \sum \frac{1}{\lambda_i} \frac{1}{(\lambda_i + k)^2} \\
 &= \sum \frac{1}{\lambda_i} \left[ 1 - 2k \frac{1}{\lambda_i + k} + k^2 \frac{1}{\lambda_i} \frac{1}{(\lambda_i + k)^2} \right] \\
 &= \sum \frac{1}{\lambda_i} \left[ \frac{(\lambda_i + k)^2 - 2k(\lambda_i + k) + k^2}{(\lambda_i + k)^2} \right] \\
 &= \sum \frac{\lambda_i}{(\lambda_i + k)^2}
 \end{aligned}$$

Finalmente, tem-se que:

$$E[L_1^2(k)] = \underbrace{\sum \frac{\lambda_i}{(\lambda_i + k)^2}}_{\gamma_1(k)} + \underbrace{k^2 \beta'(X'X + kI)^{-2} \beta}_{\gamma_2(k)}.$$

Os significados dos dois termos da decomposição,  $\gamma_1(k)$  e  $\gamma_2(k)$  são: o primeiro termo,  $\gamma_1(k)$ , é a soma das variâncias dos estimadores dos parâmetros, o segundo elemento,  $\gamma_2(k)$ , é o quadrado da distância a partir  $Z(k)\beta$  para  $\beta$  que será zero quando  $k = 0$ , desde que  $Z = I$ . Assim,  $\gamma_2(k)$  é o quadrado do viés introduzido quando  $\hat{\beta}(k)$  é usado ao invés de  $\hat{\beta}$ .

O fato que a regressão de cumeieira reduz, para  $k$  suficientemente pequeno, o erro quadrático médio  $E[(\hat{\beta}(k) - \beta)'(\hat{\beta}(k) - \beta)]$  é uma propriedade mais geral. De fato, a regressão de cumeieira diminui o erro quadrático médio

generalizado. Considere em  $\mathbb{R}^n$  outro produto interno dado pelo produto interno  $\langle\langle \beta_1, \beta_2 \rangle\rangle = \beta_1' B \beta_2$ , em que  $B$  é uma matriz simétrica positiva definida. Considere a soma de quadrados generalizada  $\beta' B \beta$  com  $B$  apenas não negativa definida. Dado um estimador  $\hat{\beta}$  qualquer pode-se considerar o erro quadrático médio generalizado dado por  $E((\hat{\beta} - \beta)' B (\hat{\beta} - \beta))$ . O resultado de Theobald (1974) nos garante que no caso do estimador de cumeieira, para valores suficientemente pequenos de  $k$ , tem-se uma diminuição no erro quadrático médio generalizado em relação ao estimador de quadrados mínimos.

Sejam as matrizes de momentos de segunda ordem dos estimadores  $\hat{\beta}_1$  e  $\hat{\beta}_2$

$$M_1 = E((\hat{\beta}_1 - \beta)(\hat{\beta}_1 - \beta)')$$

$$M_2 = E[(\hat{\beta}_2 - \beta)(\hat{\beta}_2 - \beta)']$$

e erros quadráticos médios generalizados dados por:

$$m_1 = E((\hat{\beta}_1 - \beta)' B (\hat{\beta}_1 - \beta))$$

$$m_2 = E((\hat{\beta}_2 - \beta)' B (\hat{\beta}_2 - \beta))$$

**Proposição 14.** *Theobald (1974) São equivalentes:*

- i)  $M_1 - M_2$  é não negativa definida
- ii)  $m_1 - m_2 \geq 0$  para todas as matrizes  $B$  não negativas definidas

Considerando  $\tilde{\beta}_j$  um certo estimador, tem-se:

*Demonstração.* i) implica ii)

$$\begin{aligned} BM_j &= BE[(\tilde{\beta}_j - \beta)(\tilde{\beta}_j - \beta)'] \\ &= E[B(\tilde{\beta}_j - \beta)(\tilde{\beta}_j - \beta)'] \end{aligned}$$

$$\begin{aligned}
tr[E(BM_j)] &= tr(BE[(\tilde{\beta}_j - \beta)(\tilde{\beta}_j - \beta)']) \\
&= E[tr(B(\tilde{\beta}_j - \beta)(\tilde{\beta}_j - \beta)')] \\
&= E[(\tilde{\beta}_j - \beta)'B(\tilde{\beta}_j - \beta)] \\
&= m_j
\end{aligned}$$

Portanto  $m_1 - m_2 = tr(B(M_1 - M_2))$ .

Como  $M_1 - M_2$  é simétrica, sejam  $\mu_1, \dots, \mu_p$  seus autovalores e  $q_1, \dots, q_p$  seus autovetores. Tem-se que (FERREIRA, 2008):

$$M_1 - M_2 = \sum_{i=1}^p \mu_i q_i q_i'$$

Logo

$$\begin{aligned}
tr(B(M_1 - M_2)) &= tr(B(\sum_{i=1}^p \mu_i q_i q_i')) \\
&= tr((\sum_{i=1}^p \mu_i B q_i q_i')) \\
&= \sum_{i=1}^p \mu_i tr(B q_i q_i') \\
&= \sum_{i=1}^p \mu_i tr(q_i' B q_i) \\
&= \sum_{i=1}^p \mu_i q_i' B q_i \\
&\geq 0
\end{aligned}$$

pois  $\mu_i \geq 0$  e  $q_i' B q_i \geq 0$ , se  $B$  é não negativa definida.

ii) implica i) Se  $m_1 - m_2 \geq 0 \forall B$  não negativa definida fazendo  $B =$

$q_1q_1', B = q_2q_2', \dots, B = q_pq_p'$  tem-se que  $\mu_i \geq 0$  e portanto  $M_1 - M_2$  é não negativa definida.  $\square$

Para os estimadores de cumeieira, suas matrizes de momentos de segunda ordem satisfazem:

$$\begin{aligned}
M(k) &= E[\hat{\beta}(k)\hat{\beta}(k)'] \\
&= E[((X'X + kI)^{-1}X'Y)((X'X + kI)^{-1}X'Y)'] \\
&= E[(X'X + kI)^{-1}X'YY'X(X'X + kI)^{-1}] \\
&= (X'X + kI)^{-1}X'E[YY']X(X'X + kI)^{-1} \\
&= (X'X + kI)^{-1}X'[\sigma^2I + \beta'\beta]X(X'X + kI)^{-1} \\
&= \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1} \\
&\quad + ((X'X + kI)^{-1}X\beta)((X'X + kI)^{-1}X\beta)' \\
&= \sigma^2(X'X + kI)^{-1}X'X(X'X + kI)^{-1} \\
&\quad + (X'X + kI)^{-1}X'X\beta\beta'X'X(X'X + kI)^{-1} \\
&= \sigma^2(X'X + kI)^{-1}[X'X + X'X\beta\beta'X'X](X'X + kI)^{-1}
\end{aligned}$$

Logo  $M(0) = \sigma^2(X'X)^{-1} + \beta\beta'$

**Proposição 15.** *Existe  $k_0$  tal que para  $0 \leq k \leq k_0$  então  $M(0) - M(k)$  é positiva definida.*

*Demonstração.* (THEOBALD, 1974).

$\square$

Para os estimadores de cumeieira o comportamento do erro quadrático médio em função de  $k$ , desenvolvido em detalhes em Hoerl e Kennard (1970a) é descrito na Figura 21:

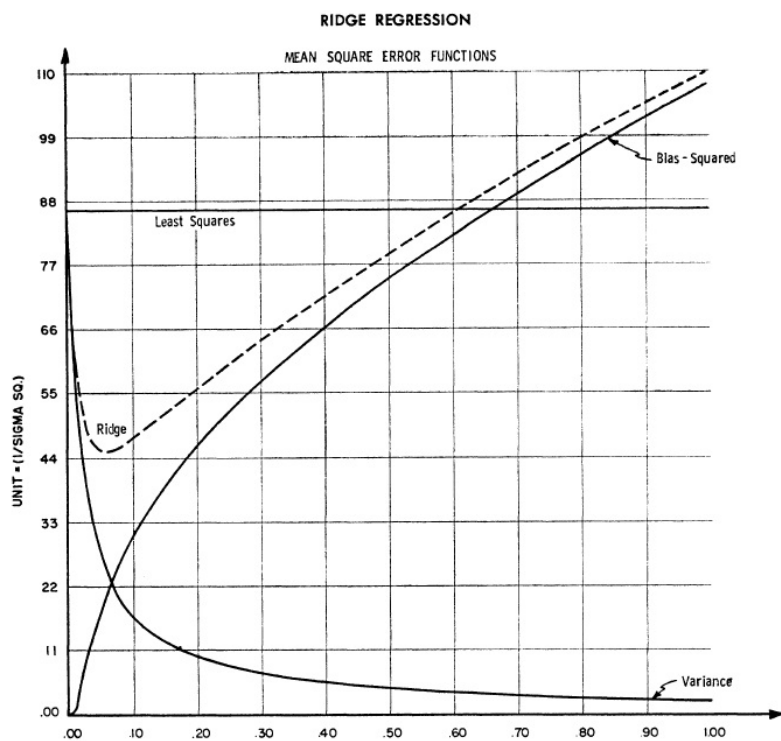


Figura 21 Comportamento do EQM. Fonte: Hoerl e Kennard (1970a)

Existe, portanto, um único valor para  $k$ , para o qual  $E[L_1^2]$  é mínimo, que será denominado  $k_{otimo}$ . Como a equação de  $E[L_1^2]$  depende dos parâmetros populacionais  $\beta$  e  $\sigma^2$ , o  $k_{otimo}$  solução da equação  $\frac{d}{dk} E[L_1^2] = 0$  é função de  $\beta$  e  $\sigma^2$ , sendo portanto ele mesmo um parâmetro populacional. A equação  $\frac{d}{dk} E[L_1^2] = 0$ , não admite solução explícita, pois envolve polinômio de ordem superior e portanto  $k_{otimo}$  não possui estimadores naturais. Vários estimadores para o  $k_{otimo}$  são propostos na literatura e inicialmente será visto o originalmente propostos por Hoerl e Kennard (1970a).

## 2.10 A procura pelo valor ótimo de $k$

Para se obter bons estimadores de  $k_{otimo}$  vários critérios de qualidade, além do erro quadrático médio, devem ser utilizados. Uma mudança de parâmetros para o problema de regressão múltipla

$$Y = X\beta + \varepsilon$$

que será adequada para a obtenção de estimadores é utilizar a transformação ortogonal  $P$  que diagonaliza  $X'X$ , isto é

$$P'(X'X)P = \Lambda = D(\lambda_1, \dots, \lambda_p).$$

Seja  $\alpha$  tal que  $\beta = P\alpha$ . Logo

$$Y = X\beta + \varepsilon = XP\alpha + \varepsilon = X^*\alpha + \varepsilon$$

$$(X^*X^*) = (XP)'XP = P'(X'X)P = \Lambda.$$

Neste caso diremos que o problema de regressão está na forma canônica. A vantagem de se usar coordenadas na forma canônica é que em relação aos parâmetros originais o erro quadrático não se altera. De fato se  $\hat{\beta}$  é um estimador de  $\beta$ , então considerando  $P\hat{\beta} = \hat{\alpha}$  como estimador de  $\alpha$  tem-se:

$$L_1^2 = \|\hat{\beta} - \beta\|^2 = \|P\hat{\alpha} - P\alpha\|^2 = \|P(\hat{\alpha} - \alpha)\|^2 = \|\hat{\alpha} - \alpha\|^2$$

, pois  $P$  é uma transformação ortogonal e portanto preserva distâncias.

Para o estimador de cumeira  $\hat{\beta}(k)$  tem-se:



$$\begin{aligned}
\hat{\alpha}(k) &= P'((X'X) + kI)^{-1}X'Y \\
&= P'((X'X) + kI)^{-1}PP'X'Y \\
&= (\Lambda + kI)^{-1}(XP)'Y \\
&= (\Lambda + kI)^{-1}X^*Y
\end{aligned}$$

e, portanto se  $\hat{\beta}(k)$  é o estimador de cumeieira para a regressão  $Y = X\beta + \varepsilon$ ,  $\hat{\alpha}(k)$  é o estimador de cumeieira para  $Y = X^*\alpha + \varepsilon$ . Como

$$\begin{aligned}
\hat{\alpha}(k) &= (\Lambda + kI)^{-1}X^*Y \\
&= (\Lambda + kI)^{-1}(X^*X^*)(X^*X^*)^{-1}X^*Y \\
&= (\Lambda + kI)^{-1}\Lambda\hat{\alpha} \\
&= \begin{pmatrix} \frac{\lambda_1}{\lambda_1 + k} & 0 & \cdots & 0 \\ 0 & \frac{\lambda_2}{\lambda_2 + k} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{\lambda_p}{\lambda_p + k} \end{pmatrix} \begin{pmatrix} \hat{\alpha}_1 \\ \vdots \\ \hat{\alpha}_p \end{pmatrix}
\end{aligned}$$

em que  $\hat{\alpha}$  é o estimador de quadrados mínimos de  $Y = X^*\alpha + \varepsilon$ . Note que

$$\begin{aligned}
\hat{\alpha} &= (X^*X^*)^{-1}X^*Y \\
&= (P'X'XP)^{-1}(XP)Y \\
&= P(X'X)^{-1}PP'X'Y \\
&= P(X'X)^{-1}X'Y \\
&= P\hat{\beta}
\end{aligned}$$

Tem-se então para as coordenadas do vetor de parâmetros  $\hat{\alpha}(k)$

$$\hat{\alpha}_i(k) = \left( \frac{\lambda_i}{\lambda_i + k} \right) \hat{\alpha}_i \quad (16)$$

isto é, tem-se um encolhimento do estimador de quadrados mínimos  $\hat{\alpha}_i$ . Como

$$\begin{aligned} E((\hat{\alpha}_i(k) - \alpha_i)^2) &= E(\hat{\alpha}_i^2(k) - 2\hat{\alpha}_i(k)\alpha_i + \alpha_i^2) \\ &= E(\hat{\alpha}_i^2(k)) - 2\alpha_i E(\hat{\alpha}_i(k)) + \alpha_i^2 \\ &= \frac{\lambda_i^2}{(\lambda_i + k)^2} E(\hat{\alpha}_i^2) - 2\alpha_i \left( \frac{\lambda_i}{\lambda_i + k} \right) E(\hat{\alpha}_i) + \alpha_i^2 \\ &= \frac{\lambda_i^2}{(\lambda_i + k)^2} \left( \frac{\sigma^2}{\lambda_i} + \alpha_i^2 \right) - 2\alpha_i \frac{\lambda_i}{\lambda_i + k} \alpha_i + \alpha_i^2 \end{aligned}$$

$$\frac{d}{dk} E((\hat{\alpha}_i - \alpha_i)^2) = \frac{-\lambda_i^2(\lambda_i \sigma^2 + \alpha_i^2)2(\lambda_i + k)}{(\lambda_i + k)^4} + \frac{2\alpha_i^2 \lambda_i}{(\lambda_i + k)^2} = 0$$

$$\frac{-\lambda_i^2 \left( \lambda_i \frac{\sigma^2}{\lambda_i} + \alpha_i \right)}{\lambda_i + k} + \alpha_i^2 \lambda_i = 0$$

$$\frac{\lambda_i \left( \lambda_i \frac{\sigma^2}{\lambda_i} + \alpha_i \right)}{\lambda_i + k} = \alpha_i^2 \Rightarrow \lambda_i + k = \frac{\lambda_i \left( \lambda_i \frac{\sigma^2}{\lambda_i} + \alpha_i \right)}{\alpha_i^2}$$

$$k = \frac{\sigma^2}{\alpha_i^2} + \lambda_i - \lambda_i = \frac{\sigma^2}{\alpha_i^2}$$

Tem-se então um  $k_{otimo}$  para cada coordenada. Um valor único para  $k$  para todas as coordenadas pode ser definido utilizando-se a média  $\frac{\alpha_1^2 + \dots + \alpha_p^2}{p}$

$$k = \frac{\sigma^2}{\frac{\sum \alpha_i^2}{p}} = \frac{p\sigma^2}{\sum \alpha_i^2}$$

Tal fato sugeriu a definição do estimador

$$\hat{k}_{otimo} = \frac{p\hat{\sigma}^2}{\sum \hat{\alpha}_i^2}$$

e correspondente estimador de cumeieira

$$\hat{\alpha}_i(\hat{k}_{otimo}) = \left( \frac{\lambda_i}{\lambda_i + \frac{p\hat{\sigma}^2}{\sum \hat{\alpha}_i^2}} \right) \hat{\alpha}_i$$

denominado estimador *HKB*. Uma crítica a este estimador é o fato de se estimar  $\alpha_i^2$  por  $\hat{\alpha}_i^2$ , pois

$$E(\hat{\alpha}_i^2) = var(\hat{\alpha}_i) + (E(\hat{\alpha}_i))^2 = \frac{\sigma^2}{\alpha_i} + \alpha_i^2$$

e, portanto com alta probabilidade  $\hat{\alpha}_i^2$  superestima  $\alpha_i^2$ .

Portanto, estimar  $\frac{\sigma^2}{\alpha_i^2}$  por  $\frac{\hat{\sigma}^2}{\hat{\alpha}_i^2}$  é um procedimento inadequado. Baseado em argumentos Bayesianos Lawless e Wang (1976) propuseram o uso da média ponderada  $\frac{\sum \lambda_i \hat{\alpha}_i^2}{p}$  obtendo então o seguinte estimador do  $k_{otimo}$ :

$$\hat{k}_{otimo} = \frac{p\hat{\sigma}^2}{\sum \lambda_i \hat{\alpha}_i^2}$$

É também possível obter bons estimadores a partir de processos iterativos. Um deles, também estudado por Lawless e Wang (1976) é obtido por: utilizando-se o estimador de quadrados mínimos  $\hat{\alpha}_i$  em 2.10 obtêm-se primeiramente o estimador de Hoerl e Kennard (1976). O processo iterativo é iniciado.

No segundo passo,  $\hat{\alpha}_i$  substitui  $\hat{\alpha}_i$  e um novo estimador é obtido

$$(\hat{\alpha}_i)_l = \frac{\lambda_i}{\lambda_i + \frac{\hat{\sigma}^2}{(\hat{\alpha}_i^2)_{l-1}}}.$$

O processo se repete até que algum tipo de convergência ocorra.

A literatura relacionada à obtenção de estimadores para o parâmetro ótimo  $k_{otimo}$ , na regressão de cumeira é extensa e continua sendo uma área de pesquisas.

### 2.11 Outros métodos de obtenção de estimadores de cumeira

Estimadores de cumeira podem ser obtidos por considerações teóricas diferentes das consideradas no início deste texto. A derivação dos estimadores de cumeira como solução do problema variacional de se obter sobre uma elipse o ponto mais próximo da origem, seja talvez a mais interessante. Mas, uma abordagem é utilizar os estimadores mistos (GRUBER, 1998).

Considere que o modelo linear de posto completo  $\mathbf{Y}_{n \times 1} = X_{n \times p} \boldsymbol{\beta}_{p \times 1} + \varepsilon_{n \times 1}$ ,  $cov(\varepsilon_{n \times 1}) = \sigma^2 I_{n \times n}$  será aplicado para analisar um determinado experimento. No entanto, o experimentador sabe que outro modelo linear foi aplicado em um experimento similar, também de posto completo  $\mathbf{Z}_{k \times 1} = R_{k \times p} \boldsymbol{\beta}_{p \times 1} + \phi_{k \times 1}$ ,  $cov(\phi_{k \times 1}) = \tau^2 I_{k \times k}$ . Isto é, os dois experimentos possuem as mesmas covariáveis diferindo apenas no número utilizado de valores destas em cada experimento.

Supondo os dois modelos independentes, isto é,  $\varepsilon_{n \times 1}$  e  $\phi_{k \times 1}$  são vetores aleatórios independentes, uma forma de incorporar ambos modelos na análise é considerar o modelo linear aumentado:

$$\begin{pmatrix} Y \\ Z \end{pmatrix} = \begin{pmatrix} X \\ R \end{pmatrix} \boldsymbol{\beta} + \begin{pmatrix} \varepsilon \\ \phi \end{pmatrix} = H_{(n+k) \times p} \boldsymbol{\beta}_{p \times 1} + \eta_{(n+k) \times 1}$$

$$\text{cov}(\eta)_{n+k \times p} = \begin{pmatrix} \sigma^2 I_{n \times n} & \\ & \tau^2 I_{k \times k} \end{pmatrix} = D$$

O estimador de Gauss-Markov de  $\beta$  para este modelo é dado por (RENCHE, 2008):

$$\hat{\beta} = (H' D H)^{-1} H' D Y.$$

Com o uso de álgebra linear para matrizes particionadas obtém-se:

$$\hat{\beta} = (X' X \tau^2 + R' R \sigma^2)^{-1} (\tau^2 X' Y + \sigma^2 R' Z).$$

Suponha agora que os valores observados no segundo experimento podem ser considerados como valores provenientes do erro experimental, isto é, como  $Z = R\beta + \phi$  pode-se afirmar que  $R\beta = 0$  e que estocasticamente  $Z = 0$ . Neste caso o estimador  $\hat{\beta}$  fica de forma:

$$\begin{aligned} \hat{\beta} &= (X' X \tau^2 + R' R \sigma^2)^{-1} (\tau^2 X' Y) \\ &= (X' X + \frac{\sigma^2}{\tau^2} R' R)^{-1} (X' Y) \\ &= (X' X + G)^{-1} X' Y \end{aligned}$$

com  $G = \frac{\sigma^2}{\tau^2} R' R$ . Portanto, o estimador *Gauss-Markov* é dado por um estimador de cumeira de Rao.

Esta abordagem é interessante, pois obtém um estimador de cumeira como o estimador de quadrados mínimos que se obtém ao incluir no modelo linear que será utilizado em uma informação *a priori*, isto é, a realização prévia de um primeiro experimento ao modelo que será utilizado.

É importante observar que este procedimento que utiliza um valor estocástico  $Z = 0$  como *priori* é diferente da abordagem determinística em que se

tem uma restrição paramétrica  $R\beta = 0$ . Pois, no caso da restrição paramétrica o estimador  $\hat{\beta}$  obtido é tal que  $R\hat{\beta} = 0$ . No caso do estimador obtido supondo a priori  $Z = 0$  estocástico  $R\hat{\beta} = R(X'X + G)^{-1}X'Y$  não é identicamente igual a zero.

Esta construção generaliza a obtenção dos estimadores que são obtidos pela perturbação na matriz  $X'X$  pela matriz  $I$  ou a matriz diagonal  $K$  por uma perturbação dada por uma matriz  $G$  positiva definida. Estes estimadores são denominados estimadores de cumeeira de Rao, (GRUBER, 1998) e estudados recentemente em Costa (2014).

Os estimadores de cumeeira também podem ser obtidos como casos particulares de estimadores de *Bayes* e também como casos particulares de estimadores minimax. Em geral os textos sobre estimadores de cumeeira adotam apenas uma derivação teórica destes, no entanto as outras abordagens são importantes na compreensão e interpretação dos resultados e do uso destes estimadores.

### 3 RESULTADOS E DISCUSSÃO

Logo após Hoerl e Kennard (1970a) apresentarem a teoria dos estimadores de cumeieira, Mayer e Wilke (1973) apresentaram um estimador de cumeieira, que é um caso particular do estimador de cumeieira de Rao que hoje leva seus nomes. Neste trabalho foi proposto um novo estimador de Rao e Toutenburg (1999) semelhante ao estimador de Mayer e Wilke (1973). As propriedades deste novo estimador serão desenvolvidas e um estudo computacional comparativo entre eles será explicado. O estudo comparativo se dará com dados de Gorman e Toman (1966) analisados anteriormente com a utilização de estimadores Hoerl e Kennard (1970a) e Mayer e Wilke (1973). Um exemplo por simulação computacional também será considerado utilizando-se o mesmo esquema desenvolvido em Marquardt(1970). Estes estimadores também são utilizados em dados de seleção genômica, anteriormente analisados em Silveira (2014) utilizando o método *PLS*.

Para uma melhor compreensão, uma exposição sumária da construção de Mayer & Wilke (1973) será apresentada.

#### 3.1 O estimador de Mayer & Wilke

Sendo  $C$  a classe de todos os estimadores lineares de  $\beta$  obtidos como transformações lineares do estimador de quadrados mínimos, isto é,

$$C = \left\{ \hat{\beta}_A = A\hat{\beta}, A : R^p \rightarrow R^p \right\}.$$

Considere também como função perda o quadrado do resíduo:

$$\begin{aligned}
L(\hat{\beta}_A) &= (Y - XA\hat{\beta})' (Y - XA\hat{\beta}) \\
&= (Y - XA\hat{\beta} + X\hat{\beta} - X\hat{\beta})' (Y - XA\hat{\beta} + X\hat{\beta} - X\hat{\beta}) \\
&= [(Y - X\hat{\beta}) - (XA\hat{\beta} - X\hat{\beta})]' [(Y - X\hat{\beta}) - (XA\hat{\beta} - X\hat{\beta})] \\
&= (Y - X\hat{\beta})' (Y - X\hat{\beta}) - (Y - X\hat{\beta})' (XA\hat{\beta} - X\hat{\beta}) \\
&\quad (XA\hat{\beta} - X\hat{\beta})' (Y - X\hat{\beta}) + (XA\hat{\beta} - X\hat{\beta})' (XA\hat{\beta} - X\hat{\beta}) \\
&= (Y - XA\hat{\beta})' (Y - XA\hat{\beta}) + \hat{\beta}' (A - I)' X' X (A - I) \hat{\beta}
\end{aligned}$$

De acordo com a Figura 22, o duplo produto  $(Y - X\hat{\beta})' (XA\hat{\beta} - X\hat{\beta}) = 0$  pois o vetor  $(Y - X\hat{\beta})$  é perpendicular ao plano  $Im(X)$  (Figura 23).



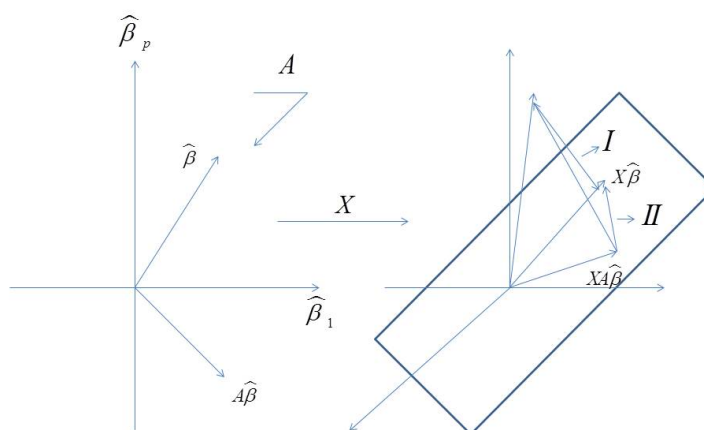


Figura 22 Reparametrização linear

Mayer e Wilke (1973) obtêm dentro da classe de estimadores  $C$  o estimador  $\hat{\beta}_{\lambda I} = \lambda \hat{\beta}$  como um estimador com propriedades que satisfazem certas condições ótimas. Este estimador é denominado estimador de Mayer & Wilke e é simplesmente um encolhimento do estimador de quadrados mínimos. É também um caso particular de um estimador de cumeeira de Rao. Estes estimadores são da forma  $\hat{\beta}_R = (kG + X'X)^{-1}X'Y$  em que  $G$  é uma matriz positiva definida (em geral o contexto deixa claro qual a matriz  $G$  utilizada) cujas propriedades foram apresentadas em Costa (2014).

O estimador de Mayer & Wilke é um estimador de cumeeira de Rao quando se toma  $G = X'X$ .

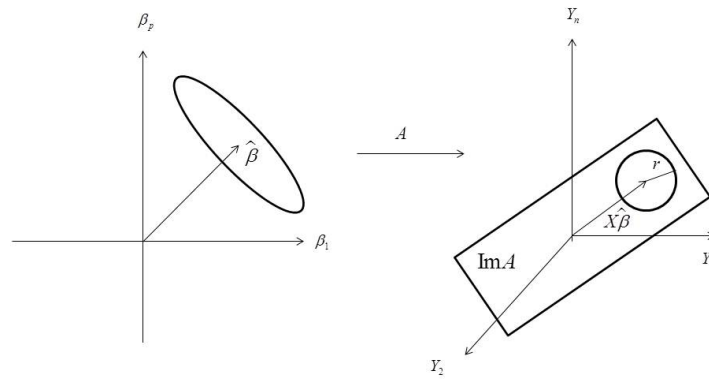


Figura 23 Elipsoide centrado em  $\hat{\beta}$

$$\begin{aligned}
 \hat{\beta}_R &= (kG + X'X)^{-1}X'Y \\
 &= (k(X'X) + X'X)^{-1}X'Y \\
 &= (kI + I)^{-1}(X'X)^{-1}X'Y \\
 &= \frac{1}{k+1}\hat{\beta} \\
 &= \lambda\hat{\beta}
 \end{aligned}$$

Geometricamente o estimador de Hoerl & Kennard e o estimador de Mayer & Wilke são obtidos respectivamente pelas intersecções de um círculo com um elipsoide e a intersecção de dois elipsoides paralelos (Figura 24).

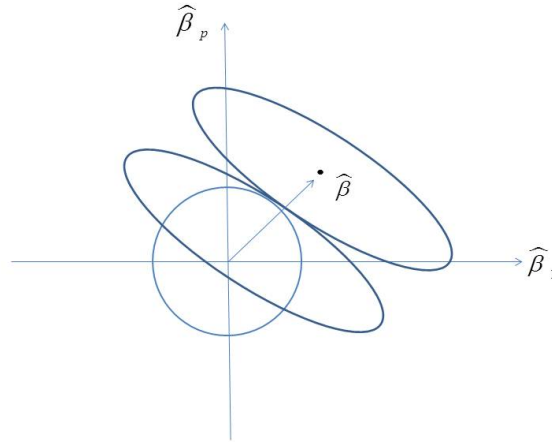


Figura 24 Interpretação geométrica do estimador de Mayer & Wilke

### 3.2 Uma proposta para um novo estimador de cuneeira de Rao

A questão que se quer abordar é para os estimadores de cuneeira de *Rao*  $\hat{\beta}_R = (X'X + kG)^{-1}X'Y$  como se fazer uma boa escolha da matriz  $G$ ?

Para o estimador de Mayer & Wilke a escolha foi  $G = X'X$  obtendo-se  $\lambda\hat{\beta}$ . No caso de quasecolinearidade esse estimador pode não ser adequado pois como  $var(\lambda\hat{\beta}) = \lambda^2 var(\hat{\beta})$  e a variância do estimador de quadrados mínimos é muito grande, o fator de encolhimento  $\lambda$  deverá ser muito pequeno acarretando um “excesso” de encolhimento, o que pode não ser muito realista.

A proposta de escolha da matriz  $G$  desenvolvida neste trabalho é baseada nos argumentos de Hoerl & Kennard no artigo “*Ridge regression: Applications to nonorthogonal problems, 1970*”. Neste artigo as curvas de cuneeira são vistas como um retrato das complexas interrelações existentes entre as covariáveis e a seus efeitos na estimação do vetor de parâmetros  $\beta$ . Ao plotar no plano bidimensional os valores das coordenadas de  $\hat{\beta}(k) = (\hat{\beta}_1(k), \dots, \hat{\beta}_p(k))$  em função de  $k$  a

ideia básica é se, por exemplo  $\hat{\beta}_1(0)$  é um valor alto e decresce rapidamente com  $k$ , o valor  $\hat{\beta}_1(0)$  deve ter sido superestimado, em razão da grande variabilidade de  $\hat{\beta}$ .

Se a covariável  $x_1$  é realmente importante na determinação da resposta  $y$ , uma pequena variação de  $k$  no estimador de cumeeira não poderia afetar demasiadamente a estimativa de  $\hat{\beta}_1(k)$ . Portanto, o que se quer é certa estabilidade, isto é, para algum intervalo de valores de  $k$ , os valores das estimativas  $\hat{\beta}(k)$  não se alterem muito. Este é um critério empírico sugerido para se obter uma estimativa do valor ótimo para  $k$  e conseqüentemente se obter o estimador de cumeeira mais apropriado. Estas ponderações são descritas de forma esquemática na Figura 25:

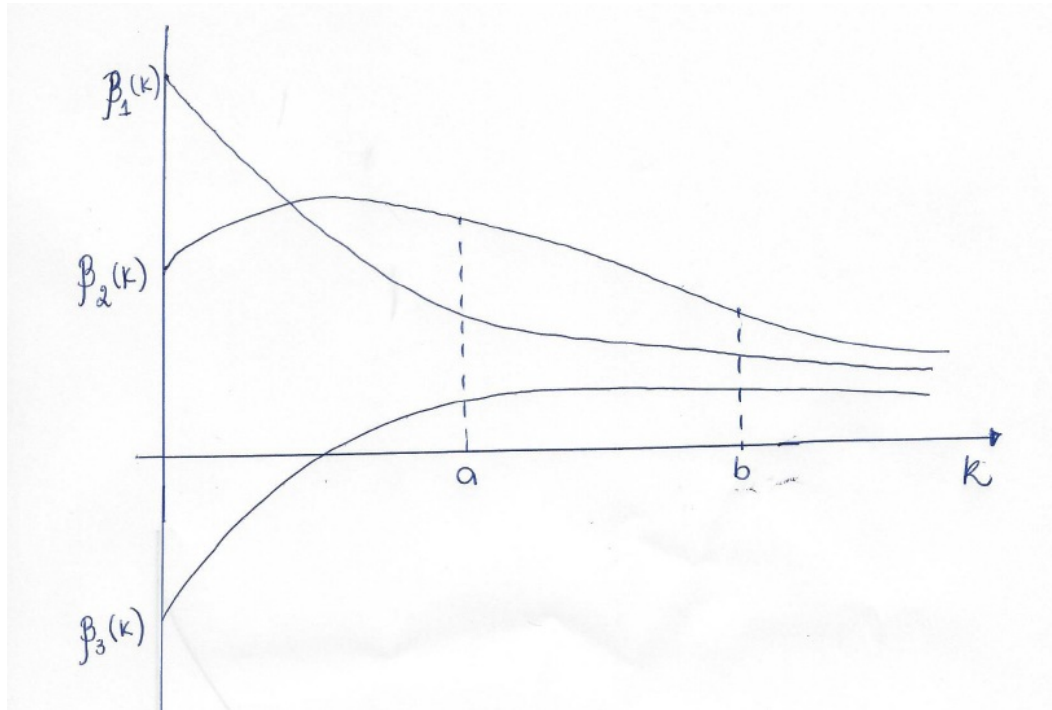


Figura 25 Estudo de curvas de cumeeira

- $\hat{\beta}_1(k)$  inicialmente superestimado, bastante instável diminuindo rapidamente significando que a covariável  $x_1$  não é expressiva na estimação.
- $\hat{\beta}_2(k)$  levemente superestimado, mas bastante estável.
- $\hat{\beta}_3(k)$  estimado inicialmente com sinal trocado, mas atingindo o sinal correto se estabilizando.
- O valor ótimo de  $k$  deve se encontrar no intervalo  $[a, b]$  em que ocorre uma certa estabilidade nas curvas.

A seguinte explicação geométrica é proposta para a interpretação do “intervalo de estabilidade”. Sem perda de generalidade considere o caso bidimensional com seu

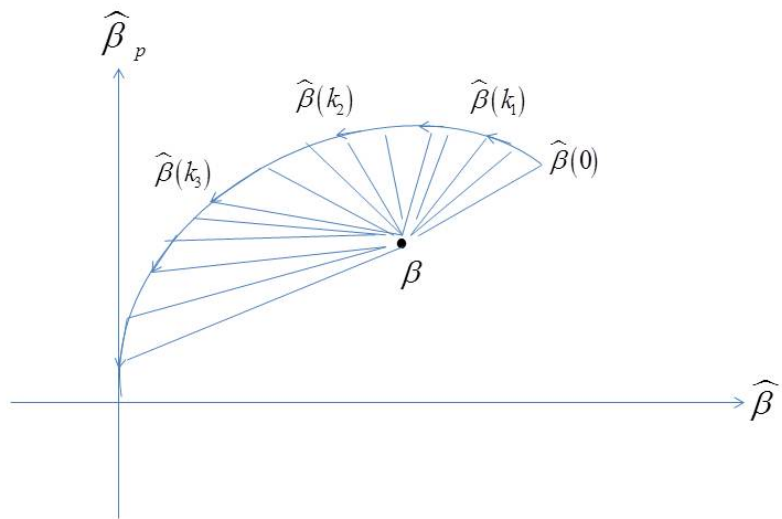


Figura 26 Região de utilização do estimador de cumeeira de Rao

verdadeiro valor do parâmetro  $\beta = (\beta_1, \beta_2)$ . Plotando-se as curvas parametrizada  $\beta(k) = (\beta_1(k), \beta_2(k))$  vamos supor que a velocidade desta curva seja constante,  $\left\| \frac{d}{dk} \hat{\beta}(k) \right\| = cte$  (Figura 26). Desta forma as distâncias iguais ao longo da curva correspondem a intervalos iguais em relação a  $k$ .

A distância  $\left\| \hat{\beta}(k) - \beta \right\|$ , entre a estimativa  $\hat{\beta}(k)$  e o verdadeiro vetor de parâmetros  $\beta$ , varia menos quando a curva  $\hat{\beta}(k)$  se aproxima de  $\beta$ , isso significa que a distância entre as coordenadas  $\left\| \hat{\beta}(k)_1 - \beta_1 \right\|$  e  $\left\| \hat{\beta}(k)_2 - \beta_2 \right\|$  também deve variar menos. Logo, no gráfico das curvas de cumeeira as duas curvas  $\hat{\beta}(k)_1$  e  $\hat{\beta}(k)_2$  ficam aproximadamente paralelas ao eixo horizontal. Portanto, tem-se uma

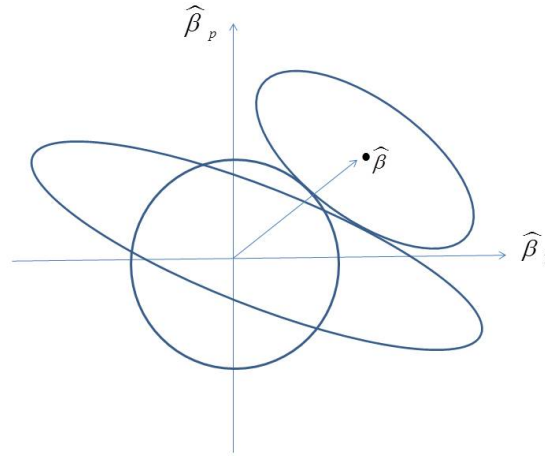


Figura 27 Os estimadores de Hoerl & Kennard e de Mayer & Wilke

certa estabilidade das curvas. Assim, o valor ótimo de  $k$ , isto é, o valor que minimiza  $E \left[ \left\| \hat{\beta}(k) - \beta \right\|^2 \right]$  pode ser estimado pelo valor  $k$  que minimiza  $\left\| \hat{\beta}(k) - \beta \right\|^2$  e este deve se encontrar no intervalo em que ocorre a estabilização. Com esta interpretação uma ideia promissora na escolha de  $G$  seria a que acelera a velocidade da curva parametrizada  $\hat{\beta}_R(k)$  para que o valor ótimo seja atingido mais rapidamente.

O estimador  $\hat{\beta}_R(k)$  é obtido pelas interseções de dois elipsóides, um definido por  $X'X$  centrado em  $\hat{\beta}$  e o outro centrado na origem. No caso do estimador de Hoerl & Kennard é uma esfera e no caso do estimador Mayer & Wilke é um elipsóide paralelo (Figura 27).

Para acelerar a curva traço de cumeeira é sugerido então  $G = (X'X)^{-1}$  definindo o estimador proposto  $\hat{\beta}(k)_{prop} = (X'X + k(X'X)^{-1})^{-1}X'Y$ . Desta forma o estimador será obtido pela intersecção entre dois elipsóides ortogonais entre si. A ortogonalidade é entendida aqui por, se uma direção define o maior eixo do elipsóide definido por  $X'X$ , esta direção irá definir a direção do menor eixo no

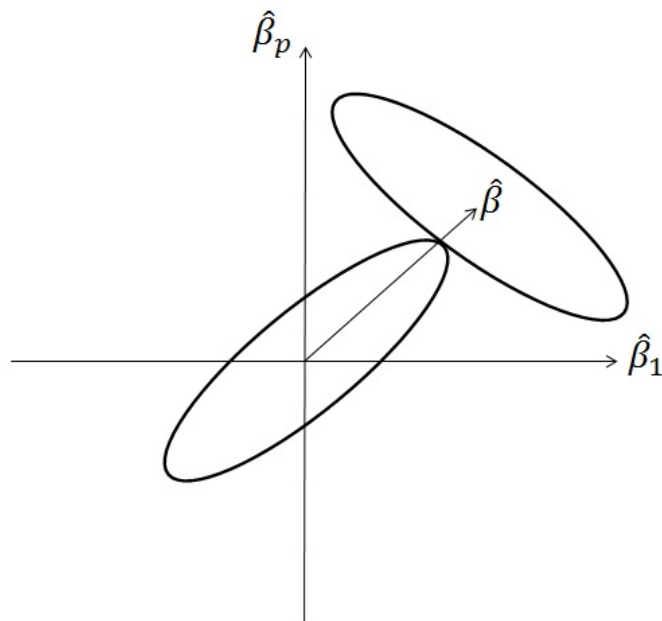


Figura 28 Região de utilização do estimador de cumeeira de Rao

elipse definido por  $(X'X)^{-1}$  (Figura 28). Note que se  $\lambda_1 \geq \lambda_2 \geq \dots \lambda_p > 0$ , são os autovalores de  $X'X$ , então  $\frac{1}{\lambda_p} \geq \frac{1}{\lambda_{p-1}} \geq \dots \frac{1}{\lambda_2} \geq \frac{1}{\lambda_1} > 0$ , são os autovalores de  $(X'X)^{-1}$ .

Outro argumento que sugere a superioridade do estimador proposto em relação ao estimador de Mayer & Wilke é sua variância total. Para o cálculo da variância total é necessário o seguinte fato trivial de álgebra linear. Se  $A$  é uma matriz ortogonal que diagonaliza  $X'X$ , isto é,  $A(X'X)A^{-1} = D_1$  então  $A(X'X)^{-1}A^{-1} = D_1^{-1}$  pois  $A(X'X)^{-1}A^{-1}D_1 = A(X'X)^{-1}A^{-1}A(X'X)A^{-1} = I$  e portanto  $A(X'X)^{-1}A^{-1} = D_1^{-1}$

A variância total do estimador proposto  $\hat{\beta}(k)_{prop}$  é:

$$var(\hat{\beta}(k)_{prop}) = tr(X'X + k(X'X)^{-1})^{-1} X' var(\mathbf{Y}) X (X'X + k(X'X)^{-1})^{-1}$$



Portanto, se  $A$  diagonaliza  $X'X$ , então:

$$\begin{aligned}
&= A(X'X + k(X'X)^{-1})^{-1}X'X(X'X + k(X'X)^{-1})^{-1}A^{-1} \\
&= A(X'X + k(X'X)^{-1})^{-1}A^{-1}AX'XA^{-1}A(X'X + k(X'X)^{-1})^{-1}A^{-1} \\
&= (A(X'X + k(X'X)^{-1})A^{-1})^{-1}(AX'XA^{-1})(A(X'X + k(X'X)^{-1})A^{-1})^{-1} \\
&= (A(X'X)A^{-1} + kA(X'X)^{-1}A^{-1})(AX'XA^{-1})(A(X'X)A^{-1} \\
&\quad + kA(X'X)^{-1}A^{-1}) \\
&= (D_1 + kD_1^{-1})^{-1}D_1(D_1 + kD_1^{-1})^{-1} \\
&= (D_1 + kD_1^{-1})^{-2}D_1
\end{aligned}$$

Logo

$$\begin{aligned}
\text{var}(\hat{\beta}(k)_{prop}) &= \text{tr}((D_1 + kD_1^{-1})^{-2}D_1) \\
&= \sum_{i=1}^p \left( \frac{\lambda_i}{\lambda_i^2 + k} \right)^2 \lambda_i
\end{aligned}$$

Recordando que o estimador de quadrados mínimos terá variância total grande quando  $X'X$  tiver um autovalor  $\lambda_i$  pequeno uma vez que a variância total depende do inverso de  $\lambda_i$ . Mas, se  $\lambda_i$  é pequeno então  $\frac{\lambda_i^3}{(k + \lambda_i)^2}$  é pequeno e portanto a variabilidade do estimador proposto não será grande. Desta forma o problema da quasecolinearidade é contornado.

### 3.3 Erro quadrático médio do estimador proposto

$$\begin{aligned}\hat{\beta}_{prop}(k) &= [X'X + k(X'X)^{-1}]^{-1}X'Y \\ W(k) &= [X'X + k(X'X)^{-1}]^{-1} \\ Z(k) &= [I + k(X'X)^{-2}]^{-1} = [X'X + k(X'X)^{-1}]^{-1}(X'X) = W(k)(X'X) \\ W(k) &= Z(k)(X'X)^{-1} \\ \hat{\beta}_{prop}(k) &= Z(k)\hat{\beta}\end{aligned}$$

$$\begin{aligned}Z(k) + k[W(k)](X'X)^{-1} &= W(k)(X'X) + k[W(k)](X'X)^{-1} \\ &= W(k)[X'X + k(X'X)^{-1}] \\ &= W(k)[W(k)]^{-1} \\ &= I\end{aligned}$$

O que implica em  $Z(k) = I - k[W(k)](X'X)^{-1}$ .

Logo o *EQM* do estimador proposto fica da forma

$$\begin{aligned}EQM(\hat{\beta}_{prop}(k)) &= E[(\hat{\beta} - \beta)'Z(k)'Z(k)(\hat{\beta} - \beta)] \\ &= tr(\sigma^2(X'X)^{-1}Z(k)'Z(k)) \\ &= \sigma^2 tr(X'X)^{-1}(Z(k)'Z(k))\end{aligned}$$

$$\begin{aligned}\beta'(Z(k) - I)'(Z(k) - I)\beta &= \beta'[-k(X'X + k(X'X)^{-1})^{-1}][-k(X'X + k(X'X)^{-1})^{-1}]\beta \\ &= k^2\beta'(X'X + k(X'X)^{-1})^{-2}(X'X)^{-2}\beta\end{aligned}$$

e

$$\begin{aligned}
Z(k)'Z(k) &= [I - kW(k)(X'X)^{-1}]^{-2} \\
&= [I - k[X'X + k(X'X)^{-1}]^{-1}(X'X)^{-1}]^{-2} \\
&= I - 2k[X'X + k(X'X)^{-1}]^{-1}(X'X)^{-1} \\
&\quad + k^2[X'X + k(X'X)^{-1}]^{-2}(X'X)^{-2}
\end{aligned}$$

$$\begin{aligned}
EQM(\hat{\beta}_{prop}(k)) &= \sigma^2 tr[(X'X)^{-1}(I - 2k[X'X + k(X'X)^{-1}]^{-1}(X'X)^{-1} \\
&\quad + k^2[X'X + k(X'X)^{-1}]^{-2}(X'X)^{-2}) \\
&\quad + k^2\beta'(X'X + k(X'X)^{-1})^{-2}(X'X)^2\beta] \\
&= \sigma^2 \left( \sum_{i=1}^p \frac{1}{\lambda_i} - 2k \sum_{i=1}^p \frac{1}{\lambda_i} \frac{\lambda_i}{\lambda_i^2 + k} \frac{1}{\lambda_i} + k^2 \sum_{i=1}^p \frac{1}{\lambda_i} \frac{\lambda_i^2}{\lambda_i^2 + k^2} \frac{1}{\lambda_i} \right) \\
&= \sigma^2 \sum_{i=1}^p \frac{1}{\lambda_i} \left[ 1 - \frac{2k}{(\lambda_i^2 + k)} + \frac{k^2}{(\lambda_i^2 + k)^2} \right] \\
&= \sigma^2 \sum_{i=1}^p \left[ \frac{\lambda_i^3}{(\lambda_i^2 + k)^2} \right]
\end{aligned}$$

De  $k^2\beta'(X'X + k(X'X)^{-1})^{-2}(X'X)^2\beta$  tem-se:

$$\sum_{i=1}^p \frac{\alpha_i^2 k^2}{(\lambda_i^2 + k)^2}$$

Finalmente, tem-se que:

$$EQM[\hat{\beta}] = \underbrace{\sigma^2 \sum_{i=1}^p \left[ \frac{\lambda_i^3}{(\lambda_i^2 + k)^2} \right]}_{\gamma_1(k)} + \underbrace{\sigma^2 \sum_{i=1}^p \frac{\alpha_i^2 k^2}{(\lambda_i^2 + k)^2}}_{\gamma_2(k)}$$

### 3.4 Um exemplo por simulação computacional

Este trabalho é em essência teórico e um estudo por simulação computacional, este feito através do software R, para o estimador proposto será objeto de pesquisas futuras. No entanto no sentido de ser mais conclusivo um exemplo simples, bidimensional estudado em Marquardt e Snee (1975) e analisado com o uso de um estimador de cumeeira de Rao em Costa, Chaves e Souza (2014) será abordado por simulação. Considere o modelo linear:

$$Y = X\beta + \varepsilon$$

$$X = \begin{pmatrix} \frac{3\sqrt{2}}{10} & \frac{4\sqrt{2}}{10} \\ \frac{4\sqrt{2}}{3\sqrt{2}} & \frac{10}{3\sqrt{2}} \\ \frac{10}{5\sqrt{2}} & \frac{10}{5\sqrt{2}} \\ \frac{10}{10} & \frac{10}{10} \end{pmatrix}, \quad X'X = \begin{pmatrix} 1 & \frac{49}{50} \\ \frac{49}{50} & 1 \end{pmatrix}$$

Os autovalores da matriz desta matriz são 1,98 e 0,02. O autovalor 0,02 caracteriza ocorrência de quase colinearidade. Seguem os estimadores utilizados:

- O estimador de quadrados mínimos  $\hat{\beta} = (X'X)^{-1}X'Y$ .
- O estimador de cumeeira de Hoerl & Kennard  $\hat{\beta}_{HK}(k) = (X'X + kI)^{-1}X'Y$ .
- O estimador de cumeeira de Mayer & Wilke  $\hat{\beta}_{MW}(k) = (X'X + kX'X)^{-1}X'Y = \frac{1}{1+k}\hat{\beta}$ .
- O estimador proposto  $\hat{\beta}_{prop}(k) = (X'X + k(X'X)^{-1})^{-1}X'Y$ .

Fixado um vetor de parâmetros  $\beta = [3, 5]$  e  $\sigma^2 = 1$  gerou-se um vetor de respostas  $y = [6, 34 \quad 3, 94 \quad 5, 96]$ . A partir desse vetor as curvas de cumeeira

são obtidas nas Figuras (29 , 30, 31).

Com base nas Figuras (29 , 30, 31), visualmente utilizou-se um critério que o valor ótimo para o parâmetro  $k$  ocorre quando as curvas de cumeeira apresentam uma certa estabilidade. Logo, obtêm-se como valores ótimos  $k = 0,3$  para o estimador de Hoerl & Kennard,  $k = 6,0$  para o estimador de Mayer & Wilke e  $k = 0,05$  para o estimador proposto.

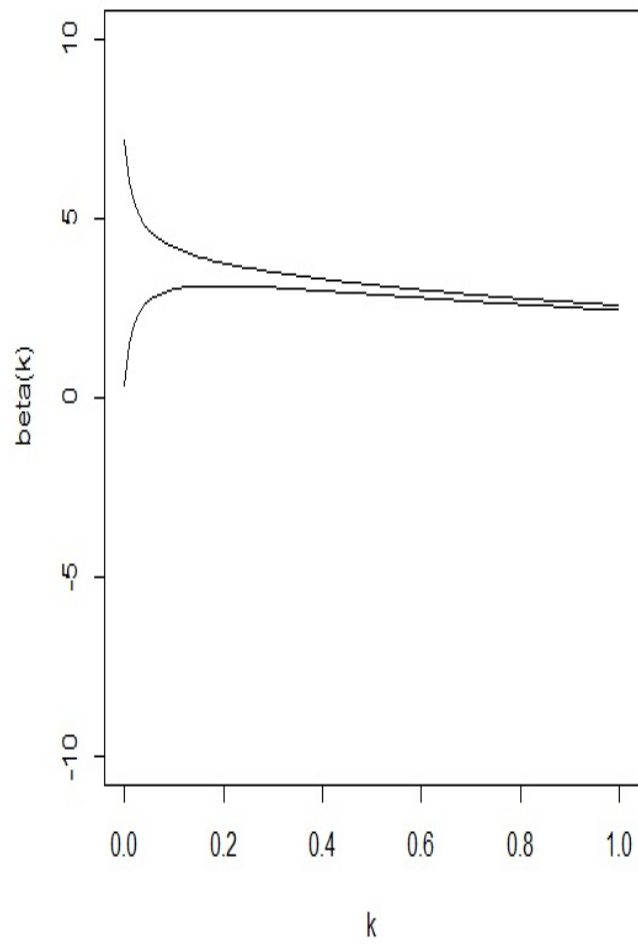


Figura 29 Curvas traço de cumeeira do estimador de Hoerl & Kennard

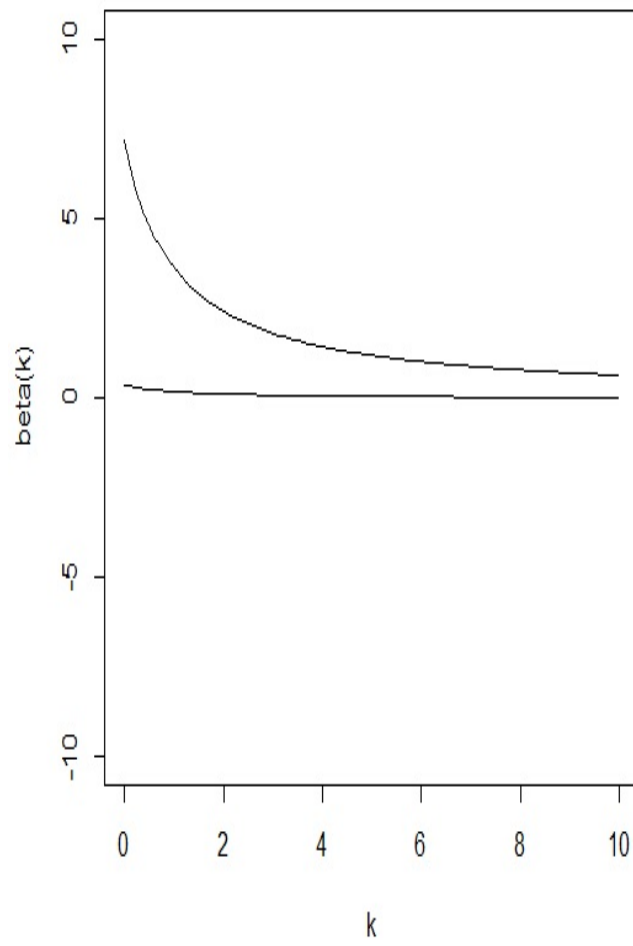


Figura 30 Curvas traço de cumeeira do estimador de Mayer& Wilke

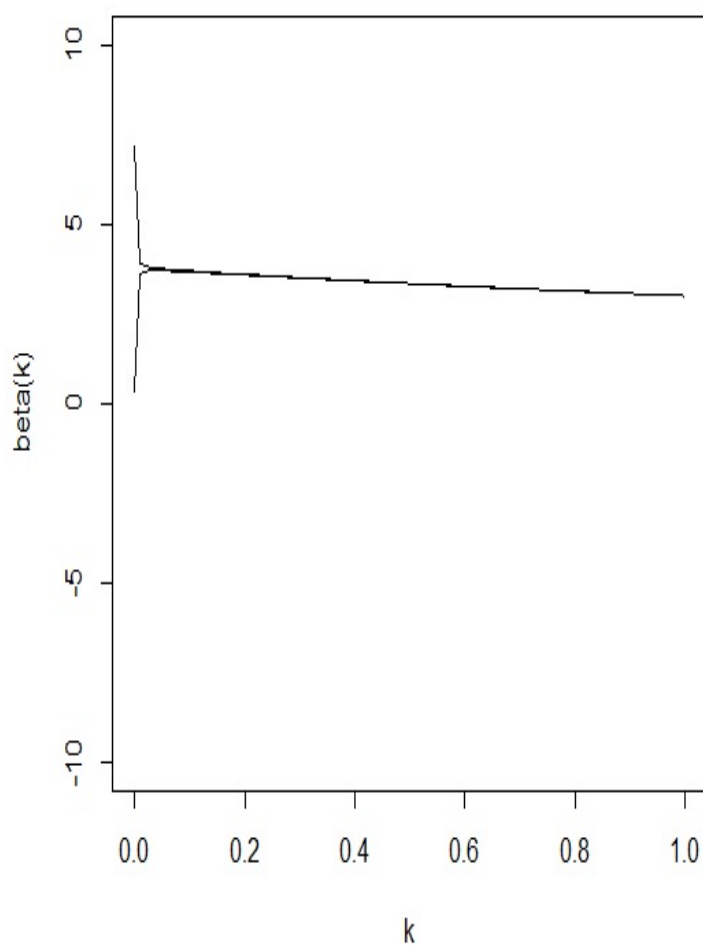


Figura 31 Curvas traço de cumeeira do estimador proposto

Tem-se então que o estimador proposto apresenta erro quadrático médio menor que o estimador de cumeeira usual. Além disso, o valor adequado de  $k$  para o estimador proposto ocorre com valores muito menores que os do estimador de



cumeeira, este fato pode representar uma vantagem significativa pois a obtenção de valores adequados de  $k$  para estimadores de cumeeira é um dos fatores limitantes de seu uso, uma vez que não se tem estimadores com propriedades suficientemente boas para o seu cálculo. O erro quadrático médio para os quatro estimadores são obtidos gerando-se 1000 valores para o vetor  $Y$  e obtendo-se então 1000 estimativas para cada um dos estimadores. Os resultados são representados na Tabela 1

Tabela 1 Erro quadrático médio

$\hat{\beta}$	EQM	=	variância	+	viés
$\hat{\beta}_{QM}$	50,500	=	50,500	+	0
$\hat{\beta}_{HK}(0, 3)$	2,776	=	0,594	+	2,182
$\hat{\beta}_{MW}(6, 0)$	13,628	=	1,188	+	12,400
$\hat{\beta}_{prop}(0, 05)$	2,472	=	0,966	+	1,506

### 3.5 Desempenho do estimador proposto aplicado a um conjunto de dados reais

No artigo de Hoerl e Kennard (1970b) foi feita uma análise do desempenho do estimador de cumeeira utilizando-se dados publicados em Gorman e Toman (1966) em que se tem um problema de regressão com dez fatores. As matrizes de correlação amostrais são:



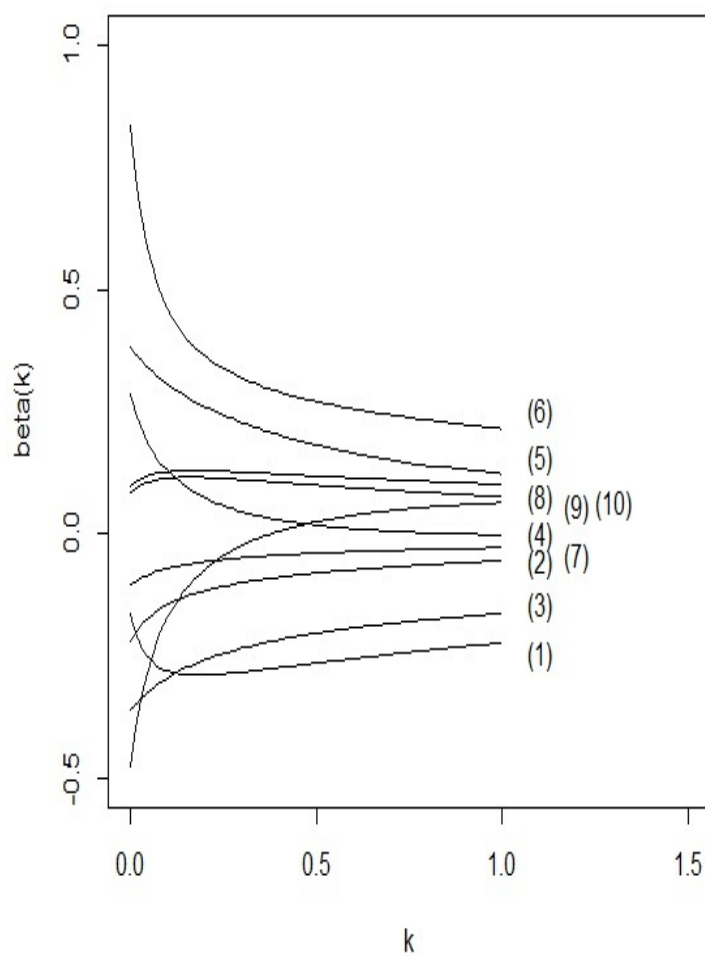


Figura 32 Curvas traço de cumeeira do estimador de Hoerl & Kennard

- Os coeficientes das estimativas de quadrados mínimos estão aparentemente superestimados uma vez que eles diminuem rapidamente em função de  $k$ . Este comportamento fica mais acentuado para os fatores 5 e 6.

- O fator 5 inicialmente estimado com valor negativo decresce em valor absoluto e rapidamente se torna positivo. Isto é coerente com o fato do fator 5 ter estimativa inicial positiva e ter correlação amostral 0,84 com o fator 6 e portanto, se espera que tenham os comportamentos semelhantes, uma vez que correlação amostral alta sugere que os dois fatores teriam o mesmo efeito.
- As correlações com os outros fatores podem ter acarretado subestimação do fator 1.
- O fator 7 está superestimado.
- O sistema estabiliza para valores de  $k$  no intervalo  $[0, 2; 0, 3]$  sugerindo então que o valor real do parâmetro deve se encontrar nessa faixa.

A partir das curvas de cumeeira Figura (33) do estimador proposto  $\hat{\beta}_{prop}(k) = (X'X + k(X'X)^{-1})^{-1} X'Y$  conclui-se que o comportamento analítico das curvas de cumeeira são essencialmente o mesmo. O fenômeno que ocorre, como era esperado, é uma maior sensibilidade das curvas em relação à  $k$ . A estabilidade ocorre para valores bem menores de  $k$  situando-se no intervalo  $[0, 05; 0, 15]$ .

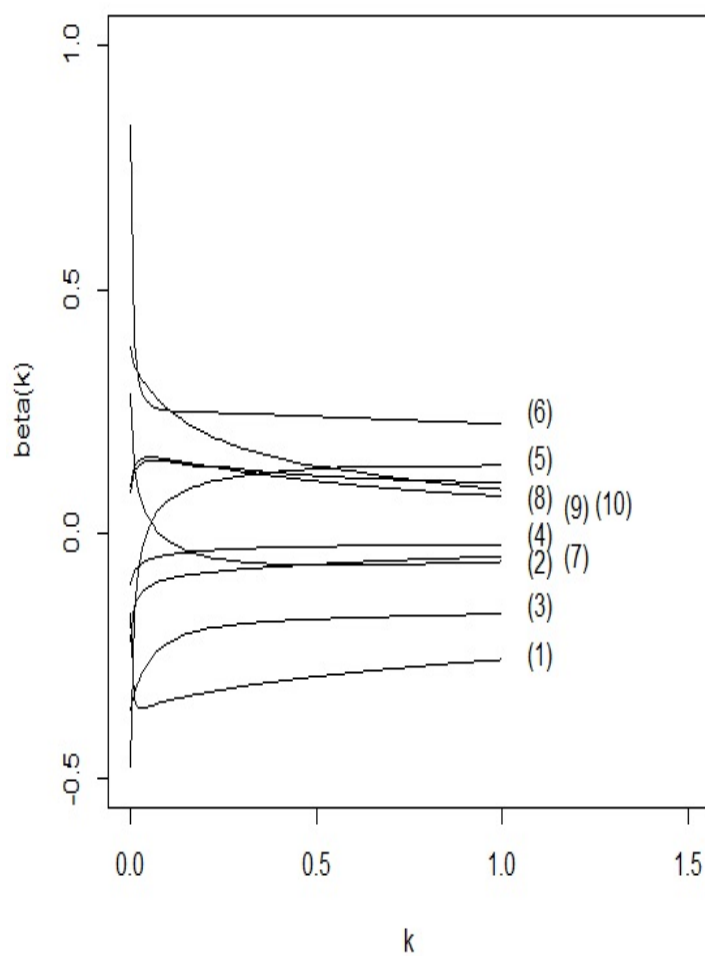


Figura 33 Curvas traço de cumeeira do estimador proposto

Para efeito de comparação foi realizada também a análise utilizando o estimador de Mayer & Wilke (Figura 34)

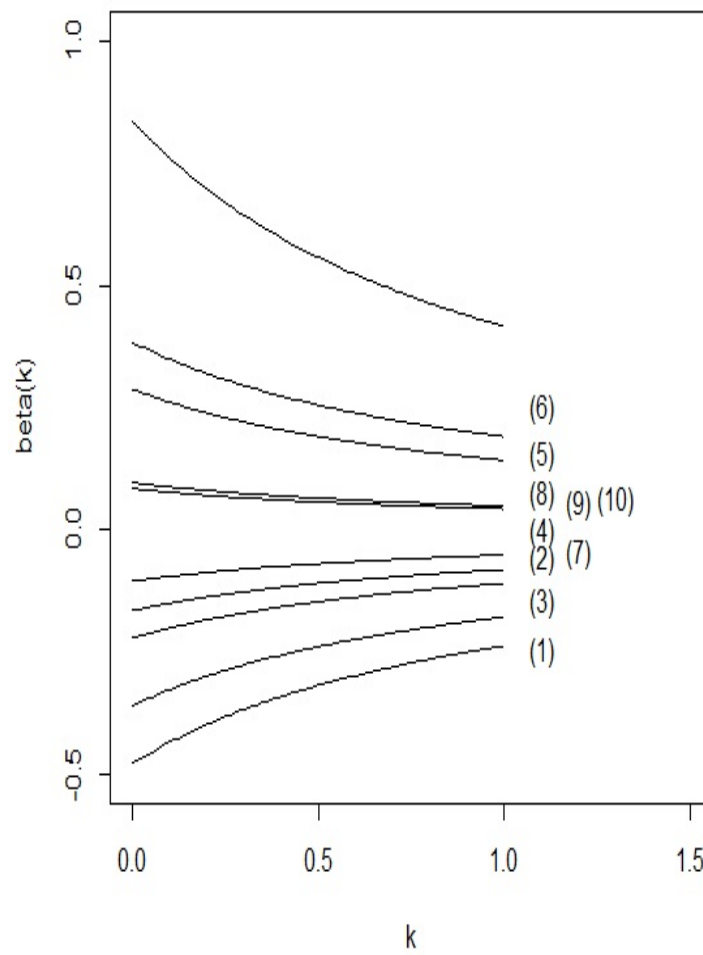


Figura 34 Curvas traço de cumeeira do estimador de Mayer & Wilke

Como era também de se esperar, a taxa de decrescimento das curvas de cumeeira neste caso são iguais dependendo de  $\frac{1}{1+k}$ . Em particular não é possível, a partir deste estimador, considerar que houve subestimação no estimador

de quadrados mínimos para algum fator, como se pode observar pelo fator 1, considerado subestimado com a utilização dos estimadores de Hoerl & Kennard e do estimador proposto. Para o estimador de Mayer & Wilke de fato não se pode falar em uma faixa de estabilidade, uma vez que a taxa de decrescimento é crescente. Isto torna difícil uma obtenção gráfica para um valor ótimo de  $k$ . Para os dados em questão poderia ser afirmado que uma certa estabilidade ocorre para  $k$  no intervalo  $[4, 0; 6, 0]$ .

### 3.6 Uma aplicação a dados de seleção genômica

Os dados utilizados são provenientes da Granja de Melhoramento de Suínos do Departamento de Zootecnia (DZO) da Universidade Federal de Viçosa (UFV), em Viçosa, Minas Gerais, Brasil. Neste experimento uma população  $F_2$  foi composta de 345 suínos provenientes do cruzamento de dois varões da raça local brasileira Piau, com 18 fêmeas de linhagem desenvolvida na UFV pelo acasalamento de animais de linha comercial (Landrace x Large White x Pietrain).

Os detalhes dos procedimentos utilizados, cuja extração do DNA foi realizada no Laboratório de Biotecnologia Animal do Departamento de Zootecnia da Universidade Federal de Viçosa, podem ser encontrados em Peixoto *et al.* (2006). A genotipagem foi realizada via tecnologia Golden Gate/Vera Code R, no Laboratório de Genética Animal (LGA), Embrapa Recursos Genéticos e Biotecnologia (CENARGEM), Brasília, DF, conforme descrito por Hidalgo *et al.* (2013). Os marcadores *SNPs* utilizados estão distribuídos da seguinte forma nos cromossomos da espécie *Sus scrofa domesticus*: SSC1(56), SSC4(54), SSC7(59), SSC8(31), SSC17(25) e SSCX(12), totalizando assim 237 marcadores.

Os dados fenotípicos (características de qualidade de carne) dos 345 indivíduos foram mensurados após o abate (realizando aproximadamente aos 105 dias de idade dos animais) e, entre estas características, optou-se por aquelas relacionadas com o pH da carne *post-mortem* (pH aos 45 minutos,  $pH_{45}$ , e às 24 horas,  $pH_u$ ). Estas medidas de pH foram realizadas pela inserção de um eletrodo de vidro (DIGIMED, DME-CV1), acoplado a um pHmetro DGIMED DM-20, previamente calibrado, no músculo *Longissimus dorsi*, retirado da região imediatamente posterior à última costela do animal. Em resumo têm-se duas variáveis respostas e 237 covariáveis e 345 dados.

Este conjunto de dados foi analisado utilizando o método PLS (*Partial*



*Least Square*) em Silveira (2014), com o objetivo da identificação dos marcadores *SNP* mais relevantes para cada uma das duas características obtendo os resultados:

Tabela 2 Identificação dos SNPs reportados como sendo os mais relevantes para os fenótipos  $pH_{(45)}$  (45 min após o abate) e  $pH_u$  (24 horas após o abate)

<i>Fenotipo</i>	<i>SNP</i>	<i>chr</i>	<i>pos-Mbp</i>	<i>Efeito</i>
$pH_{45}$	ALGA0026103	4	75,55577	-0,17937
	ALGA0026237	4	80,01721	0,157341
	ALGA0026100	4	75,53339	0,150047
	ALGA0026241	4	80,13745	0,138221
	ALGA0026109	4	75,57379	0,114538
$pH_u$	ALGA0026103	4	75,55577	-0,11968
	ALGA0026237	4	80,01721	0,108745
	ALGA0026100	4	75,57379	0,095448
	ALGA0026241	4	80,13745	0,089195
	ALGA0026109	4	75,53339	0,050324

As curvas de cumeieira para todos os fatores utilizando o estimador de Hoerl & Kennard e o estimador proposto estão plotadas respectivamente nas Figuras 35 até 38, e individualizados para os fatores relevantes obtidos por Silveira (2014) nas Figuras 39 até 42:

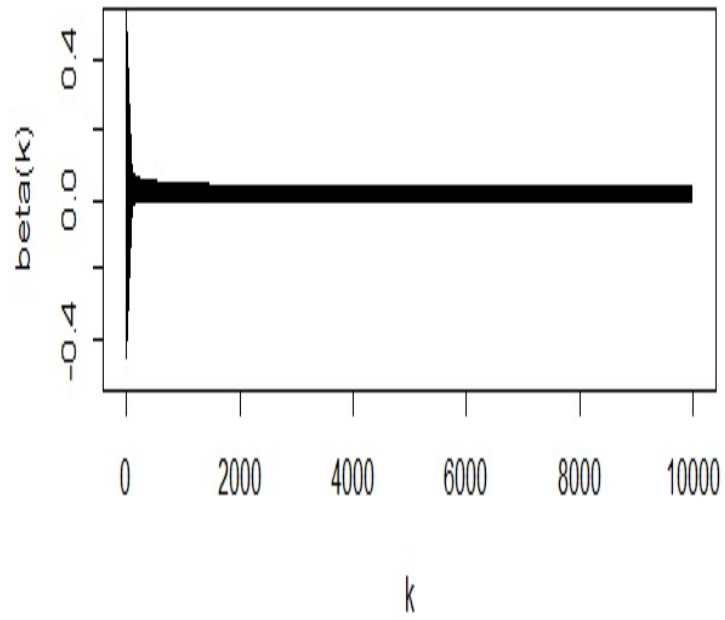


Figura 35 Curvas de cumeira resposta  $pH_{45}$  estimador Hoerl & Kennard

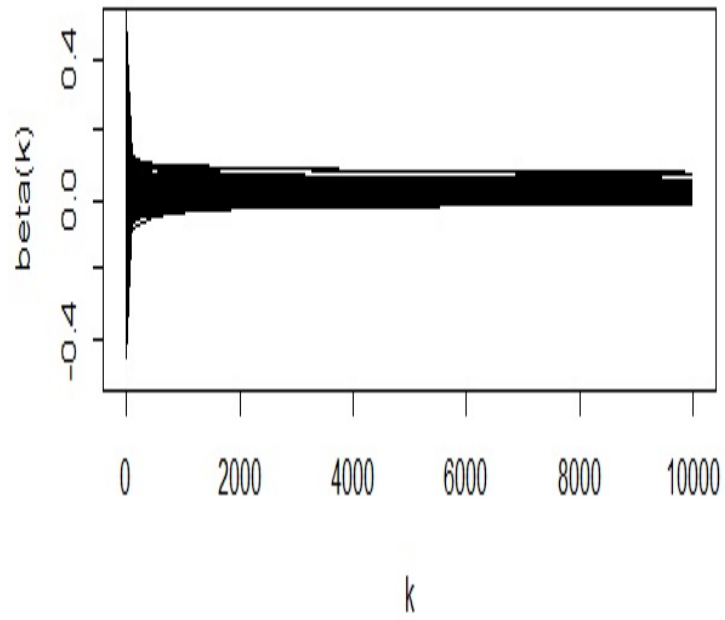


Figura 36 Curvas de cumeira resposta  $pH_{45}$  estimador Proposto

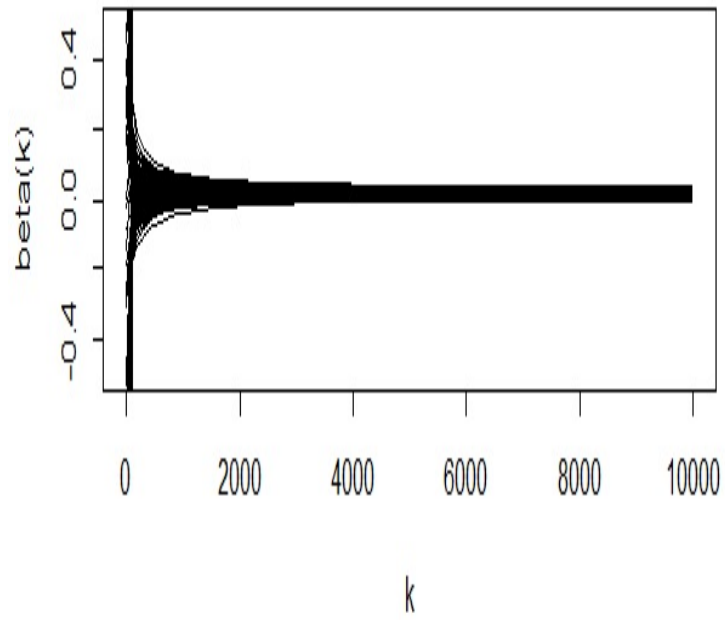


Figura 37 Curvas de cumeeira resposta  $pH_u$  estimador Hoerl & Kennard

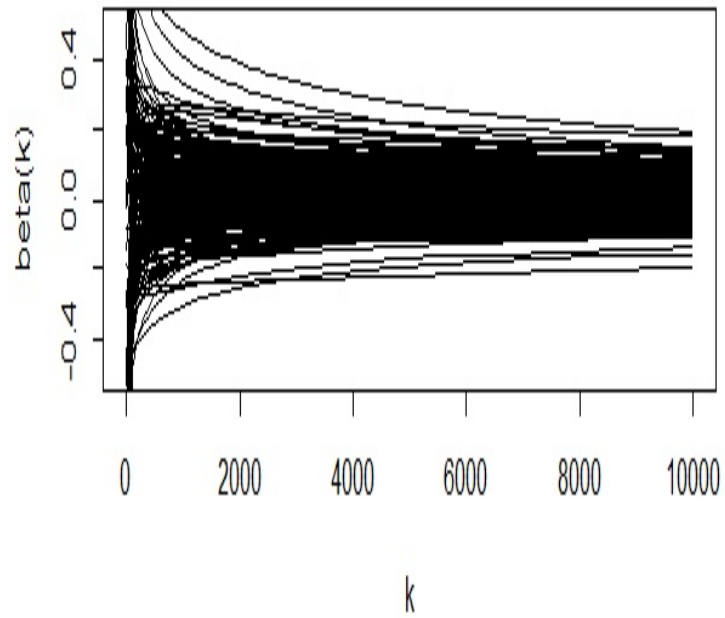


Figura 38 Curvas de cumeira resposta  $pH_u$  estimador Proposto

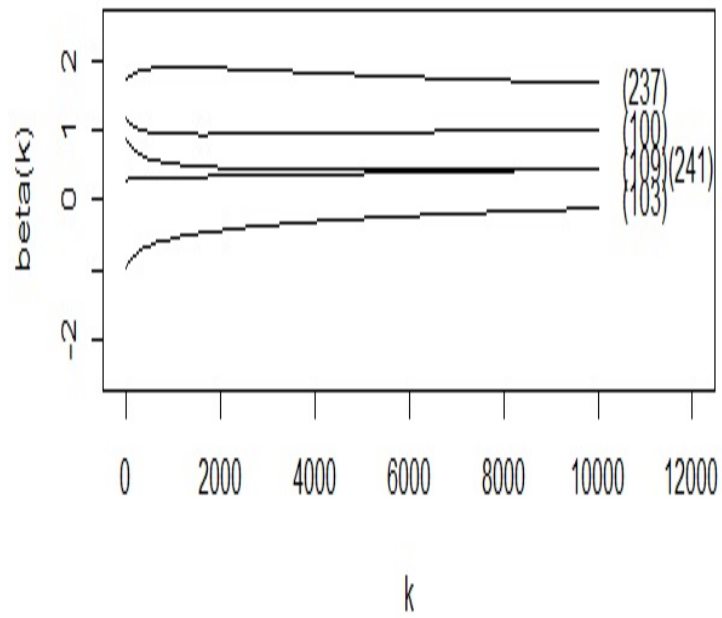


Figura 39 Curvas de cumeira resposta  $pH_u$  fatores relevantes estimador Proposto

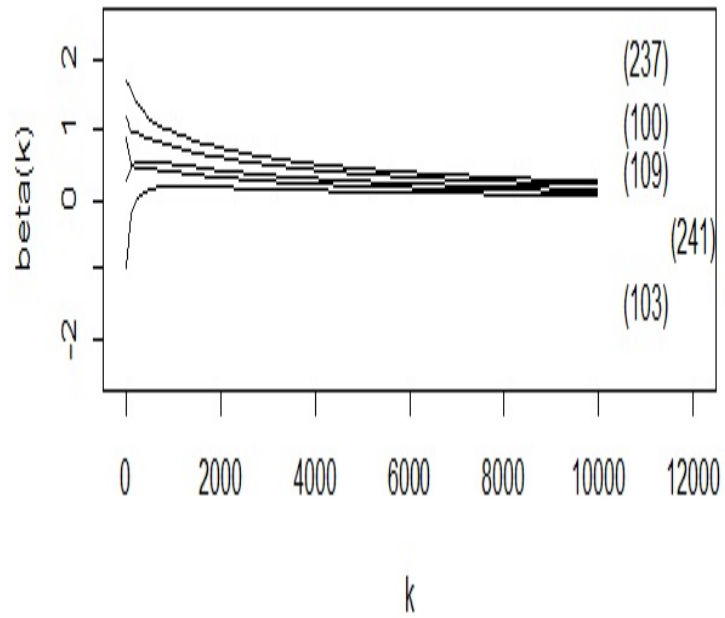


Figura 40 Curvas de cumeieira resposta  $pH_u$  fatores relevantes estimador Hoerl & Kennard



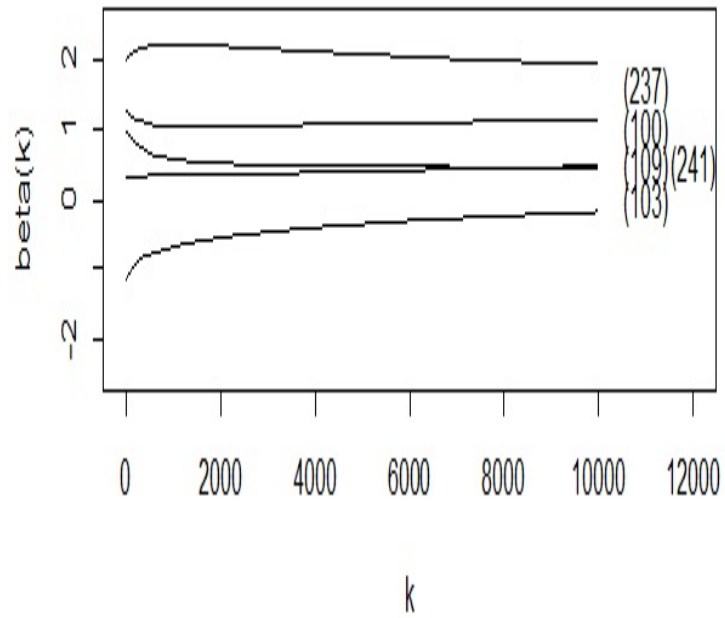


Figura 41 Curvas de cumeieira resposta  $pH_{45}$  fatores relevantes estimador Proposto

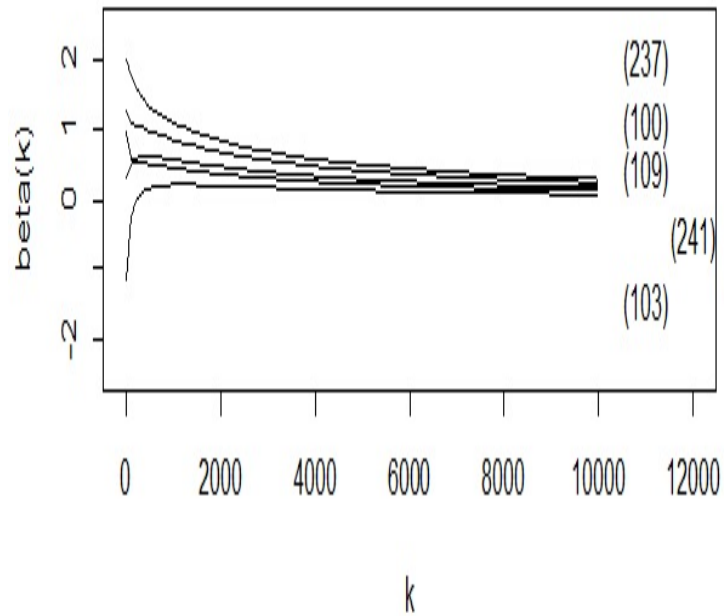


Figura 42 Curvas de cumeeira resposta  $pH_{45}$  fatores relevantes estimador Hoerl & Kennard

Pela análise das curvas de cumeeira tem-se: para as Figuras 35,36,37 e 38:

- Como era esperado pela teoria a faixa de estabilização em que deve ocorrer o valor ótimo para o parâmetro  $k$  é atingida mais rapidamente pelo estimador proposto.
- A sensibilidade para diferenciar os valores das estimativas das covariáveis é maior para o estimador proposto uma vez que para o estimador de Hoerl

& Kennard as curvas de cumeeira decrescem de maneira mais rápida para zero.

Para as Figuras 39,40,41 e 42:

- Para os fatores relevantes as curvas de cumeeira, tanto para o estimador de Hoerl & Kennard quanto o estimador proposto, sugerem que o sinal das estimativas obtidas em Silveira (2014) estão corretos, exceto para o SPN *ALGA0026103* que pelo estimador de Hoerl & Kennard deveria ser positiva para a resposta  $pH_{45}$  e  $pH_u$ . Pelo estimador proposto tem-se evidência que o sinal negativo está correto.
- A obtenção dos fatores mais relevantes utilizando os estimadores de Hoerl & Kennard e o estimador proposto não foi realizada, mas uma primeira análise gráfica indica que possivelmente seriam obtidos outros fatores relevantes diferentes dos obtidos em Silveira (2014), mas com estimativas próximas às obtidas.
- Os valores das estimativas para os fatores relevantes obtidos em Silveira (2014) estão razoavelmente próximos aos valores obtidos pelo estimador proposto com uma tendência de que as estimativas pelo estimador proposto sejam um pouco superiores. Utilizando o estimador de Hoerl & Kennard tem-se que as estimativas são substancialmente menores que as obtidas em Silveira (2014), para a mesma faixa de variação do parâmetro  $k$ .

## 4 CONCLUSÕES

- Uma abordagem essencialmente geométrica à teoria do estimador de cumeira de Hoerl & Kennard de uma maneira mais geral aos estimadores de cumeira de Rao é abrangente e enfatiza uma série de propriedades estatísticas fundamentais para estes estimadores.
- O novo estimador de Rao proposto surge naturalmente de construções geométricas.
- O novo estimador de Rao proposto apresenta propriedades promissoras como suas curvas de cumeira atingirem mais rapidamente uma certa estabilidade, o que facilita o critério gráfico de obtenção do valor ótimo para o parâmetro de cumeira. Estas propriedades justificam um trabalho futuro para estabelecer rigorosamente seu desempenho.
- O desempenho do estimador proposto na análise de dados simulados, dados anteriormente analisados na literatura e dados de seleção genômica foi satisfatório e em geral superior ao desempenho do estimador de Hoerl-Kennard e do estimador de Mayer & Wilke.

## REFERÊNCIAS

COSTA, L. A.; CHAVES, L. M.; SOUZA, D. J. Uma abordagem geométrica do estimador de cumeieira de C.R. Rao. **Revista Brasileira de Biometria**, São Paulo, v. 32, n. 1, p. 28-41, 2014.

CULE, E; LORIO, M. A semi-automatic method to guide the choice of ridge parameter in ridge regression. **Annals of Applied Statistics**, Cleveland, v. 37, n. 7, p. 704-714, 2012.

CULE, E.; LORIO, M. Ridge regression in prediction problems: automatic choice of the ridge parameter. **Genetic Epidemiology**, New York, v. 37, n. 7, p. 704-714, 2013.

FERREIRA, D. F. **Estatística multivariada**. Lavras: Ed. UFLA, 2008. 662 p.

GORMAN, J. W.; TOMAN, R. J. Selection of variables for fitting equations to data. **Technometrics**, Washington, v. 8, n. 1, p. 27-51, 1966.

GRUBER, M. H. J. **Improving efficiency by shrinkage: the James-Stein and ridge regression estimators**. 2nd ed. New York: M. Dekker, 1998. 632 p.

HIDALGO, A. M. et al., Fine mapping and single nucleotide polymorphism effects estimation on pig chromosomes 1,4,7,8,17 and X. **Genetics and Molecular Biology**, Ribeirão Preto, v. 36, n. 4, p. 511-519, dec. 2013.

HOERL, A. E.; KENNARD, R. W. Ridge regression: applications to nonorthogonal problems. **Technometrics**, Washington, v. 12, n. 1, p. 69-82, 1970a.

HOERL, A. E.; KENNARD, R. W. Ridge regression: biased estimation for nonorthogonal problems. **Technometrics**, Washington, v. 12, n. 1, p. 55-67, 1970b.

HOERL, A. E.; KENNARD, R. W. Ridge regression: iterative estimation of the biasing parameter. **Communications in Statistics - Theory and Methods**, New York, v. 5, n. 1, p. 77-88, 1976.

LAWLESS, J. F.; WANG, P. A simulation study of ridge and other regression estimators. **Communications in Statistics - Theory and Methods**, New York, v. 5, n. 4, p. 307-323, 1976.

MARQUARDT, D. W.; SNEE, R. D. Ridge regression in practice. **The American Statistician Journal**, Chicago, v. 29, n. 1, p. 3-20, 1975.

PEIXOTO, J. O. *et al.* Associations of leptin gene polymorphisms with production traits in pigs. **Journal of Animal Breeding and Genetics**, [s.l.], v.123, n. 6, p. 378-383, dez. 2006.

RAO, C. R.; TOUTENBURG, H. **Linear models, least squares and alternatives**. 2nd ed. New York: Springer-Verlag, 1999. 427 p. (Springer Series in Statistics).

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. Vienna: R Foundation for Statistical Computing, 2013.

RENCHER, C. A.; SCHAALJE, B. G. **Linear models in statistics**. 2nd ed. New Jersey: J. Wiley Sons, 2008. 672 p.

SILVEIRA, F. G. **Abordagem geométrica do método dos quadrados mínimos parciais com uma aplicação a dados de seleção genômica**. 2014. 177 f. Tese (Doutorado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, 2014.

THEOBALD, C. M. Generalizations of mean square error applied to ridge regression. **Journal of the Royal Statistical Society. Series B (Methodological)**, London, v. 36, n. 1, p. 103-106, 1974.