FERNANDA GOMES DA SILVEIRA

ABORDAGEM GEOMÉTRICA DO MÉTODO DOS QUADRADOS MÍNIMOS PARCIAIS COM UMA APLICAÇÃO A DADOS DE SELEÇÃO GENÔMICA

LAVRAS-MG

2014

FERNANDA GOMES DA SILVEIRA

ABORDAGEM GEOMÉTRICA DO MÉTODO DOS QUADRADOS MÍNIMOS PARCIAIS COM UMA APLICAÇÃO A DADOS DE 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 Doutor.

Orientador
Dr. Lucas Monteiro Chaves

Coorientador
Dr. Fabyano Fonseca e Silva

LAVRAS-MG
2014

Ficha Catalográfica Elaborada pela Coordenadoria de Produtos e Serviços da Biblioteca Universitária da UFLA

Silveira, Fernanda Gomes da.
Abordagem geométrica do método dos quadrados mínimos parciais com uma aplicação a dados de seleção genômica / Fernanda Gomes da Silveira. – Lavras : UFLA, 2014.
176 p. : il.

Tese (Doutorado) - Universidade Federal de Lavras, 2014.
Orientador: Lucas Monteiro Chaves.
Bibliografia.

1. Projeções. 2. Regressão. 3. Componentes principais. 4. Suíno. I. Universidade Federal de Lavras. II. Título.
CDD – 519.536

FERNANDA GOMES DA SILVEIRA

ABORDAGEM GEOMÉTRICA DO MÉTODO DOS QUADRADOS MÍNIMOS PARCIAIS COM UMA APLICAÇÃO A DADOS DE 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 Doutor.

APROVADA em 27 de janeiro de 2014.

Dr. Daniel Furtado Ferreira	UFLA
Dr. Márcio Balestre	UFLA
Dr. André Luís da Costa Paiva	IFMG - Campus Bambuí
Dr. Antonio Policarpo Souza Carneiro	UFV
Dr. Fabyano Fonseca e Silva	UFV

Dr. Lucas Monteiro Chaves
Orientador

LAVRAS-MG
2014

Aos meus pais, Geraldo e Maria de Lurdes,

aos meus irmãos Fábio e Renata,

pelo amor incondicional.

DEDICO

AGRADECIMENTOS

A Deus, pelo dom da vida, por estar sempre me abençoando, me guiando e dando forças para vencer os obstáculos.

Aos meus pais, Geraldo e Maria de Lurdes pela oportunidade que me proporcionaram nos estudos, pelo apoio e pela compreensão nos momentos em que não pude estar presente e por serem sempre meu esteio e minha inspiração.

Aos meus irmãos, Fábio e Renata; aos cunhados, Renata e Norton; aos sobrinhos Lucas e Bernardo, pelo incentivo, força e carinho.

A toda minha querida família pelos constantes incentivos: avós, tios e tias, primos e primas.

À Universidade Federal de Lavras, pela oportunidade de aprimoramento acadêmico.

À Fundação de Amparo à Pesquisa do Estado de Minas Gerais (FAPEMIG) e à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES), pelo apoio financeiro.

Ao Instituto Federal de Minas Gerais, campus Bambuí, pelo afastamento que tornou possível minha qualificação.

Ao professor Lucas Monteiro Chaves pela orientação, apoio e dedicação nestes quatro anos.

Ao professor Fabyano Fonseca e Silva pela co-orientação, amizade, confiança e por estar sempre bem disposto a ajudar.

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, em especial a Josi, secretária da Pós-graduação.

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.

À amizade das companheiras de repúblicas: Adriana, Alcilene, Isabelle e Larissa.

À hospedagem e o carinho com que me acolheram em suas casas para que as viagens se tornassem menos exaustivas: Alice e João Paulo, Fátima e Diego, Élida e Edna, Luzia e Fábio, Alcilene (Elza, Júlia e Tatiane).

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

“Quanto maior a dificuldade, tanto maior é o mérito em superá-la.”

Henry Ward Beecher

RESUMO

Quando, em uma regressão múltipla, tem-se relações lineares ou quase colinearidade entre as covariáveis ou, ainda, quando o número de covariáveis é maior que o número de observações, o método dos Quadrados Mínimos Ordinários pode não ser adequado. Neste contexto, o método dos Quadrados Mínimos Parciais (PLS) tem se mostrado eficiente. O método consiste em obter a redução de dimensionalidade, uma vez que a regressão é realizada em relação a componentes relevantes. O método é abordado na literatura principalmente sob dois aspectos: algorítmico e algébrico. Neste trabalho, uma abordagem geométrica, utilizando projeções ortogonais, é apresentada no sentido de explicitar todas as etapas teóricas e da construção do algoritmo PLS. No intuito de tornar o texto mais didático, a abordagem geométrica também foi aplicada na teoria da Regressão em Componentes Principais (PCR), uma vez que os métodos PLS e PCR são similares. Uma rotina foi desenvolvida usando o software R, visando também a explicitar o passo a passo da construção do algoritmo. Como em qualquer análise de redução de dimensionalidade, um passo importante na aplicação do método PLS é a determinação de um número ótimo de componentes. Para tal, foi apresentada a teoria de Graus de Liberdade e o método de Validação Cruzada. Os métodos PCR e PLS, além da regressão tradicional sem redução de dimensionalidade, foram aplicados em uma análise de seleção genômica em suínos considerando um painel de marcadores SNPs de baixa densidade e dois fenótipos relacionados com a qualidade da carne.

Palavras-chave: Abordagem Geométrica. Quadrados Mínimos Parciais. Regressão. Seleção Genômica Ampla.

ABSTRACT

When estimating the coefficients of a multiple regression model, if the vectors of predictors are highly correlated, meaning that one can be (almost) a linear combination of the others, or if the number of predictors is greater than the number of observations, the Ordinary Least Square method may be non-appropriate. In that case, the Partial Least Square (PLS) method has shown to be efficient. It consists of obtaining a reduction in dimension by restricting the regression to relevant components. It is usual the literature to be restricted to two main aspects: algorithmic and algebraic. In this work, a geometric approach, based in orthogonal projections, is used to explicit all the theory behind the PLS method as well as in the construction of the PLS algorithm. Aiming to make the text more didactic, the same approach is applied to the Principal Components Regression (PCR) method, since both PLS and PCR are similar. Also a step by step routine for the PLS algorithm was developed using the R software. As in any procedure of reduction of dimensionality, the determination of the optimal number of components is a key step. To do that, we have described and used the Degree of Freedom Method and the Cross Validation Method. Both the PLS and PCR methods, besides the usual regression with no dimensionality reduction, were applied to a genomic selection analysis of pigs, considering a panel of low density SNP markers and two phenotypes related to meat quality.

Keywords: Geometric Approach. Partial Least Squares. Regression. Genome Wide Selection.

LISTA DE FIGURAS

LISTA DE TABELAS

SUMÁRIO

INTRODUÇÃO GERAL

O método mais utilizado para obter equações de predição é o método dos quadrados mínimos. Contudo, quando existem fortes relações lineares entre as covariáveis, ou quando o número de covariáveis é maior que o número de observações em uma regressão múltipla, tem-se um problema denominado multicolinearidade. Nesta situação, as estimativas dos coeficientes de regressão por quadrados mínimos, tendem a ser instáveis, geralmente com grandes erros-padrão, podendo resultar em inferências errôneas (GARTHWAITE, 1994).

Uma forma de contornar o problema da multicolinearidade é a aplicação de métodos de redução de dimensão nas covariáveis. Dentre estes métodos destaca-se o método dos Quadrados Mínimos Parciais (*Partial Least Squares* - PLS). Este método foi introduzido por Wold em 1975, sendo considerado útil para a construção de equações de predições em situações nas quais se tem um grande número de variáveis explicativas e um número relativamente pequeno de dados amostrais (HOSKULDSSON, 1988). Resumidamente, a ideia geral do PLS é formar componentes, isto é, combinações lineares, que capturem a maior quantidade de informação possível disposta nas variáveis explicativas $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{p}$ para predizer as variáveis respostas $\mathrm{Y}_1, \ldots, \mathrm{Y}_\mathrm{k}$. O método PLS apresenta similaridades com o tradicional método de Regressão via Componentes Principais (*Principal Components Regression* - PCR). A maior diferença é dada pelo fato do PCR levar em consideração, na construção dos componentes, apenas as variáveis explicativas, enquanto que o PLS também leva em consideração as variáveis respostas (GARTHWAITE, 1994).

Muitos estudos são encontrados na literatura com aplicações utilizando o método PLS, mas são raras as referências em que se aborda a teoria do PLS,

em especial a abordagem geométrica. Assim, trabalhos contemplando esta teoria geométrica são úteis para o entendimento e desenvolvimento do método.

Como em qualquer análise de redução de dimensionalidade, um passo importante na execução do PLS é a determinação do número ótimo de componentes. No entanto, isso tem sido uma lacuna na aplicação dessa teoria. Metodologias como a teoria de graus de liberdade (*Degrees of Freedom* - DoF) e validação cruzada (*Cross Validation* - CV) foram propostas (KRÄMER; SUGIYAMA, 2011), porém, até o momento, não há relatos da comparação de tais metodologias na aplicação de dados.

A título de ilustração, considere a grande contribuição da genética molecular no melhoramento animal. A utilização direta das informações de DNA no processo de identificação de animais geneticamente superiores, denominado Seleção Genômica Ampla (*Genome Wide Selection* - GWS) consiste da análise de um grande número de marcadores de Polimorfismo de Nucleotídeo Único (*Single Nucleotide Polymorphisms* - SNP) amplamente distribuídos no genoma (MEUWISSEN; HAYES; GODDARD, 2001). Porém, a utilização dessas informações é um desafio, uma vez que, geralmente, o número de marcadores é muito maior que o número de animais genotipados (alta dimensionalidade) e tais marcadores são altamente correlacionados (multicolinearidade). Assim, o sucesso da seleção genômica ampla deve-se à escolha de metodologias que contornem essas adversidades. Dentre estas, o PLS apresenta-se como uma alternativa interessante que merece ser investigada.

Diante do exposto, objetivou-se, no capítulo 1, apresentar uma abordagem geométrica à teoria do PLS (populacional e amostral) e do PCR. E no capítulo 2, objetivou-se aplicar os métodos PCR e PLS, além da regressão tradicional sem redução de dimensionalidade, em uma análise de seleção genômica em suínos

considerando um painel de marcadores SNPs de baixa densidade e dois fenótipos relacionados com a qualidade da carne (pH medido aos 45 min e às 24 horas após o abate). Além disso, objetivou-se, ainda, testar dois diferentes métodos de determinação do número ótimo de componentes na análise PLS, a teoria de graus de liberdade e a validação cruzada, bem como sua influência na performance preditiva do método.

CAPÍTULO 1

Abordagem Geométrica da Regressão por Quadrados Mínimos Parciais

RESUMO

A abordagem geométrica do método dos Quadrados Mínimos Parciais (PLS) é natural e intuitiva. Esta abordagem explicita as relações existentes entre os métodos de regressão PLS, Quadrados Mínimos Ordinários (OLS) e Componentes Principais (PCR). Em termos de projeções ortogonais, são explicadas, neste capítulo, as etapas da construção dos algoritmos PLS, populacional e amostral, bem como as construções relativas à regressão em componentes principais. Uma rotina foi desenvolvida no software R e aplicada a um exemplo didático, explicitando o passo a passo do algoritmo PLS. A teoria dos Graus de Liberdade é apresentada como uma alternativa ao método da Validação Cruzada, uma vez que esta vem sendo proposta para determinar o número de componentes a ser utilizado na regressão.

Palavras-chave: Componentes Principais. Projeções. Redução de dimensionalidade. Vetores Aleatórios.

ABSTRACT

The geometric approach to the Partial Least Square (PLS) Method is natural and intuitive. It makes clear the similarities among the PLS, the Ordinary Least Square (OLS) and the Principal Components (PCR) regression methods. In this chapter we use the orthogonal projections to explain the step by step construction of the PLS algorithm, for sample and for population, as well as the PCR. A routine was developed in the software R and applied in a didactic example. The Degrees of Freedom theory is presented as an alternative to the Cross Validation method, since it has been proposed to choose the number of components to be used in the regression.

Keywords: Principal Components. Projections. Dimensionality Reduction. Random Vectors.

1 INTRODUÇÃO

A álgebra linear é talvez a área da matemática mais acessível. Uma razão que pode justificar tal fato é a sua natureza dupla, isto é, álgebra, com toda a sua abstração e elegância, e geometria, com todo o seu apelo intuitivo. Vetores, ângulos, subespaços vetoriais são conceitos geométricos e de fácil visualização. A álgebra é, também, a área da matemática com maior aplicabilidade em estatística, como modelos lineares e estatística multivariada. O uso de transformações lineares ortogonais, diagonalização de matrizes simétricas e diagonalização de formas quadráticas, são essenciais e de uso constante. Esses resultados de álgebra linear são encontrados em praticamente todos os livros utilizados nos cursos de graduação em matemática, engenharias e estatística.

A teoria da regressão é uma área ampla e muito importante da estatística, e, apesar de existirem muitas referências sobre o assunto, pouco se encontra na literatura sobre sua abordagem geométrica. O fato interessante a ser observado é que a abordagem feita pelos autores estatísticos, em geral, é totalmente algébrica, extremamente analítica e abstrata. Desse modo, a utilização de uma abordagem geométrica poderá dar uma contribuição ao entendimento de vários conceitos na teoria da regressão.

Dentre os métodos de regressão, a regressão via componentes principais e a regressão via método dos quadrados mínimos parciais, têm destaque pela sua aplicabilidade, em especial esta última, que vem sendo muito utilizada em áreas como quimiometria e genética. No entanto, apesar de ser recorrente o uso do método PLS em trabalhos encontrados na literatura, muito se discute sobre sua aplicação, mas muito pouco se encontra sobre sua teoria.

O objetivo deste capítulo é apresentar a teoria da regressão em compo-

nentes principais e o método dos quadrados mínimos parciais, sob o ponto de vista geométrico, e a teoria dos Graus de Liberdade e Validação Cruzada, que vêm sendo propostas para determinar o número de componentes, tomando como referências os artigos: Helland (1990), Hoskuldsson (1988), Garthwaite (1994), Phatak e Jong (1997) e Krämer e Sugiyama (2011).

2 REFERENCIAL 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, $\mathbf{Y} = (\mathrm{Y}_1, \ldots, \mathrm{Y}_\mathrm{n})$, em que cada coordenada é uma variável aleatória unidimensional. O exemplo mais usual é quando se tem uma amostra aleatória simples $\mathbf{Y} = (\mathrm{Y}_1, \ldots, \mathrm{Y}_\mathrm{n})$. Neste caso $\mathbf{Y}$ representa medidas repetidas independentes de uma mesma grandeza populacional. Outra situação é quando se tem uma variável multidimensional propriamente dita, isto é, a medida 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 característica ou valores obtidos por repetição.

Os valores que $\mathbf{Y} = (\mathrm{Y}_1, ..., \mathrm{Y}_\mathrm{n})$ assume podem ser vistos como vetores no espaço $\mathbb{R}^\mathrm{n}$, e, portanto, podem também ser apropriadamente denominados de vetores aleatórios. Neste trabalho, as variáveis aleatórias multidimensionais serão vistas como vetores aleatórios e estudam-se suas projeções em subespaços vetoriais, utilizando-se uma abordagem inteiramente geométrica.

Quando uma base do espaço vetorial é escolhida será usada a notação $\mathbb{R}^\mathrm{n}$. Caso o espaço vetorial seja abstrato, isto é, sem uma escolha particular de base, será utilizada a notação V. Uma vez fixadas bases no domínio e no contradomínio, uma transformação linear pode ser representada por uma matriz. Dessa forma, não será feita distinção entre os conceitos de transformações lineares e de matrizes.

Uma situação usual em estatística é a transformação linear de dados, como mudanças de escalas, por exemplo, ou o procedimento de se trabalhar com a mé-

dia amostral. Tais procedimentos, em termos de álgebra linear consistem em aplicar transformações lineares, fazer projeções ortogonais, calcular ângulos, etc. Da mesma forma, todos estes procedimentos podem ser aplicados em vetores aleatórios.

Considere transformações lineares: $\mathrm{L} : \mathbb{R}^{\mathrm{n}} \rightarrow \mathbb{R}^{\mathrm{m}}$, conforme na Figura 1.

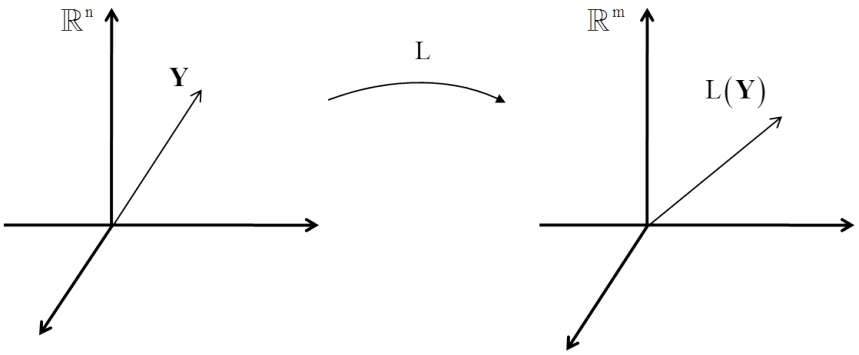


Figura 1 Transformação linear de $\mathbb{R}^{\mathrm{n}}$ em $\mathbb{R}^{\mathrm{m}}$

Considere L como uma matriz $\mathrm{L}_{\mathrm{m}\times\mathrm{n}}$ e L' sua transposta. Em termos de produto de matrizes, tem-se que a transformação linear pode ser dada por:

$$\mathrm{L}_{\mathrm{m}\times\mathrm{n}}\mathbf{Y}_{\mathrm{n}\times 1} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \dots & \mathrm{x}_{1\mathrm{n}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \dots & \mathrm{x}_{2\mathrm{n}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{\mathrm{m}1} & \mathrm{x}_{\mathrm{m}2} & \dots & \mathrm{x}_{\mathrm{mn}} \end{bmatrix} \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \vdots \\ \mathrm{Y}_{\mathrm{n}} \end{bmatrix}.$$

O produto interno do vetor aleatório $\mathbf{Y}$ com um vetor $\mathbf{x}' = (\mathrm{x}_1, \mathrm{x}_2, ..., \mathrm{x}_{\mathrm{n}})$ é dado por $\mathbf{x} \cdot \mathbf{Y} = \mathrm{x}_1\mathrm{Y}_1 + ... + \mathrm{x}_{\mathrm{n}}\mathrm{Y}_{\mathrm{n}}$. Em termos de matrizes, se $\mathbf{x}$ é considerado

uma matriz $\mathrm{n} \times 1$, ou seja, $\mathbf{x} = \begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_\mathrm{n} \end{bmatrix}$, então $\mathbf{x} \cdot \mathbf{Y}$ pode ser expresso como o produto de matrizes

$$\mathbf{x}'\mathbf{Y} = \begin{bmatrix} \mathrm{x}_1 & \mathrm{x}_2 & \dots & \mathrm{x}_\mathrm{n} \end{bmatrix} \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \vdots \\ \mathrm{Y}_\mathrm{n} \end{bmatrix}.$$

As seguintes notações serão usadas indistintamente: $\mathbf{x} \cdot \mathbf{Y} = \mathbf{x}'\mathbf{Y} =< \mathbf{x}, \mathbf{Y} >$.

O produto interno é dado por $\mathbf{x} \cdot \mathbf{Y} = \|\mathbf{x}\| \, \|\mathbf{Y}\| \cos\theta$. Geometricamente, $\mathbf{x} \cdot \mathbf{Y}$ significa o tamanho do vetor $\mathbf{x}$ vezes o tamanho do vetor aleatório obtido pela projeção do vetor **Y** no subespaço unidimensional gerado pelo vetor **x** (Figura 2).

$$\cos\theta = \frac{\|P_\mathbf{x}\mathbf{Y}\|}{\|\mathbf{Y}\|} \Rightarrow \|P_\mathbf{x}\mathbf{Y}\| = \|\mathbf{Y}\| \cos\theta,$$

$$\mathbf{x} \cdot \mathbf{Y} = \|\mathbf{x}\| \, \|\mathbf{Y}\| \cos\theta = \|\mathbf{x}\| \, \|P_\mathbf{x}\mathbf{Y}\|.$$

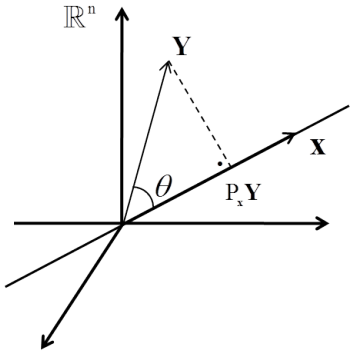


Figura 2 Interpretação geométrica do produto interno

A esperança de um vetor aleatório é o vetor dado pela esperança de cada componente, denominado $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\beta}$, em que $\boldsymbol{\beta}$ é um vetor $\mathrm{p} \times 1$. Outra característica fundamental de um vetor aleatório é sua matriz de variâncias e covariâncias, $\mathrm{cov}\left(\mathbf{Y}\right) = \boldsymbol{\Sigma} = \mathrm{E}\left(\left(\mathbf{Y} - \mathrm{E}\left(\mathbf{Y}\right)\right)\left(\mathbf{Y} - \mathrm{E}\left(\mathbf{Y}\right)\right)'\right)$, em que $\boldsymbol{\Sigma}$ é uma matriz positiva definida $\mathrm{p} \times \mathrm{p}$.

Alguns teoremas importantes envolvendo projetores ortogonais de um vetor aleatório $\mathbf{Y}$ são 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, d-dimensional, de V, denotada por $\mathrm{P_W}\left(\mathbf{Y}\right)$, também é um vetor aleatório que depende, obviamente, de $\mathbf{Y}$ (Figura 3). O vetor $\mathrm{P_W}\left(\mathbf{Y}\right)$ pode ser expresso, em relação a uma base ortogonal $\{\mathbf{u_1}, \ldots, \mathbf{u_d}\}$ de W, por

$$\mathrm{P_W}\mathbf{Y} = \left(\frac{\mathbf{Y} \cdot \mathbf{u_1}}{\mathbf{u_1} \cdot \mathbf{u_1}}\right)\mathbf{u_1} + \left(\frac{\mathbf{Y} \cdot \mathbf{u_2}}{\mathbf{u_2} \cdot \mathbf{u_2}}\right)\mathbf{u_2} + \ldots + \left(\frac{\mathbf{Y} \cdot \mathbf{u_d}}{\mathbf{u_d} \cdot \mathbf{u_d}}\right)\mathbf{u_d}.$$

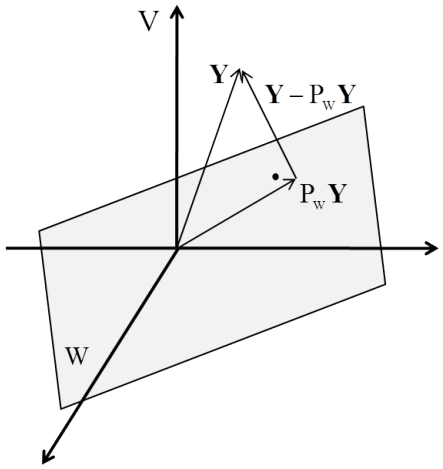


Figura 3 Projetor ortogonal de um vetor aleatório $\mathbf{Y}$ sobre um subespaço W

Um fato importante sobre projetores é sua caracterização como matrizes. Uma matriz quadrada tal que $\mathbf{A}^2 = \mathbf{A}$ é chamada matriz de projeção ou proje-

tor. Tem-se que um projetor $\mathbf{A}$ restrito à Im$\mathbf{A}$ é a identidade, ou seja, $\mathbf{A}\left(\mathbf{Ax}\right) = \mathbf{A}^2\mathbf{x} = \mathbf{Ax}$, $\forall \mathbf{x} \in \mathbb{R}^{\mathrm{n}}$. $\mathbf{I} - \mathbf{A}$ também é um projetor, pois $(\mathbf{I} - \mathbf{A})^2 = \mathbf{I} - 2\mathbf{A} + \mathbf{A}^2 = \mathbf{I} - \mathbf{A}$, em que $\mathbf{I}$ é a matriz identidade.

Tem-se que: $\mathrm{Ker}\mathbf{A} = \mathrm{Im}(\mathbf{I} - \mathbf{A})$ e $\mathrm{Im}\mathbf{A} = \mathrm{Ker}(\mathbf{I} - \mathbf{A})$.

Demonstração.

$$\begin{aligned}
\mathbf{A}((\mathrm{I} - \mathbf{A})\mathbf{x}) = \mathbf{A}(\mathbf{x} - \mathbf{Ax}) = \mathbf{Ax} - \mathbf{A}^2\mathbf{x} = \mathbf{Ax} - \mathbf{Ax} = \mathbf{0} \\
\Rightarrow \mathrm{Im}(\mathbf{I} - \mathbf{A}) \subset \mathrm{Ker}\mathbf{A} \quad \mathrm{e} \\
\mathbf{x} \in \mathrm{Ker}\mathbf{A} \Rightarrow (\mathbf{I} - \mathbf{A})\mathbf{x} = \mathbf{x} - A\mathbf{x} = \mathbf{x} - \mathbf{0} \ \Rightarrow \mathbf{x} \in \mathrm{Im}(\mathbf{I} - \mathbf{A}) \\
\Rightarrow \mathrm{Ker}\mathbf{A} \subset \mathrm{Im}(\mathbf{I} - \mathbf{A}),
\end{aligned}$$

logo, $\mathrm{Ker}\mathbf{A} = \mathrm{Im}(\mathbf{I} - \mathbf{A})$.

$$\begin{aligned}
\mathbf{y} \in \mathrm{Ker}\left(\mathbf{I} - \mathbf{A}\right) \Rightarrow \mathbf{0} = \left(\mathbf{I} - \mathbf{A}\right)\mathbf{y} = \mathbf{y} - \mathbf{Ay} \Rightarrow \mathbf{Ay} = \mathbf{y} \Rightarrow \mathbf{y} \in \mathrm{Im}\mathbf{A} \\
\Rightarrow \mathrm{Ker}\left(\mathbf{I} - \mathbf{A}\right) \subset \mathrm{Im}\mathbf{A} \quad \mathrm{e} \\
\mathbf{y} \in \mathrm{Im}\mathbf{A} \Rightarrow \mathbf{y} = \mathbf{Ax} \Rightarrow \left(\mathbf{I} - \mathbf{A}\right)\mathbf{y} = \left(\mathbf{I} - \mathbf{A}\right)\mathbf{Ax} = \mathbf{Ax} - \mathbf{A}^2\mathbf{x} \\
= \mathbf{Ax} - \mathbf{Ax} = \mathbf{0} \Rightarrow \mathbf{y} \in \mathrm{Ker}\left(\mathbf{I} - \mathbf{A}\right) \\
\Rightarrow \mathrm{Im}\mathbf{A} \subset \mathrm{Ker}\left(\mathbf{I} - \mathbf{A}\right),
\end{aligned}$$

logo, $\mathrm{Im}\mathbf{A} = \mathrm{Ker}(\mathbf{I} - \mathbf{A})$. □

Uma matriz de projeção $\mathbf{A}$ é dita um projetor ortogonal se $\mathbf{Av} - \mathbf{v}$ é perpendicular ao subespaço Im$\mathbf{A}$.

Proposição 1. *Uma matriz de projeção* $\mathrm{A}_{\mathrm{n}\times\mathrm{n}}$ *é simétrica se, e somente se, é um projetor ortogonal, isto é,*

$$\langle \mathbf{Av}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{Aw} \rangle \quad \Leftrightarrow \quad \langle \mathbf{Av} - \mathbf{v}, \mathbf{Aw} \rangle = 0, \ \ \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathrm{n}}.$$

Demonstração. ($\Rightarrow$) Se $\mathbf{A}$ é simétrica ($\mathbf{A}' = \mathbf{A}$), então:

$$\begin{aligned} \langle \mathbf{v} - \mathbf{Av}, \mathbf{Aw} \rangle &= \langle \mathbf{v}, \mathbf{Aw} \rangle - \langle \mathbf{Av}, \mathbf{Aw} \rangle = \langle \mathbf{v}, \mathbf{Aw} \rangle - \left\langle \mathbf{v}, \mathbf{A}^2\mathbf{w} \right\rangle \\ &= \langle \mathbf{v}, \mathbf{Aw} \rangle - \langle \mathbf{v}, \mathbf{Aw} \rangle = 0, \ \ \forall \mathbf{v}, \mathbf{w} \in \mathbb{R}^{\mathrm{n}}. \end{aligned}$$

Logo, $\mathbf{A}$ é um projetor ortogonal.

($\Leftarrow$) Se $\mathbf{A}$ é um projetor ortogonal, $\mathbf{Av} - \mathbf{v}$ é perpendicular à imagem de $\mathbf{A}$, isto é, $\langle \mathbf{Av} - \mathbf{v}, \mathbf{Aw} \rangle = 0 \Rightarrow \langle \mathbf{Av}, \mathbf{Aw} \rangle = \langle \mathbf{v}, \mathbf{Aw} \rangle$, da mesma forma $\mathbf{Aw} - \mathbf{w}$ é perpendicular a $\mathbf{Av}$, isto é, $\langle \mathbf{Aw} - \mathbf{w}, \mathbf{Av} \rangle = 0 \Rightarrow \langle \mathbf{Aw}, \mathbf{Av} \rangle = \langle \mathbf{w}, \mathbf{Av} \rangle$, portanto, $\langle \mathbf{Av}, \mathbf{w} \rangle = \langle \mathbf{v}, \mathbf{Aw} \rangle$. Logo, $\mathbf{A}$ é simétrica.

□

Teorema 1. *Se* $\mathbf{Y} \in \mathrm{V}$ *é vetor aleatório n-dimensional com* $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\tau}$, $\mathrm{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}$ *e* W *um subespaço d-dimensional de* V. *Então:*

- $\mathrm{E}(\mathbf{x} \cdot \mathbf{Y}) = \mathbf{x} \cdot \mathrm{E}(\mathbf{Y}) = \mathbf{x} \cdot \boldsymbol{\tau}$
- $\mathrm{var}(\mathbf{x} \cdot \mathbf{Y}) = \mathbf{x}'\boldsymbol{\Sigma}\mathbf{x}$
- $\mathrm{cov}(\mathbf{x} \cdot \mathbf{Y}, \mathbf{z} \cdot \mathbf{Y}) = \mathbf{x}'\boldsymbol{\Sigma}\mathbf{z}$
- *Se* $\mathbf{x}$ *é um autovetor de* $\boldsymbol{\Sigma}$ *relativo ao autovalor* ξ *então* $\mathrm{var}(\mathbf{x} \cdot \mathbf{Y}) = \|\mathbf{x}\|^2 \xi$.
- *Sejam* $\mathbf{x}$, $\mathbf{z}$ *autovetores relativos aos autovalores* ξ *e* η.

Se $\xi = \eta$*, então* $\mathrm{cov}\,(\mathbf{x}\cdot\mathbf{Y}, \mathbf{z}\cdot\mathbf{Y}) = (\mathbf{x}\cdot\mathbf{z})\xi$.

Se $\xi \neq \eta$*, então* $\mathrm{cov}\,(\mathbf{x}\cdot\mathbf{Y}, \mathbf{z}\cdot\mathbf{Y}) = 0$.

- *Se* W *é um subespaço d-dimensional do subespaço dos autovetores relativos ao autovalor* ξ*, tem-se:*

$$\mathrm{E}\,(\mathrm{P_W}\mathbf{Y}) = \mathrm{P_W}\mathrm{E}\,(\mathbf{Y}) = \mathrm{P_W}\boldsymbol{\tau}$$

$$\mathrm{E}\left(\|\mathrm{P_W}\mathbf{Y}\|^2\right) = \|\mathrm{P_W}\boldsymbol{\tau}\|^2 + \mathrm{d}\xi.$$

Para o caso em que $\mathrm{cov}\,(\mathbf{Y}) = \sigma^2\mathbf{I}$*, tem-se:*

- $\mathrm{var}\,(\mathbf{x}\cdot\mathbf{Y}) = \sigma^2\mathbf{x}'\mathbf{x}$
- $\mathrm{cov}\,(\mathbf{x}\cdot\mathbf{Y}, \mathbf{z}\cdot\mathbf{Y}) = \sigma^2\mathbf{x}'\mathbf{z}$
- $\mathrm{E}\left(\|\mathrm{P_W}\mathbf{Y}\|^2\right) = \|\mathrm{P_W}\boldsymbol{\tau}\|^2 + \mathrm{d}\sigma^2$

Corolário 1. *Se* $\mathbf{x}$ *e* $\mathbf{z}$ *são vetores ortogonais então* $\mathbf{x}\cdot\mathbf{Y}$ *e* $\mathbf{z}\cdot\mathbf{Y}$ *são variáveis aleatórias não correlacionadas.*

As demonstrações destes resultados encontram-se em Adão (2011).

2.2 Regressão Linear Múltipla

Considere observações $(\mathrm{Y_1}, \mathrm{Y_2}, \ldots, \mathrm{Y_n})$ não correlacionadas, tais que:

$$\mathrm{E}\,(\mathrm{Y_i}) = \mathrm{x_{i1}}\beta_1 + \mathrm{x_{i2}}\beta_2 + \ldots + \mathrm{x_{ip}}\beta_\mathrm{p} \;\text{ e }\; \mathrm{var}\,(\mathrm{Y_i}) = \sigma^2,$$

para $\mathrm{i} = 1, 2, \ldots, \mathrm{n}$, em que $\beta_1, \beta_2, \ldots, \beta_\mathrm{p}$ e σ^2 são parâmetros desconhecidos.

Tem-se, portanto, o sistema:

$$\begin{aligned}
\mathrm{E}\left(\mathrm{Y}_1\right) &= \mathrm{x}_{11}\beta_1 + \mathrm{x}_{12}\beta_2 + \cdots + \mathrm{x}_{1\mathrm{p}}\beta_{\mathrm{p}} \\
\mathrm{E}\left(\mathrm{Y}_2\right) &= \mathrm{x}_{21}\beta_1 + \mathrm{x}_{22}\beta_2 + \cdots + \mathrm{x}_{2\mathrm{p}}\beta_{\mathrm{p}} \\
&\vdots \\
\mathrm{E}\left(\mathrm{Y}_{\mathrm{n}}\right) &= \mathrm{x}_{\mathrm{n}1}\beta_1 + \mathrm{x}_{\mathrm{n}2}\beta_2 + \cdots + \mathrm{x}_{\mathrm{np}}\beta_{\mathrm{p}}.
\end{aligned}$$

Representando matricialmente este sistema de equações, tem-se:

$$\mathbf{Y} = \begin{bmatrix} \mathrm{Y}_1 \\ \mathrm{Y}_2 \\ \vdots \\ \mathrm{Y}_{\mathrm{n}} \end{bmatrix} \quad \mathrm{e} \quad \boldsymbol{\beta} = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{\mathrm{p}} \end{bmatrix}$$

$$\begin{bmatrix} \mathrm{E}\left(\mathrm{Y}_1\right) \\ \mathrm{E}\left(\mathrm{Y}_2\right) \\ \vdots \\ \mathrm{E}\left(\mathrm{Y}_{\mathrm{n}}\right) \end{bmatrix} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \ldots & \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \ldots & \mathrm{x}_{2\mathrm{p}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{\mathrm{n}1} & \mathrm{x}_{\mathrm{n}2} & \ldots & \mathrm{x}_{\mathrm{np}} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_{\mathrm{p}} \end{bmatrix},$$

isto é, denominando o vetor de médias por $\mathrm{E}\left(\mathbf{Y}\right) = \boldsymbol{\tau}$ o problema de se estimar o vetor $\boldsymbol{\tau}$ é equivalente a se obter o vetor $\boldsymbol{\beta}$ tal que $\boldsymbol{\tau} = \mathbf{X}\boldsymbol{\beta}$, em que $\mathbf{X} = \left(\mathrm{x}_{\mathrm{ij}}\right)$ é uma matriz de coeficientes conhecidos que depende do delineamento subjacente à obtenção dos dados. Como casos extremos, tem-se o caso em que os y_{i} são observações independentes de uma mesma população, ou, utilizando a linguagem da estatística experimental, todos os y_{i} são respostas de um mesmo tratamento, neste caso, $\mathrm{E}\left(\mathbf{Y}\right) = \boldsymbol{\tau}$ é um vetor com coordenadas iguais e a matriz $\mathbf{X}$ é uma matriz $\mathrm{n} \times 1$ com entradas iguais a 1. A outra situação ocorre quando os y_{i} são observações independentes de n populações diferentes ($\mathrm{p} = \mathrm{n}$), ou, utilizando a linguagem

da estatística experimental, os $\mathrm{y_i}$ são respostas relativas a n tratamentos diferentes. Neste caso, $\mathrm{E}(\mathbf{Y}) = \boldsymbol{\tau}$ é um vetor com todas as coordenadas possivelmente diferentes e a matriz $\mathbf{X}$ é uma matriz $\mathrm{n} \times \mathrm{n}$.

É claro que os valores observados no vetor de respostas $\mathbf{Y}$ são, em razão da aleatoriedade que envolve todo experimento, diferentes dos valores esperados $\mathrm{E}(\mathbf{Y})$. Note que os valores esperados não são conhecidos, mas sabe-se, por exemplo, que a repetição de um tratamento deveria ter a mesma resposta e tal fato não acontece em razão dos fatores aleatórios que ocorrem em todo experimento.

Desse modo, pode-se considerar uma variável aleatória $\boldsymbol{\varepsilon}$, tal que

$$\boldsymbol{\varepsilon} = \mathbf{Y} - \mathrm{E}(\mathbf{Y}) \Rightarrow \mathbf{Y} = \mathrm{E}(\mathbf{Y}) + \boldsymbol{\varepsilon}.$$

O vetor $\boldsymbol{\varepsilon}$ é denominado vetor de erros ou vetor de resíduos, uma vez que mede a diferença entre o valor efetivamente observado e o valor desconhecido esperado. Considerando que os componentes do vetor de erros $\boldsymbol{\varepsilon}$ sejam independentes e igualmente distribuídos, tem-se que:

$$\mathrm{E}(\boldsymbol{\varepsilon}) = \mathrm{E}(\mathbf{Y} - \mathrm{E}(\mathbf{Y})) = \mathrm{E}(\mathbf{Y}) - \mathrm{E}(\mathrm{E}(\mathbf{Y})) = \mathrm{E}(\mathbf{Y}) - \mathrm{E}(\mathbf{Y}) = \mathbf{0},$$
$$\mathrm{cov}(\boldsymbol{\varepsilon}) = \mathrm{cov}(\mathbf{Y} - \mathrm{E}(\mathbf{Y})) = \mathrm{cov}(\mathbf{Y}) = \sigma^2 \mathbf{I}_\mathrm{n},$$

em que $\mathrm{cov}(\cdot)$ é a matriz de variâncias e covariâncias e $\mathbf{I}_\mathrm{n}$ é a matriz identidade de ordem n.

O problema que se quer resolver é: como estimar os parâmetros desconhecidos β_j com base nas observações $\mathrm{y_i}$?

A equação $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ é a chamada equação do modelo linear geral. A razão do nome é que o modelo é linear em seus parâmetros β_j. O modelo linear geral é a equação do modelo, sob as condições $\mathrm{E}(\boldsymbol{\varepsilon}) = \mathbf{0}$ e $\mathrm{cov}(\boldsymbol{\varepsilon}) = \sigma^2 \mathbf{I}_\mathrm{n}$.

Geometricamente, o modelo linear de Gauss-Markov pode ser descrito como na Figura 4.

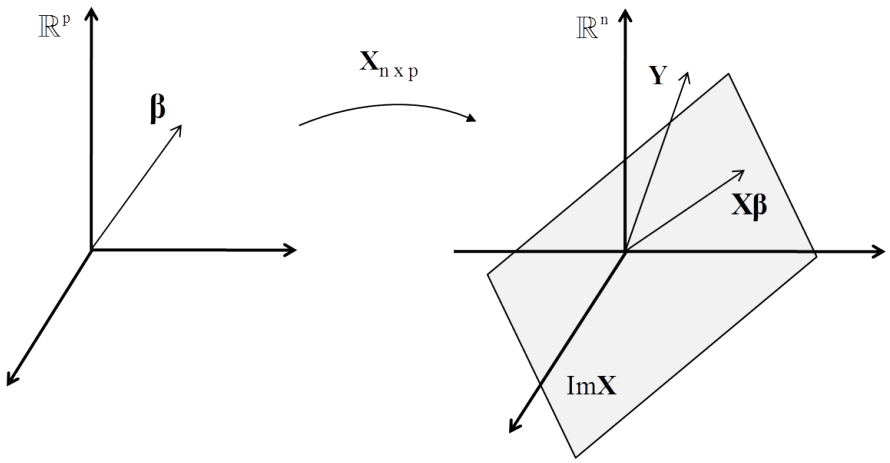


Figura 4 Representação geométrica para o modelo linear de Gauss-Markov

Seja $\mathbb{R}^n$ o espaço dos dados, $\mathbb{R}^p$ o espaço dos parâmetros, e $\mathrm{Im}\mathbf{X} = \{\mathbf{v} = \mathbf{X}\boldsymbol{\beta},\ \ \boldsymbol{\beta} \in \mathbb{R}^p\}$ o subespaço formado pelos vetores que são imagem da transformação linear **X**, isto é, o subespaço vetorial gerado pelas colunas da matriz **X**. O subespaço dos vetores que são levados no vetor nulo por **X**, é denominado kernel da transformação linear **X**, Ker**X**. Em geral, n é maior que p, pois o número de dados geralmente é maior que o número de parâmetros.

2.2.1 As equações normais e os estimadores de quadrados mínimos

No modelo linear geral: $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$ ou $\boldsymbol{\varepsilon} = \mathbf{Y} - \mathbf{X}\boldsymbol{\beta}$, o estimador de quadrados mínimos de $\boldsymbol{\beta}$ é o vetor $\hat{\boldsymbol{\beta}}$ que minimiza a soma de quadrados dos erros, isto é, o valor do vetor de parâmetros que faz a norma do vetor $\boldsymbol{\varepsilon}$, $\|\boldsymbol{\varepsilon}\|$, ser a menor possível, ou de forma equivalente, que faz a norma ao quadrado, $\|\boldsymbol{\varepsilon}\|^2$, a menor possível.

Tal problema pode ser resolvido por considerações geométricas de forma bastante simples, sem o uso de cálculo diferencial. A solução é o vetor obtido pela diferença entre o vetor observado $\mathbf{Y}$ e a projeção deste vetor no subespaço Im$\mathbf{X}$ uma vez que a projeção ortogonal minimiza a distância de um ponto a um subespaço vetorial (Figura 5).

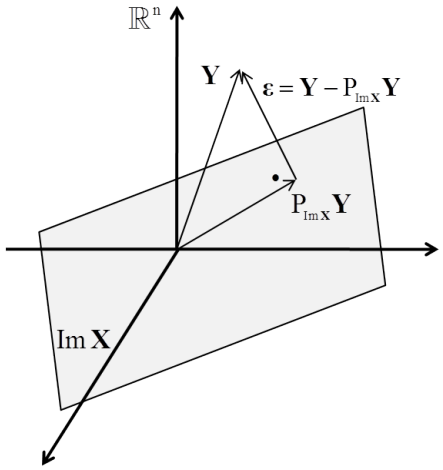


Figura 5 Representação geométrica do vetor residual

O problema, então, torna-se: como obter o projetor ortogonal no subespaço Im$\mathbf{X}$. Projetar ortogonalmente em um subespaço é equivalente a obter o subespaço complementar ortogonal deste subespaço. Tem-se, portanto, que se obter o subespaço $\mathrm{Im}\mathbf{X}^{\perp}$. Uma maneira de se obter tal subespaço é trabalhar com a matriz transposta $\mathbf{X}'$.

A transposta de $\mathbf{X}$ é uma transformação linear $\mathbf{X}' : \mathbb{R}^n \to \mathbb{R}^p$, tal que, para $\mathbf{x} \in \mathbb{R}^p$ e $\mathbf{y} \in \mathbb{R}^n$ quaisquer, tem-se $\langle \mathbf{Xx}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{X'y} \rangle$.

$\langle \mathbf{Xx}, \mathbf{y} \rangle = (\mathbf{Xx})'\mathbf{y} = \mathbf{x'X'y} = \langle \mathbf{x}, \mathbf{X'y} \rangle$. Se $\mathbf{x} \in \mathrm{Ker}\mathbf{X}$ então, para todo $\mathbf{y} \in \mathbb{R}^n$, $\langle \mathbf{Xx}, \mathbf{y} \rangle = \langle \mathbf{0}, \mathbf{y} \rangle = \langle \mathbf{x}, \mathbf{X'y} \rangle = 0$ e, consequentemente, $\mathrm{Ker}\mathbf{X}$ é perpendicular à $\mathrm{Im}\mathbf{X}'$. Da mesma forma segue que $\mathrm{Ker}\mathbf{X}'$ é perpendicular à Im$\mathbf{X}$.

Temos então a configuração geométrica essencial, mostrada na Figura 6.

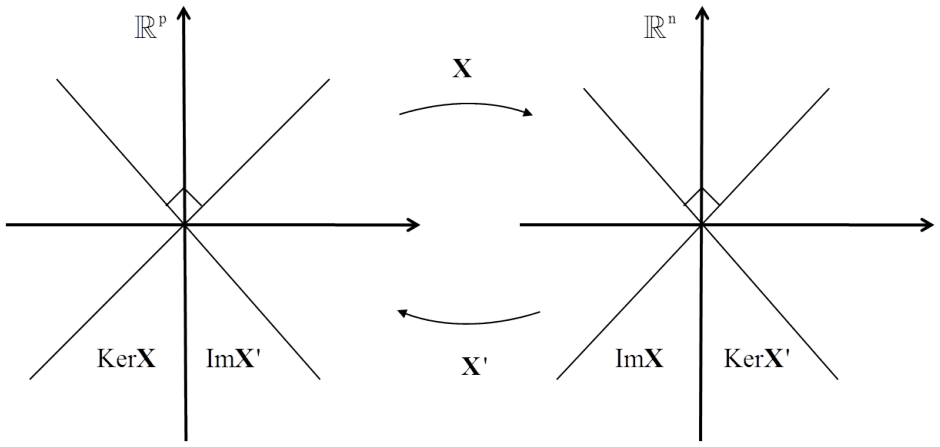


Figura 6 Representação gráfica do núcleo e da imagem de $\mathbf{X}$ e $\mathbf{X}'$

Primeiramente vamos considerar o caso mais simples em que $\mathrm{Ker}\mathbf{X} = \{0\}$. Neste caso, a matriz $\mathbf{X}'\mathbf{X}$ é inversível e define um operador $\mathbf{X}'\mathbf{X} : \mathbb{R}^p \to \mathbb{R}^p$. A ideia é definir a matriz $\mathbf{P} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'$. Como

$$\mathbf{P}^2 = \mathbf{X}\big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}'\mathbf{X}\big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}' = \mathbf{X}\big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}' = \mathbf{P},$$

tem-se que $\mathbf{P}$ é um projetor, e como

$$\mathbf{P}' = \Big(\mathbf{X}\big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}'\Big)' = \mathbf{X}\big(\mathbf{X}'\mathbf{X}\big)^{-1}\mathbf{X}' = \mathbf{P},$$

segue que $\mathbf{P}$ é um projetor ortogonal, e a projeção é no subespaço $\mathrm{Im}\mathbf{X}$.

Segue deste fato que o estimador de quadrados mínimos é dado pelo único vetor $\boldsymbol{\beta}$ que é levado por $\mathbf{X}$ no vetor $\mathbf{P}_{\mathrm{Im}\mathbf{X}}\mathbf{Y} = \Big(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\Big)\mathbf{Y}$, isto é, o estimador é obtido resolvendo-se a equação $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Novamente, utilizando que a transformação $\mathbf{X}$ é injetiva, a equação pode ser simplificada aplicando-se o operador $\mathbf{X}'$ em ambos os lados da igualdade obtendo-se

$\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \Rightarrow \mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \mathbf{X}'\mathbf{Y}$, que são as chamadas equações normais do modelo linear, com solução da forma $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$, em que OLS significa Quadrados Mínimos Ordinários (*Ordinary Least Square*), ou, simplesmente, $\hat{\boldsymbol{\beta}}$.

Quando $\mathbf{X}$ não é injetiva, o projetor ortogonal é dado por $\mathbf{P}_{\mathrm{Im}\mathbf{X}} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'$ em que $(\mathbf{X}'\mathbf{X})^{-}$ é uma inversa generalizada de $\mathbf{X}'\mathbf{X}$. Geralmente se utiliza da inversa generalizada de Moore-Penrose. Note que fica então provado que o sistema de equações normais sempre admite solução. O problema agora é que quando $\mathbf{X}$ não é injetiva existem infinitos vetores $\boldsymbol{\beta}$ em $\mathbb{R}^{\mathrm{p}}$ que são levados por $\mathbf{X}$ em $\mathbf{P}_{\mathrm{Im}\mathbf{X}}\mathbf{Y} = \left(\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\right)\mathbf{Y}$, pois se $\mathbf{w} \in \mathrm{Ker}\mathbf{X}$, $\mathbf{X}(\boldsymbol{\beta} + \boldsymbol{w}) = \mathbf{X}\boldsymbol{\beta}$, ou seja, como o núcleo de $\mathbf{X}$ contém infinitos elementos, então existem infinitos vetores que são levados na mesma imagem e, portanto, têm-se infinitas estimativas de quadrados mínimos.

Mesmo neste caso de não injetividade de $\mathbf{X}$, isto é, mesmo quando $\mathbf{X}$ não tem posto coluna completo, os vetores que são as soluções da equação $\mathbf{X}\boldsymbol{\beta} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$ são os mesmos vetores que são soluções do sistema obtido aplicando-se $\mathbf{X}'$ em ambos os lados da igualdade. Tal fato ocorre, pois como a Im$\mathbf{X}$ é perpendicular ao $\mathrm{Ker}\mathbf{X}'$, $\mathbf{X}'$ restrita à Im$\mathbf{X}$ é injetiva.

No caso de posto incompleto, escolhendo-se uma inversa generalizada de Quadrados Mínimos (RAO, 2002) fica definido uma pré-imagem, isto é, $\hat{\boldsymbol{\beta}}$ também pode ser expresso explicitamente como função de $\mathbf{Y}$ da forma (Figura 7):

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

Ocorre que, para o caso da matriz do delineamento $\mathbf{X}$ não ser de posto completo, havendo infinitas estimativas (soluções) de quadrados mínimos para os parâmetros $\boldsymbol{\beta}$, é um complicante na teoria. As propriedades e a construção geo-

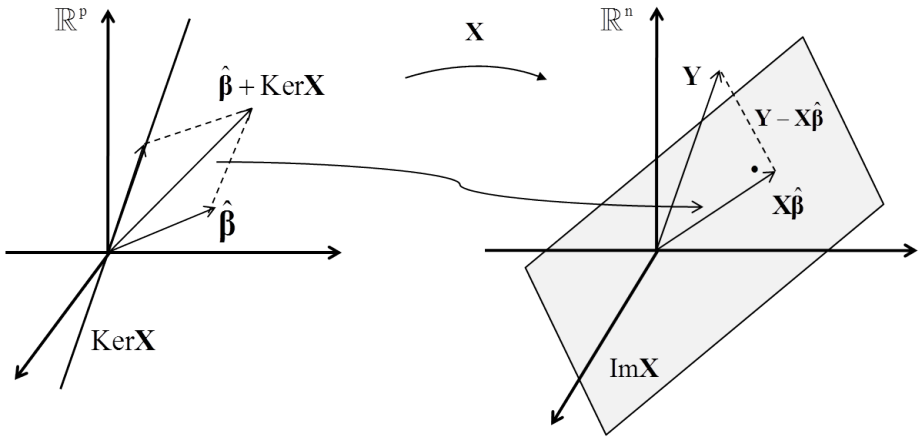


Figura 7 Modelo de Gauss-Markov quando $\mathbf{X}$ não é injetiva

métrica das inversas generalizadas podem ser vistas em Guimarães (2010).

De uma maneira geral, se o interesse é estimar a combinação linear entre os efeitos dos tratamentos, tem-se $\mathrm{x}_1\tau_1+...+\mathrm{x}_\mathrm{n}\tau_\mathrm{n}=\langle \mathbf{x}, \boldsymbol{\tau} \rangle$. Como $\langle \mathbf{x}, \boldsymbol{\tau} \rangle = \langle \mathbf{x}, \mathbf{X}\boldsymbol{\beta} \rangle = \langle \mathbf{X}'\mathbf{x}, \boldsymbol{\beta} \rangle$, uma combinação linear entre os efeitos dos tratamentos pode ser, de maneira equivalente, descrita como uma combinação linear dos parâmetros. Uma maneira de se estimar tais combinações lineares é tomar um estimador de quadrados mínimos de $\boldsymbol{\beta}$ e fazer o produto interno $\left\langle \widehat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{x} \right\rangle$. Tal procedimento, para estar bem definido, não pode depender de qual estimativa de quadrados mínimos, isto é, $\left\langle \widehat{\boldsymbol{\beta}}, \mathbf{X}'\mathbf{x} \right\rangle$ tem que ser o mesmo valor, qualquer que seja $\widehat{\boldsymbol{\beta}}$. Para que isto ocorra, o vetor $\mathbf{X}'\mathbf{x}$ tem que ser ortogonal ao $\mathrm{Ker}\mathbf{X}$. Como tal fato é verdadeiro, pois $\mathrm{Im}\mathbf{X}'$ é ortogonal ao $\mathrm{Ker}\mathbf{X}$, o procedimento pode sempre ser feito. De uma maneira geral, uma combinação linear $\langle \boldsymbol{\beta}, \mathbf{v} \rangle$ é estimável no sentido dos quadrados mínimos se $\mathbf{v}$ for perpendicular ao espaço $\mathrm{Ker}\mathbf{X}$. Caso contrário, $\langle \boldsymbol{\beta}, \mathbf{v} \rangle$ é dito não estimável, nome não muito adequado, pois não é estimável via quadrados mínimos. Mas a expressão quadrados mínimos relativa à combinação linear $\langle \boldsymbol{\beta}, \mathbf{v} \rangle$ também não é muito adequada, pois não foi feito ne-

nhum procedimento de se minimizar quadrados em relação ao valor $\langle \boldsymbol{\beta}, \mathbf{v} \rangle$.

2.2.2 Propriedades dos estimadores de quadrados mínimos

No caso de posto completo, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}$. Assim:

- *Não viesado:*

$$\begin{aligned} \mathrm{E}\left(\hat{\boldsymbol{\beta}}\right) &= \mathrm{E}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{E}(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\boldsymbol{\tau} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} = \boldsymbol{\beta}, \end{aligned}$$

  e, portanto, é um estimador não viesado.

- $\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right) = \mathrm{cov}\left((\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y}\right) = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathrm{cov}(\mathbf{Y})\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}$.
  Se $\mathrm{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$, então,

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1} = \sigma^2(\mathbf{X}'\mathbf{X})^{-1}.$$

No caso de posto incompleto, $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}$. Assim:

- *Viesado:*

$$\begin{aligned} \mathrm{E}(\hat{\boldsymbol{\beta}}) &= \mathrm{E}\left((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}\right) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathrm{E}(\mathbf{Y}) \\ &= (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\boldsymbol{\tau} = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta}, \end{aligned}$$

  logo, é um estimador viesado.

- $\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right) = \mathrm{cov}\left((\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{Y}\right) = (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathrm{cov}(\mathbf{Y})\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-}$.

Se $\mathrm{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$, então,

$$\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right) = \sigma^2(\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}(\mathbf{X}'\mathbf{X})^{-} = \sigma^2(\mathbf{X}'\mathbf{X})^{-}.$$

- Seja a combinação linear dos parâmetros, $\langle \mathbf{r}, \boldsymbol{\beta} \rangle$. Se $\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle$ é estimador de quadrados mínimos de $\langle \mathbf{r}, \boldsymbol{\beta} \rangle$, então:

  - No caso de posto completo e $\mathrm{cov}\,(\mathbf{Y}) = \sigma^2\mathbf{I}$:
    $\mathrm{E}\left(\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle\right) = \left\langle \mathbf{r}, \mathrm{E}\left(\hat{\boldsymbol{\beta}}\right) \right\rangle = \langle \mathbf{r}, \boldsymbol{\beta} \rangle$, que é não viesado, e
    $\mathrm{cov}\left(\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle\right) = \mathbf{r}'\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{r} = \sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{r}$.
  - No caso de posto incompleto e $\mathrm{cov}\,(\mathbf{Y}) = \sigma^2\mathbf{I}$:
    $\mathrm{E}\left(\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle\right) = \left\langle \mathbf{r}, \mathrm{E}\left(\hat{\boldsymbol{\beta}}\right) \right\rangle = \left\langle \mathbf{r}, (\mathbf{X}'\mathbf{X})^{-}\mathbf{X}'\mathbf{X}\boldsymbol{\beta} \right\rangle$, que é viesado, e
    $\mathrm{cov}\left(\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle\right) = \mathbf{r}'\mathrm{cov}\left(\hat{\boldsymbol{\beta}}\right)\mathbf{r} = \sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{r}$.

  Se $\langle \mathbf{l}, \boldsymbol{\beta} \rangle$ é outra combinação linear dos parâmetros, a covariância entre os estimadores é dada por:

  - No caso de posto completo e $\mathrm{cov}\,(\mathbf{Y}) = \sigma^2\mathbf{I}$:
    $\mathrm{cov}\left(\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle, \left\langle \mathbf{l}, \hat{\boldsymbol{\beta}} \right\rangle\right) = \sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-1}\mathbf{l}$.
  - No caso de posto incompleto e $\mathrm{cov}\,(\mathbf{Y}) = \sigma^2\mathbf{I}$:
    $\mathrm{cov}\left(\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle, \left\langle \mathbf{l}, \hat{\boldsymbol{\beta}} \right\rangle\right) = \sigma^2\mathbf{r}'(\mathbf{X}'\mathbf{X})^{-}\mathbf{l}$.

A vantagem de se estimar combinações lineares dos parâmetros utilizando o estimador de quadrados mínimos segue da proposição:

Proposição 2. *Entre todos os estimadores não viesados de* $\langle \mathbf{r}, \boldsymbol{\beta} \rangle$ *dados por combinações lineares dos dados, o de menor variância é* $\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle$.

Demonstração. Seja $\langle \mathbf{l}, \mathbf{Y} \rangle$ um estimador não viesado de $\langle \mathbf{r}, \boldsymbol{\beta} \rangle$. Então $\mathrm{E}\left(\langle \mathbf{l}, \mathbf{Y} \rangle\right) = \langle \mathbf{r}, \boldsymbol{\beta} \rangle$. Mas $\mathrm{E}\left(\langle \mathbf{l}, \mathbf{Y} \rangle\right) = \langle \mathbf{l}, \mathrm{E}\left(\mathbf{Y}\right) \rangle = \langle \mathbf{l}, \mathbf{X}\boldsymbol{\beta} \rangle = \langle \mathbf{X}'\mathbf{l}, \boldsymbol{\beta} \rangle$ e, portanto, $\langle \mathbf{r}, \boldsymbol{\beta} \rangle = \langle \mathbf{X}'\mathbf{l}, \boldsymbol{\beta} \rangle$. Como $\boldsymbol{\beta}$ é desconhecido, esta igualdade tem que ser válida para todo $\boldsymbol{\beta}$ e, portanto, $\mathbf{r} = \mathbf{X}'\mathbf{l}$. Como $\mathrm{var}\left(\mathbf{l}'\mathbf{Y}\right) = \|\mathbf{l}\|^2 \sigma^2$, temos que obter $\mathbf{l}$ de norma mínima na pré-imagem de $\mathbf{r}$ por $\mathbf{X}'$. Vamos projetar ortogonalmente $\mathbf{l}$ em Im $\mathbf{X}$, isto é, $\mathbf{Pl}$. Logo $\mathbf{l} = \mathbf{Pl} + (\mathbf{l} - \mathbf{Pl}) \Rightarrow \|\mathbf{l}\|^2 \geq \|\mathbf{Pl}\|^2$ e $\mathbf{X}'\mathbf{l} = \mathbf{X}'\mathbf{Pl}$.

Assim, $\mathrm{E}\left(\langle \mathbf{l} - \mathbf{Pl}, \mathbf{Y} \rangle\right) = \langle \mathbf{l} - \mathbf{Pl}, \mathrm{E}(\mathbf{Y}) \rangle = \langle \mathbf{l} - \mathbf{Pl}, \mathbf{X}\boldsymbol{\beta} \rangle = 0 \Rightarrow \langle \mathbf{l}, \mathbf{X}\boldsymbol{\beta} \rangle = \langle \mathbf{Pl}, \mathbf{X}\boldsymbol{\beta} \rangle$.

Como $\langle \mathbf{Pl}, \mathbf{Y} \rangle = \langle \mathbf{l}, \mathbf{PY} \rangle = \left\langle \mathbf{l}, \mathbf{X}\hat{\boldsymbol{\beta}} \right\rangle = \left\langle \mathbf{X}'\mathbf{l}, \hat{\boldsymbol{\beta}} \right\rangle = \left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle$, o estimador de menor variância para $\langle \mathbf{r}, \boldsymbol{\beta} \rangle$ é a combinação linear dos dados expressa por $\left\langle \mathbf{r}, \hat{\boldsymbol{\beta}} \right\rangle$.

□

2.3 Regressão Estatística

Uma das situações mais comuns em estatística é quando se tem variáveis aleatórias $\mathrm{Y}_1, \ldots, \mathrm{Y}_\mathrm{k}$ e $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{p}$ com uma distribuição conjunta

$$\mathrm{f}_{\mathrm{Y}_1,\ldots,\mathrm{Y}_\mathrm{k},\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{p}}\left(\mathrm{y}_1, \ldots, \mathrm{y}_\mathrm{k}, \mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{p}\right)$$

e, por alguma razão, as variáveis $\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{p}$ podem ser amostradas, mas as variáveis Y_i não. Após a observação das variáveis X_i, o conhecimento sobre as variáveis Y_i aumenta pois podemos calcular a função densidade de probabilidade condicional

$$\mathrm{f}_{\mathrm{Y}_1,\ldots,\mathrm{Y}_\mathrm{k}|\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{p}}\left(\mathrm{y}_1, \ldots, \mathrm{y}_\mathrm{k}; \mathrm{x}_1, \ldots, \mathrm{x}_\mathrm{p}\right).$$

Como, em geral, a densidade de probabilidade conjunta é desconhecida, a densidade de probabilidade condicional não pode ser calculada. As variáveis observáveis $\mathrm{X_i}$ são denominadas variáveis preditoras e as variáveis $\mathrm{Y_i}$ de variáveis de interesse ou respostas.

Considere a situação mais simples em que se tem apenas uma variável de interesse Y. Seja

$$\begin{aligned}&\mathrm{E}\left(\mathrm{Y}|\mathrm{X_1}{=}\mathrm{x_1},\ldots,\mathrm{X_p}{=}\mathrm{x_p}\right)\\&=\int\cdots\int \mathrm{y}\mathrm{f}_{\mathrm{Y}|\mathrm{X_1},\ldots,\mathrm{X_p}}\left(\mathrm{y};\mathrm{x_1},\ldots,\mathrm{x_p}\right)\mathrm{dy}=\mathrm{M}\left(\mathrm{x_1},\ldots,\mathrm{x_p}\right).\end{aligned}$$

Pode-se, agora, considerar a variável aleatória

$$\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)=\mathrm{E}\left(\mathrm{Y}|\mathrm{X_1},\ldots,\mathrm{X_p}\right).$$

Esta variável é a melhor variável aleatória para predizer Y, no sentido de apresentar a menor esperança do quadrado do erro, isto é, se $\mathrm{f}(\mathrm{X_1},\ldots,\mathrm{X_p})$ é uma função qualquer das variáveis $\mathrm{X_1},\ldots,\mathrm{X_p}$, então:

Proposição 3. $\mathrm{E}\left[\left(\mathrm{Y}-\mathrm{f}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)^2\right]\geq\mathrm{E}\left[\left(\mathrm{Y}-\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)^2\right].$

Demonstração.

$$\begin{aligned}&\mathrm{E}\left(\left(\mathrm{Y}-\mathrm{f}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)^2\right)\\=&\mathrm{E}\left(\left(\mathrm{Y}-\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)+\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)-\mathrm{f}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)^2\right)\\=&\mathrm{E}\left(\left(\mathrm{Y}-\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)^2\right)+\mathrm{E}\left(\left(\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)-\mathrm{f}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)^2\right)\\&+2\mathrm{E}\left(\left(\mathrm{Y}-\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)\left(\mathrm{M}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)-\mathrm{f}\left(\mathrm{X_1},\ldots,\mathrm{X_p}\right)\right)\right).\end{aligned}$$

Tem-se que

$$\mathrm{E}\left[\left(\mathrm{Y}-\mathrm{M}\left(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{m}\right)\right)\left(\mathrm{M}\left(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{m}\right)-\mathrm{f}\left(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{m}\right)\right)\right]=0.$$

Para tal, considere a notação simplificada $\mathrm{E}\left[\left(\mathrm{Y}-\mathrm{M}\right)\left(\mathrm{M}-\mathrm{f}\right)\right]$:

$$\begin{aligned}
&\mathrm{E}\left[\left(\mathrm{Y}-\mathrm{M}\right)\left(\mathrm{M}-\mathrm{f}\right)\right]\\
&=\int\int\left(\mathrm{y}-\mathrm{m}\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathrm{Y},\mathbf{X}}\left(\mathrm{y},\mathbf{x}\right)\mathrm{dydx}\\
&=\int\int\left(\mathrm{y}-\mathrm{m}\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathrm{Y}|\mathbf{X}}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dydx}\\
&=\int\left(\int\left(\mathrm{y}-\mathrm{m}\right)\mathrm{f}_{\mathrm{Y}|\mathbf{X}}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{dy}\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dx}\\
&=\int\left(\int\left(\mathrm{y}-\mathrm{E}\left(\mathrm{y}|\mathbf{x}\right)\right)\mathrm{f}_{\mathrm{Y}|\mathbf{X}}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{dy}\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dx}\\
&=\int\left(\int\mathrm{yf}_{\mathrm{Y}|\mathbf{X}}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{dy}-\int\mathrm{E}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{f}_{\mathrm{Y}|\mathbf{X}}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{dy}\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dx}\\
&=\int\left(\mathrm{E}\left(\mathrm{y}|\mathbf{x}\right)-\mathrm{E}\left(\mathrm{y}|\mathbf{x}\right)\int\mathrm{f}_{\mathrm{Y}|\mathbf{X}}\left(\mathrm{y}|\mathbf{x}\right)\mathrm{dy}\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dx}\\
&=\int\left(\mathrm{E}\left(\mathrm{y}|\mathbf{x}\right)-\mathrm{E}\left(\mathrm{y}|\mathbf{x}\right)\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dx}\\
&=\int\left(0\right)\left(\mathrm{m}-\mathrm{f}\right)\mathrm{f}_{\mathbf{X}}\left(\mathbf{x}\right)\mathrm{dx}=0
\end{aligned}$$

Logo, $\mathrm{E}\left(\left(\mathrm{Y}-\mathrm{f}\left(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{p}\right)\right)^2\right)\geq\mathrm{E}\left(\left(\mathrm{Y}-\mathrm{M}\left(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{p}\right)\right)^2\right).$

□

A variável $\mathrm{E}\left(\mathrm{Y}|\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{p}\right)=\mathrm{M}\left(\mathrm{X}_1,\ldots,\mathrm{X}_\mathrm{p}\right)$ é o preditor de menor erro quadrático médio, denominada função de regressão.

Exemplo : Caso Normal

Seja $\mathbf{x}$ uma variável aleatória multidimensional e y uma variável aleatória unidimensional. Suponha que:

$$\begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix} \sim N\left(\begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}' & \sigma_{\mathrm{y}}^2 \end{bmatrix} \right)$$

em que, $\mathbf{x} = \begin{bmatrix} \mathrm{x}_1 \\ \vdots \\ \mathrm{x}_{\mathrm{p}} \end{bmatrix}$, $\boldsymbol{\sigma} = \mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right) = \begin{bmatrix} \mathrm{cov}\left(\mathrm{x}_1, \mathrm{y}\right) \\ \vdots \\ \mathrm{cov}\left(\mathrm{x}_{\mathrm{p}}, \mathrm{y}\right) \end{bmatrix}$ e $\boldsymbol{\Sigma} = \mathrm{cov}\left(\mathbf{x}\right)$.

Dada uma matriz particionada $\mathbf{A} = \begin{bmatrix} \mathbf{A}_{11} & \mathbf{A}_{12} \\ \mathbf{A}_{21} & \mathbf{A}_{22} \end{bmatrix}$, sua inversa é dada por (RAO, 2002):

$$\mathbf{A}^{-1} = \begin{bmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & -\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \\ -\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{bmatrix}$$

ou ainda,

$$\mathbf{A}^{-1} = \begin{bmatrix} \left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1} & -\left(\mathbf{A}_{11} - \mathbf{A}_{12}\mathbf{A}_{22}^{-1}\mathbf{A}_{21}\right)^{-1}\mathbf{A}_{12}\mathbf{A}_{22}^{-1} \\ -\left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1}\mathbf{A}_{21}\mathbf{A}_{11}^{-1} & \left(\mathbf{A}_{22} - \mathbf{A}_{21}\mathbf{A}_{11}^{-1}\mathbf{A}_{12}\right)^{-1} \end{bmatrix}.$$

Assim, a inversa da matriz $\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}' & \sigma_y^2 \end{bmatrix}$ é:

$$\begin{bmatrix} \left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} & -\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} \\ -\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} & \frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} \end{bmatrix} = \begin{bmatrix} \left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} & -\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2} \\ -\frac{\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} & \frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} \end{bmatrix}.$$

Na distribuição conjunta das variáveis $\mathbf{x}$ e y, o expoente da exponencial

é:

$$
\begin{bmatrix} \mathbf{x}' & \mathrm{y}' \end{bmatrix}
\begin{bmatrix} \left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} & -\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} \\ -\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} & \frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} \end{bmatrix}
\begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix}
$$

$$
= \begin{bmatrix} \mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} - \mathrm{y}'\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} & -\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} + \mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} \end{bmatrix}
\begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix}
$$

$$
= \left(\mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} - \mathrm{y}'\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\right)\mathbf{x} + \left(-\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}} + \mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\right)\mathrm{y}
$$

$$
= \mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x} - \mathrm{y}'\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x} - \mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y} + \mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}
$$

$$
= \mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x} - \mathrm{y}'\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x} - \mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2}\mathrm{y} + \mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}
$$

$$
= \mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x} - 2\mathbf{x}'\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2}\mathrm{y} + \mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}.
$$

Fazendo,

$$
\mathbf{C} = \left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} \text{ e } \mathbf{H} = \frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2},
$$

a expressão anterior fica da forma:

$$
\mathbf{x}'\mathbf{C}\mathbf{x} - 2\mathbf{x}'\mathbf{C}\mathbf{H}\mathrm{y} + \mathrm{y}'(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B})\mathrm{y}.
$$

Logo, a marginal da variável $\mathbf{x}$ é dada por (Apêndice A):

$$
\begin{aligned}
\mathrm{f}_{\mathbf{X}}(\mathbf{x}) &\propto exp\left(\mathbf{x}'\left(\mathbf{C} - \mathbf{C}\mathbf{H}\left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right)^{-1}\mathbf{H}'\mathbf{C}\right)\mathbf{x}\right) \\
&= exp\left(\mathbf{x}'\left(\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} - \left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2}\left(\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)\frac{\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\right)\mathbf{x}\right) \\
&= exp\left(\mathbf{x}'\left(\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\left(\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)}\left(\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)\frac{\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}}{\left(\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)}\right)\mathbf{x}\right) \\
&= exp\left(\mathbf{x}'\left(\left(\boldsymbol{\Sigma} - \frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1} - \frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}}{\left(\sigma_{\mathrm{y}}^2 - \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)}\right)\mathbf{x}\right)
\end{aligned}
$$

e a distribuição condicional fica da forma:

$$
\begin{aligned}
\mathrm{f}\left(\mathrm{y}|\mathbf{x}\right) =& \frac{\mathrm{f}\left(\mathrm{y},\mathbf{x}\right)}{\mathrm{f}\left(\mathbf{x}\right)} \\
& \propto \frac{exp\left(\mathbf{x}'\left(\boldsymbol{\Sigma}-\frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x}-2\mathbf{x}'\left(\boldsymbol{\Sigma}-\frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2}\mathrm{y}+\mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}\right)}{exp\left(\mathbf{x}'\left(\left(\boldsymbol{\Sigma}-\frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}-\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}}{\left(\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)}\right)\mathbf{x}\right)} \\
& = exp\Bigg(\left(\mathbf{x}'\left(\boldsymbol{\Sigma}-\frac{\sigma\sigma'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x}-2\mathbf{x}'\left(\boldsymbol{\Sigma}-\frac{\sigma\sigma'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\sigma}{\sigma_{\mathrm{y}}^2}\mathrm{y}+\mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2-\sigma'\boldsymbol{\Sigma}^{-1}\sigma}\mathrm{y}\right) \\
& \quad -\left(\mathbf{x}'\left(\boldsymbol{\Sigma}-\frac{\sigma\sigma'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\mathbf{x}-\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\sigma\sigma'\boldsymbol{\Sigma}^{-1}}{\left(\sigma_{\mathrm{y}}^2-\sigma'\boldsymbol{\Sigma}^{-1}\sigma\right)}\mathbf{x}\right)\Bigg) \\
& = exp\left(-2\mathbf{x}'\left(\boldsymbol{\Sigma}-\frac{\boldsymbol{\sigma}\boldsymbol{\sigma}'}{\sigma_{\mathrm{y}}^2}\right)^{-1}\frac{\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2}\mathrm{y}+\mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}+\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}}{\left(\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)}\mathbf{x}\right) \\
& = exp\left(-2\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}+\mathrm{y}'\frac{1}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y}+\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathbf{x}\right).
\end{aligned}
$$

Comparando a expressão anterior com a forma geral $\left(\mathrm{y}-\mathrm{w}\right)'\frac{1}{a}\ \left(\mathrm{y}-\mathrm{w}\right)$, em relação ao termo cruzado, em que w é a média da distribuição y, tem-se:

$$
\begin{aligned}
& 2\mathrm{w}'\frac{1}{a}\mathrm{y}=2\mathbf{x}'\frac{\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\mathrm{y} \\
& \Rightarrow \mathrm{w}'=\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\ \text{ e }\ a=\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}.
\end{aligned}
$$

De onde segue que:

$$
\mathrm{f}\left(\mathrm{y}|\mathbf{x}\right)\propto exp\left(\left(\mathrm{y}-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\mathbf{x}\right)'\frac{1}{\sigma_{\mathrm{y}}^2-\boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}}\left(\mathrm{y}-\mathbf{x}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)\right).
$$

Logo, $\hat{\mathbf{Y}}=\mu_{\mathrm{y}}+\boldsymbol{\beta}'\left(\mathbf{x}-\boldsymbol{\mu}_{\mathrm{x}}\right)$ com $\boldsymbol{\beta}=\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}$.

Portanto, para o caso normal, a função de regressão é linear.

Uma questão importante para distribuições em geral é obter, entre todas as funções lineares nas covariáveis, $\alpha+\boldsymbol{\beta}'\mathbf{x}$, aquela que minimiza o erro quadrático médio $\mathrm{E}\left(\left(\mathrm{y}-\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)\right)^2\right)$. Assim:

$$
\begin{aligned}
&\mathrm{E}\left(\left(\mathrm{y}-\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)\right)^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2-2\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)\mathrm{y}+\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\mathrm{E}\left(\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)\mathrm{y}\right)+\mathrm{E}\left(\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\left(\mathrm{E}\left(\alpha\mathrm{y}\right)+\mathrm{E}\left(\boldsymbol{\beta}'\mathbf{x}\mathrm{y}\right)\right)+\mathrm{E}\left(\left(\alpha+\boldsymbol{\beta}'\mathbf{x}\right)^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\alpha\mathrm{E}\left(\mathrm{y}\right)-2\mathrm{E}\left(\left(\boldsymbol{\beta}'\mathbf{x}\right)\mathrm{y}\right)+\mathrm{E}\left(\alpha^2+2\alpha\boldsymbol{\beta}'\mathbf{x}+\left(\boldsymbol{\beta}'\mathbf{x}\right)^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\alpha\mathrm{E}\left(\mathrm{y}\right)-2\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\mathrm{y}\right)+\alpha^2+2\alpha\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right)+\mathrm{E}\left(\left(\boldsymbol{\beta}'\mathbf{x}\right)^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\alpha\mathrm{E}\left(\mathrm{y}\right)-2\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\mathrm{y}\right)+\alpha^2+2\alpha\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right)+\mathrm{E}\left(\left(\boldsymbol{\beta}'\mathbf{x}\right)^2\right)\\
&-\left(\mathrm{E}(\boldsymbol{\beta}'\mathbf{x})^2\right)+\left(\mathrm{E}(\boldsymbol{\beta}'\mathbf{x})^2\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\alpha\mathrm{E}\left(\mathrm{y}\right)-2\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\mathrm{y}\right)+\left(\alpha+\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right)\right)^2+\mathrm{var}\left(\boldsymbol{\beta}'\mathbf{x}\right)\\
=&\mathrm{E}\left(\mathrm{y}^2\right)-2\alpha\mathrm{E}\left(\mathrm{y}\right)-2\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\mathrm{y}\right)+\left(\alpha+\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right)\right)^2+\boldsymbol{\beta}'\mathrm{cov}(\mathbf{x})\boldsymbol{\beta}.
\end{aligned}
$$

Derivando em relação a α e igualando a zero, obtém-se:

$$
\begin{aligned}
&-2\mathrm{E}\left(\mathrm{y}\right)+2\left(\alpha+\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right)\right)=0\\
&\Rightarrow\alpha^*=\mathrm{E}\left(\mathrm{y}\right)-\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right).
\end{aligned}
$$

Derivando em relação a $\boldsymbol{\beta}$ e igualando a zero, obtém-se:

$$
\begin{aligned}
&-2\mathrm{E}\left(\mathbf{x}\mathrm{y}\right)+2\left(\alpha+\boldsymbol{\beta}'\mathrm{E}\left(\mathbf{x}\right)\right)\mathrm{E}\left(\mathbf{x}\right)+2\mathrm{cov}(\mathbf{x})\boldsymbol{\beta}\\
=&-2\mathrm{E}\left(\mathbf{x}\mathrm{y}\right)+2\mathrm{E}\left(\mathrm{y}\right)\mathrm{E}\left(\mathbf{x}\right)+2\mathrm{cov}(\mathbf{x})\boldsymbol{\beta}=0
\end{aligned}
$$

$$\begin{aligned}
\operatorname{cov}(\mathbf{x})\boldsymbol{\beta} =& \mathrm{E}(\mathbf{x}\mathrm{y}) - \mathrm{E}(\mathrm{y})\,\mathrm{E}(\mathbf{x}) \\
\operatorname{cov}(\mathbf{x})\boldsymbol{\beta}^{*} =& \operatorname{cov}(\mathrm{y},\mathbf{x}) \\
\boldsymbol{\beta}^{*} =& \left(\operatorname{cov}(\mathbf{x})\right)^{-1} \operatorname{cov}(\mathrm{y},\mathbf{x}) \\
\boldsymbol{\beta}^{*} =& \boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}.
\end{aligned}$$

Esta combinação linear, $\alpha^{*}+\boldsymbol{\beta}^{*'}\mathbf{x}$, tem covariância máxima com y, em relação a qualquer outra combinação linear $\alpha + \boldsymbol{\beta}'\mathbf{x}$, com $\boldsymbol{\beta}'\operatorname{cov}(\mathbf{x})\,\boldsymbol{\beta} = \boldsymbol{\beta}^{*'}\operatorname{cov}(\mathbf{x})\,\boldsymbol{\beta}^{*}$.

Demonstração.

$$\begin{aligned}
\boldsymbol{\beta}^{*'}\operatorname{cov}(\mathbf{x})\,\boldsymbol{\beta}^{*} =& \left(\left(\operatorname{cov}(\mathbf{x})\right)^{-1}\operatorname{cov}(\mathrm{y},\mathbf{x})\right)' \operatorname{cov}(\mathbf{x}) \left(\left(\operatorname{cov}(\mathbf{x})\right)^{-1}\operatorname{cov}(\mathrm{y},\mathbf{x})\right) \\
=& \left(\operatorname{cov}(\mathrm{y},\mathbf{x})\right)' \left(\operatorname{cov}(\mathbf{x})\right)^{-1}\operatorname{cov}(\mathbf{x}) \left(\operatorname{cov}(\mathbf{x})\right)^{-1}\operatorname{cov}(\mathrm{y},\mathbf{x}) \\
=& \left(\operatorname{cov}(\mathrm{y},\mathbf{x})\right)' \left(\operatorname{cov}(\mathbf{x})\right)^{-1}\operatorname{cov}(\mathrm{y},\mathbf{x}) \\
=& \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma},
\end{aligned}$$

e $\operatorname{cov}\left(\mathrm{y}, \alpha + \boldsymbol{\beta}'\mathbf{x}\right) = \operatorname{cov}\left(\mathrm{y}, \boldsymbol{\beta}'\mathbf{x}\right) = \boldsymbol{\beta}'\operatorname{cov}(\mathrm{y},\mathbf{x}) = \boldsymbol{\beta}'\boldsymbol{\sigma}$.

Assim, tem-se que maximizar $\boldsymbol{\beta}'\boldsymbol{\sigma}$, com a restrição $\boldsymbol{\beta}'\operatorname{cov}(\mathbf{x})\,\boldsymbol{\beta} = \boldsymbol{\beta}^{*'}\operatorname{cov}(\mathbf{x})\,\boldsymbol{\beta}^{*} = \boldsymbol{\sigma}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}$.

A lagrangiana é dada por:

$$\mathrm{H}(\boldsymbol{\beta}, \lambda) = \boldsymbol{\beta}'\boldsymbol{\sigma} + \lambda\left(\boldsymbol{\beta}'\operatorname{cov}(\mathbf{x})\,\boldsymbol{\beta} - \boldsymbol{\sigma}'\left(\operatorname{cov}(\mathbf{x})\right)^{-1}\boldsymbol{\sigma}\right).$$

Derivando e igualando a zero, obtém-se

$$\frac{\mathrm{dH}}{\mathrm{d}\lambda} = 0 \Rightarrow \boldsymbol{\beta}'\mathrm{cov}\left(\mathbf{x}\right)\boldsymbol{\beta} = \boldsymbol{\sigma}'(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma} \tag{1}$$

$$\frac{\mathrm{dH}}{\mathrm{d}\boldsymbol{\beta}} = 0 \Rightarrow \boldsymbol{\sigma} + \lambda\left(2\mathrm{cov}\left(\mathbf{x}\right)\boldsymbol{\beta}\right) = 0. \tag{2}$$

Segue que

$$\boldsymbol{\sigma} = -2\lambda\mathrm{cov}\left(\mathbf{x}\right)\boldsymbol{\beta} \Rightarrow \boldsymbol{\beta} = -\frac{(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}}{2\lambda}. \tag{3}$$

Substituindo (3) na equação (1), tem-se

$$\left(-\frac{(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}}{2\lambda}\right)' \mathrm{cov}\left(\mathbf{x}\right)\left(-\frac{(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}}{2\lambda}\right) = \boldsymbol{\sigma}'(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}$$

$$\Rightarrow \frac{1}{4\lambda^2}\boldsymbol{\sigma}'(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma} = \boldsymbol{\sigma}'(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}$$

$$\Rightarrow \frac{1}{4\lambda^2} = 1 \Rightarrow 4\lambda^2 = 1 \Rightarrow \lambda = \pm\frac{1}{2}.$$

Logo,

$$\begin{aligned}\boldsymbol{\beta} &= -\frac{(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}}{2\lambda} = -\frac{(\mathrm{cov}\left(\mathbf{x}\right))^{-1}\boldsymbol{\sigma}}{2\left(\pm\frac{1}{2}\right)} \\ &= \pm\left(\mathrm{cov}\left(\mathbf{x}\right)\right)^{-1}\boldsymbol{\sigma} = \pm\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma} = \pm\boldsymbol{\beta}^{*}.\end{aligned}$$

O valor máximo ocorre quando o sinal é positivo, assim, $\boldsymbol{\beta} = \boldsymbol{\beta}^{*}$.

□

Segue, então, que o melhor preditor linear (*Best Linear Predictor* - BLP) de menor erro quadrático médio é dado por:

$$\begin{aligned}\alpha^* + \mathbf{x}'\boldsymbol{\beta}^* &= \mathrm{E}(\mathrm{y}) - \boldsymbol{\beta}^{*'}\mathrm{E}(\mathbf{x}) + \boldsymbol{\beta}^{*'}\mathbf{x} \\ &= \mathrm{E}(\mathrm{y}) + \boldsymbol{\beta}^{*'}(\mathbf{x} - \mathrm{E}(\mathbf{x})).\end{aligned}$$

Neste caso, o EQM do preditor é:

$$\begin{aligned}&\mathrm{E}\left(\left(\mathrm{y} - \left(\mathrm{E}(\mathrm{y}) + \boldsymbol{\beta}^{*'}(\mathbf{x} - \mathrm{E}(\mathbf{x}))\right)\right)^2\right) \\ =&\mathrm{E}\left((\mathrm{y} - \mathrm{E}(\mathrm{y}))^2 - 2(\mathrm{y} - \mathrm{E}(\mathrm{y}))\boldsymbol{\beta}^{*'}(\mathbf{x} - \mathrm{E}(\mathbf{x})) + \left(\boldsymbol{\beta}^{*'}(\mathbf{x} - \mathrm{E}(\mathbf{x}))\right)^2\right) \\ =&\mathrm{E}(\mathrm{y} - \mathrm{E}(\mathrm{y}))^2 - 2\mathrm{E}\left((\mathrm{y} - \mathrm{E}(\mathrm{y}))\boldsymbol{\beta}^{*'}(\mathbf{x} - \mathrm{E}(\mathbf{x}))\right) + \mathrm{E}\left(\boldsymbol{\beta}^{*'}(\mathbf{x} - \mathrm{E}(\mathbf{x}))\right)^2 \\ =&\mathrm{var}(\mathrm{y}) - 2\boldsymbol{\beta}^{*'}\mathrm{cov}(\mathrm{y}, \mathbf{x}) + \boldsymbol{\beta}^{*'}\mathrm{cov}(\mathbf{x})\boldsymbol{\beta}^* \\ =&\mathrm{var}(\mathrm{y}) - 2\left((\mathrm{cov}(\mathbf{x}))^{-1}\mathrm{cov}(\mathrm{y}, \mathbf{x})\right)'\mathrm{cov}(\mathrm{y}, \mathbf{x}) \\ &+ \left((\mathrm{cov}(\mathbf{x}))^{-1}\mathrm{cov}(\mathrm{y}, \mathbf{x})\right)'\mathrm{cov}(\mathbf{x})\left((\mathrm{cov}(\mathbf{x}))^{-1}\mathrm{cov}(\mathrm{y}, \mathbf{x})\right) \\ =&\mathrm{var}(\mathrm{y}) - \left((\mathrm{cov}(\mathbf{x}))^{-1}\mathrm{cov}(\mathrm{y}, \mathbf{x})\right)'\mathrm{cov}(\mathrm{y}, \mathbf{x}) \\ =&\sigma_{\mathrm{y}}^2 - \left(\boldsymbol{\Sigma}^{-1}\boldsymbol{\sigma}\right)'\boldsymbol{\sigma}.\end{aligned}$$

Observação: $\boldsymbol{\beta}^*$ é um vetor paramétrico, não é uma estimativa. O interessante é que este vetor paramétrico depende das variâncias e covariâncias e não da forma específica da distribuição populacional $\mathrm{f}_{\mathrm{y},\mathbf{x}}(\mathrm{y}, \mathbf{x})$.

2.4 Componentes Principais

A análise de componentes principais (*Principal Component Analysis* - PCA) pode ser aplicada como um método de redução de dimensionalidade, visando a solucionar eventuais problemas de multicolinearidade da matriz **X** e a necessidade de um número excessivo de amostras para a construção de um modelo viável, sem acarretar na perda significativa de informações presentes nos dados (OTTO, 1999).

O PCA pode ser feito com base em autovalores e autovetores de uma matriz de covariância. Os elementos dos autovetores são os coeficientes requeridos para a combinação linear das variáveis originais e são conhecidos como *loadings*. Os elementos individuais das novas variáveis são derivados das variáveis explicativas e são conhecidos como os *scores*. A variância total dos dados originais é dada pela soma de suas variâncias individuais e os autovalores representam a variância associada a cada componente principal.

2.5 Regressão por Componentes Principais

A regressão via componentes principais (*Principal Components Regression* - PCR) foi introduzida por Kendall (1957) e Hotelling (1957). Este método faz uso dos procedimentos empregados na análise de componentes principais visando a contornar os problemas apresentados pela regressão linear múltipla.

A metodologia do método PCR consiste, basicamente, em eliminar componentes que não contribuam na explicação da variância presente nos dados, o que reduz a dimensionalidade dos dados originais. Após a escolha do número ótimo de componentes a serem utilizados, os coeficientes podem ser determinados por

meio do método dos quadrados mínimos ordinários.

Como em qualquer regressão múltipla, a escolha do número de variáveis a serem incluídas no modelo é de extrema importância, tendo em vista evitar a perda de informações relevantes. Segundo Roggo *et al*. (2007), cada componente descreve uma fração da variação total contida nos dados, tornando possível a determinação do número ótimo de componentes a serem incluídos na regressão.

2.6 Regressão por Quadrados Mínimos Parciais

Quadrados Mínimos Parciais (*Partial Least Squares* - PLS) foi introduzido por Wold em 1975, sendo considerado útil para a construção de equações de predições em situações nas quais se tem um grande número de variáveis explicativas e um número relativamente pequeno de dados amostrais (HOSKULDSSON, 1988). Resumidamente, a ideia geral do PLS é formar componentes que capturem a maior quantidade de informação possível disposta nas variáveis explicativas $(\mathrm{X}_1, \ldots, \mathrm{X}_\mathrm{p})$ para predizer as variáveis respostas $(\mathrm{Y}_1, \ldots, \mathrm{Y}_\mathrm{k})$. O método PLS apresenta similaridades com o método de Regressão via Componentes Principais, sendo a maior diferença entre eles dada pelo fato do PCR levar em consideração apenas as variáveis explicativas na construção dos componentes, enquanto que o PLS também leva em consideração as variáveis respostas.

3 METODOLOGIA

3.1 Componentes Principais

Foi feita uma abordagem da obtenção dos Componentes Principais por meio do cálculo diferencial, além do tradicional método multiplicadores de Lagrange. Geometricamente, objetivou-se encontrar a direção de maior variabilidade do vetor aleatório $\mathbf{X}$ (covariáveis).

3.2 Regressão por Componentes Principais

A regressão em componentes principais é colocada neste trabalho em razão das semelhanças com o PLS. Na linha deste trabalho, a exposição da teoria do PCR foi bastante geométrica e, portanto, diferente da abordagem usual na literatura.

Os componentes são obtidos pelas direções no subespaço gerado pelas observações das covariáveis que fornecem a maior variabilidade dos dados. Na construção dos componentes, utiliza-se apenas as covariáveis, e estes componentes representam as novas variáveis a serem utilizadas na regressão. Os coeficientes da regressão podem ser estimados pelo método dos quadrados mínimos.

Geometricamente, foi demonstrado que os componentes são representados como imagem dos autovetores da matriz $\mathbf{X}'\mathbf{X}$. Obtendo o subespaço gerado por estes componentes, a regressão é feita pela projeção do vetor de dados neste subespaço. Também é demonstrado que o vetor de coeficientes estimado via Regressão em Componentes Principais, $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$, é um estimador de encolhimento.

3.3 Regressão por Quadrados Mínimos Parciais

Uma ferramenta para resolver problemas de regressão nos quais as variáveis preditoras (ou covariáveis) apresentam um alto nível de colinearidade é o uso da teoria dos Quadrados Mínimos Parciais. Neste caso, a matriz das covariáveis $\mathbf{X}$ pode ser de posto incompleto. Outra situação, muito comum em aplicações na área de química, ocorre quando o número de covariáveis é muito maior que o número de observações. Em ambas as situações, as equações normais para obtenção do estimador de quadrados mínimos admitem infinitas soluções. Para a escolha de uma estimativa em particular, um método possível é o uso de inversas generalizadas. No entanto, a escolha desta inversa é, de certa forma, arbitrária, gerando situações indesejáveis para as equações de predição obtidas. No método de regressão utilizando PLS, uma estimativa única para o parâmetro de regressão é obtida.

O método é semelhante ao de Regressão em Componentes Principais, mas enquanto este só depende da matriz das covariáveis $\mathbf{X}$, o método PLS depende também do vetor de dados $\mathbf{Y}$. A ideia básica é regredir o vetor de dados $\mathbf{Y}$ em cada um dos vetores das covariáveis e, através de escolha de pesos, obter combinações lineares destas regressões. Desta forma, o vetor de dados $\mathbf{Y}$ é explicado em termos de um número menor de vetores. A escolha destas combinações lineares é justificada estatisticamente, pela maximização de covariâncias.

A estimativa do vetor de parâmetros através do método PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, admite uma expressão como uma transformação linear da estimativa $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, invariante em relação ao $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ escolhido. O método possui uma natureza geométrica sendo obtido por meio de projeções oblíquas, diferentemente do método dos quadrados mínimos, que utiliza projeções ortogonais.

A literatura referente ao método PLS, apesar de abundante, é em sua mai-

oria, aplicada em problemas oriundos da Quimiometria e Genética, por exemplo. Poucas referências existem relativas à parte teórica do método. Em geral, os artigos se restringem a apresentar os algoritmos básicos utilizados pelo método. As referências mais acessíveis relacionadas à teoria geométrica e algébrica do PLS e à descrição dos algoritmos são os artigos Helland (1990), Hoskuldsson (1988), Garthwaite (1994) e Phatak e Jong (1997). Sendo assim, o texto que segue discute o método PLS e suas propriedades, essencialmente, sob o ponto de vista dos autores destes quatro artigos, apresentando, sempre que possível, uma abordagem geométrica. Além disso, é feito um paralelo entre os algoritmos para o PLS populacional e amostral.

3.4 Determinação do número ótimo de componentes

Uma questão a ser abordada é o número de componentes que deve ser utilizado na regressão em PLS e, para isto, é necessário utilizar alguns conceitos de seleção de modelos. Dentre os métodos propostos na literatura, são abordados os conceitos de Graus de Liberdade (*Degrees of Freedom* - DoF) e Validação Cruzada (*Cross Validation* - CV).

São apresentadas as definições do DoF para modelos de regressão linear e não linear, bem como a demonstração para uma cota inferior do DoF quando $\mathrm{m} = 1$. Ainda, dois gráficos ilustram como o DoF e o Erro Quadrático Médio (EQM - utilizado na validação cruzada) se comportam em função do número de componentes. Os gráficos foram gerados no software R (R Core Team, 2013), por meio das funções *pls.model* e *pls.cv* do pacote *plsdof*, respectivamente, para DoF e CV.

3.5 Exemplo Didático

Foi elaborada uma rotina no software R (Apêndice D) com o objetivo de explicitar passo a passo os resultados do algoritmo PLS (HOSKULDSSON, 1988). Um exemplo didático foi apresentado para ilustrar tal rotina.

Inicialmente, gerou-se uma matriz de covariâncias, 5×5, da forma

$$\begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}' & \sigma_{\mathrm{y}}^2 \end{bmatrix}, \ \text{em que } \boldsymbol{\sigma} = \mathrm{cov}\left(\mathbf{X}, \mathrm{Y}\right), \quad \boldsymbol{\Sigma} = \mathrm{cov}\left(\mathbf{X}\right) \ \text{ e } \ \sigma_{\mathrm{y}}^2 = \mathrm{var}(\mathrm{Y}).$$

Utilizando esta matriz de covariâncias, gerou-se de uma Normal Multivariada, uma matriz de covariáveis $\mathbf{X}$, com 4 covariáveis, e um vetor para a variável de interesse Y, ambos com $\mathrm{n} = 6$ observações.

Obtidos $\mathbf{X}$ e Y, implementou-se o algoritmo PLS para a construção dos componentes. Para efeito de comparação, estimou-se o vetor de parâmetros $\hat{\boldsymbol{\beta}}$ pela regressão por Quadrados Mínimos Ordinários, $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, que só depende de $\mathbf{X}$ e de Y, e pela regressão por Quadrados Mínimos Parciais, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, utilizando os m componentes obtidos no algoritmo, com $\mathrm{m} = 1, 2, 3 \text{ e } 4$.

Uma vez obtidos os parâmetros da regressão, gerou-se um novo valor para cada uma das quatro covariáveis ($\mathbf{X}_{\mathrm{p}}$), e um novo valor para a variável de interesse (Y_{p}), a fim de fazer predição. Com os novos valores de $\mathbf{X}_{\mathrm{p}}$ e com os vetores $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ e $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, encontrou-se o valor predito para Y_{p} ($\hat{\mathrm{Y}}_{\mathrm{p}} = \mathbf{X}_{\mathrm{p}}\hat{\boldsymbol{\beta}}$), denominados $\hat{\mathrm{Y}}_{\mathrm{p,OLS}}$ e $\hat{\mathrm{Y}}_{\mathrm{p,PLS}}$.

4 RESULTADOS

4.1 Componentes Principais

A teoria de componentes principais está relacionada ao problema básico: dado um vetor aleatório $\mathbf{Z}$, encontrar a direção em que este vetor apresenta a maior variação. Geometricamente, o problema pode ser descrito como: dada uma direção $\boldsymbol{a} \in \mathbb{R}^n$, $\|\boldsymbol{a}\| = 1$, projeta-se $\mathbf{Z}$ na linha definida por $\boldsymbol{a}$ (Figura 8) obtendo-se um vetor aleatório totalmente descrito por sua norma $\boldsymbol{a} \cdot \mathbf{Z} = \|\mathbf{Z}\| \cos\theta$ que define uma variável aleatória unidimensional, $\mathbf{Y}$. A variância desta é a variabilidade do vetor $\mathbf{Z}$ na direção de $\boldsymbol{a}$. Nas condições do Teorema 1, tem-se:

$$\text{var}\,(\mathbf{Y}) = \text{var}\,(\boldsymbol{a} \cdot \mathbf{Z}) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}.$$

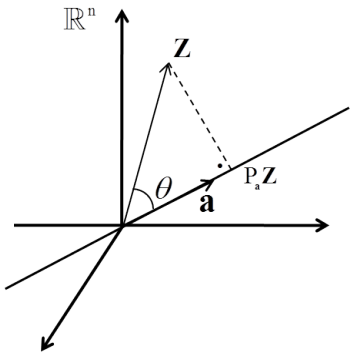


Figura 8 Representação geométrica da obtenção do componente principal

É necessário, então, maximizar var $(\boldsymbol{a} \cdot \mathbf{Z}) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a}$ com a restrição $\|\boldsymbol{a}\| = 1$. Serão apresentados dois métodos de maximização: utilizando multiplicadores de Lagrange e Cálculo Diferencial.

Utilizando multiplicadores de Lagrange, tem-se:

$$\mathbf{H}(\boldsymbol{a}, \lambda) = \boldsymbol{a}'\boldsymbol{\Sigma}\boldsymbol{a} - \lambda\left(\|\boldsymbol{a}\|^2 - 1\right)$$

$$\frac{\partial \mathbf{H}}{\partial \boldsymbol{a}} = 2\boldsymbol{\Sigma}\boldsymbol{a} - 2\lambda\boldsymbol{a} = 0 \Rightarrow \boldsymbol{\Sigma}\boldsymbol{a} = \lambda\boldsymbol{a}.$$

Portanto, os pontos críticos, $\boldsymbol{a}$, são os autovetores da matriz $\boldsymbol{\Sigma}$.

Uma outra demonstração pode ser feita através do cálculo diferencial: seja $a\left(\mathrm{t}\right) = \left(a_1\left(\mathrm{t}\right), \ldots, a_\mathrm{n}\left(\mathrm{t}\right)\right)$ uma curva parametrizada com $\|a\left(\mathrm{t}\right)\| = 1$, isto é, uma curva sobre a esfera de raio unitário, centrada na origem do espaço $\mathbb{R}^\mathrm{n}$. Vamos também supor que esta curva esteja parametrizada pelo comprimento do arco, isto é, $\left\|\frac{\mathrm{d}a}{\mathrm{dt}}\left(\mathrm{t}\right)\right\| = 1$. Definindo a função $\mathrm{f}\left(a\left(\mathrm{t}\right)\right) = a'\left(\mathrm{t}\right)\boldsymbol{\Sigma}a\left(\mathrm{t}\right)$, o que se quer obter são os pontos críticos de f.

Derivando e igualando a zero,

$$\frac{\mathrm{df}}{\mathrm{dt}}\left(0\right) = \frac{\mathrm{df}}{\mathrm{d}a}\left(a\left(0\right)\right)\frac{\mathrm{d}a}{\mathrm{dt}}\left(0\right) = 0$$

Como $\frac{\mathrm{df}}{\mathrm{d}a}\left(a\left(0\right)\right) = 2\boldsymbol{\Sigma}a\left(0\right)$ tem-se

$$2\left(\boldsymbol{\Sigma}a\left(0\right)\right)'\frac{\mathrm{d}a}{\mathrm{dt}}\left(0\right) = 0$$

e segue, portanto, que o vetor $\frac{\mathrm{d}a}{\mathrm{dt}}\left(0\right)$ é perpendicular ao vetor $\boldsymbol{\Sigma}a\left(0\right)$.

Como tal fato tem que ser verdade para qualquer curva que passa em $a\left(0\right)$ significa que $\boldsymbol{\Sigma}a\left(0\right)$ é um vetor perpendicular ao espaço tangente da esfera em $a\left(0\right)$. Neste caso $\boldsymbol{\Sigma}a\left(0\right)$ tem que ser um múltiplo de $a\left(0\right)$ e, portanto, $\boldsymbol{\Sigma}a\left(0\right) = \lambda a\left(0\right)$. Assim, os pontos críticos de f são justamente os autovetores de $\boldsymbol{\Sigma}$.

A natureza dos pontos críticos pode ser estudada obtendo-se a derivada

segunda:

$$\begin{aligned}
\mathrm{f}'(\mathrm{t}) &= 2a'(\mathrm{t})\,\boldsymbol{\Sigma}\frac{\mathrm{d}a}{\mathrm{dt}}(\mathrm{t}) \\
\mathrm{f}''(\mathrm{t}) &= 2\frac{\mathrm{d}a'}{\mathrm{dt}}(\mathrm{t})\,\boldsymbol{\Sigma}\left(\frac{\mathrm{d}a}{\mathrm{dt}}(\mathrm{t})\right) + 2a'(\mathrm{t})\,\boldsymbol{\Sigma}\frac{\mathrm{d}^2a}{\mathrm{dt}^2}(\mathrm{t}) \\
\mathrm{f}''(0) &= 2\frac{\mathrm{d}a'}{\mathrm{dt}}(0)\,\boldsymbol{\Sigma}\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right) + 2a'(0)\,\boldsymbol{\Sigma}\frac{\mathrm{d}^2a}{\mathrm{dt}^2}(0).
\end{aligned}$$

Como $\left\|\frac{\mathrm{d}a}{\mathrm{dt}}(\mathrm{t})\right\| = 1$ (parametrizado pelo comprimento do arco), tem-se:

$$1 = \left(\frac{\mathrm{d}a}{\mathrm{dt}}(\mathrm{t})\right)'\left(\frac{\mathrm{d}a}{\mathrm{dt}}(\mathrm{t})\right) \Rightarrow 0 = 2\left(\frac{\mathrm{d}^2a}{\mathrm{dt}^2}\right)'\frac{\mathrm{d}a}{\mathrm{dt}}(\mathrm{t}).$$

Logo, $\frac{\mathrm{d}^2a}{\mathrm{dt}^2}(0)$ é paralelo a $a(0)$ e, portanto, também é um autovetor de $\boldsymbol{\Sigma}$, isto é,

$$\boldsymbol{\Sigma}\left(\frac{\mathrm{d}^2a}{\mathrm{dt}^2}(0)\right) = \lambda\left(\frac{\mathrm{d}^2a}{\mathrm{dt}^2}(0)\right) = \lambda \mathrm{k} a(0),$$

em que k é negativo (note que $\frac{\mathrm{d}^2a}{\mathrm{dt}^2}(\mathrm{t})$ é a aceleração centrípeta). Assim,

$$\begin{aligned}
\mathrm{f}''(0) &= 2\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right)'\boldsymbol{\Sigma}\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right) + 2\lambda\mathrm{k}\|a(0)\|^2 \\
&= 2\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right)'\boldsymbol{\Sigma}\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right) + 2\lambda\mathrm{k},
\end{aligned}$$

e $\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right)'\boldsymbol{\Sigma}\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right) \geq 0$ pois $\boldsymbol{\Sigma}$ é positiva semidefinida. Os autovalores de $\boldsymbol{\Sigma}$ são positivos pelo mesmo motivo. A curva $a(\mathrm{t})$ pode ser sempre pensada como

$$a(\mathrm{t}) = \cos(\mathrm{t})\,\mathrm{i} + \mathrm{sen}(\mathrm{t})\,\mathrm{j} \Rightarrow \frac{\mathrm{d}^2a}{\mathrm{dt}^2}(\mathrm{t}) = -\cos(\mathrm{t})\,\mathrm{i} - \mathrm{sen}(\mathrm{t})\,\mathrm{j} = -a(\mathrm{t}).$$

Dessa forma, k$= -1$, e temos $\mathrm{f}''(0) = 2\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right)'\boldsymbol{\Sigma}\left(\frac{\mathrm{d}a}{\mathrm{dt}}(0)\right) - \lambda$.

Portanto, $a\,(0)$ é um ponto de máximo se

$$\frac{\mathrm{d}^2\mathrm{f}}{\mathrm{dt}^2}\left(a\left(\mathrm{t}\right)\right)\Big|_{\mathrm{t}=0} < 0 \Rightarrow \left(\frac{\mathrm{d}a}{\mathrm{dt}}\left(0\right)\right)' \mathbf{\Sigma} \left(\frac{\mathrm{d}a}{\mathrm{dt}}\left(0\right)\right) < \lambda,$$

para todo vetor $\frac{\mathrm{d}a}{\mathrm{dt}}\,(0)$ no espaço tangente da esfera S^{n} em $a(0)$ (Figura 9).

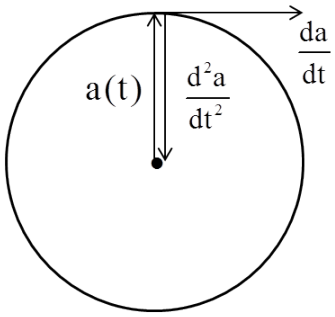


Figura 9 Representação geométrica da curva parametrizada pelo comprimento do arco sobre a esfera de raio unitário, centrada na origem do espaço $\mathbb{R}^{\mathrm{n}}$

Obtidos os autovetores $\boldsymbol{a}$ de $\mathbf{\Sigma}$, as componentes principais $\mathbf{Y}$ são obtidas como as combinações lineares $\boldsymbol{a} \cdot \mathbf{Z}$. Onde conclui-se que para o maior autovalor se tem o máximo, para o menor autovalor o mínimo, e os outros pontos críticos são selas.

4.2 Regressão por Componentes Principais: uma abordagem geométrica

Esta seção tem como referência o artigo Phatak e Jong (1997).

Para o problema de regressão em que se tem o modelo $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta}+\boldsymbol{\varepsilon}$, a ideia básica é obter a direção, no subespaço Im$\mathbf{X}$, que fornece a maior variabilidade de $\mathbf{Y}$.

Neste caso, o problema de multiplicadores de Lagrange, agora, é mais complicado uma vez que se têm duas restrições ($\|\boldsymbol{\beta}\| = 1$ e subespaço Im$\mathbf{X}$). No entanto, pode-se modificar o problema para maximizar var $(\mathbf{X}\boldsymbol{\beta} \cdot \mathbf{Y})$ restrito a $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$, isto é, maximizar $(\mathbf{X}\boldsymbol{\beta})' \boldsymbol{\Sigma} (\mathbf{X}\boldsymbol{\beta})$, restrito a $\|\boldsymbol{\beta}\| = 1$, ou seja, $\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}$ $\mathbf{X}\boldsymbol{\beta}$, restrito a $\|\boldsymbol{\beta}\| = 1$, como mostrado na Figura 10.

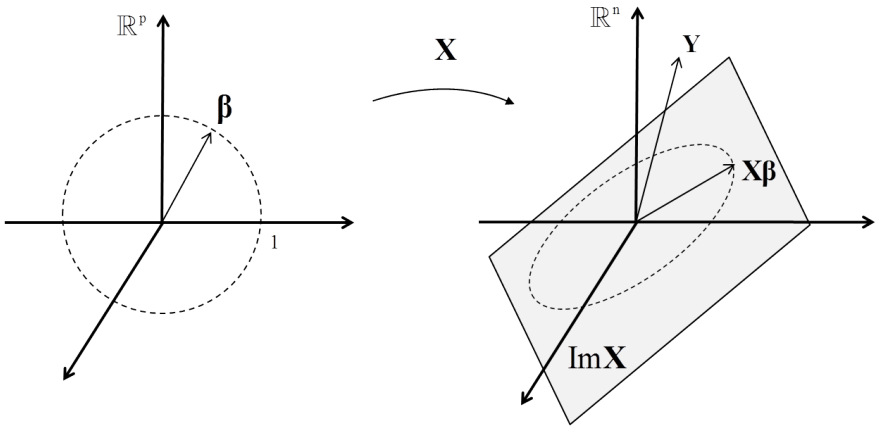


Figura 10 Representação geométrica da maximização da var $(\mathbf{X}\boldsymbol{\beta} \cdot \mathbf{Y})$ restrito a $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$

Este problema é semelhante ao anteriormente feito, com a direção de maior variação dada agora no espaço paramétrico. As direções são dadas pelos autovetores de $\mathbf{X}'\boldsymbol{\Sigma}\mathbf{X}$. O caso mais simples ocorre quando $\boldsymbol{\Sigma} = \sigma^2\mathbf{I}$, e então, as direções são dadas pelos autovetores de $\mathbf{X}'\mathbf{X}$.

Sejam $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_\mathrm{p}$ os autovetores e $\lambda_1, \lambda_2, ..., \lambda_\mathrm{p}$ os respectivos autova-

lores, da matriz $\mathbf{X}'\mathbf{X}$. Assim, $\mathbf{X}'\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}} = \lambda_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}}$ com $\boldsymbol{\gamma}_{\mathrm{i}}^{'}\boldsymbol{\gamma}_{\mathrm{i}} = 1$ e $\boldsymbol{\gamma}_{\mathrm{i}}^{'}\boldsymbol{\gamma}_{\mathrm{j}} = 0$ para $\mathrm{i} \neq \mathrm{j} : 1, ..., \mathrm{p}$ (Figura 11).

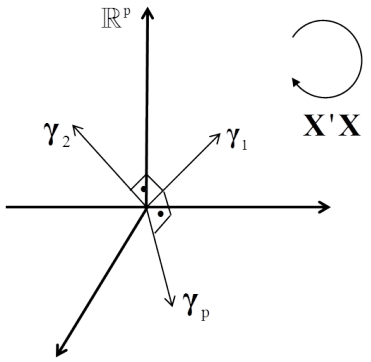


Figura 11 Representação geométrica dos autovetores $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{p}}$ da matriz $\mathbf{X}'\mathbf{X}$

Note que,

$$\mathrm{var}\left(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}} \cdot \mathbf{Y}\right) = \left(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}\right)' \left(\sigma^2\mathbf{I}\right)\left(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}\right) = \sigma^2\boldsymbol{\gamma}_{\mathrm{i}}^{'}\mathbf{X}'\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}} = \sigma^2\boldsymbol{\gamma}_{\mathrm{i}}^{'}\left(\lambda_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}}\right) = \sigma^2\lambda_{\mathrm{i}}.$$

Os vetores $\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \ldots, \mathbf{X}\boldsymbol{\gamma}_{\mathrm{p}}$ são ortogonais, pois:

$$\mathbf{X}\mathbf{X}'\left(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}\right) = \mathbf{X}\left(\mathbf{X}'\mathbf{X}\right)\boldsymbol{\gamma}_{\mathrm{i}} = \mathbf{X}\lambda_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}} = \lambda_{\mathrm{i}}\left(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}\right).$$

Assim, os $\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}$ são autovetores de $\mathbf{X}\mathbf{X}'$, portanto, ortogonais.

Ordenando os autovalores $\lambda_1 > \lambda_2 > ... > \lambda_{\mathrm{p}}$, definem-se como componentes principais as direções dadas por $\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3, ...$ Como estas direções apresentam uma ordem decrescente de variação do vetor aleatório $\mathbf{Y}$, elas "explicam" bem o vetor $\mathbf{Y}$. A ideia, então, é formar os subespaços (Figura 12):

$$\mathrm{W}_1 = span\left\{\mathbf{X}\boldsymbol{\gamma}_1\right\}$$
$$\mathrm{W}_2 = span\left\{\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2\right\}$$

$$\mathrm{W}_3 = span\left\{\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3\right\}$$
$$\vdots$$
$$\mathrm{W}_\mathrm{m} = span\left\{\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3, ..., \mathbf{X}\boldsymbol{\gamma}_\mathrm{m}\right\}$$
$$\mathrm{Im}\mathbf{X} = span\left\{\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3, ..., \mathbf{X}\boldsymbol{\gamma}_\mathrm{p}\right\}.$$

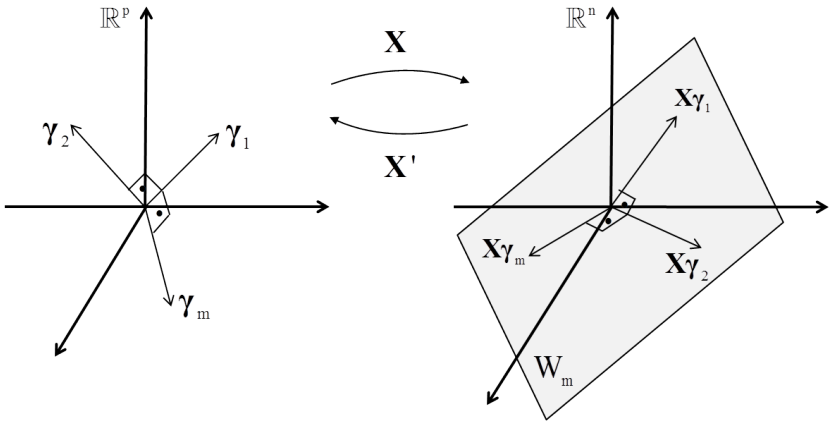


Figura 12 Subespaço W_m gerado pelos m-primeiros autovetores da matriz $\mathbf{X}'\mathbf{X}$

Como estas direções explicam a variabilidade de $\mathbf{Y}$, a ideia é projetar $\mathbf{Y}$ ortogonalmente no subespaço W_m, isto é, regredir $\mathbf{Y}$ aos m-primeiros componentes principais. A projeção define, então, o parâmetro de regressão denominado $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$, cujas equações normais são dadas por $P_{\mathrm{W}_\mathrm{m}}\mathbf{Y} = \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$. Geometricamente, fica como representado na Figura 13.

O triângulo retângulo formado pelos vetores $P_{\mathrm{Im}\mathbf{X}}\mathbf{Y}$, $P_{\mathrm{W}_\mathrm{m}}\mathbf{Y}$ e pelo vetor dado pela diferença entre eles, garante que $\|P_{\mathrm{W}_\mathrm{m}}\mathbf{Y}\| \leq \|P_{\mathrm{Im}\mathbf{X}}\mathbf{Y}\|$ e, apesar deste fato não implicar em $\left\|\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}\right\| < \left\|\hat{\boldsymbol{\beta}}\right\|$, o procedimento da regressão em componentes principais é claramente conservador, no sentido que fornece estimativas menores para o vetor de parâmetros $\boldsymbol{\beta}$. De fato, será demonstrado que $\left\|\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}\right\| < \left\|\hat{\boldsymbol{\beta}}\right\|$. Resta então, obter expressões matriciais para as equações nor-

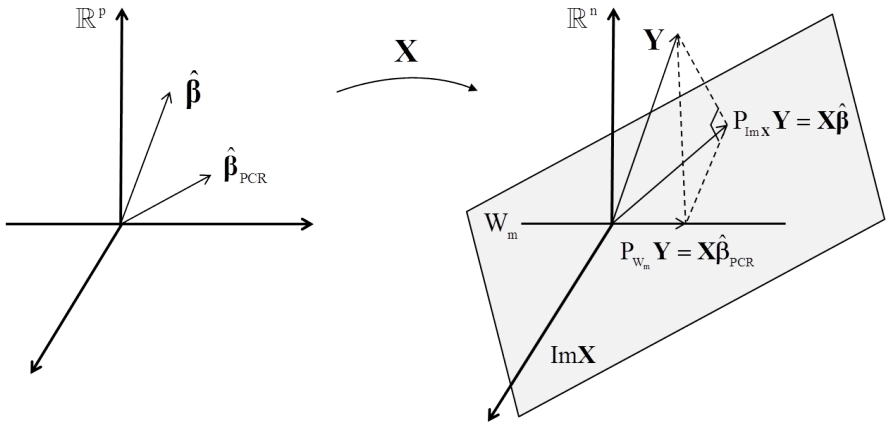


Figura 13 Projeção ortogonal de $\mathbf{Y}$ no subespaço $\mathrm{W_m}$ definindo o parâmetro de regressão $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$

mais, isto é, expressar o projetor $P_{\mathrm{W_m}}$ matricialmente.

A construção será heurística baseada no argumento: seja $f : A \to B$ uma função injetiva entre os conjuntos A e B. O que se quer é uma aplicação $g : B \to B$ que seja uma projeção em $\mathrm{Im}f$, isto é, $g\left(f\left(x\right)\right) = f\left(x\right), \forall x \in A$ e $g\left(g\left(y\right)\right) = g\left(y\right), \forall y \in B$, como mostra a Figura 14.

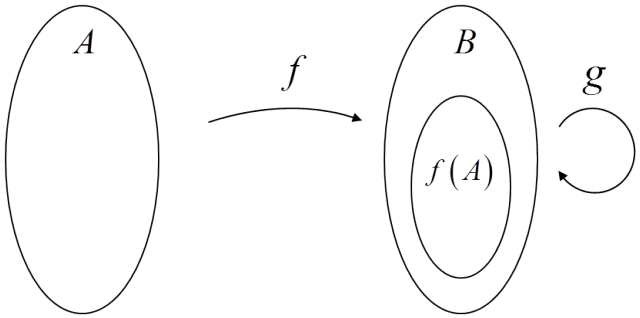


Figura 14 Representação da função injetiva f entre os conjuntos A e B e da função g como uma projeção na $\mathrm{Im}f$

A ideia básica é que a composição de uma função injetiva com uma sobrejetiva define uma bijeção. Seja $h : B \to A$ uma função sobrejetiva tal que h

restrita à imagem de f seja injetiva, $h \circ f : A \to A$ é uma bijeção (Figura 15). Define-se então,

$$g = f \circ (h \circ f)^{-1} \circ h. \tag{4}$$

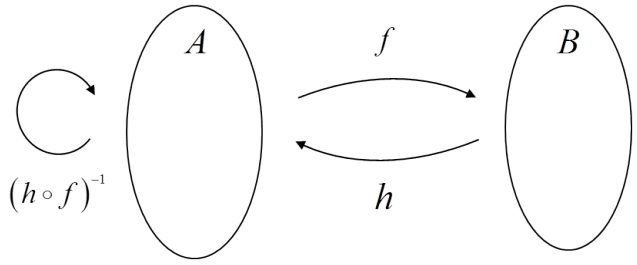


Figura 15 Representação da composição de uma função h sobrejetiva com uma função f injetiva, definindo uma bijeção $h \circ f$

$$\begin{aligned} g \circ g &= \left(f \circ (h \circ f)^{-1} \circ h\right) \circ \left(f \circ (h \circ f)^{-1} \circ h\right) \\ &= f \circ (h \circ f)^{-1} \circ (h \circ f) \circ (h \circ f)^{-1} \circ h \\ &= f \circ (h \circ f)^{-1} \circ h = g \quad \text{e} \end{aligned}$$

$$\begin{aligned} g\left(f\left(x\right)\right) &= f \circ (h \circ f)^{-1} \circ h\left(f\left(x\right)\right) \\ &= f \circ (h \circ f)^{-1} (h \circ f)\left(x\right) = f\left(x\right). \end{aligned}$$

Portanto, g é uma projeção em $\mathrm{Im}f$, como procurado. Basta agora fazer a construção de g utilizando as funções lineares convenientes para se obter a projeção linear em $\mathrm{W_m}$.

Sejam $\boldsymbol{\xi}_1 = \mathbf{X}\boldsymbol{\gamma}_1, \boldsymbol{\xi}_2 = \mathbf{X}\boldsymbol{\gamma}_2, ..., \boldsymbol{\xi}_{\mathrm{m}} = \mathbf{X}\boldsymbol{\gamma}_{\mathrm{m}}$, como na Figura 16.

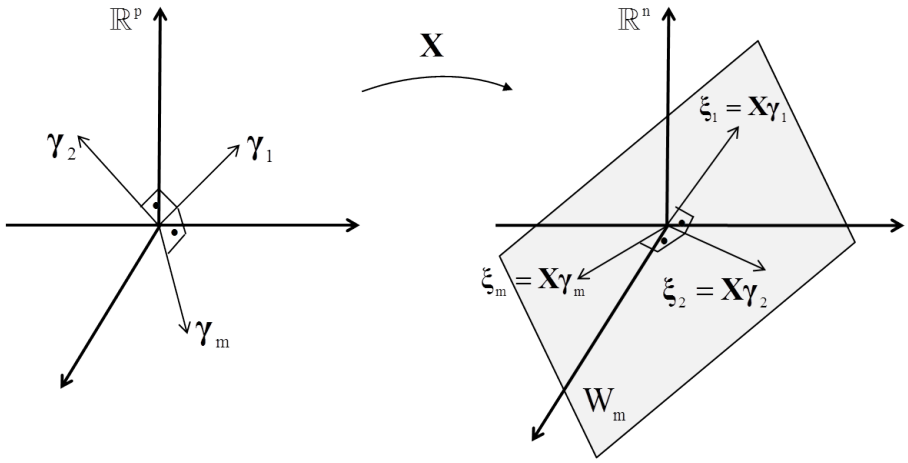


Figura 16 Representação dos vetores $\boldsymbol{\xi}_1$,$\boldsymbol{\xi}_2$, ..., $\boldsymbol{\xi}_m$ como imagem dos autovetores $\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_m$ pela matriz $\mathbf{X}$

Seja $\boldsymbol{\Xi}_m$ a matriz $n \times m$ cujas colunas são os vetores $\boldsymbol{\xi}_i$,

$$\boldsymbol{\Xi}_m = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_m] .$$

A matriz $\boldsymbol{\Xi}_m$ define uma transformação linear $\boldsymbol{\Xi}_m : \mathbb{R}^m \to \mathbb{R}^n$ tal que $\boldsymbol{\Xi}_m \mathbf{e}_i = \boldsymbol{\xi}_i$, em que $\mathbf{e}_i$ são os vetores canônicos. Desta forma, $\boldsymbol{\Xi}_m$ é injetiva de posto m e a imagem de $\boldsymbol{\Xi}_m$ é o subespaço W_m (Figura 17).

Considere agora $\boldsymbol{\Xi}_m^{'} : \mathbb{R}^n \to \mathbb{R}^m$ a transformação linear definida pela matriz transposta de $\boldsymbol{\Xi}_m$,

$$\boldsymbol{\Xi}_m^{'} = \begin{bmatrix} \boldsymbol{\xi}_1^{'} \\ \boldsymbol{\xi}_2^{'} \\ \vdots \\ \boldsymbol{\xi}_m^{'} \end{bmatrix} .$$

Como o posto linha é igual ao posto coluna, tem-se que $\boldsymbol{\Xi}_m^{'}$ é uma aplica-

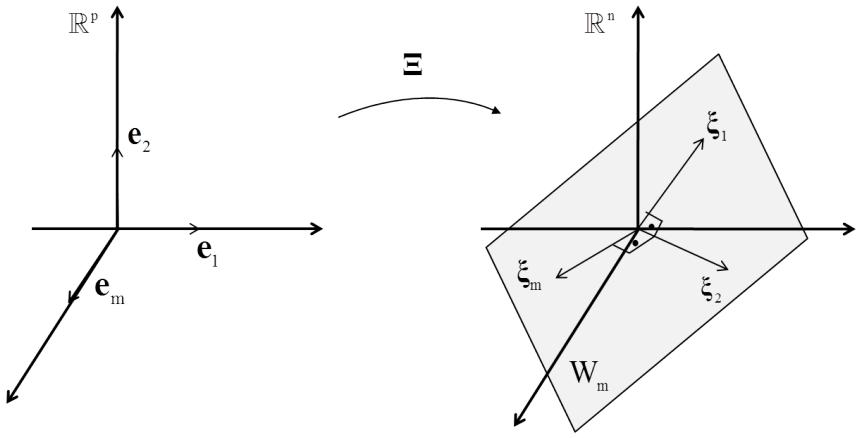


Figura 17 Representação da transformação linear $\boldsymbol{\Xi}_{\mathrm{m}} : \mathbb{R}^{\mathrm{m}} \rightarrow \mathbb{R}^{\mathrm{n}}$ em que $\boldsymbol{\Xi}_{\mathrm{m}}\mathbf{e}_{\mathrm{i}} = \boldsymbol{\xi}_{\mathrm{i}}$

ção sobrejetiva. $\boldsymbol{\Xi}'_{\mathrm{m}}$ restrito à Im$\mathbf{X}$ deve ser injetiva. De fato,

$$\begin{bmatrix} \boldsymbol{\xi}'_{\mathbf{1}} \\ \boldsymbol{\xi}'_{\mathbf{2}} \\ \vdots \\ \boldsymbol{\xi}_{\mathbf{i}} \\ \vdots \\ \boldsymbol{\xi}'_{\mathbf{m}} \end{bmatrix} [\boldsymbol{\xi}_{\mathbf{i}}] = \begin{bmatrix} 0 \\ 0 \\ \vdots \\ \|\boldsymbol{\xi}_{\mathbf{i}}\|^2 \\ \vdots \\ 0 \end{bmatrix}.$$

Podemos, então, definir o projetor

$$P_{\mathrm{W_m}} = \boldsymbol{\Xi}_{\mathbf{m}}\left(\boldsymbol{\Xi}'_{\mathbf{m}}\boldsymbol{\Xi}_{\mathbf{m}}\right)^{-1}\boldsymbol{\Xi}'_{\mathbf{m}},$$

de acordo com a equação (4). De onde seguem as equações normais do estimador

de componentes principais, $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$:

$$\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = P_{\mathrm{W_m}}\mathbf{Y} = \boldsymbol{\Xi}_{\mathrm{m}}\left(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}}\right)^{-1}\boldsymbol{\Xi}_{\mathrm{m}}^{'}\mathbf{Y}.$$

Pode-se obter uma relação entre os estimadores $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ e $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$. Seja $\boldsymbol{\Gamma}_{\mathrm{m}}$ a matriz $\mathrm{p} \times \mathrm{m}$ definida por $\boldsymbol{\Gamma}_{\mathrm{m}} = [\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{m}}]$. Tem-se que:

$$\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}} = \mathbf{X}\left[\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{m}}\right] = \left[\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, ..., \mathbf{X}\boldsymbol{\gamma}_{\mathrm{m}}\right] = \boldsymbol{\Xi}_{\mathrm{m}}.$$

Substituindo $\boldsymbol{\Xi}_{\mathrm{m}} = \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}$ na definição de $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$,

$$\begin{aligned}
\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} &= \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\left(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}}\right)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\left(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}}\right)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}(\mathbf{X}'\mathbf{X})(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\left(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}}\right)^{-1}(\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}})'\,\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{Y} \\
&= \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\left(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}}\right)^{-1}(\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}})'\,\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}.
\end{aligned}$$

Como os $\boldsymbol{\xi}_{\mathrm{i}}$ são ortogonais, $\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}} = \boldsymbol{\Lambda}_{\mathrm{m}}$ é uma matriz diagonal

$$\boldsymbol{\Lambda}_{\mathrm{m}} = \begin{bmatrix} \|\boldsymbol{\xi}_1\|^2 & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \|\boldsymbol{\xi}_{\mathrm{m}}\|^2 \end{bmatrix}.$$

Assim,

$$\begin{aligned}
&\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\left(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Xi}_{\mathrm{m}}\right)^{-1}(\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}})' \\
&= \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Lambda}_{\mathrm{m}}^{-1}(\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}})'
\end{aligned}$$

$$= [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \begin{bmatrix} \|\boldsymbol{\xi}_1\|^{-2} & 0 & 0 \\ 0 & \ddots & 0 \\ 0 & 0 & \|\boldsymbol{\xi}_{\mathrm{m}}\|^{-2} \end{bmatrix} \begin{bmatrix} \boldsymbol{\xi}_1^{'} \\ \boldsymbol{\xi}_2^{'} \\ \vdots \\ \boldsymbol{\xi}_{\mathrm{m}}^{'} \end{bmatrix}$$

$$= [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \begin{bmatrix} \boldsymbol{\xi}_1^{'} \Big/ \|\boldsymbol{\xi}_1\|^2 \\ \boldsymbol{\xi}_2^{'} \Big/ \|\boldsymbol{\xi}_2\|^2 \\ \vdots \\ \boldsymbol{\xi}_{\mathrm{m}}^{'} \Big/ \|\boldsymbol{\xi}_{\mathrm{m}}\|^2 \end{bmatrix}.$$

Portanto, $\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \begin{bmatrix} \boldsymbol{\xi}_1^{'} \Big/ \|\boldsymbol{\xi}_1\|^2 \\ \boldsymbol{\xi}_2^{'} \Big/ \|\boldsymbol{\xi}_2\|^2 \\ \vdots \\ \boldsymbol{\xi}_{\mathrm{m}}^{'} \Big/ \|\boldsymbol{\xi}_{\mathrm{m}}\|^2 \end{bmatrix} \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}.$

Utilizando a decomposição de $\mathbf{X}$ da forma:

$$\mathbf{X} = \boldsymbol{\xi}_1\boldsymbol{\gamma}_1^{'} + \boldsymbol{\xi}_2\boldsymbol{\gamma}_2^{'} + ... + \boldsymbol{\xi}_{\mathrm{p}}\boldsymbol{\gamma}_{\mathrm{p}}^{'},$$

$$[\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \begin{bmatrix} \boldsymbol{\xi}_1^{'} \Big/ \|\boldsymbol{\xi}_1\|^2 \\ \boldsymbol{\xi}_2^{'} \Big/ \|\boldsymbol{\xi}_2\|^2 \\ \vdots \\ \boldsymbol{\xi}_{\mathrm{m}}^{'} \Big/ \|\boldsymbol{\xi}_{\mathrm{m}}\|^2 \end{bmatrix} \left(\boldsymbol{\xi}_1\boldsymbol{\gamma}_1^{'} + \boldsymbol{\xi}_2\boldsymbol{\gamma}_2^{'} + ... + \boldsymbol{\xi}_{\mathrm{p}}\boldsymbol{\gamma}_{\mathrm{p}}^{'}\right) \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

$$= [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \left\{ \sum_{\mathrm{i}=1}^{\mathrm{p}} \begin{bmatrix} \boldsymbol{\xi}_1^{'} \Big/ \|\boldsymbol{\xi}_1\|^2 \\ \boldsymbol{\xi}_2^{'} \Big/ \|\boldsymbol{\xi}_2\|^2 \\ \vdots \\ \boldsymbol{\xi}_{\mathrm{m}}^{'} \Big/ \|\boldsymbol{\xi}_{\mathrm{m}}\|^2 \end{bmatrix} \boldsymbol{\xi}_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}}^{'} \right\} \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

$$= [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \left\{ \sum_{\mathrm{i}=1}^{\mathrm{p}} \frac{\boldsymbol{\xi}_{\mathrm{i}}^{'}}{\|\boldsymbol{\xi}_{\mathrm{i}}\|^2} \boldsymbol{\xi}_{\mathrm{i}} \boldsymbol{\gamma}_{\mathrm{i}}^{'} \right\} \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

$$= [\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{m}}] \, \boldsymbol{\Gamma}_{\mathrm{m}}^{'} \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$$

$$= \mathbf{X} \boldsymbol{\Gamma}_{\mathrm{m}} \boldsymbol{\Gamma}_{\mathrm{m}}^{'} \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}.$$

Logo, $\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = \mathbf{X}\left(\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\right)\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \Rightarrow \hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = \left(\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\right)\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$. Como os vetores $\boldsymbol{\gamma}_{\mathrm{i}}$ são ortonormais, $\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\boldsymbol{\Gamma}_{\mathrm{m}} = \mathbf{I}_{\mathrm{m}}$ e $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = \boldsymbol{\Gamma}_{\mathrm{m}}\left(\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\boldsymbol{\Gamma}_{\mathrm{m}}\right)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\,\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, portanto, $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ é obtido como a projeção ortogonal do vetor $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ no subespaço gerado por $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{m}}\}$, como observado na Figura 18.

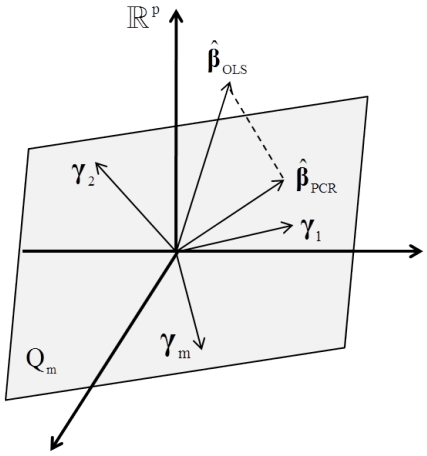


Figura 18 Representação do parâmetro $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ como a projeção ortogonal do vetor $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ no subespaço gerado por $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{m}}\}$

Em particular, fica demonstrado que $\left\|\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}\right\| < \left\|\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\right\|$, logo, $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ é um estimador de encolhimento.

Outra maneira mais geométrica de se ver esta relação é: considerando $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{p}}\}$ como base no domínio e $\{\boldsymbol{\xi}_1, \boldsymbol{\xi}_2, ..., \boldsymbol{\xi}_{\mathrm{p}}\}$ como base no contra-

domínio. A transformação linear $\mathbf{X}$ se expressa como a matriz identidade

$$\mathbf{X}\left(\sum_{\mathrm{i}=1}^{\mathrm{p}} a_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}}\right) = \sum_{\mathrm{i}=1}^{\mathrm{p}} a_{\mathrm{i}}\boldsymbol{\xi}_{\mathrm{i}}.$$

Projetar $\mathbf{Y}$ ortogonalmente na Im$\mathbf{X}$ e depois projetar ortogonalmente em $\mathrm{W_m}$ é o mesmo que projetar ortogonalmente $\mathbf{Y}$ em $\mathrm{W_m}$. A demonstração desse fato segue da aplicação do Teorema de Pitágoras nos vários triângulos retângulos da Figura 19.

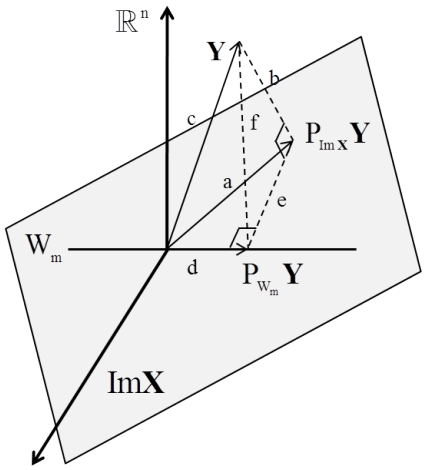


Figura 19 Projeção do vetor de dados $\mathbf{Y}$ no subespaço gerado pelos m primeiros componentes principais

Portanto, se $\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} = P_{\mathrm{Im}\mathbf{X}}\mathbf{Y} \Rightarrow \mathbf{X}\left(\sum a_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}}\right) = \sum a_{\mathrm{i}}\boldsymbol{\xi}_{\mathrm{i}}$, projetar $P_{\mathrm{Im}\mathbf{X}}\mathbf{Y}$ em $\mathrm{W_m}$ é simplesmente tomar as m primeiras coordenadas, $P_{\mathrm{W_m}}\mathbf{Y} = \sum_{\mathrm{i}=1}^{\mathrm{m}} a_{\mathrm{i}}\boldsymbol{\xi}_{\mathrm{i}}$ e $\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = P_{\mathrm{W_m}}\mathbf{Y}$ implica que $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = \sum_{\mathrm{i}=1}^{\mathrm{m}} a_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}}$ e assim, $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ é a projeção de $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ no subespaço gerado por $\{\boldsymbol{\gamma}_1, \boldsymbol{\gamma}_2, ..., \boldsymbol{\gamma}_{\mathrm{m}}\}$.

A matriz de variâncias e covariâncias de $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ é:

$$\begin{aligned}\operatorname{cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}\right) &= \operatorname{cov}\left(\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\right) \\ &= \boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\operatorname{cov}\left(\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}\right)\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'} \\ &= \sigma^2\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\left(\mathbf{X}'\mathbf{X}\right)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}.\end{aligned}$$

Matricialmente, tem-se:

$$\begin{aligned}&[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_{\mathrm{m}}]\begin{bmatrix}\boldsymbol{\gamma}_1^{'}\\ \boldsymbol{\gamma}_2^{'}\\ \vdots\\ \boldsymbol{\gamma}_{\mathrm{m}}^{'}\end{bmatrix}\left(\mathbf{X}'\mathbf{X}\right)^{-1}[\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_{\mathrm{m}}]\begin{bmatrix}\boldsymbol{\gamma}_1^{'}\\ \boldsymbol{\gamma}_2^{'}\\ \vdots\\ \boldsymbol{\gamma}_{\mathrm{m}}^{'}\end{bmatrix}\\ &= [\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_{\mathrm{m}}]\begin{bmatrix}\boldsymbol{\gamma}_1^{'}\\ \boldsymbol{\gamma}_2^{'}\\ \vdots\\ \boldsymbol{\gamma}_{\mathrm{m}}^{'}\end{bmatrix}\left[\frac{1}{\lambda_1}\boldsymbol{\gamma}_1,\frac{1}{\lambda_2}\boldsymbol{\gamma}_2,\ldots,\frac{1}{\lambda_{\mathrm{m}}}\boldsymbol{\gamma}_{\mathrm{m}}\right]\begin{bmatrix}\boldsymbol{\gamma}_1^{'}\\ \boldsymbol{\gamma}_2^{'}\\ \vdots\\ \boldsymbol{\gamma}_{\mathrm{m}}^{'}\end{bmatrix}\\ &= [\boldsymbol{\gamma}_1,\boldsymbol{\gamma}_2,\ldots,\boldsymbol{\gamma}_{\mathrm{m}}]\begin{bmatrix}1/\lambda_1 & 0 & 0\\ 0 & \ddots & 0\\ 0 & 0 & 1/\lambda_{\mathrm{m}}\end{bmatrix}\begin{bmatrix}\boldsymbol{\gamma}_1^{'}\\ \boldsymbol{\gamma}_2^{'}\\ \vdots\\ \boldsymbol{\gamma}_{\mathrm{m}}^{'}\end{bmatrix}\\ &= \left[\frac{\boldsymbol{\gamma}_1}{\lambda_1},\frac{\boldsymbol{\gamma}_2}{\lambda_2},\ldots,\frac{\boldsymbol{\gamma}_{\mathrm{m}}}{\lambda_{\mathrm{m}}}\right]\begin{bmatrix}\boldsymbol{\gamma}_1^{'}\\ \boldsymbol{\gamma}_2^{'}\\ \vdots\\ \boldsymbol{\gamma}_{\mathrm{m}}^{'}\end{bmatrix}.\end{aligned}$$

Essa matriz possui $\boldsymbol{\gamma}_{\mathrm{i}}$ como autovetores, pois,

$$\left(\left[\frac{\boldsymbol{\gamma}_1}{\lambda_1}, \frac{\boldsymbol{\gamma}_2}{\lambda_2}, \ldots, \frac{\boldsymbol{\gamma}_{\mathrm{m}}}{\lambda_{\mathrm{m}}}\right]\begin{bmatrix}\boldsymbol{\gamma}_1' \\ \boldsymbol{\gamma}_2' \\ \vdots \\ \boldsymbol{\gamma}_{\mathrm{m}}'\end{bmatrix}\right)(\boldsymbol{\gamma}_{\mathrm{i}})$$

$$= \left[\frac{\boldsymbol{\gamma}_1}{\lambda_1}, \frac{\boldsymbol{\gamma}_2}{\lambda_2}, \ldots, \frac{\boldsymbol{\gamma}_{\mathrm{m}}}{\lambda_{\mathrm{m}}}\right]\left(\begin{bmatrix}\boldsymbol{\gamma}_1' \\ \boldsymbol{\gamma}_2' \\ \vdots \\ \boldsymbol{\gamma}_{\mathrm{m}}'\end{bmatrix}(\boldsymbol{\gamma}_{\mathrm{i}})\right)$$

$$= \left[\frac{\boldsymbol{\gamma}_1}{\lambda_1}, \frac{\boldsymbol{\gamma}_2}{\lambda_2}, \ldots, \frac{\boldsymbol{\gamma}_{\mathrm{m}}}{\lambda_{\mathrm{m}}}\right]\begin{bmatrix}0 \\ 1 \\ \vdots \\ 0\end{bmatrix} = \frac{1}{\lambda_{\mathrm{i}}}\boldsymbol{\gamma}_{\mathrm{i}}$$

e, portanto, a variância total de $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}}$ é $\sum_{\mathrm{i}=1}^{\mathrm{m}} \frac{1}{\lambda_{\mathrm{i}}}$.

Para o caso em que $\operatorname{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}$, o estimador de menor variância é dado pelo estimador de Gauss-Markov (PEREIRA, 2013):

$$\begin{aligned}\hat{\boldsymbol{\beta}}_{\mathrm{GM}} =& \left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \\ \mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{GM}} =& \mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}.\end{aligned}$$

Assim, $\hat{\mathbf{Y}}$ é dado pela projeção de $\mathbf{Y}$, em que $\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$ é o projetor.

No caso usual, em que $\operatorname{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$, as projeções nos subespaços W_{m} que definem $\hat{\boldsymbol{\beta}}_{\mathrm{PCR,m}}$ são ortogonais, e quando $\mathrm{m} = \mathrm{p}$, $\hat{\boldsymbol{\beta}}_{\mathrm{PCR}} = \hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$. Portanto, no caso $\operatorname{cov}(\mathbf{Y}) = \boldsymbol{\Sigma}$, é natural utilizar a mesma projeção. O projetor $\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$ é idempotente, mas não é simétrico, logo, não é um pro-

jetor ortogonal. No entanto, em relação ao produto interno $\langle\langle \mathbf{x}, \mathbf{y} \rangle\rangle = \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$, esse projetor é simétrico. Assim, o projetor $\mathbf{X}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}$ é ortogonal em relação ao produto interno $\langle\langle \mathbf{x}, \mathbf{y} \rangle\rangle = \mathbf{x}'\boldsymbol{\Sigma}^{-1}\mathbf{y}$.

Deseja-se então, maximizar a variância da projeção de $\mathbf{Y}$ na direção de $\mathbf{X}\boldsymbol{\beta}$, isto é, maximizar $\mathrm{var}\left(\langle\langle \mathbf{X}\boldsymbol{\beta}, \mathbf{Y} \rangle\rangle\right)$, restrito a $\boldsymbol{\beta}'\boldsymbol{\beta} = 1$:

$$\begin{aligned}
\mathrm{var}\left((\mathbf{X}\beta)'\,\boldsymbol{\Sigma}^{-1}\mathbf{Y}\right) =& \mathrm{var}\left(\boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y}\right) \\
=& \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathrm{cov}\left(\mathbf{Y}\right)\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \\
=& \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\boldsymbol{\Sigma}\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta} \\
=& \boldsymbol{\beta}'\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\beta}.
\end{aligned}$$

Sejam $\boldsymbol{\gamma}_{\mathrm{i}}$ autovetores de $\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}$:

$$\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\left(\boldsymbol{\gamma}_{\mathrm{i}}\right) = \lambda_{\mathrm{i}}\boldsymbol{\gamma}_{\mathrm{i}} \ \text{ e } \ \boldsymbol{\gamma}_{\mathrm{i}}^{'}\boldsymbol{\gamma}_{\mathrm{i}} = 1.$$

Os vetores $\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}$ são ortogonais, pois

$$\begin{aligned}
\langle\langle \mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}, \mathbf{X}\boldsymbol{\gamma}_{\mathrm{j}} \rangle\rangle =& \left(\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}}\right)'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\gamma}_{\mathrm{j}} = \boldsymbol{\gamma}_{\mathrm{i}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\gamma}_{\mathrm{i}} \\
=& \boldsymbol{\gamma}_{\mathrm{i}}^{'}\left(\lambda_{\mathrm{j}}\boldsymbol{\gamma}_{\mathrm{j}}\right) = \lambda_{\mathrm{j}}\boldsymbol{\gamma}_{\mathrm{i}}^{'}\boldsymbol{\gamma}_{\mathrm{j}} = 0.
\end{aligned}$$

Definindo o subespaço $\mathrm{W_m} = span\left\{\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3, ..., \mathbf{X}\boldsymbol{\gamma}_{\mathrm{m}}\right\}$, tem-se:

$$\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR,m}} = P_{\mathrm{W_m}}\mathbf{Y},$$

em que $P_{\mathrm{W_m}}$ é a projeção ortogonal, em relação ao produto interno $\langle\langle\ ,\ \rangle\rangle$, no subespaço gerado por $\mathbf{X}\boldsymbol{\gamma}_1, \mathbf{X}\boldsymbol{\gamma}_2, \mathbf{X}\boldsymbol{\gamma}_3, ..., \mathbf{X}\boldsymbol{\gamma}_{\mathrm{m}}$.

Da mesma forma, como feito para o caso em que $\mathrm{cov}\left(\mathbf{Y}\right) = \sigma^2\mathbf{I}$, obtém-

se:

$$P_{\mathrm{W_m}} = \boldsymbol{\Xi}_{\mathrm{m}}\Big(\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Sigma}^{-1}\boldsymbol{\Xi}_{\mathrm{m}}\Big)^{-1}\boldsymbol{\Xi}_{\mathrm{m}}^{'}\boldsymbol{\Sigma}^{-1}.$$

Como $\boldsymbol{\Xi}_{\mathrm{m}} = \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}$, segue que

$$\begin{aligned}
\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PCR,m}} =& \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big(\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \\
=& \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big(\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{Y} \\
=& \mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big(\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\hat{\boldsymbol{\beta}}_{\mathrm{GM}}.
\end{aligned}$$

Logo,

$$\hat{\boldsymbol{\beta}}_{\mathrm{PCR,m}} = \boldsymbol{\Gamma}_{\mathrm{m}}\Big(\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{m}}\Big)^{-1}\boldsymbol{\Gamma}_{\mathrm{m}}^{'}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\hat{\boldsymbol{\beta}}_{\mathrm{GM}}.$$

Observe que, se $\mathrm{m} = \mathrm{p}$, $\boldsymbol{\Gamma}_{\mathrm{p}}$ é uma matriz $\mathrm{p} \times \mathrm{p}$ ortogonal. Assim:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}_{\mathrm{PCR,p}} =& \boldsymbol{\Gamma}_{\mathrm{p}}\Big(\boldsymbol{\Gamma}_{\mathrm{p}}^{'}\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\boldsymbol{\Gamma}_{\mathrm{p}}\Big)^{-1}\boldsymbol{\Gamma}_{\mathrm{p}}^{'}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\hat{\boldsymbol{\beta}}_{\mathrm{GM}} \\
=& \boldsymbol{\Gamma}_{\mathrm{p}}(\boldsymbol{\Gamma}_{\mathrm{p}})^{-1}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)^{-1}\Big(\boldsymbol{\Gamma}_{\mathrm{p}}^{'}\Big)^{-1}\boldsymbol{\Gamma}_{\mathrm{p}}^{'}\left(\mathbf{X}'\boldsymbol{\Sigma}^{-1}\mathbf{X}\right)\hat{\boldsymbol{\beta}}_{\mathrm{GM}} \\
=& \hat{\boldsymbol{\beta}}_{\mathrm{GM}}.
\end{aligned}$$

4.3 Regressão por Quadrados Mínimos Parciais: uma abordagem geométrica

4.3.1 PLS Populacional

Esta seção tem como referência o artigo Helland (1990).

Seja uma situação em que se quer predizer uma variável aleatória y a partir de p variáveis aleatórias $\mathrm{x}_1, \cdots, \mathrm{x}_\mathrm{p}$:

$$\mathrm{E}\begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix} = \begin{bmatrix} \mathbf{0} \\ 0 \end{bmatrix} \quad \mathrm{e} \quad \mathrm{cov}\begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\Sigma} & \sigma \\ \sigma' & \sigma_y^2 \end{bmatrix},$$

$$\text{em que, } \mathbf{x} = \begin{bmatrix} \mathrm{x}_1 \\ \vdots \\ \mathrm{x}_\mathrm{p} \end{bmatrix}, \quad \boldsymbol{\sigma} = \mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right) = \begin{bmatrix} \mathrm{cov}\left(\mathrm{x}_1, \mathrm{y}\right) \\ \vdots \\ \mathrm{cov}\left(\mathrm{x}_\mathrm{p}, \mathrm{y}\right) \end{bmatrix} \quad \mathrm{e} \quad \boldsymbol{\Sigma} = \mathrm{cov}\left(\mathbf{x}\right).$$

Uma pergunta interessante: Qual é o vetor $\boldsymbol{\eta} \in \mathbb{R}^\mathrm{p}$ que define a direção de maior covariância com a variável y? Isto é: se $\mathbf{Z} = \begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix}$, deseja-se o vetor $\boldsymbol{\eta}$ em $\mathbb{R}^\mathrm{p}$, $\|\boldsymbol{\eta}\| = 1$, que, ao se projetar o vetor aleatório $\mathbf{Z}$ na direção de $\boldsymbol{\eta}$, fornece a maior covariância com $\mathbf{e}_{\mathrm{p}+1} \cdot \mathbf{Z} = \mathrm{y}$. Tem-se, então, o problema de maximização $\max_{\|\boldsymbol{\eta}\|=1} \mathrm{cov}\left(\mathbf{e}_{\mathrm{p}+1} \cdot \mathbf{Z}, \boldsymbol{\eta}^* \cdot \mathbf{Z}\right)$, em que $\boldsymbol{\eta}^* = \begin{bmatrix} \boldsymbol{\eta} \\ 0 \end{bmatrix}$. Utilizando o Teorema 1 sobre

vetores aleatórios, tem-se:

$$\operatorname{cov}\left(\mathbf{e}_{\mathrm{p}+1}\cdot\mathbf{Z},\begin{bmatrix}\boldsymbol{\eta}\\0\end{bmatrix}\cdot\mathbf{Z}\right)=\left\langle\begin{pmatrix}\mathbf{0}\\1\end{pmatrix},\begin{bmatrix}\boldsymbol{\Sigma}&\boldsymbol{\sigma}\\\boldsymbol{\sigma}'&\sigma_{\mathrm{y}}^2\end{bmatrix}\begin{bmatrix}\boldsymbol{\eta}\\0\end{bmatrix}\right\rangle$$
$$=\left\langle\begin{bmatrix}\mathbf{0}\\1\end{bmatrix},\begin{bmatrix}\boldsymbol{\Sigma}\boldsymbol{\eta}\\\boldsymbol{\sigma}'\boldsymbol{\eta}\end{bmatrix}\right\rangle=\boldsymbol{\sigma}'\boldsymbol{\eta}=\|\boldsymbol{\eta}\|\,\|\boldsymbol{\sigma}\|\cos\theta=\|\boldsymbol{\sigma}\|\cos\theta$$

Para se ter covariância máxima, o ângulo deve ser zero ($\cos 0=1$). Logo, $\boldsymbol{\eta}$ deve estar na mesma direção de $\boldsymbol{\sigma}$. Assim, a direção que fornece a covariância máxima é dada pelo próprio vetor de covariâncias $\boldsymbol{\sigma}$. A projeção de $\mathbf{Z}$ em $\boldsymbol{\sigma}$ é $\boldsymbol{\sigma}'\mathbf{Z}=\boldsymbol{\sigma}'\mathbf{x}$.

A ideia é, como uma primeira aproximação, substituir a variável aleatória de interesse y por um múltiplo da variável aleatória preditora $\boldsymbol{\sigma}'\mathbf{x}$. Usando como critério de predição o Erro Quadrático Médio (EQM), devemos minimizar a expressão $\mathrm{E}\left[\left(\mathrm{y}-\alpha\left(\boldsymbol{\sigma}'\mathbf{x}\right)\right)^2\right]$:

$$\begin{aligned}&\mathrm{E}\left[\left(\mathrm{y}-\alpha\left(\boldsymbol{\sigma}'\mathbf{x}\right)\right)^2\right]\\&=\mathrm{E}\left[\mathrm{y}^2\right]-2\alpha\mathrm{E}\left[\mathrm{y}\left(\boldsymbol{\sigma}'\mathbf{x}\right)\right]+\alpha^2\mathrm{E}\left[\left(\boldsymbol{\sigma}'\mathbf{x}\right)^2\right]\\&=\sigma_y^2-2\alpha\operatorname{cov}\left(\mathrm{y},\boldsymbol{\sigma}'\mathbf{x}\right)+\alpha^2\operatorname{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)\end{aligned}$$

Derivando e igualando a zero, obtém-se:

$$\begin{aligned}&-2\operatorname{cov}\left(\mathrm{y},\boldsymbol{\sigma}'\mathbf{x}\right)+2\alpha\operatorname{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)=0\\&\alpha=\tfrac{\operatorname{cov}(\mathrm{y},\boldsymbol{\sigma}'\mathbf{x})}{\operatorname{var}(\boldsymbol{\sigma}'\mathbf{x})}\end{aligned}$$

Note que, a escala da variável y é diferente das escalas das variáveis $\mathrm{x_i}$, e com esta ponderação pela $\operatorname{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)$, este problema fica minimizado.

Assim, a variável que prediz y com menor erro quadrático médio é a va-

riável $\frac{\mathrm{cov}(\mathrm{y},\boldsymbol{\sigma}'\mathbf{x})}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\boldsymbol{\sigma}'\mathbf{x}$.

Denominando $\mathrm{q}_1 = \mathrm{cov}\left(\mathrm{y}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right)$, tem-se:

$$\begin{aligned}
&\mathrm{cov}\left(\mathrm{y}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) \\
&= \mathrm{cov}\left(\mathbf{e}_{\mathrm{p}+1} \cdot \begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix}, \begin{bmatrix} \frac{\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \\ 0 \end{bmatrix} \cdot \begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix}\right) \\
&= \left\langle \mathbf{e}_{\mathrm{p}+1}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}' & \sigma_{\mathrm{y}}^2 \end{bmatrix} \begin{bmatrix} \frac{\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \\ 0 \end{bmatrix} \right\rangle \\
&= \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} = \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\mathrm{cov}(\mathbf{x})\boldsymbol{\sigma}} = \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma}\boldsymbol{\sigma}}.
\end{aligned}$$

Assim, a parte de y que não foi explicada por $\frac{\mathrm{cov}(\mathrm{y},\boldsymbol{\sigma}'\mathbf{x})}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\boldsymbol{\sigma}'\mathbf{x}$, denominada por f_1, é dada por:

$$\mathrm{f}_1 = \mathrm{y} - \frac{\mathrm{cov}\left(\mathrm{y},\boldsymbol{\sigma}'\mathbf{x}\right)}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)}\boldsymbol{\sigma}'\mathbf{x} = \mathrm{y} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)}\boldsymbol{\sigma}'\mathbf{x} = \mathrm{y} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma}\boldsymbol{\sigma}}\boldsymbol{\sigma}'\mathbf{x}.$$

No sentido de se obter uma sequência de variáveis preditoras, considera-se como primeira variável preditora de y, apenas a variável $\frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}$. Assim, o vetor de correlações $\boldsymbol{\sigma} = \mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right)$, renomeado por $\mathbf{w}_1$, pode ser substituído pelo vetor $\mathrm{cov}\left(\mathbf{x}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right)$. Denomina-se este vetor por:

$$\mathbf{p}_1 = \frac{\mathrm{cov}\left(\mathbf{x}, \boldsymbol{\sigma}'\mathbf{x}\right)}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)} = \begin{bmatrix} \frac{\mathrm{cov}(\mathrm{x}_1,\boldsymbol{\sigma}'\mathbf{x})}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \\ \vdots \\ \frac{\mathrm{cov}(\mathrm{x}_\mathrm{p},\boldsymbol{\sigma}'\mathbf{x})}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \end{bmatrix}.$$

A i-ésima coordenada do vetor $\mathbf{p}_1$ é dada por:

$$
\begin{aligned}
&\mathrm{x_i} = \mathbf{e_i} \cdot \mathbf{Z} \\
&\boldsymbol{\sigma}'\mathbf{x} = (\boldsymbol{\sigma}, 0) \cdot \mathbf{Z} \\
&\mathrm{cov}\left(\mathrm{x_i}, \boldsymbol{\sigma}'\mathbf{x}\right) = \mathrm{cov}\left(\mathbf{e_i} \cdot \mathbf{Z}, \boldsymbol{\sigma} \cdot \mathbf{Z}\right) \\
&= \left\langle \mathbf{e_i}, \begin{bmatrix} \boldsymbol{\Sigma} & \boldsymbol{\sigma} \\ \boldsymbol{\sigma}' & \sigma_{\mathrm{y}}^2 \end{bmatrix} \begin{bmatrix} \boldsymbol{\sigma} \\ 0 \end{bmatrix} \right\rangle = \left\langle \mathbf{e_i}, \begin{bmatrix} \boldsymbol{\Sigma}\boldsymbol{\sigma} \\ \boldsymbol{\sigma}'\boldsymbol{\sigma} \end{bmatrix} \right\rangle \\
&= \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{cov}\left(\mathrm{x_i}, \mathrm{x_j}\right) \sigma_{\mathrm{j}} = \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{cov}\left(\mathrm{x_i}, \mathrm{x_j}\right) \mathrm{cov}\left(\mathrm{x_j}, \mathrm{y}\right) \\
&\Rightarrow \mathrm{cov}\left(\mathrm{x_i}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) = \frac{1}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{cov}\left(\mathrm{x_i}, \mathrm{x_j}\right) \mathrm{cov}\left(\mathrm{x_j}, \mathrm{y}\right).
\end{aligned}
$$

Ou ainda,

$$
\begin{aligned}
&\mathrm{cov}\left(\mathrm{x_i}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) = \mathrm{cov}\left(\mathrm{x_i}, \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{x_j} \frac{\mathrm{cov}\left(\mathrm{x_j}, \mathrm{y}\right)}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) \\
&= \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{cov}\left(\mathrm{x_i}, \mathrm{x_j} \frac{\mathrm{cov}\left(\mathrm{x_j}, \mathrm{y}\right)}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}\right) = \sum_{\mathrm{j}=1}^{\mathrm{p}} \frac{\mathrm{cov}\left(\mathrm{x_j}, \mathrm{y}\right)}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \mathrm{cov}\left(\mathrm{x_i}, \mathrm{x_j}\right) \\
&= \sum_{\mathrm{j}=1}^{\mathrm{p}} \frac{\mathrm{cov}\left(\mathrm{x_i}, \mathrm{x_j}\right)}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \sigma_{\mathrm{j}} = \frac{1}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} (\boldsymbol{\Sigma}\boldsymbol{\sigma})_{\mathrm{i}}.
\end{aligned}
$$

Logo, $\mathbf{p_1} = \frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})}$.

A componente de $\mathbf{x}$ perpendicular ao vetor $\boldsymbol{\sigma}$ não foi utilizada na predição de y. Deve-se então, encontrar α tal que $\mathbf{x} - \alpha\mathbf{p}_1$ seja perpendicular à $\boldsymbol{\sigma}$:

$$
\begin{aligned}
&\langle \boldsymbol{\sigma}, \mathbf{x} - \alpha\mathbf{p}_1 \rangle = \langle \boldsymbol{\sigma}, \mathbf{x} \rangle - \alpha \langle \boldsymbol{\sigma}, \mathbf{p}_1 \rangle \\
&\Rightarrow \langle \boldsymbol{\sigma}, \mathbf{x} \rangle - \alpha \left\langle \boldsymbol{\sigma}, \frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \right\rangle = 0 \\
&\Rightarrow \langle \boldsymbol{\sigma}, \mathbf{x} \rangle - \alpha \frac{\boldsymbol{\sigma}'\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} = 0 \\
&\Rightarrow \boldsymbol{\sigma}'\mathbf{x} - \alpha = 0 \\
&\Rightarrow \alpha = \boldsymbol{\sigma}'\mathbf{x}.
\end{aligned}
$$

Desta forma, denominando $\mathrm{t}_1 = \boldsymbol{\sigma}'\mathbf{x}$, tem-se o vetor $\mathbf{e}_1 = \mathbf{x} - \mathrm{t}_1\mathbf{p}_1$, $\mathbf{e}_1 \perp \boldsymbol{\sigma}$. Note que, a preditora de y pode ser reescrita como $\frac{\mathrm{t}_1}{\mathrm{var}(\mathrm{t}_1)}$ e

$$\mathbf{e}_1 = \mathbf{x} - (\boldsymbol{\sigma}'\mathbf{x})\frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\mathrm{var}(\boldsymbol{\sigma}'\mathbf{x})} \ .$$

O processo agora é repetido, substituindo-se y por sua parte não explicada f_1 e $\mathbf{x}$ por sua componente ainda não utilizada $\mathbf{e}_1$. A direção de maior covariância entre f_1 e os vetores aleatórios no subespaço perpendicular a $\boldsymbol{\sigma}$, é dada pelo vetor $\mathbf{w}_2$:

$$\begin{aligned}
\mathbf{w}_2 &= \mathrm{cov}\left(\mathbf{e}_1, \mathrm{f}_1\right) = \mathrm{cov}\left(\mathbf{e}_1, \mathrm{y} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)}\boldsymbol{\sigma}'\mathbf{x}\right) \\
&= \mathrm{cov}\left(\mathbf{e}_1, \mathrm{y}\right) - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)}\mathrm{cov}\left(\mathbf{e}_1, \boldsymbol{\sigma}'\mathbf{x}\right) \overset{\mathbf{e}_1 \perp \boldsymbol{\sigma}}{=} \mathrm{cov}\left(\mathbf{e}_1, \mathrm{y}\right) \\
&= \mathrm{cov}\left(\mathbf{x} - \mathbf{p}_1\mathrm{t}_1, \mathrm{y}\right) = \mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right) - \mathrm{cov}\left(\mathbf{p}_1\mathrm{t}_1, \mathrm{y}\right) \\
&= \boldsymbol{\sigma} - \mathbf{p}_1\mathrm{cov}\left(\boldsymbol{\sigma}'\mathbf{x}, \mathrm{y}\right) = \boldsymbol{\sigma} - \frac{\boldsymbol{\Sigma}\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma}\boldsymbol{\sigma}}\boldsymbol{\sigma}'\boldsymbol{\sigma}.
\end{aligned}$$

Definindo $\mathrm{t}_2 = \mathbf{e}_1{}'\mathbf{w}_2$, a variável preditora de f_1 será $\frac{\mathrm{t}_2}{\mathrm{var}(\mathrm{t}_2)}$. Calculando-se a covariância entre f_1 e seu preditor $\frac{\mathrm{t}_2}{\mathrm{var}(\mathrm{t}_2)}$, temos a grandeza q_2, dada por

$$\mathrm{q}_2 = \frac{\mathrm{cov}\left(\mathrm{f}_1, \mathrm{t}_2\right)}{\mathrm{var}\left(\mathrm{t}_2\right)}.$$

Desta forma, a parte não explicada de f_1 será dada por $\mathrm{f}_2 = \mathrm{f}_1 - \mathrm{q}_2\mathrm{t}_2$.

Define-se agora, o vetor dado pela covariância entre o vetor aleatório $\mathbf{e}_1$ e a variável preditora de f_1, $\frac{\mathrm{t}_2}{\mathrm{var}(\mathrm{t}_2)}$, denominado $\mathbf{p}_2$:

$$\mathbf{p}_2 = \frac{\mathrm{cov}\left(\mathbf{e}_1, \mathrm{t}_2\right)}{\mathrm{var}\left(\mathrm{t}_2\right)}.$$

Novamente, a componente do vetor $\mathbf{e}_1$ que não foi utilizada para explicar

f_1 será dada pelo vetor $\mathbf{e}_2 = \mathbf{e}_1 - \mathbf{p}_2\mathrm{t}_2$.

Resumindo:

Para $a = 1$:

$$\mathbf{w}_1 = \mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right) = \boldsymbol{\sigma}$$

$$\mathrm{t}_1 = \mathbf{x}'\mathbf{w}_1 = \boldsymbol{\sigma}'\mathbf{x}$$

$$\mathbf{p}_1 = \mathrm{cov}\left(\mathbf{x}, \frac{\mathrm{t}_1}{\mathrm{var}\left(\mathrm{t}_1\right)}\right) = \mathrm{cov}\left(\mathbf{x}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)}\right) = \frac{\boldsymbol{\Sigma\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}$$

$$\mathrm{q}_1 = \mathrm{cov}\left(\mathrm{y}, \frac{\mathrm{t}_1}{\mathrm{var}\left(\mathrm{t}_1\right)}\right) = \mathrm{cov}\left(\mathrm{y}, \frac{\boldsymbol{\sigma}'\mathbf{x}}{\mathrm{var}\left(\boldsymbol{\sigma}'\mathbf{x}\right)}\right) = \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}$$

$$\mathbf{e}_1 = \mathbf{x} - \mathbf{p}_1\mathrm{t}_1 = \mathbf{x} - \frac{\boldsymbol{\Sigma\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\sigma}'\mathbf{x}$$

$$\mathrm{f}_1 = \mathrm{y} - \mathrm{q}_1\mathrm{t}_1 = \mathrm{y} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\sigma}'\mathbf{x}$$

Para $a = 2$:

$$\mathrm{t}_2 = \mathbf{e}_1{}'\mathbf{w}_2 = \mathbf{x}'\left(\boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\Sigma\sigma}\right) = \mathbf{x}'\boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\mathrm{x}'\boldsymbol{\Sigma\sigma}$$

$$\mathrm{var}\left(\mathrm{t}_2\right) = \left(\boldsymbol{\sigma} - \frac{\|\boldsymbol{\sigma}\|^2}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\Sigma\sigma}\right)' \boldsymbol{\Sigma} \left(\boldsymbol{\sigma} - \frac{\|\boldsymbol{\sigma}\|^2}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\Sigma\sigma}\right)$$

$$\mathbf{w}_2 = \mathrm{cov}\left(\mathbf{e}_1, \mathrm{f}_1\right) = \boldsymbol{\sigma} - \frac{\boldsymbol{\Sigma\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\sigma}'\boldsymbol{\sigma}$$

$$\mathbf{p}_2 = \frac{\mathrm{cov}\left(\mathbf{e}_1, \mathrm{t}_2\right)}{\mathrm{var}\left(\mathrm{t}_2\right)} = \frac{\mathrm{cov}\left(\mathbf{x}, \mathrm{t}_2\right)}{\mathrm{var}\left(\mathrm{t}_2\right)}$$

$$= \mathrm{cov}\left(\mathbf{x}, \mathbf{x}'\boldsymbol{\sigma} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\mathrm{x}'\boldsymbol{\Sigma\sigma}\right) = \boldsymbol{\Sigma\sigma} - \frac{\boldsymbol{\sigma}'\boldsymbol{\sigma}}{\boldsymbol{\sigma}'\boldsymbol{\Sigma\sigma}}\boldsymbol{\Sigma}^2\boldsymbol{\sigma}$$

$$\mathrm{q}_2 = \frac{\mathrm{cov}\left(\mathrm{f}_1, \mathrm{t}_2\right)}{\mathrm{var}\left(\mathrm{t}_2\right)}$$

$$\mathbf{e}_2 = \mathbf{e}_1 - \mathbf{p}_2\mathrm{t}_2$$

$$\mathrm{f}_2 = \mathrm{f}_1 - \mathrm{q}_2\mathrm{t}_2$$

Um algoritmo para a predição da variável aleatória de interesse y, em termos das covariáveis fica, então, definido.

Algoritmo para o PLS Populacional (HELLAND, 1990):

1. Defina os valores iniciais para $\mathbf{x}$ residual $(\mathbf{e}_a)$ e y residual (f_a) :

$$\mathbf{e}_0 = \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}$$
$$\mathrm{f}_0 = \mathrm{y} - \mu_{\mathrm{y}}$$

   Para $a = 1, 2, \cdots$, execute os passos $(2) - (4)$ abaixo:

2. Introduza *scores* t como combinações lineares do $\mathbf{x}$ residual a partir do último passo; para fazer os *scores* mais relacionados com y, use covariâncias com y residual como coeficiente/pesos $(\mathbf{w})$:

$$\mathbf{w}_a = \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-1}\right)$$
$$\mathrm{t}_a = \mathbf{e}_{a-1}{}'\mathbf{w}_a$$

3. Determine $\mathbf{x}$ *loadings* $(\mathbf{p}_a)$ e y *loading* (q_a) por quadrados mínimos:

$$\mathbf{p}_a = \frac{\mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{t}_a\right)}{\mathrm{var}\left(\mathrm{t}_a\right)}$$
$$\mathrm{q}_a = \frac{\mathrm{cov}\left(\mathrm{f}_{a-1}, \mathrm{t}_a\right)}{\mathrm{var}\left(\mathrm{t}_a\right)}$$

4. Encontre novos resíduos:

$$\mathbf{e}_a = \mathbf{e}_{a-1} - \mathbf{p}_a \mathrm{t}_a$$
$$\mathrm{f}_a = \mathrm{f}_{a-1} - \mathrm{q}_a \mathrm{t}_a$$

Deve-se observar que os vetores $\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_\mathrm{A}$ e $\mathbf{p}_1, \mathbf{p}_2, \cdots \mathbf{p}_\mathrm{A}$ são vetores populacionais. Além disso, q_a são grandezas populacionais, isto é, fixada uma determinada população, os números q_a ficam determinados. Portanto, as grandezas aleatórias são os vetores $\mathbf{e}_a$, os pesos t_a e as variáveis f_a. Duas iterações do algoritmo estão descritas na Figura 20.

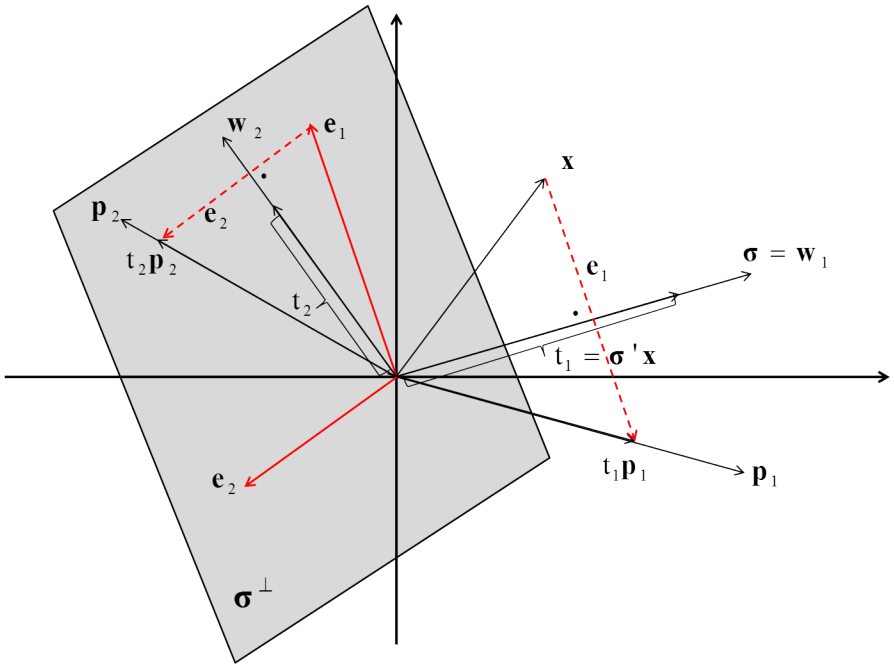


Figura 20 Representação geomética do algoritmo PLS Populacional com duas iterações

Da construção geométrica do algoritmo, seguem os seguintes fatos fundamentais:

1. Os vetores $\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_\mathrm{A}$ são ortogonais.

2. As variáveis aleatórias $\mathrm{t}_1, \mathrm{t}_2, \cdots \mathrm{t}_\mathrm{A}$ são não correlacionadas. Para o caso em que $\mathbf{\Sigma} = \mathbf{I}$, este fato segue diretamente da construção geométrica, pois

as variáveis $\mathrm{t}_1, \mathrm{t}_2, \cdots \mathrm{t}_\mathrm{A}$ são obtidas por projeções nos vetores $\mathbf{w}_\mathrm{i}$, que são ortogonais.

3. O vetor aleatório $\mathbf{e}_a$ é não correlacionado à variável t_a, pois t_a é obtida por uma projeção em $\mathbf{w}_a$ e o vetor $\mathbf{e}_a$ é perpendicular ao vetor $\mathbf{w}_a$. De fato, temos mais do que isso: como $\mathbf{e}_a$ é perpendicular a $\mathbf{w}_a$, por construção, e $\mathbf{w}_a$ é perpendicular a $\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_{a-1}$, temos que $\mathbf{e}_a$ é perpendicular a todos eles, $\mathbf{w}_1, \mathbf{w}_2, \cdots \mathbf{w}_a$, implicando então, que $\mathbf{e}_a$ é não correlacionado às variáveis $\mathrm{t}_1, \mathrm{t}_2, \cdots \mathrm{t}_a$. Da mesma forma, este resultado segue da construção geométrica se $\mathbf{\Sigma} = \mathbf{I}$.

Estas propriedades podem ser demonstradas algebricamente (HOSKULDSSON, 1988).

As propriedades 2 e 3 permitem as seguintes simplificações no algoritmo:

•

$$\begin{aligned}
\mathbf{w}_a &= \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-1}\right) = \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-2} - \mathrm{q}_{a-1}\mathrm{t}_{a-1}\right) \\
&= \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-2}\right) - \mathrm{q}_{a-1}\mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{t}_{a-1}\right) \\
&\overset{\text{P.3}}{=} \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-2}\right) = \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-3} - \mathrm{q}_{a-2}\mathrm{t}_{a-2}\right) \\
&= \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-3}\right) - \mathrm{q}_{a-2}\mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{t}_{a-2}\right) \\
&\overset{\text{P.3}}{=} \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-3}\right) = \cdots = \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{y}\right)
\end{aligned}$$

Logo, $\mathbf{w}_a = \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{f}_{a-1}\right) = \mathrm{cov}\left(\mathbf{e}_{a-1}, \mathrm{y}\right)$.

•

$$\begin{aligned}\operatorname{cov}(\mathbf{e}_{a-1}, \mathrm{t}_a) &= \operatorname{cov}(\mathbf{e}_{a-2} - \mathbf{p}_{a-1}\mathrm{t}_{a-1}, \mathrm{t}_a) \\ &= \operatorname{cov}(\mathbf{e}_{a-2}, \mathrm{t}_a) - \operatorname{cov}(\mathrm{t}_{a-1}, \mathrm{t}_a)\,\mathbf{p}_{a-1} \\ &\overset{\text{P.2}}{=} \operatorname{cov}(\mathbf{e}_{a-2}, \mathrm{t}_a) = \operatorname{cov}(\mathbf{e}_{a-3} - \mathbf{p}_{a-2}\mathrm{t}_{a-2}, \mathrm{t}_a) \\ &= \operatorname{cov}(\mathbf{e}_{a-3}, \mathrm{t}_a) - \operatorname{cov}(\mathrm{t}_{a-2}, \mathrm{t}_a)\,\mathbf{p}_{a-2} \\ &\overset{\text{P.2}}{=} \operatorname{cov}(\mathbf{e}_{a-3}, \mathrm{t}_a) = \cdots = \operatorname{cov}(\mathbf{x}, \mathrm{t}_a)\end{aligned}$$

Logo, $\mathbf{p}_a = \frac{\operatorname{cov}(\mathbf{e}_{a-1}, \mathrm{t}_a)}{\operatorname{var}(\mathrm{t}_a)} = \frac{\operatorname{cov}(\mathbf{x}, \mathrm{t}_a)}{\operatorname{var}(\mathrm{t}_a)}$.

•

$$\begin{aligned}\operatorname{cov}(\mathrm{f}_{a-1}, \mathrm{t}_a) &= \operatorname{cov}(\mathrm{f}_{a-2} - \mathrm{q}_{a-1}\mathrm{t}_{a-1}, \mathrm{t}_a) \\ &= \operatorname{cov}(\mathrm{f}_{a-2}, \mathrm{t}_a) - \mathrm{q}_{a-1}\operatorname{cov}(\mathrm{t}_{a-1}, \mathrm{t}_a) \\ &\overset{\text{P.2}}{=} \operatorname{cov}(\mathrm{f}_{a-2}, \mathrm{t}_a) = \operatorname{cov}(\mathrm{f}_{a-3} - \mathrm{q}_{a-2}\mathrm{t}_{a-2}, \mathrm{t}_a) \\ &= \operatorname{cov}(\mathrm{f}_{a-3}, \mathrm{t}_a) - \mathrm{q}_{a-2}\operatorname{cov}(\mathrm{t}_{a-2}, \mathrm{t}_a) \\ &\overset{\text{P.2}}{=} \operatorname{cov}(\mathrm{f}_{a-3}, \mathrm{t}_a) = \cdots = \operatorname{cov}(\mathrm{y}, \mathrm{t}_a)\end{aligned}$$

Logo, $\mathrm{q}_a = \frac{\operatorname{cov}(\mathrm{f}_{a-1}, \mathrm{t}_a)}{\operatorname{var}(\mathrm{t}_a)} = \frac{\operatorname{cov}(\mathrm{y}, \mathrm{t}_a)}{\operatorname{var}(\mathrm{t}_a)}$.

Portanto, $\mathbf{e}_{a-1}$ pode ser substituído por $\mathbf{x}$ na definição de $\mathbf{p}_a$, e f_{a-1} pode ser substituído por y nas definições de $\mathbf{w}_a$ e q_a, nos passos 2 e 3 do algoritmo.

Assim, tem-se que o algoritmo fica da forma:

$$\mathbf{e}_0 = \mathbf{x} - \boldsymbol{\mu}_{\mathbf{x}}$$

$$\mathrm{f}_0 = \mathrm{y} - \mu_{\mathrm{y}}$$

$$\mathrm{t}_a = \mathbf{e}_{a-1}{}'\mathbf{w}_a$$

$$\begin{aligned}
\mathbf{w}_a &= \operatorname{cov}(\mathbf{e}_{a-1}, \mathrm{y}) \\
\mathbf{p}_a &= \frac{\operatorname{cov}(\mathbf{x}, \mathrm{t}_a)}{\operatorname{var}(\mathrm{t}_a)} \\
\mathrm{q}_a &= \frac{\operatorname{cov}(\mathrm{y}, \mathrm{t}_a)}{\operatorname{var}(\mathrm{t}_a)} \\
\mathbf{e}_a &= \mathbf{e}_{a-1} - \mathbf{p}_a \mathrm{t}_a \\
\mathrm{f}_a &= \mathrm{f}_{a-1} - \mathrm{q}_a \mathrm{t}_a.
\end{aligned}$$

Apesar do algoritmo nessa versão ser mais simples, como o objetivo é a motivação para se obter o algoritmo para o caso amostral, a primeira versão é a mais útil.

No passo A do algoritmo, obtemos a representação:

$$\begin{aligned}
&\mathbf{p}_1\mathrm{t}_1 + \mathbf{p}_2\mathrm{t}_2 + \mathbf{p}_3\mathrm{t}_3 + \cdots + \mathbf{p}_\mathrm{A}\mathrm{t}_\mathrm{A} \\
&= (\mathbf{e}_0 - \mathbf{e}_1) + (\mathbf{e}_1 - \mathbf{e}_2) + (\mathbf{e}_2 - \mathbf{e}_3) + \cdots + (\mathbf{e}_{\mathrm{A}-1} - \mathbf{e}_\mathrm{A}) \\
&= \mathbf{e}_0 - \mathbf{e}_\mathrm{A} = (\mathbf{x} - \boldsymbol{\mu}_\mathbf{x}) - \mathbf{e}_\mathrm{A},
\end{aligned}$$

logo, $\mathbf{x} = \boldsymbol{\mu}_\mathbf{x} + \mathbf{p}_1\mathrm{t}_1 + \cdots + \mathbf{p}_\mathrm{A}\mathrm{t}_\mathrm{A} + \mathbf{e}_\mathrm{A}$.

$$\begin{aligned}
&\mathrm{q}_1\mathrm{t}_1 + \mathrm{q}_2\mathrm{t}_2 + \mathrm{q}_3\mathrm{t}_3 + \cdots + \mathrm{q}_\mathrm{A}\mathrm{t}_\mathrm{A} \\
&= (\mathrm{f}_0 - \mathrm{f}_1) + (\mathrm{f}_1 - \mathrm{f}_2) + (\mathrm{f}_2 - \mathrm{f}_3) + \cdots + (\mathrm{f}_{\mathrm{A}-1} - \mathrm{f}_\mathrm{A}) \\
&= \mathrm{f}_0 - \mathrm{f}_\mathrm{A} = (\mathrm{y} - \mu_\mathrm{y}) - \mathrm{f}_\mathrm{A},
\end{aligned}$$

logo, $\mathrm{y} = \mu_\mathrm{y} + \mathrm{q}_1\mathrm{t}_1 + \cdots + \mathrm{q}_\mathrm{A}\mathrm{t}_\mathrm{A} + \mathrm{f}_\mathrm{A}$.

Temos então, um preditor para a variável resposta y, dado por

$$\hat{\mathrm{y}}_{\mathrm{A,PLS}} = \mu_\mathrm{y} + \mathrm{q}_1\mathrm{t}_1 + \cdots \mathrm{q}_\mathrm{A}\mathrm{t}_\mathrm{A}.$$

Este preditor é, de fato, um preditor linear, pois utilizando a segunda versão do

algoritmo, tem-se:

$$\hat{y}_{\mathrm{A,PLS}} = \mu_{\mathrm{y}} + \boldsymbol{\beta}^{'}_{\mathrm{A,PLS}} \left(\mathbf{x} - \boldsymbol{\mu}_{\mathrm{x}}\right)$$

em que (HELLAND, 1990),

$$\boldsymbol{\beta}_{\mathrm{A,PLS}} = \mathbf{W}_{\mathrm{A}} \left(\mathbf{W}_{\mathrm{A}}{}'\boldsymbol{\Sigma}\mathbf{W}_{\mathrm{A}}\right)^{-1} \mathbf{W}_{\mathrm{A}}{}'\boldsymbol{\sigma}.$$

Explicitando essas matrizes:

$$\mathbf{W_A} = [\mathbf{w}_1, \ldots, \mathbf{w}_{\mathrm{A}}]$$

$$\boldsymbol{\Sigma}\left(\mathbf{W_A}\right) = \boldsymbol{\Sigma}\left[\mathbf{w}_1, \ldots, \mathbf{w}_{\mathrm{A}}\right] = \left[\boldsymbol{\Sigma}\mathbf{w}_1, \ldots, \boldsymbol{\Sigma}\mathbf{w}_{\mathrm{A}}\right]$$

$$\mathbf{W_A}{}'\boldsymbol{\Sigma}\mathbf{W_A} = \begin{bmatrix} \mathbf{w}_1{}' \\ \vdots \\ \mathbf{w}_{\mathrm{A}}{}' \end{bmatrix} \left[\boldsymbol{\Sigma}\mathbf{w}_1, \ldots, \boldsymbol{\Sigma}\mathbf{w}_{\mathrm{A}}\right] = \begin{bmatrix} \mathbf{w}_1{}'\boldsymbol{\Sigma}\mathbf{w}_1 & \cdots & \mathbf{w}_1{}'\boldsymbol{\Sigma}\mathbf{w}_{\mathrm{A}} \\ \vdots & \ddots & \vdots \\ \mathbf{w}_{\mathrm{A}}{}'\boldsymbol{\Sigma}\mathbf{w}_1 & \cdots & \mathbf{w}_{\mathrm{A}}{}'\boldsymbol{\Sigma}\mathbf{w}_{\mathrm{A}} \end{bmatrix}.$$

Casos particulares:

1º Caso: Se $\mathbf{w}_{\mathrm{i}}{}'\boldsymbol{\Sigma}\mathbf{w}_{\mathrm{j}} = 0, \forall \mathrm{i} \neq \mathrm{j}$, temos uma matriz diagonal e poderemos calcular a inversa:

$$\left(\mathbf{W_A}{}'\boldsymbol{\Sigma}\mathbf{W_A}\right)^{-1} = \begin{bmatrix} \frac{1}{\mathbf{w}_1{}'\boldsymbol{\Sigma}\mathbf{w}_1} & 0 & \cdots & 0 \\ 0 & \frac{1}{\mathbf{w}_2{}'\boldsymbol{\Sigma}\mathbf{w}_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \frac{1}{\mathbf{w}_{\mathrm{A}}{}'\boldsymbol{\Sigma}\mathbf{w}_{\mathrm{A}}} \end{bmatrix}$$

$$\mathbf{W_A}{}'\boldsymbol{\sigma} = \begin{bmatrix} \mathbf{w}_1{}' \\ \vdots \\ \mathbf{w}_{\mathrm{A}}{}' \end{bmatrix} \boldsymbol{\sigma} = \begin{bmatrix} \mathbf{w}_1{}'\boldsymbol{\sigma} \\ \vdots \\ \mathbf{w}_{\mathrm{A}}{}'\boldsymbol{\sigma} \end{bmatrix}$$

$$\left(\mathbf{W_A}'\mathbf{\Sigma}\mathbf{W_A}\right)^{-1}\mathbf{W_A}'\boldsymbol{\sigma} = \begin{bmatrix} \frac{\mathbf{w}_1'\boldsymbol{\sigma}}{\mathbf{w}_1'\mathbf{\Sigma}\mathbf{w}_1} \\ \vdots \\ \frac{\mathbf{w}_\mathrm{A}'\boldsymbol{\sigma}}{\mathbf{w}_\mathrm{A}'\mathbf{\Sigma}\mathbf{w}_\mathrm{A}} \end{bmatrix}$$

$$\begin{aligned}\mathbf{W_A}\left(\mathbf{W_A}'\mathbf{\Sigma}\mathbf{W_A}\right)^{-1}\mathbf{W_A}'\boldsymbol{\sigma} &= [\mathbf{w}_1, \ldots, \mathbf{w}_\mathrm{A}] \begin{bmatrix} \frac{\mathbf{w}_1'\boldsymbol{\sigma}}{\mathbf{w}_1'\mathbf{\Sigma}\mathbf{w}_1} \\ \vdots \\ \frac{\mathbf{w}_\mathrm{A}'\boldsymbol{\sigma}}{\mathbf{w}_\mathrm{A}'\mathbf{\Sigma}\mathbf{w}_\mathrm{A}} \end{bmatrix} \\ &= \left[\left(\frac{\mathbf{w}_1'\boldsymbol{\sigma}}{\mathbf{w}_1'\mathbf{\Sigma}\mathbf{w}_1}\right)\mathbf{w}_1 + \cdots + \left(\frac{\mathbf{w}_\mathrm{A}'\boldsymbol{\sigma}}{\mathbf{w}_\mathrm{A}'\mathbf{\Sigma}\mathbf{w}_\mathrm{A}}\right)\mathbf{w}_\mathrm{A}\right].\end{aligned}$$

2º Caso: Se o número de passos A do algoritmo é igual a n, $\mathrm{A} = \mathrm{n}$, a matriz $\mathbf{W}_\mathrm{A}$ é quadrada e inversível, assim,

$$\begin{aligned}\boldsymbol{\beta}_{\mathrm{A,PLS}} &= \mathbf{W}_\mathrm{A}\left(\mathbf{W}_\mathrm{A}'\mathbf{\Sigma}\mathbf{W}_\mathrm{A}\right)^{-1}\mathbf{W}_\mathrm{A}'\boldsymbol{\sigma}, \\ \boldsymbol{\beta}_{\mathrm{n,PLS}} &= \mathbf{W}_\mathrm{A}{\mathbf{W}_\mathrm{A}}^{-1}\mathbf{\Sigma}^{-1}\left(\mathbf{W}_\mathrm{A}'\right)^{-1}\mathbf{W}_\mathrm{A}'\boldsymbol{\sigma} = \mathbf{\Sigma}^{-1}\boldsymbol{\sigma},\end{aligned}$$

que é o mesmo coeficiente do melhor preditor linear, no caso em que $\begin{bmatrix} \mathbf{x} \\ \mathrm{y} \end{bmatrix}$ tem distribuição Normal.

3º Caso: O preditor para apenas 1 passo, fica da forma:

$$\hat{\mathbf{Y}}_{\mathrm{A,PLS}} = \mu_\mathrm{y} + \mathrm{q}_1\mathrm{t}_1 = \mu_\mathrm{y} + \frac{\|\boldsymbol{\sigma}\|^2}{\boldsymbol{\sigma}'\mathbf{\Sigma}\boldsymbol{\sigma}}\mathbf{x}'\boldsymbol{\sigma} = \mu_\mathrm{y} + \frac{\|\boldsymbol{\sigma}\|^2}{\boldsymbol{\sigma}'\mathbf{\Sigma}\boldsymbol{\sigma}}\boldsymbol{\sigma}'\left(\mathbf{x} - \boldsymbol{\mu}_\mathrm{x}\right).$$

4.3.2 PLS Amostral

Esta seção tem como referência o artigo Hoskuldsson (1988).

Sejam duas matrizes $\mathbf{X}_{\mathrm{n\times p}}$ e $\mathbf{Y}_{\mathrm{n\times k}}$, centradas na média, em que p é o número de covariáveis, k o número de variáveis respostas e n o número de observações (repetições):

$$\mathbf{X}=\begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{2\mathrm{p}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{\mathrm{n}1} & \mathrm{x}_{\mathrm{n}2} & \cdots & \mathrm{x}_{\mathrm{np}} \end{bmatrix}, \quad \mathbf{Y}=\begin{bmatrix} \mathrm{y}_{11} & \mathrm{y}_{12} & \cdots & \mathrm{y}_{1\mathrm{k}} \\ \mathrm{y}_{21} & \mathrm{y}_{22} & \cdots & \mathrm{y}_{2\mathrm{k}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{y}_{\mathrm{n}1} & \mathrm{y}_{\mathrm{n}2} & \cdots & \mathrm{y}_{\mathrm{nk}} \end{bmatrix}.$$

A matriz de covariâncias amostrais, a menos de uma constante, é dada por $\mathbf{X}'\mathbf{Y}$:

$$\mathbf{X}'\mathbf{Y} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{21} & \cdots & \mathrm{x}_{\mathrm{n}1} \\ \mathrm{x}_{12} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{\mathrm{n}2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{1\mathrm{p}} & \mathrm{x}_{2\mathrm{p}} & \cdots & \mathrm{x}_{\mathrm{np}} \end{bmatrix} \begin{bmatrix} \mathrm{y}_{11} & \mathrm{y}_{12} & \cdots & \mathrm{y}_{1\mathrm{k}} \\ \mathrm{y}_{21} & \mathrm{y}_{22} & \cdots & \mathrm{y}_{2\mathrm{k}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{y}_{\mathrm{n}1} & \mathrm{y}_{\mathrm{n}2} & \cdots & \mathrm{y}_{\mathrm{nk}} \end{bmatrix}$$

$$= \begin{bmatrix} \hat{\mathrm{cov}}\left(\mathbf{x}_1,\mathbf{y}_1\right) & \hat{\mathrm{cov}}\left(\mathbf{x}_1,\mathbf{y}_2\right) & \cdots & \hat{\mathrm{cov}}\left(\mathbf{x}_1,\mathbf{y}_{\mathrm{k}}\right) \\ \hat{\mathrm{cov}}\left(\mathbf{x}_2,\mathbf{y}_1\right) & \hat{\mathrm{cov}}\left(\mathbf{x}_2,\mathbf{y}_2\right) & \cdots & \hat{\mathrm{cov}}\left(\mathbf{x}_2,\mathbf{y}_{\mathrm{k}}\right) \\ \vdots & \vdots & \ddots & \vdots \\ \hat{\mathrm{cov}}\left(\mathbf{x}_{\mathrm{p}},\mathbf{y}_1\right) & \hat{\mathrm{cov}}\left(\mathbf{x}_{\mathrm{p}},\mathbf{y}_2\right) & \cdots & \hat{\mathrm{cov}}\left(\mathbf{x}_{\mathrm{p}},\mathbf{y}_{\mathrm{k}}\right) \end{bmatrix}.$$

Observe que as colunas da matriz $\mathbf{X}'\mathbf{Y}$, a menos de uma constante, estimam os vetores $\boldsymbol{\sigma}_{\mathrm{j}} = \mathrm{cov}\left(\mathbf{X}, \mathbf{Y}_{\mathrm{j}}\right)$ com $\mathrm{j} = 1, \ldots, \mathrm{k}$.

4.3.2.1 Componentes com covariância máxima

Considere dois componentes $\mathbf{f}$ e $\mathbf{g}$ definidos pelas colunas das matrizes $\mathbf{X}$ e $\mathbf{Y}$:

$$\mathbf{f} = \mathbf{X}\mathbf{d}, \quad \|\mathbf{d}\| = 1$$

$$\mathbf{g} = \mathbf{Y}\mathbf{e}, \quad \|\mathbf{e}\| = 1.$$

Note que,

$$\mathbf{f} = \mathbf{X}\mathbf{d} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{2\mathrm{p}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{\mathrm{n}1} & \mathrm{x}_{\mathrm{n}2} & \cdots & \mathrm{x}_{\mathrm{np}} \end{bmatrix} \begin{bmatrix} \mathrm{d}_1 \\ \mathrm{d}_2 \\ \vdots \\ \mathrm{d}_\mathrm{p} \end{bmatrix} = \mathrm{d}_1 \begin{bmatrix} \mathrm{x}_{11} \\ \mathrm{x}_{21} \\ \vdots \\ \mathrm{x}_{\mathrm{n}1} \end{bmatrix} + \cdots + \mathrm{d}_\mathrm{p} \begin{bmatrix} \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{2\mathrm{p}} \\ \vdots \\ \mathrm{x}_{\mathrm{np}} \end{bmatrix},$$

logo, $\mathbf{f} = \mathbf{X}\mathbf{d}$ é uma combinação linear das colunas de $\mathbf{X}$, isto é, é um vetor definido como uma média ponderada das realizações de todas as covariáveis. Da mesma forma, $\mathbf{g} = \mathbf{Y}\mathbf{e}$ é uma combinação linear das colunas de $\mathbf{Y}$.

A covariância amostral entre os dois componentes é dada por:

$$\mathrm{cov}\,(\mathbf{f}, \mathbf{g}) = \frac{\mathbf{f}'\mathbf{g}}{\mathrm{n}}.$$

Uma boa escolha para $\mathbf{f}$ e $\mathbf{g}$ será aquela que fornecer covariância máxima. Aqui, a covariância deve ser máxima em valor absoluto, visto que a covariância pode ser positiva ou negativa.

Para obter os vetores que realizam essa maximização da covariância, é necessário fazer a decomposição em valores singulares (Apêndice B) da matriz de covariâncias amostrais $\mathbf{X}'\mathbf{Y}$. Para tal, considere $\mathbf{X}'\mathbf{Y}$ como uma transformação linear de $\mathbb{R}^\mathrm{k}$ em $\mathbb{R}^\mathrm{p}$, $\mathrm{k} < \mathrm{p}$. Existe uma base $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_\mathrm{k}\}$ em $\mathbb{R}^\mathrm{k}$, e uma base

$\{\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_\mathrm{p}\}$ em $\mathbb{R}^\mathrm{p}$, ambas ortonormais, tais que $\mathbf{X}'\mathbf{Y}\left(\mathbf{v}_\mathrm{i}\right) = a_\mathrm{i}\mathbf{z}_\mathrm{i}$, sendo a ordem dos vetores nas bases definida por $a_1^2 \geq a_2^2 \geq \cdots \geq a_\mathrm{k}^2$, conforme mostra a Figura 21. Desta forma, tem-se a igualdade de matrizes $\mathbf{X}'\mathbf{Y} = \sum_{\mathrm{i}=1}^{\mathrm{k}} a_\mathrm{i}\mathbf{z}_\mathrm{i}\mathbf{v}_\mathrm{i}{}'$, pois

$$\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_\mathrm{i}\mathbf{z}_\mathrm{i}\mathbf{v}_\mathrm{i}{}'\right)\left(\mathbf{v}_\mathrm{j}\right) = \sum_{\mathrm{i}=1}^{\mathrm{k}} \left(a_\mathrm{i}\mathbf{z}_\mathrm{i}\mathbf{v}_\mathrm{i}{}'\right)\left(\mathbf{v}_\mathrm{j}\right) = \sum_{\mathrm{i}=1}^{\mathrm{k}} \left(a_\mathrm{i}\mathbf{z}_\mathrm{i}\right)\left(\mathbf{v}_\mathrm{i}{}'\mathbf{v}_\mathrm{j}\right) = a_\mathrm{j}\mathbf{z}_\mathrm{j} = \mathbf{X}'\mathbf{Y}\left(\mathbf{v}_\mathrm{j}\right).$$

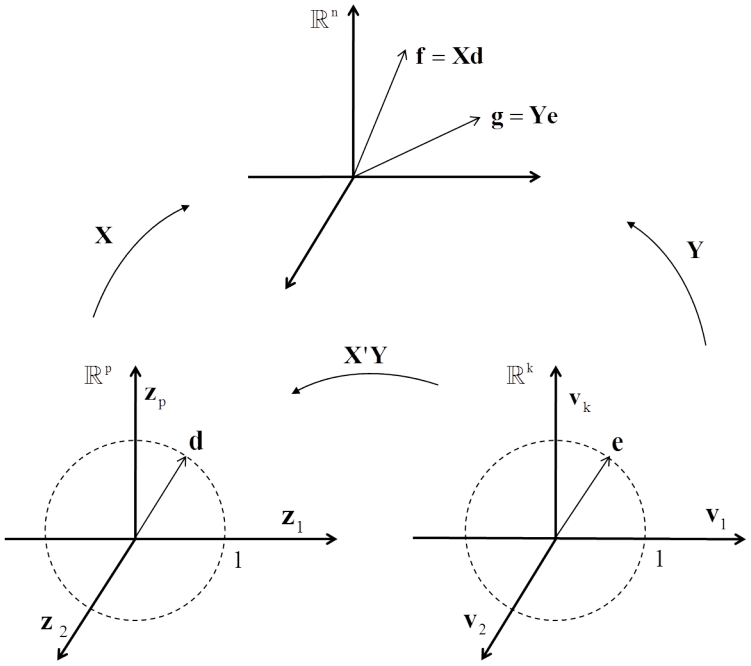


Figura 21 Representação de $\mathbf{X}'\mathbf{Y}$ como uma transformação linear de $\mathbb{R}^\mathrm{k}$ em $\mathbb{R}^\mathrm{p}$

O problema da maximização, $\max\left(\mathbf{f}'\mathbf{g}\right)$, $\|\mathbf{d}\| = \|\mathbf{e}\| = 1$, pode ser colocado da forma:

$$\begin{aligned}\max\left(\mathbf{f}'\mathbf{g}\right) &= \max\left(\left(\mathbf{X}\mathbf{d}\right)'\left(\mathbf{Y}\mathbf{e}\right)\right) = \max\left(\mathbf{d}'\mathbf{X}'\mathbf{Y}\mathbf{e}\right) \\ &= \max\left(\mathbf{d}'\left(\mathbf{X}'\mathbf{Y}\right)\mathbf{e}\right) = \max\left(\mathbf{d}'\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_{\mathrm{i}}\mathbf{z}_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}{}'\right)\mathbf{e}\right) \\ &= \max\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_{\mathrm{i}}\left(\mathbf{d}'\left(\mathbf{z}_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}{}'\right)\mathbf{e}\right)\right) = \max\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_{\mathrm{i}}\left(\mathbf{d}'\mathbf{z}_{\mathrm{i}}\right)\left(\mathbf{v}_{\mathrm{i}}{}'\mathbf{e}\right)\right).\end{aligned}$$

Assim, os vetores $\mathbf{d}$ e $\mathbf{e}$ que realizam essa maximização são os vetores $\mathbf{d} = \mathbf{z}_1$ e $\mathbf{e} = \mathbf{v}_1$, pois,

$$\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_{\mathrm{i}}\left(\mathbf{z}_1{}'\mathbf{z}_{\mathrm{i}}\right)\left(\mathbf{v}_{\mathrm{i}}{}'\mathbf{v}_1\right)\right) = a_1.$$

Demonstração. Objetiva-se maximizar (ou minimizar) a expressão:

$$\begin{aligned}\max\left(\left(\mathbf{X}\mathbf{d}\right)'\left(\mathbf{Y}\mathbf{e}\right)\right) &= \max\left(\mathbf{d}'\left(\mathbf{X}'\mathbf{Y}\right)\mathbf{e}\right) \\ &= \max\left(\mathbf{d}'\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_{\mathrm{i}}\mathbf{z}_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}^{'}\right)\mathbf{e}\right) \\ &= \max\left(\sum_{\mathrm{j}=1}^{\mathrm{k}} \mathrm{d}_{\mathrm{j}}\mathbf{z}_{\mathrm{j}}\right)^{'}\left(\sum_{\mathrm{i}=1}^{\mathrm{k}} a_{\mathrm{i}}\mathbf{z}_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}^{'}\right)\left(\sum_{\mathrm{s}=1}^{\mathrm{k}} \mathrm{e}_{\mathrm{s}}\mathbf{v}_{\mathrm{s}}\right) \\ &= \max\sum_{\mathrm{j,i,s}} \mathrm{d}_{\mathrm{j}}a_{\mathrm{i}}\mathrm{e}_{\mathrm{s}}\mathbf{z}_{\mathrm{j}}^{'}\left(\mathbf{z}_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}^{'}\right)\mathbf{v}_{\mathrm{s}} \\ &= \max\sum_{\mathrm{j,i,s}} \mathrm{d}_{\mathrm{j}}a_{\mathrm{i}}\mathrm{e}_{\mathrm{s}}\left(\mathbf{z}_{\mathrm{j}}^{'}\mathbf{z}_{\mathrm{i}}\right)\left(\mathbf{v}_{\mathrm{i}}^{'}\mathbf{v}_{\mathrm{s}}\right) \\ &= \max\sum_{\mathrm{i}} \mathrm{d}_{\mathrm{i}}a_{\mathrm{i}}\mathrm{e}_{\mathrm{i}}\left(\mathbf{z}_{\mathrm{i}}^{'}\mathbf{z}_{\mathrm{i}}\right)\left(\mathbf{v}_{\mathrm{i}}^{'}\mathbf{v}_{\mathrm{i}}\right) \\ &= \max_{\|\mathbf{d}\|=\|\mathbf{e}\|=1}\sum_{\mathrm{i}} \mathrm{d}_{\mathrm{i}}a_{\mathrm{i}}\mathrm{e}_{\mathrm{i}}\end{aligned}$$

Assim, obtém-se a função lagrangiana com duas restrições, $\|\mathbf{d}\| = \|\mathbf{e}\| = 1$:

$$\mathrm{H}\left(\mathrm{d}_1, \ldots, \mathrm{d}_\mathrm{k}, \mathrm{e}_1, \ldots, \mathrm{e}_\mathrm{k}, \lambda, \alpha\right) = \sum \mathrm{d}_\mathrm{i} a_\mathrm{i} \mathrm{e}_\mathrm{i} + \lambda \left(\sum \mathrm{d}_\mathrm{i}^2 - 1 \right) + \alpha \left(\sum \mathrm{e}_\mathrm{i}^2 - 1 \right).$$

De onde segue o sistema de equações:

$$\begin{cases} \frac{\partial \mathrm{H}}{\partial \mathrm{d}_\mathrm{j}} = a_\mathrm{j} \mathrm{e}_\mathrm{j} + 2\lambda \mathrm{d}_\mathrm{j} = 0 \\ \frac{\partial \mathrm{H}}{\partial \mathrm{e}_\mathrm{l}} = a_\mathrm{l} \mathrm{d}_\mathrm{l} + 2\alpha \mathrm{e}_\mathrm{l} = 0 \\ \sum \mathrm{d}_\mathrm{i}^2 = 1 \\ \sum \mathrm{e}_\mathrm{i}^2 = 1 \end{cases}$$

Multiplicando a 1ª equação por d_j e a 2ª equação por e_j, tem-se:

$$\mathrm{d}_\mathrm{j} a_\mathrm{j} \mathrm{e}_\mathrm{j} + 2\lambda \mathrm{d}_\mathrm{j}^2 = 0$$
$$\mathrm{e}_\mathrm{j} a_\mathrm{j} \mathrm{d}_\mathrm{j} + 2\alpha \mathrm{e}_\mathrm{j}^2 = 0$$

Igualando as duas equações, segue que:

$$\lambda \mathrm{d}_\mathrm{j}^2 = \alpha \mathrm{e}_\mathrm{j}^2 \Rightarrow \lambda \sum \mathrm{d}_\mathrm{j}^2 = \alpha \sum \mathrm{e}_\mathrm{j}^2 \Rightarrow \lambda = \alpha \Rightarrow \mathrm{d}_\mathrm{j} = \left|\mathrm{e}_\mathrm{j}\right|.$$

Assim, volta-se ao problema da maximização, mas agora com apenas uma restrição:

$$\max_{\|\mathbf{d}\|=1} \sum a_\mathrm{i} \mathrm{d}_\mathrm{i}^2.$$

Dada a função lagrangiana:

$$\mathrm{H}\left(\mathrm{d}_1, \ldots, \mathrm{d}_\mathrm{k}, \lambda\right) = \sum a_\mathrm{i} \mathrm{d}_\mathrm{i}^2 + \lambda \left(\sum \mathrm{d}_\mathrm{i}^2 - 1 \right),$$

segue que

$$\frac{\partial \mathrm{H}}{\partial \mathrm{d_j}} = 2a_\mathrm{j}\mathrm{d_j} + 2\lambda \mathrm{d_j} = 0.$$

Logo, $a_\mathrm{j}\mathrm{d_j} = \lambda \mathrm{d_j}$.

Supondo todos os a_i diferentes entre si, $a_\mathrm{i} \neq a_\mathrm{j}$ para $\mathrm{i} \neq \mathrm{j}$, a solução consiste em todos os $\mathrm{d_j}$ serem zero, exceto um deles que deve ser 1, pois $\|\mathbf{d}\| = 1$. Assim, o máximo ocorre para $\mathrm{i} = 1$, pois a_1 é maior que os outros. Logo,

$$\max_{\|\mathbf{d}\|=\|\mathbf{e}\|=1} \left(\mathbf{d}' \left(\mathbf{X}'\mathbf{Y}\right) \mathbf{e}\right) = a_1.$$

□

Portanto, foram obtidos os vetores de correlação máxima $\mathbf{f} = \mathbf{X}\mathbf{z}_1$ e $\mathbf{g} = \mathbf{Y}\mathbf{v}_1$. O algoritmo PLS é justamente um processo iterativo para se obter outros pares de componentes que também maximizam a covariância com mais restrições (em ordem decrescente de magnitude).

4.3.2.2 O algoritmo PLS Multivariado

No sentido de melhor explicar os passos do algoritmo, seu desenvolvimento será paralelamente relacionado ao algoritmo PLS Populacional, anteriormente descrito. Com o objetivo de simplificar a notação, todo vetor da forma $\frac{\mathbf{a}}{\mathbf{a}'\mathbf{a}}$ receberá o mesmo nome $\mathbf{a}$.

O algoritmo inicia-se da seguinte forma: tome o vetor $\mathbf{u}$ como a primeira coluna da matriz de respostas $\mathbf{Y}$, isto é, $\mathbf{y}_1$. Faça $\frac{\mathbf{u}}{\mathbf{u}'\mathbf{u}}$ e seja $\mathbf{w} = \frac{\mathbf{X}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$. Logo, $\mathbf{w}$ é uma combinação linear das linhas da matriz $\mathbf{X}$, ou seja, é um vetor de covariâncias

amostrais entre a primeira variável resposta e as covariáveis:

$$\mathbf{w} = \mathbf{X}'\mathbf{u} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{21} & \cdots & \mathrm{x}_{\mathrm{n}1} \\ \mathrm{x}_{12} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{\mathrm{n}2} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{1\mathrm{p}} & \mathrm{x}_{2\mathrm{p}} & \cdots & \mathrm{x}_{\mathrm{np}} \end{bmatrix} \begin{bmatrix} \mathrm{y}_{11} \\ \mathrm{y}_{21} \\ \vdots \\ \mathrm{y}_{\mathrm{n}1} \end{bmatrix}$$

$$= \mathrm{y}_{11} \begin{bmatrix} \mathrm{x}_{11} \\ \mathrm{x}_{12} \\ \vdots \\ \mathrm{x}_{1\mathrm{p}} \end{bmatrix} + \cdots + \mathrm{y}_{\mathrm{n}1} \begin{bmatrix} \mathrm{x}_{\mathrm{n}1} \\ \mathrm{x}_{\mathrm{n}2} \\ \vdots \\ \mathrm{x}_{\mathrm{np}} \end{bmatrix} = \begin{bmatrix} \hat{\mathrm{cov}}\left(\mathbf{x}_1, \mathbf{y}_1\right) \\ \hat{\mathrm{cov}}\left(\mathbf{x}_2, \mathbf{y}_1\right) \\ \vdots \\ \hat{\mathrm{cov}}\left(\mathbf{x}_{\mathrm{p}}, \mathbf{y}_1\right) \end{bmatrix}.$$

Observe que $\mathbf{w} = \frac{\mathbf{X}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$ é o equivalente amostral do vetor $\mathbf{w}_1 = \boldsymbol{\sigma} = \mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right)$, no algoritmo PLS populacional univariado, com $\mathrm{y} = \mathrm{y}_1$.

Normalize $\mathbf{w}$ e seja $\mathbf{t} = \mathbf{X}\mathbf{w}$:

$$\mathbf{t} = \mathbf{X}\mathbf{w} = \begin{bmatrix} \mathrm{x}_{11} & \mathrm{x}_{12} & \cdots & \mathrm{x}_{1\mathrm{p}} \\ \mathrm{x}_{21} & \mathrm{x}_{22} & \cdots & \mathrm{x}_{2\mathrm{p}} \\ \vdots & \vdots & \ddots & \vdots \\ \mathrm{x}_{\mathrm{n}1} & \mathrm{x}_{\mathrm{n}2} & \cdots & \mathrm{x}_{\mathrm{np}} \end{bmatrix} \begin{bmatrix} \hat{\mathrm{cov}}\left(\mathbf{x}_1, \mathbf{y}_1\right) \\ \hat{\mathrm{cov}}\left(\mathbf{x}_2, \mathbf{y}_1\right) \\ \vdots \\ \hat{\mathrm{cov}}\left(\mathbf{x}_{\mathrm{p}}, \mathbf{y}_1\right) \end{bmatrix} = \begin{bmatrix} \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{x}_{1\mathrm{i}} \hat{\mathrm{cov}}\left(\mathrm{x}_{\mathrm{i}}, \mathbf{y}_1\right) \\ \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{x}_{2\mathrm{i}} \hat{\mathrm{cov}}\left(\mathrm{x}_{\mathrm{i}}, \mathbf{y}_1\right) \\ \vdots \\ \sum_{\mathrm{i}=1}^{\mathrm{p}} \mathrm{x}_{\mathrm{ni}} \hat{\mathrm{cov}}\left(\mathrm{x}_{\mathrm{i}}, \mathbf{y}_1\right) \end{bmatrix}.$$

Veja que, a i-ésima entrada de $\mathbf{t}$ corresponde a uma estimativa da variável $\mathrm{t} = \boldsymbol{\sigma}'\mathbf{x} = \left(\mathrm{cov}\left(\mathbf{x}, \mathrm{y}\right)\right)'\mathbf{x}$, com $\mathrm{y} = \mathrm{y}_1$.

Em seguida, faça $\mathbf{c} = \frac{\mathbf{Y}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$:

$$\mathbf{c} = \mathbf{Y}'\mathbf{t} = \begin{bmatrix} y_{11} & y_{21} & \cdots & y_{n1} \\ y_{12} & y_{22} & \cdots & y_{n2} \\ \vdots & \vdots & \ddots & \vdots \\ y_{1k} & y_{2k} & \cdots & y_{nk} \end{bmatrix} \begin{bmatrix} \sum_{i=1}^{p} x_{1i} \hat{cov}(x_i, \mathbf{y}_1) \\ \sum_{i=1}^{p} x_{2i} \hat{cov}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{i=1}^{p} x_{ni} \hat{cov}(x_i, \mathbf{y}_1) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{j=1}^{n} \sum_{i=1}^{p} y_{j1} x_{ji} \hat{cov}(x_i, \mathbf{y}_1) \\ \sum_{j=1}^{n} \sum_{i=1}^{p} y_{j2} x_{ji} \hat{cov}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{j=1}^{n} \sum_{i=1}^{p} y_{jk} x_{ji} \hat{cov}(x_i, \mathbf{y}_1) \end{bmatrix}.$$

Fazendo correspondência com o PLS populacional, a i-ésima entrada do vetor $\mathbf{c}$ é uma estimativa do parâmetro $q_1 = \mathrm{cov}(y, \boldsymbol{\sigma}'\mathbf{x}) = \mathrm{cov}(y, t_1)$, com $y = y_1$.

Normalize $\mathbf{c}$ e tome $\mathbf{u} = \mathbf{Y}\mathbf{c}$:

$$\mathbf{u} = \mathbf{Y}\mathbf{c} = \begin{bmatrix} y_{11} & y_{12} & \cdots & y_{1k} \\ y_{21} & y_{22} & \cdots & y_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ y_{n1} & y_{n2} & \cdots & y_{nk} \end{bmatrix} \begin{bmatrix} \sum_{j=1}^{n} \sum_{i=1}^{p} y_{j1} x_{ji} \hat{cov}(x_i, \mathbf{y}_1) \\ \sum_{j=1}^{n} \sum_{i=1}^{p} y_{j2} x_{ji} \hat{cov}(x_i, \mathbf{y}_1) \\ \vdots \\ \sum_{j=1}^{n} \sum_{i=1}^{p} y_{jk} x_{ji} \hat{cov}(x_i, \mathbf{y}_1) \end{bmatrix}$$

$$= \begin{bmatrix} \sum_{l=1}^{k}\sum_{j=1}^{n}\sum_{i=1}^{p} y_{1l}y_{j1}x_{ji}\hat{cov}\left(x_i,\mathbf{y}_1\right) \\ \sum_{l=1}^{k}\sum_{j=1}^{n}\sum_{i=1}^{p} y_{2l}y_{j1}x_{ji}\hat{cov}\left(x_i,\mathbf{y}_1\right) \\ \vdots \\ \sum_{l=1}^{k}\sum_{j=1}^{n}\sum_{i=1}^{p} y_{nl}y_{j1}x_{ji}\hat{cov}\left(x_i,\mathbf{y}_1\right) \end{bmatrix}.$$

Este vetor não tem uma interpretação em termos do PLS populacional.

Nesta etapa, o algoritmo define um *loop* obtendo sequências de vetores $\mathbf{u}_n = \mathbf{YY'XX'u}_{n-1}$, $\mathbf{c}_n = \mathbf{Y'XX'Yc}_{n-1}$, $\mathbf{t}_n = \mathbf{XX'YY't}_{n-1}$ e $\mathbf{w}_n = \mathbf{X'YY'Xw}_{n-1}$. Essas sequências são convergentes em razão de uma série de propriedades de álgebra linear, relativas ao diagrama da Figura 22.

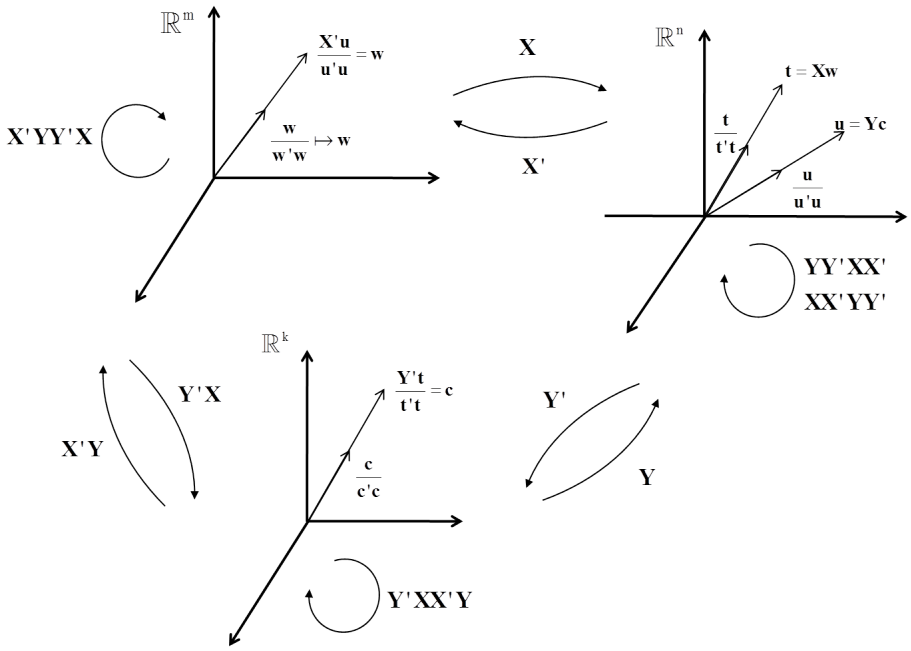


Figura 22 Representação geométrica das sequências de vetores $\mathbf{u}_n = \mathbf{YY'XX'u}_{n-1}$, $\mathbf{c}_n = \mathbf{Y'XX'Yc}_{n-1}$, $\mathbf{t}_n = \mathbf{XX'YY't}_{n-1}$ e $\mathbf{w}_n = \mathbf{X'YY'Xw}_{n-1}$

Observe que:

$$\mathbf{X'YY'X} = \left(\sum_{i=1}^{k} a_i \mathbf{z}_i \mathbf{v}_i^{'}\right)\left(\sum_{j=1}^{k} a_j \mathbf{v}_j \mathbf{z}_j^{'}\right) = \sum_{i=1}^{k}\sum_{j=1}^{k} a_i a_j \mathbf{z}_i \mathbf{v}_i^{'} \mathbf{v}_j \mathbf{z}_j^{'} = \sum_{i=1}^{k} a_i^2 \mathbf{z}_i \mathbf{z}_i^{'} \ .$$

A partir dessa decomposição, os autovalores de $\mathbf{X'YY'X}$ são $a_1^2 \geq a_2^2 \geq \cdots \geq a_k^2$, com respectivos autovetores $\mathbf{z}_1, \mathbf{z}_2, \ldots, \mathbf{z}_k$, pois:

$$\mathbf{X'YY'X}\left(\mathbf{z}_j\right) = \sum_{i=1}^{k} a_i^2 \mathbf{z}_i \mathbf{z}_i^{'} \mathbf{z}_j = a_j^2 \mathbf{z}_j.$$

- As matrizes $\mathbf{X'YY'X}$, $\mathbf{YY'XX'}$, $\mathbf{Y'XX'Y}$ e $\mathbf{XX'YY'}$ possuem os mesmos autovalores, pois:

  se $\mathbf{r}$ é um autovetor da matriz $\mathbf{X'YY'X}$, então

  $$\mathbf{XX'YY'}\left(\mathbf{Xr}\right) = \mathbf{X}\left(\mathbf{X'YY'X}\right)\mathbf{r} = \mathbf{X}\left(\lambda\mathbf{r}\right) = \lambda\left(\mathbf{Xr}\right).$$

  Se $\mathbf{r}$ é um autovetor da matriz $\mathbf{XX'YY'}$, então

  $$\mathbf{Y'XX'Y}\left(\mathbf{Y'r}\right) = \mathbf{Y'}\left(\mathbf{XX'YY'}\right)\mathbf{r} = \mathbf{Y'}\left(\lambda\mathbf{r}\right) = \lambda\left(\mathbf{Y'r}\right).$$

  Se $\mathbf{r}$ é um autovetor da matriz $\mathbf{Y'XX'Y}$, então

  $$\mathbf{YY'XX'}\left(\mathbf{Yr}\right) = \mathbf{Y}\left(\mathbf{Y'XX'Y}\right)\mathbf{r} = \mathbf{Y}\left(\lambda\mathbf{r}\right) = \lambda\left(\mathbf{Yr}\right).$$

  Portanto, o maior autovalor para todas as matrizes é dado por a_1^2.

- Aplicando o método das potências (Apêndice C) para o cálculo do autovetor relativo ao maior autovalor às sequências $\mathbf{u}_n = \mathbf{YY'XX'u}_{n-1}$, $\mathbf{c}_n = \mathbf{Y'XX'Yc}_{n-1}$, $\mathbf{t}_n = \mathbf{XX'YY't}_{n-1}$, e $\mathbf{w}_n = \mathbf{X'YY'Xw}_{n-1}$, obtém-se

os vetores limites das sequências $\mathbf{u} = \lim \mathbf{u}_{\mathrm{n}}$, $\mathbf{c} = \lim \mathbf{c}_{\mathrm{n}}$, $\mathbf{t} = \lim \mathbf{t}_{\mathrm{n}}$ e $\mathbf{w} = \lim \mathbf{w}_{\mathrm{n}}$. Assim:

$$\mathbf{YY'XX'u} = a_1^2\mathbf{u}$$
$$\mathbf{Y'XX'Yc} = a_1^2\mathbf{c}$$
$$\mathbf{XX'YY't} = a_1^2\mathbf{t}$$
$$\mathbf{X'YY'Xw} = a_1^2\mathbf{w}.$$

Ocorrendo a convergência pela aplicação de algum critério de parada, ficam definidos os vetores $\mathbf{w}$ e $\mathbf{c}$, relativos ao maior autovalor a_1^2. Pelo problema de maximização das componentes com maior covariância, os vetores relativos ao maior autovalor são dados explicitamente a partir da decomposição em valores singulares $\mathbf{X'Y}$, isto é, $\mathbf{w} = \mathbf{z}_1$ e $\mathbf{c} = \mathbf{v}_1$. Portanto, as componentes com maior covariância são dadas por $\mathbf{t} = \mathbf{Xw}$ e $\mathbf{u} = \mathbf{Yc}$.

O próximo passo do algoritmo é definido por: primeiramente se calcula a covariância amostral entre as componentes $\mathbf{t} = \mathbf{Xw}$ e $\mathbf{u} = \mathbf{Yc}$, dada por $\mathrm{b} = \frac{\mathbf{u't}}{\mathbf{t't}}$.

Como o vetor $\mathbf{t}$ é a combinação linear das colunas de $\mathbf{X}$ que possue maior covariância com uma combinação linear das colunas de $\mathbf{Y}$, a ideia agora é obter uma nova matriz $\mathbf{X}$ de tal forma que as colunas desta nova matriz $\mathbf{X}$ sejam ortogonais ao vetor $\mathbf{t}$. A projeção ortogonal no subespaço ortogonal ao vetor $\mathbf{t}$ é dada por $\mathbf{I} - \frac{\mathbf{tt'}}{\mathbf{t't}}$. Veja:

$$\left(\mathbf{I} - \frac{\mathbf{tt'}}{\mathbf{t't}}\right)\left(\mathbf{I} - \frac{\mathbf{tt'}}{\mathbf{t't}}\right) = \mathbf{I} - \frac{\mathbf{tt'}}{\mathbf{t't}} - \frac{\mathbf{tt'}}{\mathbf{t't}} + \frac{\mathbf{tt'tt'}}{\mathbf{t't}} = \mathbf{I} - \frac{\mathbf{tt'}}{\mathbf{t't}}.$$

Portanto, a nova matriz $\mathbf{X}$ é dada por:

$$\left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\right)\mathbf{X}.$$

Assim,

$$\left(\mathbf{I} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\right)\mathbf{X} = \mathbf{X} - \frac{\mathbf{t}\mathbf{t}'}{\mathbf{t}'\mathbf{t}}\mathbf{X} = \mathbf{X} - \mathbf{t}\frac{\mathbf{t}'\mathbf{X}}{\mathbf{t}'\mathbf{t}} = \mathbf{X} - \mathbf{t}\mathbf{p}',$$

em que $\mathbf{p} = \frac{\mathbf{X}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$.

O vetor $\mathbf{p}$ representa o vetor de covariâncias amostrais entre as covariáveis dadas pelas colunas de $\mathbf{X}$ e o vetor de estimativas das variáveis preditoras $\mathbf{t}$, e a nova matriz $\mathbf{X}$ representa as componentes das covariáveis que não foram utilizadas na predição, sendo denominada matriz residual e que será utilizada na próxima iteração do algoritmo.

Uma construção semelhante é feita para a matriz $\mathbf{Y}$. O melhor preditor das variáveis respostas $\mathbf{y} = \begin{bmatrix} \mathbf{y}_1 \\ \mathbf{y}_2 \\ \vdots \\ \mathbf{y}_\mathrm{k} \end{bmatrix}$ é o vetor $\mathrm{cov}\,(\mathbf{t}, \mathbf{y}) = \begin{bmatrix} \mathrm{cov}\,(\mathbf{t}, \mathbf{y}_1) \\ \mathrm{cov}\,(\mathbf{t}, \mathbf{y}_2) \\ \vdots \\ \mathrm{cov}\,(\mathbf{t}, \mathbf{y}_\mathrm{k}) \end{bmatrix}$, que é estimado pelo vetor $\mathbf{c}$, em que,

$$\mathbf{c}_{\mathrm{k}\times 1} = \frac{\mathbf{Y}'\mathbf{t}}{\mathbf{t}'\mathbf{t}} = \begin{bmatrix} \mathbf{Y}_1' \\ \mathbf{Y}_2' \\ \vdots \\ \mathbf{Y}_\mathrm{k}' \end{bmatrix}_{\mathrm{k}\times\mathrm{n}} \begin{bmatrix} \mathrm{t}_1 \\ \mathrm{t}_2 \\ \vdots \\ \mathrm{t}_\mathrm{n} \end{bmatrix}_{\mathrm{n}\times 1} = \begin{bmatrix} \frac{\mathbf{Y}_1'\mathbf{t}}{\mathbf{t}'\mathbf{t}} \\ \frac{\mathbf{Y}_2'\mathbf{t}}{\mathbf{t}'\mathbf{t}} \\ \vdots \\ \frac{\mathbf{Y}_\mathrm{k}'\mathbf{t}}{\mathbf{t}'\mathbf{t}} \end{bmatrix}_{\mathrm{k}\times 1}.$$

Segue que,

$$\mathbf{tc}' = \begin{bmatrix} \mathrm{t}_1 \\ \mathrm{t}_2 \\ \vdots \\ \mathrm{t}_\mathrm{n} \end{bmatrix}_{\mathrm{n}\times 1} \left[\frac{\mathbf{Y}_1'\mathbf{t}}{\mathbf{t}'\mathbf{t}}, \frac{\mathbf{Y}_2'\mathbf{t}}{\mathbf{t}'\mathbf{t}}, \ldots, \frac{\mathbf{Y}_\mathrm{k}'\mathbf{t}}{\mathbf{t}'\mathbf{t}} \right]_{1\times \mathrm{k}} = \left[\frac{\mathbf{Y}_1'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\mathbf{t}, \frac{\mathbf{Y}_2'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\mathbf{t}, \ldots, \frac{\mathbf{Y}_\mathrm{k}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}\mathbf{t} \right]_{\mathrm{n}\times \mathrm{k}}.$$

As colunas da matriz $\mathbf{tc}'$ representam a parte que o vetor $\mathbf{t}$ explica de $\mathbf{Y}$. Tem-se, portanto, duas fontes de explicação da variável $\mathbf{Y}$, a saber, a parte de $\mathbf{t}$ que explica $\mathbf{u}$, que é uma combinação linear das colunas de $\mathbf{Y}$, e a parte de $\mathbf{t}$ que explica $\mathbf{Y}$. Portanto, a quantidade de informação que o vetor $\mathbf{t}$ tem do vetor $\mathbf{u}$ e dos vetores coluna da matriz $\mathbf{Y}$ é obtida pelo produto $\mathrm{b}\frac{\mathbf{t}'\mathbf{Y}_\mathrm{i}}{\mathbf{t}'\mathbf{t}}$.

Assim, ponderando a quantidade de informação pelo produto, a parte não explicada de $\mathbf{Y}$ é a nova matriz $\mathbf{Y}$, dada pela diferença:

$$\mathbf{Y} - \mathrm{b}\mathbf{t}\left(\frac{\mathbf{t}\mathbf{Y}'}{\mathbf{t}'\mathbf{t}}\right) = \mathbf{Y} - \mathrm{b}\mathbf{t}\mathbf{c}'.$$

Obtidas as novas matrizes $\mathbf{X}$ e $\mathbf{Y}$, o processo é repetido, ficando o algoritmo da forma:

Algoritmo para o PLS Amostral (HOSKULDSSON, 1988):

1. Seja $\mathbf{u}$ a primeira coluna de $\mathbf{Y}$
2. $\mathbf{w} = \frac{\mathbf{X}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$
3. Normalize $\mathbf{w}$
4. $\mathbf{t} = \mathbf{X}\mathbf{w}$
5. $\mathbf{c} = \frac{\mathbf{Y}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$

6. Normalize $\mathbf{c}$

7. $\mathbf{u} = \mathbf{Y}\mathbf{c}$

8. Se houver convergência vá para o passo 9, caso contrário, volte ao passo 2

9. $\mathbf{X}$-*loadings*: $\mathbf{p} = \frac{\mathbf{X}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$

10. $\mathbf{Y}$-*loadings*: $\mathbf{q} = \frac{\mathbf{Y}'\mathbf{u}}{\mathbf{u}'\mathbf{u}}$

11. Regressão ($\mathbf{u}$ sobre $\mathbf{t}$): $\mathrm{b} = \frac{\mathbf{u}'\mathbf{t}}{\mathbf{t}'\mathbf{t}}$

12. Matrizes Residuais: $\mathbf{X} \leftarrow \mathbf{X} - \mathbf{t}\mathbf{p}'$ e $\mathbf{Y} \leftarrow \mathbf{Y} - \mathrm{b}\mathbf{t}\mathbf{c}'$

Realizadas A iterações deste algoritmo, obtêm-se os vetores:

$$\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_\mathrm{A},$$

$$\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_\mathrm{A},$$

$$\mathbf{c}_1, \mathbf{c}_2, \ldots, \mathbf{c}_\mathrm{A},$$

$$\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_\mathrm{A},$$

$$\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_\mathrm{A}.$$

- É importante observar que da construção do algoritmo segue que a sequência dos vetores de componentes $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_\mathrm{A}$, formam uma sequência de vetores ortogonais.

- Os vetores $\mathbf{w}_1, \mathbf{w}_2, \ldots, \mathbf{w}_\mathrm{A}$, são vetores mutuamente ortogonais.

Demonstração. Do algoritmo, sabe-se que:

$$\begin{aligned}
\mathbf{X}_2 &= \left(\mathbf{I} - \frac{\mathbf{t}_1\mathbf{t}_1^{'}}{\mathbf{t}_1^{'}\mathbf{t}_1}\right)\mathbf{X}_1 \\
\mathbf{X}_3 &= \left(\mathbf{I} - \frac{\mathbf{t}_2\mathbf{t}_2^{'}}{\mathbf{t}_2^{'}\mathbf{t}_2}\right)\mathbf{X}_2 = \left(\mathbf{I} - \frac{\mathbf{t}_2\mathbf{t}_2^{'}}{\mathbf{t}_2^{'}\mathbf{t}_2}\right)\left(\mathbf{I} - \frac{\mathbf{t}_1\mathbf{t}_1^{'}}{\mathbf{t}_1^{'}\mathbf{t}_1}\right)\mathbf{X}_1 \\
&\vdots \\
\mathbf{X}_{\mathrm{j}} &= \left(\mathbf{I} - \frac{\mathbf{t}_{\mathrm{j}-1}\mathbf{t}_{\mathrm{j}-1}^{'}}{\mathbf{t}_{\mathrm{j}-1}^{'}\mathbf{t}_{\mathrm{j}-1}}\right)\mathbf{X}_{\mathrm{j}-1} \\
&= \left(\mathbf{I} - \frac{\mathbf{t}_{\mathrm{j}-1}\mathbf{t}_{\mathrm{j}-1}^{'}}{\mathbf{t}_{\mathrm{j}-1}^{'}\mathbf{t}_{\mathrm{j}-1}}\right)\left(\mathbf{I} - \frac{\mathbf{t}_{\mathrm{j}-2}\mathbf{t}_{\mathrm{j}-2}^{'}}{\mathbf{t}_{\mathrm{j}-2}^{'}\mathbf{t}_{\mathrm{j}-2}}\right)\ldots\left(\mathbf{I} - \frac{\mathbf{t}_1\mathbf{t}_1^{'}}{\mathbf{t}_1^{'}\mathbf{t}_1}\right)\mathbf{X}_1.
\end{aligned}$$

Assim, para $\mathrm{i} < \mathrm{j}$, tem-se:

$$\mathbf{X}_{\mathrm{j}} = \mathbf{Z}\left(\mathbf{I} - \frac{\mathbf{t}_{\mathrm{i}}\mathbf{t}_{\mathrm{i}}^{'}}{\mathbf{t}_{\mathrm{i}}^{'}\mathbf{t}_{\mathrm{i}}}\right)\mathbf{X}_{\mathrm{i}},$$

em que $\mathbf{Z}$ é o produto de matrizes $\left(\mathbf{I} - \frac{\mathbf{t}_{\mathrm{k}}\mathbf{t}_{\mathrm{k}}^{'}}{\mathbf{t}_{\mathrm{k}}^{'}\mathbf{t}_{\mathrm{k}}}\right)$, $\mathrm{k} = \mathrm{i}+1, \ldots, \mathrm{j}$.

De onde segue que, para $\mathrm{i} < \mathrm{j}$:

$$\begin{aligned}
\mathbf{X}_{\mathrm{j}}\mathbf{w}_{\mathrm{i}} &= \mathbf{Z}\left(\mathbf{X}_{\mathrm{i}} - \frac{\mathbf{t}_{\mathrm{i}}\mathbf{t}_{\mathrm{i}}^{'}}{\mathbf{t}_{\mathrm{i}}^{'}\mathbf{t}_{\mathrm{i}}}\mathbf{X}_{\mathrm{i}}\right)\mathbf{w}_{\mathrm{i}} = \mathbf{Z}\left(\mathbf{X}_{\mathrm{i}}\mathbf{w}_{\mathrm{i}} - \frac{\mathbf{t}_{\mathrm{i}}\mathbf{t}_{\mathrm{i}}^{'}}{\mathbf{t}_{\mathrm{i}}^{'}\mathbf{t}_{\mathrm{i}}}\mathbf{X}_{\mathrm{i}}\mathbf{w}_{\mathrm{i}}\right) \\
&= \mathbf{Z}\left(\mathbf{t}_{\mathrm{i}} - \frac{\mathbf{t}_{\mathrm{i}}\mathbf{t}_{\mathrm{i}}^{'}}{\mathbf{t}_{\mathrm{i}}^{'}\mathbf{t}_{\mathrm{i}}}\mathbf{t}_{\mathrm{i}}\right) = \mathbf{Z}\left(\mathbf{t}_{\mathrm{i}} - \mathbf{t}_{\mathrm{i}}\right) = 0.
\end{aligned}$$

Sabendo que $\mathbf{X}_{\mathrm{j}}^{'}\mathbf{Y}_{\mathrm{j}}\mathbf{Y}_{\mathrm{j}}^{'}\mathbf{X}_{\mathrm{j}}\mathbf{w}_{\mathrm{j}} = a_{\mathrm{j}}\mathbf{w}_{\mathrm{j}}$, obtém-se:

$$\mathbf{w}_{\mathrm{j}}^{'}\mathbf{w}_{\mathrm{i}} = \left(\frac{\mathbf{X}_{\mathrm{j}}^{'}\mathbf{Y}_{\mathrm{j}}\mathbf{Y}_{\mathrm{j}}^{'}\mathbf{X}_{\mathrm{j}}\mathbf{w}_{\mathrm{j}}}{a_{\mathrm{j}}}\right)^{'}\mathbf{w}_{\mathrm{i}} = \frac{\mathbf{w}_{\mathrm{j}}^{'}\mathbf{X}_{\mathrm{j}}^{'}\mathbf{Y}_{\mathrm{j}}\mathbf{Y}_{\mathrm{j}}^{'}\mathbf{X}_{\mathrm{j}}}{a_{\mathrm{j}}}\mathbf{w}_{\mathrm{i}}$$

$$= \frac{\mathbf{w}_{\mathrm{j}}^{'}\mathbf{X}_{\mathrm{j}}^{'}\mathbf{Y}_{\mathrm{j}}\mathbf{Y}_{\mathrm{j}}^{'}}{a_{\mathrm{j}}}\left(\mathbf{X}_{\mathrm{j}}\mathbf{w}_{\mathrm{i}}\right) = 0.$$

□

- Os vetores $\mathbf{w}_{\mathrm{i}}$ são ortogonais aos vetores $\mathbf{p}_{\mathrm{j}}$, para $\mathrm{i} < \mathrm{j}$.

  Demonstração.

$$\mathbf{w}_{\mathrm{i}}^{'}\mathbf{p}_{\mathrm{j}} = \mathbf{w}_{\mathrm{i}}^{'}\frac{\mathbf{X}_{\mathrm{j}}^{'}\mathbf{t}_{\mathrm{j}}}{\mathbf{t}_{\mathrm{j}}^{'}\mathbf{t}_{\mathrm{j}}} = \mathbf{w}_{\mathrm{i}}^{'}\mathbf{X}_{\mathrm{j}}^{'}\frac{\mathbf{t}_{\mathrm{j}}}{\mathbf{t}_{\mathrm{j}}^{'}\mathbf{t}_{\mathrm{j}}} = \left(\mathbf{X}_{\mathrm{j}}\mathbf{w}_{\mathrm{i}}\right)'\frac{\mathbf{t}_{\mathrm{j}}}{\mathbf{t}_{\mathrm{j}}^{'}\mathbf{t}_{\mathrm{j}}} = 0.$$

□

- A sequência de vetores $\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{\mathrm{A}}$, também é uma sequência de vetores ortogonais.

4.3.2.3 Regressão em PLS

Uma das questões mais importantes em análise de regressão é reduzir o número de variáveis independentes. Uma situação típica em que é vantajosa a seleção de variáveis é que a variância dos coeficientes de regressão aumenta quando novas variáveis são introduzidas no modelo (HOSKULDSSON, 1988). A regressão PLS pode ser vista como um bom método de análise de regressão porque os componentes são selecionados de modo que eles descrevem as variáveis respostas. A principal característica é que o método PLS é capaz de reduzir o número de variáveis, em termos de componentes, durante o processo de estimação. Vamos considerar aqui, a regressão via método PLS.

Um modelo de regressão linear pode ser escrito como:

$$\mathbf{Y}_{\mathrm{n\times k}} = \mathbf{X}_{\mathrm{n\times p}}\boldsymbol{\beta}_{\mathrm{p\times k}} + \mathbf{e}_{\mathrm{n\times k}},$$

em que $\boldsymbol{\beta}$ é o vetor dos coeficientes de regressão, $\mathbf{X}$ é a matriz do delineamento e $\mathbf{e}$ é a matriz residual.

Sendo obtidos os vetores $\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{\mathrm{A}}$, em A iterações, a matriz do delineamento $\mathbf{X}$ pode ser decomposta da forma

$$\mathbf{X}_{\mathrm{n\times p}} = \sum_{\mathrm{i=1}}^{\mathrm{A}} \mathbf{t}_{\mathrm{i}}\mathbf{p}_{\mathrm{i}}^{'} + \mathbf{X}_0,$$

onde $\mathbf{X}_0$ é a matriz residual do processo.

Note que se o número de iterações A é igual ao posto de $\mathbf{X}$, então $\mathbf{X}_0 = 0$.

Desprezando a matriz residual $\mathbf{X}_0$, tem-se

$$\mathbf{X}_{\mathrm{n\times p}} = \sum_{\mathrm{i=1}}^{\mathrm{A}} \mathbf{t}_{\mathrm{i}}\mathbf{p}_{\mathrm{i}}^{'} = \mathbf{T}_{\mathrm{n\times A}}\mathbf{P}_{\mathrm{A\times p}}^{'}$$

com $\mathbf{T}_{\mathrm{n\times A}} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_{\mathrm{A}})$ em que $\mathbf{t}_{\mathrm{i}}$ é a i-ésima coluna de $\mathbf{T}$ e $\mathbf{P}_{\mathrm{n\times A}} = (\mathbf{p}_1, \mathbf{p}_2, \ldots, \mathbf{p}_{\mathrm{A}})$ em que $\mathbf{p}_{\mathrm{i}}$ é a i-ésima coluna de $\mathbf{P}$.

Como os vetores $\mathbf{p}_{\mathrm{i}}$ são ortonormais, multiplica-se a igualdade $\mathbf{X}_{\mathrm{n\times p}} = \mathbf{T}_{\mathrm{n\times A}}\mathbf{P}_{\mathrm{A\times p}}^{'}$ por $\mathbf{P}_{\mathrm{p\times A}}$, obtendo

$$\mathbf{X}_{\mathrm{n\times p}}\mathbf{P}_{\mathrm{p\times A}} = \mathbf{T}_{\mathrm{n\times A}}\mathbf{P}_{\mathrm{A\times p}}^{'}\mathbf{P}_{\mathrm{p\times A}} = \mathbf{T}_{\mathrm{n\times A}}\mathbf{I}_{\mathrm{A\times A}} = \mathbf{T}_{\mathrm{n\times A}}.$$

Observe que $\mathbf{P}$ é uma inversa generalizada à direita de $\mathbf{P}'$. De fato, qualquer inversa generalizada poderia ser utilizada (JONG, 1993).

Assim, $\mathbf{T} = \mathbf{XP}$, e a equação de regressão pode ser escrita como

$$\mathbf{Y}_{\mathrm{n\times k}} = \mathbf{X}_{\mathrm{n\times p}}\mathbf{P}_{\mathrm{p\times A}}\mathbf{P}'_{\mathrm{A\times p}}\boldsymbol{\beta}_{\mathrm{p\times k}} + \mathbf{e}_{\mathrm{n\times k}} = \mathbf{T}_{\mathrm{n\times A}}\left(\boldsymbol{\beta}_{\mathrm{PLS,A}}\right)_{\mathrm{A\times k}} + \mathbf{e}_{\mathrm{n\times k}},$$

com $\mathbf{T}$ a nova matriz de covariáveis, denominada matriz de componentes PLS, e $\boldsymbol{\beta}_{\mathrm{PLS,A}} = \mathbf{P}'\boldsymbol{\beta}$ a nova matriz de parâmetros de regressão, denominada matriz de parâmetros PLS.

Com os novos parâmetros PLS, temos um problema de regressão usual, porém, agora, com o número de parâmetros muito menor que o problema original. O vetor de parâmetros, $\boldsymbol{\beta}_{\mathrm{PLS,A}}$, é estimado pelo método dos quadrados mínimos, isto é,

$$\left(\hat{\boldsymbol{\beta}}_{\mathrm{PLS,A}}\right)_{\mathrm{A\times k}} = \left(\mathbf{T}'\mathbf{T}\right)^{-1}_{\mathrm{A\times A}}\mathbf{T}'_{\mathrm{A\times n}}\mathbf{Y}_{\mathrm{n\times k}}. \tag{5}$$

Obtidas as estimativas dos coeficientes de regressão, a equação de predição para novos valores $\mathbf{x_0}$ das covariáveis, retornando às variáveis originais não centradas, fica da forma:

$$\hat{\mathbf{y}} - \hat{\boldsymbol{\mu}}_{\mathbf{y}} = \left(\mathbf{t} - \hat{\boldsymbol{\mu}}_{\mathbf{t}}\right)'\hat{\boldsymbol{\beta}}_{\mathrm{PLS,A}} = \left(\mathbf{x}_0 - \hat{\boldsymbol{\mu}}_{\mathbf{x}}\right)'\mathbf{P}\hat{\boldsymbol{\beta}}_{\mathrm{PLS,A}}. \tag{6}$$

4.3.3 Uma alternativa ao algoritmo PLS

Esta seção tem como referências os artigos Garthwaite (1994) e Phatak e Jong (1997), e será utilizada a mesma notação. A descrição deste algoritmo é feita para apenas uma resposta, $\mathrm{k} = 1$.

Como anteriormente exposto, a matriz de covariáveis $\mathbf{X}_{\mathrm{n}\times\mathrm{p}}$, é considerada como uma transformação do espaço dos parâmetros $\mathbb{R}^{\mathrm{p}}$ no espaço dos dados $\mathbb{R}^{\mathrm{n}}$, e o vetor da variável resposta é $\mathbf{Y}_{\mathrm{n}\times 1}$, conforme Figura 23.

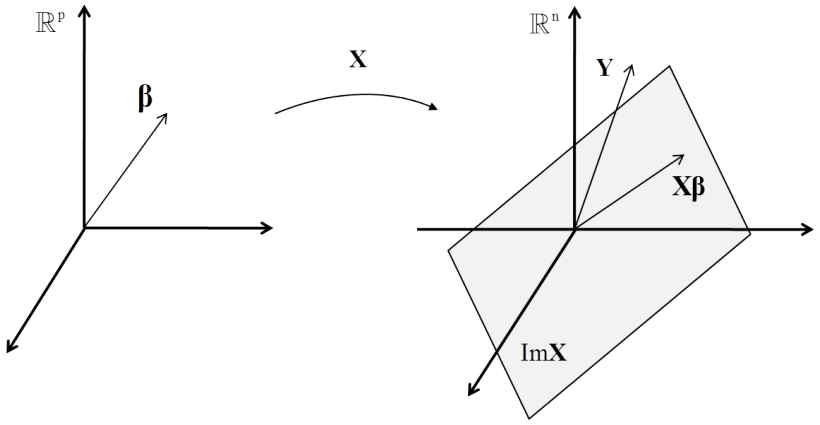


Figura 23 Representação da matriz de covariáveis $\mathbf{X}_{\mathrm{n}\times\mathrm{p}}$ como uma transformação do espaço de parâmetros $\mathbb{R}^{\mathrm{p}}$ no espaço dos dados $\mathbb{R}^{\mathrm{n}}$

A matriz $\mathbf{X}$ aplicada ao vetor canônico $\mathbf{e}_{\mathrm{i}} = (0, ..., 0, 1, 0, ..., 0)$ tem como imagem a i-ésima coluna de $\mathbf{X}$, denotada por $\mathbf{X}_{\mathrm{i}}$ (Figura 24).

Como é usual, utilizam-se variáveis centradas na média:

$$\mathbf{U}_1 = \mathbf{Y} - \bar{\mathbf{Y}} \ \text{ e } \ \mathbf{V}_{1\mathrm{j}}{=}\mathbf{X}_{\mathrm{j}}{-}\bar{\mathbf{X}}_{\mathrm{j}} \ \text{ para } \ \mathrm{j} = 1, ..., \mathrm{p},$$

sendo $\bar{\mathbf{Y}}' = (\bar{\mathrm{y}}, \bar{\mathrm{y}}, ..., \bar{\mathrm{y}})'$ e $\mathbf{1}' = (1, 1, ..., 1)'$, vetores linha n-dimensionais. Observe, como na Figura 25, que os vetores $\mathbf{Y} - \bar{\mathbf{Y}}$ e $\mathbf{X}_{\mathrm{j}}{-}\bar{\mathbf{X}}_{\mathrm{j}}$ são ortogonais ao

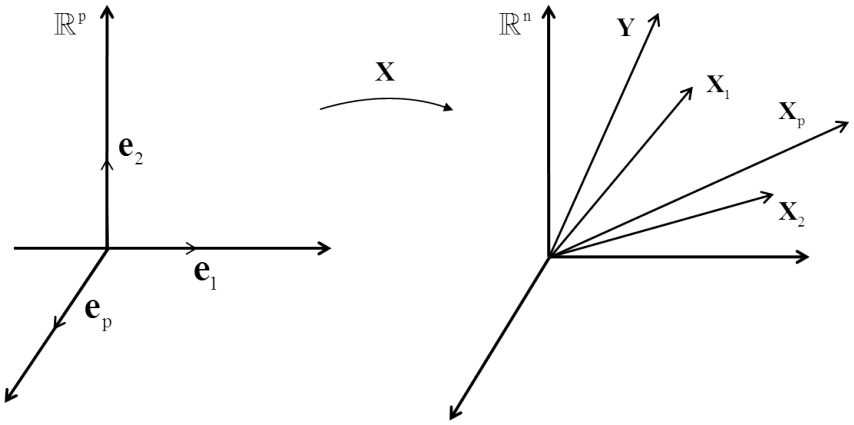


Figura 24 Representação das colunas de $\mathbf{X}$ como imagem dos vetores canônicos $\mathbf{e}_i$

subespaço gerado pelo vetor $\mathbf{1}$.

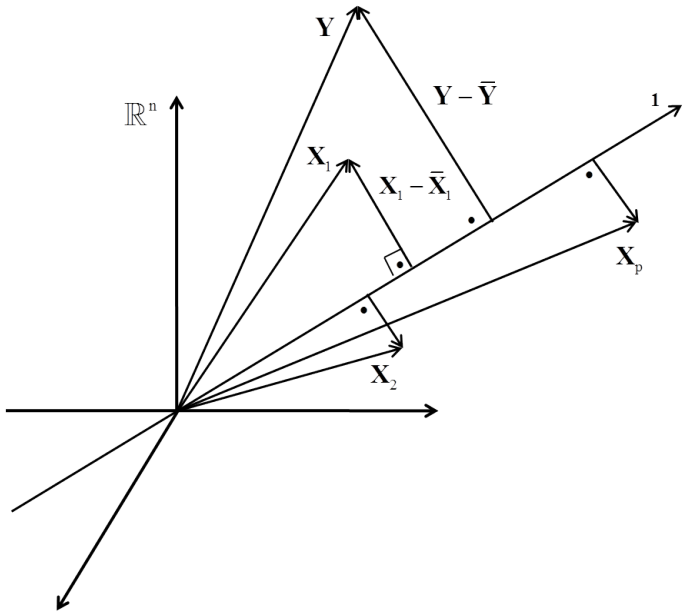


Figura 25 Representação das variáveis centradas através da projeção ortogonal no subespaço gerado pelo vetor $\mathbf{1}$

A ideia é regredir o vetor $\mathbf{U}_1$ em cada um dos vetores $\mathbf{V}_{1j}$ (Figuras 26 e 27). Tal procedimento é uma forma de medir o quanto de informação o vetor da covariável $\mathbf{X}_j$ explica o vetor de dados, obtendo-se o vetor $\mathbf{U}_{1j} = \frac{\mathbf{u}_1' \mathbf{v}_{1j}}{\mathbf{v}_{1j}' \mathbf{v}_{1j}} \mathbf{v}_{1j}$.

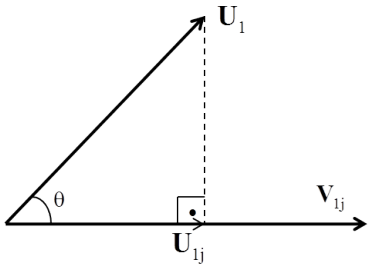


Figura 26 Projeção ortogonal do vetor $\mathbf{U}_1$ em cada um dos vetores $\mathbf{V}_{1j}$

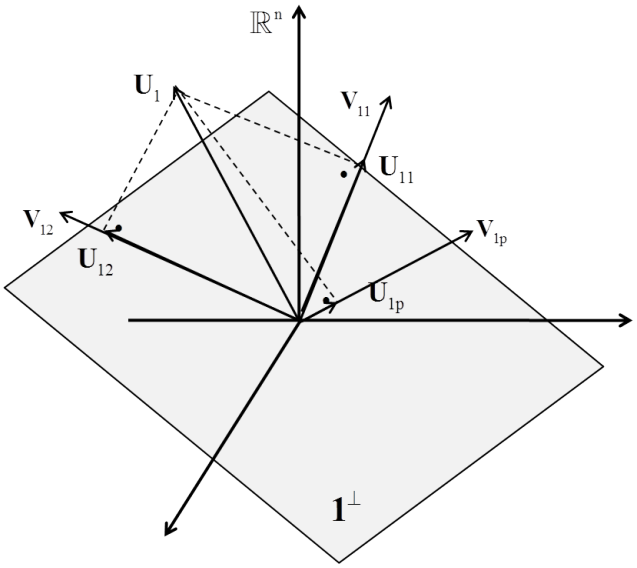


Figura 27 Projeção ortogonal do vetor $\mathbf{U}_1$ em cada um dos vetores $\mathbf{V}_{1j}$

Uma maneira de se coletarem todas estas informações sobre o vetor de dados é através de uma média ponderada dos vetores $\mathbf{U}_{1j}$, utilizando pesos apro-

priadamente escolhidos, $\mathrm{w_{1j}} \geq 0$ e $\sum_{\mathrm{j=1}}^{\mathrm{p}} \mathrm{w_{1j}} = 1$, como na Figura 28.

$$\mathbf{T}_1 = \sum_{\mathrm{j=1}}^{\mathrm{p}} \mathrm{w_{1j}} \mathbf{U}_{1\mathrm{j}}. \tag{7}$$

Uma escolha usual para os pesos é a média simples, $\mathrm{w_{1j}} = 1/\mathrm{p}$.

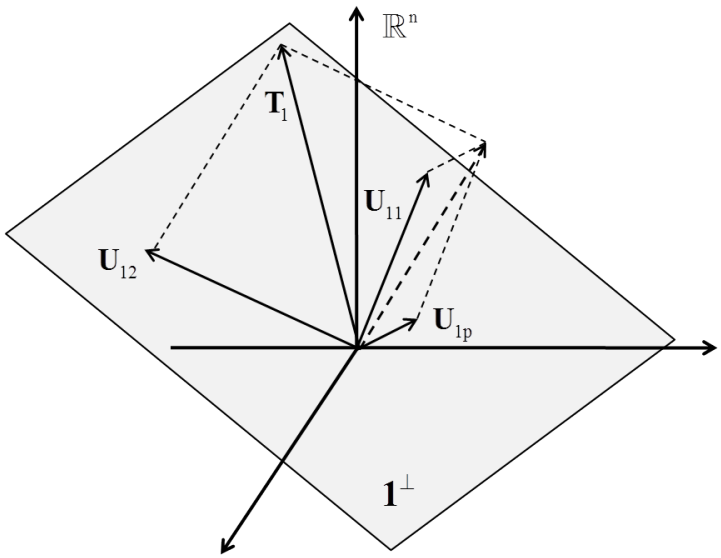


Figura 28 Construção do vetor $\mathbf{T}_1$ como média ponderada dos vetores $\mathbf{U}_{1\mathrm{j}}$

A quantidade de informação que $\mathbf{T}_1$ contém do vetor de dados $\mathbf{U}_1$ é obtida pela regressão do vetor $\mathbf{U}_1$ em $\mathbf{T}_1$, $P_{\mathbf{T}_1}\mathbf{U}_1 = \frac{\mathbf{u}_1'\mathbf{t}_1}{\mathbf{t}_1'\mathbf{t}_1}\mathbf{t}_1$. A parte não explicada é definida pelo vetor de resíduo $\mathbf{U}_2 = \mathbf{U}_1 - P_{\mathbf{T}_1}\mathbf{U}_1$. Novamente, regride-se $\mathbf{V}_{1\mathrm{j}}$ em $\mathbf{T}_1$, $P_{\mathbf{T}_1}\mathbf{V}_{1\mathrm{j}} = \frac{\mathbf{v}_{1\mathrm{j}}'\mathbf{t}_1}{\mathbf{t}_1'\mathbf{t}_1}\mathbf{t}_1$, e toma-se os vetores de resíduos $\mathbf{V}_{2\mathrm{j}} = \mathbf{V}_{1\mathrm{j}} - P_{\mathbf{T}_1}\mathbf{V}_{1\mathrm{j}}$, que são uma medida do quanto o vetor $\mathbf{T}_1$ não explica as covariáveis, conforme Figura 29.

Procede-se a regressão do vetor $\mathbf{U}_2$ nos vetores $\mathbf{V}_{2\mathrm{j}}$, que são denotados por $\mathbf{U}_{2\mathrm{j}} = P_{\mathbf{V}_{2\mathrm{j}}}\mathbf{U}_2 = b_{2\mathrm{j}}\mathbf{V}_{2\mathrm{j}} = \frac{\mathbf{U}_2'\mathbf{V}_{2\mathrm{j}}}{\mathbf{V}_{2\mathrm{j}}'\mathbf{V}_{2\mathrm{j}}}\mathbf{V}_{2\mathrm{j}}$ (Figura 30).

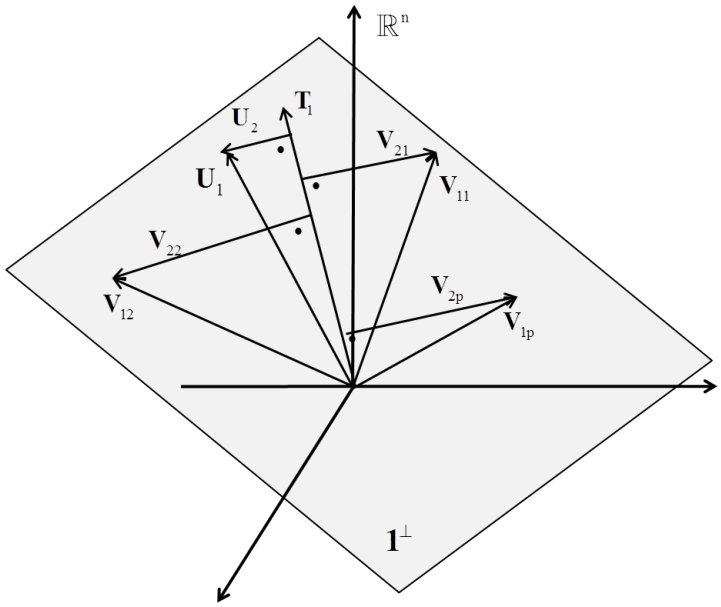


Figura 29 Construção dos vetores $\mathbf{U}_2$ e $\mathbf{V}_{2j}$ através da projeção dos vetores $\mathbf{U}_1$ e $\mathbf{V}_{1j}$, respectivamente, em $\mathbf{T}_1$

Utilizando pesos w_{2j}, encontramos o vetor $\mathbf{T}_2$ fazendo a média ponderada dos vetores $\mathbf{U}_{2j}$ (Figura 31).

Observe que, por construção, os vetores $\mathbf{T}_1$ e $\mathbf{T}_2$ são ortogonais (Figura 32).

O procedimento pode ser repetido para a construção de outros vetores $\mathbf{T}_3$,$\mathbf{T}_4$, ..., $\mathbf{T}_m$.

Suponha que $\mathbf{T}_i$ $(i \geq 1)$ tenha sido construído a partir das variáveis $\mathbf{U}_i$ e $\mathbf{V}_{ij}$ com $j = 1, ..., p$. Para obter o componente $\mathbf{T}_{i+1}$, as variáveis $\mathbf{V}_{(i+1)j}$ e $\mathbf{U}_{(i+1)}$ devem ser determinadas. Com este propósito, é feita uma regressão entre $\mathbf{V}_{ij}$ e $\mathbf{T}_i$ e $\mathbf{V}_{(i+1)j}$ fica definido por:

$$\mathbf{V}_{(i+1)j} = \mathbf{V}_{ij} - \frac{\mathbf{v}_{ij}^{'}\mathbf{t}_i}{\mathbf{t}_i^{'}\mathbf{t}_i}\mathbf{t}_i,$$

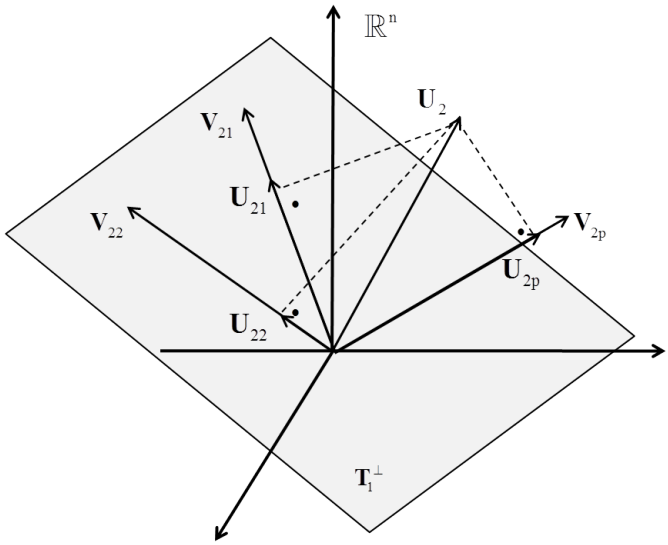


Figura 30 Construção dos vetores $\mathbf{U}_{2j}$ como projeção do vetor $\mathbf{U}_2$ nos vetores $\mathbf{V}_{2j}$

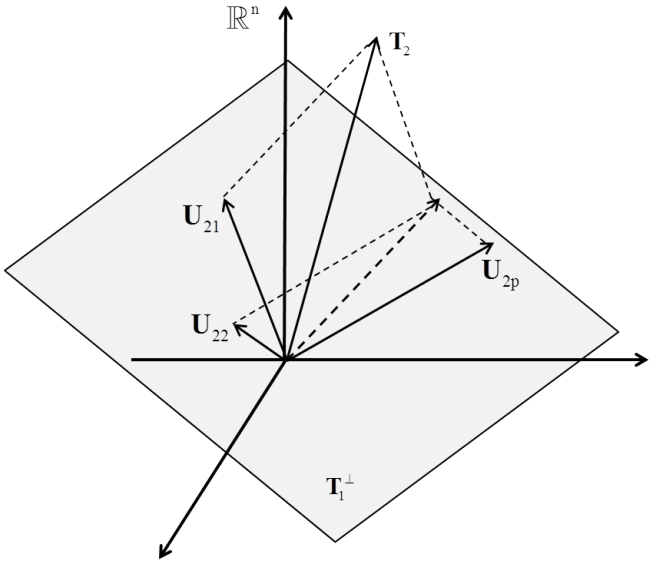


Figura 31 Construção do vetor $\mathbf{T}_2$ como média ponderada dos vetores $\mathbf{U}_{2j}$

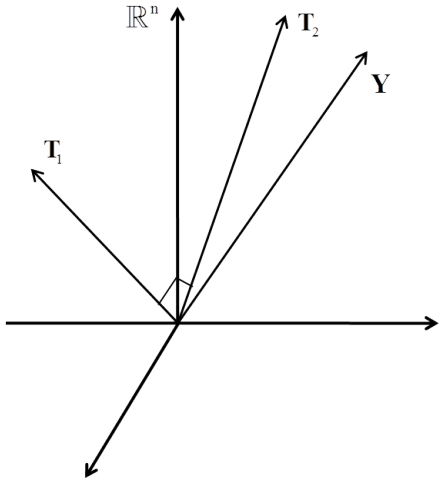


Figura 32 Componentes ortogonais $\mathbf{T}_1$ e $\mathbf{T}_2$

sendo $\mathbf{t}_\mathrm{i}$ o vetor de valores de $\mathbf{T}_\mathrm{i}$, $\mathbf{v}_{(\mathrm{i}+1)\mathrm{j}}$ os resíduos da regressão e $\frac{\mathbf{v}_{\mathrm{ij}}^{'}\mathbf{t}_\mathrm{i}}{\mathbf{t}_\mathrm{i}^{'}\mathbf{t}_\mathrm{i}}$, o coeficiente da regressão entre $\mathbf{V}_{\mathrm{ij}}$ e $\mathbf{T}_\mathrm{i}$.

De forma análoga, $\mathbf{U}_{(\mathrm{i}+1)}=\mathbf{U}_\mathrm{i}-\frac{\mathbf{u}_\mathrm{i}^{'}\mathbf{t}_\mathrm{i}}{\mathbf{t}_\mathrm{i}^{'}\mathbf{t}_\mathrm{i}}\mathbf{t}_\mathrm{i}$ e $\mathbf{u}_{(\mathrm{i}+1)}$ são os resíduos da regressão entre $\mathbf{U}_\mathrm{i}$ e $\mathbf{T}_\mathrm{i}$. Assim, a j-ésima regressão entre $\mathbf{U}_{\mathrm{i}+1}$ e $\mathbf{V}_{(\mathrm{i}+1)\mathrm{j}}$ resulta num coeficiente de regressão dado por:

$$\mathrm{b}_{(\mathrm{i}+1)\mathrm{j}}=\frac{\mathbf{u}_{(\mathrm{i}+1)}^{'}\mathbf{v}_{(\mathrm{i}+1)\mathrm{j}}}{\mathbf{v}_{(\mathrm{i}+1)\mathrm{j}}^{'}\mathbf{v}_{(\mathrm{i}+1)\mathrm{j}}}.$$

Analogamente à equação 7, define-se $\mathbf{T}_{(\mathrm{i}+1)}$ como sendo:

$$\mathbf{T}_{(\mathrm{i}+1)}=\sum_{\mathrm{j}=1}^{\mathrm{p}}\mathrm{w}_{(\mathrm{i}+1)\mathrm{j}}\mathrm{b}_{(\mathrm{i}+1)\mathrm{j}}\mathbf{V}_{(\mathrm{i}+1)\mathrm{j}}.$$

Os vetores $\mathbf{T}_1$,$\mathbf{T}_2$, ..., $\mathbf{T}_\mathrm{m}$ são ortogonais e, portanto, como vetores aleatórios, são não correlacionados (no caso em que $\mathrm{cov}(\mathbf{Y}) = \sigma^2\mathbf{I}$).

Construídos os vetores $\mathbf{T}_1$,$\mathbf{T}_2$, ..., $\mathbf{T}_\mathrm{m}$ fica definido na Im$\mathbf{X}$, um subespa-

ço m-dimensional, Im$\mathbf{T}$. Projeta-se ortogonalmente o vetor $\mathbf{Y}$ no subespaço Im$\mathbf{T}$ obtendo-se $\hat{\mathbf{Y}}_{\mathrm{PLS}} = P_{\mathrm{Im}\mathbf{T}}\mathbf{Y}$. O processo de estimação do $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ é o usual utilizando o método dos quadrados mínimos, isto é, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ é obtido como uma solução particular, bem determinada, das equações normais $\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{PLS}} = P_{\mathrm{Im}\mathbf{T}}\mathbf{Y}$. Uma vez obtida a estimativa $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, a equação de predição fica da forma:

$$\mathbf{y} - \bar{\mathbf{y}} = \hat{\boldsymbol{\beta}}_{\mathrm{PLS}}^{'} \begin{bmatrix} \mathrm{t}_1 \\ \vdots \\ \mathrm{t}_{\mathrm{m}} \end{bmatrix}.$$

Para obter a expressão matricial das equações normais que definem o $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ e também obter uma relação linear entre $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ e $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, foi utilizada uma formalização do método em termos de matrizes, isto é, transformações lineares.

Suponha que m componentes $\mathbf{t}_1$,$\mathbf{t}_2$, ..., $\mathbf{t}_{\mathrm{m}}$ tenham sido construídos. Considere a matriz $\mathbf{T}_{\mathrm{n}\times\mathrm{m}}$, cujas colunas são definidas pelos vetores $\mathbf{t}_{\mathrm{i}}$, $\mathbf{T} = [\mathbf{t}_1,\mathbf{t}_2, ..., \mathbf{t}_{\mathrm{m}}]$. $\mathbf{T}$ define uma transformação linear do $\mathbb{R}^{\mathrm{m}}$ para o $\mathbb{R}^{\mathrm{n}}$, tal que $\mathbf{T}\mathbf{e}_{\mathrm{i}} = \mathbf{t}_{\mathrm{i}}$, conforme Figura 33.

A imagem de $\mathbf{T}$, Im$\mathbf{T}$, é um subespaço da imagem de $\mathbf{X}$ gerado pelos vetores $\mathbf{t}_1$,$\mathbf{t}_2$, ...$\mathbf{t}_{\mathrm{m}}$. Portanto, a projeção ortogonal sobre a Im$\mathbf{T}$, é dada pela matriz $\mathbf{T}(\mathbf{T}'\mathbf{T})^{-1}\mathbf{T}'$, e a projeção do vetor de dados $\mathbf{Y}$ na Im$\mathbf{T}$ é dada por:

$$\hat{\mathbf{y}}_{\mathrm{PLS}}^{\mathrm{m}} = \mathbf{T}_{\mathrm{m}}\big(\mathbf{T}_{\mathrm{m}}'\mathbf{T}_{\mathrm{m}}\big)^{-1}\mathbf{T}_{\mathrm{m}}'\mathbf{y}, \tag{8}$$

conforme Figura 34.

Seja o vetor $\mathbf{r}_{\mathrm{i}}$ definido como uma solução qualquer da equação $\mathbf{X}\mathbf{r}_{\mathrm{i}} = \mathbf{t}_{\mathrm{i}}$, i=1,...,m. Da mesma forma, pode-se definir a matriz $\mathbf{R}_{\mathrm{p}\times\mathrm{m}}$, cujas colunas são os vetores $\mathbf{r}_{\mathrm{i}}$, i=1,...,m, $\mathbf{R}_{\mathrm{m}} = [\mathbf{r}_1, \mathbf{r}_2, ..., \mathbf{r}_{\mathrm{m}}]$. Assim, $\mathbf{R}\mathbf{e}_{\mathrm{i}} = \mathbf{r}_{\mathrm{i}}$.

Como $\mathbf{T}\mathbf{e}_{\mathrm{i}} = \mathbf{t}_{\mathrm{i}}$ e $\mathbf{X}\mathbf{R}\mathbf{e}_{\mathrm{i}} = \mathbf{t}_{\mathrm{i}}$, tem-se que $\mathbf{T}\mathbf{e}_{\mathrm{i}} = \mathbf{X}\mathbf{R}\mathbf{e}_{\mathrm{i}}$. Logo, $\mathbf{T} = \mathbf{X}\mathbf{R}$

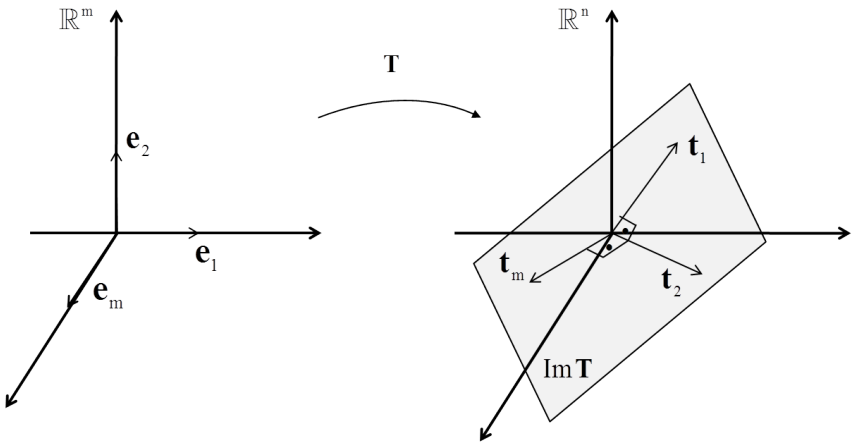


Figura 33 Representação geométrica dos componentes como imagem da transformação linear **T** do $\mathbb{R}^m$ para o $\mathbb{R}^n$

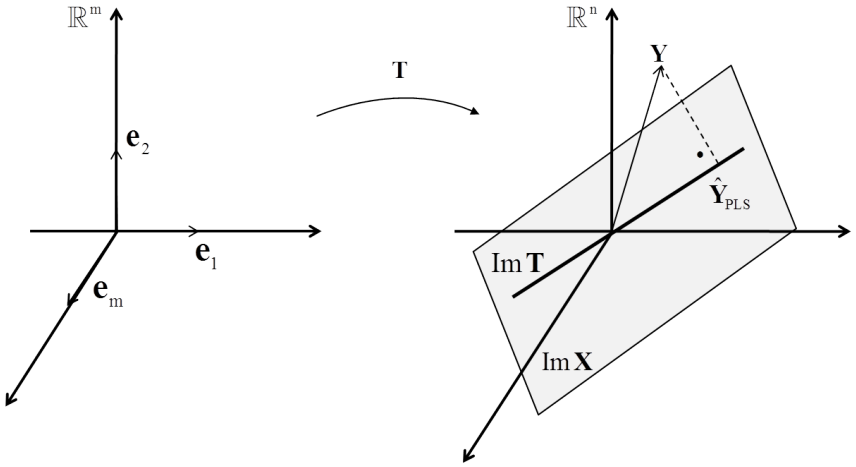


Figura 34 Projeção do vetor de dados **Y** no subespaço Im**T** gerado pelos componentes

e obtém-se o diagrama comutativo da Figura 35.

A matriz $\mathbf{T}$ é claramente uma aplicação linear injetiva, pois, suas colunas são linearmente independentes. Assim, restrita à imagem de **R**, segue que **X** também é uma aplicação injetiva. Vetorialmente, $\mathrm{Im}\mathbf{R} \cap \mathrm{Ker}\mathbf{X} = \{\mathbf{0}\}$.

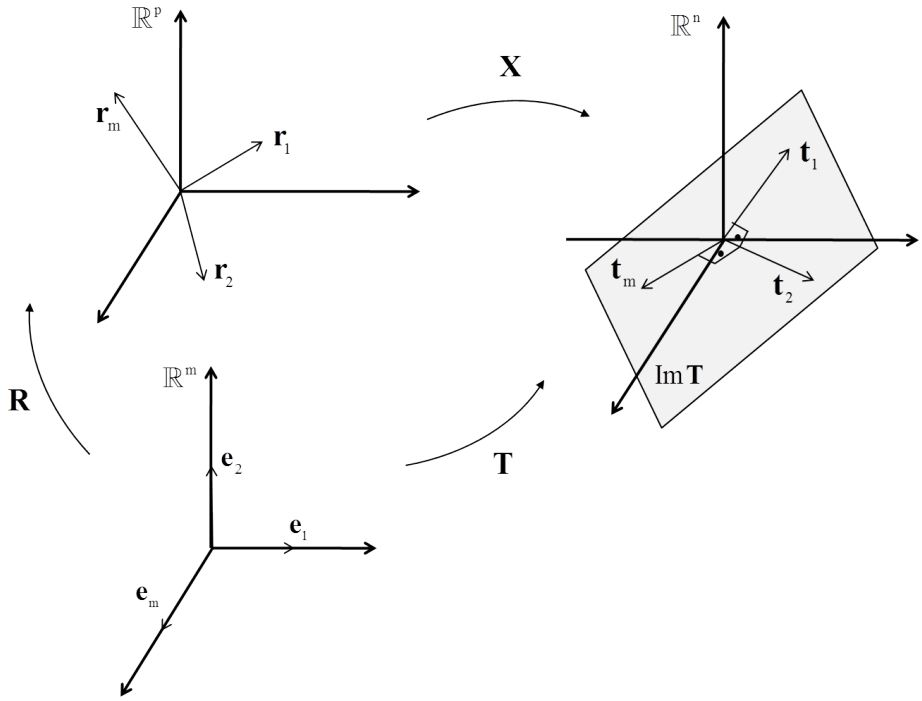


Figura 35 Representação geomética das transformações lineares **T**, **R** e **X**

Substituindo $\mathbf{T}$ por $\mathbf{XR}$ e $\mathbf{X}'\mathbf{y}$ por $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$, na equação (8), obtemos:

$$\hat{\mathbf{y}}^{\mathrm{m}}_{\mathrm{PLS}} = \mathbf{X}\mathbf{R}_{\mathrm{m}}\left(\mathbf{R}_{\mathrm{m}}^{'}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}_{\mathrm{m}}^{'}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}.$$

Observe que a substituição de $\mathbf{X}'\mathbf{y}$ por $\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ é verdadeira em relação a qualquer $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ escolhido.

Utilizando o fato de **X** ser injetiva quando restrita à imagem de **R**, temos que $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ está bem definido e é dado por:

$$\hat{\boldsymbol{\beta}}^{\mathrm{m}}_{\mathrm{PLS}} = \mathbf{R}_{\mathrm{m}}\left(\mathbf{R}_{\mathrm{m}}^{'}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}_{\mathrm{m}}^{'}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}. \tag{9}$$

A escolha particular dos vetores $\mathbf{r}_{\mathrm{i}}$, não afeta a estimativa de $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, pois

se $\mathbf{A}$ é uma transformação linear inversível, $\mathbf{A} : \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{m}}$, a mudança de coordenadas $\tilde{\mathbf{R}} = \mathbf{RA}$ é tal que, substituindo na equação (9), é obtida a mesma estimativa:

$$\begin{aligned}
\hat{\boldsymbol{\beta}}^{\mathrm{m}}_{\mathrm{PLS}} &= \tilde{\mathbf{R}}_{\mathrm{m}}\left(\tilde{\mathbf{R}}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\tilde{\mathbf{R}}_{\mathrm{m}}\right)^{-1}\tilde{\mathbf{R}}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \\
&= \mathbf{R}_{\mathrm{m}}\mathbf{A}\left(\mathbf{A}'\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\mathbf{A}\right)^{-1}\mathbf{A}'\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \\
&= \mathbf{R}_{\mathrm{m}}\mathbf{A}\left(\mathbf{A}'\right)^{-1}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}(\mathbf{A})^{-1}\mathbf{A}'\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}} \\
&= \mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}.
\end{aligned}$$

Este fato reflete que apenas o subespaço gerado pelos vetores $\mathbf{r}_{\mathrm{i}}$ é importante, e não uma escolha particular da base.

É possível obter uma relação linear entre $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$ e $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$ que possui uma interpretação geométrica muito interessante. Observe que:

$$\begin{aligned}
&\left(\mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\right)\left(\mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\right) \\
&= \mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X} \\
&= \mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{I}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X} \\
&= \mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}.
\end{aligned}$$

Logo, essa transformação é um projetor. Como ela não é simétrica, esse projetor não é um projetor ortogonal e, sim, um projetor oblíquo, portanto, como todo projetor, a projeção se dá ao longo do kernel (Figura 36).

Como as matrizes $\mathbf{R}$ e $\mathbf{X}'\mathbf{X}$ são injetivas, $\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}$ é inversível e essas transformações não têm kernel, portanto, o kernel da transformação

$$\mathbf{R}_{\mathrm{m}}\left(\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X}\mathbf{R}_{\mathrm{m}}\right)^{-1}\mathbf{R}^{'}_{\mathrm{m}}\mathbf{X}'\mathbf{X},$$

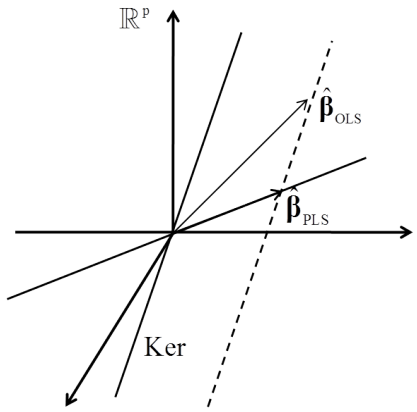


Figura 36 Vetor $\hat{\boldsymbol{\beta}}_{\text{PLS}}$ como projeção oblíqua do vetor $\hat{\boldsymbol{\beta}}_{\text{OLS}}$

é exatamente o kernel de $\mathbf{R}_{\text{m}}^{'}$. Note também que, o kernel de $\mathbf{X}$ está contido no kernel de $\mathbf{R}_{\text{m}}^{'}\mathbf{X}'\mathbf{X}$.

4.4 Determinação do número ótimo de componentes (variáveis latentes)

Um fator fundamental na utilização do PLS é a escolha do número ótimo de componentes a serem utilizados no modelo. Não existe uma regra formal para determinar esse número, porém, na literatura, um critério de decisão utilizado para a metodologia PCR é adotar uma percentagem da variação total explicada que se deseja obter e usar o número de componentes que atinja esse valor pré-determinado. Essa abordagem empírica vem sendo usada na prática, geralmente fixando uma porcentagem de 70% (AZEVEDO, 2012). Porém, por se tratar de um enfoque empírico, alternativas provenientes de derivações teóricas ou de recursos computacionais intensivos são mais adequadas, como por exemplo as teorias Graus de Liberdade e Validação Cruzada.

4.4.1 Graus de Liberdade e seleção de modelos

Esta seção tem como referência o artigo Krämer e Sugiyama (2011).

Em regressão, o Grau de Liberdade quantifica a complexidade intrínseca de um modelo. Para modelos de Regressão Linear, em que os valores ajustados são funções lineares dos dados, $\mathbf{y}$, isto é, $\hat{\mathbf{y}}_\lambda = \mathbf{H}_\lambda \mathbf{y}$ com $\mathbf{H}_\lambda \in \mathbb{R}^{\mathrm{n}\times\mathrm{n}}$, em que $\mathbf{H}$ é uma matriz de projeção (*hat-matrix*), o grau de liberdade é definido como o traço da matriz de projeção:

$$\mathrm{DoF}\left(\lambda\right) = tr\left(\mathbf{H}_\lambda\right). \tag{10}$$

Observe que o traço de $\mathbf{H}$ é igual à dimensão do subespaço definido pelo modelo (em geral, o subespaço gerado pelos vetores colunas da matriz de covariáveis). No caso em que a matriz de covariáveis é de posto completo, o DoF é exatamente igual ao número de parâmetros. Como exemplo, essa definição aplica-se à Regressão de Componentes Principais.

Na regressão PLS, tem-se, conforme a equação (8), que

$$\hat{\mathbf{y}}^{\mathrm{m}}_{\mathrm{PLS}} = \bar{\mathbf{y}} + \mathbf{T}_{\mathrm{m}}\left(\mathbf{T}'_{\mathrm{m}}\mathbf{T}_{\mathrm{m}}\right)^{-1}\mathbf{T}'_{\mathrm{m}}\mathbf{y} = \bar{\mathbf{y}} + \mathbf{P}_{\mathbf{T}}\mathbf{y}.$$

Como as componentes dependem não somente das covariáveis, mas também dos dados, a regressão PLS não é linear. Portanto, a definição (10) não pode ser aplicada. É necessário, então, usar uma generalização proposta por Efron (2004).

Definição 4.1. *Faça $\hat{f}_\lambda$ ser uma estimativa da verdadeira função de regressão f, parametrizada por λ. Defina o vetor de valores ajustados como $\hat{\mathbf{y}}_\lambda = \left(\hat{f}_\lambda(\mathbf{x}_1), \ldots, \hat{f}_\lambda(\mathbf{x}_\mathrm{n})\right)'$. Os graus de liberdade são*

$$\mathrm{DoF}(\lambda) = \mathrm{E}\left(tr\left(\frac{\partial \hat{\mathbf{y}}_\lambda}{\partial \mathbf{y}}\right)\right).$$

Assumindo $\mathbf{X}$ fixada e a esperança E tomada com relação a $\mathrm{y}_1, \ldots \mathrm{y}_\mathrm{n}$.

Neste sentido, o grau de liberdade é uma medida da sensibilidade dos valores ajustados como função dos dados.

Para o caso PLS, o DoF fica definido como:

$$\mathrm{DoF}(\mathrm{m}) = 1 + \mathrm{E}\left(tr\left(\frac{\partial \mathrm{P}_{\mathbf{T}}\mathbf{y}}{\partial \mathbf{y}}\right)\right).$$

A constante 1 corresponde à estimação do intercepto, uma vez que a matriz $\mathbf{X}$ foi considerada centrada, e m é o número de componentes utilizados no modelo.

Um estimador não viesado para o Grau de Liberdade da regressão PLS com m componentes latentes, $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_\mathrm{m})$, é dado por:

$$\hat{\mathrm{DoF}}(\mathrm{m}) = 1 + tr\left(\frac{\partial \mathrm{P}_{\mathbf{T}}\mathbf{y}}{\partial \mathbf{y}}\right).$$

Uma conjectura importante, afirma que $\mathrm{DoF}(\mathrm{m}) \geq \mathrm{m}+1$ (FRANK; FRIEDMAN, 1993; MARTENS; NAES, 1989). Para exemplificar, será calculado explicitamente uma cota inferior para o DoF quando $\mathrm{m} = 1$.

Seja $\mathbf{S} = \frac{1}{\mathrm{n}-1}\mathbf{X}'\mathbf{X} \in \mathbb{R}^{\mathrm{p}\times\mathrm{p}}$ a matriz de covariância amostral do grupo $\mathbf{X}$, e $\mathbf{s} = \frac{1}{\mathrm{n}-1}\mathbf{X}'\mathbf{y} \in \mathbb{R}^{\mathrm{p}}$, a covariância amostral entre os grupos $\mathbf{X}$ e $\mathbf{Y}$.

Teorema 2. *Se o maior autovalor* $\lambda_{\max}$ *da matriz de covariância amostral* $\mathbf{S}$ *satisfaz*

$$\lambda_{\max} \leq \frac{1}{2} tr\left(\mathbf{S}\right),$$

então

$$\hat{\mathrm{DoF}}\left(\mathrm{m}=1\right) \geq 1 + \frac{tr\left(\mathbf{S}\right)}{\lambda_{\max}}.$$

Demonstração. Sabendo que o primeiro componente é definido por $\mathbf{t}_1 = \mathbf{Xs}$, tem-se:

$$\frac{\mathbf{t}_1}{\|\mathbf{t}_1\|} = \frac{\mathbf{t}_1}{\sqrt{\mathbf{t}_1{}'\mathbf{t}_1}} = \frac{\mathbf{Xs}}{\sqrt{\left(\mathbf{Xs}\right)'\mathbf{Xs}}} = \frac{\mathbf{Xs}}{\sqrt{\mathbf{s}'\mathbf{X}'\mathbf{Xs}}} = \frac{\mathbf{Xs}}{\sqrt{\mathbf{s}'\mathbf{Ss}}},$$

de onde segue que a projeção de $\mathbf{y}$ no primeiro componente $\mathbf{t}_1$, é dada por:

$$\begin{aligned}\left(\mathbf{y}\cdot\frac{\mathbf{t}_1}{\|\mathbf{t}_1\|}\right)\frac{\mathbf{t}_1}{\|\mathbf{t}_1\|} = \left(\mathbf{y}\cdot\frac{\mathbf{Xs}}{\sqrt{\mathbf{s}'\mathbf{Ss}}}\right)\frac{\mathbf{Xs}}{\sqrt{\mathbf{s}'\mathbf{Ss}}} &= \frac{\left(\mathbf{y}'\mathbf{X}\right)\mathbf{s}}{\sqrt{\mathbf{s}'\mathbf{Ss}}}\frac{\mathbf{Xs}}{\sqrt{\mathbf{s}'\mathbf{Ss}}} \\ = \frac{\mathbf{s}'\mathbf{s}}{\sqrt{\mathbf{s}'\mathbf{Ss}}}\frac{\mathbf{Xs}}{\sqrt{\mathbf{s}'\mathbf{Ss}}} &= \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{Ss}}\mathbf{Xs}.\end{aligned}$$

Assim,

$$\hat{\mathbf{y}}_1 = \bar{\mathbf{y}} + \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{Ss}}\mathbf{Xs}. \tag{11}$$

Calculando a derivada de (11) em relação a $\mathbf{y}$ e sendo a derivada de uma forma quadrática dada por $\frac{\partial}{\partial\mathbf{y}}\left(\mathbf{y}'\mathbf{Ay}\right) = 2\left(\mathbf{Ay}\right)'$, tem-se:

$$\frac{\partial}{\partial\mathbf{y}}\left(\bar{\mathbf{y}}\right) = \begin{bmatrix} 1/\mathrm{n} & \cdots & 1/\mathrm{n} \\ \vdots & \ddots & \vdots \\ 1/\mathrm{n} & \cdots & 1/\mathrm{n} \end{bmatrix} \Rightarrow tr\left(\frac{\partial}{\partial\mathbf{y}}\left(\bar{\mathbf{y}}\right)\right) = \frac{1}{\mathrm{n}}\mathrm{n} = 1, \tag{12}$$

e,

$$
\begin{aligned}
\frac{\partial P_{\mathbf{t}_1}\mathbf{y}}{\partial \mathbf{y}} =& \frac{\partial}{\partial \mathbf{y}}\left(\frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\mathbf{X}\mathbf{s}\right) \\
=& \frac{\partial}{\partial \mathbf{y}}\left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\mathbf{X}\mathbf{X}'\mathbf{y}\right) \\
=& \left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\frac{\partial}{\partial \mathbf{y}}\left(\mathbf{X}\mathbf{X}'\mathbf{y}\right) + \frac{\partial}{\partial \mathbf{y}}\left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\left(\mathbf{X}\mathbf{X}'\mathbf{y}\right) \\
=& \left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\mathbf{X}\mathbf{X}' + \left(\frac{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)\left(2\mathbf{X}\mathbf{X}'\mathbf{y}\right)' - \left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)\left(2\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)'}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)^2}\right)\mathbf{X}\mathbf{X}'\mathbf{y} \\
=& \left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\mathbf{X}\mathbf{X}' + \left(\frac{2\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)\mathbf{y}'\mathbf{X}\mathbf{X}' - 2\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)^2}\right)\mathbf{X}\mathbf{X}'\mathbf{y} \\
=& \left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\mathbf{X}\mathbf{X}' + 2\left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)} - \frac{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)^2}\right)\mathbf{X}\mathbf{X}'\mathbf{y} \\
=& \left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\mathbf{X}\mathbf{X}' + 2\left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)} - \frac{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)^2}\right) \\
=& \left(\frac{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}}{\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}}\right)\mathbf{X}\mathbf{X}' + 2\left(1 - \frac{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)}\frac{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)}{\left(\mathbf{y}'\mathbf{X}\mathbf{X}'\mathbf{X}\mathbf{X}'\mathbf{y}\right)}\right) \\
=& \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\mathbf{X}\mathbf{X}' + 2\left(1 - \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\right).
\end{aligned}
$$

Usando $tr(\mathbf{AB}) = tr(\mathbf{BA})$, segue que:

$$
\begin{aligned}
\text{DôF}\left(\mathrm{m}=1\right) =& 1 + tr\left(\frac{\partial \mathrm{P}_{\mathbf{T}}\mathbf{y}}{\partial \mathbf{y}}\right) \\
=& 1 + tr\left(\frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\mathbf{X}\mathbf{X}' + 2\left(1 - \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\right)\right) \\
=& 1 + \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}tr\left(\mathbf{X}'\mathbf{X}\right) + 2 - 2\frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} \\
=& 3 + \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\left(tr\left(\mathbf{S}\right) - 2\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\right).
\end{aligned}
$$

Note que a constante 1 da fórmula é exatamente a derivada do vetor $\bar{\mathbf{y}}$, calculada em (12).

Afirmação: $\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} \leq \lambda_{\max}$.

Seja $\mathbf{A}$ uma matriz ortogonal, $\mathbf{A}'\mathbf{A} = \mathbf{A}\mathbf{A}' = \mathbf{I}$, que diagonaliza $\mathbf{S}$:

$$\mathbf{A}\mathbf{S}\mathbf{A}' = \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} = \mathbf{\Lambda}.$$

Tem-se que $\mathbf{A}$ também diagonaliza $\mathbf{S}^2$. Note que

$$\mathbf{S}\mathbf{v} = \lambda\mathbf{v} \Rightarrow \mathbf{S}\mathbf{S}\mathbf{v} = \lambda\mathbf{S}\mathbf{v} \Rightarrow \mathbf{S}^2\mathbf{v} = \lambda^2\mathbf{v},$$

logo,

$$\begin{aligned} \mathbf{A}\mathbf{S}^2\mathbf{A}' =& \mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{A}\mathbf{S}\mathbf{A}' = \mathbf{\Lambda}\mathbf{\Lambda} \\ &= \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \begin{pmatrix} \lambda_1 & & \\ & \ddots & \\ & & \lambda_p \end{pmatrix} \\ &= \begin{pmatrix} \lambda_1^2 & & \\ & \ddots & \\ & & \lambda_p^2 \end{pmatrix}. \end{aligned}$$

Fazendo $\mathbf{s} = \mathbf{A}'\mathbf{w}$, em que $\mathbf{w} = \mathbf{A}\mathbf{s}$, tem-se:

$$\begin{aligned} \frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} &= \frac{\mathbf{w}'\mathbf{A}\mathbf{S}^2\mathbf{A}'\mathbf{w}}{\mathbf{w}'\mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{w}} = \frac{\mathbf{w}'\mathbf{\Lambda}^2\mathbf{w}}{\mathbf{w}'\mathbf{\Lambda}\mathbf{w}} \\ &= \frac{\sum \lambda_i^2\mathbf{w}_i^2}{\sum \lambda_i\mathbf{w}_i^2} = \frac{\lambda_{max}^2}{\lambda_{max}} \frac{\sum \left(\frac{\lambda_i}{\lambda_{max}}\right)^2 \mathbf{w}_i^2}{\sum \frac{\lambda_i}{\lambda_{max}}\mathbf{w}_i^2}. \end{aligned}$$

Como $\frac{\lambda_{\mathrm{i}}}{\lambda_{\max}} \leq 1 \Rightarrow \left(\frac{\lambda_{\mathrm{i}}}{\lambda_{\max}}\right)^2 \leq \frac{\lambda_{\mathrm{i}}}{\lambda_{\max}}$, segue que

$$\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} \leq \lambda_{\max}.$$

Portanto,

$$tr\left(\mathbf{S}\right) - 2\frac{\mathbf{s}'\mathbf{S}^2\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} \geq tr\left(\mathbf{S}\right) - 2\lambda_{\max}.$$

Além disso, $\frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} \geq \frac{1}{\lambda_{\max}}$.

Novamente, fazendo $\mathbf{s} = \mathbf{A}'\mathbf{w}$,

$$\begin{aligned}\frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} &= \frac{\mathbf{w}'\mathbf{A}\mathbf{A}'\mathbf{w}}{\mathbf{w}'\mathbf{A}\mathbf{S}\mathbf{A}'\mathbf{w}} = \frac{\mathbf{w}'\mathbf{w}}{\mathbf{w}'\mathbf{\Lambda}\mathbf{w}} \\ &= \frac{\sum \mathbf{w}_{\mathrm{i}}^2}{\sum \lambda_{\mathrm{i}}\mathbf{w}_{\mathrm{i}}^2} = \frac{1}{\lambda_{\max}}\frac{\sum \mathbf{w}_{\mathrm{i}}^2}{\sum \frac{\lambda_{\mathrm{i}}}{\lambda_{\max}}\mathbf{w}_{\mathrm{i}}^2}.\end{aligned}$$

Como $\frac{\lambda_{\mathrm{i}}}{\lambda_{\max}} \leq 1$, segue que:

$$\frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}} \geq \frac{1}{\lambda_{\max}}.$$

Assim,

$$\begin{aligned}\mathrm{D\hat{o}F}\left(\mathrm{m}=1\right) \geq & 3 + \frac{\mathbf{s}'\mathbf{s}}{\mathbf{s}'\mathbf{S}\mathbf{s}}\left(tr\left(\mathbf{S}\right) - 2\lambda_{\max}\right) \\ \geq & 3 + \frac{1}{\lambda_{\max}}\left(tr\left(\mathbf{S}\right) - 2\lambda_{\max}\right) \\ = & 1 + \frac{tr\left(\mathbf{S}\right)}{\lambda_{\max}}.\end{aligned}$$

□

Observe que $tr(\mathbf{S})$ é igual a soma dos autovalores e, portanto, $\frac{tr(\mathbf{S})}{\lambda_{\max}} > 1$. Logo, $\mathrm{D\hat{o}F}\left(\mathrm{m}=1\right) \geq 2$ e, $\mathrm{DoF}\left(\mathrm{m}=1\right) \geq 2$, mostrando que a conjectura é

verdadeira para $\mathrm{m} = 1$.

Pode-se utilizar o conceito de Grau de Liberdade para determinar um número adequado de componentes a ser utilizado na regressão, determinando o DoF para cada valor de m. Se para um acréscimo em m tem-se um acréscimo considerável no valor do DoF, esse valor de m é justificado. Porém, se para um acréscimo em m, o acrésimo no valor do DoF é pequeno, o princípio da parcimônia recomenda que esse acréscimo na complexidade do modelo é desnecessário. Traçando-se o gráfico DoF em função do número de componentes, o número adequado de componentes ocorre quando a curva tende a se estabilizar em um valor constante, como ilustrado na Figura 37.

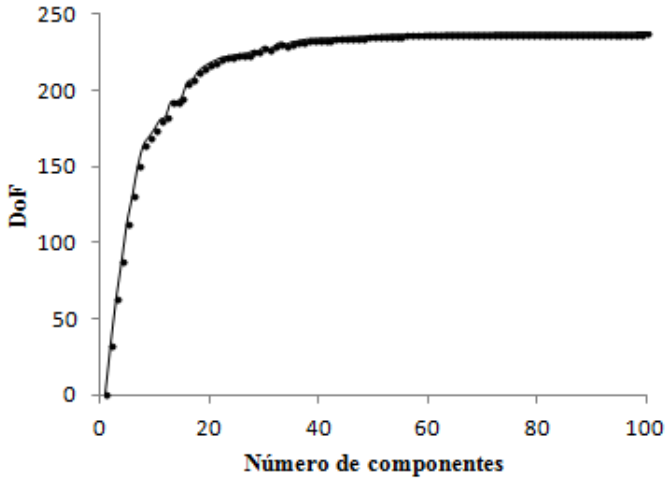


Figura 37 Representação de uma curva gerada por valores dos graus liberdade, DoF, em função do número de componentes m

4.4.2 Validação Cruzada

Uma forma mais prática de se determinar o número de componentes, é por meio da validação cruzada (PÉREZ-CABAL *et al*., 2012). Este método consiste em dividir o conjunto de dados originais em N subconjuntos (*folds*), e realizar N análises, de forma que em cada uma delas um dos subconjuntos é retirado a fim de ser utilizado como recurso para validar a análise realizada. Assim, os valores preditos pelas equações estimadas em cada análise podem ser diretamente comparados com os valores observados que foram retirados. No contexto de PLSR, o número ótimo de componentes será aquele que fornecer o menor erro quadrático entre o conjunto de valores observados e preditos, como ilustra a Figura 38.

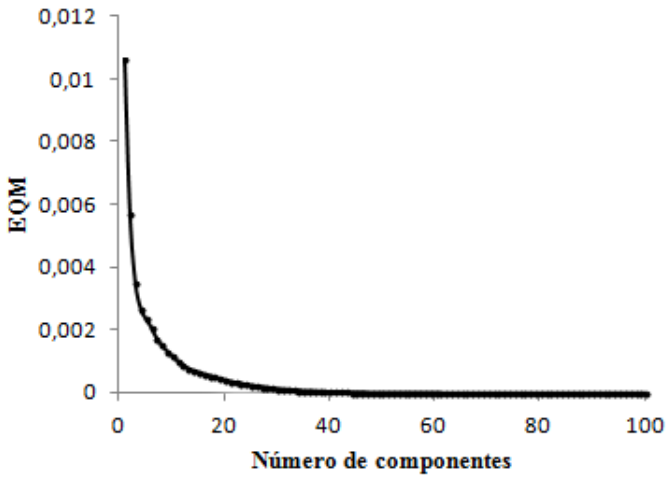


Figura 38 Representação de uma curva gerada por valores de erro quadrático médio em função do número de componentes m

4.5 Exemplo

Matriz de covariâncias simulada conforme a seção 3.5:

```
 6.2810455  8.194268 0.3211009  5.693020 20.811639
 8.1942679 10.856983 1.0961611  7.626992 27.924481
 0.3211009  1.096161 4.5260950  7.049471  6.085769
 5.6930197  7.626992 7.0494713 25.766833 26.599556
20.8116386 27.924481 6.0857694 26.599556 77.622607
```

Autovalores e autovetores da matriz de covariâncias:

```
$values
1.043224e+02 1.758224e+01 2.610955e+00 5.375069e-01
4.210175e-04

$vectors
-0.227191  0.178466 -0.292529 -0.457806  0.788272
-0.304403  0.223537 -0.140214 -0.695880 -0.594524
-0.080458 -0.327128  0.851982 -0.371038  0.151556
-0.342879 -0.865085 -0.362872 -0.015131 -0.046416
-0.855384  0.250587  0.192913  0.410200  0.006555
```

Matriz de correlação:

```
1.00000000 0.9922922 0.06022318 0.4475032 0.9425322
0.99229225 1.0000000 0.15637163 0.4560035 0.9619144
0.06022318 0.1563716 1.00000000 0.6527760 0.3246830
0.44750325 0.4560035 0.65277602 1.0000000 0.5947710
0.94253219 0.9619144 0.32468300 0.5947710 1.0000000
```

Matriz de covariáveis $\mathbf{X}_{6\times4}$:

```
 0.9391754 2.0814841 4.776478 8.597991
 2.1972739 4.1631504 5.466290 6.725786
 3.5575236 5.4117197 3.282917 6.600031
-0.8903883 0.1328435 5.472369 4.877636
-0.5096611 0.2326990 3.388711 0.957349
 1.9463446 3.6372080 5.319583 7.467699
```

Vetor da variável resposta $\mathrm{Y}_{6\times1}$:

```
 7.820750
11.335945
16.909518
 3.731732
-1.821228
 9.446128
```

Vetor de parâmetros estimado via OLS, $\hat{\boldsymbol{\beta}}_{\mathrm{OLS}}$:

```
-42.710500
 33.456902
 -9.939255
  3.077167
```

Vetor de respostas estimado via OLS, $\hat{\mathrm{Y}}_{\mathrm{OLS}}$:

```
 8.510179
11.804963
16.795406
 3.091492
-1.182066
 8.667023
```

Utilizando 1 componente: $\mathrm{m} = 1$

Vetor dos componentes:

```
 9.723363
 9.813350
 9.514253
 6.064935
 2.297160
10.018880
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função dos componentes:

```
1.088373
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
0.2557785
0.4469298
0.5175329
0.8071838
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função dos componentes:

```
10.582641
10.680581
10.355052
 6.600909
 2.500166
10.904274
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
10.582641
10.680581
10.355052
 6.600909
 2.500166
10.904274
```

Utilizando 2 componentes: $\mathrm{m} = 2$

Matriz dos componentes:

```
 9.723363 -1.1645658
 9.813350  0.5490979
 9.514253  3.2962070
 6.064935 -3.6298704
 2.297160 -2.0393287
10.018880  0.1271250
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função dos componentes:

```
1.088373
1.480342
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
 1.0619346
 1.3092326
-0.3592859
 0.6364624
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função dos componentes:

```
 8.8586862
11.4934339
15.2345645
 1.2274611
-0.5187373
11.0924624
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
 7.4786659
10.1006417
13.8842227
 0.3666752
-0.8447693
 9.6704998
```

Utilizando 3 componentes: $\mathrm{m} = 3$

Matriz dos componentes:

```
 9.723363 -1.1645658 -1.52574743
 9.813350  0.5490979  0.79694695
 9.514253  3.2962070  0.28010446
 6.064935 -3.6298704  0.13097361
 2.297160 -2.0393287  1.30889216
10.018880  0.1271250  0.05475768
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função dos componentes:

```
1.08837257
1.48034155
0.09640233
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
 1.0341975
 1.3699283
-0.3146987
 0.5830576
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função dos componentes:

```
 8.711601
11.570261
15.261567
 1.240087
-0.392557
11.097741
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
 7.3327488
10.1769183
13.9079179
 0.3829437
-0.7165427
 9.6756510
```

Utilizando 4 componentes: $\mathrm{m} = 4$

Matriz dos componentes:

```
 9.723363 -1.1645658 -1.52574743 -0.004186004
 9.813350  0.5490979  0.79694695  0.004877645
 9.514253  3.2962070  0.28010446  0.031876755
 6.064935 -3.6298704  0.13097361  0.038476500
 2.297160 -2.0393287  1.30889216 -0.016407845
10.018880  0.1271250  0.05475768 -0.050515983
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função dos componentes:

```
 1.08837257
 1.48034155
 0.09640233
48.11779953
```

Vetor de parâmetros estimado via PLS, $\hat{\boldsymbol{\beta}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
-35.14784
 16.25217
-22.32000
 17.91708
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função dos componentes:

```
 8.510179
11.804963
16.795406
 3.091492
-1.182066
 8.667023
```

Vetor de respostas estimado via PLS, $\hat{\mathrm{Y}}_{\mathrm{PLS}}$, em função das covariáveis originais:

```
 48.258491
-11.070420
  7.891424
 -1.296101
-36.787790
  5.768917
```

Novos valores de $\mathbf{X}_{\mathrm{p}}$ e Y_{p} gerados a fim de fazer predição:

```
Xp
3.289261 5.040320 3.607594 8.119808

Yp
12.87691
```

Valores preditos para Y_{p} ($\hat{\mathrm{Y}}_{\mathrm{p}}$) via OLS e PLS:

```
OLS  17.23269
PLS  10.81509 - 1 componente
PLS  15.09155 - 2 componentes
PLS  15.01617 - 3 componentes
PLS  37.04379 - 4 componentes
```

5 CONCLUSÕES

A abordagem geométrica apresentou-se como uma maneira eficiente e intuitiva para explicitar os passos teóricos da teoria e do algoritmo do método dos Quadrados Mínimos Parciais. Além disso, permite uma abordagem unificada das teorias de regressão em Quadrados Mínimos Ordinários, Componentes Principais e Quadrados Mínimos Parciais.

REFERÊNCIAS

ADÃO, A. S. **Introdução à teoria geométrica dos delineamentos experimentais.** 2011. 78 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG, 2011.

AZEVEDO, C. F. **Métodos de redução de dimensionalidade aplicados na seleção genômica para características de carcaça em suínos.** 2012. 50 f. Dissertação (Mestrado em Estatística Aplicada e Biometria) - Universidade Federal de Viçosa, Viçosa, MG, 2012.

EFRON, B. The Estimation of Prediction Error: covariance penalties and cross-validation. **Journal of the American Statistical Association**, [s.l.], v. 99, n. 467, p. 619-632, set. 2004.

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

FRANK, I.; FRIEDMAN, J. A Statistical View of Some Chemometrics Regression Tools. **Technometrics**, [s.l.], v. 35, n. 2, p. 109-135, maio 1993.

GARTHWAITE, P. H. An Interpretation of Partial Least Squares. **Journal of the American Statistical Association**, [s.l.], v. 89, n. 425, p. 122-127, mar. 1994.

GUIMARÃES, P. H. S. **Uma abordagem geométrica da Teoria de Inversas Generalizadas.** 2010. 67 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG, 2010.

HELLAND, I. S. Partial Least Squares Regression and Statistical Models. **Scandinavian Journal of Statistics**, [s.l.], v. 17, n. 2, p. 97-114, 1990.

HOSKULDSSON, A. PLS Regression Methods. **Journal of Chemometrics**, [s.l.], v. 2, p. 211-228, 1988.

HOTELLING, H. The relations of the newer multivariate statistical methods to factor analysis. **British Journal of Mathematical and Statistical Psychology**, [s.l.], v. 10, n. 2, p. 69-79, nov. 1957.

JONG, S. SIMPLS: an alternative approach to partial least squares regression. **Chemometrics and Intelligent Laboratory Systems**, [s.l.], v. 18, p. 251-263, 1993.

KALMAN, D. A Singularly Valuable Decomposition: the SVD of a Matrix. **The College Mathematics Journal**, [s.l.], v. 27, n. 1, p. 2-23, jan. 1996.

KENDALL, M. G. **A Course in Multivariate Analysis.** London: Griffin, 1957.

KRÄMER, N.; SUGIYAMA, M. The Degrees of Freedom of Partial Least Squares Regression. **Journal of the American Statistical Association**, [s.l.], v. 106, n. 494, p. 697-705, fev. 2011.

MARTENS, H.; NAES, T. **Multivariate Calibration.** New York: Wiley, 1989.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome wide densemarker maps. **Genetics**, [s.l.], v. 157, p. 1819-1829, abr. 2001.

OTTO, M. **Chemometrics.** Weinheim: Wiley, 1999. 328 p.

PEREIRA, L. S. **Abordagem geométrica à teoria dos modelos de Gauss-Markov.** 2013. 130 f. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG, 2013.

PÉREZ-CABAL, M. A. *et al*. Accuracy of genome-enabled prediction in a dairy cattle population using different cross-validation layouts. **Frontiers in Genetics - Livestock Genomics**, [s.l.], v. 3, p. 1-7, fev. 2012.

PHATAK, A.; JONG, S. The geometry of partial least squares. **Journal of the Chemometrics**, [s.l.], v. 11, n. 4, p. 311-338, jul. 1997.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2013. Disponível em: <http://www.R-project.org.>. Acesso em: 14 nov. 2013.

RAO, C. R. **Linear Statistical Inference and Its Applications.** New York: Wiley, 2002. 625 p.

ROGGO, Y. *et al*. A review of near infrared spectroscopy and chemometrics in harmaceutical technologies. **Journal of Pharmaceutical and Biomedical Analysis**, [s.l.], v. 44, n. 3, p. 683-700, jul. 2007.

WOLD, H. Path models with Latent Variables: the NIPALS Approach. In: H. B. *et al*. (Ed.). **Quantitative Sociology**: international Perspectives on Mathematical and Statistical Modeling. [New York]: Academic Press, 1975. p. 307-357.

APÊNDICES

APÊNDICE A - Completamento de Quadrados

Deseja-se o completamento de quadrados em y para o seguinte termo:

$$\mathbf{x}'\mathbf{C}\mathbf{x} - 2\mathbf{x}'\mathbf{C}\mathbf{H}\mathrm{y} + \mathrm{y}'(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B})\mathrm{y}.$$

Ou seja, deseja-se determinar w tal que:

$$(\mathrm{y} - \mathrm{w})' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) (\mathrm{y} - \mathrm{w}) + g(\mathbf{x}) = \mathbf{x}'\mathbf{C}\mathbf{x} - \mathbf{2}\mathbf{x}'\mathbf{C}\mathbf{H}\mathrm{y} + \mathrm{y}'(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B})\mathrm{y},$$

em que $g(\mathbf{x})$ é uma função que não depende de y.

Tem-se que:

$$\begin{aligned}
&(\mathrm{y} - \mathrm{w})' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) (\mathrm{y} - \mathrm{w}) \\
&= \mathrm{y}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{y} - 2\mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{y} + \mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{w}.
\end{aligned}$$

Observe que, $\mathbf{x}'\mathbf{C}\mathbf{H}\mathrm{y} = \mathrm{w}'(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B})\mathrm{y}$. Logo,

$$\mathrm{w} = \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right)^{-1} \mathbf{H}'\mathbf{C}\mathbf{x}.$$

Portanto,

$$\begin{aligned}
&\mathbf{x}'\mathbf{C}\mathbf{x} - 2\mathbf{x}'\mathbf{C}\mathbf{H}\mathrm{y} + \mathrm{y}'(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B})\mathrm{y} \\
=&\mathbf{x}'\mathbf{C}\mathbf{x} + \mathrm{y}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{y} - 2\mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{y} \\
=&\mathbf{x}'\mathbf{C}\mathbf{x} + \mathrm{y}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{y} - 2\mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{y}
\end{aligned}$$

$$
\begin{aligned}
&+ \mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{w} - \mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{w} \\
&= (\mathrm{y} - \mathrm{w})' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) (\mathrm{y} - \mathrm{w}) + \mathbf{x}'\mathbf{C}\mathbf{x} - \mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{w}.
\end{aligned}
$$

Assim, a distribuição marginal de $\mathbf{X}$ é dada por:

$$
\begin{aligned}
\mathrm{f}_{\mathbf{X}} (\mathbf{x}) \propto & exp \left(\mathbf{x}'\mathbf{C}\mathbf{x} - \mathrm{w}' \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \mathrm{w}\right) \\
\propto & exp \Bigg(\mathbf{x}'\mathbf{C}\mathbf{x} - \mathbf{x}'\mathbf{C}\mathbf{H} \big(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\big)^{-\mathbf{1}} \left(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\right) \times \\
& \big(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\big)^{-1} \mathbf{H}'\mathbf{C}\mathbf{x} \Bigg) \\
\propto & exp \left(\mathbf{x}'\mathbf{C}\mathbf{x} - \mathbf{x}'\mathbf{C}\mathbf{H} \big(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\big)^{-\mathbf{1}} \mathbf{H}'\mathbf{C}\mathbf{x} \right) \\
\propto & exp \left(\mathbf{x}' \left(\mathbf{C} - \mathbf{C}\mathbf{H} \big(\mathbf{H}'\mathbf{C}\mathbf{H} + \mathbf{B}\big)^{-1} \mathbf{H}'\mathbf{C} \right) \mathbf{x} \right).
\end{aligned}
$$

APÊNDICE B - Decomposição em Valores Singulares

Teorema 3. *(Decomposição Espectral ou de Jordan). Seja* $\mathbf{L}$ *uma transformação linear,* $\mathbf{L} : \mathbb{R}^{\mathrm{n}} \to \mathbb{R}^{\mathrm{n}}$, *simétrica. Então, existe uma base ortonormal de autovetores* $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{\mathrm{n}}\}$, *tal que* $\mathbf{L}(\mathbf{v}_{\mathrm{i}}) = \lambda_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}$, *com* λ_{i} *autovalores de* $\mathbf{L}$. *De forma equivalente, para toda matriz simétrica* $\mathbf{L}_{\mathrm{n}\times\mathrm{n}}$ *existe uma decomposição da forma* $\mathbf{L} = \mathbf{V}\mathbf{D}\mathbf{V}'$, *em que* $\mathbf{D}$ *é a matriz diagonal formada pelos autovalores* λ_{i} *de* $\mathbf{L}$, *e* $\mathbf{V}_{\mathrm{n}\times\mathrm{n}} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{\mathrm{n}}]$ *é a matriz ortogonal formada pelos autovetores de* $\mathbf{L}$ *em suas colunas.*

Esse teorema pode ser generalizado da forma:

Teorema 4. *(Decomposição em valores singulares). Seja* $\mathbf{L}$ *uma transformação linear,* $\mathbf{L} : \mathbb{R}^{\mathrm{m}} \to \mathbb{R}^{\mathrm{n}}$, *de posto* r. *Então existe uma base ortonormal de* $\mathbb{R}^{\mathrm{n}}$, $\{\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{\mathrm{n}}\}$, *e uma base ortonormal de* $\mathbb{R}^{\mathrm{m}}$, $\{\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{\mathrm{m}}\}$, *tal que* $\mathbf{L}(\mathbf{v}_{\mathrm{i}}) = \sigma_{\mathrm{i}}\mathbf{u}_{\mathrm{i}}$ *e* $\mathbf{L}'(\mathbf{u}_{\mathrm{i}}) = \sigma_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}$, $\mathrm{i} = 1, \ldots, \mathrm{r}$ *e* $\sigma_{\mathrm{i}} = 0$, *para* $\mathrm{i} \geq \mathrm{r} + 1$. *De forma equivalente, toda matriz* $\mathbf{L}_{\mathrm{n}\times\mathrm{m}}$ *de posto* r *pode ser decomposta em* $\mathbf{L} = \mathbf{V}\mathbf{D}\mathbf{U}'$, *em que* $\mathbf{D}_{\mathrm{r}\times\mathrm{r}}$ *é a matriz diagonal formada pelos valores singulares* σ_{i} *de* $\mathbf{L}$, *e* $\mathbf{V}_{\mathrm{n}\times\mathrm{r}} = [\mathbf{v}_1, \mathbf{v}_2, \ldots, \mathbf{v}_{\mathrm{n}}]$ *e* $\mathbf{U}_{\mathrm{m}\times\mathrm{r}} = [\mathbf{u}_1, \mathbf{u}_2, \ldots, \mathbf{u}_{\mathrm{m}}]$ *são matrizes ortonormais por coluna.*

O teorema da decomposição em valores singulares admite a seguinte interpretação geométrica (KALMAN, 1996):

Seja $\mathbf{L}_{\mathrm{n}\times\mathrm{m}}$, de posto r, e seja $\mathbf{v}$ um vetor na esfera unitária de $\mathbb{R}^{\mathrm{n}}$, isto é, $\mathbf{v} = \sum_{\mathrm{i}=1}^{\mathrm{n}} \mathbf{x}_{\mathrm{i}}\mathbf{v}_{\mathrm{i}}$ e $\|\mathbf{v}\|^2 = \mathbf{x}_1^2 + \ldots + \mathbf{x}_{\mathrm{n}}^2 = 1$. Então, existem bases tais que

$\mathbf{L}(\mathbf{v}_i) = \sigma_i \mathbf{u}_i$. Logo, para $i = 1, \ldots, r$,

$$\mathbf{L}\mathbf{v} = \mathbf{L}\left(\sum_{i=1}^{r} \mathbf{x}_i \mathbf{v}_i\right) = \sum_{i=1}^{r} \mathbf{x}_i \mathbf{L}(\mathbf{v}_i) = \sum_{i=1}^{r} \mathbf{x}_i \sigma_i \mathbf{u}_i.$$

Fazendo $\mathbf{y}_i = \mathbf{x}_i \sigma_i$, segue que:

$$\frac{\mathbf{y}_1^2}{\sigma_1^2} + \ldots + \frac{\mathbf{y}_r^2}{\sigma_r^2} = \frac{\mathbf{x}_1^2 \sigma_1^2}{\sigma_1^2} + \ldots + \frac{\mathbf{x}_r^2 \sigma_r^2}{\sigma_r^2} = \mathbf{x}_1^2 + \ldots + \mathbf{x}_r^2 \leq 1.$$

Portanto, a imagem pela transformação $\mathbf{L}$ da esfera unitária no $\mathbb{R}^n$ é um elipsóide, como pode ser observado pela Figura 39. Este elipsóide será sólido se o posto da transformação for menor que n.

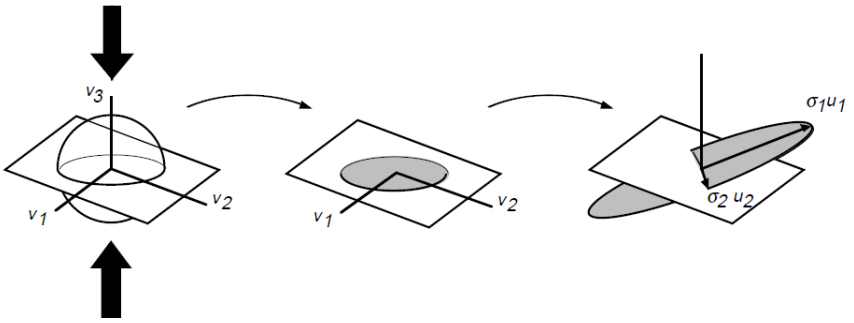


Figura 39 Como a transformação $\mathbf{L}$ deforma uma esfera em um elipsóide

APÊNDICE C - Método das Potências

O Método das Potências, ou *Power Method*, é um método para obtenção dos autovalores e autovetores de uma matriz. Sejam $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_n$ os autovetores de uma matriz $\mathbf{A}_{n\times n}$, associados aos autovalores $\lambda_1, \lambda_2, \cdots, \lambda_n$, respectivamente, em que $\lambda_1 \geq \lambda_2 \geq \cdots \geq \lambda_n$. Como os autovetores formam uma base ortonormal, pode-se obter um vetor arbitrário $\mathbf{v}^{(0)}$ como uma combinação linear dos vetores desta base:

$$\mathbf{v}^{(0)} = c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_n\mathbf{x}_n, \tag{13}$$

em que $c_1, c_2, \ldots, c_n$ são constantes reais.

Pré-multiplicando por $\mathbf{A}$ ambos os lados da igualdade (13), e utilizando que $\mathbf{A}\mathbf{x}_i = \lambda_i\mathbf{x}_i$, tem-se:

$$\begin{aligned}
\mathbf{A}\mathbf{v}^{(0)} =& \mathbf{A}\left(c_1\mathbf{x}_1 + c_2\mathbf{x}_2 + \ldots + c_n\mathbf{x}_n\right)\\
=& c_1\mathbf{A}\mathbf{x}_1 + c_2\mathbf{A}\mathbf{x}_2 + \ldots + c_n\mathbf{A}\mathbf{x}_n\\
=& c_1\lambda_1\mathbf{x}_1 + c_2\lambda_2\mathbf{x}_2 + \ldots + c_n\lambda_n\mathbf{x}_n.
\end{aligned}$$

Repetindo o processo k vezes, obtém-se:

$$\begin{aligned}
\mathbf{A}^2\mathbf{v}^{(0)} =& c_1\lambda_1^2\mathbf{x}_1 + c_2\lambda_2^2\mathbf{x}_2 + \ldots + c_n\lambda_n^2\mathbf{x}_n\\
&\vdots\\
\mathbf{A}^k\mathbf{v}^{(0)} =& c_1\lambda_1^k\mathbf{x}_1 + c_2\lambda_2^k\mathbf{x}_2 + \ldots + c_n\lambda_n^k\mathbf{x}_n\\
=& \lambda_1^k\left(c_1\mathbf{x}_1 + c_2\frac{\lambda_2^k}{\lambda_1^k}\mathbf{x}_2 + \ldots + c_n\frac{\lambda_n^k}{\lambda_1^k}\mathbf{x}_n\right).
\end{aligned}$$

Como λ_1 é o autovalor dominante, $\left(\frac{\lambda_i}{\lambda_1}\right)^k \to 0$ quando $k \to \infty$, logo,

$$\mathbf{A}^k\mathbf{v}^{(0)} \approx c_1\lambda_1^k\mathbf{x}_1,$$

para k suficientemente grande (Figura 40).

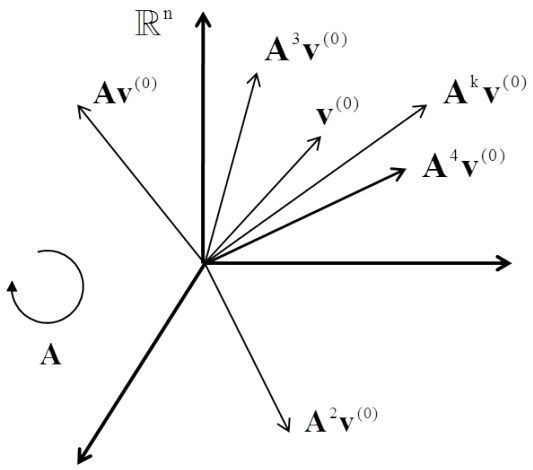


Figura 40 Representação geométrica do método das potências

O método das potências normaliza os produtos $\mathbf{A}\mathbf{v}^{(k-1)}$ para evitar *overflow* ou *underflow* (FERREIRA, 2008).

APÊNDICE D - Rotina Computacional para Algoritmo PLS

```
Rotina utilizada no Exemplo mostrando o passo
a passo do Algoritmo Amostral PLS:

library(MBESS)
library(MASS)
library(Matrix)

# gerando uma matriz de covariâncias

RandomSigma <- function(p = 5, df = 10)
{
  T <- diag(sqrt(rgamma(c(rep(1, p)), df/2, 1/2)))
  T[lower.tri(T)] <- 0.98*rgamma((p * (p - 1)/2),
                                      df/2, 1/2)
  Sigma <-   T %*% t(T)
  return(Sigma)
}
p <- 5
df <- 2
Sigma <- RandomSigma(p, df)
Sigma
eigen(Sigma)
Rho <- diag(diag(Sigma)^-0.5)  %*% Sigma %*%
                          diag(diag(Sigma)^-0.5)
```

```
Rho

nxy=nrow(Sigma)
nxy

# vetor de médias populacional de X e Y
mpXY=as.matrix(c(1,2,3,4,5))
mpXY

# gerando X e Y de uma normal multivariada
XY=mvrnorm(6,mpXY,Sigma)
XY
n = dim(XY)[1]
n

X = as.matrix(XY[1:n,1:4])
X
m = dim(X)[2]
m
X1=sum(X[1:n,1])/n          # média da covariável X1
X1
X2=sum(X[1:n,2])/n          # média da covariável X2
X2
X3=sum(X[1:n,3])/n          # média da covariável X3
X3
X4=sum(X[1:n,4])/n          # média da covariável X4
```

```
X4

maX=(c(X1,X2,X3,X4)) # vetor de médias amostrais de X
maX

Y = as.matrix(XY[1:n,5])
Y
k = dim(Y)[2]
k

maY=sum(Y[1:n,1])/n  # vetor de médias amostrais de Y
maY

# simulação PLS

nc = 4               # variando o nº de componentes

# os vetores armazenados
T = U = matrix(0,n,nc)
C = matrix(0,k,nc)
P = W = matrix(0,m,nc)
Cont = matrix(0,1,nc)

# O algoritmo:

contnumbcomp = 1          # contador do nº componentes
```

```
u = as.matrix(Y[,1])      # passo 1 - o chute inicial
while (contnumbcomp<=nc){
flag = 0                  # indicador de convergência
cont = 0
while (flag<1){
cont = cont + 1
w = t(X)%*%u/((t(u)%*%u)[1,1])     # passo 2
w = w/sqrt((t(w)%*%w)[1,1])        # passo 3
t = X%*%w                          # passo 4
c = t(Y)%*%t/((t(t)%*%t)[1,1])     # passo 5
c = c/sqrt((t(c)%*%c)[1,1])        # passo 6
if (max(sqrt((u - Y%*%c)^2)) > 0.0001) u = Y%*%c
else flag = 1                      # passos 7 e 8
}
p = t(X)%*%t/((t(t)%*%t)[1,1])     # passo 9
q = t(Y)%*%u/((t(u)%*%u)[1,1])     # passo 10
Cont[1,contnumbcomp] = cont
W[1:m,contnumbcomp] = w
T[1:n,contnumbcomp] = t
C[1:k,contnumbcomp] = c
U[1:n,contnumbcomp] = u
P[1:m,contnumbcomp] = w
b = (t(u)%*%t)[1]/((t(t)%*%t)[1])  # passo 11
X = X - t%*%t(p)                   # passo 12
Y = Y - b*(t%*%t(c))               # passo 12
contnumbcomp = contnumbcomp + 1
```

```
}

# chamando as matrizes de dados X e Y novamente:
X = as.matrix(XY[1:n,1:4])
X
Y = as.matrix(XY[1:n,5])
Y

# estimando via OLS
beta_ols=solve(t(X)%*%X)%*%t(X)%*%Y
beta_ols
y_ols=X%*%beta_ols
y_ols

# estimando via PLS
beta_pls_T=solve(t(T)%*%T)%*%t(T)%*%Y
beta_pls_T
beta_pls_X=P%*%beta_pls_T
beta_pls_X

y_pls=T%*%beta_pls_T
y_pls

y_pls=X%*%P%*%beta_pls_T
y_pls
```

```
# gerando novas matrizes de dados X e Y para predição
XYp=mvrnorm(1,mpXY,Sigma)
XYp
Xp = XYp[1:4]
Xpcentrado=Xp-maX

Yp = XYp[5]
Yp

# Eq. de predição: y = ym + (x-xm)beta
y_pls=maY+Xpcentrado%*%P%*%beta_pls_T
y_pls
y_ols=maY+Xpcentrado%*%beta_ols
y_ols

# verificando ortonormalidade de T

t1 = as.matrix(T[1:n,1])
t2 = as.matrix(T[1:n,2])
t3 = as.matrix(T[1:n,3])

t(t1)%*%t2
t(t1)%*%t3
t(t2)%*%t3
```

CAPÍTULO 2

Quadrados Mínimos Parciais aplicado à seleção genômica para qualidade de carne em suínos

RESUMO

A principal contribuição da genética molecular no melhoramento animal é a utilização direta das informações de DNA no processo de identificação de animais geneticamente superiores. No âmbito da seleção genômica, uma vez que o número de marcadores é geralmente muito maior que o número de animais genotipados e tais marcadores são altamente correlacionados, métodos estatísticos baseados na redução de dimensionalidade apresentam grande aplicabilidade. Dentre esses métodos, destacam-se a Regressão em Componentes Principais (PCR) e os Quadrados Mínimos Parciais (PLS). Para a aplicação de tais métodos, a determinação do número ótimo de componentes a ser utilizado ainda se caracteriza como uma questão relevante, no entanto, metodologias como a teoria de graus de liberdade (DoF) e a validação cruzada (CV) têm sido propostas. Diante do exposto, objetivou-se aplicar os métodos PCR e PLS, além da regressão tradicional, sem redução de dimensionalidade, em uma análise de seleção genômica em suínos, considerando um painel de marcadores SNPs de baixa densidade e dois fenótipos relacionados à qualidade da carne (pH medido aos 45 min e às 24 horas após o abate). Objetivou-se, ainda, testar as teorias DoF e CV na determinação do número ótimo de componentes na análise PLS, e sua influência na performance preditiva do método. Os resultados mostraram que a metodologia PLS é eficiente para a seleção genômica por possibilitar predições satisfatórias para o pH da carne suína utilizando apenas informações genotípicas, além de identificar uma região relevante no cromossomo 4. Os métodos DoF e CV foram compatíveis para a determinação do número ótimo de componentes na análise PLS, sendo este de eficiência superior ao PCR e à regressão múltipla tradicional.

Palavras-chave: Graus de Liberdade. Marcadores SNPs. pH. Validação Cruzada.

ABSTRACT

The main contribution of molecular genetics in animal breeding is the direct use of DNA information in the identification process of genetically superior animals. Within the genomic selection, since the number of markers is generally much larger than the number of genotyped animals and such markers are highly correlated, statistical methods based on dimensionality reduction have wide applicability. Among these methods the Principal Components Regression (PCR) and Partial Least Squares (PLS) are highlighted. To further the application of such methods, determining the optimal number of components to be used is characterized as a relevant issue, however, methodologies such as the theory of degrees of freedom (DoF) and cross-validation (CV) have been proposed. Given these facts, we sought to apply PCR and PLS methods, beyond traditional regression without dimensionality reduction , in an analysis of genomic selection in pigs considering a panel of low-density SNP markers and two phenotypes related to meat quality (pH measured at 45 minutes and at 24 hours after slaughter). We still aimed to test the DoF and CV theories to determine the optimal number of components in the PLS analysis , and its influence on the predictive performance of the method. The results showed that the PLS method is efficient for genomic selection for enabling satisfactory predictions for swine meat pH using only genotypic information, and for identifying an important region on chromosome 4. The DoF and CV methods were compatible for determining the optimal number of components in the PLS analysis, and this method demonstrated itself as being more efficient than PCR and traditional multiple regression.

Keywords: Degrees of Freedom. SNPs Markers. pH. Cross Validation.

1 INTRODUÇÃO

A partir do início do século XXI, os avanços biotecnológicos na área de automação do processo de genotipagem permitiram o desenvolvimento de novas classes de marcadores moleculares (JENKINS; GIBSON, 2002), dentre os quais se destacam os SNPs (*Single Nucleotide Polymorphisms*). Diante da abundância destes marcadores, Meuwissen; Hayes; Goddard (2001) idealizaram a seleção genômica ampla (*Genome Wide Selection* - GWS). A principal contribuição da genética molecular no melhoramento animal é a utilização direta das informações de DNA no processo de identificação de animais geneticamente superiores.

Devido à alta densidade dos marcadores SNPs no genoma, é possível assumir que alguns deles estejam em desequilíbrio de ligação com locos de características quantitativas (*Quantitative Trait Loci* - QTL), possibilitando sua utilização direta na estimação do valor genético genômico de indivíduos sujeitos a seleção, inclusive de indivíduos que ainda não foram fenotipados. No âmbito da seleção genômica, uma vez que o número de marcadores é geralmente muito maior que o número de animais genotipados (alta dimensionalidade) e tais marcadores são altamente correlacionados (multicolinearidade devida ao desequilíbrio de ligação), métodos estatísticos baseados na redução de dimensionalidade apresentam grande aplicabilidade.

Dentre estes métodos destacam-se a regressão em Componentes Principais (*Principal Components Regression* - PCR) e os Quadrados Mínimos Parciais (*Partial Least Squares* - PLS), os quais são recomendados para situações em que se tem mais covariáveis do que observações (HOSKULDSSON, 1988) e também alta correlação entre as covariáveis, uma vez que garantem que a correlação entre qualquer par de variáveis latentes, ou componentes (combinações lineares das

covariáveis), seja igual a zero. A principal diferença entre estes dois métodos é que o PCR leva em consideração apenas as variáveis explicativas na construção dos componentes, enquanto que o PLS também leva em consideração as variáveis respostas (GARTHWAITE, 1994).

Embora o PLS apresente uma vantagem teórica sobre o PCR por considerar a variável resposta no processo de formação dos componentes, a determinação do número ótimo de tais componentes ainda se caracteriza como um problema relevante para a aplicação do referido método, uma vez que diferentes números de componentes podem levar a diferentes resultados. Metodologias como a teoria de graus de liberdade (*Degrees of Freedom* - DoF), a qual é baseada em uma sofisticada teoria estatística (KRÄMER; SUGIYAMA, 2011); e a validação cruzada (*Cross Validation* - CV), a qual é de caráter empírico, têm sido propostas. Porém, até o momento, comparações entre essas metodologias sob o enfoque de predição genômica ainda não foram reportadas na literatura e merecem ser averiguadas.

Atualmente, fenótipos relacionados com a qualidade da carne são relevantes para a seleção genômica em suínos e, dentre estes, destacam-se as medidas de pH, como o pH inicial (45 minutos após o abate) e o pH último (24 horas após abate). A importância de tais fenótipos diz respeito ao fato dos mesmos estarem diretamente relacionados com a retenção de água, maciez, suculência e aparência da carne, que por sua vez relacionam-se com a aceitabilidade e a palatabilidade (BENEVENUTO JUNIOR, 2001). Além disso, como são fenótipos avaliados após o abate, a identificação de indivíduos superiores destinados a seleção pode ser realizada precocemente por meio de painéis de marcadores SNPs identificados a partir de análises estatísticas específicas como PCR e PLS.

Diante do exposto, objetivou-se aplicar os métodos PCR e PLS, além da regressão tradicional sem redução de dimensionalidade, em uma análise de sele-

ção genômica em suínos, considerando um painel de marcadores SNPs de baixa densidade e dois fenótipos relacionados com a qualidade da carne (pH medido aos 45 min e às 24 horas após o abate). Além disso, objetivou-se, ainda, testar dois diferentes métodos de determinação do número ótimo de componentes na análise PLS, a teoria de graus de liberdade e a validação cruzada, bem como sua influência na performance preditiva do método.

2 MATERIAL E MÉTODOS

2.1 Descrição dos dados utilizados

Os dados utilizados no presente estudo 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, a população F_2 foi composta de 345 suínos provenientes do cruzamento de dois varrões 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 (CENARGEN), 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) de 345 indivíduos foram mensurados após o abate (realizado aproximadamente aos 105 dias de idade dos animais) e, dentre 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 DIGIMED DM-20, previa-

mente calibrado, no músculo *Longissimus dorsi*, retirado da região imediatamente posterior à última costela do animal.

2.2 Quadrados Mínimos Parciais - PLS

A metodologia do PLS consiste em formar componentes $\mathbf{t}_{\mathrm{i}}, (\mathrm{i} = 1, \ldots, \mathrm{m})$, que capturem a maior quantidade de informação possível disposta nas variáveis explicativas $\mathrm{X}_1, \ldots, \mathrm{X}_{\mathrm{p}}$, neste caso, os marcadores SNPs (os quais assumem os valores 0, 1 e 2, respectivamente aos genótipos aa, aA e AA), visando a predizer a variável dependente Y, neste caso, os dois fenótipos considerados (pH_{45} e pH_{u}). Sob esse enfoque, a equação de regressão é expressa por:

$$\hat{\mathbf{y}} = \hat{\boldsymbol{\beta}}_0 + \hat{\boldsymbol{\beta}}_1\mathbf{t}_1 + \hat{\boldsymbol{\beta}}_2\mathbf{t}_2 + \ldots + \hat{\boldsymbol{\beta}}_{\mathrm{m}}\mathbf{t}_{\mathrm{m}}, \tag{1}$$

em que $\mathbf{t}_{\mathrm{i}}$ é o vetor coluna que compõe a matriz de componentes $\mathbf{T}$ e $\hat{\boldsymbol{\beta}}_{\mathrm{i}}$ é a estimativa dos coeficientes da regressão entre $\mathbf{Y}$ e $\mathbf{T}$, $\forall \mathrm{i} = 1, \ldots, \mathrm{m}$.

A correlação de qualquer par de componentes é igual a 0, isto é: $cor(\mathbf{t}_{\mathrm{i}}, \mathbf{t}_{\mathrm{i}'}) = 0, \forall \mathrm{i} \neq \mathrm{i}'$. Assim, o método PLS reduz o número de termos na equação de regressão, uma vez que o número de componentes na equação (1) geralmente é menor que o número de variáveis $\mathbf{X}$.

Para simplificar os cálculos, utilizam-se variáveis centradas de $\mathbf{Y}$ e $\mathbf{X}_{\mathrm{j}}$ denotadas respectivamente por $\mathbf{U}_1$ e $\mathbf{V}_{1\mathrm{j}}$, em que:

$$\mathbf{U}_1 = \mathbf{Y} - \bar{\mathbf{Y}} \ \text{ e } \ \mathbf{V}_{1\mathrm{j}} = \mathbf{X}_{\mathrm{j}} - \bar{\mathbf{X}}_{\mathrm{j}} \ \text{ para } \ \mathrm{j} = 1, \ldots, \mathrm{p},$$

sendo seus vetores de valores dados por $\mathbf{u}_1 = \mathbf{y} - \bar{\mathrm{y}}\mathbf{1}$ e $\mathbf{v}_{1\mathrm{j}} = \mathbf{x}_{\mathrm{j}} - \bar{\mathrm{x}}_{\mathrm{j}}\mathbf{1}$, em que $\mathbf{1}$ é um vetor coluna n-dimensional, $\mathbf{1}' = (1, 1, ..., 1)'$.

Os componentes são determinados de forma sequencial, sendo que primeiramente é realizada uma regressão de $\mathbf{U}_1$ em função de $\mathbf{V}_{11}$, $\mathbf{V}_{12}$ e, assim por diante, até $\mathbf{V}_{1\mathrm{p}}$. Tais regressões são independentes, portanto, no desenvolvimento do método PLS, as correlações entre os $\mathbf{V}_{1\mathrm{j}}$ são ignoradas.

As equações de regressão resultantes do método dos quadrados mínimos ordinários para $\mathrm{j} = 1, \ldots, \mathrm{p}$, são iguais a:

$$\mathbf{U}_{1\mathrm{j}} = \frac{\mathbf{u}_1^{'}\mathbf{v}_{1\mathrm{j}}}{\mathbf{v}_{1\mathrm{j}}\mathbf{v}_{1\mathrm{j}}}\mathbf{v}_{1\mathrm{j}}.$$

Segundo Garthwaite (1994), geralmente define-se $\mathbf{T}_1$ como uma média ponderada, dada por:

$$\mathbf{T}_1 = \sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{w}_{1\mathrm{j}}\mathbf{U}_{1\mathrm{j}}, \tag{2}$$

em que $\mathrm{w}_{1\mathrm{j}}$ é um peso apropriadamente escolhido, com $\mathrm{w}_{1\mathrm{j}} \geq 0$ e $\sum_{\mathrm{j}=1}^{\mathrm{p}} \mathrm{w}_{1\mathrm{j}} = 1$.

Como o componente $\mathbf{T}_1$ é uma média ponderada dos $\mathbf{U}_{1\mathrm{j}}$, $\mathbf{T}_1$ não contém todas as informações existentes em $\mathbf{X}_\mathrm{j}$. A informação de $\mathbf{X}_\mathrm{j}$ ausente em $\mathbf{T}_1$ pode ser estimada pelos resíduos obtidos na regressão entre $\mathbf{V}_{1\mathrm{j}}$ e $\mathbf{T}_1$. De modo análogo, a variabilidade em $\mathbf{Y}$ que não está sendo explicada por $\mathbf{T}_1$ pode ser estimada pelos resíduos obtidos na regressão entre $\mathbf{U}_1$ e $\mathbf{T}_1$. Estes resíduos são denotados por $\mathbf{V}_{2\mathrm{j}}$ para $\mathbf{V}_{1\mathrm{j}}$ e por $\mathbf{U}_2$ para $\mathbf{U}_1$.

O segundo componente, $\mathbf{T}_2$, é construído do mesmo modo que $\mathbf{T}_1$, porém substituindo $\mathbf{U}_1$ e $\mathbf{V}_{1\mathrm{j}}$ por $\mathbf{U}_2$ e $\mathbf{V}_{2\mathrm{j}}$, respectivamente. Desta forma, o procedimento se estende analogamente para os componentes $\mathbf{T}_2, \ldots, \mathbf{T}_\mathrm{m}$.

Generalizando, suponha que $\mathbf{T}_\mathrm{i}$ seja construído a partir das variáveis $\mathbf{U}_\mathrm{i}$ e $\mathbf{V}_{\mathrm{ij}}$ com $\mathrm{j} = 1, \ldots, \mathrm{p}$. Para obter o componente $\mathbf{T}_{\mathrm{i}+1}$, as variáveis $\mathbf{V}_{(\mathrm{i}+1)\mathrm{j}}$, e $\mathbf{U}_{(\mathrm{i}+1)\mathrm{j}}$, devem ser determinadas. Com este propósito, é feita uma regressão entre

$\mathbf{V}_{ij}$ e $\mathbf{T}_i$ e, assim, $\mathbf{V}_{(i+1)j}$ é definido por:

$$\mathbf{V}_{(i+1)j}=\mathbf{V}_{ij}-\frac{\mathbf{v}_{ij}^{'}\mathbf{t}_i}{\mathbf{t}_i^{'}\mathbf{t}_i}\mathbf{t}_i$$

sendo $\mathbf{t}_i$ o vetor de valores de $\mathbf{T}_i$, $\mathbf{V}_{(i+1)j}$ os resíduos da regressão e $\frac{\mathbf{v}_{ij}^{'}\mathbf{t}_i}{\mathbf{t}_i^{'}\mathbf{t}_i}$ o coeficiente da regressão entre $\mathbf{V}_{ij}$ e $\mathbf{T}_i$.

De forma análoga, $\mathbf{U}_{(i+1)}=\mathbf{U}_i-\frac{\mathbf{u}_i\mathbf{t}_i}{\mathbf{t}_i\mathbf{t}_i}\mathbf{t}_i$ e $\mathbf{u}_{(i+1)}$ são os resíduos da regressão entre $\mathbf{U}_i$ e $\mathbf{T}_i$. Assim, a j-ésima regressão entre $\mathbf{U}_{i+1}$ e $\mathbf{V}_{(i+1)j}$ resulta num coeficiente de regressão dado por:

$$b_{(i+1)j}=\frac{\mathbf{u}_{(i+1)}^{'}\mathbf{v}_{(i+1)j}}{\mathbf{v}_{(i+1)j}^{'}\mathbf{v}_{(i+1)j}}.$$

Analogamente à equação 2, define-se $\mathbf{T}_{(i+1)}$ como sendo:

$$\mathbf{T}_{(i+1)}=\sum_{j=1}^{p} w_{(i+1)j}b_{(i+1)j}\mathbf{V}_{(i+1)j}.$$

O método é repetido a fim de obter $\mathbf{T}_1,\mathbf{T}_2,...,\mathbf{T}_m$ e, após a obtenção dos m componentes, os coeficientes da equação de regressão (1) são determinados por meio do método dos quadrados mínimos ordinários.

Conforme já relatado, a principal característica do método PLS é que a correlação entre qualquer par de componentes é igual a 0, e isso se deve ao fato dos resíduos provenientes da regressão não serem correlacionados com o regressor, ou seja, $\mathbf{V}_{(i+1)j}$ é não correlacionado com $\mathbf{T}_i$. Assim, como cada componente $\mathbf{T}_{(i+1)},\ldots,\mathbf{T}_m$ é uma combinação linear de $\mathbf{V}_{(i+1)j}$, os componentes são não correlacionados com $\mathbf{T}_i$. Essa característica é de suma importância para a seleção genômica ampla, pois o método PLS torna-se uma alternativa eficiente para

se obter preditores apesar da multicolinearidade presente nos dados de marcadores (covariáveis $\mathbf{X}_j$). Além disso, os coeficientes da equação de regressão (1) podem ser estimados por uma simples regressão feita entre $\mathbf{Y}$ e $\mathbf{T}_i$. Ainda, adicionando-se componentes à equação, os componentes anteriores não têm seus coeficientes alterados. Deve-se ressaltar que, após os coeficientes da equação (1) serem estimados, é possível expressar o modelo em relação a $\mathbf{X}_j$, ao invés dos componentes $\mathbf{T}_i$, ou seja, expressar o modelo em termos de suas variáveis originais e com interpretação biológica.

A necessidade dessa descrição detalhada do método PLS é justificada pelo fato desta teoria ser pouco difundida na área de genética e melhoramento. Por outro lado, a teoria do PCR é amplamente utilizada e já dispõe de vasta literatura abordando aspectos teóricos (ver Capítulo 1) e aplicações na referida área (AZEVEDO *et al*., 2013), portanto, sua descrição aprofundada não é destacada no presente trabalho. Quanto ao método da regressão múltipla tradicional, isto é, sem redução de dimensionalidade, vale ressaltar que em situações nas quais o número de marcadores é maior que o número de observações fenotípicas, a aplicação da mesma só é possível via utilização de inversas generalizadas, como a inversa de Moore-Penrose.

2.3 Número ótimo de componentes (variáveis latentes) no PLS

Os graus de liberdade (DoF) quantificam a complexidade intrínseca de um método de regressão (VAN DER VOET, 1999). Segundo Krämer e Sugiyama (2011), a complexidade da análise PLS depende da colinearidade das variáveis preditoras, de forma que, quanto maior a colinearidade, menor é a complexidade e, portanto, menor o DoF (menor o número de componentes). Estes autores apresen-

tam uma estimativa não-viesada dos DoF para PLS com m componentes latentes, $\mathbf{T} = (\mathbf{t}_1, \mathbf{t}_2, \ldots, \mathbf{t}_\mathrm{m})$, a qual é dada por:

$$\mathrm{D\hat{o}F}\,(\mathrm{m}) = 1 + tr\left(\frac{\partial \mathrm{P}_{\mathbf{T}}\mathbf{y}}{\partial \mathbf{y}}\right). \tag{3}$$

O termo constante 1 corresponde à estimativa do intercepto β_0, que consome um grau de liberdade. Para calcular o traço da derivada em (3) de forma explícita é necessário usar um algoritmo baseado na decomposição ortonormal de Lanczos (LANCZOS, 1950) de $\mathbf{X}$. Para $\mathbf{y}$ e m fixos, a decomposição é única (ELDÉN, 2004).

Uma forma mais prática de se determinar o número de componentes é por meio da validação cruzada. Esse método consiste em dividir o conjunto de dados originais em N subconjuntos (*folds*), e realizar N análises, de forma que em cada uma delas um dos subconjuntos é retirado a fim de ser utilizado como recurso para validar a análise realizada. Assim, os valores preditos pelas equações estimadas em cada análise podem ser diretamente comparados com os valores retirados (observados). No contexto de PLS, o número ótimo de componentes será aquele que fornecer a maior correlação entre o conjunto de valores observados e preditos. Embora esse método seja teoricamente mais simples, o mesmo é mais complexo sob o ponto de vista computacional.

2.4 Capacidade Preditiva

A população original de 345 indivíduos foi fracionada em duas diferentes populações, ($\mathrm{N} = 2$): população de treinamento e população de validação. A fim de separar os grupos levando em consideração o menor parentesco, utilizou-se o programa RENUMF90 para prover informações de parentesco que permitiram

compor dois grupos com maior parentesco dentro de cada grupo e menor parentesco entre cada grupo. A população de treinamento, contendo 175 indivíduos, foi utilizada para estimar os efeitos dos marcadores SNPs, considerando os métodos PLS, PCR e regressão múltipla tradicional (sem redução de dimensionalidade). A população de validação, contendo 170 indivíduos, não foi utilizada para estimar os efeitos de marcadores, mas sim, para avaliar o potencial preditivo de cada um dos métodos. Dessa forma, os valores preditos para os fenótipos da população de validação ($\hat{\mathbf{Y}}_{\mathrm{p}}$) foram obtidos da seguinte forma: $\hat{\mathbf{Y}}_{\mathrm{p}} = \mathbf{X}_{\mathrm{p}}\hat{\boldsymbol{\beta}}$, sendo $\mathbf{X}_{\mathrm{p}}$ os genótipos dos marcadores SNPs dos indivíduos da população de validação e $\hat{\boldsymbol{\beta}}$ o vetor de estimativas dos efeitos de marcadores provenientes da população de treinamento.

A correlação entre o vetor de fenótipos predito e observado na população de validação é denominada de capacidade preditiva. Sob o ponto de vista prático, essa grandeza permite calcular a eficiência da seleção dos indivíduos da população de validação utilizando apenas informações genotípicas ($\mathbf{X}_{\mathrm{p}}$) e um vetor de estimativas ($\hat{\boldsymbol{\beta}}$) proveniente da população de treinamento. Em outras palavras, a capacidade preditiva quantifica a habilidade de se predizer o mérito genético dos indivíduos da população de validação sem a necessidade de se utilizar fenótipos desta população.

Considerando os fenótipos pH_{45} e pH_{u} do presente estudo, nota-se que, realmente, a seleção genômica constitui uma ferramenta importante para o melhoramento genético de características de qualidade de carne em suínos, pois se a capacidade preditiva for alta, é possível predizer a qualidade de carne e, assim, identificar os animais superiores, sem a necessidade de abate.

Em posse da escolha do melhor método (PLS, PCR e regressão múltipla tradicional) por meio da capacidade preditiva, os efeitos dos marcadores (em valor absoluto) estimados por este método foram utilizados para a construção dos

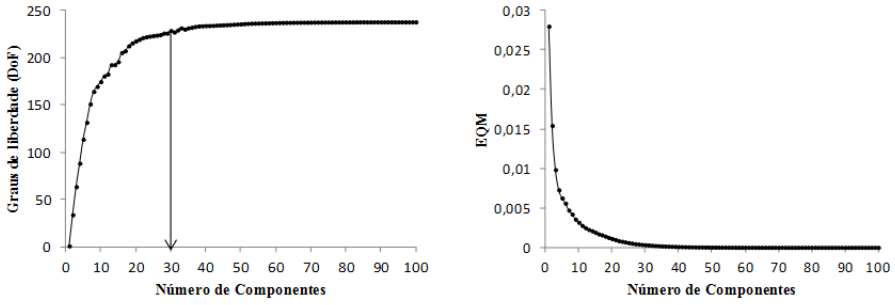
gráficos *Manhattan plot*, nos quais cada ponto representa um marcador SNP com sua exata localização no genoma (eixo X mostrando sua localização no cromossomo) e a magnitude de seu efeito (eixo Y). Estes gráficos são importantes para identificar possíveis regiões cromossômicas (QTLs) diretamente associadas com as características de interesse.

Todas as análises foram realizadas no software R (R DEVELOPMENT CORE TEAM, 2013) por meio das funções *pcr*, *pls.model* e *pls.cv* do pacote *plsdof*.

3 RESULTADOS E DISCUSSÃO

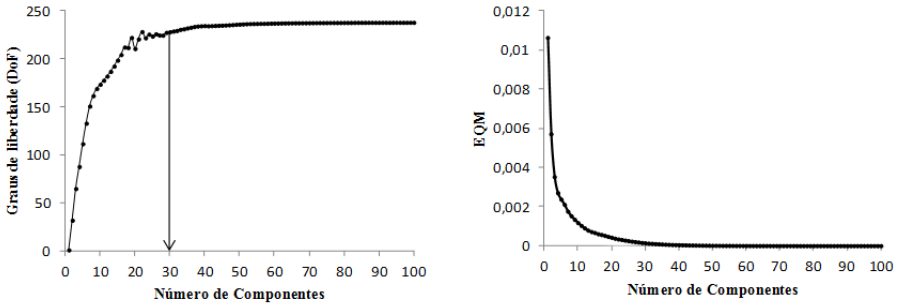
As Figuras 1 e 2 mostram, respectivamente, os resultados da análise de determinação do número ótimo de componentes no PLS para as variáveis pH_{45} e pH_{u}. Por meio destas figuras nota-se que ambos os métodos, DoF e CV, proporcionaram o mesmo número ótimo, o qual pode ser resumido em 30 componentes. Esse valor foi escolhido devido ao fato de se observar a estabilização das curvas a partir deste número de componentes. Ao utilizar números de componentes acima deste valor, tem-se um aumento na complexidade do modelo, mas sua performance permanece inalterada.

Os resultados obtidos na análise de determinação do número de componentes concordam com aqueles mostrados por Krämer e Sugiyama (2011), os quais simularam vários cenários envolvendo diferentes números de componentes e observaram alta concordância entre os métodos DoF e CV quanto à escolha do número ótimo de componentes. Porém, da mesma forma que constatado no presente trabalho, o método CV, por necessitar de maior demanda computacional, é mais oneroso em relação ao tempo de computação e, portanto, em termos práti-



(a) (b)

Figura 1 Determinação do número ótimo de componentes na análise PLS utilizando a teoria de graus de liberdade (a) e validação cruzada (b) para a variável pH da carne suína aos 45 min após o abate



(a) (b)

Figura 2 Determinação do número ótimo de componentes na análise PLS utilizando a teoria de graus de liberdade (a) e validação cruzada (b) para a variável pH da carne suína 24 horas após o abate

cos, recomenda-se o método DoF. Além disso, sob o ponto de vista estatístico, tal método deve ser exaltado devido ao seu maior embasamento teórico em relação ao CV que, por sua vez, pode ser caracterizado como um método empírico baseado em esforço computacional.

As capacidades preditivas obtidas pelos métodos de redução dimensional

(PLS e PCR) e Quadrados Mínimos Ordinários (*Ordinary Least Square* - OLS), para as características de pH_{45} e pH_{u}, são apresentadas na Figura 3. Devido ao fato da análise de número ótimo de componentes no PLS ter mostrado resultados consistentes (DoF e CV indicando 30 componentes), esse mesmo número também foi utilizado na análise PCR a fim de proporcionar uma situação na qual tais métodos sejam diretamente comparáveis.

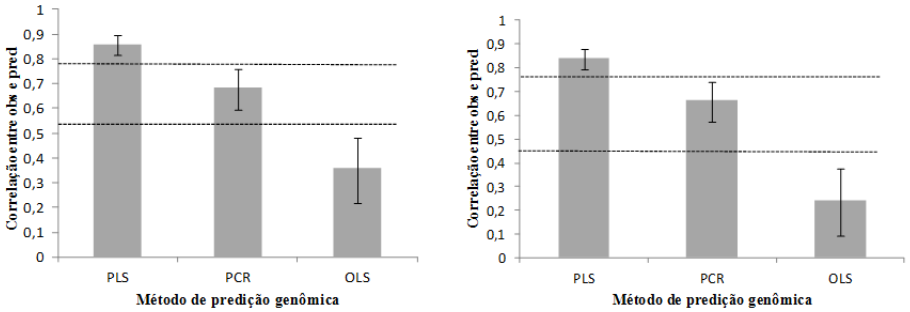


Figura 3 Capacidades preditivas obtidas pelos métodos PLS, PCR e OLS (regressão múltipla tradicional) para as características de pH_{45} (a) e pH_{u} (b)

Nota-se na Figura 3 que o método PLS mostrou-se mais eficiente para ambas as características, seguido pelo método PCR e, por fim, o método OLS. Nota-se, ainda, que essas diferenças na capacidade preditiva são estatisticamente significativas devido ao fato dos intervalos de confiança de 95% para a correlação não se sobreporem.

De forma geral, embora os métodos PLS e PCR sejam semelhantes no que diz respeito à condição de correlação nula entre os componentes e à possibilidade de serem empregados em situações envolvendo número de variáveis maior que o número de observações, como na presente situação (175 observações individuais e 237 marcadores), observou-se uma nítida vantagem do método PLS em relação a

sua performance preditiva. Certamente, essa vantagem deve-se ao fato do método PLS levar em consideração a variável resposta no processo de composição dos componentes, enquanto o método PCR considera apenas as próprias covariáveis. O fraco desempenho do método OLS pode estar associado à não correção dos problemas de multicolinearidade e à utilização de uma matriz inversa generalizada (Moore-Penrose), no caso $\mathrm{n} < \mathrm{p}$, a qual pode, em alguns casos, levar a problemas de sobreparametrização (*overfitting*) se comparada com as inversas clássicas utilizadas nos métodos PLS e PCR.

Ainda em relação a Figura 3, vale ressaltar que os valores de capacidade preditiva obtidos pelo método PLS para os fenótipos pH_{45} e pH_{u}, os quais foram respectivamente 0,85 e 0,84, realmente comprovam a eficiência da seleção para a análise de características de qualidade de carne em suínos. Até o momento, não há relatos na literatura de valores de acurácia de seleção genômica para pH de carne suína, fato este que impossibilita a comparação direta dos resultados obtidos no presente trabalho com outros relatos científicos. Porém, de forma geral, estes valores acima de 80% permitem inferir que seja possível identificar animais geneticamente superiores para pH da carne sem a necessidade de abatê-los, ou seja, utilizando apenas suas informações genotípicas. Tal fato representa um grande avanço tecnológico na área de melhoramento de suínos, pois além de reduzir o intervalo de geração por não necessitar que o animal atinja a idade de abate para inferir sobre seu mérito genético, também é possível evitar que bons animais sejam abatidos e, assim, não destinados à reprodução, por não se conhecer de antemão o mérito de tais animais.

Além da análise preditiva inerente à seleção genômica, também é de interesse a identificação de marcadores SNPs mais relevantes para cada característica, pois nas regiões cromossômicas destes marcadores é possível inferir sobre a pre-

sença de QTLs e/ou genes de efeito maior que podem vir a ser utilizados para fins de seleção. Para tanto, faz-se necessário identificar tais marcadores e suas posições específicas em cada cromossomo, o que pode ser facilmente implementado por análises gráficas como os *Manhattan plots* apresentados na Figura 4.

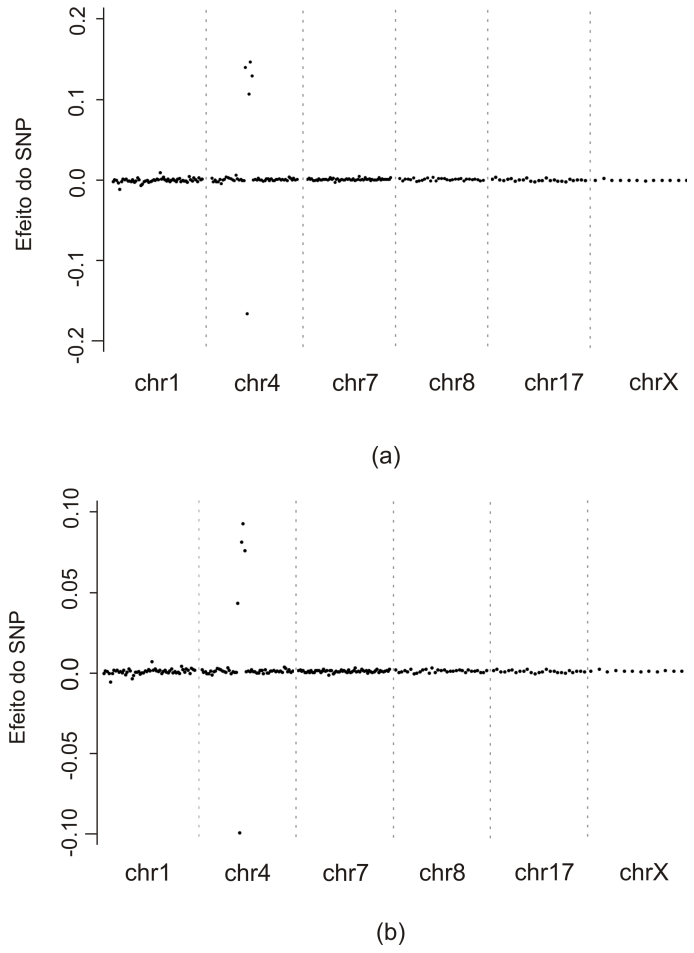


Figura 4 *Manhattan plot* (efeitos estimados dos SNPs ao longo das posições genômicas) para as características pH da carne suína aos 45 min (a) e 24 horas (b) após o abate

De acordo com a Figura 4, nota-se que para ambos os fenótipos (pH_{45} e

Tabela 1 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)

Fenótipo	SNP	chr	pos_Mbp	Efeito
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
	ALGA0026109	4	75,57379	0,095448
	ALGA0026241	4	80,13745	0,089195
	ALGA0026100	4	75,53339	0,050324

pH_{u}) detectou-se uma região sugestiva de QTL no cromossomo 4, pois é visível a alta influência dos SNPs localizados nessa região sobre os fenótipos em questão. Por meio destes gráficos é possível concluir que a mesma região cromossômica controla o pH da carne suína em diferentes tempos. Para um maior detalhamento dessa região, tem-se informações complementares na Tabela 1. Nessa, observa-se que os SNPs mais relevantes para pH_{45} são os mesmos para pH_{u}, os quais situam-se em torno da região entre 75 e 80 Mbp do cromossomo SSC4.

Essa região realmente apresenta influência sobre o pH da carne suína, uma vez que várias outras pesquisas têm reportado QTLs nessa mesma região cromossômica. Tal afirmação é comprovada por meio da Figura 5, proveniente da base de dados PigQTLdb (NATIONAL ANIMAL GENOME RESEARCH PROGRAM, 2014), na qual verifica-se alta densidade de QTLs relatados por várias pesquisas científicas. Dentre essas, destacam-se os resultados encontrados por Stratz *et al*. (2012) e Ma *et al*. (2013), que também utilizaram populações F_2 envolvendo linhas comerciais e identificaram QTLs para pH_{45} nas posições 79,5 e 88,26 Mbp,

respectivamente. Também destacam-se os resultados obtidos por Wimmers *et al.* (2006) e Ponsuksili *et al.* (2010), os quais reportaram QTLs para $\mathrm{pH_u}$ nas regiões 82 e 67 Mbp, respectivamente. Nesses trabalhos foram utilizadas diferentes populações derivadas da raça Duroc.

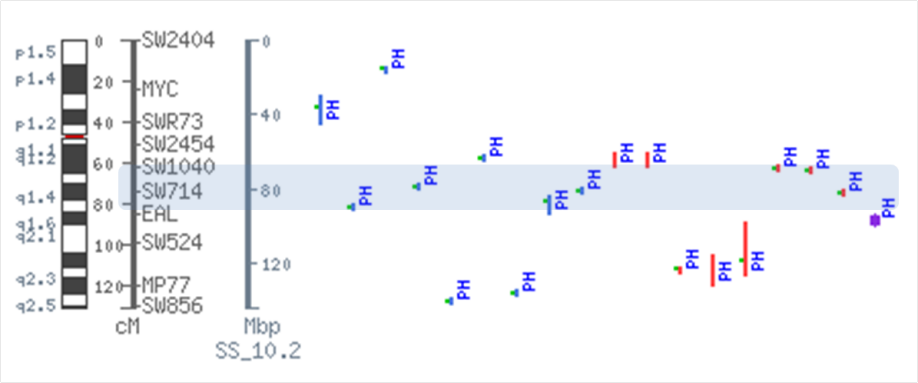


Figura 5 Densidade de QTLs reportados na região intermediária do cromossomo SSC4 proveniente da base de dados PigQTLdb.

Em resumo, embora a comprovação da importância desta região do SSC4 para o pH da carne seja interessante para pesquisas científicas na área de genética suína, ainda são necessárias pesquisas complementares a fim de identificar os genes de efeito maior nesta região. Dessa forma, uma vez constatada a presença desses genes e, compreendidos seus mecanismos de atuação, estes podem ser diretamente empregados com objetivos de seleção por meio de genotipagens específicas para os mesmos, as quais permitem inferir sobre o pH da carne, simplesmente por meio dos genótipos observados para estes genes.

4 CONCLUSÕES

- A metodologia Quadrados Mínimos Parciais é eficiente para a seleção genômica por possibilitar predições satisfatórias (capacidade preditiva acima de 80%) do pH da carne suína, utilizando apenas informações genotípicas baseadas em marcadores SNPs.

- Os métodos Graus de Liberdade e Validação Cruzada foram equivalentes na determinação do número ótimo de componentes na análise por Quadrados Mínimos Parciais.

- O método da Regressão em Componentes Principais apresenta vantagens sobre a regressão múltipla tradicional em relação à predição genômica, porém, este é de eficiência inferior ao método Quadrados Mínimos Parciais.

- A análise Quadrados Mínimos Parciais possibilitou identificar uma região (em torno de 70 e 80 Mbp) relevante no cromossomo 4 para o controle genético do pH da carne suína.

REFERÊNCIAS

AZEVEDO, C. F. *et al*. Regressão via componentes independentes aplicada à seleção genômica para características de carcaça em suínos. **Pesquisa Agropecuária Brasileira**, Brasília, v. 48, n. 6, p. 619-626, jun. 2013.

BENEVENUTO JUNIOR, A. A. **Avaliação de rendimento de carcaça e de qualidade da carne de suínos comerciais, nativos e cruzados.** 2001. 94 f. Dissertação (Mestrado em Ciência e Tecnologia de Alimentos) - Universidade Federal de Viçosa, Viçosa, MG, 2001.

ELDÉN, L. Partial least-squares vs. Lanczos bidiagonalization–I: analysis of a projection method for multiple regression. **Computational Statistics and Data Analysis**, [s.l.], v. 46, n. 1, p. 11-31, 2004.

GARTHWAITE, P. H. An Interpretation of Partial Least Squares. **Journal of the American Statistical Association**, [s.l.], v. 89, n. 425, p. 122-127, mar. 1994.

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**, [s.l.], v. 36, n. 4, p. 511-519, dez. 2013.

HOSKULDSSON, A. PLS Regression Methods. **Journal of Chemometrics**, [s.l.], v. 2, p. 211-228, 1988.

JENKINS, S.; GIBSON, N. High-throughput SNP genotyping. **Comparative and Functional Genomics**, [s.l.], v. 3, n. 1, p. 57-66, fev. 2002.

KRÄMER, N.; SUGIYAMA, M. The Degrees of Freedom of Partial Least Squares Regression. **Journal of the American Statistical Association**, [s.l.], v. 106, n. 494, p. 697-705, fev. 2011.

LANCZOS, C. An Iteration Method for the Solution of the Eigenvalue Problem of Linear Differential and Integral Operators. **Journal of Research of the**

National Bureau of Standards, [s.l.], v. 45, n. 4, p. 225-280, out. 1950.

MA, J. *et al*. Genome-Wide Association Study of Meat Quality Traits in a White Duroc $\times$ Erhualian F2 Intercross and Chinese Sutai Pigs. **PLOS ONE**, [s.l.], v. 8, n. 5, p. 1-11, maio 2013.

MEUWISSEN, T. H. E.; HAYES, B. J.; GODDARD, M. E. Prediction of total genetic value using genome wide dense marker maps. **Genetics**, [s.l.], v. 157, p. 1819-1829, abr. 2001.

NATIONAL ANIMAL GENOME RESEARCH PROGRAM. **PigQTLdb**. Disponível em: <http://www.animalgenome.org/QTLdb/>. Acesso em: 8 jan. 2014.

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.

PONSUKSILI, S. *et al*. Identification of expression QTL (eQTL) of genes expressed in porcine *M. longissimus dorsi* and associated with meat quality traits. **BMC Genomics**, [s.l.], v. 11, n. 572, p. 1-14, 2010.

R DEVELOPMENT CORE TEAM. **R**: a language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, 2013. Disponível em: <http://www.R-project.org.>. Acesso em: 14 nov. 2013.

STRATZ, P. *et al*. A two-step approach to map quantitative trait loci for meat quality in connected porcine F_2 crosses considering main and epistatic effects. **Animal Genetics**, [s.l.], v. 44, p. 14-23, 2012.

VAN DER VOET, H. Pseudo-degrees of freedom for complex predictive models: the example of partial least squares. **Journal of Chemometrics**, [s.l.], v. 13, n. 3, p. 195-208, jul. 1999.

WIMMERS, K. *et al*. QTL for microstructural and biophysical muscle properties and body composition in pigs. **BMC Genetics**, [s.l.], v. 7, n. 15, p. 1-14, mar.

2006.

APÊNDICE

APÊNDICE A - Rotina Computacional usada na análise dos dados

```
memory.limit(size = 100000 )
dados=read.csv("all_gws_2013.csv",h=T,sep=";")

snp=as.matrix(dados[,-(1:13)])
ph45=as.matrix(dados[,2])

snp0=as.matrix(snp[-(176:345),])
ph450=as.matrix(ph45[-(176:345)])

snp1=as.matrix(snp[(176:345),])
ph451=as.matrix(ph45[(176:345)])

library("plsdof")

# Regressão Múltipla Tradicional

beta_ols=ginv(t(snp0)%*%snp0)%*%t(snp0)%*%ph450

gbv_ols=snp1%*%beta_ols
cor.test(gbv_ols,ph451)
```

```
# Regressão em Componentes Principais

pcr=pcr(snp0,ph450,m=30)

coef=pcr$coefficients[,30] #manhattan-plot
coef=as.matrix(coef)
max(coef)

gbv_pcr=snp1%*%coef
cor.test(gbv_pcr,ph451)

# Quadrados Mínimos Parciais: DoF

pls=pls.model(snp0,ph450,100,compute.DoF=TRUE,
                        compute.jacobian=TRUE)
a=information.criteria(pls$RSS, pls$DoF, pls$yhat,
          pls$sigmahat, nrow(snp),criterion="bic")
plot(a$DoF)
as.matrix(a$DoF)

coef=pls$coefficients[,30] #manhattan-plot
coef=as.matrix(coef)
max(coef)

gbv_dof=snp1%*%coef
cor.test(gbv_dof,ph451)
```

```
# Quadrados Mínimos Parciais: CV

pls_cv=pls.cv(snp,ph45,m=100)
plot(pls_cv$cv.error)
```