CARACTERIZAÇÃO DOS MELHORES ESTIMADORES LINEARES NÃO TENDENCIOSOS NO MODELO LINEAR GERAL

ROSA MARIA SALANI MOTA

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ÁREA: ESTATÍSTICA

ORIENTADOR:

Prof. Dr. Euclydes Custódio de Lima Filho

- SÃO PAULO, DEZEMBRO DE 1982 -

Aos meus pais, Osvaldo e Horayda, que me educaram com dedicação e carinho.

As minhas filhas, Bruna e Sula

Ao João Maurício, companheiro, incentivador e, sobretudo, amigo.

A Inã

AGRADECIMENTOS

Este trabalho é resultado da ajuda recebida de muitos, de diversas formas e nas diversas etapas de sua elaboração.

Quero a todos externar os melhores agradecimentos, em especial

Ao professor Euclydes Custodio de Lima e Filho, pela sugestão do tema e por ter proporciohado uma formação básica tão necessária para a realização desta monografia.

Ao professor Adolpho Walter Pimazoni Canton do Departamento de Estatística da Universidade de São Paulo.

Ao professor Plínio Amarantes Quirino Simões pelo seu apoio constante no esclarecimento de conceitos matemáticos e pelo seu estímulo sem os quais não seria possível a realização deste trabalho.

Aos amigos Vera, Franklin, Carlinhos, Wagner, pelo estímulo nas horas difíceis.

Aos colegas do Curso de Pós-Graduação e as amigas da biblioteca do Instituto de Matemática e Estatística da Universidade de São Paulo pela convivência amiga.

Ao Sr. João Baptista Esteves de Oliveira, Alice Perran Taborga e Luzia do Carmo Namiki que colaboraram nesta monografia com o seu trabalho datilográfico.

CONTEÚDO

INTRODUÇÃO

Nas mais diversas áreas científicas e tecnológicas, o pesquisador não raras vezes se depara com o problema da estimação de parâmetros de um modelo matemático cuja estrutura é da forma

$$y = X\beta + e$$

sendo,

$y = (y_i)$ um vetor nx1 de observações

X uma matriz nxp de constantes conhecidas

β um vetor px1 de parâmetros

e

$e =$ (e_i) um vetor nx1 onde,

cada e_i é considerado uma perturbação (ou erro) associado ao i-ésimo valor observado e, satisfaz

$E(e_i) = 0$ para todo $i=1, \ldots, n$.

De um modo geral, o problema da estimação é o de encontrar, com base no vetor das observações, y, um vetor $\hat{y}$ que de algum modo seja um "bom" estimador da $E(y) = X\beta$.

A solução clássica para esse problema, foi primeiramente es tudada por Legendre e Gauss, em um trabalho onde, eles utilizam o então chamado Princípio dos Mínimos Quadrados que consiste em achar um vetor $\hat{y} \in C(X)$ (espaço gerado pelas colunas da matriz X), tal que

$$\min_{\eta \in C(X)} ||y - \eta||^2 = \min_{\eta \in C(X)} (y - X\beta)^t (y - X\beta) = ||y - \hat{y}||^2.$$

Gauss (1821) e Markoff (1912) ocuparam-se do problema adicionando ao modelo linear a condição de que, o erro médio quadrático para cada observação y_i, $i=1, \ldots, n$, fosse igual a σ^2 onde, σ^2 é uma constante real positiva e desconhecida. Além disso consideraram que a correlação existente entre cada par de observações fosse nula.

Tais restrições reduziram bastante a gama de situações nas quais o modelo pudesse ser aplicado.

A partir de 1935, vários trabalhos foram desenvolvidos, (ver, por exemplo Aitken (1935)) com o objetivo de estudar uma forma de encontrar o "melhor" estimador linear de uma função $\lambda^t \beta$, $\lambda \in R^p$, admitindo que a matriz de variância e covariância do vetor das observações, y, fosse dada por

$$D(y) = \sigma^2 V$$

onde

$\sigma^2 > 0$ é uma constante real desconhecida

e

V é uma matriz nxn de constantes reais conhecidas e de posto n.

A partir de 1940 vários autores entre os quais, Goldman e Zelen (1964), mostram que a situação geral (X e V não necessariamente de posto completo) pode ser reduzida através de transformadores adequadas à situação inicial proposta por Gauss-Markoff.

Entretanto, como tais reduções são complexas, tornou-se necessário encontrar-se um modo mais simples de se obter um "melhor" estimador linear de uma função paramétrica, $\lambda^t \beta$, $\lambda \in R^p$, cuja representação seja válida para todas as situações do modelo linear.

Zyskind e Martin (1969) assim como Rao (1971, 1972, 1973, 1975, 1978) e outros ocuparam-se em obter uma teoria unificada para a representação de tais estimadores no modelo linear cuja situação fosse geral.

Entretanto, nestes artigos, os autores tiveram que se preocupar em obter também um suporte computacional eficiente sobretudo para o cálculo de "Inversas Generalizadas" o que, ainda tornam tais representações um tanto difíceis de serem obtidas.

Considerando a importância prática do tema, procuraremos apresentar de forma didática uma caracterização do "melhor" estimador de $X \beta$, sem fazer referência a inversas generalizadas, utilizando o argumento apresentado por Rao (1967) denominado ajustamento por covariância.

Este trabalho foi organizado em quatro capítulos, da seguinte forma:

A notação e os pré-requisitos necessários para uma melhor compreensão do texto, foram apresentados no Capítulo 1.

O segundo capítulo é dedicado a um estudo detalhado da Aproximação dos Mínimos Quadrados sugerido por Legendre e Gauss, e, onde, obtemos que o vetor ajustado, $\hat{y}$, de y por vetores de $C(X)$ (espaço gerado pelas coluna da matriz X) é igual à projeção ortogonal de y sobre esse espaço.

Ainda neste segundo capítulo, admitindo $E(e) = \emptyset$ observamos que o conjunto de estimadores de uma função paramétrica $\lambda^t\beta$, $\lambda \in R^p$, é dado por funções lineares das observações, y, da forma

$$a^t y + \rho^t y \qquad \text{onde,}$$

a é um vetor nx1 pertencente a $C(X)$

e

ρ é um vetor nx1 pertencente a $C(X)^{\perp}$ (espaço complemento ortogonal do espaço $C(X)$ em relação à R^n). Completando a estrutura do modelo linear $y = X\beta + e$ com as restrições do Modelo Linear de Gauss-Markoff ($E(e\ e^t) = \sigma^2 I_n$) obtemos que o "melhor" estimador de $X\beta$ é único e, é igual ao vetor ajustado $\hat{y}$ de y por vetores de $C(X)$ obtido pela aproximação dos mínimos quadrados.

No terceiro capítulo mostramos que, com as restrições $E(e) = \emptyset$, $E(e\ e^t) = \sigma^2\ V$, a singularidade de V induz em algumas restrições no vetor das observações, y, e no vetor dos parâmetros, β. Além disso, com essas restrições, é possível encontrarmos um conjunto de funções lineares da forma $c^t y$, $c \in R^n$, $c \neq \emptyset$, tais que

$$E(c^t y) = 0$$

e

$$\text{var } (c^t y) = 0$$

Assim, como poderemos observar uma função paramétrica $\lambda^t\beta$, $\lambda \in R^p$, pode admitir distintos "melhores" estimadores como função das observações embora, o valor numérico de todos eles sejam iguais.

Ainda, neste terceiro capítulo mostramos que, dependendo da relação existente entre a matriz X_{nxp} e a matriz V_{nxn} o "melhor" estimador de uma função paramétrica $\lambda^t\beta$,$\lambda \in R^p$, pode ser obtido através da solução dos mínimos quadrados.

Utilizando o argumento apresentado por Rao (1973) finalizamos o terceiro capítulo caracterizando o "melhor" estimador de $X\beta$ como sendo a projeção das observações, y, sobre o espaço C(X) segundo a direção dos vetores do espaço C(VZ) onde, Z é uma matriz de posto coluna completo tal que $C(Z) = C(X)^{\perp}$.

Finalmente no quarto capítulo, com o objetivo de facilitarmos o entendimento da teoria, apresentamos dois exemplos de modelos lineares obtidos em experimentos aleatorizados.

CAPÍTULO 1

ALGUNS RESULTADOS DE ÁLGEBRA LINEAR

Com o objetivo de facilitarmos o entendimento dos capítulos subseqüentes, descreveremos aqui alguns resultados da teoria das matrizes, assim como também alguns resultados de Álgebra Linear relacionada com projetores e projetores ortogonais.

Para atingirmos o propósito deste capítulo, por conveniência, admitiremos que os espaços vetoriais sejam todos espaços vetoriais reais com produto interno denotado por (x,y) onde x e y são ambos vetores do mesmo espaço.

1.1 - ALGUMAS DEFINIÇÕES E NOTAÇÕES

DEFINIÇÃO 1.1 - Seja E um espaço vetorial, onde

$$E \subset R^n.$$

Então, para todo vetor $y \in E$

$$y = \begin{pmatrix} y_1 \\ \vdots \\ y_n \end{pmatrix} \text{ com } y_i \in R, \quad i = 1,\dots,n \qquad (1.1)$$

e ainda

$$y^t = \text{vetor transposto de } y \longleftrightarrow y^t = (y_1 : \ldots : y_n) \qquad (1.2)$$

DEFINIÇÃO 1.2 - Sejam E_1 e E_2 dois espaços vetoriais.

Chama-se operador linear ou transformação linear de E_1 em E_2 a toda aplicação

$$T: E_1 \longrightarrow E_2 \text{ tal que}$$

a. $T(x,y) = T(x) + T(y)$ para todo x e $y \in E_1$;

b. $T(\alpha x) = \alpha T(x)$ para todo $x \in E_1$ e $\alpha \in R$.

OBSERVAÇÃO - Um operador linear

$T: E_1 \rightarrow E_1$ é também chamada de

endomorfismo de E_1 (1.3)

DEFINIÇÃO 1.3 - Chama-se núcleo de um operador linear

$$T: E_1 \longrightarrow E_2,$$

ao conjunto de todos os vetores $x \in E_1$, tais que

$$T(x) = \emptyset \text{ (vetor nulo)} \qquad (1.4)$$

NOTAÇÃO -

$$\text{núcleo de } T = Ker(T) = \left\{x \in E_1 \mid T(x) = \emptyset\right\} \qquad (1.5)$$

o conjunto

$$Im(T) = \left\{y \in E_2 \mid \text{existe } x \in E_1 \text{ tal que } T(x) = y\right\} \text{ é denominado imagem de } T \qquad (1.6)$$

Observe que Ker(T) e a Im(T)

são sub-espaços de E_1 e E_2, respectivamente. (1.7)

DEFINIÇÃO 1.4 - Seja E um espaço vetorial e T um endomorfismo de E.

Chamamos de transformação adjunta de T à transformação T^t tal que

$$(T(x),y) = (x,T^t(y)) \text{ para todo } x \text{ e } y \in E$$

e ainda, se $[T]_\beta = A =$ matriz de T com relação à base β de E (ver Halmos (1978), pg. 109, temos

$$[T^t]_\beta = A^t$$

onde

$$A^t = \text{matriz transposta de } A. \qquad (1.8)$$

DEFINIÇÃO 1.5 - Seja E um espaço vetorial e T um endomorfismo de E.

Dizemos que T é auto-adjunta (ou simétrica) se, e somente se,

$$T = T^t$$

ou equivalentemente $[T]_\beta = A = A^t$, ou seja, a matriz de T com relação a

uma base β de E é uma matriz simétrica. (1.9)

DEFINIÇÃO 1.6 - Dizemos que um endomorfismo T de E é idempotente se, e somente se,

$$T^2 = T \circ T = T \qquad (1.10)$$

ou equivalentemente se

$$[T]_\beta = A$$

então T é idempotente se, e somente se,

$$A^2 = A \text{ ou seja, a matriz } A \text{ é idempotente.} \qquad (1.11)$$

DEFINIÇÃO 1.7 - Seja E um espaço vetorial de dimensão finita e T um endomorfismo de E.

Dizemos que T é positiva definida (p.d.) (ou positiva semi-definida (p.s.d.)) se

$$T = T^t$$

e para todo $y \in E$ $(T(y),y) \geq 0$, onde

$$(T(y),y) = 0 \longleftrightarrow y = \emptyset \qquad (1.12)$$

(ou $T = T^t$ e para todo $y \in E$ $(T(y),y) \geq 0$, onde

$$(T(y),y) = 0 \text{ para algum } y \neq 0). \qquad (1.13)$$

DEFINIÇÃO 1.8 - Seja E um espaço vetorial de dimensão finita e T um endomorfismo de E.

Dizemos que T é positiva se, e somente se,

T é positiva definida ou positiva semi-definida. (1.14)

Equivalentemente se

$$[T]_\beta = A$$

então T é positiva se, e somente se,

$A = B^t B$ para alguma matriz B tal que

$$\text{posto } (B) = \text{posto } (A)$$

e neste caso, diremos também que

$$A \text{ é uma matriz positiva.} \qquad (1.15)$$

DEFINIÇÃO 1.9 - Seja E um espaço vetorial e T um endomorfismo de E. Um auto valor de T é um escalar $\lambda \in R$ tal que exista um vetor não nulo $x \in E$ com

$$T(x) = \lambda x. \qquad (1.16)$$

Se λ é um auto valor de T então para todo $x \in E$ tal que

$T(x) = \lambda x$ dizemos que x é um auto vetor de T associado ao auto valor λ (1.17)

Equivalentemente, se $[T]_\beta = A$ então λ é um auto valor de T e $x \in E$ é um auto vetor de T associado ao auto valor λ se, e somente se, $Ax = \lambda x$.

Neste caso, costumamos dizer também que λ é um auto valor da matriz A e x é um auto vetor de A associado ao auto valor λ (1.18)

DEFINIÇÃO 1.10 - Seja E um espaço vetorial e x e y dois vetores de E.

Dizemos que x e y são ortogonais se, e somente se

$$(x,y) = 0. \qquad (1.19)$$

Se S é um conjunto de vetores em E, dizemos que S é um conjunto ortonormal se, dois quaisquer vetores distintos de S são ortogonais e para todo vetor $x \in S$ temos que

$$(x,x) = 1 \tag{1.20}$$

DEFINIÇÃO 1.11 - Seja E um espaço vetorial e S um sub-espaço de E.

Dizemos que o vetor $x \in E$ é ortogonal a S se e somente,se

$$(x,y) = 0 \text{ para todo vetor } y \in S. \tag{1.21}$$

DEFINIÇÃO 1.12 - Seja E um espaço vetorial e S um sub-espaço de E.

Dizemos que o sub-espaço $S^{\perp}$ de E é o espaço complemento ortogonal de S em E se e somente se,

$$\text{para todo } y \in S^{\perp} \longrightarrow y \text{ é ortogonal a } S. \tag{1.22}$$

1.2 - ALGUNS ASPECTOS SOBRE MATRIZES

Considerando A uma matriz $m \times n$ de números reais, temos

DEFINIÇÃO 2.1 - O espaço vetorial gerado pelas colunas da matriz A (espaço coluna de A) é igual ao conjunto de vetores $y \in R^m$ tal que

$$y = Ax \text{ para algum vetor } x \in R^n$$

simbolicamente

$$C(A) = \left\{ y \in R^m / y = Ax \text{ para algum } x \in R^n \right\}.$$

DEFINIÇÃO 2.2 - O espaço vetorial gerado pelas linhas de A (espaço linha de A) é igual ao conjunto de vetores $y \in R^n$ tal que

$$y^t = x^t A \text{ para algum } x \in R^m$$

simbolicamente

$$L(A) = \left\{ y \in R^n \mid y^t = x^t A \text{ para algum } x \in R^n \right\}.$$

Observemos que

$$L(A) = C(A^t) \tag{1.23}$$

e também, se T é uma aplicação linear de R^n em R^m tal que a matriz de T com relação as bases β_n e β_m de R^n e R^m respectivamente, é dada por A, então $Im(T) = C(A)$.

DEFINIÇÃO 2.3 - O núcleo da matriz A é igual ao conjunto dos vetores $x \in R^n$ tal que $Ax = \emptyset$, simbolicamente

$$Ker(A) = \left\{ x \in R^n \mid Ax = \emptyset \right\}.$$

PROPOSIÇÃO 2.1 - O Ker(A) é igual ao espaço complemento ortogonal de L(A) (1.24)

PROVA - Como, para todo $y \in Ker(A) \longrightarrow Ay = \emptyset$ e para todo

$$x \in L(A) \longleftrightarrow \text{(por 1.22)}$$

existe $z \in R^m$ tal que $x = A^t z$, temos

$$(y,x) = (y, A^t z) = (Ay, z) = (\emptyset, z) = 0 \longleftrightarrow$$

$$y \in (L(B))^{\perp} = (C(A^t))^{\perp}$$

e portanto

$$Ker(A) = (C(A^t))^{\perp}.$$

COROLÁRIO 2.1 - Se A é uma matriz $m \times n$ de números reais, então

$$R^n = L(A) \overset{\perp}{\oplus} Ker(A) \tag{1.25}$$

onde $\overset{\perp}{\oplus}$ = soma direta de sub-espaços ortogonais.

COROLÁRIO 2.2 - Se A é uma matriz $m \times n$ simétrica, então,

a) $C(A^t) = L(A) = C(A)$ (1.26)

b) $Ker(A) = (C(A))^{\perp}$ (1.27)

e assim

$$R^m = C(A) \overset{\perp}{\oplus} Ker(A). \tag{1.28}$$

PROVA - Trivial.

COROLÁRIO 2.3 - Se A é uma matriz $m \times n$, então

$$Ker(A^t) = C(A)^{\perp} \tag{1.29}$$

e assim

$$R^m = C(A) \overset{\perp}{\oplus} Ker(A^t). \tag{1.30}$$

PROVA - Trivial.

PROPOSIÇÃO 2.2 - Se E_1 e E_2 são dois sub-espaços de R^n, então

a) - $(E_1+E_2)^{\perp} = E_1^{\perp} \cap E_2^{\perp}$ (1.31)

b) - $(E_1^{\perp}+E_2^{\perp}) = (E_1 \cap E_2)^{\perp}$ (1.32)

PROVA

a) - como $E_1 \subset E_1+E_2$ e $E_2 \subset E_1+E_2 \rightarrow (E_1+E_2)^{\perp} \subset E_1^{\perp}$ e

$$(E_1+E_2)^{\perp} \subset E_2^{\perp} \rightarrow (E_1+E_2)^{\perp} \subset E_1^{\perp} \cap E_2^{\perp}.$$

inversamente, como para todo vetor

$$y \in E_1^{\perp} \cap E_2^{\perp} \rightarrow y \in E_1 \text{ e } y \in E_2$$

e para todo vetor $x \in E_1+E_2 \rightarrow$ existe $x_1 \in E_1$ e $x_2 \in E_2$ tal que

$$x = x_1 + x_2$$

e (por (1.21)) $(y,x_1) = 0$, $(y,x_2) = 0$, então

$$(y,x) = (y,x_1+x_2) = (y,x_1) + (y,x_2) = 0 \rightarrow$$

$$\text{(por(1.21))} \; y \in (E_1+E_2)^{\perp}$$

e portanto, concluímos que

$$(E_1+E_2)^{\perp} = E_1^{\perp} \cap E_2^{\perp}.$$

b - Substituindo E_1 por $E_1^{\perp}$ e E_2 por $E_2^{\perp}$ em a) temos

$$(E_1+E_2)^{\perp} = (E_1^{\perp})^{\perp} \cap (E_2^{\perp})^{\perp}$$

e como E_1 e E_2 são sub-espaços de R^n, n: finito, então

$$(E_1^{\perp}+E_2^{\perp})^{\perp} = E_1 \cap E_2$$

e portanto

$$E_1^{\perp} + E_2^{\perp} = (E_1 \cap E_2)^{\perp}.$$

PROPOSIÇÃO 2.3 - Se A e B são matrizes $n \times p$ e $p \times m$, respectivamente, então

$$\text{posto }(AB) = \text{posto}(A) - \dim(C(A^t) \cap (C(B))^\perp). \quad (1.33)$$

PROVA - Sendo AB uma matriz $n \times m$, então

$$\dim(C(A)) = \text{posto}(A) = n - \dim(C(A)^\perp)$$

e

$$\dim(C(AB)) = \text{posto}(AB) = n - \dim(C(AB)^\perp)$$

mas, como

$$C(AB)^\perp = \text{Ker}(AB)^t = \{x \in R^n | x^t AB = \emptyset\} =$$

$$= \{x \in R^n | x^t A = \emptyset\} \cup \{x \in R^n | x^t A \neq 0 \text{ e } x^t AB = \emptyset\}$$

onde esses dois conjuntos são disjuntos e

$$\{x \in R^n | x^t A = \emptyset\} = C(A)^\perp$$

$$\{x \in R^n | x^t A \neq \emptyset \text{ e } x^t AB = \emptyset\} = \{x \in R^n | A^t x \neq \emptyset \rightarrow (A^t x)^t B = \emptyset\} \longleftrightarrow$$

$$A^t x \in C(A^t) \text{ e } A^t x \in C(B)^\perp$$

temos que

$$\dim(C(AB)^\perp) = \dim C(A)^\perp + \dim\{x \in R^n | A^t x \in C(A^t) \cap C(B)^\perp\}$$

e portanto

$$\text{posto}(AB) = n - \dim(C(A)^\perp) - \dim(C(A^t) \cap C(B)^\perp) =$$

$$= \text{posto}(A) - \dim(C(A^t) \cap C^\perp(B)).$$

EXEMPLO 2.1 - Se A é uma matriz $m \times n$, então

$$\text{posto}(A^t A) = \text{posto}(A) = \text{posto}(A^t) = \text{posto}(AA^t)$$

pois

$$\text{posto}(A^tA) = \text{posto}(A^t) - \dim(C(A)\cap C(A)^{\perp}) = \text{posto}(A^t)$$

e

$$\text{posto}(AA^t) = \text{posto}(A) - \dim(C(A^t)\cap C(A^t)^{\perp}) = \text{posto}(A)$$

e como $\text{posto}(A) = \text{posto}(A^t)$, então

$$\text{posto}(A^tA) = \text{posto}(A) = \text{posto}(A^t) = \text{posto}(AA^t).$$

PROPOSIÇÃO 2.4 - Supondo as matrizes $\mathbb{C}p\times m$ e $Am\times n$ onde

$$\text{posto}(A) = n \le m.$$

Se $E_1 = C(A)\cap \text{Ker}(\mathbb{C})$ então $E_1^{\perp}\cap C(A) = C(P_A\mathbb{C}^t)$, onde

$$P_A = A(A^tA)^{-1}A^t \tag{1.34}$$

PROVA - Sendo $E_1^{\perp}$ = complemento ortogonal de E_1 em R^m, então

$$E_1 = C(A)\cap\text{Ker}(\mathbb{C}) \longleftrightarrow E_1^{\perp} = (C(A)\cap\text{Ker}(\mathbb{C}))^{\perp} = \text{por } (1.32) =$$

$$= C(A)^{\perp} + (\text{Ker}(\mathbb{C}))^{\perp} = \text{por } (1.24) = C(A)^{\perp} + C(\mathbb{C}^t)$$

e

$$E_1^{\perp}\cap C(A) = (C(A)^{\perp} + C(A^t))\cap C(A) =$$

$$= C(A)^{\perp}\cap C(A) + C(\mathbb{C}^t)\cap C(A) = C(\mathbb{C}^t)\cap C(A)$$

assim, como para todo $y\in R^m$, $y\in C(A) \longleftrightarrow y = Ax$ para algum

$$x\in R^n \longleftrightarrow y = P_Ay \quad \text{e} \quad y - P_Ay = (I_m - P_A)y = \emptyset$$

onde I_m é a matriz identidade de ordem m, e

$$PA = A(A^tA)^{-1}A^t, \text{ temos,}$$

$$y \in E_1^{\perp} \cap C(A) \longleftrightarrow y \in E_1^{\perp} \text{ e } y \in C(A) \longleftrightarrow$$

$$y \in C(\mathbb{C}^t) \text{ e } y \in C(A) \longleftrightarrow$$

$$y = P_A y \text{ e } y = (I_m - P_A)y + \mathbb{C}^t\beta$$

para algum $\beta \in R^m \longleftrightarrow$

$$y = P_A y = P_A(I_m - P_A)y + P_A(\mathbb{C}^t\beta) \longleftrightarrow y = (P_A\mathbb{C}^t)\beta$$

para algum vetor $\beta \in R^m$ e portanto $E_1^{\perp} \cap C(A) = C(P_A\mathbb{C}^t)$, onde

$$P_A = A(A^tA)^{-1}A^t.$$

PROPOSIÇÃO 2.5 - Para toda matriz $A m \times n$ temos

$$a_0)\ C(A^tA) = C(A^t) \tag{1.35}$$

$$b_0)\ C(AA^t) = C(A) \tag{1.36}$$

PROVA - Trivial usando o Exemplo 2.1 anterior lembrando que $C(A^tA) \subset C(A^t)$ e também $C(AA^t) \subset C(A)$.

1.3 - PROJETORES E PROJETORES ORTOGONAIS

Definição 3.1 Projeções - Sejam E_1 e E_2 dois sub-espaços do espaço vetorial E tal que $E = E_1 \oplus E_2$.

Seja P um endomorfismo de E.

Dizemos que P é uma projeção de E no sub-espaço E_1 segundo a direção dos vetores e E_2 se, e somente se,

para todo vetor $y \in E$ tal que $y = y_1 + y_2$ com $y_1 \in E_1$ e $y_2 \in E_2$

$$P(y) = y_1.$$

É claro que $I_m(P) = E_1$ e o $Ker(P) = E_2$.

TEOREMA 3.1 - Seja E um espaço vetorial e P um endomorfismo de E. Então,

$$P \text{ é uma projeção de } E \longleftrightarrow P \text{ é idempotente.} \qquad (1.37)$$

PROVA - Supondo E_1 e E_2 dois sub-espaços de E tal que

$$E = E_1 \oplus E_2$$

e P projeção de E em E_1 segundo E_2. Temos,

$$\text{para todo } y_1 \in E_1 \text{ que } P(y_1) = y_1$$

e

$$\text{para todo } y_2 \in E_2 \text{ que } P(y_2) = \emptyset,$$

assim, como para todo $y \in E$, y é univocamente determinado por

$$y = y_1 + y_2 \text{ com } y_1 \in E_1 \text{ e } y_2 \in E_2$$

então

$$P^2(y) = P(P(y)) = P(P(y_1+y_2)) = P(y_1) = y_1 = P(y)$$

ou seja $P^2 = P$.

Reciprocamente, supondo de $P^2 = P$, mostremos que P é uma projeção.

Seja

$$E_1 = \{y \in E / P(y) = y\} \quad e \quad E_2 = Ker(P).$$

Claramente, E_1 e E_2 são sub-espaços de E tais que $E = E_1 \oplus E_2$ e $I_m(P) = E_1$, pois, para todo $y \in E$ vale a identidade

$$y = P(y) + y - P(y) = P(y) + (I_E - P)(y)$$

I_E = operador identidade de E, escrevendo

$$y_1 = P(y) \quad e \quad y_2 = (I_E - P)(y)$$

temos

$$P(y_1) = P(P(y)) = P^2(y) = P(y) = y_1$$

e

$$P(y_2) = P(I_E - P)(y) = P(y) - P^2(y) = \emptyset$$

ou seja,

$$y = y_1 + y_2 \text{ com } y_1 \in E_1 \text{ e } y_2 \in E_2$$

e ainda para todo $y \in E_1 \cap E_2$ tem-se que

$$y \in E_1 \text{ e } y \in E_2 \longrightarrow \text{por definição} \longrightarrow P(y) = y \text{ e } P(y) = \emptyset$$

e portanto $y = \emptyset \longrightarrow E = E_1 \oplus E_2$ onde $P(E) = Im(E) = E_1$, ou seja P é uma projeção de E na $Im(P) = E_1$ segundo a direção de vetores de $E_2 = Ker(P_2)$.

TEOREMA 3.2 - Se P é uma projeção de um espaço vetorial E,então, E é a soma direta da Im(P) e do Ker(P). Reciprocamente, se E é a soma direta de dois sub-espaços E_1 e E_2, então, existe uma única projeção P, de E, cuja imagem é E_1 e cujo núcleo é E_2.

PROVA - Supondo P uma projeção de E então por (1.37) $P^2 = P$ e assim demonstrado de maneira idêntica à recíproca do Teorema 3.1 obteremos que à recíproca do Teorema 3.1, obteremos que

$$E = Im(P) \oplus Ker(P)$$

recíprocamente.

Se $E = E_1 \oplus E_2$ e P é uma aplicação de E em E tal que

$$Im(P) = E_1 \quad e \quad Ker(P) = E_2$$

é trivial que P é um endomorfismo de E tal que $P^2 = P$ e portanto P é uma projeção de E em E_1 segundo a direção de vetores de E_2.

Suponto P* uma outra projeção de $E = E_1 \oplus E_2$ tal que

$$Im(P^*) = E_1 \quad e \quad Ker(P^*) = E_2,$$

temos, para todo $y \in E$, y é univocamente determinado para

$$y = y_1 + y_2 \text{ com } y_1 \in E_1 \text{ e } y_2 \in E_2$$

e

$$P^*(y) = y_1 = P(y) \text{ para todo } y \in E \longleftrightarrow P^* = P$$

ou seja P é a única projeção de E tal que $Im(P) = E_1$ e $Ker(P) = E_2$.

PROPRIEDADE 3.1 - Seja E um espaço vetorial. Um endomorfismo P de E é uma projeção $\longleftrightarrow$ $(I_E - P)$ é também uma projeção de E.

(1.38)

Além disso, se

$$E_1 = Im(P) \quad e \quad E_2 = Ker(P)$$

então

$$E_1 = Ker(I_E - P) \quad e \quad E_2 = Im(I_E - P) \tag{1.39}$$

PROVA - Como P é uma projeção $\longleftrightarrow$ $P^2 = P$ (por 1.37), então

$$P^2 = P \longleftrightarrow P(I_E - P) = \emptyset \longleftrightarrow (I_E - P)^2 = (I_E - P) \longleftrightarrow I_E - P$$

é uma projeção de E, além disso, se $Im(P) = E_1$ e $Ker(P) = E_2$, temos $P(E_1) = E_1$ e $P(E_2) = \emptyset$, assim

$$(I_E - P)(E_1) = I_E(E_1) - P(E_1) = \emptyset$$

e

$$(I_E - P)(E_2) = I_E(E_2) - P(E_2) = E_2$$

ou seja

$$Im(I_E - P) = E_2 \quad e \quad Ker(I_E - P) = E_1.$$

TEOREMA 3.3 - Se um espaço vetorial E é a soma direta de uma família $(M_i)_{i=1,\dots,s}$ de sub-espaços de E, então existe uma família $(P_i)_{i=1,\dots,s}$ de endomorfismo de E que satisfaz as condições

a) $P_i^2 = P_i$, $i=1,\dots,s$

b) $Im(P_i) = M_i$, $i=1,\dots,s$

c) $P_iP_j = \emptyset$, se $i \neq j$, i e $j=1,\dots,s$

d) $\sum_{i=1}^{s} P_i = I_E$.

PROVA - Supondo $E = M_1 \oplus M_2 \oplus \dots \oplus M_s$ temos que para todo $y \in E$ existe um único $y_i \in M_i$ tal que

$$y = \sum_{i=1}^{s} y_i.$$

Definindo P_i, $i=1,\dots,s$ uma aplicação de E em E, tal que $P_i(y) = y_i$ é fácil de ver que P_i, $i=1,\dots,s$ é um endomorfismo de E, e satisfaz

a) $P_i^2 = P_i$ $(i=1,\dots,s)$ pois

$P_i^2(y) = P_i(P_i(y)) = P_i(y_i) = y_i = P_i(y)$ para todo $y \in E$

b) $Im(P_i) = M_i$, pois, $P_i(y) = y_i \in M_i$ para todo $y \in E \longrightarrow$

$P_i(E) = \sum_{i=1}^{s} P_i(M_i) = P_i(M_i) = M_i$

e ainda, o $Ker(P_i) = M_1 \oplus \dots \oplus M_{i-1} \oplus M_{i+1} \oplus \dots + M_s$, pois como para todo $y \in E \longrightarrow P_i(y) = y_i \in M_i$ então se $P(y) = \emptyset$, temos que $y_i = 0$ e reciprocamente se $y_i = 0$ e $y = y_1 + \dots + y_{i-1} + y_{i+1} + \dots y_n$ com $y_j \in M_j$, $j=1,\dots,s$ e $j \neq i$ então $P_i(y) = \emptyset$, isto é, o vetor y $y \in Ker(P_i)$ é na realidade uma soma de vetores dos sub-espaços M_j com $j=1,\dots,s$; $j \neq i$.

Assim, como

$$\mathrm{Im}(P_j) = M_j, \text{ para } j = 1,\dots,s$$

e

$$\mathrm{Ker}(P_i) = \oplus \sum_{\substack{j=1 \\ j\neq i}}^{s} M_j$$

então para todo $y \in E$ temos

c) $P_iP_j(y) = P_i(y_j) = \emptyset \longrightarrow P_iP_j = \emptyset$

para i e $j = 1,\dots,s$, $i \neq j$, e também, como

d) $y = P_1(y) + \dots + P_s(y) \longleftrightarrow y = \left(\sum_{i=1}^{s} P_i\right)(y)$, temos

$$I_E = \sum_{i=1}^{s} P_i .$$

TEOREMA 3.4 - Seja E um espaço vetorial e $(P_i)_{i=1,\dots,s}$ uma família de endomorfismo de E tal que

a) $P_i^2 = P_i$, $i=1,\dots,2$

b) $P_iP_j = \emptyset$ se $i \neq j$, i e $j=1,\dots,s$

c) $I_E = \sum_{i=1}^{s} P_i$

Se, $M_i = \mathrm{Im}(P_i)$ temos

d_1) $E = M_1 \oplus \dots \oplus M_s$

d_2) a família de projeções associada à decomposição do espaço $E = M_1 \oplus \ldots \oplus M_s$ é a própria família $(P_i)_{i=1,\ldots,s}$.

PROVA - Supondo P_i, $i=1,\ldots,s$ endomorfismos de E que satisfazem as condições a), b) e c) e

$$Im(P_i) = M_i$$

certamente $E = \sum_{i=1}^{s} M_i$, pois, por c) $I_E = \sum_{i=1}^{s} P_i$, então para todo $y \in E$

$$I_E(y) = y = \left(\sum_{i=1}^{s} P_i\right)(y) = \sum_{i=1}^{s} P_i(y)$$

onde $P_i(y) \in M_i$, $i=1,\ldots,s$, e esta expressão para y é única, pois, se

$$y = \sum_{i=1}^{s} y_i$$

com $y_i \in M_i$, digamos $y_i = P_i(y_i)$, então, usando a) e b), temos

$$P_j^2(y_j) = P_j(P_j(y_j)) = P_j(y_j) = y_j$$

e

$$P_j(y) = \sum_{i=1}^{s} P_j(y_i) = \sum_{i=1}^{s} P_j(P_i(y_i)) = P_j^2(y_j) = y_j$$

o que prova que $E = M_1 \oplus \ldots \oplus M_s$ e ainda, como

$$Ker(P_i) = \oplus \sum_{j=1}^{s} M_j, \text{ para } j \neq i$$

e por a) $P_i^2 = P_i$ onde $M_i = Im(P_i)$ temos que

P_i é a projeção de E sobre M_i segundo a direção dos vetores do

$$Ker(P_i) = \oplus \sum_{j=1}^{s} M_j, \text{ para } j \neq i.$$

Geometricamente observamos então que projeções podem ser usadas para descrever decomposições do espaço vetorial E em somas diretas e vice-versa, ou seja, a família de decomposições em soma direta estão associadas a famílias de projeções.

Prosseguindo no espírito do Teorema 3.2 e Propriedade 3.1, apresentamos, sem demonstração, condições sob as quais algumas combinações algébricas de projeções são elas próprias projeções. A demonstração pode ser encontrada por exemplo, em Halmos (1978) pg. 81 e Jacy Monteiro (1969), pg. 118.

TEOREMA 3.5 - Seja E um espaço vetorial.

Se P_1 e P_2 são projeções de E, então

a) $Im(P_1) \subset Im(P_2) \longleftrightarrow P_2P_1 = P_1$ (1.40)

b) $P_1 + P_2$ é uma projeção $\longleftrightarrow P_2P_1 = P_1P_2 = \emptyset$ e, satisfeita esta condição, então

$$Im(P_1+P_2) = Im(P_1) \oplus Im(P_2)$$

$$Ker(P_1+P_2) = Ker(P_1) \cap Ker(P_2). \quad (1.41)$$

c) P_1-P_2 é uma projeção $\longleftrightarrow P_1P_2 = P_2P_1 = P_2$, e, satisfeita esta condição, então

$$Im(P_1-P_2) = Im(P_1) \cap Ker(P_2)$$

$$Ker(P_1-P_2) = Ker(P_1) \oplus Im(P_2) \quad (1.42)$$

d) Se $P_1P_2 = P_2P_1 = P$, então,

d_1) P é uma projeção de E, onde

$$Im(P) = Im(P_1) \cap Im(P_2)$$

e o

$$Ker(P) = Ker(P_1) + Ker(P_2) \qquad (1.43)$$

d_2) $P_1 + P_2 - P$ é uma projeção de E onde

$$Im(P_1+P_2-P) = Im(P_1) + Im(P_2)$$

e o

$$Ker(P_1+P_2-P) = Ker(P_1) \cap Ker(P_2) \qquad (1.44)$$

Enunciaremos agora um teorema de grande importância no estuo de estimação em modelos lineares no qual utilizaremos os sub-espaços gerados pelas colunas ou linhas de uma matriz.

TEOREMA 3.6 - Se E_1 e E_2 são sub-espaços de E tal que $E = E_1 \oplus \oplus E_2$, então, uma condição necessária e suficiente para que E_1 e E_2 sejam invariantes sob um endomorfismo V de E é que

$$PV = VP \text{ ou seja, } V \text{ e } P \text{ comutam,}$$

onde

$$P \text{ é uma projeção de E em } E_1 \text{ segundo } E_2. \qquad (1.45)$$

PROVA - Supondo que $PV = VP$ onde P é uma projeção de E em E_1 segundo E_2 temos,

se $y \in E_1$, $V(y) = VP(y) = PV(y) \longrightarrow V(y) \in E_1$

se $y \in E_2$, $V(y) = VP(y) = V(\emptyset) = \emptyset \longrightarrow V(y) \in E_2$

e portanto, E_1 e E_2 são invariantes sob V.

Admitindo agora que E_1 e E_2 são invariantes sob V temos

para todo $y \in E_1 \longrightarrow V(y) \in E_1$ e $P(y) = y \longrightarrow PV(y) =$

$$= V(y) = VP(y)$$

para todo $y \in E_2$; $V(y) \in E_2$ e $P(y) = \emptyset \longrightarrow P(V(y)) =$

$$= PV(y) = \emptyset = VP(y),$$

assim, como para todo $y \in E$, $y = y_1 + y_2$ com $y_1 \in E_1$ e $y_2 \in E_2$, então

$$PV(y) = PV(y_1) + PV(y_2) = VP(y_1) + VP(y_2) = VP(y)$$

e portanto

$$PV = VP$$

Halmos (1978) pg. 98 afirma que

TEOREMA 3.7 - Se E é um espaço vetorial de dimensão finita. Correspondente a qualquer endomorfismo V em E, existe uma transformação linear inversível T em E tal que

TV é uma projeção de E sobre a Im(TV) segundo a direção dos vetores do Ker(V).

Uma classe especial de projetores poderá ser encontrada quando o espaço vetorial E é decomposto em soma direta de um sub-espaço E_1 de E e do seu complemento ortogonal em E.

Tal classe é definida por:

DEFINIÇÃO 3.2 Projetores Ortogonais - Seja E um espaço vetorial com produto interno e E_1 um sub-espaço de E.

Então, se $E = E_1 \overset{\perp}{\oplus} E_2$ onde $E_2 = E_1^{\perp}$ é o único sub-espaço de

E que completa ortogonalmente E_1 em E e, P é um endomorfismo de E.

Dizemos que P é uma projeção ortogonal de E em E_1 se, e somente se,

P projeta vetores de E em E_1 segundo a direção dos vetores de $E_1^{\perp}$.

PROPRIEDADE 3.2 - P é uma projeção ortogonal de E no sub-espaço E_1 de E $\longleftrightarrow$ (I-P) é uma projeção ortogonal de E no sub-

$$\text{espaço } E_1^{\perp} \qquad (1.46)$$

PROVA - Imediata usando a Propriedade 3.1.

TEOREMA 3.8 - Seja E um espaço vetorial com produto interno e P um endomorfismo de E.

Então, P é uma projeção ortogonal de E se, e somente se, $P^2 = P$ e $P = P^t$, ou seja,

$$P \text{ é simétrica e Idempotente,} \qquad (1.47)$$

e ainda, se P é uma projeção ortogonal de E, então

$$\forall y \in E \ \|P(y)\| \leq \|y\|, \qquad (1.48)$$

onde $\| \ \|$ = norma do vetor.

PROVA - Supondo P uma projeção ortogonal de E, temos por (1.37) $P^2 = P$ e ainda, como $\forall y \in E$ temos

$$(P(y), y_2) = 0, \text{ para todo } y_2 \in (\mathrm{Im}(P))^{\perp}$$

então,

$$(P(y),y_2) = (y,P^t(y_2)) = o \ (\forall y \in E) \longleftrightarrow P^t(y_2) = \emptyset$$

para todo $y_2 \in (\text{Im } P)^\perp$

de modo análogo, como por (1.46) I-P é também uma projeção ortogonal de E sobre o Ker P = $(\text{Im } P)^\perp$ segundo a direção dos vetores da (Im P), temos

$$\forall y \in E \quad e \quad y_1 \in \text{Im}(P)$$

$$((I-P)(y),y_1) = (y,(I-P)^t(y_1)) = 0 \longleftrightarrow$$

$$(I-P)^t(y_1) = y_1 - P^t(y_1) = 0 \longleftrightarrow P^t(y_1) = y_1$$

logo, como para todo $y \in E \longrightarrow y = y_1 + y_2$ com $y_1 \in \text{Im}(P)$ e $y_2 \in (\text{Im}P)^\perp$ temos

$$P^t(y) = P^t(y_1+y_2) = P^t(y_1) + P^t(y_2) = P^t(y) = y_1 = P(y)$$

para todo $y \in E \longleftrightarrow P^t = P$,

inversamente, supondo $P^2 = P = P^t$ temos, $P^2 = P \longleftrightarrow$ por (1.31) P é uma projeção de E na Im(P) segundo a direção dos vetores do Ker(P).

Mostrando que KerP = $(\text{Im } P)^\perp$ temos: se

$$y \in \text{Im}(P) \quad e \quad x \in \text{Ker}(P)$$

então,

$$(x,y) = (x,P(y)) = (P^t(x),y) = (P(x),y) = (\emptyset,y) = 0$$

ou seja (KerP) = $(\text{Im } P)^\perp$, ainda,

Supondo P uma projeção ortogonal de E temos

$$\forall y \in E,\ 0 \leq \|P(y)\| = (P(y),P(y)) = (P(y),y)$$

e portanto por (1.14) P é positiva.

Como I - P é também uma projeção ortogonal de E então I-P é positiva e portanto

$$\|y\|^2 - \|P(y)\| = (y,y) - (P(y),P(y)) = (y,y) - (P(y),y) =$$

$$= ((I-P)(y),y) \geq 0 \longrightarrow \|P(y)\| \leq \|y\|.$$

COROLÁRIO 3.1 - Se E é um espaço vetorial com produto interno e P é uma projeção ortogonal de E então os auto valores de P são iguais a zero ou um e, os auto vetores de P associação a auto valores distintos são ortogonais. (1.49)

PROVA - Supondo P uma projeção ortogonal de E e $\lambda \in R$ temos que, se λ é tal que

$$Px = \lambda x \text{ onde } x \neq \emptyset,\ x \in E \longrightarrow P(P(x)) = P^2x = Px = \lambda x$$

e

$$P(P(x)) = P(\lambda x) = \lambda P(x) = \lambda^2 x \longleftrightarrow \lambda x = \lambda^2 x \longleftrightarrow$$

$$(\lambda - \lambda^2)x = \emptyset \longleftrightarrow \lambda - \lambda^2 = 0$$

(pois,

$$x \neq \emptyset) \longleftrightarrow \lambda = \lambda^2 \longleftrightarrow \lambda = 1 \text{ ou } \lambda = 0$$

e também, se

$$P(x_1) = \lambda_1 x_1 \text{ para algum } x_1 \in E \text{ e } x_1 \neq \emptyset$$

e

$$P(x_2) = \lambda_2 x_2 \text{ para algum } x_2 \in E \text{ e } x_2 \neq \emptyset$$

temos

$$\lambda_1(x_1,x_2) = (\lambda_1 x_1,x_2) = (P(x_1),x_2) = (x_1,P(x_2)) =$$

$$= (x_1,\lambda_2 x_2) = \lambda_2(x_1,x_2),$$

assim,

se $\lambda_1 \neq \lambda_2$ então $(x_1,x_2) = 0$ ou seja os auto vetores de P associado a auto valores distintos são ortogonais.

TEOREMA 3.9 - Seja E um espaço vetorial com produto interno. Se P é um endomorfismo de E tal que

$$P^2 = P \text{ e } \|P(y)\| \leq \|y\| \text{ para todo } y \in E \text{ então } P = P^t. \quad (1.50)$$

PROVA - Como $P^2 = P \longrightarrow$ (por 1.37) P é uma projeção de E na Im(P) segundo o Ker(P) basta então mostrarmos que

$$(\text{Im } P)^\perp = \text{Ker}(P) \text{ ou } (\text{Ker}(P))^\perp = \text{Im}(P).$$

Para este propósito, temos:

$$\forall x \in (\text{Ker } P)^\perp \text{ e } y = P(x) - x$$

temos que $x \in (\text{Ker } P)^\perp$ e $y \in \text{Ker}(P)$ assim

$$P(x) = x + y \text{ com } (x,y) = 0 \longrightarrow \|P(x)\|^2 =$$

$$= (P(x),P(x)) = (x+y,x+y) = (x,x) +$$

$$+ (y,y) = \|x\|^2 + \|y\|^2 \leq \|x\|^2 \longleftrightarrow \|y\| \leq$$

$$\leq 0 \longleftrightarrow y = \emptyset$$

ou seja $\|P(x)\| = \|x\| \longleftrightarrow x \in \text{Im}(P)$ e portanto $(\text{Ker}(P))^\perp \subset \text{Im}(P)$.

Como $\dim(\text{Im}(B)) = \dim(E) - \dim(\text{Ker}(B))$ e a

$$\dim(\mathrm{Ker}(P))^{\perp} = \dim(E) - \dim(\mathrm{Ker}(P))$$

temos $\dim(\mathrm{Im}(P)) = \dim(\mathrm{Ker}(P))$

e portanto $\mathrm{Im}(P) = (\mathrm{Ker}(P))^{\perp}$.

Observe que o Teorema 3.5 permanece válido se a palavra projeção é qualificada no contexto como ortogonal.

Isto é uma consequência imediata da caracterização precedente de projeções ortogonais e do fato que soma e diferença de transformações auto-adjuntas é auto-adjunta, enquanto o produto de duas transformações auto-adjunta é auto-adjunta se, e somente se elas comutam. (1.51)

Generalizando o teorema da soma para projeções ortogonais, temos

TEOREMA 3.10 - Seja E um espaço vetorial com produto interno. Seja ainda $P_1, \ldots, P_s$ projeções ortogonais de E.

Uma condição necessária e suficiente para que

$$P = \sum_{i=1}^{s} P_i \text{ seja uma projeção ortogonal de E}$$

é que $P_i P_j = \emptyset$, para $i \neq j$, i e $j = 1, \ldots, s$

ou seja, a imagem de P_i e P_j para $i \neq j$ são ortogonais (1.52)

PROVA - Supondo P_i, $i = 1, \ldots, s$ e $P = \sum_{i=1}^{s} P_i$ projeções ortogonais de E, temos por (1.48) para todo $x \in E$

$$\|P(x)\| \leq \|x\|$$

mas, como

$$\|P(x)\| \le \|x\| \longleftrightarrow \|P(x)\|^2 \le \|x\|^2 \longleftrightarrow$$

$$\|P(x)\|^2 = (P(x),P(x)) = (P^tP(x)),x) =$$

$$= (P^2(x),x) = (P(x),x) = \left(\sum_{i=1}^{s} P_i(x),x\right) =$$

$$= \sum_{i=1}^{s}(P_i(x),x) = \sum_{i=1}^{s}\|P_i(x)\|^2 \le \|x\|^2$$

então, se $x \in Im(P_i)$ para algum i, $i=1,\ldots,s$ temos

$P_i(x) = x$ e portanto

$$\sum_{\substack{j=1\\ j\neq i}}^{s} \|P_j(x)\|^2 \le 0 \longleftrightarrow$$

$$\|P_j(x)\|^2 = 0 \text{ para todo } j,\ j\neq i,\ i \text{ e } j=1,\ldots,s \longleftrightarrow$$

$$P_j(x) = 0 \longleftrightarrow x \in Ker(P_j) = (Im(P_j))^{\perp}$$

para todo j, $j\neq i$, i e $j=1,\ldots,s$

ou seja, $Im(P_i) \subset (Im(P_j))^{\perp}$ para todo j, $j\neq i$, i e $j=1,\ldots,s$

reciprocamente,

é imediata usando indução finita sobre $s \ge 2$ e o teorema 3.5 para projeções ortogonais.

Finalizando nosso estudo de projeções temos,

TEOREMA 3.11 - Se P_1 e P_2 são projeções ortogonais de E, então as seguintes condições são mutuamente equivalentes

a) $\|P_1(x)\| \le \|P_2(x)\|$ para todo $x \in E$

b) $Im(P_1) \subset Im(P_2)$

c) $P_2P_1 = P_1$

d) $P_1P_2 = P_1$.

PROVA - Supondo, P_1 e P_2 projeções ortogonais de E temos

pelo Teorema 3.8 $P_1 = P_1^t$ e $P_2 = P_2^t$

pelo Teorema 3.5 $Im(P_1) \subset Im(P_2) \longleftrightarrow P_2P_1 = P_1$ e portanto, é trivial que

$$Im(P_1) \subset Im(P_2) \longleftrightarrow P_2P_1 = P_1 \longleftrightarrow P_1P_2 = P_1$$

mostremos que a) $\longleftrightarrow$ b).

Supondo que $\|P_1(x)\| \leq \|P_2(x)\|$ para todo $x \in E$, então, se

$$x \in Ker(P_2) \longrightarrow P_2(x) = \emptyset$$

$$\longrightarrow \|P_2(x)\| = 0 \geq \|P_1(x)\| \longrightarrow \|P_1(x)\| = 0 \longrightarrow$$

$$\longrightarrow P_1(x) = 0 \longrightarrow x \in (P_1) \text{ ou seja } Ker(P_2) \subset Ker(P_1)$$

e portanto

$$Im(P_1) = (Ker(P_1))^{\perp} \subset (Ker(P_2))^{\perp} = Im(P_2)$$

supondo que $Im(P_1) \subset Im(P_2)$,

então, por a) b) e c)

$$P_2P_1 = P_1P_2 = P_1 \longleftrightarrow \text{(por 1.42)}$$

$P_2 - P_1$ é uma projeção e também satisfaz por hipótese

$$(P_2 - P_1)^t = P_2^t - P_1^t = P_2 - P_1$$

assim, para todo $y \in E$

$$\|P_2(y)\|^2 - \|P_1(x)\|^2 = (P_2(y), P_2(y)) - (P_1(y), P_1(y)) =$$

$$= (P_2(y), y) - (P_1(y), y) = ((P_2 - P_1)(y), y) =$$

$$= ((P_2 - P_1)(y), (P_2 - P_1)(y)) = \|(P_2 - P_1)(y)\|^2 \geq 0 \longrightarrow$$

$$\|P_2(y)\|^2 \geq \|P_1(y)\|^2 \text{ para todo } y \in E,$$

ou seja

$$\|P_2(y)\| \geq \|P_1(y)\| \text{ para todo } y \in E.$$

1.4 - PROJETOR ORTOGONAL ASSOCIADO A UMA MATRIZ

DEFINIÇÃO 4.1 - Considerando uma matriz $X_{n \times p}$; $X = (x_{(j)})$, onde $x_{(j)}$, $j = 1, \ldots, p$, são vetores $n \times 1 \in R^n$, e

$P: R^n \longrightarrow R^n$ um operador projeção ortogonal de R^n.

Então, se

$Im(P) = C(X)$ dizemos que P é uma projeção ortogonal de R^n sobre o espaço coluna de X .

DEFINIÇÃO 4.2 - Seja X uma matriz $n \times p$ e P um operador projeção ortogonal de R^n sobre $C(X)$.

Define-se $\cos(y, C(x))$ para todo $y \in R^n$ ao cosseno do ângulo entre o vetor y e a sua projeção ortogonal sobre $C(X)$, ou seja

$$\cos(y, C(X)) = \cos(y, P(y)).$$

Uma forma matricial para apresentar o operador projeção ortogonal sobre $C(X)$ é dada por:

TEOREMA 4.1 - Para toda matriz X $n \times p$ de posto $r \leq \min(n,p)$ existe uma matriz B $p \times n$ tal que

$$X^t XB = X^t$$

e, se

$$P_x = XB$$

então

P_X representa a matriz do operador projeção ortogonal sobre $C(X)$ e ainda, $P_x = XB$ é única (1.53)

PROVA - Como, por (1.35) temos

$C(X^tX) = C(X^t)$ então, para todo $y \in R^n$,
existe $b \in R^p$ tal que $X^tXb = X^ty$.

Fazendo y sucessivamente igual a

$$y_j = (d_{ij}),\ j=1,\ldots,n,\ \text{onde } d_{ij} = \begin{cases} 1 & \text{se } i=j \\ 0 & \text{se } i \neq j \end{cases}$$

temos que existe b_j, $j=1,\ldots,n, \in R^p$ tal que

$$X^t y_j = X^t y_j = x^t_{(j)}$$

onde $x_{(j)}$ = j-ésima lina de x, assim, tomando $B = (b_1 : \ldots : b_n)$ temos que existe uma matriz B $p \times n$ tal que

$$X^t XB = X^t.$$

Fazendo $P_x = XB$ temos que P_x é a matriz do operador projeção ortogonal sobre C(X), pois,

a) $P_x^t = P_x$ pois,

$$X^tXB = X^t \longleftrightarrow B^tX^tX = X$$

assim,

$$P_x = XB = B^tX^tXB = B^tX^t = (XB)^t = P_x^tX$$

b) $P_x^2 = P_x$, pois

$$(P_x)^2 = (XB)(XB) = (XB)^t(XB) = B^tX^tXB = B^tX^t = P_x$$

c) posto(P_x) = posto X, pois

como

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

$$\leq \text{posto}(XB) \leq \text{posto}(X),$$

então

$$\text{posto}(BX) = \text{posto}(X)$$

e como $C(P_x) \subset C(X)$ concluímos que $C(P_x) = C(X)$ e portanto, por a), b) e c) concluímos que $P_x = XB$ é a matriz do operador projeção ortogonal sobre C(X).

Supondo que exista uma outra matriz $P_x^* = XB_1$ onde B_1 é uma matriz p×n tal que $X^tXB_1 = X^t$, temos então que, P_x^* é também uma matriz do operador projeção ortogonal sobre C(X). Temos para todo $y \in R^n$

$$X^t P_X^* y = X^t y = X^t P_X y \longleftrightarrow$$

$$\longleftrightarrow X^t(P_X^* y - P_X y) = \emptyset \longleftrightarrow$$

$$\longleftrightarrow P_X^* y - P_X y \in Ker(X^t) = C(X)^{\perp}$$

e como

$$P_X^* y - P_X y = XB_1 y = X(B_1 y - By) \in C(X)$$

temos

$$P_X^* y - P_X y = \emptyset \longrightarrow P_X^* y = P_X y \text{ para todo } y \in R^n$$

então, concluímos que $P_X^* = P_X$.

DEFINIÇÃO 4.3 - Seja X uma matriz $n \times p$.

Define-se matriz do operador projeção ortogonal sobre C(X) à matriz

$P_X = XB$ onde B é uma matriz $p \times n$ tal que

$$X^t XB = X^t.$$

Como, a relação entre X e B é tal que

$$(XB)^t = XB \text{ e}$$

$$XBX = (XB)^t X = B^t X^t X = X$$

então, concluímos que B é uma inversa generalizada de X de mínimos quadrados (ver Rao & Mitra (1969). (1.54)

Assim, um aspecto importante da matriz do operador projeção ortogonal sobre C(X) é que ela é obtida da aproximação de mínimos quadrados e portanto aproxima o vetor $y \in R^n$ por vetores pertencentes ao sub-espaço $C(X)$ de R^n. Tal fato, será estudado com detalhes no Capítulo 2.

Algumas das propriedades a respeito de $P_x = XB$ são conseqüências imediatas das propriedades dos operadores projeção ortogonal e portanto faremos apenas citá-las.

PROPRIEDADES DE P_x - Se X é uma matriz $n \times p$ de posto $r \leq \min(n,p)$ e $P_x = XB$ é a matriz do operador projeção ortogonal sobre $C(X)$ então, P_x satifaz as seguintes propriedades:

I) $I - P_x$ é a matriz do operador projeção ortogonal sobre $C(X)^\perp$ (1.55)

II) Para todo vetor $y \in R^n$

$P_x y$ é a projeção ortogonal de y em $C(X)$ (1.56)

e

$(I-P_x)y$ é a projeção ortogonal de y em $\overset{\perp}{C}(X)$ (1.57)

III) Se A é uma matriz $n \times r$ de posto r tal que $C(A) = C(X)$ então

$$P_A = P_x \tag{1.58}$$

PROVA - Supondo $C(X) = C(A)$ onde A é uma matriz $n \times r$ de posto r temos

$$C(X) = C(A) \Longrightarrow C(X)^\perp = C(A)^\perp$$

e como para todo vetor $y \in R^n$, y é univocamente determinado por

$$y = P_x y + (I-P_n)y = P_A y + (I-P_A)y$$

(usando (1.56) e (1.57))

onde $P_x y \in C(X)$ e $P_A y \in C(A)$

e $(I-P_x)y \in \overset{\perp}{C}(X)$ e $(I-P_A)y \in \overset{\perp}{C}(A)$

então por hipótese vale a igualdade acima apenas quando

$P_x y = P_A y$ para todo $y \in R^n$ ou seja $P_x = P_A$.

IV) Se as colunas da matriz A formam uma base ortonormal de C(X) então, $P_x = AA^t$ (1.59)

PROVA - Supondo que as colunas da matriz A formam uma base ortogonal de C(X) temos

$$\therefore \quad A^t A = I_r \text{ e } C(A) = C(X).$$

Como por definição $P_A = A\mathbb{C}$ é a matriz do operador projeção ortogonal sobre C(A) apenas quando $\mathbb{C}$ satisfaz $A^t A\mathbb{C} = A^t$, então, se P_A é a matriz do operador projeção ortogonal sobre C(A) temos que

$$P_A = AA^t$$

e assim, como C(X) = C(A) onde

$$P_A = AA^t \longrightarrow \text{(por 1.58)}$$

$$P_x = AA^t.$$

V) Os auto-valores de P_x são iguais a zero ou um. (1.60)

Estudemos agora as partições de C(X) como fonte geradora de sub-espaços que decompõem este $C(X)$ em soma direta de sub-espaços ortogonais.

PROPOSIÇÃO 4.1 - Se X é uma matriz $n \times p$ e X_i, $i=1,\ldots,s$, são matrizes $n \times p_i$ respectivamente onde $p = \sum_{i=1}^{s} p_i$, então,

$$X = (X_1 : X_2 : \ldots : X_s) \leftrightarrow C(X) = \sum_{i=1}^{s} C(X_i) \qquad (1.61)$$

PROVA - Supondo $X = (X_1 : \ldots : X_s)$ temos, para todo $\lambda \in R^n$ tal que $\lambda \in C^{\perp}(X)$ então,

$$\lambda^t = \lambda^t (X_1 : \ldots : X_s) = \emptyset \longleftrightarrow$$

$$\longleftrightarrow (\lambda^t X_1 : \ldots \lambda^t X_2 : \ldots \lambda^t X_s) = \emptyset \longleftrightarrow$$

$$\longleftrightarrow \lambda^t X_i = \emptyset \text{ para todo } i=1,\ldots,s \longleftrightarrow$$

$$\longleftrightarrow \lambda \in \bigcap_{i=1}^{s} C(X_i)^{\perp} \longleftrightarrow \text{(por (1.31))}$$

$$\lambda \in \left(\sum_{i=1}^{s} C(X_i) \right)^{\perp}$$

e portanto $\sum_{i=1}^{s} C(X_i) \subset C(X)$

Como

$$\dim \left(\sum_{i=1}^{s} C(X_i) \right) = \sum_{i=1}^{s} \dim C(X_i) - \sum_{i=1}^{s-1} \sum_{j=i+1}^{s} \dim(C(X_i) \cap C(X_j)) =$$

$$= \sum_{i=1}^{s} \text{posto}(X_i) - \sum_{i=1}^{s} \sum_{j=i+1}^{s} \dim(C(X_i) \cap C(X_j)) =$$

$$= \text{posto}(X) = \dim(C(X))$$

concluímos que $C(X) = \sum_{i=1}^{s} C(X_i)$ reciprocamente é trivial que $X = (X_1 : \ldots : X_s)$ onde X_i são matrizes $n \times p_i$ com

$$p = \sum_{i=1}^{s} p_i .$$

TEOREMA 4.2 - Se $X = (X_1 : X_2)$ onde X_i, $i=1,2$, são matrizes $n \times p_i$ respectivamente, com $p_1 + p_2 = p$ então, se

P_{12} é o operador projeção ortogonal de R^n sobre $C(X)$

e

P_1 é o operador projeção ortogonal de R^n sobre $C(X_1)$

temos que

$$C(X) = Im(P_1) \overset{\perp}{\oplus} Im(P_{12}-P_1)$$

onde

$$Im(P_{12}-P_1) = Im(P_{12}) \cap Ker(P_1)$$

e

$$Ker(P_{12}-P_1) = Im(I-P_{12}) \overset{\perp}{\oplus} Im(P_1). \qquad (1.62)$$

PROVA - Supondo $X = (X_1 : X_2)$ e

P_{12} o operador projeção ortogonal de R^n sobre $C(X)$

e

P_1 o operador projeção ortogonal de R^n sobre $C(X1)$

temos

Como $X = (X_1 : X_2) \longrightarrow C(X_1) \subset C(X) \longrightarrow Im(P_1) \subset Im(P_{12}) \longrightarrow$

$\longrightarrow$ (teor 3.11)

$$P_{12}P_1 = P_1P_{12} = P_1 \longrightarrow \text{(extensão de (1.42)}$$

para projeções ortogonais)

$P_{12}-P_1$ é uma projeção ortogonal onde

$$Im(p_{12}-P_1) = Im(P_{12}) \cap Ker(P_1)$$

e

$$Ker(P_{12}-P_1) = Im(I-P_{12}) \overset{\perp}{\oplus} Im(P_1)$$

e ainda, como

$$P_{12}P_1 = P_1P_{12} = P_1 \longleftrightarrow$$

$$\longleftrightarrow (I-P_{12})P_1 = P_1(I-P_{12}) = \emptyset \longrightarrow \text{(extensão de}$$

(1.41) para projeções ortogonais)

$P_1+(P_{12}-P_1) = P_{12}$ é uma projeção ortogonal sobre a

$$Im(P_1+(P_{12}-P_1)) = Im(P_1) \overset{\perp}{\oplus} Im(P_{12}-P_1)$$

e como

$$Im(P_1+(P_{12}-P_1)) = Im(P_{12}) = C(X)$$

então concluímos que

$$C(X) = Im(P_1) \overset{\perp}{\oplus} Im(P_{12}-P_1)$$

onde

$$Im(P_{12}-P_1) = Im(P_{12}) \cap Ker(P_1).$$

Observe que tal decomposição pode ser estendida para uma partição de C(X) em mais de dois sub-espaços de C(X) como por exemplo:

PROPOSIÇÃO 4.2 - Se $X = (X_1 : X_2 : X_3)$, onde X_i, i=1,2,3, são matrizes $n \times p_i$ respectivamente, com

$$\sum_{i=1}^{3} p_i = p$$

então, se

P_{123} é o operador projeção ortogonal sobre C(X)

P_{12} é o operador projeção ortogonal sobre $C(X_1 : X_2)$

P_1 é operador projeção ortogonal sobre $C(X_1)$,

temos que

$$C(X) = Im(P_1) \overset{\perp}{\oplus} Im(P_{12}-P_1) \overset{\perp}{\oplus} Im(P_{123}-P_{12})$$

onde

$$Im(P_{12}-P_1) = Im(P_{12}) \cap Ker(P_1)$$

$$Im(P_{123}-P_{12}) = Im(P_{123}) \cap Ker(P_{12}). \qquad (1.63)$$

PROVA - Análoga ao Teorema 4.2.

Da mesma forma que existe $B_{p \times n}$ tal que $X^t = XB = X^t$ podemos demonstrar que existe uma matriz $C_{p \times n}$ tal que

$$XX^tC^t = X$$

e

$P_{X^t} = X^t C^t$ é a única matriz do operador projeção ortogonal sobre $C(X^t) = L(X)$. (1.64)

DEFINIÇÃO 4.4 - Seja X uma matriz $n \times p$.

Define-se matriz do operador projeção ortogonal sobre $C(X^t) = L(X)$ = espaço linha de X à matriz $P_{X^t} = X^t C^t$, onde C é uma matriz $p \times n$ tal que $XX^t C^t = X$. (1.65)

Considerando, a matriz $G = CXB$, onde B e C são matrizes $p \times n$ satisfazendo $X^t XB = X^t$ e $XX^t C^t = X$, respectivamente (1.66) temos

PROPOSIÇÃO 4.3 - A matriz G definida em (1.65) é única para qualquer escolha de B e C. (1.67)

PROVA - Supondo B_1 e B_2 quaisquer duas matrizes tais que

$$X^t XB_1 = X^t \text{ e } X^t XB_2 = X^t$$

a C_1 e C_2 quaisquer duas matrizes tais que

$$XX^t C_1^t = X \text{ e } XX^t C_1^t = X$$

temos por (1.53) e (1.64) que

$$XB_1 = XB_2 \text{ e } C_1 X = C_2 X$$

assim, se $G_{ij} = C_i XB_j$ para $i,j = 1,2$, então

$$C_1 XB_1 = C_2 XB_1 = C_2 XB_2 = C_1 XB_2$$

ou seja,

$G_{ij} = G_{i'j'}$ para todo $i,j,i',j' = 1,2$ e assim prova-se que G é única. Mas, ainda G satisfaz as seguintes propriedades:

I) $XGX = X$ pois

como para toda matriz C tal que

$$XX^tC^t = X \tag{1.68}$$

e para toda matriz B tal que

$$X^tXB = X^t \tag{1.69}$$

então se $G = C\,B$ como em (1.66) temos

$$XGX = XCXBX = \text{por } (1.68)\ XBX = \text{por } (1.69)\ X.$$

II) $GXG = G$ pois

$$G = CXB \longrightarrow CXBXCXB = (1.69)\ CXCXB = (1.68)\ CXB = G$$

III) $(XG)^t = XG$ pois

$$(XG)^t = G^tX^t = B^tX^tC^tX^t = B^tX^t = XB = XCXB = XG$$

IV) $(GX)^t = GX$, pois

$$(GX)^t = X^tG^t = X^tB^tX^tC^t = X^tC^t = CX = CXBX = GX$$

que são as condições para a única inversa de Moore-Penrose (ver Rao & Mitra (1971)).

A inversa de Moore-Penrose é importante na solução de sistema lineares consistentes, como veremos:

PROPRIEDADE 4.1 - Seja X uma matriz $n \times p$ se, $Xy = c$ é um sistema linear consistente então a solução geral de $Xy = c$ é dado por

$$y = X^{+}c + (I_p - X^{+}X)z \text{ para todo vetor } z \in R^p, \text{ e onde}$$

$$X^{+} = CXB \text{ é a única inversa de Moore-Penrose.}$$

PROVA - Sabe-se que a solução geral do sistema $Xy = c$ consiste de uma solução particular arbitrária combinada com cada solução de $Xy = \emptyset$ que são os vetores pertencentes ao núcleo de X. Assim, se posto X = r então, posto $(I_p - X^{+}X) = p - r$ e portanto o espaço coluna de $(I_p - X^{+}X)$ é de dimensão p-r, e assim como

$$x(I_p - X^{+}X) = \emptyset \text{ e a } \dim(C(I_p - X^{+}X)) = \dim N(X)$$

então, podemos concluir que

$$C(I_p - X^{+}X) = N(X)$$

e portanto, a solução geral de $Xy = c$ é

$$y = X^{+}c + (I_p - X^{+}X)z \text{ para qualquer vetor } z \in R^p.$$

Um outro aspecto de grande importância de X^{+} é dado por (Rao & Mitra (1971) onde eles mostram que $X^{+} = CXB$ é a solução de Mínima Norma do sistema de equações consistente $Xy = c$.

000o000

CAPÍTULO 2
CARACTERIZAÇAO DO BLUE
NO MODELO DE GAUSS-MARKOFF

2.1 - APROXIMAÇÃO DOS MÍNIMOS QUADRADOS - Seja

$y = (y_i)$, $i=1 \ldots n$, um vetor de R^n

$X = (x_{(j)})$ uma matriz $n\chi p$ conhecida de posto $r < p$

onde

$x_{(j)}$, $j=1, \ldots, p$, são vetor $n\times 1$ de R^n

e

$\beta = (\beta_j)$, $j=1, \ldots, p$, um vetor de R^p

Considerando o modelo

$y = X\beta + e$ onde e é um vetor $n\times 1$ de perturbação ou erro (2.1)

temos

DEFINIÇAO 1.1 - A aproximação dos mínimos quadrados de y por vetores de $C(X)$ consiste em achar um vetor $\eta \in C(X)$ tal que

$$\min_{\eta \in C(X)} ||y - \hat{\eta}||^2 = ||y - \hat{\eta}||^2 \tag{2.2}$$

Tal aproximação é dada por

TEOREMA 1.1 - Considerando uma matriz $X_{nxp}=(x_{(j)})$ onde, $x_{(j)}$, $j=1, \ldots, p$, são vetores de R^n.

e

P_X uma matriz do operador projeção ortogonal de R^n sobre $C(X)$

Então, para todo $y \in R^n$

$$\min_{\eta \in C(X)} ||y - \eta||^2 = y^t (I - P_X) y$$

e este mínimo é atingido quando

$$\hat{\eta} = P_X y = \text{projeção ortogonal de } y \text{ sobre } C(X) \tag{2.3}$$

PROVA - Por Cauchy-Schawarz, para qualquer vetor $c \in R^n$ e $\eta \in C(X)$

$$||y - \eta||^2 \geqslant \frac{[c^t (y - \eta)]^2}{c^t c} \tag{2.4}$$

Fazendo $c = (I - P_X) y$ onde

P_X é uma matriz do operador projeção ortogonal de R^n sobre $C(X)$ temos, por (1.55) que

$I - P_X$ é uma matriz do operador projeção ortogonal de R^n sobre $C(X)^{\perp}$

e portanto

$$c^t c = y^t (I - P_X)^t (I - P_X) y = y^t (I - P_X) y \tag{2.5}$$

e

$$c^t (y - \eta) = y^t (I - P_X) (y - \eta) =$$

$$= y^t [(I - P_X) y - (I - P_X) \eta] =$$

$$= y^t \, (I - P_X) \, y \qquad (2.6)$$

Substituindo (2.5) e (2.6) em (2.4)

(2.4) fica

$$\|y - \eta\|^2 \geqslant \frac{(y^t \, (I - P_X) \, y)^2}{y^t \, (I - P_X) \, y} = y^t \, (I - P_X) \, y \qquad (2.7)$$

Fazendo $\hat{\eta} = P_X \, y$ = projeção ortogonal de

y sobre $C(X)$ então

$$\|y - \hat{\eta}\|^2 = \|y - P_X \, y\|^2 = \|(I - P) \, y\|^2 =$$

$$= ((I-P_X) \, y, \, (I-P_X) \, y) = (y, (I-P_X) \, y) =$$

$$= y^t \, (I - P_X) \, y \qquad (2.8)$$

assim, por (2.7) e (2.8)

$$\min_{\eta \in C(X)} \|y - \eta\|^2 = \|y - \hat{\eta}\| = y^t \, (I - P_X) \, y$$

onde

$\hat{\eta} = P_X \, y$ = projeção ortogonal de

y sobre $C(X)$

Apesar deste resultado ser matematicamente elegante, podemos notar que esta formulação independe da base gerada pelos vetores coluna de X, portanto, despreza aspectos do problema nos quais envolve os atuais vetores $x_{(j)}$, $j = 1, \ldots, p$, que são de grande importância, como por exemplo, nos problemas de regressão.

Considerando, a aproximação dos mínimos quadrados dos vetores $y \in R^n$ por vetores de $C(X)$ ao vetor

$\hat{y} = X\hat{\beta} \in C(X)$ temos

DEFINIÇAO 2.2 - Seja o modelo $y = X\beta + e$ dado por (2.1)

A aproximação dos mínimos quadrados de y por uma combinação linear das colunas da matriz X - $\sum_{j=1}^{p} x_{(j)} \beta_j$ - consiste em achar aqueles coeficientes $\hat{\beta}_1, \ldots, \hat{\beta}_p$ os quais minimizam a soma dos quadrados dos desvios de y pela sua aproximação ou seja, os $\hat{\beta}_1, \ldots, \hat{\beta}_p$ escolhidos de tal forma que

$$\min_{\beta_1,\ldots,\beta_p} \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij} \beta_j)^2 =$$

$$= \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} x_{ij} \hat{\beta}_j)^2 \tag{2.9}$$

Em forma matricial temos,

Se $\min_{\eta \in C(X)} ||y - \eta||^2 = \min_{\beta \in R^p} ||y - X\beta||^2 = ||y - X\hat{\beta}||^2$

então a aproximação dos mínimos quadrados de y por vetores de $C(X)$ é dado por $X\hat{\beta} = \hat{y}$ (2.10)

2.2 - COMPUTAÇÃO DOS COEFICIENTES DE MÍNIMOS QUADRADOS E EQUAÇÕES NORMAIS

Considerando a função

$Q(\beta) = (y - X\beta)^t (y - X\beta)$ onde

y, X e β são dados pelo modelo (2.1) temos que

$$Q(\beta) = \sum_{i=j}^{n} (y_i - \sum_{j=1}^{p} x_{ij} \beta_j)^2 \geqslant 0$$

e é contínua para todo vetor $\beta \in R^p$.

Assim, derivando $Q(\beta)$ com relação a β_k, $k = 1, \ldots, p$, e, igualando a zero cada uma dessas derivadas parciais temos

$$\frac{\partial Q(\beta)}{\partial \beta_k} = -2 \sum_{i=j}^{n} x_{ik} (y_i - \sum_{j=1}^{p} x_{ij} \beta_j) = 0 \quad \leftrightarrow$$

$$\sum_{i=j}^{n} x_{ij} y_i = \sum_{i=1}^{n} x_{iK} (\sum_{j=1}^{p} x_{ij} \beta_j) \quad \leftrightarrow$$

$$X^t_{(k)} y = X^t_{(k)} X \beta$$

onde $X_{(k)}$ = k-ésima coluna de X, ($k = 1, \ldots, p$).

Em forma matricial, se

$$\frac{\partial (Q(\beta))}{\partial \beta} = \emptyset$$ então β satisfaz o sistema de e-quações $X^t X \beta = X^t y$

que é denominado de equações normais que é sempre consistente desde que por (1.35) temos $C(X^t X) = C(X^t)$.

Ainda, como $\forall \beta \in R^p$ e $\hat{\beta} \in R^p$ tal que $X^t X\hat{\beta} = X^t y \rightarrow$

$$Q(\beta) = (y-X\beta)^t (y-X\beta) = (y-X\hat{\beta} + X(\hat{\beta}-\beta))^t (y-X\beta + X(\hat{\beta}-\beta)) =$$

$$= (y-X\hat{\beta})^t (y-X\beta) + (\hat{\beta}-\beta)^t X^t X(\hat{\beta}-\beta) \geqslant Q(\hat{\beta})$$

concluímos que $Q(\beta)$ admite um mínimo em β dado por $\hat{\beta}$ para todo $\hat{\beta} \in R^p$ tal que satisfaz as equações normais $X^t X\beta = X^t y$

e também, se $\hat{\beta}_1$ e $\hat{\beta}_2$ são dois vetores de R^p tais que

$$X^t X \hat{\beta}_1 = X^t y \quad e \quad X^t X \hat{\beta}_2 = X^t y$$

então

$$X^t X \hat{\beta}_1 = X^t X \hat{\beta}_2 = X^t y \leftrightarrow$$

$$X^t X (\hat{\beta}_1 - \hat{\beta}_2) = \emptyset \rightarrow (\hat{\beta}_1 - \hat{\beta}_2)^t X^t X (\hat{\beta}_1 - \hat{\beta}_2) = 0 \leftrightarrow$$

$$(X(\hat{\beta}_1 - \hat{\beta}_2))^t (X(\beta_1 - \beta_2)) = 0 \leftrightarrow X(\hat{\beta}_1 - \hat{\beta}_2) = \emptyset \leftrightarrow$$

$$X \hat{\beta}_1 = X \hat{\beta}_2$$

ou seja, $Q(\beta)$ é invariante para toda solução das equações normais $X^t X\beta = X^t y$ (2.12)

Reciprocamente, supondo que $Xt = \tilde{y}$ onde $t = \hat{\beta} + \rho$ com $\rho \in R^p$ e $\hat{\beta} \in R^p$ qualquer solução das equações normais seja outra aproximação do vetor y então,

$$Q(t) = (y - X(\hat{\beta}+\rho))^t (y - X(\hat{\beta}+\rho)) =$$

$$= [(y - X\hat{\beta}) - X\rho]^t [(y - X\hat{\beta}) - X\rho] =$$

$$= (y - X\hat{\beta})^t (y - X\hat{\beta}) - 2 (y - X\hat{\beta})^t X\rho + \rho^t X^t X\rho =$$

$$= (y - X\hat{\beta})^t (y - X\hat{\beta}) + \rho^t X^t X \rho \text{ pois } (y - X\hat{\beta})^t X = \emptyset$$

e assim, $Q(t) \geq Q(\hat{\beta})$

onde

$$Q(t) = Q(\hat{\beta}) \leftrightarrow \rho^t X^t X \rho = 0 \leftrightarrow$$
$$X\rho = \emptyset$$

e neste caso

$$\tilde{y} = X\, t = X\, \hat{\beta} = \hat{y} \tag{2.13}$$

Assim, por (2.11), (2.12) e (2.13) podemos concluir que

TEOREMA 2.1 - Considerando a função $Q(\beta) = (y - X\beta)^t\, (y - X\, \beta)$ onde y, X e β são dados pelo modelo (2.1)
então

$$\min_{\beta \in R^p} \; Q(\beta) = Q(\hat{\beta}) \leftrightarrow \hat{\beta}$$

é solução das equações normais $X^t\, X\, \beta = X^t\, y$ que são sempre consistentes e ainda, $X\, \hat{\beta}$ é o único vetor em $C(X)$ tal que

$$\min_{\beta \in R^p} \; Q(\beta) = Q(\hat{\beta}) \tag{2.14}$$

COROLÁRIO 2.1 - Quando a matriz X_{nxp} do modelo linear (2.1) é de posto p, então, o vetor invariante $X\, \hat{\beta}$ é explicitamente dado por um conjunto de funções lineares das coordenadas do vetor y da forma

$$X\, \hat{\beta} = X\, (X^t X)^{-1}\, X^t y \tag{2.15}$$

PROVA - Trivial

Expressões análogas para o vetor $X\, \hat{\beta}$ no caso de posto $X = r < p$ poderão ser obtidas por:

PROPOSIÇÃO 2.1 - Considerando a matriz X_{nxp} do modelo linear (2.1) e,

$P_x = XB$ a matriz do operador projeção ortogonal sobre $C(X)$ temos então que, By é solução das equações normais e portanto,

$$X \hat{\beta} = \underbrace{X \hat{B} y}_{P_v y} \quad \text{para todo } \hat{\beta} \in R^p \text{ tal que}$$

$$X^t X \hat{\beta} = X^t y \qquad (2.16)$$

PROVA - Como por definição $P_x = XB$ é a matriz do operador projeção ortogonal sobre $C(X)$ se e somente se X é uma matriz pxn tal que

$$X^t X B = = X^t$$

então, para todo $y \in R^n$ temos

$$X^t X . By = X^t y \text{ onde } By \in R^p \qquad (2.17)$$

Observando que as estruturas das equações (2.17) coincidem com a estrutura das equações normais, temos que, se

$$y = X \beta + e \quad \text{como em } (2.1)$$

então $$X^t X B y = X^t y \rightarrow$$

By é solução das equações normais e por (2.12)

$$X B y = X \hat{\beta} \text{ para todo vetor}$$

$$\hat{\beta} \in R^p \quad \text{tal que } X^t X \hat{\beta} = X^t y$$

PROPOSIÇÃO 2.2 - A aproximação dos mínimos quadrados $\hat{y} = X \hat{\beta}$ para o vetor de dados y é a projeção ortogonal de y sobre $C(X)$.

O vetor de resíduos é a projeção ortogonal de y sobre o complemento ortogonal de $C(X)$.

PROVA - por (2.16) $X\hat{\beta} = X B y = P_X y$ onde P_X é a matriz do operador projeção ortogonal sobre $C(X)$

e

$\hat{e} = y - X\hat{\beta} = (I - P_X)\, y$ onde $I - P_X$ é a matriz do operador projeção ortogonal sobre $C(X)^{\perp}$

Geometricamente, falando a linguagem de espaços vetoriais, começa-se com um vetor $y \in R^n$ e p vetores $x_{(1)}, \dots, x_{(p)} \in R^p$ tal que $C(X) = C(x_{(1)} : \dots : x_{(p)})$.

O problema é então escolher um vetor $\hat{y} = X\hat{\beta} \in C(X)$ tal que $\hat{y}$ esteja o mais próximo de y. Note que a distância mínima requerida é a Euclidiana de um vetor em relação a um sub-espaço. Portanto, $X\hat{\beta}$ é a projeção ortogonal de y em $C(X)$.

Considerando

$$X\hat{\beta} = P_X y \qquad \text{e}$$

$$y - X\hat{\beta} = (I - P_X)\, y$$

como anteriormente obviamente temos que $(X\hat{\beta})^t (y - X\hat{\beta}) = 0$ e assim, como para todo vetor $y \in R^n$

$$y = P_X y + (I - P_X)\, y = X\hat{\beta} + (y - X\hat{\beta})$$

então

$$||y||^2 = y^t y = ||X\hat{\beta}||^2 + ||y - X\hat{\beta}||^2$$

que é o teorema de Pitágoras.

Esta decomposição é importante em prover uma medida de ajustamento de $X\hat{\beta}$ para y.

Evidentemente o ajuste é "melhor" quando os desvios $y - X\hat{\beta}$ são pequenos, isto é, quando a perpendicular a $C(X)$, $y - X\hat{\beta}$, é pequena.

Para uma medida absoluta do ajustamento faz-se uma relação entre o comprimento desta perpendicular e o comprimento de y.

Uma medida óbvia é:

$$\frac{||x\hat{\beta}||^2}{||y||^2} = 1 - \frac{||y - x\beta||^2}{||y||^2} = \cos (y, C(X))$$

(ver 1.4) o qual está sempre entre zero e um.

Evidentemente quando o ângulo entre y e $C(X)$ é pequeno, o cosseno é grande e assim y está bem ajustado por $P_X y = X\hat{\beta}$

Nos casos extremos, quando o ângulo entre y e $C(X)$ é zero significa que $y \in C(X)$ e assim y é perfeitamente ajustado por $P_X y = y$; por outro lado um ângulo reto significa que y é perpendicular a $C(X)$ e a aproximação trivial e $X\hat{\beta} = \emptyset$ e neste caso a aproximação é sem utilidade.

2.3 - ESTIMABILIDADE

Quais combinações lineares dos parâmetros β, $\lambda^t\beta$, são determináveis através de combinações lineares dos dados?

Para responder esta questão, defina-se

DEFINIÇÃO 3.1 - A função linear paramétrica $\lambda^t\ \beta$ onde λ é um vetor de R^p é linearmente estimável se, e somente se, existe um vetor $a \in R^n$ tal que

$$E\ (a^t y) = \lambda^t\ \beta \text{ para todo } \beta \in R^p \qquad (2.18)$$

Assim, dizemos que $\lambda^t\ \beta$ é linearmente estimável se existir um vetor $a \in R^n$ tal que a esperança de $a^t y$ é igual a $\lambda^t\ \beta$ independente do valor do parâmetro β .

Se tal uma função não existir então $\lambda^t\ \beta$ é dita ser não estimável.

Supondo que no modelo linear $y = X\ \beta + e$ dado por (2.1) temos $E(e) = \emptyset$ (ou seja, a média dos erros é nula) então,

PROPOSIÇAO 3.1 - $\lambda^t\ \beta$ é linearmente estimável se, e somente se $\lambda^t = a^t\ X$ para algum vetor $a \in R^n$. (2.19)

PROVA - Supondo $\lambda^t\beta$ uma função linear linearmente estimável então por (2.18) existe $a \in R^n$ tal que

$$E(a^t\ y) = \lambda^t\ \beta \text{ para todo } \beta \in R^p$$

mas

$$E(a^t\ y) = E(a^t(X\ \beta + e)) =$$

$$= E(a^t\ X\ \beta + a^t\ e) = a^t\ X\ \beta + a^t\ E(e) = a^t X\beta = \lambda^t\ \beta$$

para todo $\beta \in R^p \leftrightarrow a^t\ X = \lambda^t$

Formas equivalentes de enunciarmos esta proposição são:

COROLARIO 3.1 - $\lambda^t \beta$ é linearmente estimável se, e somente se $\lambda \in C(X^t)$ (2.20)

COROLARIO 3.2 - $\lambda^t\beta$ é linearmente estimável se, e somente se, posto (X) = posto $(X^t : \lambda)$ (2.21)

COROLARIO 3.3 - O número de funções paramétricas linearmente estimáveis e independentes é igual ao posto de X (2.22)

PROVA - Imediata lembrando que qualquer base para $C(X^t)$ consiste exatamente de posto (X) vetores linearmente independentes.

COROLARIO 3.4 - Qualquer função linear de funções paramétricas linearmente estimáveis é estimável. (2.23)

PROVA - Supondo que $\lambda_i^t \beta$, $i = 1, \ldots, m$, sejam funções paramétricas linearmente estimáveis

e

$\lambda^t \beta = \sum_{i=1}^{m} b_i \, (\lambda_i^t \beta)$ com $b_i \in R$, $i=1,\ldots, m$ temos, como $\lambda_i^t \beta$ é linearmente estimável $\rightarrow$ por (2.20) $\lambda_i \in C(X^t)$ para todo $i=1, \ldots, m$ mas

$$\lambda_i \in C(X^t) \rightarrow \sum_{i=1}^{m} b_i \lambda_i \in C(X^t)$$

para todo $b_i \in R$, $i=1, \ldots, m$.

assim como

$$\left(\sum_{i=1}^{m} b_i \lambda_i\right)^t \beta = \left(\sum_{i=1}^{m} b_i \lambda_i^t\right) \beta =$$

$$= \sum_{i=1}^{m} b_i \, (\lambda_i^t \, \beta) = \lambda^t \, \beta \text{ para todo } \beta \in R^p$$

onde

$\sum_{i=1}^{m} b_i \, \lambda_i \in C(X^t)$ então, por (2.20) temos que

$$\lambda^t \, \beta = \sum_{i=1}^{m} b_i (\lambda_i^t \, \beta) \text{ é}$$

é uma função paramétrica linearmente estimável.

COROLÁRIO 3.5 - Se posto (X) = p então, β é estimável e portanto cada β_i é também estimável (2.24)

PROVA - Como, por (1.33) temos posto (X) = posto $(X^tX) = p$ onde $X^t \, X$ é uma matriz $p \times p$, então,

$$\beta = (X \; X)^{-1} \, (X^tX) \; \beta = ((X^tX)^{-1} \; X^t) \; X \; \beta$$

e portanto, existe uma matriz

$$A_{n\times p} = X(X^tX)^{-1}$$

tal que $E(A^t \; y) = A^t \; X \; \beta = X \; \beta$

ou seja β é linearmente estimável e ainda, como, para cada j=1, ..., p

$$\beta_j = (\delta_{ij})\beta \quad \text{onde} \quad \delta_{ij} = \left|\begin{matrix} 1 & \text{se } i=j \\ 0 & \text{se } i\neq j \end{matrix}\right. \quad j=1,\ldots,p$$

então por (2.23) temos que

β_j é linearmente estimável

PROPOSIÇÃO 3.2 - A função $\lambda^t \beta$, $\lambda \in R^p$, é linearmente estimável $\leftrightarrow$ existe um único

$$X\rho \in C(X) \quad \text{tal que}$$

$$X^t X \rho = \lambda \text{ é consistente} \qquad (2.25)$$

PROVA - Supondo $\lambda^t \beta$ uma função paramétrica linearmente estimável então, por (2.19) existe um vetor $a \in R^n$ tal que

$$\lambda^t = a^t X \qquad \text{ou}$$

$$\lambda = X^t a \in C(X^t)$$

mas como por (1.35) temos

$$C(X^t) = C(X^t X)$$

então

$$\lambda = X^t a \in C(X^t) \rightarrow \exists \quad \rho \in R^p$$

tal que

$$\lambda = X^t a = X^t X \rho \qquad (2.26)$$

Observando que o conjunto de equações (2.26) é um conjunto de equações cuja estrutura coincide com o das equações normais que é sempre consistente e ainda $X\rho$ é o único vetor em

$C(X)$ tal que

$$X^t X \rho = X^t a \qquad \text{(ver 2.14)}$$

concluímos então, que, se $\lambda^t \beta$ é estimável então existe um único $X\rho \in C(X)$ tal que $X^t X \rho =$ é consistente

Inversamente

Supondo que existe um único $b = X\rho \in C(X)$ tal que $X^t X\rho =$ $= X^t b = \lambda$ é consistente então, como

$X^t b = \lambda$ onde $b \in C(X) \in R^n$ temos por (2.19) que $\lambda^t \beta$ é estimável.

PROPOSIÇAO 3.3 - A aproximação de mínimos quadrados para $\lambda^t \beta$ onde, $\lambda^t \beta$ é uma função paramétrica linearmente estimável, é dada por

$$\widehat{\lambda^t \beta} = \lambda^t \hat{\beta} = \rho^t X^t y \qquad \text{com}$$

$$X^t X \rho = \lambda \qquad \text{e}$$

$\widehat{\lambda^t \beta}$ é não tendencioso para $\lambda^t \beta$.

PROVA - Supondo $\lambda^t \beta$ uma função paramétrica linearmente estimável então, por (2.26) existe um único $X\rho \in C(X)$ tal que

$$X^t X\rho = \lambda \text{ é consistente}$$

assim, se $\hat{\beta}$ é tal que $X^t X \hat{\beta} = X^t y$.

então

$$X^t X \rho = \lambda \rightarrow \lambda^t \hat{\beta} = \rho^t X^t X \hat{\beta} = \rho^t X^t y$$

e

$$\widehat{\lambda^t \beta} = \widehat{\rho^t X^t X \beta} = \rho^t X^t y$$

e ainda

$$E(\lambda^t \hat{\beta}) = E(\widehat{\lambda^t \beta}) = E(\rho^t X^t y) = \rho^t X^t E(y) =$$

$$= \rho^t X^t X \beta = \lambda^t \beta \quad \text{para}$$

todo $\beta \in R^p$.

PROPOSIÇAO 3.4 - Para todo vetor $c \in R^n$ temos que

$$E(c^t y) = 0 \ (\forall \ \beta \in R^p) \leftrightarrow c \in C(X)^{\perp} \qquad (2.27)$$

PROVA - $E(c^t y) = 0 \ (\forall \ \beta \in R^p) \leftrightarrow c^t X \beta = 0 \leftrightarrow c^t X = \emptyset$ para todo $\beta \in R^p \leftrightarrow$

$$X^t c = \emptyset \leftrightarrow \text{por (1.29) } c \in C(X)^{\perp}$$

Assim, observando os resultados das proposições 3.2 e 3.4 segue que a classe total de estimadores não tendencioso para uma função paramétrica linearmente estimável, $\lambda^t \beta$, $\lambda \in R^p$, é dada por

$\rho^t X^t y + c^t y$ onde $X\rho$ é o único vetor em $C(X)$ tal que $X^t X \rho = \lambda$ é consistente e $c \in C(X)^{\perp}$

(2.28)

2.4 - Completando agora a estrutura do modelo $y = X\beta + e$ impondo as condições de Gauss-Markoff:-

X, β fixo onde

X é uma matriz $n \times p$ de constantes conhecidas

e é um vetor $p \times 1$ desconhecido tal que

$E(e) = \emptyset$

e

$$E(e\, e^t)\, \sigma^2 = \sigma^2 I_n > 0$$

(assim os e_i, $i=1, \ldots, n$, são não correlacionados e tem a mesma variância σ^2)

Pergunta-se qual é o "melhor" estimador da função paramétrica linearmente estimável $\lambda^t \beta$, $\lambda \in R^p$, na classe dos estimadores não tendenciosos para $\lambda^t \beta$?

Para responder esta questão, defina-se

DEFINIÇÃO 4.1 - (BLUE)

Um estimador linear não tendencioso $a^t y$ é dito ser o melhor estimador não tendencioso (BLUE) da sua esperança se, e somente se, sua variância não excede a variância de qualquer outro estimador não tendencioso da esperança de $a^t y$.

Considerando, então, a função linear

$\rho^t X^t y + c^t y$ com $c \in C(X)^{\perp}$ e $X^t X \rho = \lambda$ consistente

temos

a) $E(\rho^t X^t y + c^t y) = E(\rho^t X^t y) + E(c^t y) =$

$$= \rho^t X^t X \beta + c^t X \beta = \lambda^t\beta \text{ para todo } \beta \in R^p \qquad (2.29)$$

b) $\text{cov}\ (\rho^t X^t y,\ c^t y) = \rho^t X^t(I_n\ \sigma^2)c = \sigma^2\ \rho^t X^t\ c = 0$ (2.30)

c) $\text{var}\ (\rho^t X^t y + c^t y) = \text{var}\ (\rho^t X^t y) + \text{var}\ (c^t y)$

e portanto $\text{var}\ (\rho^t X^t y + c^t y) \geqslant \text{var}(\rho^t X^t y)$

onde

$$\text{var}(\rho^t X^t y + c^t y) = \text{var}\ (\rho^t X^t y) \leftrightarrow$$

$$\text{var}(c^t y) = \sigma^2 c^t c = 0 \leftrightarrow c^t c = 0 \leftrightarrow c = 0$$

e neste caso, temos que

$$\rho^t X^t y + c^t y = \rho^t X^t y \qquad (2.31)$$

assim por (2.29), (2.30) e (2.31) concluímos que o BLUE de $\lambda^t \beta$ é dado por um único vetor da forma $\rho^t x^t y$ onde ρ é tal que $X^t X \rho = \lambda$ é consistente (2.32)

Como, pela proposição 3.3, $\rho^t X^t y = \lambda^t \hat{\beta}$ onde $\hat{\beta}$ é qualquer solução das equações normais conclui-se que

TEOREMA 4.1 - TEOREMA DE GAUSS-MARKOFF - O BLUE de uma função paramétrica linearmente estimável, $\lambda^t \beta$, $\lambda \in R^p$, é o único vetor $\lambda^t \hat{\beta} = \widehat{\lambda^t\beta}$ onde $\hat{\beta}$ é qualquer vetor que minimiza $Q(\beta)$, isto é, $\hat{\beta}$ é qualquer solução das equações normais

$$X^t X \beta = X^t y \qquad (2.33)$$

COROLÁRIO 4.1 - O BLUE de uma combinação linear de funções estimáveis é igual a mesma combinação linear dos BLUE's das funções estimáveis . (2.34)

PROVA - Trivial

COROLARIO 4.2 - O BLUE de $X\beta$ é $X\hat{\beta}$ onde $\hat{\beta}$ é tal que $X^t X \hat{\beta} = X^t y$ (2.35)

PROVA - Sendo

$$X\beta = \begin{bmatrix} x_{(1)} \\ \cdot \\ \cdot \\ \cdot \\ x_{(n)} \end{bmatrix} \beta = \begin{bmatrix} x_{(1)}\beta \\ \cdot \\ \cdot \\ \cdot \\ x_{(n)}\beta \end{bmatrix} = (x_{(i)}\ \beta)$$

onde $x_{(i)} = (x_{i1} : \ldots : x_{ip})$, $i=1, \ldots, n$, é a i-ésima linha de X temos

como $x_{(i)} \in C(X^t)$ por (2.20)

$x_{(i)}\beta$ é linearmente estimável para todo $i=1,\ldots, n$

e, por (2.33) o

BLUE de $x_{(i)}\ \beta$, $i=1,\ldots, n$, é igual a $\widehat{x_{(i)}\ \beta} = x_{(i)}\ \hat{\beta}$

onde

$\hat{\beta}$ é tal que $X^t X \hat{\beta} = X^t y$

então

O BLUE de $X\ \beta = (x_{(i)}\ \beta)$ é dado por

$(x_{(i)}\ \hat{\beta}) = (x_{(i)})\ \hat{\beta} = X\ \hat{\beta}$ onde $\hat{\beta}$ é tal que

$X^t X \hat{\beta} = X^t y$

Como, por (2.16) para todo $\hat{\beta}$ tal que $X^t X \hat{\beta} = X^t y \rightarrow$ $X \hat{\beta} = P_X y$ onde P_X é a matriz do operador projeção ortogonal de R^n sobre $C(X)$ então, uma forma equivalente de enunciarmos o corolário 4.2 é

COROLÁRIO 4.3 - O BLUE de $X\beta$ é igual a projeção ortogonal de y em $C(X)$ ou seja

$$X \hat{\beta} = P_X y = \hat{y}, \text{ que é o vetor ajustado}$$

e onde, P_X é a matriz do operador projeção ortogonal de R^n sobre $C(X)$. (2.36)

COROLARIO 4.4 - Se $\lambda^t \beta$ e $\mu^t \beta$ são funções paramétricas linearmente estimáveis então, a covariância de seus BLUE's é dada por

$$\text{cov } (\mu^t \hat{\beta}, \lambda^t \hat{\beta}) = \sigma^2 \rho_1^t \lambda = \sigma^2 \mu^t \rho_2 \quad \text{onde}$$

ρ_1 e ρ_2 são tais que

$$X^t X \rho_1 = \mu$$

$$X^t X \rho_2 = \lambda \qquad (2.37)$$

e

$$\text{cov } (\mu^t \hat{\beta}, \lambda^t \hat{\beta}) = 0 \leftrightarrow$$

$(X \rho_1)^t (X \rho_2) = 0$, ou seja, os projetados do vetor y sobre $C(X)$ para as combinações lineares $\lambda^t \beta$ e $\mu^t \beta$ são ortogonais. (2.38)

PROVA - Supondo $\lambda^t\beta$ e $\mu^t\beta$ funções paramétricas linearmente estimáveis temos por (2.32) e (2.33) que o BLUE de $\lambda^t\beta$ e $\mu^t\ \beta$ são dados respectivamente por

$$\rho_2^t\ X^t\ y = \lambda^t\ \hat{\beta} \qquad e$$

$$\rho_1^t\ X^t\ y = \mu^t\ \hat{\beta}$$

para todo $\hat{\beta}$ tal que $X^t\ X\ \hat{\beta} = X^t\ y$

e para ρ_1 e $\rho_2 \in R^p$ tal que

$X^t\ X\ \rho_1 = \mu$ é consistente

e

$X^t\ X\ \rho_2 = \lambda$ é consistente

e portanto,

$$\text{cov}\ (\mu^t\hat{\beta},\ \lambda^t\hat{\beta}\ ) = \text{cov}\ (\rho_1^t\ X^t\ y,\ \rho_2^t\ X^t\ y) =$$

$$= \rho_1^t\ X^t\ (\sigma^2\ In)\ X\ \rho_2 = \sigma^2\ \rho_1^t\ X^t\ X\ \rho_2 =$$

$$= \sigma^2\ \mu^t\ \rho_2 = \sigma^2\ \rho_1^t\ \lambda$$

e $\qquad \text{cov}\ (\mu^t\ \hat{\beta}, \lambda^t\beta\ ) = 0 \quad \leftrightarrow$

$$\sigma^2\ \rho_1^t\ X^t\ X\rho_2 = 0 \leftrightarrow (X\rho_1)^t (X\rho_2) = \rho_1^t\ X^t\ X\rho_2 = 0$$

assim como $\quad X\ \hat{\beta} = P_x\ y$ temos

$$\mu^t\ \hat{\beta} = \rho_1^t\ X^t\ X\ \hat{\beta} = \rho_1^t\ X^t\ X\ \hat{\beta} = \rho_1^t\ X^t(\ P_x\ y) =$$

$$= (X\ \rho_1)^t\ (P_x\ y)$$

e

$$\lambda^t \hat{\beta} = \rho_2^t X^t X \hat{\beta} = \rho_2^t X^t (P_x y) = (X \rho_2)^t (P_x y)$$

onde P_x é a matriz do operador projeção ortogonal sobre C(X) então

$$\text{cov} (\mu^t \hat{\beta}, \lambda^t \hat{\beta}) = 0 \leftrightarrow$$ os projetados de y sobre C(X) para as combinações lineares $\lambda^t \beta$ e $\mu^t \beta$ são ortogonais.

000o000

CAPÍTULO 3

CARACTERIZAÇÃO DE BLUE'S NO MODELO DE GAUSS - MARKOFF GENERALIZADO (GMG)

Considerando o modelo linear

$$y = X\beta + e$$

onde

X é uma matriz $n \times p$ conhecida de posto $r \leq p$

β é um vetor $p \times 1$ de parâmetros

e é um vetor $n \times 1$ de perturbação ou erro

então, se

$$E(e) = \emptyset$$

e

$$E(e\, e^t) = \sigma^2 V$$

onde

$\sigma^2 \in R$ é não conhecido

V é uma matriz $n \times n$ conhecida, simétrica e positiva de posto $s \leq n$ (3.1)

temos, que o modelo acima denomina-se modelo de Gauss-Markoff Geral (G.M.F.).

Simbolicamente temos o modelo de G.M.G. dado por
y, $X\beta$, $\sigma^2 V$)

3.1 - CONSIDERAÇÕES A RESPEITO DO MODELO DE G.M.G.

Considerando o modelo de G.M.G. $(y, X\beta, \sigma^2 V)$ temos

$$E(y) = X\beta \qquad e$$

$$D(y) = E[(y-X\beta)(y-X\beta)^t] = \sigma^2 V$$

($D(y)$ = matriz de variância e covariância de y)

Como V é singular temos então que as variáveis y são linearmente dependentes e portanto, deverá ocorrer certas restrições lineares sobre a variável dependente y, o qual terão que ser observadas certificando-se então a consistência (ou admissibilidade) do modelo.

Uma tal restrição é dada por

LEMA 1.1 - Se $(y, X\beta, \sigma^2 V)$ é um modelo de G.M.G. então

$$y \in C(V:X) \text{ com probabilidade } 1 \qquad (3.2)$$

PROVA - Considerando o modelo linear $y = X\beta + e$ dado em (3.1) temos, para todo vetor $c_{n\times 1} \in C(V)^{\perp} \rightarrow$

$$c^t V = V c = \emptyset$$

assim

$$E(c^t e) = \emptyset$$

e

$$var(c^t e) = \sigma^2 c^t V c = \emptyset. \quad (\sigma^2 > 0)$$

então $c^t e = 0$ com probabilidade 1 para todo vetor $c \in C(V)^{\perp}$ ou seja, o vetor $e \in C(V)$ com probabilidade 1 .

Logo, como $y = X\beta + e$ então $y \in C(V:X)$ com probabilidade 1.

O lema 3.1 especifica então que $y \in C(V:X)$ que é a única afirmação que pode ser feita quando y é não observado. Porém, quando temos uma observação de y, então na sua estrutura teremos que

$$y = X\beta + P_V e$$

onde

P_V é a matriz do operador projeção ortogonal sobre $C(V)$ e portanto o modelo é admissível se y e β são tais que

$$y - X\beta \in C(V) \tag{3.3}$$

o qual é automáticamente satisfeito quando V é não singular.

Observando que se Z é uma matriz $n \times n-r$ (r=posto X) tal que $C(z) = C(X)^{\perp}$ temos

$$C(V:X) = C(VZ:X) = C(VZ) \oplus C(X) \tag{3.4}$$

pois, para todo $\lambda \in R^n$ tal que $\lambda \in C(V:X)^{\perp}$

$$\rightarrow \lambda^t V = \emptyset \text{ e } \lambda^t X = \emptyset \rightarrow \lambda^t VZ = \emptyset \text{ e } \lambda^t X = \emptyset$$

$$\rightarrow \lambda \in C(VZ:X)^{\perp}$$

e, para todo vetor $\lambda \in R^n$ tal que $\lambda \in (VZ) \cap C(X) \rightarrow$ existe $\lambda_1 \in R^P$ e $\lambda_2 \in R^{n-r}$ tal que

$$\lambda = X\lambda_1 \quad e \quad \lambda = VZ\lambda_2 \quad \leftrightarrow \quad \left\| \begin{array}{l} Z^t\lambda = Z^t X \lambda_1 = \emptyset \\ e \\ \lambda = VZ\lambda_2 \end{array} \right.$$

$$\rightarrow \quad \lambda_2^t Z^t \lambda = \lambda_2^t Z^t V Z \lambda_2 = \emptyset \quad \leftrightarrow \quad C^t Z \lambda_2 = \emptyset$$

onde C é uma matriz $n \times s$ tal que $V = CC^t$ (V: é positiva)

$$\leftrightarrow \quad \lambda = V Z \lambda_2 = \emptyset$$

assim, temos $C(VZ:X) = C(VZ) \oplus C(X) \subset C(V:X)$ e

$$\dim C(VZ:X) = \dim C(VZ) + \dim C(X) = \text{por } (1.33)$$

$$\dim C(V) - \dim(C(V) \cap C(Z)^{\perp}) + \dim(C(X)) =$$

$$= \dim C(V) + \dim(C(X)) - \dim(C(V) \cap C(X)) =$$

$$= \dim C(V:X)$$

se, e somente se,

$$C(V:X) = C(VZ) \oplus C(X)$$

Considerando, também que K é uma matriz $n \times n-s$, (s=posto(V)) tal que as colunas de K formam uma base ortonormal para

$$C(V)^{\perp} \qquad (3.5)$$

então, temos que

$$C(V:X) = C(V) \overset{\perp}{\oplus} C(K\ K^t\ X) \tag{3.6}$$

pois,

Considerando P_V a matriz do operador projeção ortogonal sobre $C(V)$ e

$P_{(V:X)}$ a matriz do operador projeção ortogonal sobre $C(V:X)$ temos

por (1.55) $I-P_V$ é a matriz do operador projeção ortogonal sobre $C(V)^{\perp}$ e portanto,

por (1.59) $I-P_V = K\ K^t$

assim, como por (1.62)

$$C(V:X) = C(P_V) \overset{\perp}{\oplus} C(P_{(V:X)} - P_V)$$

onde, sendo $(P_{(V:X)} - P_V)\ (V:X) =$

$$= P_{(V:X)}(V:X) - P_V(V:X) =$$

$$= (V:X)-(P_V(V):P_VX) = (V:X) - (V:P_VX) =$$

$$= (\emptyset : X-P_VX) = (\emptyset : (I-P_V)(X)) =$$

$= (\emptyset : K\ K^t\ X)$ então

$$C(P_{(V:X)} - P_V) = C(K\ K^t\ X)$$

e portanto

$$C(V:X) = C(V) \overset{\perp}{\oplus} C(K\ K^t\ X)$$

Assim por (3.4) e (3.6) concluímos que

LEMA 1.2 - Se $(y, X\beta, \sigma^2 V)$ é um modelo de G.M.G. e Z é uma matriz $n\times n-r$ (r=posto X) tal que $C(Z) = C(X)^{\perp}$ e K é uma matriz $n\times n-s$ (s=posto V) tal que as colunas de K formam uma base ortonormal para $C(V)^{\perp}$ então

$$y \in C(X) \oplus C(VZ) = C(K\, K^t\, X) \overset{\perp}{\oplus} C(V)$$

com probabilidade 1 $\qquad$ (3.7)

PROVA - use (3.2), (3.4) e (3.6).

Vejamos o que ocorre com o modelo de G.M.G. se o transformarmos pela matriz K^t onde K é dada por (3.5).

LEMA 1.3 - Se $(y, X\beta, \sigma^2 V)$ é um modelo de G.M.G. e K é uma matriz como especificada em (3.5) então

$$K^t y = d \text{ constante com probabilidade 1} \qquad (3.8)$$

e

$$K^t X\beta = d \text{ para todo } \beta \in R^p \text{ se}$$

$$C(K) \not\subset C(X)^{\perp} \qquad (3.9)$$

PROVA - Como K é uma matriz tal que as colunas de K formam uma base ortonormal para $C(V)^{\perp}$ então temos que

$$VK = K^t\, V = \emptyset \qquad (3.10)$$

assim, como

$$E(K^t y) = K^t X\beta \text{ para todo } \beta \in R^p$$

e

$$var(K^t y) = K^t V K \sigma^2 = \emptyset \text{ (por 3.10))}$$

então

$$K^t y = K^t X\beta = d \text{ constante com probabilidade 1.}$$

Note que se $C(K) \subset C(X)^{\perp}$ então $K^t X = \emptyset$ e portanto não haveria restrição sobre o parâmetro β.

Assim, o resultado (3.8) e (3.9) mostram que quando V é singular existe uma restrição na variável aleatória y e no vetor dos parametros β.

Contudo, elas são conhecidas apenas quando temos uma observação de y.

LEMA 1.4 - Para todo vetor $q \in R^{n-s}$ tal que

$$q^t\, d = 0 \qquad e \qquad Kq \neq \emptyset$$

onde

K é uma matriz dada por (3.5) e

d satisfaz (3.8) e (3.9)

então

$$q^t\, K^t\, y = 0 \text{ com probabilidade 1} \quad (3.10)$$

e

$$q^t\, K^t\, X\beta = 0 \qquad (3.11)$$

PROVA - use (3.8) e (3.9).

Como $q^t K^t y = 0$ com probabilidade 1

$$y \in C(Kq)^1$$

então, o lema 1.4 mostra que de fato y está ligado a um determinado sub-espaço e também, como as únicas restrições sobre o parâmetro β são dadas através de (3.9) segue-se que as únicas funções de β que são iguais a zero são aquelas cujas expressões são dadas por (3.11).

Assim, se $\lambda^t\beta = 0$ para todo $\beta \in R^p$ tal que

$$K^t X \beta = d \quad \text{(constante)}$$

então $\lambda^t = q^t K^t X$ para algum $q \in R^{n-s}$ tal que $q^t d=0$ e

$$Kq \neq \emptyset \tag{3.12}$$

Observar que

$$\text{posto}(Kq) = \left|\begin{array}{ll} \text{posto}(K) - 1 & \text{se } d \neq \emptyset \\ \text{ou} & \\ \text{posto}(K) & \text{se } d = \emptyset \end{array}\right.$$

pois $\text{posto}(Kq) = \text{posto}(K) - \dim(C(K^t) \cap C(K^t y))$

$$= \text{posto}(K) - \dim(C(K^t y))$$

Como a singularidade de V induz restrições no vetor de parâmetros β, como fica então o problema de estimabilidade?

Zyskind e Martin (1969) provam que

TEOREMA 1.1 - Se $(y, X\beta, \sigma^2 V)$ é um modelo de G.M.G. então, toda função linear dos parâmetros, $\lambda^t \beta$, $(\lambda \in R^p)$ é linearmente estimável se, e somente se, $\lambda \in C(X^t)$

PROVA - Se, $\lambda \in R^p$ é tal que $\lambda \in C(X^t)$ então, existe um vetor $a \in R^n$ tal que $\lambda^t = a^t X$. Assim, $E(a^t y) = a^t X\beta = \lambda^t \beta$ $\rightarrow$

$$\lambda^t \beta \text{ é estimável}$$

inversamente,

Supondo $\lambda^t \beta$ uma função paramétrica linearmente estimável então, existe um vetor $a \in R^n$ tal que $E(a^t y) = \lambda^t \beta$ para todo $\beta \in R^p$ tal que

$$K^t X\beta = d$$

assim, como

$$E(a^t y) = a^t X\beta = \lambda^t \beta \leftrightarrow (a^t X - \lambda^t)\beta = 0$$

por (3.12) temos $a^t X - \lambda^t = q^t K^t K$ para algum $q \in R^{n-s}$ onde

$$q^t d = 0 \quad \text{e} \quad Kq \neq \emptyset \tag{3.13}$$

assim (3.13) $\rightarrow a^t X - \lambda^t = q^t K^t X \leftrightarrow$

$$\lambda^t = (a^t - q^t K^t) X \leftrightarrow \lambda \in C(X^t)$$

COROLÁRIO 1.1 - Se $E(a^t y) = \lambda^t \beta$ para todo vetor $\beta \in R^p$ tal que $K^t X\beta = d$ (constante), então, $X^t a - \lambda \in C(X^t K)$.

PROVA - imediata usando (3.13)

Deste modo, observamos que a condição $a^t X = \lambda^t$ é apenas necessária para que $E(a^t y) = \lambda^t \beta$ que é bastante diferente da condição usual que estabelece que se $a^t y$ é não tendencioso para $\lambda^t \beta \rightarrow X^t a = \lambda$; condição válida apenas quando β é não restrito.

Tal afirmação foi usada erroneamente por Rao (1971).

A falsidade desta afirmação foi verificada por Zyskind(1967) e no trabalho de Zyskind e Martin (1969) como citado anteriormente.

Uma nota interessante no modelo quando V é singular é

PROPRIEDADE 1.1 - Se $E(a^t y) = \lambda^t \beta$ para todo $\beta \in R^p$ tal que $K^t X\beta = d$ (constante), então, existe $b \in R^n$

$$b = a - Kq \text{ (K e q dados no lema 1.4)}$$

tal que

$$b^t X = \lambda^t \text{ (e portanto } E(b^t y) = \lambda^t \beta) \tag{3.14}$$

e

$$a^t y = b^t y \text{ com probabilidade 1} \tag{3.15}$$

PROVA - Supondo $b = a - Kq$ então

$$b^t X = a^t X - q^t K^t X$$

$$b^t X - \lambda^t = a^t X - q^t K^t X - \lambda^t = (a^t - q^t K^t) X - \lambda^t = 0 \text{ por (3.13) e}$$

portanto, $b^t X = \lambda^t \rightarrow b^t X\beta = \lambda^t\beta$ para todo $\beta \in R^p$

$$\rightarrow E(b^t y) = \lambda^t\beta$$

assim, $E(a^t y - b^t y) = E(a^t y) - E(b^t y) = 0$

e

$$var(a^t y - b^t y) = var((a^t - b^t)y) =$$

$$= var(q^t K^t y) = 0$$

então, $a^t y = b^t y$ com probabilidade 1.

Assim, quando V é singular temos que uma particular função paramétrica linearmente estimável, $\lambda^t\beta$, pode admitir distintos BLUE's como função das observações embora, o valor numérico de todos êles sejam iguais.

Dito de outra forma, com V singular podemos ter um conjunto de funções lineares das observações identicamente nulas que, quando adicionadas ou subtraídas do BLUE de $\lambda^t\beta$ o novo funcional linear também será BLUE de $\lambda^t\beta$.

Portanto, para acharmos um BLUE da função paramétrica linearmente estimável, $\lambda^t\beta$, é suficiente determinarmos um vetor $b \in R^n$ tal que

$$b^t X = \lambda^t$$

e

$$b^t V b \text{ seja mínimo}$$

embora, esta condição não seja uma condição necessária.

TEOREMA 1.2 - Se $\lambda \in R^p$ é tal que $\lambda^t\beta$ é uma função paramétrica linearmente estimável no modelo de G.M.G. $(y, X\beta, \sigma^2 V)$, então, existem vetores $b \in R^n$ e $m \in R^p$ tal que

$$V\,b + X\,m = \emptyset \qquad (3.16)$$

e

$$X^t\,b = \lambda \qquad (3.17)$$

e neste caso $b^t y$ é um BLUE de $\lambda^t \beta$.

PROVA - Supondo $\lambda^t \beta$ uma função paramétrica linearmente estimável, então, pelo teorema 1.1 existe um vetor $b \in R^n$ tal que $X^t b = \lambda$. Mas, como $X^t b = \lambda \to b^t y$ é um estimador não tendencioso de $\lambda^t \beta$ então, utilizando o processo de Multiplicador de Lagrange para determinar

$$b \in R^n \text{ tal que}$$

$$b^t\,V\,b \text{ é mínimo}$$

sujeito à condição de

$$X^t\,b = \lambda^t$$

temos,

O Lagrangiano L com um vetor de multiplicador $2m \in R^p$ é

$$L = b^t\,V\,b + 2(b^t\,X - \lambda^t)m =$$

$$= \sum_{i=1}^{n} \sum_{j=1}^{n} b_i b_j \sigma_{ij} + 2 \sum_{i=1}^{n} \sum_{k=1}^{p} b_i x_{ik} m_k - 2 \sum_{k=1}^{p} \lambda_k m_k$$

onde σ_{ij} pertence a i-ésima linha e j-ésima coluna de V.

assim, sendo

$$\frac{\partial L}{\partial b_u} = 0 \quad \text{para} \quad u=1,\dots,n$$

dado por

$$2 \sum_{k=1}^{n} \sigma_{uj}\, b_j + 2 \sum_{k=1}^{p} x\, u_k m_k = 0 \quad \text{para} \quad u=1,\dots,n$$

em forma matricial temos

$$V\, b + X\, m = \emptyset$$

Considerando o sistema de equações

$$\left|\begin{array}{l} Vb + Xm = \emptyset \\ X^t\, b = \lambda \end{array}\right.$$

temos que (3.16) e (3.17) satisfazem,

1 - <u>O sistema de equações é consistente</u> pois, para todo vetor ν_1 e $\nu_2 \in R^n$ tal que

$$(\nu_1^t : \nu_2^t) \begin{pmatrix} V & X \\ X^t & \emptyset \end{pmatrix} = \emptyset \quad \rightarrow$$

$$\rightarrow \left|\begin{array}{ll} \nu_1^t\, V + \nu_2^t\, X^t = \emptyset & \quad (3.18) \\ \text{e} & \\ \nu_1^t\, X = \emptyset & \quad (3.19) \end{array}\right.$$

multiplicando (3.18) por ν_1 à esquerda

$$\rightarrow \left|\begin{array}{l} \nu_1^t\, V\, \nu_1 + \nu_2^t\, X^t\, \nu_1 = \emptyset \\ \text{e} \\ \nu_1^t\, X = \emptyset \end{array}\right.$$

$$\rightarrow \nu_1^t \, V \, \nu_1 = \emptyset \tag{3.20}$$

mas, como V é positiva, então existe uma matriz $C_{n \times s}$ tal que $V = C^t C$ assim, (3.20) fica

$$\nu_1^t \, V \, \nu_1 = \emptyset \;\leftrightarrow\; (C \, \nu_1)^t (C \, \nu_1) = \emptyset \leftrightarrow C \, \nu_1 = \emptyset$$

e portanto $V \, \nu_1 = \emptyset$ ou $\nu_1^t \, V = \emptyset$ (3.21)

Substituindo (3.21) em (3.18) temos

$$\nu_2^t \, X^t = \emptyset \tag{3.22}$$

como, por hipótese $X^t b = \lambda$ é consistente então, por (3.22)

$$\nu_2^t \, X^t = \emptyset \;\rightarrow\; \nu_2^t \, \lambda = 0$$

Logo

$$(\nu_1^t \,:\, \nu_2^t) \begin{pmatrix} V & X \\ X^t & \emptyset \end{pmatrix} = \emptyset \rightarrow (\nu_1^t \,:\, \nu_2^t) \begin{pmatrix} \emptyset \\ \lambda \end{pmatrix} = \emptyset$$

assim, existe $b \in R^n$ e $m \in R^p$ tal que satisfazem (3.16) e (3.17) e neste caso, como b é tal que

$$E(b^t y) = \lambda^t \, \beta$$

então $b^t y$ é um BLUE de $\lambda^t \beta$ pois

2 - Invariância das Soluções

Supondo b_1 e $b_2 \in R^n$ e

m_1 e $m_2 \in R^p$

tais que satisfazem

$$\left|\begin{array}{ll} V\ b_1 + Xm_1 & = \emptyset \\ X^t\ b_1 & = \lambda \end{array}\right.$$

e

$$\left|\begin{array}{ll} V\ b_2 + Xm_2 & = \emptyset \\ X^t\ b_2 & = \lambda \end{array}\right.$$

então, $V(b_1-b_2)+X(m_1-m_2) = \emptyset$ (3.23)

e $X^t(b_1 - b_2) = \emptyset$ (3.24)

multiplicando (3.23) à esquerda por $(b_1^t - b_2^t)$ temos

$$(b_1^t - b_2^t)V(b_1 - b_2)+(b_1^t - b_2^t)X(m_1 - m_2) = \emptyset \quad (3.25)$$

mas, como por (3.24) $X^t(b_1-b_2) = 0 \quad \rightarrow$

$$\rightarrow \quad (b_1^t - b_2^t)\ X\ (m_1 - m_2) = \emptyset \quad (3.26)$$

então, substituindo (3.26) em (3.25) temos

$$(b_1^t - b_2^t)\ V\ (b_1 - b_2) = \emptyset \rightarrow V(b_1 - b_2) = \emptyset$$

(usando o mesmo argumento $V = CC^t$ anterior)

Assim $V(b_1 - b_2) = \emptyset \rightarrow V\ b_2 = V\ b_2$

e também por (3.23)

$$V(b_1-b_2)=\emptyset \to X(m_1-m_2)=\emptyset \to X\ m_1=X\ m_2$$

claramente temos também que

$$b_1Vb_1 = b_1Vb_2 = b_2Vb_1 = b_2Vb_2 \tag{3.27}$$

3 - Minimização Global

Se $b \in R^n$ satisfaz o sistema de equações (3.16) e (3.17) então, para todo $\rho \in R^n$ tal que $X^t(b+\rho) = \lambda$ temos

$$X^t\ \rho = \emptyset \tag{3.28}$$

e

$$\rho^tVb = -\rho^tX\ m \to \rho^tVb = 0 = b^tV\rho \tag{3.29}$$

e portanto

$$\mathrm{cov}(b^ty,\ \rho^ty) = \sigma^2\ b^tV\rho = 0 \tag{3.30}$$

assim,

$$\mathrm{var}((b+\rho)^ty)=\mathrm{var}(b^ty + \rho^ty) =$$

$$= \mathrm{var}(b^ty)+2\mathrm{cov}(b^ty,\rho^ty)+\mathrm{var}(\rho^ty) =$$

$$= (\text{por}(3.30))\ \mathrm{var}(b^ty) + \mathrm{var}(\rho^ty) \to$$

$$\mathrm{var}((b+\rho)^ty) \geq \mathrm{var}(b^ty)$$

e portanto b^ty é um BLUE de $\lambda^t\beta$ onde

$$\mathrm{var}((b+\rho)^ty) = \mathrm{var}(b^ty) \leftrightarrow \mathrm{var}(\rho^ty) = 0 \leftrightarrow$$

$$\rho^t V = 0 \leftrightarrow V\rho = \emptyset \tag{3.31}$$

assim, por (3.29) e (3.31) temos

$$(b+\rho)^t y \text{ é também um BLUE de } \lambda^t\beta \leftrightarrow$$

$$\rho \in C(X)^{\perp} \cap C(V)^{\perp} \tag{3.32}$$

Como por (3.28) e (3.30) temos que para todo $b \in R^n$ satisfazendo (3.16) e (3.17) e para todo $\rho \in R^n$ tal que $X^t\rho = \emptyset$

$$E(\rho^t y) = \rho^t X \beta = 0 \text{ para todo } \beta \in R^p$$

e $$\text{cov}(b^t y, \rho^t y) = 0$$

então, podemos concluir que, ρ não modifica a estimabilidade de $\lambda^t\beta$ e não aumenta a informação que $b^t y$ tem sobre $\lambda^t\beta$. Tal caso, não é o mesmo que veremos na seção 3.3 deste capítulo em que encontraremos uma forma de melhorar a informação sobre $\lambda^t\beta$ através de vetores de $C(X)^{\perp}$ usando o argumento apresentado por Rao(1967) denominado ajustamento por covariância.

Supondo $b \in R^n$ satisfazendo (3.16) e (3.17) e

$$a \in R^n \text{ tal que } E(a^t y) = \lambda^t\beta$$

temos

$$E((a-b)^t y) = E(a^t y - b^t y) =$$

$$= E(a^t y) - E(b^t y) = 0$$

mas, como por (3.13)

$$E((a-b)^t y) = 0 \leftrightarrow (a-b)^t X = q^t K^t X$$

onde K é uma matriz tal que as suas colunas formam uma base ortonormal de $C(V)^{\perp}$ e q é tal que $K q \neq 0$ e $q^t K^t X\beta = 0$ temos,

$$(a-b)^t X = q^t K^t X \leftrightarrow$$

$$(a^t - b^t - q^t K^t) X = \emptyset \leftrightarrow a-b-Kq = \rho$$

onde

ρ é tal que $X^t \rho = \emptyset$ e portanto,

$$a-b = \rho + Kq \in C(X)^{\perp} + C(K).$$

Como $a = b+(a-b)$ onde

$$\mathrm{cov}(b^t y, (a-b)^t y) = \mathrm{cov}(b^t y, \rho^t y + q^t K^t y) =$$

$$= \mathrm{cov}(b^t y, \rho^t y) + \mathrm{cov}(b^t y, q^t K^t y) = \text{por } (3.29)$$

$$= \mathrm{cov}(b^t y, q^t K^t y) = b^t V Kq\, \sigma^2 = 0$$

então

$$\mathrm{var}(a^t y) = \mathrm{var}(b^t y + (a-b)^t y) =$$

$$= \mathrm{var}(b^t y) + \mathrm{var}((a-b)^t y) \geq \mathrm{var}(b^t y)) \qquad (3.33)$$

onde, $a^t y$ é também um BLUE de $\lambda^t \beta$ se, e somente se

$$\mathrm{var}((a-b)^t y) = (a-b)^t V (a-b) \sigma^2 = 0 \leftrightarrow$$

$$(a-b)^t V (a-b) = 0 \leftrightarrow V(a-b) = 0$$

$$Va = Vb = -Xm \quad (\text{por } (3.17)) \tag{3.34}$$

(Observe que se V é não singular então, a única escolha para a é a=b)

Portanto, concluímos que

COROLÁRIO 1.2 - Considerando o modelo de G.M.G. $(y, X\beta, \sigma^2 V)$.

Uma função $a^t y$ das observações é BLUE da sua esperança se, e somente se,

$$Va \in C(X) \tag{3.35}$$

Tal resultado foi também provado por Zyskind (1967) onde, no seu desenvolvimento êle obtém explicitamente um conjunto de equações para o qual o vetor a pertence.

Rao (1965) mostra também que uma estatística é um estimador não tendencioso de variância mínima da sua esperança se é não correlacionado com toda estatística não viciada para o zero.

COROLÁRIO 1.3 - Se Z é uma matriz n×n-r (r=posto X) tal que $C(Z) = C(X)^{\perp}$ então, $a^t y$ é BLUE da sua esperança se, e somente se,

$$Z^t \, V \, a = \emptyset \tag{3.36}$$

PROVA - Use (3.35)

COROLÁRIO 1.4 - Funções lineares de BLUE's é também BLUE da sua esperança (3.37)

PROVA - trivial.

COROLÁRIO 1.5 - Se V é não singular então, existe um único $a=V^{-1}Xm$; $m \in R^p$, tal que $a^t y$ é BLUE de $\lambda^t\beta$ (3.38)

e reciprocamente, se $E(a^t y) = \lambda^t\beta$ e $a = V^{-1}Xm$ para algum $m \in R^p$ então $a^t y$ é BLUE de $\lambda^t\beta$. (3.39)

PROVA - Supondo $a \in R^n$ tal que

$$a^t X = \lambda^t$$

e

$$V_a = Xm \text{ para algum } m \in R^p$$

como V é não singular, temos

$$a^t X = \lambda^t\beta$$

e

$$a = V^{-1} Xm$$

onde $a^t y$ é BLUE de λ^t

ainda, como para todo $\rho \in R^n$ então $(a+\rho)^t y = (a^t+\rho^t)y$ é também um BLUE de $\lambda^t\beta \leftrightarrow$ (por (3.32))

$$\rho \in C(X)^{\perp} \cap C(V)^{\perp}$$

então $\rho \in C(V)^{\perp} \to V\rho = \emptyset$ mas, como V é não singular e $V\rho = \emptyset$ temos que $\rho=\emptyset$ e portanto $a = V^{-1}Xm$ é o único vetor em $C(V^{-1}X)$ tal que $a^t y$ é BLUE de $\lambda^t\beta$

<u>reciprocamente</u>

como $E(a^t y) = \lambda^t\beta$ e $a = V^{-1}Xm$ para algum $m \in R^p$ então,

$$Va = V V^{-1} Xm = Xm \in C(X)$$

e portanto por (3.35) temos que

$$a^t y \text{ é BLUE de } \lambda^t \beta \, .$$

PROPRIEDADE 1.1 - O BLUE de $X\beta$ é $A^t y$ onde A é uma matriz $n \times n$ satisfazendo

$$VA + XM = \emptyset \text{ para alguma matriz } M_{p \times n}$$

e

$$X^t A = X^t \tag{3.40}$$

PROVA - Considerando $\lambda_i^t = x_{(i)}$ onde, $x_{(i)}$, $i = 1, \dots, n$, é a i-ésima linha de X temos pelo teorema 1.1 - 3.1 que $\lambda_i^t \beta$ é estimável assim, se $a_{(i)} \in R^n$, $i=1, \dots, n$ é tal que

$$V a_{(i)} + X m_{(i)} = \emptyset \text{ para algum } m_{(i)} \in R^p, \ i=1, \dots, n$$

e

$$X^t a_{(i)} = \lambda_i$$

temos que $a_{(i)}^t y$ é BLUE de $\lambda_i^t \beta = x_{(i)} \beta$ e portanto, tomando

$$A = (a_{(i)}) \text{ que é uma matriz } n \times n$$

e

$$M = (m_{(i)}) \text{ que é uma matriz } n \times p$$

temos que

$$V A + X M = \emptyset$$

e

$$X^t A = X^t$$

onde $A^t y$ é BLUE de $X\beta$.

COROLÁRIO 1.6 - Se V é inversível e $A = V^{-1} X M$ onde M é tal que

$$(X^t V^{-1} X) M = X^t$$

então

$$A^t y \text{ é BLUE de } X\beta$$

onde, A^t é a matriz do operador projeção sobre $C(X)$ segundo a direção de vetores de $C(VZ)$ onde Z é uma matriz tal que

$$C(Z) = C(X)^{\perp} \tag{3.41}$$

e ainda, se

$$X^t V^{-1} X \text{ é inversível então}$$

$$A^t y = X(X^t V^{-1} X)^{-1} X^t V^{-1} y \quad \text{(Aitken 1934)} \tag{3.42}$$

PROVA - Utilizando (3.40) é trivial que $A^t y$ é BLUE de $X\beta$. Mostrando que A^t é a matriz do operador sobre $C(X)$ segundo a direção de $C(VZ)$ temos,

como $(X^t V^{-1} X) M = X^t \leftrightarrow M^t (X^t V^{-1} X) = X$

então $\quad X M = M^t (X^t V^{-1} X) M = M^t X^t = (X M)^t$

assim $\quad A^2 = V^{-1} X M V^{-1} X M = V^{-1} M^t X^t V^{-1} X M =$

$$= V^{-1} M^t X^t = V^{-1} X M = A$$

e portanto $\quad A^2 = A \Longrightarrow (A^t)^2 = A^t$

e ainda, como

$$A^t X = X$$

e

$$A^t VZ = (VA)^t Z = M^t X^t Z = \emptyset$$

temos, que A^t é a matriz do operador projeção sobre $C(X)$ segundo a direção de vetores de

$$C(VZ)$$

Assim, como por (3.6) $C(V:X) = C(X) \oplus C(VZ)$ temos que o BLUE de $X\beta$ é a projeção de y em $C(X)$ segundo a direção de vetores de $C(VZ)$.

3.2 - CONDIÇÕES EM QUE O MELHOR ESTIMADOR DE MÍNIMOS QUADRADOS SIMPLES (SLSE) É TAMBÉM UM BLUE NO MODELO DE G.M.G.

DEFINIÇÃO 2.1 - Chamamos de Melhor Estimador dos Mínimos Quadrados Simples (SLSE) de uma função paramétrica linearmente estimável, $\lambda^t\beta$, ao estimador não tendencioso de menor variância de $\lambda^t\beta$ no modelo $(y, X\beta, \sigma^2 I)$.

Assim, se o BLUE de $\lambda^t\beta$ é o estimador não tendencioso de menor variância de $\lambda^t\beta$ no modelo de G.M.G. temos,

TEOREMA 2.1 - O SLSE de $X\beta$ é também o BLUE de $X\beta$ se, e somente se,

$$V\,X = X\,Q \text{ para alguma matriz } Q(p\times p) \tag{3.43}$$

PROVA - Supondo $A^t y$ um BLUE de $X\beta$ onde

$$X^t\,A = X^t$$

e

$$V\,A = X\,M \text{ para alguma matriz } M_{p\times n}$$

e sendo o SLSE de $X\beta$ dado por $P_x y$ onde P_x é a matriz do operador projeção ortogonal sobre $C(X)$ (ver (2.36)) então, se

SLSE ≡ BLUE temos (por (3.40)) que

$$P_x X = X$$

e

$$V P_x = X M \text{ para alguma matriz } M_{p\times n}$$

e portanto,

$$V P_x X = V X = X M X = X Q$$

onde $Q = M X$ é uma matriz $p\times p$

Inversamente, se $V X = X Q$, tomando uma matriz B $p\times n$ tal que $P_x = X B$ é a matriz do operador projeção ortogonal sobre $C(X)$, então

$$P_x X = X$$

e

$$V P_x = V X B = X Q B = X M$$

onde $M = QB$ é uma matriz $p\times n$

→ (teor 1.2, 3.1) $P_x y$ é o BLUE de $X\beta$ e portanto por (2.36) o SLSE de $X\beta$ é também o BLUE de $X\beta$.

COROLÁRIO 2.1 - Se V é não singular então SLSE ≡ BLUE de $X\beta$ se, e somente se, existe uma matriz $R_{p\times p}$ tal que

$$V^{-1} X = X R \tag{3.44}$$

PROVA - trivial

Observando que $V X = X Q$ se, e somente se, $C(X)$ é um sub--espaço invariante da matriz V (Hoffman, Kunze (1979) pg 254) então, temos

TEOREMA 2.2 - Para qualquer função das observações, $\omega^t y$, temos que o BLUE e o SLSE da $E(\omega^t y)$ são iguais se, e somente se, $C(X)$ é um sub-espaço invariante da matriz V. (3.45)

PROVA - como, $\omega^t y$ é SLSE da $E(\omega^t y) \leftrightarrow \omega \in C(X)$ então, se $\omega^t y$ é também BLUE da $E(\omega^t y)$ temos por (3.35) que $V\omega \in C(X)$.

Assim, se para todo $W \in C(X) \rightarrow V\omega \in C(X)$ então $C(X)$ é um sub-espaço invariante de V.

Inversamente

como, $\omega^t y$ é SLSE da $E(\omega^t y) \leftrightarrow \omega \in C(X)$ e por hipótese

$$\omega \in C(X) \rightarrow V\omega \in C(X)$$

então, se

$\omega^t y$ é SLSE da $E(\omega^t y)$ temos por (3.35) que $\omega^t y$ é também BLUE de $E(\omega^t y)$.

Zyskind (1967) mostra também que

TEOREMA 2.3 - Para qualquer função paramétrica linearmente estimável temos que o seu BLUE e o SLSE são iguais se, e somente se V possui r auto vetores ortogonais que formam uma base ortonormal para $C(X)$. (3.46)

PROVA - Supondo $\lambda^t\beta$ uma função paramétrica linearmente estimável temos por (3.45) que o SLSE de $\lambda^t\beta$ é também o BLUE de $\lambda^t\beta$ se e somente se C(X) é um sub-espaço invariante através da matriz V e portanto (Hoffman, Kunze (1979) pg 284)

$C(X)^{\perp}$ é também um sub-espaço invariante através da matriz V.

Disto, se 0_1^t é uma matriz n×r tal que suas colunas formam uma base ortonormal para C(X) e

0_o^t é uma matriz n×(n-r) tal que suas colunas formam uma base ortonormal para $C(X)^{\perp}$

temos então que, existem matrizes simétricas W_1 r×r e W_o n-r×n-r tais que

$$V\,0_1^t = 0_1^t\,W_1$$

e (3.47)

$$V\,0_o^t = 0_o^t\,W_o$$

mas, como $0_1\,0_1^t = I_r$,

$$0_o\,0_o^t = I_{n-r} \quad \text{e}$$

$$0_1\,0_o^t = 0_o\,0_1^t = \emptyset$$

então (3.47) é equivalente a

$$0_1\,V\,0_1^t = W_1$$

$$0_o\,V\,0_o^t = W_o \qquad (3.48)$$

$$0_1\,V\,0_o^t = 0_o\,V\,0_1^t = \emptyset$$

tomando $0 = (0_1^t : 0_o^t)$ então

$$0^t \; V \; 0 = \begin{pmatrix} 0_1 \; V \; 0_1^t & 0_1 \; V \; 0_o^t \\ 0_o \; V \; 0_1^t & 0_o \; V \; 0_o^t \end{pmatrix} \tag{3.49}$$

substituindo (3.48) em (3.49) temos

$$0^t \; V \; 0 = \begin{pmatrix} W_1 & \emptyset \\ \emptyset & W_o \end{pmatrix} \tag{3.50}$$

mas, com W_1 e W_o são matrizes simétricas então existem matrizes ortonormais

R_1 r×r e R_o n-r×n-r tais que

$$R_1 \; W_1 \; R_1^t = D_1$$

e (3.51)

$$R_o \; W_o \; R_o^t = D_o$$

onde D_1 e D_o são matrizes diagonais.

assim, se

$$S^t = \begin{pmatrix} R_1 & \emptyset \\ \emptyset & R_o \end{pmatrix} 0^t = \begin{pmatrix} R_1 & 0_1 \\ R_o & 0_o \end{pmatrix}$$

temos que S^t é uma matriz ortonormal e satisfaz

$$S^t \; V \; S = \begin{pmatrix} R_1 & \emptyset \\ \emptyset & R_o \end{pmatrix} \; 0^t \; V \; 0 \begin{pmatrix} R^t & \emptyset \\ \emptyset & R_o^t \end{pmatrix} = \text{por (3.50)}$$

$$= \begin{pmatrix} R_1 & \emptyset \\ \emptyset & R_o \end{pmatrix} \begin{pmatrix} W_1 & \emptyset \\ \emptyset & W_o \end{pmatrix} \begin{pmatrix} R_1^t & \emptyset \\ \emptyset & R_o^t \end{pmatrix} =$$

$$= \begin{pmatrix} R_1 \; W_1 \; R_1^t & \emptyset \\ \emptyset & R_o \; W_o \; R_o^t \end{pmatrix} = \text{por } (3.51)$$

$$= \begin{pmatrix} D_1 & \emptyset \\ \emptyset & D_o \end{pmatrix} \quad \text{que é uma matriz diagonal}$$

e portanto, as colunas de S formam um conjunto de auto vetores de V onde,

sendo $\quad S = (0_1^t \; R_1^t \; : \; 0_o^t \; R_o^t)$

então, os primeiros r auto vetores de V são combinações lineares de vetores da base de C(X) e portanto pertencem a C(X) e, os restantes n-r são combinações lineares de vetores da base de $C(X)^{\perp}$ e portanto pertencem a $C(X)^{\perp}$.

mas, como S^t é ortonormal $\rightarrow$

$$S^tS = S\,S^t = I_n$$

então $\quad S^tS = I_n \rightarrow (R_1 0_1)(0_1^t \; R_1^t) = I_r \quad \rightarrow$

as colunas de $0_1^t\ R_1^t$ são vetores ortonormais e portanto formam uma base ortonormal de C(X), e assim, concluímos que V possui r auto vetores ortogonais que formam uma base ortonormal para C(X).

Inversamente

Supondo que V possui r auto vetores ortogonais que formam uma base ortonormal para C(X)

então, V também possui n-r auto vetores ortogonais que formam uma base ortonormal para $C(X)^{\perp}$.

Seja $0 = (0_1 : 0_o)$ a matriz ortonormal cujas colunas de 0_1 e 0_o são formadas respectivamente por esses vetores das bases de C(X) e $C(X)^{\perp}$.

Assim, como

$$0^t\ V\ 0 = \begin{pmatrix} D_1 & \emptyset \\ \emptyset & D_o \end{pmatrix} \quad \text{onde } D_1 \text{ é uma matriz}$$

r×r diagonal e D_o é uma matriz

n-r × n-r diagonal temos

$$V = 0 \begin{pmatrix} D_1 & \emptyset \\ \emptyset & D_o \end{pmatrix} 0^t \leftrightarrow V = 0_1 D_1 0_1^t + 0_o^t D_o 0_o^t \tag{3.52}$$

daí, como para todo vetor $\omega \in C(X)$

temos

$$0_o^t\ \omega = \emptyset \tag{3.53}$$

então, aplicando V em ω temos por (3.52) e (3.53) que

$$V\omega = 0_1 D_o 0_1^t \omega + 0_o D_o 0_o^t \omega = 0_1 D_1 0_1^t \omega = 0_1 \phi$$

onde $\phi = D_1\ 0_1^t\ \omega \in R^r$

e portanto

$$V\omega \in C(X) \tag{3.54}$$

assim, como

$$\omega \in C(X) \rightarrow \omega^t y \text{ é SLSE da } E(\omega^t y)$$

e

$$V\omega \in C(X) \rightarrow \omega^t y \text{ é BLUE da } E(\omega^t y)$$

então, por (3.52) e (3.54) concluímos que

para toda função paramétrica linearmente estimável o seu SLSE e o seu BLUE são iguais.

COROLÁRIO 2.2 - Se ω é um vetor de C(X) e é uma combinação linear dos auto vetores de V pertencentes a C(X) então $\omega^t y$ são ambos SLSE e BLUE da $E(\omega^t y)$. (3.55)

PROVA - trivial

COROLÁRIO 2.3 - Para qualquer função paramétrica linearmente estimável, o seu BLUE e o SLSE são iguais se, e somente se,

$$V = X\ T_1\ X^t + Z\ T_2\ Z^t + a^2\ I_n \tag{3.56}$$

onde

T_1 e T_2 são matrizes p×p e n-r×n-r respectivamente e $a \in R$.

PROVA - Como, por (3.46) SLSE $\equiv$ BLUE $\rightarrow$

V possui r auto vetores ortogonais que formam uma base ortonormal para C(X) e portanto, V também possui n-r auto vetores que formam uma base para $C(X)^{\perp}$

então, temos que

$V = 0 \, D \, 0^t$ onde $0 = (0_1 : 0_o)$ é uma

matriz ortonormal cujas colunas de 0_1 e 0_o são formados respectivamente por esses vetores da base de C(X) e $C(X)^{\perp}$ e D é uma matriz n×n diagonal

assim, fazendo $D = a^2 \, I_n + Q$ onde $a \in R$ e Q é uma matriz diagonal temos que

$$V = 0(a^2 I_n + Q) \, 0^t =$$

$$= a^2 \, I_n + 0 \, Q \, 0^t \quad \text{pois} \quad 0 \, 0^t = I_n$$

e ainda, se $Q = \begin{pmatrix} Q_1 & \emptyset \\ \emptyset & Q_o \end{pmatrix}$.

onde

Q_1 e Q_o são matrizes diagonais r×r e n-r×n-r respectivamente então

$$V = a^2 \, I_n + 0_1 \, Q_1 \, 0_1^t + 0_o \, Q_o \, 0_o^t$$

e como, $C(0_1) = C(X) \rightarrow 0_1 = XC_1$ onde

C_1 é uma matriz p×r de posto r

e

$C(0_o) = C(X)^{\perp} = C(Z) \rightarrow 0_o = Z \, C_o$

onde C_o é uma matriz $n-r \times n-r$ de posto $n-r$

temos,

$$V = a^2 I_n + XC_1\ Q_1\ C_1^t\ X^t + Z\ C_o\ Q_o\ C_o^t\ Z^t =$$

$$= X\ T_1\ X^t + Z\ T_o\ Z^t + a^2\ I_n$$

onde

$T_1 = C_1\ Q_1\ C_1^t$ é uma matriz $p \times p$

$T_o = C_o\ Q_o\ C_o^t$ é uma matriz $n-r \times n-r$

e $\quad a \in R$

Inversamente, supondo

$$V = a^2\ I_n + X\ T_1\ X^t + Z\ T_o\ Z^t \quad \text{onde}$$

Z é uma matriz $n \times n-r$ tal que $C(Z) = C(X)^{\perp}$

$a \in R$

T_1 é uma matriz $p \times p$

T_o é uma matriz $n-r \times n-r$

temos,

como $\omega^t y$ é SLSE da $E(\omega^t y) \leftrightarrow \omega = X\rho$ para algum $\rho \in R^p$

e $\omega^t y$ é BLUE da $E(\omega^t y) \leftrightarrow Z^t V\omega = \emptyset$ (ver 3.36)

então, temos que para todo $\omega \in R^n$ tal que $\omega = X\rho$ para algum $\rho \in R^p$

$$\rightarrow Z^t\ V\ \omega = Z^t\ V\ X\rho = a^2\ Z^t\ X\rho + Z^t\ X\ T_1\ X^t\ X\rho +$$

$$+ Z^t Z\, T_o Z^t X\rho = \emptyset \qquad \text{pois} \quad Z^t X = \emptyset$$

e portanto SLSE ≡ BLUE para toda função paramétrica linearmente estimável.

3.3 - CARACTERIZAÇÃO DOS BLUE's NO MODELO DE G.M.G.

O problema presente, pode ser tratado por vários métodos como, por exemplo,

Rao (1971, 72, 75 e 78) apresenta uma Teoria unificada para a representação de BLUE's no Modelo de G.M.G. Nestes artigos, Rao demonstra a existência de uma matriz T tal que $C(T) = C(V:X)$ e conclui que o BLUE de qualquer função paramétrica linearmente estimável, $\lambda^t\beta$, é dado por $\lambda^t\,\hat{\beta}$ para todo $\hat{\beta}$ que satisfaz

$$X^t\, T^-\, X\beta \;=\; X^t\, T^-\,\beta$$

onde

T^- é uma g-inversa de T

Rao (1971, 72) apresenta também um método denominado Método da Inversa da Matriz Particionada

$$M = \begin{pmatrix} V & X \\ X^t & \emptyset \end{pmatrix}$$ obtida a partir da

minimização de $b^t\, V\, b$ restrito a $X^t b = \lambda$ pelo método do multiplicador de Lagrange e conclui que,

sendo $$M^- = \begin{pmatrix} C_1 & C_2 \\ C_3 & -C_4 \end{pmatrix}$$

uma g-inversa de M, então, o BLUE de $\lambda^t \beta$ é $\lambda^t \hat{\beta}$ onde

$$\hat{\beta} = C_2^t \, y \qquad \text{ou} \quad C_3 \, y$$

Rao (1973) apresenta um conjunto L_p tal que para todo $L \in L_p$ então

$$L^t \, y \;\; \text{é BLUE da sua esperança.}$$

Zyskind e Martin (1969) constroem uma sub-classe de g-inversa de V que depende da relação entre X e V e diz que,

se V^- é uma g-inversa de V que pertence a esta sub-classe então, o BLUE de $\lambda^t \beta$ é $\lambda^t \hat{\beta}$ para todo $\hat{\beta}$ tal que

$$X^t \, V^- \, X \, \hat{\beta} = X^t \, V^- \, y$$

Goldman e Zelen (1964) mostram que um modelo linear com matriz de cavariância singular pode ser transformado através de uma matriz ortonormal para um modelo equivalente ao modelo linear com restrição nos parâmetros e cuja matriz de variância e covariância dos erros é $\sigma^2 I$.

Neste trabalho, apresentaremos uma caracterização do BLUE de $X\beta$, sem relacionarmos com qualquer inversa generalizada de V (singular) encontrando, uma forma de melhorar a informação sobre $X\beta$ através de vetores de $C(X)^\perp$ usando o argumento apresentado por Rao (1967) denominado ajustamento por covariância.

Para este propósito, supondo

$$a^t y \;\; \text{o SLSE de} \;\; \lambda^t \beta = a^t X \beta \tag{3.57}$$

e

Z uma matriz n×n-r, r=posto(X), tal que $C(Z) = C(X)^{\perp}$, e portanto $Z^t X = \emptyset$ (3.58)

temos que

por (3.57) $E(a^t y) = a^t X\beta = \lambda^t \beta$ para todo $\beta \in R^p$ (3.59)

e por (3.58), para todo $\rho \in R^{n-r}$ tal que $Z\rho \in \sigma(Z) \rightarrow$

$$E(\rho^t Z^t y) = \rho^t Z^t X\beta = 0 \text{ para todo } \beta \in R^p. \tag{3.60}$$

assim, por (3.59) e (3.60) o novo funcional linear $a^t y - \rho^t Z^t y$ é tal que

$$E(a^t y - \rho^t Z^t y) = E(a^t y) - E(\rho^t Z^t y) = \lambda^t \beta \tag{3.61}$$

$$\text{para todo } \beta \in R^p$$

onde, sendo

$$var(a^t y - \rho^t Z^t y) = var(a^t y) - 2\, cov(a^t y, \rho^t Z^t y) +$$

$$+ var(\rho^t Z^t y) \tag{3.62}$$

temos que,

se

$$cov(a^t y, \rho^t Z^t y) \neq 0 \tag{3.63}$$

então

$$var(a^t y - \rho^t Z^t y) \leq var(a^t y) \leftrightarrow$$

$$0 \leq var\, (\rho^t Z^t y) \leq 2\, cov\, (a^t y, \rho^t Z^t y)$$

e neste caso, supondo

$$var(\rho^t Z^t y) = cov(a^t y, \rho^t Z^t y)$$

então (3.62) fica

$$var(a^ty - \rho^tZ^ty) = var(a^ty) - cov(a^ty, \rho^tZ^ty)$$

e portanto $\lambda^t\beta$ é melhor ajustado por

$$a^ty - \rho^tZ^ty$$

se $$cov(a^ty, \rho^tZ^ty) = 0 \qquad (3.64)$$

então (3.62) fica

$$var(a^ty - \rho^tZ^ty) = var(a^ty) + var(\rho^tZ^ty) \geq var(a^ty)$$

e portanto

$\lambda^t\beta$ é melhor ajustado por a^ty

e ainda

$$cov(a^ty, \rho^tZ^ty) = 0 \quad \text{para todo } \rho \in R^{n-r} \rightarrow$$

$$\rightarrow a^t V Z = \emptyset \qquad (3.65)$$

e portanto

(3.65) $\rightarrow$ (por(3.36)) que o BLUE de $\lambda^t\beta$ é o SLSE a^ty de $\lambda^t\beta$.

Mostrando que existe $\rho \in R^{n-r}$, $\rho \neq \emptyset$ tal que o BLUE de $\lambda^t\beta$ é dado por

$a^ty - \rho^tZ^ty$ temos,

Como, por (3.36) o BLUE de $\lambda^t\beta$ é

$$a^t y - \rho^t Z^t y \leftrightarrow \quad Z^t V(a-Z\rho) = \emptyset \quad \leftrightarrow$$

$$Z^t V Z\rho = Z^t V a \tag{3.66}$$

Sendo V uma matriz positiva e portanto

$$V = CC^t \text{ onde } C \text{ é uma matriz } n{\times}s \tag{3.67}$$

então, substituindo (3.67) em (3.66) temos

$$Z^t\ C\ C^t\ Z\rho \ = Z^t\ C\ C^t\ a \quad \leftrightarrow$$

$$(C^t Z)^t (C^t Z)\rho \ = (C^t Z)^t (C^t a) \tag{3.68}$$

Observando que (3.68) é um conjunto de equações normais que são sempre consistentes para todo $C^t a \in R^s$
então, podemos concluir que existe

$$\rho \neq \emptyset,\ \rho \in R^{n-r} \text{ tal que } \rho \text{ satisfaz (3.66)} \tag{3.69}$$

e ainda, como por (2.12)

$$C^t\ Z\ \rho_1 = C^t Z\ \rho_2 \text{ para todo } \rho_1 \text{ e } \rho_2 \in R^{n-r}$$

tal que ρ_1 e ρ_2 satisfazem (3.66)

temos que

$$C^t\ Z\ \rho_1 = C^t\ Z\ \rho_2 \rightarrow a^t\ V\ Z\ \rho_1 = a^t\ V\ Z\ \rho_2$$

ou seja

$$\mathrm{cov}(a^t y,\ \rho_1^t\ Z^t\ y) = \mathrm{cov}(a^t y,\ \rho_2^t\ Z^t\ y)$$

para todo ρ_1 e $\rho_2 \in R^{n-r}$ tal que satisfazem (3.66) (3.70)

e portanto, para todo $\rho \in R^{n-r}$ tal que ρ satisfaz (3.66) temos que

$$\begin{aligned} & var(a^t y) - var(a^t y - \rho^t Z^t y) = \\ & = 2\ cov(a^t y, \rho^t Z^t y) - var(\rho^t Z^t y) = \\ & = 2\ cov(a^t y, \rho^t Z^t y) - \rho^t Z^t V Z \rho \sigma^2 = \\ & = 2\ cov(a^t y, \rho^t Z^t y) - \rho^t Z^t V a \sigma^2 = \\ & = 2\ cov(a^t y, \rho^t Z^t y) - cov(a^t y, \rho^t Z^t y) = \\ & = cov(a^t y, \rho^t Z^t y) = \sigma^2 a^t V Z \rho = \sigma^2 \rho^t Z^t V Z \rho \geq 0 \end{aligned}$$

ou seja,

$$\begin{aligned} 0 & \leq var(a^t y) - var(a^t y - \rho^t Z^t y) = \\ & = cov(a^t y, \rho^t Z^t y) \leq var(a^t y) \end{aligned} \tag{3.71}$$

assim (3.71) $\rightarrow$

$$0 \leq var(a^t y - \rho^t Z^t y) = var(a^t y) - cov(a^t y, \rho^t Z^t y)$$

e $$var\ (a^t y - \rho^t Z^t y) \leq var\ (a^t y) \tag{3.72}$$

e portanto, por (3.69), (3.70) e (3.72) concluímos que existe $\rho \neq 0$, $\rho \in R^{n-r}$ tal que satisfaz

$$Z^t V(a - Z\rho) = 0$$

e, para todo ρ satisfazendo a relação acima temos que

$a^t y - \rho^t Z^t y$ será efetivamente o BLUE de $\lambda^t\beta = a^t X\beta$ quando passamos de $V=I_n$ para V singular com a variância ajustada pela $cov(a^t y, \rho^t Z^t y)$.

Da mesma forma, temos

TEOREMA 3.1 - Sob as condições do modelo de G.M.G. o BLUE de $X\beta$ é dado por $A^t y$ onde,

$$A = I - ZB \tag{3.73}$$

onde Z é uma matriz $n \times n-r$ tal que $C(Z) = C(X)^{\perp}$ e

B é uma matriz tal que satisfaz $Z^t VZB = Z^t V$

ou

$$A = P_X - (I-P_X)\,U \tag{3.74}$$

onde P_X é a matriz do operador projeção ortogonal sobre $C(X)$ e U é uma matriz tal que satisfaz

$$(I-P_X)\ V\ (I-P_X)\ U = (I-P_X)\ V\ P_X$$

PROVA - Vimos por (3.68) que

$Z^t\ V\ Z\ \rho = Z^t\ V\ a = Z^t C(C^t a)$ é consistente para todo vetor $C^t a \in R^s$.

Assim, tomando $C^t a_{(i)}$ = i-ésima coluna de C^t $i=1,\ldots,n$

então, existe uma matriz $B = (\rho_{(i)})$
onde $\rho_{(i)} \in R^{n-r}$; $\rho_{(i)} \neq \emptyset$ $i=1,\ldots,n$

satisfazendo

$$Z^t\ V\ Z\ B = (Z^t\ V\ Z\ \rho_{(i)})$$

$$= (Z^t C\ C^t\ a_{(i)}) = Z^t C\ (C^t a_{(i)}) = Z^t C\ C^t = Z^t V \quad (3.75)$$

assim, se

$$A = I - Z\ B$$

temos que

$$X^t\ A = X^t - X^t\ Z\ B = X^t$$

e

$$V\ A = V - V\ Z\ B \quad (3.76)$$

mas, como por (3.75) temos $Z^t(V\ Z\ B - V) = \emptyset$

e como

$$Z^t(V\ Z\ B - V) = \emptyset \leftrightarrow V\ Z\ B - V = X\ M \text{ para}$$

alguma matriz $M_{p\times n}$

$$\leftrightarrow \quad V\ Z\ B = V + X\ M \quad (3.77)$$

então, substituindo (3.77) em (3.76) temos

$$X^t\ A = X^t$$

e

$$V\ A = -X\ M \text{ para alguma matriz } M_{p\times n}$$

e portanto, por (3.40) concluímos que

$A^t y$ é o BLUE de $X\beta$ onde

$A = I - ZB$ para toda matriz B tal que

$$Z^t\ V\ Z\ B = Z^t V$$

Considerando, P_X a matriz do operador projeção ortogonal sobre C(X) temos por (1.55) que $I-P_X$ é a matriz do operador projeção ortogonal sobre $C(X)^\perp$

assim tomando $Z = I-P_X$ (3.75) fica

$$(I-P_X)\ V(I-P_X)B = (I-P_X)V \tag{3.78}$$

multiplicando à direita de (3.78) por P_X e fazendo $BP_X = U$ temos

$$(I-P_X)V(I-P_X)U = (I-P_X)V\ P_X \tag{3.79}$$

onde, tomando $A = P_X - (I-P_X)U$

obtemos que

$$X^t\ A = X^tP_X - X^t(I-P_X)U = (P_X\ X)^t = X^t$$

e por (3.79)

$$(I-P_X)V\ A = (I-P_X)V\ P_X-(I-P_X)V(I-P_X)U = \emptyset \tag{3.80}$$

mas, como $C(I-P_X) = C(X)^\perp = C((I-P_X)^t)$

então $(I-P_X)\ V\ A = \emptyset\ \rightarrow\ V\ A = X\ M$ para alguma matriz $M_{p\times n}$

e portanto

$$X^t\ A = X^t$$

e $$V\ A = X\ M \text{ para alguma matriz } M_{p\times n}$$

concluímos por (3.40) que

A^ty é BLUE $X\beta$ onde

A é dada por (3.74)

PROPRIEDADE 3.1 - Sob as condições do modelo de G.M.G.

Sendo $A_1 = I-ZB$

e

$A_2 = P_x - (I-P_x)U$

dadas respectivamente por (3.73) e (3.74)

então, $A_1^t\, y = A_2^t\, y$ com probabilidade 1 (3.81)

PROVA - Como A_1 e A_2 dadas respectivamente por (3.73) e (3.74) são tais que

$A_1^t y$ e $A_2^t y$ são duas representações do BLUE de $X\beta$

então,

$$E(A_1^t\, y - A_2^t\, y) = E(A_1^t\, y) - E(A_2^t\, y) = \emptyset \qquad (3.82)$$

e ainda,

como

$$A_1^t - A_2^t = (I-ZB)^t - (P_x+(I-P_x)U)^t =$$

$$= (I-U)^t(I-P_x)-B^tZ^t$$

onde $C(I-P_x) = C(Z) = C(X)^{\perp}$

então $C(A_1^t - A_2^t) \subset C(Z^t) = C(I-P_x)$ (3.83)

e $$V(A_1-A_2) = V\,A_1 - V\,A_2 = XM_1 - XM_2 =$$

$$= X(M_1 - M_2)$$

$$\rightarrow \quad C(V(A_1-A_2)) \subset C(X) \qquad (3.84)$$

então, por (3.83) e (3.84) temos

$$\mathrm{var}((A_1^t - A_2^t)y) = \sigma^2 (A_1^t - A_2^t)\ V(A_1 - A_2)\ \emptyset \tag{3.85}$$

e portanto por (3.82) e (3.85) concluímos que

$$A_1^t y = A_2^t y \quad \text{com probabilidade 1.}$$

PROPRIEDADE 3.2 - Sob as condições do modelo de G.M.G.

Se $\quad A_1 = P_x - (I - P_x)U_1$

e

$\quad A_2 = P_x - (I - P_x)U_2$

onde P_x é a matriz do operador projeção ortogonal sobre C(X) e

U_1 e U_2 são duas soluções de

$$(I - P_x)V\ (I - P_x)U = (I - P_x)\ V\ P_x$$

então,

$$A_1^t y = A_2^t y \text{ com probabilidade 1} \quad (3.85)$$

PROVA - Como por (3.74) temos que

$A_1^t y$ e $A_2^t y$ são duas representações do BLUE de $X\beta$

então,

$$E(A_1^t y - A_2^t y) = E(A_1^t y) - E(A_2^t y) = \emptyset \tag{3.86}$$

e ainda, como

$A_1 - A_2 = (I - P_x)\ (U_1 - U_2)$ então

$$var((A_1^t - A_2^t)y) = (U_1 - U_2)^t (I-P_x)\ V(A_1-A_2)$$

mas, como $C(I-P_x) = C(X)^{\perp}$

então, por (3.36) temos que

$$(I-P_x)V\ (A_1-A_2) = (I-P_x)\ V\ A_1-(I-P_x)\ V\ A_2 = \emptyset$$

e portanto $var(A_1^t y - A_2^t y) = \emptyset$ (3.87)

ou seja,

por (3.86) e (3.87) temos que

$$A_1^t y = A_2^t y \quad \text{com probabilidade 1.}$$

Assim, por (3.81) e (3.85) apesar da falta da unicidade da representação do BLUE de $X\beta$, concluímos que a estimativa de $X\beta$ é sempre a mesma.

Rao (1973) afirma que

COROLÁRIO 3.1 - Sob a condição do modelo de G.M.G. temos que se

$$A_1 = (I - Z\ B)$$

e

$$A_2 = P_x - (I-P_x)U$$

como definidas por (3.73) e (3.74) respectivamente

então, se

posto $(V:X) = n$ teremos $A_1 = A_2$ (3.88)

ou seja,

existe uma única representação do BLUE de X no modelo de G.M.G.

PROVA - Como, por (3.4) observamos que

$$C(V:X) = C(X:VZ) = C(X) \oplus C(VZ)$$

então, se posto $(V:X) = n \rightarrow$

$\rightarrow$ posto $(X:VZ) = n$

e como por (3.7) posto $(X:VZ) = \dim (C(X:VZ)) =$

$= \dim C(X) + \dim (C(VZ)) = \text{posto}(X) + \dim (C(Z^tV)) =$

$= \text{posto } X + \text{posto } (Z^tV) =$ (por 1.33) $\text{posto}(X)+\text{posto}(Z)+$

$- \dim(C(Z) \cap C(V)^\perp) = r+n-r - \dim(C(Z) \cap C(V)^\perp) =$

$= n - \dim (C(Z) \cap C(V)^\perp) = n \leftrightarrow$

$\dim (C(Z) \cap C(V)^\perp) = 0 \leftrightarrow C(Z) \cap C(V)^\perp = \emptyset$

e como A_1 e A_2 definidas respectivamente (3.73) e (3.74) são tais que

$$X^t(A_1-A_2)=\emptyset \leftrightarrow C(A_1-A_2) \subset C(Z)$$

e, por (3.81)

$$\text{var}((A_1-A_2)^ty) = (A_1-A_2)^tV(A_1-A_2)^{\;2} = \emptyset \leftrightarrow$$

$$V(A_1-A_2) = \emptyset \text{ pois V é uma matriz positiva}$$

e, $V(A_1-A_2)=\emptyset \leftrightarrow C(A_1-A_2) \subset C(V)^1$

temos $C(A_1-A_2) \subset C(Z) \cap C(V)^1 = \emptyset \leftrightarrow$

$$C(A_1 - A_2) = \emptyset \leftrightarrow A_1 = A_2$$

Assim por (3.88) observamos que mesmo quando V é singular é possível obtermos uma única representação para o BLUE de $X\beta$.

PROPRIEDADE 3.3 - Sob as condições do modelo de G.M.G. sendo a BLUE de $X\beta$ dado por $A^t y$ onde

$$A = P_x - (I-P_x)U$$

como definida em (3.74)

então

$$E(A^t y) = X\beta$$

e

$$var(A^t y) = var(P_x y) - cov(P_x y, U^t(I-P_x)y) \qquad (3.89)$$

onde,

$cov(P_x y, U^t(I-P_x)y)$ independe da solução U e também é uma matriz positiva (3.90)

PROVA - Sendo $A = P_x-(I-P_x)U$

onde U é tal que

$$(I-P_x)V(I-P_x)U = (I-P_x)\ V\ P_x \qquad (3.91)$$

então, multiplicando à esquerda de (3.91) por U^t temos,

$$U^t(I-P_x)\ V(I-P_x)U = U^t(I-P_x)\ VP_x$$

ou seja,

$$var(U^t(I-P_x)y) = cov(U^t(I-P_x)y,\ P_xy) \qquad (3.92)$$

e portanto,

$cov(U^t(I-P_x)y,\ P_xy) = cov(P_xy,\ U^t(I-P_x)y)$ é

sempre uma matriz positiva para toda solução U de (3.91).

Mostrando que $cov(P_xy,\ U^t(I-P_x)y)$ independe da solução de (3.91) temos;

como V é positiva $\rightarrow$ $V = C\,C^t$ para alguma matriz $C_{n\times s}$.

temos então que

(3.91) é equivalente ao conjunto de equações.

$$[C^t(I-P_x)]^t[C^t(I-P_x)]U = [C^t(I-P_x)]^t(C^tP_x) \qquad (3.93)$$

observando que este conjunto de equações forma um conjunto de n equações lineares cuja estrutura coincide com o da equações normais que são sempre consistentes e também

$C^t(I-P_x)U$ é invariante para toda solução desse conjunto de sistema de equações então temos que

$$C^t(I-P_x)\ U_1 = C^t(I-P_x)\ U_2 \qquad (3.94)$$

para toda matriz U_1 e U_2 tais que satisfazem (3.93).

Assim, como $C^t(I-P_x)\ U_2 = C^t(I-P_x)\ U_1 \leftrightarrow$

$$U_2^t\ (I-P_x)\ C = U_1^t\ (I-P_x)\ C \qquad (3.95)$$

então, por (3.94) e (3.95) temos que

$$U_2^t\,(I-P_x)\;V(I-P_x)U_2 = U_1^t(I-P_x)\;V(I-P_x)\;U_2 =$$

$$= U_2^t(I-P_x)\;V(I-P_x)\;U_1 = U_1^t(I-P_x)\;V(I-P_x)\;U_1 \qquad (3.96)$$

e portanto, por (3.92) e (3.96) temos

$$\mathrm{cov}(P_x y,\; U_1^t(I-P_x)y) = \mathrm{cov}(P_x y,\; U_2^t(I-P_x)y)$$

para toda solução U_1 e U_2 de (3.91)

Mostrando agora que

$$\mathrm{var}(A^t y) = \mathrm{var}(P_x y) - \mathrm{cov}(P_x y,\; {}^t(I-P_x)y)$$

temos,

como $A^t y = P_x y - U^t(I-P_x)y$ onde U satisfaz (3.91)

então,

$$\mathrm{var}(A^t y) = \mathrm{var}(P_x y) - 2\mathrm{cov}(P_x y, U^t(I-P_x)y) + \mathrm{var}(U^t(I-P_x)y)$$

e, por (3.92) temos

$$\mathrm{var}(A^t y) = \mathrm{var}(P_x y) - 2\mathrm{cov}(P_x y, U^t(I-P_x)y) + \mathrm{cov}(P_x y, U^t(I-P_x)y)$$

ou seja,

$$\mathrm{var}(A^t y) = \mathrm{var}(P_x y) - \mathrm{cov}(P_x y, U^t(I-P_x)y)$$

Assim, sendo o SLSE de $X\beta$ dado por P_Xy e o BLUE de X dado por $A^ty = P_Xy - U^t(I-P_X)y$ que podemos considerar como

$$P_Xy - (\text{perturbação em } (I-P_X))$$

notamos que, como

$$cov(P_Xy,\ U^t(I-P_X)y) = \sigma^2 P_X\ V(I-P_X)\ U \text{ é}$$

positiva e independe de U

então, existem vetores em $C(I-P_X)U= C(X)^{\perp}$ da forma $(I-P_X)\ U$ tais que não pertencem ao núcleo de V e, a imagem desses vetores através de V são projetados por P_X em vetores não nulos de $C(X)$;

Assim, na estimação de $X\beta$ no modelo de G.M.G. há uma contribuição importante associada à perturbação $(I-P_X)U$ onde, sendo a

$$cov(P_Xy, U^t(I-P_X)y) = var(P_Xy)-var(A^ty)$$

uma matriz positiva então, temos que realmente o ajustamento de P_Xy por $U^t(I-P_X)y$ reduz a variância de P_Xy.

Assim, a variância de SLSE de $X\beta$ com $V = I_n$ é $var(P_Xy) = \sigma^2P_X$, e quando passamos a V singular o BLUE de $X\beta$ é dado por $A^ty = P_Xy - U^t(I-P_X)y$ com a

$var(P_Xy) = \sigma^2\ P_XVP_X$ ajustada pela covariância de

$(P_Xy,\ U^t(I-P_X)y)$

COROLÁRIO 3.2 - Sob as condições do modelo de G.M.G. consideran do

$A^t y = P_x y - U^t(I-P_x)y$ como dado em (3.74) o BLUE de $X\beta$

então,

$$\text{SLSE} \equiv \text{BLUE se, e somente se, } (I-P_x)VP_x = \emptyset \qquad (3.97)$$

PROVA - Sendo o SLSE de $X\beta$ dado por $P_x y$ e o

BLUE de $X\beta$ dado por

$$A^t y = P_x y - U^t(I-P_x)y \quad \text{onde}$$

U satisfaz $(I-P_x)V(I-P_x)U = (I-P_x)VP_x$

então, o

SLSE $\equiv$ BLUE se e somente se

$$\text{var}(A^t y) = \text{var}(P_x y) \leftrightarrow \text{(por (3.89))}$$

$$\text{var}(P_x y) - \text{cov}(P_x y, U^t(I-P_x)y) = \text{var}(P_x y) \leftrightarrow$$

$$\text{cov}(P_x y,\ U^t(I-P_x)y) = \sigma^2 P_x\ V\ (I-P_x)U = \emptyset$$

$$\leftrightarrow \text{(por (3.92))}\ U^t(I-P_x)V(I-P_x)U = \emptyset \qquad (3.98)$$

mas, como $V = C\ C^t$ para alguma matriz $C_{n\times s}$

então (3.98) fica

$$[C^t(I-P_x)U]^t[C^t(I-P_x)U] = \emptyset \quad \leftrightarrow$$

$$C^t(I-P_x)U = \emptyset \leftrightarrow V(I-P_x)U = \emptyset$$

mas, como $V(I-P_X)U = \emptyset \rightarrow (I-P_X)\ V(I-P_X)U = \emptyset$

então, temos que $(I-P_X)\ V\ P_X = \emptyset$

Inversamente, se $P_X\ V(I-P_X) = \emptyset$ temos

$$\sigma^2\ P_X\ V(I-P_X)U = \emptyset \quad \text{e por (3.92)}$$

$$\sigma^2\ U^t(I-P_X)\ V(I-P_X)U = \emptyset$$

ou seja $\qquad var(U^t(I-P_X)y) = \emptyset$

assim, como $E(U^t(I-P_X)y) = U^t(I-P_X)X\beta = \emptyset$

para todo $\beta \in R^p$

e $\qquad var(U^t(I-P_X)y) = \emptyset$

então $U^t(I-P_X)y = \emptyset$ com probabilidade 1

e portanto

$$A^ty = P_Xy + U^t(I-P_X)y = P_Xy$$

ou seja, SLSE $\equiv$ BLUE

COROLÁRIO 3.3 - Sob as condições do modelo de G.M.G. SLSE $\equiv$ BLUE se e somente se $VP_X = P_XV$ (3.99)

onde P_X é a matriz do operador projeção ortogonal sobre C(X).

PROVA - Como, por (3.97) temos

$$\text{SLSE} \equiv \text{BLUE} \leftrightarrow P_XV(I-P_X) = \emptyset \quad \text{então}$$

$$SLSE \equiv BLUE \leftrightarrow P_x\ V\ (I-P_x) = \emptyset \leftrightarrow$$

$$P_xV = P_x\ VP_x = (P_x\ VP_x)^t = (P_xV)^t = V\ P_x$$

COROLÁRIO 3.4 - Sob as condições do modelo de G.M.G.

$$SLSE \equiv BLUE \leftrightarrow V = X\ T_1X^t + (I-P_x)T_2(I-P_x) + a^2I_n$$

onde T_1 e T_2 são matrizes p×p e n×n respectivamente, a∈ R e P_x é a matriz do operador projeção ortogonal (3.100)

PROVA - Como, por (3.56) vemos que

$$SLSE \equiv BLUE \leftrightarrow V = X\ Q_1\ X^t + X\ Q_2\ Z^t + a^2\ I_n \qquad (3.101)$$

onde Q_1 e Q_2 são matrizes p×p e n-r×n-r respectivamente, a ∈ R e Z é uma matriz n×n-r tal que $C(Z) = C(X)^{\perp}$

então, sendo P_x a matriz do operador projeção ortogonal sobre C(X) temos, por (1.55) que $(I-P_x)$ é a matriz do operador projeção ortogonal sobre $C(X)^{\perp}$.

e portanto, sendo $C(I-P_x) = C(X)^{\perp} = C(Z)$

então, existe uma matriz $F_{n\times n-r}$ de posto n-r tal que $(I-P_x)F=Z$ (3.102)

assim, substituindo (3.102) em (3.101) e lembrando que

$(I-P_x) = (I-P_x)^t$ temos

$$SLSE \equiv BLUE \leftrightarrow V = XQ_1X^t + (I-P_x)FT_2F^t(I-P_x) + a^2I_n$$

$$\leftrightarrow V = XT_1X^t + (I-P_x)T_2(I-P_x) + a^2I_n$$

onde $T_1 = Q_1$, $Q_2 = F\ T_2\ F^t$ são matrizes p×p e n×n respectivamente e a $\in$ R.

Considerando, novamente o BLUE de $X\beta$ dado por $A^t y$ onde

$$A = I-ZB \text{ é definida em } (3.73)$$

e, definindo a matriz $C_{n\times n}$ por

$C = A\ P_x$ onde P_x é a matriz do operador projeção ortogonal sobre $C(X)$

temos,

TEOREMA 3.2 - Sob a condição do modelo de G.M.G. definindo a matriz

$$C = A\ P_x \tag{3.103}$$

onde, P_x é a matriz do operador projeção ortogonal sobre $C(X)$

$$A = I - ZB \quad \text{onde}$$

B satisfaz $Z^t\ V\ Z\ B = Z^t\ V$

e $Z_{n\times n-r}$ e tal que $C(Z) = C(X)^{\perp}$

então

$$C^t y \text{ é um BLUE de } X\beta \tag{3.104}$$

e $\quad C^t y = A^t y$ com probabilidade 1

PROVA - Como $A = I-ZB$ definida em (3.73) é tal que por (3.76)

$$X^t\ A = X^t$$

e $\qquad V\,A = XM$ para alguma matriz $M_{p\times n}$

então $C = A\,P_x \rightarrow C = P_x - Z\,B\,P_x$ é tal que

$$X^tC = X^t\,A\,P_x = X^tP_x = (P_xX)^t = X^t$$

e

$$VC = V\,A\,P_x = X\,M\,P_x = X\,Q \tag{3.105}$$

onde $a = M\,P_x$ é uma matriz $p\times n$ e portanto por (3.40)

temos que C^ty é também um BLUE de XB

mostrando que $A^ty = C^ty$ com probabilidade 1 temos,

a. como $E(A^ty) = X\beta = E(C^ty)$

então $E(A^ty - C^ty) = E(A^ty) - E(C^ty) = \emptyset$

b. como

$$A-C = I-ZB - P_x + Z\,B\,P_x =$$

$$= (I-P_x)\,(I-ZB)$$

e

$$(A^t-C^t)V = A^tV - C^tV = M^tX^t-Q^tX^t =$$

$$= (M^t - Q^t)\,X^t$$

então

$$var(A^ty-C^ty) = var((A^t-C^t)y) =$$

$$= (A^t-C^t)\,V\,(A-C)\,\sigma^2 =$$

$$= (M^t-Q^t)X^t(I-P_x)(I-ZB)\,\sigma^2 = \emptyset$$

e portanto, por a. e b. concluímos que

$$A^t y = C^t y \quad \text{com probabilidade } 1$$

Observando com mais detalhes a matriz

$$C = P_X - Z\ B\ P_X \quad \text{onde}$$

B é tal que $Z^t\ V\ Z\ B = Z^t V$

temos que

$$C^2 = (P_X - Z\ B\ P_X)(P_Z - Z\ B\ P_X) =$$

$$= P_X^2 - P_X\ Z\ B\ P_X - Z\ B\ P_X^2 + ZBP_X\ Z\ B\ P_X =$$

$$= P_X - Z\ B\ P_X = C$$

mas, como $C^2 = C \leftrightarrow (C^t)^2 = C^t$ (3.106)

e C^t é tal que por (3.105)

$$C^t\ X = X$$

e $VC = X\ Q$ para alguma matriz Q

$$\rightarrow C^t X = X \text{ e } Z^t\ V\ C = C^t\ V\ Z = \emptyset \qquad (3.107)$$

então, concluímos por (3.106) e (3.107) que

$C^t y$ é a matriz do operador projeção

sobre C(X) segundo a direção de vetores de C(V Z) onde, Z é uma matriz n×n-r tal que $C(Z) = C(X)^{\perp}$.

Deste modo, provamos que

COROLÁRIO 3.5 - Sob a condição do modelo de G.M.G. um BLUE de $X\beta$ é dado pela projeção $C^t y$ sobre $C(X)$ segundo a direção de vetores de $C(VZ)$ (3.108)

onde $$C = P_x - Z\,B\,P_x$$

onde B satisfaz $Z^t\,V\,Z\,B = Z^t\,V$

Z é uma matriz $n \times n-r$ tal que $C(Z) = C(X)^{\perp}$

e P_x é a matriz do operador projeção ortogonal sobre $C(X)$

PROPRIEDADE 3.4 - Sob as condições do modelo de G.M.G., o BLUE de $X\beta$ dado por

$C^t y$ onde C é definida em (3.103) é tal que

$$E(C^t y) = X\beta$$

e
$$\mathrm{var}(C^t y) = \mathrm{var}(P_x y) - \mathrm{cov}(P_x y,\ P_x\,B^t\,Z^t y) \qquad (3.109)$$

$\mathrm{cov}(P_x y,\ P_x\,B^t\,Z^t\,y)$ é uma matriz positiva e é única para todo B tal que

$$Z^t\,V\,Z\,B = Z^t\,V$$

PROVA - Mostrando que $\mathrm{cov}(P_x y,\ P_x\,B^t\,Z^t\,y)$ é uma matriz positiva e é única para todo B tal que $Z^t\,V\,Z\,B = Z^t\,V$ temos, como $\quad cov(P_x y,\ P_x\ B^t\ Z^t y) =$

$$= \sigma^2 P_x\ V\ Z\ B\ P_x = \sigma^2\ P_x (Z^t\ V\ Z\ B)^t\ B\ P_x =$$

$$= \sigma^2\ P_x\ B^t\ Z^t\ V\ Z\ B\ P_x = var(P_x\ B^t\ Z^t\ y) \tag{3.110}$$

e sendo a $var(P_x\ B^t\ Z^t\ y)$ uma matriz positiva então a

$cov(P_x\ y),\ P_x\ B^t\ Z^t\ y)$ é uma matriz positiva.

e ainda, como

$$Z^t\ V\ Z\ B = Z^t\ V \tag{3.111}$$

onde $V = S\ S^t$ para alguma matriz $S_{n \times s}$ temos que (3.111) é equivalente ao sistema de equações

$$(S^t Z)^t\ (S^t Z)\ B = (S^t Z)\ S^t \tag{3.112}$$

Observando que (3.112) é um conjunto de n equações lineares cuja estrutura coincidem com o da equações normais, que é sempre consistente e também, se B_1 e B_2 são duas soluções desse sistema então

$$S^t\ Z\ B_1 = S^t\ Z\ B_2 \quad \text{temos}$$

$$S^t\ Z\ B_1 = S^t\ Z\ B_2 \ \rightarrow\ V\ Z\ B_1 = V\ Z\ B_2 \ \rightarrow$$

$$\rightarrow\ P_x\ V\ Z\ B_1 = P_x\ V\ Z\ B_2$$

$$\rightarrow\ P_x\ V\ B_1\ P_x = P_x\ V\ Z\ B_2\ P_x$$

$$\rightarrow\ cov(P_x y,\ P_x B_1^t Z^t y) = cov(P_x y, P_x B_2^t Z^t y)$$

onde, B_1 e B_2 são duas soluções quaisquer de (3.111).

Mostrando que

$$var(C^t y) = var(P_X y) - cov(P_X y,\ P_X\ B^t\ Z^t y)$$

temos,

como $$C^t y) = P_X y - P_X\ B^t\ Z^t\ y \quad \text{então}$$

$$var(C^t y) = var(P_X y) - 2\ cov(P_X y,\ P_X\ B^t\ Z^t\ y) +$$
$$+\ var(P_X\ B^t\ Z^t\ y)$$

e como por (3.110)

$$cov(P_X y,\ P_X\ B^t\ Z^t\ y) = var\ (P_X\ B^t\ Z^t\ y)$$

então,

$$var(C^t y) = var(P_X y) - cov(P_X y,\ P_X\ B^t\ Z^t y)$$

Assim, sendo o SLSE de $X\beta$ dado por $P_X y$ e o BLUE de $X\beta$ dado por

$$C^t y = P_X y - P_X\ B^t\ Z^t y$$

que podemos considerar como

$$P_X y - (\text{perturbação em } Z) \text{ onde } Z \text{ é tal que } C(Z) = C(X)^{\perp}$$

notamos que, como

$$cov(P_X y,\ P_X\ B^t\ Z^t y) = \sigma^2\ P_X\ V\ Z\ B\ P_X \text{ é}$$

única e também positiva

então, existem vetores da forma $Z\ B\ P_X$ em $C(Z)$ tais que a imagem desses vetores através de V são projetados por P_X em vetores não nulos de C(X) e ainda, como $cov(P_Xy, P_XB^tZ^ty) = var(P_Xy) - var(C^ty)$ é uma matriz positiva temos que realmente o ajustamento de P_Xy por $P_XB^tZ^ty$ reduz a variância de P_Xy.

Assim, a variância do SLSE de $X\beta$ com $V = I_n$ é $var(P_Xy) = \sigma^2P_X$ e quando passamos a V singular o BLUE de $X\beta$ é dado por $P_Xy - P_XB^tZ^ty$ com a $var(P_Xy) = \sigma^2P_XVP_X$ ajustada pela covariância de

$(P_Xy,\ P_X\ B^t\ Z^t\ y)$.

No teorema 3.1 observamos que sendo $A = P_X - (I-P_X)U$ onde

$U = B\ P_X$ e

satisfaz $(I-P_X)\ V\ (I-P_X)U\ =\ (I-P_X)\ V\ P_X$

então, A^ty é um BLUE de $X\beta$ e satisfaz

$$A^tX = X$$

e $$A^t\ V\ Z = \emptyset \quad \text{onde } Z \text{ é tal que } C(Z) = C(X)^{\perp}$$

Assim, notando que

$$A^2 = (P_X - (I-P_X)\ B\ P_X)(P_X - (I-P_X)\ B\ P_X) =$$

$$= P_X^2 - P_X\ (I-P_X)\ B\ P_X - (I-P_X)\ B\ P_X^2 +$$

$$+ (I-P_X)\ B\ P_X(I-P_X)B\ P_X = P_X-(I-P_X)\ B\ P_X = A$$

então, temos que $(A^t)^2 = A^t$ e portanto,

$A^t = P_X - U^t(I-P_X)$ é também uma matriz do operador projeção sobre C(X) segundo a direção dos vetores de C(VZ).

Mas, como o operador projeção sobre C(X) segundo a direção de vetores de C(VZ) é único, concluímos que

para todo $y \in C(V:X) = C(X) \oplus C(VZ)$ então

$$A^t y = C^t y$$

onde $$C = P_X - Z\,B\,P_X$$

para B satisfazendo $Z^t\,V\,Z\,B = Z^t\,V$

e portanto, temos

COROLÁRIO 3.6 - Sob a condição do modelo de G.M.G. sendo P_X a matriz do operador projeção ortogonal sobre C(X)

$Z_{n\times n-r}$ uma matriz tal que $C(Z) = C(X)^{\perp}$

então, se

$$A = P_X - (I-P_X)U$$

onde $U = BP_X$ satisfaz

$$(I-P_X)\,V(I-P_X)U = (I-P_X)\,VP_X$$

e

$$C = P_X - Z\,B\,P_X$$

onde B satisfaz $Z^t\,V\,Z\,B = Z^t\,y$

temos que $A^t y = C^t y$ = projeção de y em C(X) segundo a direção de vetores de C(VZ)

e ainda

$$A^t y = C^t y \text{ é um BLUE de } X\beta.$$

COROLÁRIO 3.7 - Se no modelo linear $y = X\beta + e$ temos que

$$E(e\ e^t) = \sigma^2\ I_n$$

então, o BLUE de X é dado por

$$P_x y = \text{projeção ortogonal de } y \text{ em } C(X)$$

PROVA - Sendo, o BLUE de $X\beta$ dado por

$$A^t y = P_x y - U^t(I-P_x)y = \text{projeção de } y \text{ em } C(X)$$

segundo a direção de vetores de C(VZ) onde U é tal que
$(I-P_x)\ V(I-P_x)U = (I-P_x)\ V\ P_x$
então, para $V = I$ temos,

$$(I-P_x)V(I-P_x)U = (I-P_x)^2\ U = (I-P_x)\ U =$$

$$= (I-P_x)P_x = \emptyset$$

assim, $A^t y = P_x y$ = projeção ortogonal de y sobre C(X).

COROLÁRIO 3.8 - Se no modelo linear $y = X\beta + e$ temos que
V é não singular
então, o BLUE de $X\beta$ é dado por

$$A^t y = M^t\ X^t\ V^{-1} y \quad \text{onde } M \text{ é tal que } X^t\ V^{-1}\ X\ M = X^t$$

e

$A^t y$ = projeção de y sobre C(X) segundo a direção de vetores de C(VZ).

PROVA - como o BLUE de $X\beta$ é dado por

$A^t y = P_x y - U^t (I-P_x) y$ = projeção de y sobre C(X) segundo a direção dos vetores de C(VZ)

onde $(I-P_x)\ V(I-P_x)U = (I-P_x)\ V\ P_x$

então, se V é não singular temos

$$(I-P_x)\ V\ (I-P_x)U = (I-P_x)\ V\ P_x \quad \leftrightarrow$$

$$\leftrightarrow \quad (I-P_x)\ [\ V(I-P_x)U - VP_x] = \emptyset \quad \leftrightarrow$$

$$V(I-P_x)U - VP_x = -XM \text{ para alguma matriz } M_{p\times n}$$

$$\leftrightarrow \quad V(I-P_x)U = VP_x - XM \leftrightarrow$$

$$(I-P_x)U = P_x - V^{-1}\ XM$$

e ainda,

$$X^t P_x - (X^t V^{-1} X)M = X^t - (X^t V^{-1} X)M = X^t (I-P_x)U$$

$$\rightarrow X^t - (X^t V^{-1} X)M = \emptyset \rightarrow \quad (X^t V^{-1} X)M = X^t$$

ou seja, o BLUE de $X\beta$ é dado por

$$A^t y = P_x y - P_x y + M^t\ X^t\ V^{-1}\ y = M^t\ X^t\ V^{-1} y =$$

= projeção de y em C(X) segundo a direção de vetores de C(VZ)

onde M é tal que $(X^t\ V^{-1} X)M = X^t$.

0000000

CAPITULO 4

ESTRUTURA DE COVARIÂNCIA DE EXPERIMENTOS ALEATORIZADOS

Com a finalidade de exemplificar a teoria exposta do Modelo Linear Geral, utilizaremos como base a idéia dos experimentos aleatorizados proposta por Fisher (1935) e amplamente estudada por Kempthorne e colaboradores (1961). A motivação é termos possibilidade de evitar o grande número de pressuposições que o modelo infinito exige, ou seja, normalidade, independencia, homogeneidade de variância e a imposição do modelo que os dados deverão obedecer. Nesse sentido é que os modelos lineares obtidos em experimentos aleatorizados são ditos derivados, ou seja, não são impostos mas sim obtidos com uma base física estrutural.

Com os exemplos que estudaremos atenderemos a dois objetivos, sendo um deles, uma estrutura de covariância que permite - nos obter estimadores ótimos (BLUE) através da solução simples do mínimos quadrados (exemplo do teorema 1, Zyskind (1967), e teorema (4.1) Rao (1973) o outro sendo que os experimentos aleatorizados oferece-nos uma classe de modelos lineares em que a estrutura de covariância é do tipo

$$\sigma^2 V = aI + bJ \qquad \text{onde,}$$

$$J = [1]\ [1]^t$$

([1] representa um vetor cujas coordenadas são iguais a um) ,

e, em alguns casos esta estrutura oferece matrizes singulares.

EXEMPLO 1 -

Uma ilustração primitiva de uma estrutura de covariância diferente de $\sigma^2 I$ é a que ocorre na retirada de uma amostra casual simples de tamanho n sem reposição de uma população finita de tamanho N.

Definindo a variável aleatória δ_i^j para relacionarmos os elementos amostrais onde,

$$\delta_i^j = \begin{cases} 1 & \text{se a resposta do j-ésimo elemento da população corresponde a i-ésima observação amostral} \\ 0 & \text{caso contrário} \end{cases}$$

para $i=1, \ldots, n$ e $j=1, \ldots, N$

temos que as propriedades de δ_i^j as quais são necessárias para o nosso propósito são: -

1 - $\text{prob}\ (\delta_i^j = 1) = \dfrac{1}{N}$ para todo $i=1, \ldots, n$
$j=1, \ldots, N$

2 - dado que $\delta_i^j = 1$ então

$$\text{prob } (\delta_i^{j'} = 1) = 0 \text{ para todo } j' \neq j \tag{4.2}$$

e

$$\text{prob}(\delta_{i'}^{j} = 1) = 0 \text{ para todo } i' \neq i \tag{4.3}$$

3 - dado que $\delta_i^j = 1$ então

$$\text{prob } (\delta_{i'}^{j'} = 1) = \frac{1}{N-1} \text{ para todo } i' \neq i \text{ e } j' \neq j \tag{4.4}$$

4 - $$\sum_{j=1}^{N} \delta_i^j = 1 = \sum_{i=1}^{N} \delta_i^j \tag{4.5}$$

e portanto,

$$E(\delta_i^j) = E\left[(\delta_i^j)^2 \right] = \text{prob } (\delta_i^j = 1) = \frac{1}{N} \tag{4.6}$$

para todo $i=1, \ldots , n$

$j=1, \ldots , N$

$$E(\delta_i^j \ \delta_i^{j'}) = \text{prob } (\delta_i^j \ \delta_i^{j'} = 1) =$$

$$= \text{prob}(\delta_i^{j'} = 1 / \delta_i^j = 1) \text{ prob } (\delta_i^j = 1) =$$

(por 4.2))

$$= 0 \quad \text{para todo } j' \neq j \tag{4.7}$$

analogamente

por (4.3)

$$E\,(\delta_i^j \quad \delta_{i'}^j) = 0 \tag{4.8}$$

para todo $i' \neq i$

por (4.4)

$$E\,(\,\delta_i^j \quad \delta_{i'}^{j'}\,) = \frac{1}{N(N-1)} \tag{4.9}$$

para todo $i' \neq i$ e $j' \neq j$

Assim, se

L_y representa a resposta do j-ésimo elemento da população $j = 1, \ldots, N$

e

y_i representa a resposta da i-ésima observação amostral, $i=1, \ldots, n$

e denotando por

$$u = \frac{1}{N} \sum_{j=1}^{N} L_j \qquad \text{(média populacional)}$$

$$\sigma^2 = \frac{1}{N} \sum_{j=1}^{N} (L_j - u)^2 \quad \text{(variância populacional)}$$

temos que, quando não há erro de medida então,

$$Y_i = \sum_{j=1}^{N} \delta_i^j \; L_j \tag{4.10}$$

e como, vale a identidade

$$L_j = u + (L_j - u) \qquad (4.11)$$

temos:

Substituindo (4.11) em (4.10) e usando (4.5)

$$y_i = \sum_{j=1}^{N} \delta_i^j (u + (L_j - u)) = u + \sum_{j=1}^{N} \delta_i^j (L_j - u)$$

ou seja,

$$y_i = u + e_i \qquad \text{para } i=1, \ldots, n \qquad (4.12)$$

onde

$e_i = \sum_{j=1}^{N} \delta_i^j (L_j - u)$ representa o

erro aleatório e, é tal que para todo $i=1, \ldots, n$

$$E(e_i) = \sum_{j=1}^{N} (L_j - u) E(\delta_i^j) = \text{por } (4.6)) =$$

$$= \frac{1}{N} \sum_{j=1}^{N} (L_j - u) = 0 \qquad (4.13)$$

$$\text{var}(e_i) = E(e_i^2) = E\left[\left(\sum_{j=1}^{N} (L_j - u) \quad (\delta_i^j)\right]^2\right.$$

$$= \sum_{j=1}^{N} \sum_{j'=1}^{N} (L_j - u)(L_{j'} - u) E(\delta_i^j \delta_i^{j'}) =$$

$$=(\text{por } (4.7)) = \sum_{j=1}^{N} (L_j - u)^2 \ E\left[(\delta_i^j)^2\right] =$$

$$= (\text{por } (4.6)) = \frac{1}{N} \sum_{j=1}^{N} (L_j - u)^2 = \sigma^2 \tag{4.14}$$

e para todo $i' \neq i$

$$\text{cov } (e_i, e_{i'}) = E\,(e_i\, e_{i'}) =$$

$$= E \left| \left(\sum_{j=1}^{N} \delta_i^j \, (L_j - u) \right) \left(\sum_{j'=1}^{N} \delta_{i'}^{j'} \, (L_{j'} - u) \right) \right| =$$

$$= \sum_{j=1}^{N} \sum_{j'=1}^{N} (L_j - u)(L_{j'} - u)\, E(\delta_i^j \; \delta_{i'}^{j'}) =$$

$$= \text{por } (4.8) = \sum_{\substack{j,j' \\ j' \neq i}} (L_j - u)(L_{j'} - u)\, E(\delta_i^j \; \delta^{j'}_{i'}) =$$

$$= \text{por}(4.9)) = \frac{1}{N(N-1)} \sum_{\substack{j,j' \\ j' \neq j}} (L_j - u)(L_{j'} - u) =$$

$$= - \frac{1}{N(N-1)} \sum_{j=1}^{N} (L_j - u)^2 + \frac{N}{N-1} \left(\frac{1}{N} \sum_{j=1}^{N} (L_j - u) \right)^2 =$$

$$= - \frac{1}{N(N-1)} \sum_{j=1}^{N} (L_j - u)^2 = - \frac{\sigma^2}{N-1} = \sigma^2 \rho \qquad \text{onde}$$

$$\rho = - \frac{1}{N-1} \tag{4.15}$$

Em notação matricial, com

$$y^t = (y_1 : \ldots : y_n)$$

$$e^t = (e_1 : \ldots : e_n)$$

temos o modelo linear

$y = [1]\ u + e$ onde

$X = [1]$ representa um vetor nx1 cujas coordenadas são todas iguais a um

e,

$E(e) = \emptyset$ (usando (4.13))

$E(e\ e^t) = \sigma^2\ V$

Sendo $V = \begin{vmatrix} 1 & \rho & \cdots & \rho \\ \rho & 1 & \cdots & \rho \\ \vdots & \vdots & & \vdots \\ \rho & \rho & \cdots & 1 \end{vmatrix}$ uma matriz nxn de posto n

(por (4.14) e (4.15))

Observando que

$V = (1 - \rho)\ I_n + \rho\ [1]\ \ [1]^t$ onde $(\rho = - \dfrac{1}{N - 1})$

e é tal que o vetor $[1]$ é um auto vetor de V então, segue que como

$$X = [1]$$

o S L S E

$$\bar{y} = \frac{1}{N} \sum_{i=1}^{N} y_i = \frac{1}{n} \left([1]^t \; y \right) \quad \text{de } u$$

é também o seu BLUE (Zyskind (1967))

onde,

$$\text{var } (\bar{y}) = \frac{N-n}{n-1} \quad \frac{\sigma^2}{n} .$$

De uma forma geral, sempre que a matriz V for da forma

$$V = a\, I + jb\, J$$

e a matriz X_{nxp} for da forma

$$X = \left([1] \; : \; \psi \right) \qquad \text{onde}$$

ψ é uma matriz nx(p-1)

podemos tomar, as matrizes

$$\underset{pxp}{T_1} = b \begin{vmatrix} 1 & \emptyset \\ \emptyset & \emptyset \end{vmatrix}$$

$$\underset{nxn}{T_2} = \emptyset \quad \text{(matriz nula)}$$

e

$$K^2 = a \quad \text{(constante real positiva)}$$

e satisfazem

$$V = K^2 I + X T_1 X^t + (I - P_x) T_2 (I - P_x)$$

que é a condição de Zyskind (1967) e Rao para que o S L S E $\equiv$ BLUE (4.16)

EXEMPLO 2 - Planejamento Completamente Casualizado

Admitamos que temos N unidades experimentais as quais aplicaremos aleatoriamente t tratamentos com a condição única que queremos r réplicas de cada tratamento, ou seja, N = rt.

Denotando a j-ésima réplica do i-ésimo tratamento pelo par (i,j), i=1,..., t e j=1, ... , r, então, cada par (i,j) será aplicado aleatóriamente às unidades experimentais.

A cada possível combinação tratamento e unidade experimental temos uma resposta conceitual L_{ik}, i=1, ..., t, k=1, ..., N. Ou seja,

L_{ik} é igual a resposta da

k-ésima unidade experimental ao i-ésimo tratamento.

Definindo-se a variável aleatória δ_{ij}^k para relacionarmos os resultados experimentais com a respostas conceituais L_{ik} onde

$$\delta_{ij}^k = \begin{cases} 1 & \text{se a j-ésima réplica do i-ésimo tratamento está aplicado na k-ésima unidade experimental} \\ 0 & \text{caso contrário} \end{cases}$$

para i=1, ..., t, j=1, ..., r e k=1, ..., N

temos então N t r = N^2 dessas variáveis aleatórias, cujas propriedades as quais são necessárias para o nosso propósito são:-

1 - $\text{prob}\ (\delta_{ij}^{k} = 1) = \frac{1}{N}$

para todo i=1, ..., t, j=1, ..., r, k=1, ...,N

(4.17)

2 - dado que $\delta_{ij}^{k} = 1$ então

$\text{prob}\ (\delta_{i'j'}^{k} = 1) = 0$ para todo $(i', j') \neq (i,j)$

(4.18)

$\text{prob}\ (\delta_{ij}^{k'} = 1) = 0$ para todo $k' \neq k$ (4.19)

3 - dado que $\delta_{ij}^{k} = 1$ então

$\text{prob}\ (\delta_{i'j'}^{k'}) = \frac{1}{N-1}$ para todo $(i', j') \neq (i,j)$ e $k' \neq k$

(4.20)

4 - $\sum_{k=1}^{N} \delta_{ij}^{k} = 1 = \sum_{i=1}^{t} \sum_{j=1}^{r} \delta_{ij}^{k}$ (4.21)

e portanto,

$$E(\delta_{ij}^{k}) = E\ (\delta_{ij}^{k})^2 = \text{prob}(\delta_{ij}^{k} = 1) = (\text{por}(4.17)) = \frac{1}{N}$$

para todo i=1, ..., t, j=1, ..., r, k=1, ..., N

(4.22)

$$E(\delta_{ij}^{k}\ \delta_{i'j'}^{k}) = \text{prob}\ (\delta_{ij}^{k}\ \delta_{i'j'}^{k} = 1) =$$

$$= \text{prob}\ (\ \delta_{i'j'}^{k} = 1 / \delta_{ij}^{k} = 1).\ \text{prob}(\ \delta_{ij}^{k} = 1) =$$

$$= (\text{por } (4.18)) = \text{o} \quad \text{para todo } (i',\ j') \neq (i,j) \tag{4.23}$$

analogamente

$$\text{por } (4.19))\ E(\ \delta_{ij}^{k}\ \delta_{ij}^{k'}\ ) = 0 \quad \text{para todo } k' \neq k \tag{4.24}$$

$$\text{por } (4.20))\ E(\ \delta_{ij}^{k}\ \delta_{i'j'}^{k'}) = \frac{1}{N(N-1)}$$

$$\text{para todo } (i',\ j') \neq (i,j) \quad \text{e} \quad k' \neq k \tag{4.25}$$

Considerando agora

$L_{..}$ a média sobre o experimento total

$$L_{..} = \frac{1}{Nt} \sum_i \sum_k L_{ik}$$

$L_{i.}$ a média das respostas do i-ésimo tratamento

$$L_{i.} = \frac{1}{N} \sum_k L_{ik}$$

$L_{.k}$ a média das respostas da k-ésima unidade experimental

$$L_{.k} = \frac{1}{t} \sum_i L_{ik}$$

temos, a seguinte decomposição, proposta fisicamente, das respostas conceituais

$$L_{ik} = L_{..} + (L_{i.} - L_{..}) + (L_{.k} - L_{..}) +$$

$$+ (L_{ik} - L_{i.} - L_{.k} + L_{..}) \qquad (4.26)$$

onde

$L_{i.} - L_{..}$ representa o desvio de $L_{i.}$ com relação à $L_{..}$ ou seja, o efeito do i-ésimo tratamento

$L_{.k} - L_{..}$ representa o desvio de $L_{.k}$ com relação à $L_{..}$ ou seja, o efeito da k-ésima unidade experimental e

$$L_{ik} - L_{i.} - L_{.k} + L_{..} =$$

$$= (L_{ik} - L_{..}) - (L_{i.} - L_{..}) + (L_{.k} - L_{..})$$

representa o efeito de interação do i-ésimo tratamento e da k-ésima unidade experimental

Supondo, que os efeitos das unidades experimentais e tratamentos são aditivos temos então que, a resposta do i-ésimo tratamento independe da unidade experimental ou seja,

$$L_{ik} - L_{i.} - L_{.k} + L_{..} = 0$$

e como consequência dessa suposição obtemos que a identidade (4.26) torna-se

$$L_{ik} = L_{..} + (L_{i.} - L_{..}) + (L_{.k} - L_{..}) =$$

$$= u + t_i + e_k \qquad (4.27)$$

onde $u = L_{..}$

$$t_i = L_{i.} - L_{..}$$

$$e_k = L_{.k} - L_{..}$$

Assim, se

y_{ij} representa a observação da j-ésima réplica do i-ésimo tratamento então

$$y_{ij} = \sum_{k=1}^{N} \delta_{ij}^{k} L_{ik} \qquad (4.28)$$

que são variáveis aleatórias obtidas como combinação linear das variáveis aleatórias amostrais δ_{ij}^{k}.

Substituindo a identidade populacional (4.27) na identidade das observações (4.28)

temos que

$$y_{ij} = \sum_{k=1}^{N} \delta_{ij}^{k} L_{ik} = u + t_i + \sum_{k=1}^{N} \delta_{ij}^{k} e_k$$

ou seja,

tomando $e_{ij} = \sum_{k=1}^{N} \delta_{ij}^{k} e_k$ que representa o erro aleatório da i-ésima observação na j-ésima réplica então

$$y_{ij} = u + t_i + e_{ij} \qquad \text{para } i=1, \ldots, t$$
$$j=1, \ldots, r$$

onde

$$\sum_{i=1}^{t} t_i = 0$$

$$\sum_i \sum_j e_{ij} = \sum_k e_k = 0$$

e e_{ij} para todo (i=1, ..., t e j=1, ... , r é tal que

por (4.22)

$$E(e_{ij}) = E(\sum_{k=1}^{N} \delta_{ij}^{k} e_k) = \sum_{k=1}^{N} e_k \; E(\delta_{ij}^{k}) =$$

$$= \frac{1}{N} \sum_{k=1}^{N} e_k = 0 \qquad (4.29)$$

$$\text{var}(e_{ij}) = E\left[(e_{ij})^2\right] = E\left[(\sum_{k=1}^{N} \delta_{ij}^{i} \; e_k)^2\right] =$$

$$= \sum_k \sum_{k'} e_k e_{k'} \; E(\delta_{ij}^{k} \; \delta_{ij}^{k'}) = (\text{por}(4.24)) =$$

=

$$= \sum_{k=1}^{N} e_k^2 \; E\left[(\delta_{ij}^{k})^2\right] = \text{por } (4.22) = \frac{1}{N} \sum_k c_k^2 = \sigma^2$$

(4.30)

e para todo par (i', j') tal que (i', j') ≠ (i,j)

$$\text{cov}\,(e_{ij},\, e_{i'j'}) = E\left[\left(\sum_{k=1}^{N} \delta_{ij}^{k}\; e_k\right)\left(\sum_{k'=1}^{N} \delta_{i'j'}^{k'}\; e_{k'}\right)\right] =$$

$$= \sum_{k=1}^{N} \sum_{k'=1}^{N} e_k\, e_{k'}\; E(\delta_{ij}^{k}\; \delta_{i'j'}^{k'}) = \quad (\text{por}(4.23)$$

$$= \sum_{\substack{k,k' \\ k \neq k'}} e_k\, e_{k'}\; E(\delta_{ij}^{k}\; \delta_{i'j'}^{k'}) = (\text{por } (4.25)) =$$

$$= \frac{1}{N(N-1)} \sum_{\substack{k,k' \\ k \neq k'}} e_k\, e_{k'} = -\frac{1}{N(N-1)} \sum_k e_k^2 + (\sum_k e_k)(\sum_{k'} e_{k'}) =$$

$$= -\frac{1}{N(N-1)} \sum_k e_k^2 = -\frac{\sigma^2}{N-1} = \sigma^2\,\rho \quad \text{onde } \rho = -\frac{1}{N-1} \tag{4.31}$$

Em notação matricial com

$$y^t = (y_{11} : \dots : y_{1r} : \dots : y_{t1} : \dots : y_{tr})$$

e

$$e = (e_{11} : \dots : e_{1r} : \dots : e_{t1} : \dots : e_{tr})$$

o modelo derivado para o planejamento completamente casualizado é portanto um modelo linear dado por

$$y = X\beta + e$$

com

$$\beta^t = (u\colon\ t_1\colon\ \ldots\colon\ t_t)$$

$X = \left([1] : \psi\right)$ uma matriz Mxt+1 de posto t

sendo $[1]$ um vetor Nx1 cujas coordenadas são todas iguais a um

e

$$\psi = \begin{vmatrix} x_{(1)} & \emptyset & \ldots & \emptyset \\ \emptyset & x(2) & \ldots & \emptyset \\ & & & \\ \emptyset & \emptyset & & x(t) \end{vmatrix}$$

onde cada $x_{(I)}$, i=1, ..., t, representa um vetor rx1 cujas coordenadas são todas iguais a um

e $\emptyset$ representa um vetor rxs nulo.

e,

(por, (4.29))

$$E(e) = \emptyset$$

(por (4.30) e (4.31))

$$E(e\ e^t) = \sigma^2\ V$$

Sendo $V = \begin{vmatrix} 1 & \rho \ldots & \rho \\ \rho & 1 & \rho \\ \vdots & \vdots & \vdots \\ \rho & \rho & 1 \end{vmatrix}$ uma matriz NxN de posto N-1

Observando que as matrizes X e V do modelo linear derivado do Planejamento Casualizado são semelhantes ao caso deduzido em (4.16) com, a restrição de que V é singular então, temos que

$$V = \begin{vmatrix} 1 & \rho \ldots\ldots & \rho \\ \rho & 1 \ldots\ldots & \rho \\ \vdots & \vdots & \vdots \\ \rho & \rho \ldots\ldots & 1 \end{vmatrix} = (1 - \rho)\ I_N + \rho \left(1\right) \left(1\right)^t =$$

$$= X\ T_1\ X^t + (I - P_x)\ T_2\ (I - P_x) + K^2\ I$$

onde T_2 é uma matriz nula NxN

$K^2 = (1 - \rho) = \dfrac{N}{N-1} \in R$

$T_1 = \rho \begin{vmatrix} 1 & \emptyset \\ \emptyset & \emptyset \end{vmatrix}$ é uma matriz $(t+1) \times (t+1)$

Assim, para obtermos o BLUE de uma função paramétrica linearmente estimável no modelo linear derivado de Planejamento Completamente Casualizado é suficiente trabalharmos com o modelo linear cujas hipóteses são as do Modelo Linear de Gauss-Markoff ou seja,

$E(e) = \emptyset$

$E(e\ e^t) = \sigma^2\ I$

000o000

B I B L I O G R A F I A

AITKEN; A.C. 1935. On least squares and linear combination of observations. *Proceedings of the Royal Society of Edinburg, Section A*, 55:42-47.

FISHER, R.A. 1966. *The design of experiments.* Edinburg, Oliver & Boyd. 248p.

GAUSS, C.F. 1855. *Methode des Moindres Carres.* (Trans. J. Bertrand)

GOLDMAN, A.J. & ZELEN, M. 1964. Weak generalized inverses and minimum variance linear unbiased estimation. *Journal of Research of the National Bureau of Standards, Section B*, *68B*:151-172.

HALMOS, P.R. 1978. *Espaços vetoriais de dimensão finita.* Rio de Janeiro, Campus. 199p.

HOFFMAN, K. & KUNZE, R. 1979. *Algebra linear.* 2.ed. Rio de Janeiro, Livros Técnicos e Científicos. 514p.

KEMPTHORNE, O. 1952. *The design and analysis of experiments* New York, John Wiley. 631p.

KEMPTHORNE, O. 1976. *Best linear unbiased estimation with arbitrary variance matrix.* (Unpublished mimeographed notes)

KEMPTHORNE, O. et alii. 1961. *Analysis of variance procedu res*. Ohio, Patterson Air Force Base. (Aeronautical Research Laboratory. Technical Report, 149)

KRUSKAL, W. 1975. The geometry of generalized inverses. *Journal of the Royal Statistical Society, Series B, 37*(2): 272-283.

MARKOFF, A.A. 1912. *Warscheinlichkeitsrechnung*. 2.ed. Leipzig, s.c.p. (Trans. H. Liebmann)

MONTEIRO, L.H.J. 1969. *Algebra linear*. 5.ed. São Paulo, Nobel. v.1

RAO, C.R. 1965. *Linear statistical inference and its appli cations*. New York, John Wiley. 522p. (Wiley Series in Probability and Mathematical Statistics)

RAO, C.R. 1967. Least squares theory using an estimated dispersion matrix and its application to measurement of signals. In: BERKELEY SYMPOSIUM ON MATHEMATICAL STATISTICS AND PROBABILITY, 5., Berkeley, 1965. *Proceedings*. Berkeley, University of California Press, 1967. v.1,p.355-372.

RAO, C.R. 1971. Unified theory of linear estimation. *Sankhyã, Series A, 33*(4):371-394.

RAO, C.R. 1972. Some recent results in linear estimation. *Sankhyã, Series B, 34*(4):369-378.

RAO, C.R. 1973a. Representations of best linear unbiased estimators in the Gaus-Markoff model with a singular disper sion matrix. *Journal of Multivariate Analysis*, 3(1):276--292.

RAO, C.R. 1973b. Unified theory of least squares. *Communications in Statistics*, 1(1):1-8.

RAO, C.R. 1975. On a unifed theory of estimation in linear models a review of recent results. In: GANI, J., ed. *Perspectives in probability and statistics*. London, Academic Press. p.89-104.

RAO, C.R. 1978. A note on the unified theory of least squares. *Communications in Statistics-Theory and Methods*, A7 (5):409-411.

RAO, C.R. & MITRA, S. 1971. *Generalized inverse of matrices and its applications*. New York, John Wiley. 240p. (Wiley Series in Probability and Mathematical Statistics)

SCHEFFE, H. 1959. *The analysis of variance*. 477p.

ZYSKIND, G. 1967. On canonical forms, non-negative covariance matrices and best and simple least squares linear estimators in linear models. *Annals of Mathematical Statistics*, 38(4):1092-1109.

ZYSKIND, G. et alii. 1964. *Research on analysis of variance and related topics*. Ohio, Patterson Air Force Base. (Aerospace Research Laboratories. Technical Report, 64-193)

ZYSKIND, G. & MARTIN, F.B. 1969. On best linear estimation and a general Gauss-Markov theorem in linear models with arbitrary nonnegative covariance structure. *SIAM Journal on Applied Mathematics*, 17(6):1190-1202.