MODELOS LINEARES COM MATRIZ DE COVARIANCIA ARBITRARIA - CONDIÇÕES PARA QUE O ESTIMADOR SIMPLES DE MINIMOS QUADRADOS SEJA BLUE.

ADEMIR JOSE PETENATE

ORIENTADOR

JOSE FERREIRA DE CARVALHO

Dissertação apresentada no Instituto de Matemática, Estatística e Ciência da Computação como requisito parcial para a obtenção do título de Mestre em Estatística.

FEVEREIRO, 1979

"As coisas estão no mundo
só que eu preciso aprender."

Paulinho da Viola

SUMARIO

Este trabalho tem como finalidades:

1) Revisar alguns resultados clássicos em Modelos Lineares, discutir as restrições impostas ao espaço de parâmetros quando a matriz de covariâncias é singular e interpretar geometricamente alguns resultados importantes.

2) Apresentar um conjunto de condições equivalentes sobre o modelo linear geral para que o estimador simples de mínimos quadrados de funções estimáveis seja BLUE, quando a matriz de covariância é arbitrária.

3) Discutir alguns experimentos finitos aleatorizados (amostra aleatória simples sem reposição, experimentos completamente ao acaso e blocos completamente ao acaso) e mostrar que para esses experimentos o estimador simples de mínimos quadrados de qualquer função estimável é BLUE, embora a matriz de covariância dos erros seja distinta de $\sigma^2 I$ e possivelmente singular.

SUMMARY

The scope of this work is:

1) To review some classical results in Linear Models, to discuss the restrictions imposed on the parameter space when the covariance matrix is singular and to provide geometric interpretation of some relevant results.

2) To present a set of equivalent conditions on the general linear model in order that the simple least squares estimator of any estimable function be BLUE, when the covariance matrix is arbitrary.

3) To discuss some finite randomized experiments (simple random sampling without replacement, the complete randomized design and the complete randomized block design) and to show that for these experiments the Simple Least Squares estimator of any estimable function is BLUE, although the covariance matrix of the errors is not $\sigma^2 I$ and may be singular.

INDICE

PAGINA

INTRODUÇÃO

O objetivo da presente dissertação é apresentar, para o modelo linear geral, alguns resultados importantes sobre estimação de funções lineares dos parâmetros, quando a matriz de covariância é arbitrária e conhecida, a menos de um fator escalar.

Embora a suposição de que a matriz V é conhecida possa parecer uma restrição forte, existem muitas situações importantes em que as condições físicas do experimento fornecem informação suficiente para supormos uma determinada estrutura para a matriz de covariância. Além disso. em alguns casos, é possível se determinar a matriz de covariância a menos de um fator escalar, como por exemplo, para alguns experimentos finitos aleatorizados.

Mais concretamente, apresentaremos um número de condições necessárias e suficientes sobre o modelo linear geral de tal forma que os estimadores simples de mínimos quadrados sejam BLUE's (Best Linear Unbiased Estimators).

Esses resultados serão apresentados por dois caminhos distintos: na sua forma original, como estabelecidos por Zyskind (1967), e de uma forma mais simplificada, utilizando um instrumental matemático pouco sofisticado e que foram estabelecidos por Kempthorne (197?). Daremos, sempre que possível, algumas interpretações geométricas para esses resultados.

Verificaremos também, para alguns modelos derivados de experimentos aleatorizados - experimentos esses muito utiliza dos na prática - que podemos tratar o problema da estimação de pa râmetros como se fossem válidas as hipóteses do teorema de Gauss-Markoff, embora os erros sejam correlacionados e, em alguns casos, a matriz de covariância seja singular.

CAPITULO 1

MODELOS LINEARES

1.1 - A QUESTÃO DO MODELO

Na investigação científica, quando estamos diante de um fenômeno que pode ser mensurado, um dos primeiros passos é procurar obter resultados experimentais e relacioná-los de alguma forma.

Consideremos a situação bastante geral em que medimos uma característica Y, sob certas condições experimentais, condições essas estabelecidas por um conjunto de variáveis $X_{(1)}, X_{(2)}, \ldots, X_{(p)}$, previamente fixadas, as quais chamaremos de variáveis explanatórias. Denotaremos por y o valor observado da característica Y.

O que geralmente se deseja é conhecer o comportamento da característica Y de tal forma que, fixada uma particular condição experimental, possamos prever a resposta y. Esse problema, na linguagem matemática, traduz-se na determinação de uma função f tal que

$$y = f(X_{(1)},\ldots,X_{(p)}) \quad .$$

Na quase totalidade dos casos é impossível a determinação de uma tal função f e os motivos variam desde a complexidade de tal função até a existência de fatores que influenciam na resposta y e que, por vários motivos, não foram considerados.

O que geralmente conseguimos determinar é uma função f de tal forma que

$$y \cong f(X_{(1)},\ldots,X_{(p)}) \quad .$$

Chamaremos de

$$e = y - f(X_{(1)},\ldots,X_{(p)})$$

o erro devido à falta de ajuste.

Suponhamos que para o estudo da dependência funcional de y em relação a $X_{(1)},\ldots,X_{(p)}$ dispomos de um conjunto de n observações da característica Y, y_i, $i=1,2,\ldots,n$, sendo y_i o valor observado da variável Y sob a condição experimental $(X_{i1},X_{i2},\ldots,X_{ip})$ e X_{ij} o valor da variável $X_{(j)}$ sob a qual foi observado o valor y_i.

É a partir do conjunto de dados disponíveis $\{(y_i,X_{i1},\ldots,X_{ip}),\ i=1,2,\ldots,n\}$ que procuraremos estabelecer um modelo matemático que explique o comportamento de Y como função de $X_{(1)},\ldots,X_{(p)}$. Obviamente jogamos com dois critérios básicos na pesquisa do modelo, critérios esses que nem sempre se complementam. Se por um lado desejamos que o modelo a ser ajustado explique o melhor possível o comportamento real do fenômeno, por outro lado desejamos que o modelo, o qual se traduz numa forma funcional, seja o mais simples possível.

Se escrevemos

$$y_i = f(X_{i1},\ldots X_{ip}) + e_i$$

então e_i é o erro devido à falta de ajuste dos dados em relação à função f e poderemos utilizar os valores $e_1,\ldots,e_n$ para medir a "qualidade" de ajuste do modelo.

Cada família de funções que podemos ajustar dependerá de um conjunto de parâmetros $\beta_1,\beta_2,\ldots,\beta_k$. Então podemos escrever:

$$y_i = f(X_{i1},\ldots,X_{ip}\ ;\ \beta_1,\beta_2,\ldots,\beta_k) + e_i\ .$$

Escolhida uma determinada relação funcional f, os valores $e_1,\ldots,e_n$ dependerão do valor particular de $\beta = (\beta_1,\ldots,\beta_k)$ a ser utilizado. Coloquemos o problema de ajuste sob o ponto de vista algébrico.

Representemos o conjunto de dados $y_1,\ldots,y_n$ pelo vetor $y = (y_1,\ldots,y_n)$ o qual é um ponto do espaço R^n. Consideremos que o vetor $\beta = (\beta_1,\ldots,\beta_k)$ é um ponto pertencente a um subconjunto de R^k. O vetor

$$f(X;\beta) = (f(X_{11},\ldots,X_{1p};\beta_1,\ldots,\beta_k),\ldots,f(X_{n1},\ldots,X_{np};\beta_1,\ldots,\beta_k)\ )$$

é também um ponto de R^n. Para valores diferentes de β podemos ter valores distintos de $e_1,\ldots,e_n$. Então, para uma dada forma funcional, o problema de ajuste da função pode ser reduzido à escolha de um determinado valor $\hat{\beta}$ de tal forma que o ponto $f(X;\hat{\beta})$ esteja o mais "próximo" possível do ponto y. Naturalmente, para se medir a proximidade desses pontos, temos que estabelecer uma função distância conveniente em R^n.

De um modo geral, se pretendemos compreender o comportamento de uma característica Y através dos meios propostos na discussão efetuada acima, temos que responder convenientemente as seguintes questões:

1) Quais variáveis explanatórias utilizar?

2) Que forma funcional ajustar?

3) Qual função distância utilizar?

Uma das famílias de funções com estrutura razoavelmente simples, tendo $X_{(1)},\ldots,X_{(p)}$ como argumentos e $\beta_1,\ldots,\beta_k$ como parâmetros, é a família de funções que podem ser colocadas na forma

$$f(X_{(1)},\ldots,X_{(p)}\ ;\ \beta_1,\ldots,\beta_k) = \sum_{i=1}^{n} \beta_i f_i(X_{(1)},\ldots,X_{(p)}) \qquad (1.1.1)$$

que é uma função linear nos parâmetros. Essa família de funções desempenha um papel importante no tratamento estatístico de dados.

Quando queremos ajustar um conjunto de dados a um conjunto de variáveis explanatórias, as aproximações mais simples devem ser procuradas entre funções do tipo (1.1.1). Funções mais complexas tornam o tratamento matemático extremamente árduo e de resultados pouco práticos.

Ao longo desse capítulo discutiremos o ajuste do modelo

$$y_i = \sum_{j=1}^{p} X_{ij}\beta_j + e_i \qquad i = 1,\ldots,n$$

com a suposição que

$$e_i = y_i - \sum_{j=1}^{p} X_{ij}\beta_j \qquad i = 1,\ldots,n$$

são variáveis aleatórias com valor esperado igual a zero. Apresentaremos brevemente alguns resultados clássicos em Modelos Linea res relativos à estimação de parâmetros sem nos preocuparmos com aspectos relativos à distribuição de probabilidades das variáveis aleatórias envolvidas.

1.2 - EQUAÇÕES NORMAIS

Dispomos, por hipótese, de um conjunto de dados $\{(y_i, X_{i1}, \ldots, X_{ip}),\ i=1,\ldots,n\}$ para o qual vamos ajustar um modelo da forma

$$y_i = \sum_{j=1}^{p} X_{ij}\beta_j + e_i \quad , \qquad i = 1,\ldots,n \ .$$

Utilizando notação matricial podemos escrever o modelo suposto na forma

$$y = X\beta + e$$

sendo

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} \qquad \text{- vetor de observações,}$$

$$X = \begin{bmatrix} X_{11} & X_{12} & \cdots & X_{1p} \\ X_{21} & X_{22} & \cdots & X_{2p} \\ \vdots & & & \\ X_{n1} & X_{n2} & \cdots & X_{np} \end{bmatrix} \qquad \text{- matriz de planejamento,}$$

$$\beta = \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_p \end{bmatrix} \quad \text{- vetor de parâmetros,}$$

$$e = \begin{bmatrix} e_1 \\ e_2 \\ \vdots \\ e_n \end{bmatrix} \quad \text{- vetor de erros.}$$

Essa notação facilitará enormemente a exposição dos resultados.

Denotando por E o operador esperança matemática temos por hipótese que

$$E(e) = \emptyset \quad .$$

A covariância entre e_i e e_j, $\text{cov}(e_i, e_j)$, será denotada por $\sigma^2 V_{ij}$. Portanto a matriz de covariância dos erros é

$$\text{cov}(e) = E(ee') = \sigma^2 \begin{bmatrix} V_{11} & V_{12} & \cdots & V_{1n} \\ V_{21} & V_{22} & \cdots & V_{2n} \\ \vdots & & & \\ V_{n1} & V_{n2} & \cdots & V_{nn} \end{bmatrix}$$

O vetor y de observações é um ponto do espaço R^n, espaço esse que denominaremos de espaço de dados.

Sejam $X_1,\ldots,X_p$ os vetores formados pelas colunas da matriz X e C(X) o espaço vetorial gerado pelas combinações lineares de $X_1,\ldots,X_p$. C(X) é um subespaço de R^n.

Podemos escrever

$$X\beta = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_p X_p \;.$$

Portanto $X\beta$ é um ponto pertencente a C(X).

O parâmetro $E(y) = X\beta$ é desconhecido. Com base no vetor de observações y queremos encontrar um vetor $X\hat{\beta}$, portanto um vetor pertencente a C(X), que esteja o mais próximo de y segundo uma norma estabelecida. Chamaremos $X\hat{\beta}$ de estimativa de $E(y) = X\beta$.

O critério clássico utilizado para se encontrar o vetor $\hat{\beta}$ de tal forma que $X\hat{\beta}$ seja a estimativa de $X\beta$ é minimizar a soma dos quadrados dos erros. Esse critério é uma aplicação do Princípio de Mínimos Quadrados, princípio esse introduzido independentemente por Legendre e Gauss. A função distância que é utilizada é a norma euclidiana.

Do ponto de vista algébrico queremos encontrar um vetor $X\hat{\beta}$ de tal forma que a distância entre o vetor y e o vetor $X\hat{\beta}$ seja a menor distância entre o vetor y e o subespaço C(X). Não faremos aqui nenhuma suposição sobre a estrutura da matriz de covariância dos erros e, por conseguinte, não discutiremos ainda a qualidade da estimativa conseguida. O objetivo, por enquanto, será obter um sistema de equações, através do processo de minimização referido acima, o qual se mostra extremamente fértil para o desenvolvimento e a compreensão da teoria de Modelos Lineares do ponto de vista da Estatística.

Seja a função

$$\Theta(\beta) = \sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (y_i - \sum_{j=1}^{p} X_{ij}\beta_j)^2 .$$

Nosso objetivo é encontrar um vetor β tal que $\Theta(\beta)$ seja mínimo.

Derivando $\Theta(\beta)$ em relação a β_k, $k = 1,\ldots,p$ e igualando cada equação a zero temos:

$$\frac{\partial\Theta(\beta)}{\partial\Theta_k} = -2 \sum_{i=1}^{n} X_{ik}(y_i - \sum_{j=1}^{n} X_{ij}\beta_j) = 0 \qquad k = 1,\ldots,p .$$

Portanto

$$\sum_{j=1}^{p} (\sum_{i=1}^{n} X_{ik}X_{ij})\beta_j = \sum_{i=1}^{n} X_{ik}y_i \qquad k = 1,\ldots,p .$$

As p equações acima são chamadas Equações Normais.

Em termos matriciais as equações normais tomam a seguinte forma:

$$X'X\beta = X'y \qquad (1.2.1) .$$

Qualquer vetor β que satisfaça (1.2.1) é um ponto crítico de $\Theta(\beta)$. Podemos mostrar que existe pelo menos um vetor β satisfazendo (1.2.1) e que qualquer β satisfazendo (1.2.1) torna $\Theta(\beta)$ mínimo. Antes disso enunciaremos um teorema que será utilizado ao longo dessa dissertação para provar consistência de sistema de equações lineares.

Teorema 1.1 - O sistema de equações

$Ax = b$ é consistente se e somente se qualquer a tal que $a'A = \emptyset \Rightarrow a'b = 0$.

Demonstração:

(=>) $Ax = b$ é consistente então

$$a'A = \emptyset \quad \Rightarrow \quad a'Ax = \emptyset \quad \Rightarrow \quad a'b = 0$$

(<=) O conjunto dos vetores a tal que $a'A = 0$ é o subespaço complemento ortogonal do espaço coluna de A que denotaremos por $C(A)^{\perp}$. Então, se qualquer $a \in C(A)^{\perp} \Rightarrow a'b = 0$, b é ortogonal a $C(A)^{\perp}$. Então $b \in C(A)$ e $Ax = b$ é consistente.

Teorema 1.2 - As equações normais são consistentes. Qualquer β solução das equações normais é um ponto de mínimo de $\Theta(\beta)$.

Demonstração:

Qualquer a tal que

$$a'X'X = \emptyset \quad \Rightarrow \quad a'X'Xa = 0 \quad \Rightarrow \quad a'X' = \emptyset \quad \Rightarrow \quad a'X'y = 0 .$$

Então as equações normais são consistentes.

Seja $\hat{\beta}$ qualquer solução das equações normais.

Temos

$$e = (y-X\beta) = (y-X\hat{\beta}) + (X\hat{\beta}-X\beta)$$

$$\Theta(\beta) = \sum_{i=1}^{n} e_i^2 = e'e = (y-X\beta)'(y-X\beta) =$$

$$= \{(y-X\hat{\beta}) + (X\hat{\beta}-X\beta)\}'\{(y-X\hat{\beta}) + (X\hat{\beta}-X\beta)\} =$$

$$= (y-X\hat{\beta})'(y-X\hat{\beta}) + (X\hat{\beta}-X\beta)'(X\hat{\beta}-X\beta) + 2(y-X\hat{\beta})'(X\hat{\beta}-X\beta) =$$

$$= (y-X\hat{\beta})'(y-X\hat{\beta}) + (X\hat{\beta}-X\beta)'(X\hat{\beta}-X\beta)$$

desde que $\hat{\beta}$ satisfaz $X'X\beta = X'y$.

$(y-X\hat{\beta})'(y-X\hat{\beta})$ não depende de β e para todo β temos que $(X\hat{\beta}-X\beta)'(X\hat{\beta}-X\beta) \geq 0$ e é igual a zero se $\beta = \hat{\beta}$.

Portanto $\Theta(\beta)$ é mínimo para qualquer β solução das equações normais.

Se $X'X$ for não singular existe um único vetor β que é solução das equações normais. Se $X'X$ for singular a solução das equações normais não é única. No entanto estamos interessados em estimar $E(y) = X\beta$. O teorema seguinte estabelece a unicidade da estimativa de $E(y)$ para qualquer β solução das equações normais.

<u>Teorema 1.3</u> - Se $\hat{\beta}_1$ e $\hat{\beta}_2$ são dois vetores soluções das equações normais então

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

Demonstração:

Sejam $\hat{\beta}_1$ e $\hat{\beta}_2$ soluções de $X'X\beta = X'y$. Então

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

$$X'X(\hat{\beta}_1-\hat{\beta}_2) = \emptyset$$

$$(\hat{\beta}_1-\hat{\beta}_2)'X'X(\hat{\beta}_1-\hat{\beta}_2) = 0 \quad => \quad X(\hat{\beta}_1-\hat{\beta}_2) = \emptyset .$$

Portanto

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

A invariância de $X\hat{\beta}$ implica na invariância de $(X\hat{\beta})'(X\hat{\beta})$, conhecida como soma dos quadrados devido aos parâmetros, e também na invariância de $(y-X\hat{\beta})'(y-X\hat{\beta})$, conhecida como soma dos quadrados dos resíduos.

Ao vetor $X\hat{\beta}$ tal que $\hat{\beta}$ é solução de $X'X\beta = X'y$ chamaremos de Estimador Simples de Mínimos Quadrados (ESMQ)* de $X\beta$.

1.3 - INTERPRETAÇÃO GEOMETRICA DAS EQUAÇÕES NORMAIS

O vetor y pode ser escrito como

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

onde $\hat{\beta}$ é solução das equações normais.

---

* Nos textos de língua inglesa é usada a notação SLE (Simple Least Squares estimator.)

Temos que $X\hat{\beta} \in C(X)$. Como $\hat{\beta}$ satisfaz $X'X\beta = X'y$ temos que

$$X'X\hat{\beta} = X'y \Rightarrow X'(y-X\hat{\beta}) = \emptyset$$

ou seja, $(y-X\hat{\beta})$ é ortogonal a todas as colunas da matriz X e portanto $(y-X\hat{\beta})$ é ortogonal a $C(X)$. Então $(y-X\hat{\beta}) \in C(X)^{\perp}$, complemento ortogonal de $C(X)$.

Podemos demonstrar que a decomposição de $y \in R^n$ em $X\hat{\beta} \in C(X)$ e $(y-X\hat{\beta}) \in C(X)^{\perp}$ é única.

Sejam

$$y_1 \in C(X) \quad \text{e} \quad y_2 \in C(X)^{\perp}$$

tal que

$$y = y_1 + y_2$$

$$y_1 \in C(X) \Rightarrow y_1 = X\tilde{\beta} \text{ para algum } \tilde{\beta}.$$

$$y_2 = y - y_1 = y - X\tilde{\beta} .$$

Assim

$$y = X\hat{\beta} + (y-X\hat{\beta}) = X\tilde{\beta} + (y-X\tilde{\beta}) .$$

Então

$$(y-X\tilde{\beta}) - (y-X\hat{\beta}) = (X\hat{\beta}-X\tilde{\beta}) = X(\hat{\beta}-\tilde{\beta}) .$$

Portanto, $(X\hat{\beta}-X\tilde{\beta})$ é um ponto de $C(X)$ e também de $C(X)^{\perp}$ e, como $C(X)$ e $C(X)^{\perp}$ são subespaços ortogonais, o único vetor comum aos dois subespaços é o vetor zero. Então

$$X\hat{\beta}-X\tilde{\beta}= \emptyset \Rightarrow X\hat{\beta} = X\tilde{\beta} .$$

Portanto

$$y_1 = X\tilde{\beta} = X\hat{\beta} \qquad e$$

$$y_2 = (y-X\tilde{\beta}) = (y-X\hat{\beta}) \qquad .$$

O vetor $X\hat{\beta}$ tal que $(y-X\hat{\beta})$ é ortogonal ao subespaço $C(X)$ é chamado de projeção ortogonal de y em $C(X)$.

O vetor $(y-X\hat{\beta})$, o erro estimado, pertence a $C(X)^{\perp}$. Denominaremos $C(X)^{\perp}$ de espaço de erros.

Até agora procuramos caracterizar algebricamente a estimativa de $X\beta$, $X\hat{\beta}$, tal que $\hat{\beta}$ minimiza a soma dos quadrados dos erros. Procuraremos explicitar, de uma forma geral, a estimativa $X\hat{\beta}$ deduzida acima. Para isso precisamos do seguinte teorema:

<u>Teorema 1.4</u> - Dada uma matriz qualquer X de dimensão nXp é sempre possível encontrar-se uma matriz B, de dimensão pXn, satisfazendo

$$X'XB = X'.$$

Além disso a matriz XB é única.

<u>Demonstração</u>:

Consideremos as equações normais $X'X\beta = X'y$. Substituamos sucessivamente o vetor y pelo vetor f_i, $i=1,\ldots,n$, onde f_i é um vetor com zeros em suas coordenadas, exceto na i-ésima linha onde o componente é a unidade, ou seja, consideremos as n equações normais

$X'X\beta = X'f_i \qquad i = 1,2,\ldots n$.

A consistência das equações normais garante a existência de uma matriz B tal que

$X'XB = X'I = X'$.

Além disso, a unicidade de $X\hat{\beta}$ garante a unicidade da matriz XB.

Proposição 1 - By é solução das equações normais.

Demonstração:

$X'XBy = X'y$.

Portanto expressamos o vetor $X\hat{\beta}$ como combinação linear do vetor de observações y e XBy é único.

As matrizes B e XB possuem algumas propriedades interessantes. Antes de apresentá-las definiremos a inversa generalizada de uma matriz.

Definição:

Dada qualquer matriz A, dizemos que a matriz ∂ é uma inversa generalizada de A se ∂ satisfaz

$$A\partial A = A.$$

Propriedades da matriz B e da matriz XB:

1. XB é simétrica

$X'XB = X'$

$B'X'X = X$

$B'X' = B'X'XB = XB$

2. <u>XB é idempotente</u>

$$XBXB = (XB)'XB = B'X'XB = XB$$

3. <u>posto (X) = posto (XB)</u>

$$\text{posto }(X) = \text{posto }(X') = \text{posto }(X'XB) \leq$$
$$\leq \text{posto }(XB) \leq \text{posto }(X) \; .$$

4. <u>B é uma inversa generalizada de X</u>

$$B'X' = XB \;\Rightarrow\; B'X'X = XBX \;\Rightarrow\; X = XBX \; .$$

Vimos que, dada uma matriz X de dimensão nXp, qualquer vetor $y \in R^n$ pode ser decomposto de maneira única como

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

para qualquer $\hat{\beta}$ solução de $X'X\beta = X'y$, sendo que o vetor $X\hat{\beta}$ é a projeção ortogonal de y em C(X) e o vetor $(y-X\hat{\beta})$ é a projeção ortogonal de y em $C(X)^{\perp}$.

Por outro lado temos:

$$X\hat{\beta} = XBy$$
$$y - X\hat{\beta} = y - XBy = (I-XB)y \; .$$

Então XB é a matriz do operador projeção ortogonal em C(X) e (I-XB) é a matriz do operador projeção ortogonal em $C(X)^{\perp}$.

Estabeleceremos a seguinte definição:

<u>Definição</u>:

A toda matriz quadrada A, tal que para qualquer y de dimensões apropriadas Ay é a projeção ortogonal de y no espaço coluna de A, chamaremos de matriz do operador projeção ortogonal.

Já estabelecemos a unicidade do vetor projeção ortogonal. Sejam P_1 e P_2 duas matrizes do operador projeção ortogonal em um mesmo subespaço. Então

$$P_1 y = P_2 y \qquad \text{qualquer } y \, .$$

Portanto

$$P_1 = P_2 \, .$$

Outro aspecto a ser ressaltado em estimação de mínimos quadrados é que o importante é o espaço gerado pelas colunas da matriz de planejamento e não a particular parametrização utilizada, ou seja, o problema de se estimar E(Y) no modelo $y=X\beta+e$ é o mesmo se utilizarmos o modelo $y = W\Theta+e$ desde que C(X) seja idêntico a C(W). A estimativa de mínimos quadrados de E(y) é obtida pela projeção ortogonal do vetor de observações y no espaço coluna de X. Se W é qualquer matriz de dimensão n x k, tal que $C(W) = C(X)$, então a única projeção ortogonal de y em C(X) é dada por $W\hat{\Theta}$, onde $\hat{\Theta}$ é qualquer vetor satisfazendo $W'W\Theta = W'y$ e consequentemente $W\hat{\Theta} = X\hat{\beta}$.

Vimos até aqui que dado o modelo

$$y = X\beta+e \quad ,$$

para qualquer $\hat{\beta}$ solução das equações normais, $X\hat{\beta}$ é a projeção ortogonal de y em C(X) e $X\hat{\beta}$ é único. Por outro lado, se $X\beta$ é a pro jeção ortogonal de y em $C(X)$ então $\hat{\beta}$ é solução das equações normais. Caracterizamos também a matriz do operador projeção ortogonal de R^n em $C(X)$ como a matriz XB onde B é qualquer matriz satisfazendo $X'XB = X'$. Além disso XB é única, simétrica e idempotente.

Os resultados acima estabelecem condições gerais para que uma matriz real A seja matriz de um operador projeção ortogonal. Podemos estabelecer o seguinte teorema:

Teorema 1.5 - Para que uma matriz real A seja matriz de um operador projeção ortogonal é necessário e suficiente que A seja simétrica e idempotente.

Demonstração:

A necessidade já foi demonstrada acima.

Seja U um espaço vetorial n-dimensional e S um subespaço de U. Seja A a matriz do operador linear T tal que $T(U) = S$. Suponhamos que $A' = A$ e $A^2 = A$.

Para qualquer $y \in U$, y pode ser escrito como

$$y = Ay + (y-Ay) .$$

Temos que

$$(Ay)'(y-Ay) = y'A'y - y'A'Ay = y'Ay - y'Ay = 0 .$$

Portanto Ay é ortogonal a $(y-Ay)$ e o operador T é o operador projeção ortogonal de U em S sendo A sua matriz.

Ao lado de seu interesse em problemas de estimação de parâmetros em modelos lineares, as equações normais são de grande utilidade, como vimos, para caracterizar alguns conceitos importantes da álgebra linear.

1.4 - FUNÇÕES ESTIMAVEIS

Um conceito importante em modelos lineares é o de funções estimáveis. Antes de apresentarmos a definição formal de funções estimáveis procuraremos, através de um exemplo, motivar a definição.

Seja o seguinte modelo linear

$$y_1 = \mu + \alpha_1 + e_1$$

$$y_2 = \mu + \alpha_1 + e_2$$

$$y_3 = \mu + \alpha_2 + e_3 \quad .$$

Então temos:

$$y' = (y_1, y_2, y_3)$$

$$\beta' = (\mu, \alpha_1, \alpha_2)$$

$$e' = (e_1, e_2, e_3)$$

$$X = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \end{bmatrix} \quad .$$

Suponhamos que $E(e) = \emptyset$, $E(ee') = V$ e $|V| \neq 0$.

Qualquer combinação linear dos parâmetros pode ser escrita como $\lambda'\beta$, sendo $\lambda' = (\lambda_1, \lambda_2, \lambda_3)$. Consideremos a função

$\lambda'\beta = \mu$. Então $\lambda' = (1, 0, 0)$. O problema que será colocado é: será que existe alguma combinação linear das observações, b'y, tal que b'y é uma estimativa não viciada para $\lambda'\beta = \mu$. Para que a resposta seja afirmativa a seguinte igualdade deve ser satisfeita

$$E(a'y) = \mu$$

$$E(a'y) = a'X\beta = (a_1+a_2+a_3)\mu + (a_1+a_2)\alpha_1 + a_3\alpha_2 = \mu$$

$$\mu,\ \alpha_1, \alpha_2 \text{ quaisquer.}$$

Então

$$a_3 = 0$$

$$a_1 + a_2 = 0$$

e a igualdade não se verifica.

Portanto, não existe uma função linear das observações que seja uma estimativa não viciada para $\lambda'\beta = \mu$.

Por outro lado, se considerarmos a função paramétrica $\lambda'\beta = \alpha_1-\alpha_2$ podemos verificar facilmente que $a'y = y_1-y_3$ é uma estimativa não viciada para $\alpha_1-\alpha_2$.

Portanto, para algumas funções lineares dos parâmetros, existe uma estimativa não viciada formada por uma combinação linear das observações, sendo que existem situações onde isso não ocorre. Esses fatos nos sugerem a seguinte definição:

<u>Definição</u>: Uma função linear paramétrica $\lambda'\beta$ é dita estimável se existe uma combinação linear das observações b'y que é uma estimativa não viciada para $\lambda'\beta$.

Estabeleceremos algumas condições equivalentes para que uma função linear paramétrica seja estimável. Essas condições nos permitirão caracterizar o espaço de funções estimáveis.

Suponhamos que para o modelo linear

$$y = X\beta + e \quad , \quad E(e) = \emptyset \quad , \quad E(ee') = \sigma^2 V$$

a matriz V seja não singular. Nessas condições não temos restrições no espaço de parâmetros.

Seja $\lambda'\beta$ uma função estimável e $a'y$ uma estimativa não viciada para $\lambda'\beta$. Então

$$E(a'y) = a'X\beta = \lambda'\beta \quad \forall \beta \in R^p \quad \Rightarrow \quad a'X = \lambda' \ .$$

Portanto o vetor λ pertence ao espaço gerado pelas linhas da matriz X ($\lambda \in C(X')$).

Consideremos que agora a matriz V é arbitrária (possivelmente singular). Então existe uma matriz $T = \begin{bmatrix} T_1 \\ T_2 \end{bmatrix}$ não singular com T_1 de dimensão (k x n), k = posto V, tal que

$$\begin{bmatrix} T_1 \\ T_2 \end{bmatrix} V \begin{bmatrix} T_1' & T_2' \end{bmatrix} = \begin{bmatrix} T_1VT_1' & T_1VT_2' \\ T_2VT_1' & T_2VT_2' \end{bmatrix} = \begin{bmatrix} I_k & \emptyset \\ \emptyset & \emptyset \end{bmatrix} \ .$$

Se aplicamos a transformação T ao modelo $y = X\beta + e$ obtemos

$$Ty = \begin{bmatrix} T_1y \\ T_2y \end{bmatrix} = \begin{bmatrix} T_1X\beta \\ T_2X\beta \end{bmatrix} + \begin{bmatrix} T_1e \\ T_2e \end{bmatrix} \ .$$

$Cov(T_2e) = T_2VT_2' = \emptyset$ e temos que $T_2y = T_2X\beta$ com pro babilidade 1.

Nessas condições o espaço de parâmetros está restrito a uma variedade linear q.t.p.

Portanto, quando a matriz de covariância é arbitrária, é falso o teorema que estabelece que se $a'y$ é não viciado para $\lambda'\beta$ então $a'X = \lambda'$.

Por outro lado, qualquer β solução do sistema $T_2X\beta = T_2y$ pode ser escrita como:

$$\beta = (T_2X)^- T_2y + (I-(T_2X)^- T_2X)\gamma$$ sendo γ arbitrário e

$(T_2X)^-$ uma inversa generalizada de T_2X.

Se $a'y$ é uma estimativa não viciada para $\lambda'\beta$ então

$$(a'X-\lambda')\beta = 0$$

para todo β solução de $T_2X = T_2y$. Então

$$(a'X-\lambda')\beta = 0 \quad => \quad (a'X-\lambda')\ \{(T_2X)^- T_2y + (I-(T_2X)^- T_2X)\gamma\} = 0 \quad \forall\ \gamma$$

então

$$(a'X-\lambda')(I-(T_2X)^- T_2X) = \emptyset$$

$$=> (a'X-\lambda') = (a'X-\lambda')(T_2X)^- T_2X \ .$$

Assim

$$\lambda' = a'X - (a'X-\lambda')(T_2X)^- T_2X =$$

$$= (a' - (a'X - \lambda')(T_2X)^- T_2)X =$$

$$= b'X \ .$$

Portanto, se $\lambda'\beta$ é estimável então $\lambda \in C(X')$, V arbitrária.

Por outro lado, se $\lambda \in C(X')$ temos que

$$\lambda' = b'X.$$

Então a função paramétrica $\lambda'\beta$ é estimável pois $E(b'y) = b'X\beta = \lambda'\beta$.

Assim fica demonstrada a seguinte condição:

Condição 1 - Uma condição necessária e suficiente para que uma função linear $\lambda'\beta$ seja estimável é que o vetor λ pertença ao espaço gerado pelas linhas da matriz X.

Teorema 1.6 - Seja

$$b' = a' - (a'X-\lambda')(T_2X)^{-}T_2$$

com $a'y$ não viciado para $\lambda'\beta$. Então

$$b'y = a'y \text{ com probabilidade } 1.$$

Demonstração:

$$b'y = a'y - (a'X-\lambda')(T_2X)^{-}T_2y$$

$$\mathrm{Var}\{(a'X-\lambda')(T_2X)^{-}T_2y\} = cT_2VT_2'c = \emptyset \quad ,$$

$$c = (a'X-\lambda')(T_2X)^{-} .$$

Condição 2 - $\lambda'\beta$ é estimável se e somente se existe um vetor ρ tal que $X'X\rho = \lambda$.

Demonstração:

$C(X') = C(X'X)$. Portanto λ pertence ao espaço linha de X se e somente se λ pertence ao espaço coluna de X'X.

As funções $(X'_i)'\beta$, $i=1,2,\ldots,n$, onde X'_i é a i-ésima linha da matriz X, são funções estimáveis e para qualquer função estimável $\lambda'\beta$ temos que $\lambda = \sum_{i=1}^{n} a_i X'_i$. Portanto o conjunto das funções estimáveis forma um espaço vetorial de dimensão igual ao posto linha da matriz X.

Consideremos novamente o modelo linear

$$y_1 = \mu + \alpha_1 + e_1$$

$$y_2 = \mu + \alpha_1 + e_2$$

$$y_3 = \mu + \alpha_2 + e_3 \quad .$$

Para a função paramétrica $\lambda'\beta = \alpha_1 - \alpha_2$ existe uma combinação linear das observações que é uma estimativa não viciada para $\lambda'\beta$. De fato, como a matriz X não é de posto linha completo, existe mais que uma combinação linear das observações com essa propriedade. Qualquer combinação linear da forma

$$by_1 + (1-b)y_2 - y_3 \quad , \quad b \in R$$

é uma estimativa não viciada para $\alpha_1 - \alpha_2$.

Suponhamos que $E(ee') = \sigma^2 I$. Então

$$Var(\ (b,\ 1-b,\ -1)y) = \sigma^2 (b,\ 1-b,\ -1) \begin{bmatrix} b \\ 1-b \\ -1 \end{bmatrix} =$$

$$= b^2 + (1-b)^2 + 1 \; .$$

$Var((b, 1-b, -1)y)$ é uma função quadrática de b. Dentre os valores possíveis de b podemos escolher um b_0 tal que $V((b_0, 1-b_0, -1)y)$ seja mínima. No exemplo acima $b_0 = \frac{1}{2}$.

Estabeleceremos então como critério procurar dentre os estimadores lineares não viciados de $\lambda'\beta$ aquele que possui a propriedade de ter a menor variância.

Definição - Um estimador linear $a'y$ de uma função paramétrica estimável $\lambda'\beta$ é dito Melhor Estimador Linear não-Viciado se sua variância não excede a variância de qualquer outro estimador linear não viciado para $\lambda'\beta$. Denotaremos esse estimador por BLUE (Best Linear Unbiased Estimator).

1.5 - O TEOREMA DE GAUSS-MARKOFF

Seja o modelo

$$y = X\beta + e \quad , \quad E(e) = \emptyset \quad , \quad E(ee') = \sigma^2 I$$

e $\lambda'\beta$ uma função estimável. Estamos interessados em encontrar um vetor b, de dimensão n x 1, tal que $b'y$ seja BLUE para $\lambda'\beta$. Nosso objetivo então é encontrar um vetor b que minimize $E(a'y-\lambda'\beta)^2$ sujeito à condição que $E(a'y) = \lambda'\beta$. Antes precisamos dos dois lemas seguintes.

Lema 1.5.1 - Se $\lambda'\beta$ é estimável existe um único estimador linear não viciado para $\lambda'\beta$, $b^{*\prime}y$, com $b^* \in C(X)$. Se $b'y$ é um estimador não viciado para $\lambda'\beta$ então b^* é a projeção ortogonal de b em C(X).

Demonstração:

Se $\lambda'\beta$ é estimável existe um vetor b, de dimensão nx1, tal que $b'X = \lambda'$ e existe um vetor ρ, de dimensão px1, tal que $X'X\rho = \lambda$. Então temos

$$X'X\rho = X'b.$$

Essas equações têm estrutura de equações normais.

Fazendo $b^* = X\rho$ temos:

1. $b^{*\prime}y$ é não viciado para $\lambda'\beta$.

2. Se ρ_1 e ρ_2 são duas soluções de $X'X\rho = \lambda$

$$X\rho_1 = X\rho_2 \qquad \text{e} \qquad b^* \text{ é único.}$$

3. Para qualquer vetor b tal que $E(b'y) = \lambda'\beta$ temos que b^* é a projeção ortogonal de b em C(X).

Lema 1.5.2 - Se $\lambda'\beta$ é uma função estimável o BLUE de $\lambda'\beta$ é $\rho'X'y$ para algum ρ satisfazendo $X'X\rho = \lambda$. Além disso $\rho'X'y$ é único.

Demonstração:

$$E(\rho'X'y) = \rho'X'X\beta = \lambda'\beta$$

Consideremos $(\rho'X'+\delta')y$ outro estimador qualquer não viciado para $\lambda'\beta$. Então

$$E((\rho'X'+\delta')y) = E(\rho'X'y) + E(\delta'y) = \lambda'\beta+\delta'X\beta$$

$$\Rightarrow \delta' X\beta = 0 \quad \forall\ \beta \quad \Rightarrow \delta' X = \emptyset\ .$$

Por outro lado

$$\mathrm{Var}\{(\rho' X' + \delta')y\} = \sigma^2 (X\rho+\delta)'(X\rho+\delta) =$$

$$= \sigma^2 (\rho' X' X\rho + 2\rho' X'\delta + \delta'\delta) =$$

$$= \sigma^2 (\rho' X' X\rho + \delta'\delta) \geq \sigma^2 \rho' X' X\rho = \mathrm{Var}(\rho' X' y)\ .$$

Portanto existe um único estimador linear $b^{*\prime}y = \rho' X' y$ que é BLUE de $\lambda'\beta$.

<u>Teorema (Gauss-Markoff)</u> - Seja o modelo

$$y = X\beta + e \quad , \quad E(e) = \emptyset \quad , \quad E(ee') = \sigma^2 I\ .$$

O BLUE de $\lambda'\beta$, $\lambda'\beta$ estimável, é o único vetor $\lambda'\hat{\beta}$ onde $\hat{\beta}$ é qualquer solução das equações normais.

<u>Demonstração</u>:

O BLUE de $\lambda'\beta$ é $\widehat{\lambda'\beta} = \rho' X' y$, ρ satisfazendo $X' X\rho = \lambda$.

Seja $\hat{\beta}$ tal que $X' X\hat{\beta} = X' y$. Então

$$\widehat{\lambda'\beta} = \rho' X' y = \rho' X' X\hat{\beta} = (X' X\rho)'\hat{\beta} = \lambda'\hat{\beta}.$$

<u>Corolário</u> - Se $(\lambda_1'\beta, \ldots, \lambda_k'\beta)$ são funções estimáveis então qualquer combinação linear $\Theta = \sum_{i=1}^{k} h_i \lambda_i'\beta$ é estimável e o BLUE de Θ é $\sum_{i=1}^{k} h_i \lambda_i'\hat{\beta}$, $\hat{\beta}$ qualquer solução das equações nor-

mais.

Demonstração:

Já mostramos anteriormente que o conjunto das funções estimáveis é um espaço vetorial. Portanto, qualquer combinação linear de funções estimáveis é estimável. Por outro lado, fazendo $\sum_{i=1}^{k} h_i\lambda_i'\beta = \lambda'\beta$, $\lambda' = \Sigma h_i\lambda_i'$, temos que o BLUE de $\lambda'\beta$ é $\lambda'\hat{\beta}$, sendo $\hat{\beta}$ qualquer solução das equações normais.

A hipótese básica do teorema de Gauss-Markoff é que os erros sejam não correlacionados e que tenham a mesma variância. Se a matriz de covariância é da forma σ^2V, V simétrica positiva definida, uma transformação linear conveniente no modelo torna possível explicitar a estrutura do BLUE de $\lambda'\beta$, $\lambda'\beta$ estimável.

Seja o modelo linear

$$y = X\beta+e \quad , \quad E(e) = \emptyset \quad , \quad E(ee') = \sigma^2V \quad , \quad |V| \neq 0.$$

Como V é simétrica positiva definida existe uma matriz T não singular tal que

$$TVT' = I.$$

Consideremos a seguinte transformação linear

$$z = Ty = TX\beta + Te = W\beta+f$$

$$E(f) = E(Te) = \emptyset$$

$$E(ff') = \sigma^2 TVT' = \sigma^2 I \ .$$

Qualquer linha da matriz X pode ser escrita como uma combinação linear das linhas da matriz W. Então qualquer função que é estimável sob o modelo $y = X\beta+e$ também é estimável sob o modelo $z = W\beta+f$. O modelo $z = W\beta+f$ satisfaz as condições do Teorema de Gauss-Markoff. Então, se $\lambda'\beta$ é estimável, o BLUE de $\lambda'\beta$ é $\lambda'\hat{\beta}$ onde $\hat{\beta}$ é qualquer solução de

$$W'W'\beta = W'z.$$

As equações normais acima podem ser reescritas como:

$$X'T'TX\beta = X'T'Ty$$

ou

$$X'V^{-1}X\beta = X'V^{-1}y \qquad (1.5.1) \ .$$

As equações (1.5.1) são conhecidas como equações de Aitken.

O teorema de Gauss-Markoff pode ser enunciado então de uma forma mais geral.

<u>Teorema</u> - Seja o modelo linear

$$y = X\beta+e \quad , \quad E(e) = \emptyset \quad , \quad E(ee') = \sigma^2 V \quad , \quad |V| \neq 0 \ .$$

O BLUE de qualquer função estimável $\lambda'\beta$ é $\lambda'\hat{\beta}$ onde $\hat{\beta}$ é qualquer solução de

$$X'V^{-1}X\beta = X'V^{-1}y.$$

1.6 - INTERPRETAÇÃO GEOMETRICA DO TEOREMA DE GAUSS-MARKOFF

Seja o modelo linear

$$y = X\beta+e \quad , \quad E(e) = \emptyset \quad , \quad E(ee') = \sigma^2 I \ .$$

Qualquer matriz B satisfazendo $X'XB = X'$, cuja existência foi demonstrada anteriormente, é uma inversa generalizada de X e B' é uma inversa generalizada de X'.

Seja $\lambda'\beta$ estimável. Então o sistema $X'b = \lambda$ é consistente. Qualquer vetor b solução de $X'b = \lambda$ pode ser escrito como:

$$b = B'\lambda + (I-B'X')\gamma = B'\lambda + (I-XB)\gamma \ , \quad \gamma \text{ arbitrário.}$$

Temos:

$$B'\lambda = B'X'b = XBb.$$

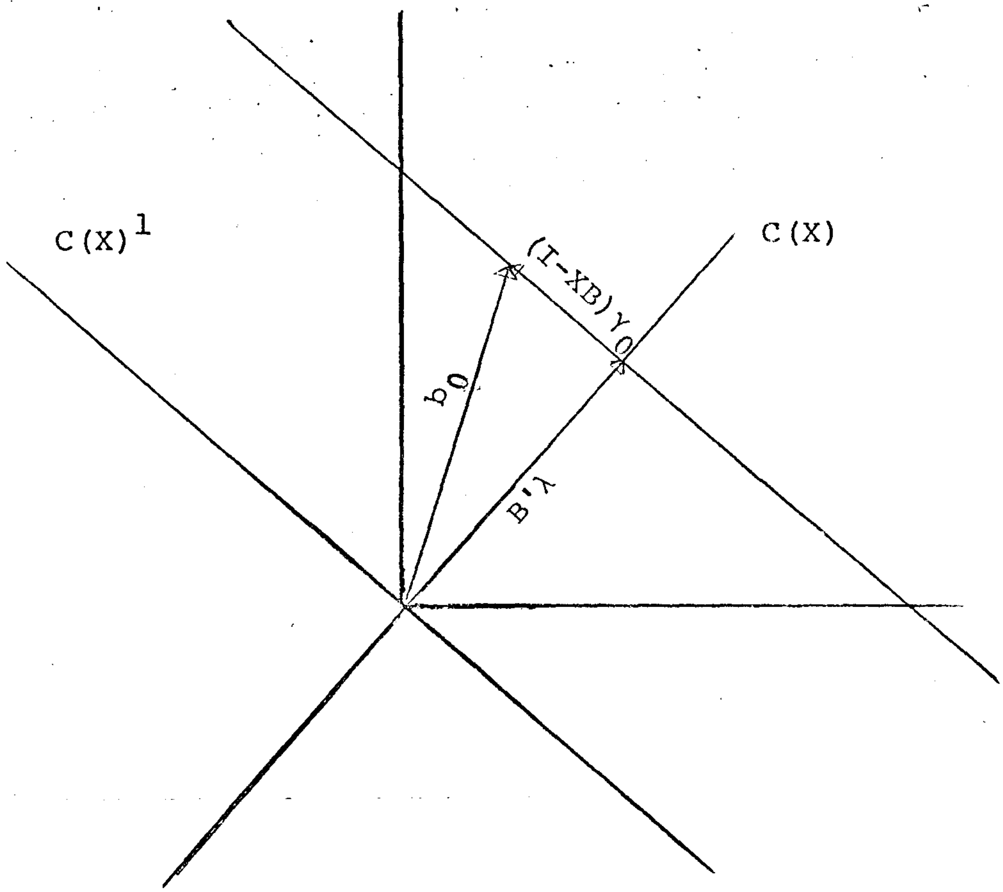
Portanto, $B'\lambda$ é a projeção ortogonal de qualquer solução de $X'b = \lambda$ em $C(X)$. Além disso $B'\lambda$ é único.

O operador $(I-XB)$ projeta o vetor γ ortogonalmente em $C(X)^{\perp}$. A decomposição de qualquer vetor b satisfazendo $X'b = \lambda$ em $b_1 + b_2 = B'\lambda + (I-XB)\gamma$, $b_1 \in C(X)$ e $b_2 \in C(X)^{\perp}$ é única.

A matriz de covariância dos erros é $\sigma^2 I$. Então $Var(b'y) = \sigma^2 b'b$. Portanto, o BLUE de $\lambda'\beta$ é dado por $b'y$ onde b é o vetor de menor norma satisfazendo $X'b = \lambda$ e esse vetor é obtido fazendo-se $\gamma = \emptyset$. Então a estimativa de menor variância dentre as combinações lineares de y que sejam não viciadas para $\lambda'\beta$ é dada por $\lambda'By$ e é única. Mas By é precisamente uma solução de $X'X\beta=X'y$. Portanto o BLUE de $\lambda'\beta$ é $\lambda'\hat{\beta}$ onde $\hat{\beta}$ é qualquer solução das equações normais e isto é o mesmo que afirma o Teorema de Gauss-Markoff.

O conjunto formado pelos vetores b da forma $B'\lambda + (I-XB)\gamma$, γ arbitrário, é uma variedade linear U paralela a $C(X)^1$ e a distância entre U e $C(X)^1$ é dada por $\| B'\lambda \|$ (considerando a norma euclidiana).

O diagrama abaixo em R^2 ilustra essas idéias.



CAPÍTULO 2

CONDIÇÕES PARA QUE O ESMQ SEJA BLUE

Definimos no capítulo anterior como ESMQ (Estimador Simples de Mínimos Quadrados) de $\lambda'\beta$, $\lambda'\beta$ estimável, ao vetor $\lambda'\hat{\beta}$ onde $\hat{\beta}$ é qualquer solução de $X'X\beta = X'y$.

Quando a matriz de covariância dos erros, $\sigma^2 V$, é não singular o BLUE (melhor estimador não viciado) de $\lambda'\beta$ é $\lambda'\tilde{\beta}$ onde $\tilde{\beta}$ é qualquer solução de $X'V^{-1}X\beta = X'V^{-1}y$, resultado esse estabelecido pelo teorema de Gauss-Markoff. Se $V = I$ então $\lambda'\hat{\beta} = \lambda'\tilde{\beta}$.

Apresentaremos nesse capítulo a condição necessária e suficiente para que o ESMQ de $\lambda'\beta$, $\lambda'\beta$ estimável, seja BLUE, considerando-se o modelo linear geral com matriz de covariância arbitrária.

Essa condição foi apresentada pela primeira vez, na forma geral, por Zyskind (1967) e ela é conhecida como a "CONDIÇÃO DOS AUTOVETORES".

Para a demonstração do teorema principal desse capítulo precisamos da forma canônica do modelo linear que será estabelecido logo em seguida.

Considere o modelo linear

$$y = X\beta + e, \quad E(e) = \emptyset \quad E(ee') = \sigma^2 V \qquad (2.1)$$

Seja Θ_1' uma base ortonormal para $C(X)$, o subespaço r-dimensional gerado pelas colunas da matriz X, e seja Θ_0' qualquer base ortonormal para o subespaço $C(X)^{\perp}$, complementar ortogonal de $C(X)$, de dimensão n-r.

Então $\Theta_0 X = \emptyset$, $\emptyset$ matriz nula.

A matriz $\Theta = \begin{bmatrix} \Theta_1 \\ \Theta_2 \end{bmatrix}$ aplicada ao modelo linear produz a seguinte transformação:

$$\Theta y = \begin{bmatrix} \Theta_1 y \\ \Theta_0 y \end{bmatrix} = \begin{bmatrix} z_1 \\ z_0 \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} \Theta_1 e \\ \Theta_0 e \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} f_1 \\ f_0 \end{bmatrix} .$$

Nós nos referimos à forma

$$\begin{bmatrix} z_1 \\ z_0 \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} f_1 \\ f_2 \end{bmatrix} \qquad (2.2)$$

como forma canônica do modelo linear.

Para a demonstração do teorema principal desse capítulo usaremos o seguinte lema:

<u>Lema 2.1</u> - Seja a matriz Θ como a definida acima. Então qualquer função linear estimável $\lambda'\beta$ pode ser escrita como $\rho'\Theta_1 X\beta$.

<u>Demonstração</u>:

$\lambda'\beta$ é estimável se e somente se λ puder ser escrito como uma combinação linear das linhas de X. O que temos que provar essencialmente é que as r linhas de $\Theta_1 X$ geram o espaço linha de X.

O sistema

$(\Theta X)'b_i = X'_i$, $i = 1, 2, \ldots, n$ (X'_i - i-ésima coluna de X')

é consistente pois

$$a'(\Theta X)' = \emptyset \Rightarrow a'X'\Theta' = \emptyset \Rightarrow a'X = \emptyset \Rightarrow a'X'_i = 0, \quad i=1,2,\ldots n.$$

Portanto qualquer linha de X pode ser escrita como combinação linear das linhas de ΘX. Mas $\Theta X = \begin{bmatrix} \Theta_1 X \\ \emptyset \end{bmatrix}$. Portanto as r linhas de $\Theta_1 X$ geram o espaço linha da matriz X.

O teorema seguinte, devido a Zyskind (1967), estabelece a condição necessária e suficiente sobre o modelo linear geral para que os estimadores simples de mínimos quadrados de todas as funções lineares estimáveis sejam também BLUE's.

Teorema 2.1 - Uma condição necessária e suficiente para que todos os estimadores lineares simples de mínimos quadrados sejam também BLUE's das correspondentes funções estimáveis $\lambda'\beta$ no modelo linear $y = X\beta+e$, $E(e) = \emptyset$, $E(ee') = \sigma^2 V$, V simétrica positiva semidefinida, posto(X) = r, é que exista um subconjunto de r autovetores ortogonais de V, os quais formam uma base para o espaço coluna da matriz X.

Demonstração (suficiência):

Suponha que V tem r autovetores ortogonais, os quais formam uma base para $C(X)$. Então V tem também $(n-r)$ autovetores ortogonais os quais formam uma base para $C(X)^{\perp}$. Sejam Θ_1' e Θ_0' duas matrizes cujas colunas são formadas respectivamente por cada um dos dois conjuntos de vetores acima.

A matriz $\Theta' = (\Theta_1', \Theta_0')$ é tal que

$\Theta V \Theta' = D$, sendo D uma matriz diagonal e $\Theta_0 X = \emptyset$.

Consideremos a forma canônica obtida pela seguinte transformação ortogonal:

$$\Theta y = \begin{bmatrix} \Theta_1 y \\ \Theta_0 y \end{bmatrix} = \begin{bmatrix} z_1 \\ z_0 \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} \Theta_1 e \\ \Theta_0 e \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} f_1 \\ f_0 \end{bmatrix}$$

Temos que

$$\mathrm{Cov}\begin{bmatrix} Z_1 \\ Z_0 \end{bmatrix} = \sigma^2 \Theta V \Theta' = \sigma^2 D = \sigma^2 \begin{bmatrix} D_1 & \emptyset \\ \emptyset & D_2 \end{bmatrix}.$$

Seja $\lambda'\beta = \rho'\Theta_1 X\beta$ uma função estimável. O estimador simples de mínimos quadrados de $\rho'\Theta_1 X\beta$ é $\rho'\Theta_1 X\hat{\beta}$, $\hat{\beta}$ solução de $X'X\beta = X'y$. Mas

$$\rho'\Theta_1 X\hat{\beta} = \rho'\Theta_1(X\hat{\beta}+y-y) = \rho'\Theta_1(y-(y-X\hat{\beta})) = \rho'\Theta_1 y.$$

$\rho'\Theta_1 y$ é único e $\rho'\Theta_1 \in C(X)$.

Seja $a'y$ um estimador não viciado para $\rho'\Theta_1 X\beta$. Então, como vimos no capítulo 1, existe um vetor b satisfazendo $\lambda'=b'x$ e $\mathrm{Var}(a'y) = \mathrm{Var}(b'y)$. Além disso o vetor b pode ser decomposto de maneira única como

$$b = b_1 + b_2 \ , \quad b_1 \in C(X), \quad b_2 \in C(X)^{\perp} \quad e \quad b_1 = \rho'\Theta_1 .$$

Podemos escrever $b_2 = \alpha'\Theta_0$. Então

$$b = \rho'\Theta_1 + \alpha'\Theta_0 \ , \quad \rho'\Theta_1 \in C(X) \quad e \quad \alpha'\Theta_0 \in C(X)^{\perp}$$

$$Var(b'y) = Var(\ (\rho'\Theta_1 + \alpha'\Theta_0)y) =$$

$$= Var(\rho'\Theta_1 y) + Var(\alpha'\Theta_0 y) + 2Cov(\rho'\Theta_1 y, \alpha'\Theta_0 y)$$

$$= Var(\rho'\Theta_1 y) + Var(\alpha'\Theta_0 y) \geq Var(\rho'\Theta_1 y).$$

Portanto o ESMQ de $\rho'\Theta_1 X\beta$ é também BLUE.

(Necessidade) - Suponha que para qualquer função estimável $\lambda'\beta$ o estimador simples de mínimos quadrados é também BLUE. Qualquer matriz ortogonal $\Theta' = (\Theta_1', \Theta_0')$ com Θ_1 de dimensão rXn e Θ_0 de dimensão (n-r)Xn construída de tal forma que $\Theta_0 X = \emptyset$ quando aplicada ao modelo linear $y = X\beta + e$ produz:

$$\Theta y = \begin{bmatrix} \Theta_1 y \\ \Theta_0 y \end{bmatrix} = \begin{bmatrix} z_1 \\ z_0 \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} \Theta_1 e \\ \Theta_0 e \end{bmatrix} = \begin{bmatrix} \Theta_1 X\beta \\ \emptyset \end{bmatrix} + \begin{bmatrix} f_1 \\ f_0 \end{bmatrix}$$

Seja $\rho'\Theta_1 X\beta$ uma função estimável.

O ESMQ de $\rho'\Theta_1 X\beta$ é $\rho'\Theta_1 y = \rho' Z_1$.

Temos por hipótese que $\rho' Z_1$ é BLUE de $\rho'\Theta_1 X\beta$.

A função $\rho' Z_1 + \alpha' Z_0$ é um estimador não viciado para

$\rho'\Theta_1 X\beta$ qualquer α pois $E(\rho'Z_1 + \alpha'Z_0) = E(\rho'Z_1) = \rho'\Theta_1 X\beta$.

$$Var\ (\rho'Z_1 + \alpha'Z_0) = Var(\rho'Z_1)+Var(\alpha'Z_0)+2Cov(\rho'Z_1,\alpha'Z_0) =$$

$$= \rho'\Theta_1 V\Theta_1'\rho+\alpha'\Theta_0 V\Theta_0'\alpha+2\rho'\Theta_1 V\Theta_0'\alpha \geq \rho'\Theta_1 V\Theta_1'\rho \quad \forall\ \alpha$$

Portanto $\rho'\Theta_1 V\Theta_0'\alpha = 0 \quad \forall\alpha \Rightarrow cov(Z_1, Z_0) = \emptyset$.

Então $cov \begin{bmatrix} Z_1 \\ Z_0 \end{bmatrix} = \begin{bmatrix} {}_rW_{1r} & \emptyset \\ \emptyset & {}_{n-r}W_{0n-r} \end{bmatrix}$

Seja agora uma matriz ortogonal R_1 de dimensão rXr e uma matriz ortogonal R_0 de dimensão $(n-r)X(n-r)$ tal que $R_1W_1R_1'=D_1$ e $R_0W_0R_0'=D_0$. Existem R_1 e R_0 nessas condições pois W_1 e W_0 são matrizes simétricas positivas semidefinidas.

Seja a matriz

$$S = \begin{bmatrix} R_1 & \emptyset \\ \emptyset & R_0 \end{bmatrix} \begin{bmatrix} \Theta_1 \\ \Theta_0 \end{bmatrix} = \begin{bmatrix} R_1 & \Theta_1 \\ R_0 & \Theta_0 \end{bmatrix}.$$

S é ortogonal

$$\begin{bmatrix} R_1 & \Theta_1 \\ R_0 & \Theta_0 \end{bmatrix} (\Theta_1'R_1',\ \Theta_0'R_0') = \begin{bmatrix} R_1\Theta_1\Theta_1'R_1' & R_1\Theta_1\Theta_0'R_0' \\ R_0\Theta_0\Theta_1'R_1' & R_0\Theta_0'\Theta_0'R_0 \end{bmatrix} =$$

$$= \begin{bmatrix} {}_rI_r & \emptyset \\ \emptyset & {}_{n-r}I_{n-r} \end{bmatrix} = I .$$

Multiplicando-se a matriz V à esquerda por S e à direita por S' temos:

$$SVS' = \begin{bmatrix} R_1 \Theta_1 \\ R_0 \Theta_0 \end{bmatrix} V \, (\Theta_1'R_1', \; \Theta_0'R_0') =$$

$$= \begin{bmatrix} R_1\Theta_1V_1\Theta_1'R_1' & R_1\Theta_1V_1\Theta_0'R_0' \\ R_0\Theta_0V\Theta_1'R_1 & R_0\Theta_0V\Theta_0'R_0' \end{bmatrix} =$$

$$= \begin{bmatrix} R_1W_1R_1' & \emptyset \\ \emptyset & R_0W_0R_0' \end{bmatrix} = \begin{bmatrix} D_1 & \emptyset \\ \emptyset & D_0 \end{bmatrix} .$$

As linhas da matriz S formam um conjunto de autovetores de V. Além disso as linhas $R_1\Theta_1$ são combinações lineares das linhas de Θ_1 e geram $C(X)$ e as linhas de $R_0\Theta_0$ são combinações lineares das linhas de Θ_0 e geram $C(X)^{\perp}$. Então as colunas de $\Theta_1'R_1'$ e $\Theta_0'R_0'$ formam bases ortogonais de $C(X)$ e $C(X)^{\perp}$ respectivamente.

Nós nos referiremos daqui por diante ao teorema 2.1 como a "condição dos autovetores".

Se podemos de algum modo determinar que V tem um subconjunto de r autovetores formando uma base para o espaço coluna da matriz X, então nós sabemos que todos os estimadores simples de mínimos quadrados são também BLUE's.

A condição dos autovetores, na forma do teorema 2.1, não é necessariamente simples de ser aplicada na prática. Entre tanto é possível se estabelecer algumas condições equivalentes que tornam mais fácil de se verificar a validade do teorema 2.1. Além disso essas condições equivalentes possibilitam uma interpretação geométrica mais clara da "condição dos autovetores".

Equivalência 1 - Um conjunto de r autovetores de V pertence ao subespaço C(X) r-dimensional, e portanto gera C(X) se e somente se C(X) é um subespaço invariante de V.

Demonstração:

(=>) O subespaço C(X) tem dimensão r. Sejam $\theta_1^{(1)}, \theta_1^{(2)}, \ldots, \theta_1^{(r)}$ r autovetores de V tal que $\theta_1^{(i)}$, $i = 1,2, \ldots, r$ pertence a C(X) e a coleção de vetores $\{\theta_1^{(i)}, i = 1,2, \ldots r\}$ gera C(X). Portanto, para qualquer vetor b tal que $b \in C(X)$ temos que

$$b = \sum_{i=1}^{r} \alpha_i \theta_1^{(i)} , \qquad \alpha_i \text{ constantes.}$$

Temos então:

$$Vb = V \sum_{i=1}^{r} \alpha_i \theta_1^{(i)} = \sum_{i=1}^{r} \alpha_i V \theta_1^{(i)} = \sum_{i=1}^{r} \alpha_i \lambda_i \theta_1^{(i)} =$$

$$= \sum_{i=1}^{r} \alpha_i ' \theta_1^{(i)}$$

portanto $Vb \in C(X)$.

(<=) Suponha que qualquer vetor $b \in C(X)$, $Vb \in C(X)$.

Sejam Θ_1 e Θ_0 matrizes com dimensão rXn e n-rXn respectivamente tais que as linhas de Θ_1 e Θ_0 geram $C(X)$ e $C(X)^{\perp}$ respectivamente.

$V\Theta_1'$ é então uma matriz cujas colunas pertencem a $C(X)$. Portanto,

$$\Theta_0 V \Theta_1' = \emptyset \quad \Rightarrow \quad \Theta_1 V \Theta_0' = \emptyset .$$

Pela mesma construção desenvolvida na prova da necessidade do teorema 2.1, a matriz V tem um subconjunto de r autovetores formando uma base para $C(X)$.

Equivalência 2 - $C(X)$ é um subespaço invariante de V se e somente se $VX = XQ$ para alguma matriz Q.

Demonstração:

(=>) Se $C(X)$ é um subespaço invariante de V então para qualquer vetor $b \in C(X)$ o vetor $Vb \in C(X)$.

Para qualquer $y \in R^p$ o vetor $Xy \in C(X)$.

Portanto,

$VXy = Xq$ para algum $q \in R^p$.

Façamos $y = e_i$, $i = 1, 2, \ldots, p$ onde e_i é um vetor com coordenadas iguais a zero exceto na i-ésima linha, a qual tem a unidade como elemento.

Temos então que:

$VX = XQ$ para alguma matriz Q.

(<=) Para qualquer vetor b tal que $b \in C(X)$, exis te $y \in R^p$ tal que $b = Xy$.

Então,

$$Vb = VXy = XQy \in C(X).$$

Portanto $C(X)$ é um subespaço invariante de V.

Antes de enunciarmos a equivalência 3 precisamos de um resultado conhecido da Álgebra Linear, o qual iremos enunciar sem demonstrar.

Teorema 2.2 - Seja Z um espaço vetorial n-dimensional e T um operador linear sobre Z. Sejam $W_1, W_2, \ldots, W_k$ subespaços de Z tal que $Z = W_1 \oplus W_2 \oplus \ldots \oplus W_k$ e $E_1, E_2, \ldots, E_k$ operadores sobre Z tais que

a) cada E_i é uma projeção $(E_i{}^2 = E_i)$,

b) $E_i E_j = \emptyset$ se $i \neq j$,

c) $I = E_j + E_2 + \ldots + E_k$,

d) a imagem de E_i é W_i .

Então, uma condição necessária e suficiente para que cada subespaço seja invariante sob T é que T comute com cada uma das projeções E_i, isto é,

$$TE_i = E_i T \quad i = 1, 2, \ldots, k.$$

Equivalência 3 - Seja P matriz do operador projeção ortogonal de R^n em $C(X)$. Então $C(X)$ é invariante sob V se e somente se

$$PV = VP.$$

Demonstração:

O ESMQ de $E(y) = X\beta$ é $X\hat{\beta} = Py$ onde $\hat{\beta}$ é qualquer vetor satisfazendo $X'X\beta = X'y$ e P é a matriz do operador projeção ortogonal do espaço de dados, R^n, no espaço coluna de X. Já verificamos anteriormente que P é simétrica e idempotente e que $C(P) = C(X)$. A matriz $(I-P)$ é a matriz do operador projeção ortogonal do espaço de dados no espaço complemento ortogonal do espaço coluna de X, $C(X)^{\perp}$.

Podemos escrever

$$R^n = C(X) \oplus C(X)^{\perp} \quad ,$$

com P e (I-P) satisfazendo:

a) P e (I-P) são projeções $(P^2 = P,\ (I-P)^2 = I-P)$,

b) $P(I-P) = \emptyset$,

c) $I = P + (I-P)$,

d) qualquer $y \in R^n$ $P_y \in C(X)$ e $(I-P)y \in C(X)^{\perp}$.

Aplicando o Teorema 2.2 temos que $C(X)$ é invariante sob V se e somente se $PV = VP$.

Se dispomos de um algoritmo para calcularmos a matriz P podemos verificar se os estimadores simples de mínimos quadrados são também BLUE's verificando se o produto VP é simétrico pois

$$PV = VP = (PV)' \ .$$

Podemos resumir o Teorema 2.1 com as condições equivalentes ao mesmo no seguinte teorema:

Teorema 2.3 - Seja o modelo $y = X\beta + e$ onde X é uma matriz conhecida nXp de posto $r \leq p$, $E(e) = \emptyset$, $E(ee') = \sigma^2 V$, V simétrica positiva semidefinida. Cada uma das seguintes condições é necessária e suficiente para que o ESMQ de qualquer função paramétrica linear estimável seja BLUE .

1. Existe um subconjunto de r autovetores de V formando uma base para o espaço coluna de X.

2. Uma reparametrização de posto completo existe tal que $E(y) = X\beta = W\Theta$ onde as colunas de W são autovetores ortogonais de V.

3. C(X) é um subespaço invariante de V.

4. $VX = XQ$ para alguma matriz Q.

5. O produto VP é simétrico.

Se a matriz de covariância é singular, os valores permitidos para o vetor β ficam restritos a um subconjunto de R^p.

Já estabelecemos no capítulo 1 que quando a matriz de covariância é arbitrária é falsa a afirmativa

$$E(a'y) = \lambda'\beta \Rightarrow a'X = \lambda' .$$

A falsidade desse resultado foi estabelecida por Zyskind (1967) embora Rao (1971) tenha utilizado esse resultado, fato esse apontado por Kempthorne (197?).

A restrição no espaço de parâmetros, provocada por singularidade na matriz de covariância, parece ser um fato ainda não muito compreendido.

Se estabelecemos a "condição dos autovetores" na forma

$$\text{ESMQ} \equiv \text{BLUE se e somente se } VX = XQ$$

para alguma matriz Q

podemos concluir imediatamente que ESMQ $\equiv$ BLUE => $V(C(X)) \subset C(X)$.

Nessas condições é falso o Teorema 1 estabelecido por Watson (1967). A segunda parte do teorema diz que se o ESMQ é BLUE para todas as funções estimáveis então $V(C(X))=C(X)$, sendo a matriz de covariância arbitrária.

Mostraremos a falsidade dessa afirmativa através da construção de um contra exemplo.

Seja o modelo linear

$$y_1 = \mu + e_1$$

$$y_2 = \mu - \alpha_1 + e_2$$

$$y_3 = \mu + \alpha_1 + e_3$$

$$E(e_i) = 0 \quad , \quad V = E(ee') = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix}$$

Então V é singular e posto $(V) = 1$.

A matriz de planejamento do modelo é

$$X = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix}$$

Seja a matriz $Q = \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix}$. Então,

$$VX = \begin{bmatrix} 1 & 1 & 1 \\ 1 & 1 & 1 \\ 1 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 3 & 0 \\ 3 & 0 \end{bmatrix}$$

e

$$XQ = \begin{bmatrix} 1 & 0 \\ 1 & -1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 3 & 0 \\ 0 & 0 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 3 & 0 \\ 3 & 0 \end{bmatrix}.$$

Portanto $VX = XQ$ para alguma matriz Q e temos que o ESMQ de qualquer função estimável é BLUE. Mas

$$V(C(X)) \neq C(X).$$

CAPITULO 3

BLUE COM MATRIZ DE COVARIANCIA ARBITRARIA

Nesse capítulo apresentaremos condições sobre o vetor b de tal forma que b'y seja estimativa não viciada de menor variância para uma dada função estimável $\lambda'\beta$ sem impor condições sobre a matriz de covariância de erros.

Dado o modelo linear $y = X\beta+e$, $E(e) = \emptyset$, $E(ee') = \sigma^2 V$, sempre que a matriz V é positiva definida, o teorema de Gauss-Markoff estabelece que o BLUE de uma função estimável $\lambda'\beta$ é $\lambda'\hat{\beta}$, $\hat{\beta}$ solução de $X'V^{-1}X\beta = X'V^{-1}y$. Portanto, sendo $(X'V^{-1}X)^-$ uma inversa generalizada de $X'V^{-1}X$, o conjunto de vetores b tal que b'y é BLUE de $\lambda'\beta$ é dado por

$\lambda' \{(X'V^{-1}X)^- X'V^{-1}y + (I-(X'V^{-1}X)^-(X'V^{-1}X))\gamma\}$, γ um vetor arbitrário.

O Teorema de Gauss-Markoff, na sua versão clássica, não trata da situação em que a matriz V é singular. Modelos lineares em que a matriz V é singular surgem frequentemente no tratamento de problemas práticos. Matrizes com uma estrutura de covariância singular são induzidas por experimentos finitos aleatorizados, como por exemplo o caso de blocos aleatorizados. Trataremos de alguns desses experimentos no próximo capítulo.

Mesmo para o modelo $y = X\beta + e$, $E(e) = \emptyset$, $E(ee')=\sigma^2 I$, a introdução de restrições nos parâmetros da forma $A\beta = c$ conduz ao modelo

$$z = \binom{y}{c} = \binom{X}{A}\beta + \binom{e}{\emptyset} = W\beta + f$$

e a matriz de covariância de f é

$$\sigma^2 V = \begin{bmatrix} In & \emptyset \\ \emptyset & \emptyset \end{bmatrix}.$$

Diversas tentativas têm sido feitas para se produzir uma teoria geral sobre BLUE de funções estimáveis.

Uma dessas tentativas está apresentada num artigo de Zyskind e Martin (1969). Nesse artigo o teorema de Gauss-Morkoff é generalizado no sentido de se construir uma classe de inversas generalizadas da matriz V, G_V^*, de tal forma que para qualquer função estimável $\lambda'\beta$, o BLUE de $\lambda'\beta$ é dado por $\lambda'\hat{\beta}$, onde $\hat{\beta}$ é solução das equações

$$X'V^* X\beta = X'V^* y, \qquad V^* \in G_V^* .$$

É interessante notar que em muitas situações o conjunto G_V^* está estritamente contido no conjunto G_V das inversas generalizadas da matriz V. Como exemplo, para o modelo

$$\begin{matrix} y_1 = \alpha_1 + e_1 \\ y_2 = \alpha_2 + e_2 \end{matrix} \qquad E(e) = \emptyset \qquad E(ee') = I ,$$

a introdução do contraste $\alpha_1 + \alpha_2 = 0$ produz o modelo

$$\begin{bmatrix} y_1 \\ y_2 \\ 0 \end{bmatrix} = \begin{bmatrix} 10 \\ 01 \\ 11 \end{bmatrix} \begin{bmatrix} \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \\ 0 \end{bmatrix} \qquad \text{sendo}$$

$$V = E(ff') = \begin{bmatrix} I_2 & \emptyset \\ \emptyset & \emptyset \end{bmatrix}, \qquad f = \begin{bmatrix} e_1 \\ e_2 \\ 0 \end{bmatrix}$$

A própria matriz V é uma inversa generalizada de si mesma. Se considerarmos a função estimável $\lambda'\beta = \alpha_1$, soluções de $X'VX\beta = X'Vy$ não conduzem ao BLUE de $\lambda'\beta$.

O procedimento para o cálculo de uma matriz V*, na forma proposta por Zyskind e Martin (1969), envolve um algoritmo não muito simples.

Em um de seus artigos, Rao (1971) procura também dar um tratamento unificado para a teoria de estimação linear. No capítulo 3 do referido artigo, Rao estabelece que o BLUE de $\lambda'\beta$ é $\lambda'\hat{\beta}$ onde $\hat{\beta} = c_2'y$ ou $\hat{\beta} = c_3y$, onde c_2 e c_3 são matrizes tais que a matriz $\begin{bmatrix} c_1 & c_2 \\ c_3 & -c_4 \end{bmatrix}$ é uma inversa generalizada da matriz $\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix}$, X matriz de planejamento e V matriz de covariância dos erros.

Uma contribuição recente nesse assunto, e sobre a qual nos prenderemos com detalhes, está contida em um artigo não publicado de Kempthorne (197?). A força do citado artigo está em desen volver os resultados de uma forma bem elementar, utilizando um instrumental matemático pouco sofisticado. No capítulo 1 foram desenvolvidas algumas idéias sobre modelos lineares e funções estimáveis a partir das informações contidas nas equações normais. O desenvolvimento apresentado no trabalho de Kempthorne utiliza em grande parte as equações normais. As condições para a igualdade entre o BLUE e o ESMQ de uma função estimável, estabelecidas por Zyskind e que tratamos no capítulo 2, são deduzidas como consequência direta dos teoremas que estabelecem as condições sobre o vetor b de tal forma que $b'y$ seja BLUE de $\lambda'\beta$.

Seja o modelo $y = X\beta + e$, $E(e) = \emptyset$, $E(ee') = \sigma^2 V$, V simétrica, positiva semidefinida e seja $\lambda'\beta$ uma função estimável. Então existe um vetor b satisfazendo $X'b = \lambda'$ e $b'y$ é um estimador não-viciado para $\lambda'\beta$. A variância de $b'y$ é dada por $\sigma^2 b'Vb$. Para determinarmos o BLUE de $\lambda'\beta$ temos então que encontrar um vetor b tal que $b'Vb$ seja mínimo sujeito à condição que $b'X = \lambda'$.

Serão apresentadas duas soluções para o problema, uma delas utilizando multiplicadores de Lagrange e a outra utilizando argumentos puramente algébricos.

3.1. SOLUÇÃO UTILIZANDO MULTIPLICADORES DE LAGRANGE

Teorema 3.1 - Seja $\lambda'\beta$ uma função estimável. Então o BLUE de $\lambda'\beta$ é $b'y$ onde o vetor b satisfaz as seguintes equações:

$$\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix}$$

X - matriz de planejamento
V - matriz de covariância dos erros.

<u>Demonstração</u> - O vetor b satisfaz $b'X = \lambda'$. Temos então que minimizar Var $(b'y) = b'Vb$ sujeito à condição $b'X = \lambda'$.

Seja o Lagrangeano:

$\alpha = b'Vb + 2(b'X-\lambda')m$, m vetor de multiplicadores.

Derivando-se α em relação a b_i, $i = 1, 2, \ldots, n$, e igualando-se ao zero temos:

$$\frac{\partial\alpha}{\partial b_i} = 2 \sum_{j=1}^{n} V_{ij}\, b_j + 2 \sum_{k=1}^{p} X_{ik} m_k = 0 \qquad i = 1,2,\ldots,n.$$

Em termos matriciais essas equações tornam-se

$$\frac{\partial\alpha}{\partial b} = 2\,Vb + 2Xm = \emptyset\ .$$

Juntando-se essas equações com as restrições temos que um ponto crítico da função α deve satisfazer

$$\begin{aligned} V\,b + Xm &= \emptyset \\ X'\,b &= \lambda \end{aligned} \qquad \text{ou}$$

$$\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix} \qquad (3.1.1).$$

a) O sistema (3.1.1) é consistente.

<u>Prova</u>: $(a_1', a_2') \begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix} = \emptyset \Rightarrow a_1'V + a_2'X' = \emptyset$

$$a_1'X = \emptyset$$

Então $a_1'Va_1 = -a_2'X'a_1 = \emptyset$.

Portanto $a_1'V = \emptyset$ e $a_2'X' = \emptyset$.

Então $(a_1', a_2') \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix} = a_2'\lambda = a_2'X'b = 0$ e o sistema (3.1.1) é consistente.

b) Se $\begin{bmatrix} b_1 \\ m_1 \end{bmatrix}$ e $\begin{bmatrix} b_2 \\ m_2 \end{bmatrix}$ são soluções de (3.1.1) então $b_1'Vb_1 = b_2'Vb_2$.

<u>Prova</u>: Sejam $\begin{bmatrix} b_1 \\ m_1 \end{bmatrix}$ e $\begin{bmatrix} b_2 \\ m_2 \end{bmatrix}$ soluções de (3.1.1).

Então:

$$Vb_1 + Xm_1 = \emptyset = Vb_2 + Xm_2$$

$$X'b_1 = \lambda = X'b_2.$$

Temos

1) $V(b_1-b_2) = X(m_2-m_1)$

2) $X'(b_1-b_2) = \emptyset$

Multiplicando 1 por $(b_1-b_2)'$ temos

$$(b_1-b_2)'V(b_1-b_2) = (b_1-b_2)'X(m_2-m_1) = 0 .$$

$$\Rightarrow V(b_1-b_2) = \emptyset \Rightarrow Vb_1 = Vb_2 .$$

Portanto,

$$b_1'Vb_1 = b_1'Vb_2 = b_2'Vb_2.$$

Além disso

$$Vb_1 + Xm_1 = Vb_2 + Xm_2 \Rightarrow Xm_1 = Xm_2.$$

c) Qualquer solução de (3.1.1) é mínimo global.

<u>Prova</u>: Seja b_0 tal que

$$Vb_0 + X_m = \emptyset$$

$$X'b_0 = \lambda.$$

Consideremos o vetor $(b_0+\delta)'$ satisfazendo

$$(b_0+\delta)'X = \lambda'.$$

Então

$$b_0'X + \delta'X = \lambda' = b_0'X \Rightarrow \delta'X = \emptyset$$

$$\mathrm{Var}\,\{(b_0 + \delta)'y\} = \sigma^2\,(b_0 + \delta)'V(b_0 + \delta) =$$

$$= \sigma^2\,(b_0'Vb_0+\delta'Vb_0+b_0'V\delta+\delta'V\delta) =$$

$$= \sigma^2 \ (b_0'Vb_0 + 2\delta'Vb_0 + \delta'V\delta) =$$

$$= \sigma^2 \ (b_0'Vb_0 - 2\delta'Xm_0 + \delta'V\delta) =$$

$$= \sigma^2 \ (b_0'Vb_0 + \delta'V\delta) \geq \sigma^2 b_0'Vb_0.$$

Portanto

Var $\{(b_0 + \delta)'y\} \geq$ Var $(b_0'y)$ com igualdade se e somente se $V\delta = \emptyset$.

Assim completamos a demonstração do Teorema 3.1.

Se tomamos $\begin{bmatrix} c_1 & c_2 \\ c_3 & -c_4 \end{bmatrix}$ como uma inversa generalizada de $\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix}$ o resultado estabelecido por Rao (1971), sobre o qual nos referimos no início desse capítulo, pode ser obtido a partir do teorema 3.1 da seguinte forma:

O BLUE de $\lambda'\beta$ é $b'y$ onde b' satisfaz

$$\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix}.$$

Seja $\begin{bmatrix} c_1 & c_2 \\ c_3 & -c_4 \end{bmatrix}$ uma inversa generalizada de $\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix}$.

Então $\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} c_1 & c_2 \\ c_3 & -c_4 \end{bmatrix} \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix}$ é uma solução de (3.1.1). Portanto temos que

$b = c_2\lambda$ e o BLUE de $\lambda'\beta$ é dado por $\lambda'c_2'y$.

<u>Corolário 3.1.2</u> - Uma função linear das observações $b'y$ é BLUE de $E(b'y)$ se e somente se $Vb \in C(X)$.

<u>Demonstração</u>:

(=>) $b'y$ é BLUE de $b'X\beta = \lambda'\beta$. Então, pelo teorema 3.1, o vetor b satisfaz

$$Vb = -Xm.$$

Portanto $Vb \in C(X)$.

(<=)

$b'y$ é um estimador não-viciado para $b'X\beta = \lambda'\beta$. Suponhamos que $Vb \in C(X)$.

Então $Vb = -Xm$ para algum m.

Portanto b satisfaz

$$\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix}$$

e $b'y$ é BLUE de $\lambda'\beta$.

O resultado do corolário (3.1.2) foi estabelecido por Zyskind (1967) mediante argumentos puramente algébricos.

<u>Corolário 3.1.3</u> - O estimador simples de mínimos quadrados de $X\beta$ é BLUE se e somente se existe uma matriz Q tal que

$$VX = XQ.$$

<u>Demonstração</u>:

Se $\begin{bmatrix} c_1 & c_2 \\ c_3 & -c_4 \end{bmatrix}$ é uma inversa generalizada de $\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix}$

então $\begin{bmatrix} c_1' & c_3' \\ c_2' & -c_4' \end{bmatrix}$ é também uma inversa generalizada de $\begin{bmatrix} V & X \\ X' & \emptyset \end{bmatrix}$.

Então

$$\begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} c_1' & c_3' \\ c_2' & -c_4 \end{bmatrix} \begin{bmatrix} \emptyset \\ \lambda \end{bmatrix}$$ é também solução de

(3.1.1) e temos

$b = c_3'\lambda$ e o BLUE de $\lambda'\beta$ é dado por $b'y = \lambda' c_3 y$.

Corolário 3.1.1 - O BLUE de $X\beta$ é $A'y$ onde A é uma matriz satisfazendo

$$\begin{aligned} VA + XM &= \emptyset \\ X'A &= X'. \end{aligned}$$

Demonstração: Substituindo sucessivamente o vetor λ em (3.1.1) por $X'i$, $i = 1, 2, \ldots, n$, onde $X'i$ é a i-ésima coluna da matriz X' obtemos o sistema

$$\begin{aligned} VA + XM &= \emptyset \\ X'A &= X' \end{aligned} \qquad (3.1.2).$$

Então o BLUE de $X\beta$ é dado por $A'Y$, A satisfazendo (3.1.2).

(=>) O estimador simples de mínimos quadrados de $X\beta$ é XBy onde B é uma matriz que satisfaz $X'XB = X'$.

Se XBy é BLUE de $X\beta$ então

$VXB + XM = \emptyset$ para algum M

$VXB = -XM$

$VXBX = -XMX$

$VX = XQ$ com $Q = -MX$.

(<=) Se $VX = XQ$ então

$VXB = XQB$

$VXB + XM = \emptyset$ com $M = -QB$.

Portanto XBy é BLUE de $X\beta$ e temos que o ESMQ de $X\beta$ é BLUE.

O corolário (3.1.3) é exatamente a "condição dos autovetores" estabelecida por Zyskind, a qual nós tratamos no capítulo 2.

Corolário 3.1.4 - Uma função $b'y$ é ESMQ e BLUE de sua esperança se e somente se ambos os vetores b e Vb pertencem a $C(X)$.

Demonstração:

(=>) Se $b'y$ é ESMQ de $b'X\beta$ então $b \in C(X)$ (Lema 1.4.1 e 1.4.2). Pelo corolário 3.1.2, se $b'y$ é BLUE de $b'X\beta$ então $Vb \in C(X)$.

(<=) Seja $b'y$ um estimador não viciado de $b'X\beta$.

a) Suponhamos que $b \in C(X)$. Então $b = XBb$ e $b'y = (XBb)'y = b'XBy = b'X\hat{\beta}$, $\hat{\beta}$ solução de $X'X\beta = X'y$. Portanto $b'y$ é ESMQ de $b'X\beta$.

b) Se $Vb \in C(X)$, pelo corolário 3.1.2 temos que $b'y$ é BLUE de $b'X\beta$, portanto o ESMQ de $b'X\beta$ é BLUE.

O corolário 3.1.4 é uma condição equivalente à "condição dos autovetores" e foi também estabelecido por Zyskind (1967).

3.2 - SOLUÇÃO ALGEBRICA

A demonstração do teorema que explicita a forma do BLUE de $X\beta$ necessita do lema seguinte:

Lema 3.2.1 - Seja E um espaço vetorial n-dimensional, f um vetor fixo em E e G uma matriz simétrica positiva semidefinida. Seja g uma função definida em E da seguinte forma

$$g(x) = 2x'f + x'Gx \qquad x \in E \qquad (3.2.1).$$

Uma condição necessária e suficiente para que $g(x)$ tenha um mínimo global em E é que o sistema de equações $Gx + f = \emptyset$ seja consistente. O mínimo de $g(x)$ é $g(x_0)$, onde x_0 é qualquer solução de $Gx + f = \emptyset$.

Demonstração:

(=>) Se G é não singular então $Gx + f = \emptyset$ é consistente.

Suponhamos que G é não singular. Então existe um vetor b, $b \neq \emptyset$, tal que $b'G = \emptyset$ e portanto $b'Gb = 0$.

$g(b) = 2b'f + b'Gb = 2b'f.$

Se $b'f > 0$ tomamos $x = -kb$, $k > 0$ e arbitrário e nós temos que

$g(-kb) = 2(-kb'f) + (-kb')G(-kb) = -2kb'f$ é arbitrariamente pequeno.

Se $b'f < 0$ tomamos $x = kb'$ com $k > 0$ e arbitrário e temos o mesmo resultado.

Portanto $g(x)$ é limitada inferiormente se e somente

$$b'G = \emptyset \quad \Rightarrow b'f = 0$$

que é precisamente a condição de consistência do sistema $Gx+f = \emptyset$.

(<=) Suponhamos que o sistema $Gx + f = \emptyset$ é consistente, ou seja, existe x_0 tal que $Gx_0 + f = \emptyset$.

Então

$$g(x_0+\delta) = 2(x_0+\delta)'f + (x_0+\delta)'G(x_0+\delta) =$$

$$= 2x_0'f + 2\delta'f + x_0'Gx_0 + 2\delta'Gx_0 + \delta'G\delta =$$

$$= 2x_0'f + x_0'Gx_0 + 2\delta'(f+Gx_0) + \delta G\delta =$$

$$= 2x_0'f + x_0'Gx_0 + \delta'G\delta \geq 2x_0'f + x_0'Gx_0.$$

Portanto

$g(x_0 + \delta) \geq g(x_0)$ qualquer δ

e $g(x_0) = 2x_0'f + x_0'Gx_0 = -2x_0'Gx_0 + x_0'Gx_0 =$

$$= -x_0'Gx_0.$$

Sejam x_0 e x_1 duas soluções de $Gx + f = \emptyset$.

Então

$$Gx_0 = Gx_1 \quad e \quad x_0'G = x_1'G$$

$$x_0'Gx_0 = x_0'Gx_1 = x_1'Gx_1.$$

Portanto $g(x)$ é invariante qualquer x solução de $Gx + f = \emptyset$.

Estamos agora em condições de demonstrar o seguinte Teorema.

Teorema 3.2 - Seja o modelo linear $y = X\beta+e$, $E(e) = \emptyset$ $E(ee') = \sigma^2 V$. O BLUE de $X\beta$ é dado por $A'y$ sendo

$$A = XB + (I-XB)N \ ,$$

N qualquer solução do sistema.

$$(I-XB)V(I-XB)N + (I-XB)VXB = \emptyset \qquad (3.2.2)$$

e B tal que $XBX = X$.

Demonstração:

Seja $\lambda'\beta$ uma função estimável. Portanto existe b tal que $X'b = \lambda$ e $b'y$ é um estimador não viciado para $\lambda'\beta$. Qualquer b solução de $X'b = \lambda$ pode ser escrito como

$$b = B'\lambda + (I-B'X')\gamma =$$

$$= B'\lambda + (I-XB)\gamma \quad , \quad \gamma \text{ arbitrário } (3.2.3).$$

$$Var\ (b'y) = b'Vb = \{\lambda'B+\gamma'(I-XB)\}V\{B'\lambda+(I-XB)\gamma\} =$$

$$= \lambda'BVB'\lambda + 2\gamma'(I-XB)VB'\lambda + \gamma'(I-XB)V(I-XB)\gamma.$$

XBb_0 é único para qualquer b_0 solução de $\lambda' = b'X$.

$\lambda'BVB'\lambda = (XBb_0)'VXBb_0$. Portanto $\lambda'BVB'\lambda$ é invariante.

Para encontrarmos b, da forma (3.2.3), que minimiza Var(by) temos que determinar o vetor γ que minimiza

$$2\gamma'(I-XB)VB'\lambda + \gamma'(I-XB)V(I-XB)\gamma \qquad (3.2.4.)$$

Fazendo

$$x = \gamma$$

$$f = (I-XB)VB'\lambda$$

$$G = (I-XB)V(I-XB)$$

a expressão a ser minimizada assume a forma (3.2.1) que é a expressão do Lema (3.2.1). Então temos que verificar se

$$(I-XB)V(I-XB)\gamma + (I-XB)VB'\lambda = \emptyset \text{ é consistente.}$$

Podemos reescrever a expressão acima na forma

$$\underbrace{(I-XB)Z}_{W'}\ \underbrace{Z'(I-XB)}_{W}\underbrace{\gamma}_{\gamma} = \underbrace{(I-XB)Z}_{W'}\ \underbrace{(-Z'B'\lambda)}_{z} \qquad (3.2.5)$$

que é a forma das equações normais. Portanto a expressão (3.2.4) tem um mínimo global e esse mínimo é alcançado para qualquer γ solução de (3.2.5).

Seja γ_0 uma solução.

$W\gamma_0 = Z'(I-XB)\gamma_0$ é único assim como

$$ZZ'(I-XB)\gamma_0 = V(I-XB)\gamma_0.$$

Então o BLUE de $\lambda'\beta$ é $b'y$, $b = B'\lambda + (I-XB)\gamma_0$ e γ_0 qualquer solução de

$$(I-XB)V(I-XB)\gamma = (I-XB)VB'\lambda.$$

Se fazemos λ sucessivamente como as linhas da matriz X temos que o BLUE de $X\beta$ é $A'y$ sendo

$$A = XB + (I-XB)N$$

onde N é qualquer solução de

$$(I-XB)V(I-XB)N + (I-XB)VXB = \emptyset.$$

Alguns corolários podem ser deduzidos imediatamente do teorema 3.2, relacionando as idéias desenvolvidas no capítulo 2. O corolário 3.2.3 será de particular interesse pois fornece um método prático para se verificar a validade da "condição dos autovetores" para alguns planejamentos de experimentos aleatorizados que veremos no próximo capítulo.

Corolário 3.2.1 - Os estimadores simples de mínimos quadrados coincidem com os BLUES se e somente se

$$(I-XB)VXB = \emptyset.$$

Demonstração:

(=>) O estimador simples de mínimos quadrados de $X\beta$ é XBy que é também BLUE. Então, pelo teorema 3.2, $A = XB$ e podemos tomar $N = \emptyset$. Portanto

$$(I-XB)VXB = \emptyset.$$

(<=) Se $(I-XB)VXB = \emptyset$ então $N = \emptyset$ é solução de (3.2.2) e portanto $A = XB$ é BLUE e ESMQ de $X\beta$.

Uma observação importante merece registro agora.

Quando consideramos que $(I-XB)VXB = \emptyset$

então

$$(I-XB)V(I-XB)N = \emptyset$$

$$N'(I-XB)V(I-XB)N = \emptyset \quad \text{e}$$

$$V(I-XB)N = \emptyset$$

Portanto se consideramos os estimadores

XBy e $(XB + N'(I-XB))y$ temos que

$$E\{(XB + N'(I-XB))y\} = E(XBy) + E(N'(I-XB)y) =$$

$$= XBX\beta + N'(I-XB)X\beta = X\beta = E(XBy)$$

$$Var\{(XB + N'(I-XB))y\} = (XB + N'(I-XB)V(XB + N'(I-XB))' =$$

$$= XBVXB + XBV(I-XB)N + N'(I-XB)VXB + N'(I-XB)V(I-XB)N =$$

$$= XBVXB = Var(XBy).$$

Então não existe uma única matriz A sob a condição que $(I-XB)VXB = \emptyset$ mas se A_1 e A_2 são soluções $A_1y = A_2y$ com probabilidade 1. Portanto, quando escrevemos que os ESMQ coincidem com os BLUES queremos dizer que ESMQ e BLUE são equivalentes com probabilidade 1.

<u>Corolário 3.2.2</u> - $(I-XB)VXB = \emptyset$ se e somente se $VX = XQ$ para alguma matriz Q.

<u>Demonstração</u>:

$(\Rightarrow)$ $(I-XB)VXB = \emptyset \Rightarrow VXB = XBVXB$

$\Rightarrow VXBX = XBVXBX \Rightarrow VX = XBVX.$

Portanto

$$VX = XQ \quad \text{com } Q = BVX$$

$(\Leftarrow)$ $VX = XQ \Rightarrow VXB = XQB$

$$(I-XB)VXB = (I-XB)XQB = \emptyset.$$

<u>Corolário 3.2.3</u> - $(I-XB)VXB = \emptyset$ se e somente se $V = XT_1X' + (I-XB)T_2(I-XB) + kI$ para alguma constante k e matrizes T_1 e T_2.

<u>Demonstração</u>:

$(\Rightarrow)$ Podemos escrever V como

$$V = \{XB + (I-XB)\}V\{XB + (I-XB)\} \qquad (3.2.6)$$

$$= XBVXB + XBV(I-XB) + (I-XB)VXB + (I-XB)V(I-XB).$$

Assim

$$V = XBVXB + (I-XB)V(I-XB).$$

Da mesma forma que em (3.2.6) podemos escrever

$$(V-kI) = \{XB + (I-XB)\}\ (V-kI)\ \{XB + (I-XB)\}.$$

Assim

$$V-kI = XB(V + kI)XB + (I-XB)(V + kI)(I-XB).$$

Portanto

$$V = kI + XB(V-kI)XB + (I-XB)(V + kI)(I-XB) =$$

$$= kI + XB(V-kI)B'X' + (I-XB)(V-kI)(I-XB) =$$

$$= kI + XT_1X' + (I-XB)T_2(I-XB)$$

(<=)

$$(I-XB)VXB = (I-XB)\{kI + XT_1X' + (I-XB)T_2(I-XB)\}XB$$

$$= (I-XB)kIXB+(I-XB)XT_1X'XB+(I-XB)T_2(I-XB)XB$$

$$= \emptyset$$

Se para um modelo particular conhecemos a matriz V e existem matrizes T_1 e T_2 e uma constante k de tal forma que V possa ser escrito como $XT_1X' + (I-XB)T_2(I-XB) + kI$ então para determinarmos o BLUE de qualquer função estimável $\lambda'\beta$ tudo o que precisamos é de uma solução do sistema $X'X\beta = X'y$.

CAPITULO 4

ALGUNS EXPERIMENTOS FINITOS ALEATORIZADOS

Alguns modelos lineares interessantes e de muita utilização, onde é válida a "condição dos autovetores", são modelos derivados de experimentos finitos aleatorizados.

Devido ao próprio processo de aleatorização, esses modelos geralmente apresentam erros correlacionados e, sob algumas condições, a matriz de covariância dos erros é singular.

Examinaremos com detalhes três desses modelos: amostra aleatória simples sem reposição, experimentos completamente aleatorizados e blocos aleatorizados.

Pode-se notar, na apresentação dos modelos a seguir, que poucas hipóteses são feitas e toda hipótese é uma decorrência direta do modo como o experimento é levado a cabo. Contraste-se tal dedução com os postulados usados nos livros textos de Modelos Lineares ou de Planejamento de Experimentos em que se começa por

$$y = X\beta + e, \quad E(e) = 0, \quad E(ee') = \sigma^2 I$$

e o vetor de erros tem distribuição normal multivariada. A nosso ver, o modelo acima deficilmente se realizará num caso prático.

É boa fortuna no entanto que, conforme provaremos em alguns casos a seguir, procedimentos obtidos com o modelo suposto sejam idênticos aos obtidos pelo modelo deduzido finito.

Mesmo assim, somos de opinião que as apresentações de

Planejamento de Experimentos baseadas nos modelos supostos são nocivas, pelo que implicam de divórcio com a realidade.

4.1 - AMOSTRA ALEATORIA SIMPLES

Quando desejamos estimar o parâmetro μ, média populacional, e o esquema amostral utilizado é a amostragem aleatória simples, sabemos que a média amostral é o melhor estimador linear não viciado para μ, se a amostra é retirada com reposição ou a população tem um número infinito de elementos. Sabemos também que, mesmo quando o número de elementos da população é finito e retiramos a amostra sem reposição, a média amostral ainda é BLUE para μ.

Colocaremos o problema de se estimar μ, através de uma amostra aleatória simples sem reposição, sob a forma do modelo linear geral e verificaremos que, utilizando a teoria exposta nos capítulos anteriores, a "condição dos autovetores" é válida para esse modelo.

Consideremos uma população com N indivíduos e uma variável Y que assume o valor Y_j no j-ésimo indivíduo da população, $j = 1, 2, \ldots, N$. Denotaremos por μ a média populacional

$$\left(\mu = \frac{1}{N} \sum_{j=1}^{N} Y_j \right).$$

Seja a seguinte igualdade algébrica

$$Y_j = \mu + (Y_j - \mu) \qquad j = 1, 2, \ldots, N .$$

Retiramos da população uma amostra aleatória simples de tamanho n, sem reposição.

Seja y_i, $i = 1, 2, \ldots, n$ o valor da variável Y no i-ésimo elemento da amostra. Cada amostra possível $(y_1, y_2, \ldots, y_n)$ tem probabilidade $\frac{1}{\binom{N}{n}}$ de ocorrer.

Podemos relacionar cada resultado y_i da amostra com os valores Y_j da população utilizando variáveis aleatórias amostrais δ_j^i definidas da seguinte forma:

$\delta_j^i = 1$ se o i-ésimo elemento da amostra é o j-ésimo elemento da população

$\delta_j^i = 0$ caso contrário .

A utilização dessas variáveis aleatórias em planejamento de experimentos é devida a Kempthorne (1952).

Temos n.N dessas variáveis, todas com a mesma distribuição

$$P(\delta_j^i = 1) = \frac{1}{N}$$

$$P(\delta_j^i = 0) = \frac{N-1}{N} \quad .$$

Algumas propriedades dessas variáveis são facilmente estabelecidas:

1) $P(\delta_j^i \, \delta_{j'}^{i'} = 1) = P(\delta_{j'}^{i'} = 1 \, / \, \delta_j^i = 1) \; . \; P(\delta_j^i = 1) =$

$$= \frac{1}{N-1} \cdot \frac{1}{N} \qquad \begin{matrix} i \neq i' \\ j \neq j' \end{matrix}$$

2) $P(\delta_j^i \ \delta_{j'}^i = 1) = 0 \qquad j \neq j'$

3) $P(\delta_j^i \ \delta_j^{i'} = 1) = 0 \qquad i \neq i'$

4) $E(\delta_j^i) = E(\delta_j^i)^2 = P(\delta_j^i = 1) = \dfrac{1}{N}$

5) $E(\delta_j^i \ \delta_{j'}^i) = 0 \qquad j \neq j'$

6) $E(\delta_j^i \ \delta_j^{i'}) = 0 \qquad i \neq i'$

7) $E(\delta_j^i \ \delta_{j'}^{i'}) = \dfrac{1}{N(N-1)} \qquad \begin{matrix} i \neq i' \\ j \neq j' \end{matrix}$

8) $\sum_{j=1}^{N} \delta_j^i = 1$.

Cada elemento y_i da amostra pode então ser escrito como:

$y_i = \sum_{j=1}^{N} \delta_j^i \ Y_j$, então

$$y_i = \sum_{j=1}^{N} \delta_j^i(\mu+(Y_j-\mu)) = \mu + \sum_{j=1}^{N} \delta_j^i(Y_j-\mu).$$

Fazendo

$$e_i = \sum_{j=1}^{N} \delta_j^i(Y_j-\mu) \qquad \text{temos:}$$

$$y_i = \mu+e_i, \quad i = 1, 2, \ldots, n$$

que é o modelo linear deduzido.

Escrevamos o modelo sob a forma $y = X\beta+e$

sendo
$$y = \begin{bmatrix} y_1 \\ y_2 \\ \cdot \\ \cdot \\ \cdot \\ y_n \end{bmatrix}, \quad X = \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix}, \quad \beta = \mu, \quad e = \begin{bmatrix} e_1 \\ e_2 \\ \cdot \\ \cdot \\ \cdot \\ e_n \end{bmatrix}$$

$$E(e_i) = E\left(\sum_{j=1}^{N} \delta_j^i(Y_j-\mu)\right) = \sum_{j=1}^{N} (Y_j-\mu)E(\delta_j^i) =$$

$$= \frac{1}{N}\sum_{j=1}^{N} (Y_j-\mu) = 0 \quad , \quad i = 1, 2, \ldots, n$$

$$\operatorname{Var}(e_i) = E\left(\sum_{j=1}^{N} \delta_j^i(Y_j-\mu)\right)^2 =$$

$$= \sum_{j=1}^{N} (Y_j-\mu)^2 E(\delta_j^i)^2 + \sum_{\substack{j=1 \\ j\neq k}}^{N} \sum_{k=1}^{N} (Y_j-\mu)(Y_k-\mu) E(\delta_j^i\ \delta_k^i) =$$

$$= \frac{1}{N} \sum_{j=1}^{N} (Y_j-\mu)^2 = \sigma^2$$

$$\mathrm{Cov}(e_i,\ e_j) = E(e_i e_j) = E\left(\sum_{k=1}^{N} \delta_k^i (Y_k-\mu) \sum_{p=1}^{N} \delta_p^j (Y_p-\mu) \right) =$$

$$= \sum_{k=1}^{N} \sum_{p=1}^{N} (Y_k-\mu)(Y_p-\mu) E(\delta_k^i\ \delta_p^j) =$$

$$= \sum_{k=1}^{N} (Y_k-\mu)^2 E(\delta_k^i\ \delta_k^j) + \sum_{\substack{k=1 \\ k\neq p}}^{N} \sum_{p=1}^{N} (Y_k-\mu)(Y_p-\mu) E(\delta_k^i\ \delta_p^j) =$$

$$= 0 + \frac{1}{N(N-1)} \sum_{\substack{k=1 \\ k\neq p}}^{N} \sum_{p=1}^{N} (Y_k-\mu)(Y_p-\mu) =$$

$$= \frac{1}{N(N-1)} \left(\sum_{k=1}^{N} (Y_k-\mu) \right)^2 - \frac{1}{N(N-1)} \sum_{k=1}^{N} (Y_k-\mu)^2 =$$

$$= - \frac{1}{N(N-1)} \sum_{k=1}^{N} (Y_k-\mu)^2 = - \frac{1}{N-1} \sigma^2 .$$

A matriz de covariância dos erros é formada então por

$$V = \mathrm{Cov}(e) = E(ee') = \sigma^2 \begin{bmatrix} 1\rho\rho\ldots\rho \\ \rho 1\rho\ldots\rho \\ \cdot \\ \cdot \\ \cdot \\ \rho\rho\rho\ldots 1 \end{bmatrix}, \quad \rho = -\frac{1}{N-1} \cdot$$

Podemos reescrever a matriz $\frac{1}{\sigma^2} V$ da seguinte forma:

$$\frac{1}{\sigma^2} V = \begin{bmatrix} 1 & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ -\frac{1}{N-1} & 1 & \cdots & -\frac{1}{N-1} \\ \vdots & & & \\ -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & 1 \end{bmatrix} = \begin{bmatrix} \frac{N-1}{N-1} & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ -\frac{1}{N-1} & \frac{N-1}{N-1} & \cdots & -\frac{1}{N-1} \\ \vdots & & & \\ -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & \frac{N-1}{N-1} \end{bmatrix} =$$

$$= \begin{bmatrix} \frac{N}{N-1} & 0 & \cdots & 0 \\ 0 & \frac{N}{N-1} & \cdots & 0 \\ \vdots & & & \\ 0 & 0 & \cdots & \frac{N}{N-1} \end{bmatrix} + \begin{bmatrix} -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \\ \vdots & & & \\ -\frac{1}{N-1} & -\frac{1}{N-1} & \cdots & -\frac{1}{N-1} \end{bmatrix} =$$

$$= \frac{N}{N-1} I + (- \frac{1}{N-1}) J ,$$ J matriz quadrada com "1" em todas as posições.

Portanto

$$V = \sigma^2((1-\rho)I + \rho J).$$

Pelo corolário 3.2.3 do capítulo 3 temos que o estimador simples de mínimos quadrados (ESMQ) de uma função estimável é BLUE se e somente se

$$V = XT_1X' + (I-XB)T_2(I-XB) + kI$$ para alguma constante k e matrizes T_1 e T_2.

Fazendo $T_1 = \rho\sigma^2$, $T_2 = \emptyset$, $k = (1-\rho)\sigma^2$ temos:

$$XT_1X' = \rho\sigma^2 \begin{bmatrix} 1 \\ 1 \\ \cdot \\ \cdot \\ \cdot \\ 1 \end{bmatrix} (1\ 1 \ldots 1) = \rho\sigma^2 J$$

$$kI = (1-\rho)\sigma^2 I.$$

Então V satisfaz as condições do corolário 3.2.3.

O estimador simples de mínimos quadrados de μ é:

$$\hat{\mu} = (X'X)^{-1}X'y = \left[(1\ 1 \ldots 1) \begin{bmatrix} 1 \\ 1 \\ \vdots \\ 1 \end{bmatrix} \right]^{-1} (1\ 1 \ldots 1) \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} =$$

$$= \frac{1}{n} \sum_{i=1}^{n} y_i = \bar{y} \quad .$$

Portanto $\bar{y}$ é o BLUE de μ.

De uma forma geral, sempre que a matriz V for da forma $aI + bJ$ e a matriz X puder ser colocada na forma

$$\left[\begin{array}{c|c} 1 & \\ 1 & \\ \vdots & \psi \\ 1 & \end{array} \right], \quad \text{podemos tomar}$$

$$T_1 = b\begin{pmatrix} 1 & \emptyset \\ \emptyset & \emptyset \end{pmatrix}, \qquad T_2 = \emptyset \qquad \text{e} \qquad k = a$$

e temos:

$$XT_1X' = bJ \quad , \qquad kI = aI.$$

Então V pode ser escrita sob a forma

$$V = XT_1X' + (I-XB)T_2(I-XB) + kI$$

e sob essas condições ESMQ = BLUE.

4.2 - PLANEJAMENTO COMPLETAMENTE ALEATORIZADO

Suponhamos que temos n=rt unidades experimentais e t tratamentos que serão aplicados às unidades com a restrição que cada tratamento seja aplicado em r unidades. Denotaremos a j-ésima replicação do i-ésimo tratamento pelo par (i, j), i=1,2,...t; j=1,2,...,r.

Cada par (i, j) será aplicado aleatoriamente às unidades experimentais. Isso pode ser feito, por exemplo, sorteando-se sequencialmente ao acaso os pares (i, j) e aplicando-os sucessivamente às unidades experimentais, as quais deverão estar previamente ordenadas. A alocação aleatória dos pares (i, j) às unidades experimentais permite controlar fatores desconhecidos do pesquisador que podem estar influenciando nos resultados do experimento.

A cada possível combinação tratamento-unidade experimental temos uma "resposta conceitual" Y_{ik}, k = 1, 2, ... , n ; i = 1, 2, ... , t , onde Y_{ik} é a resposta da k-ésima unidade experimental ao tratamento i.

Estamos denominando Y_{ik} de resposta conceitual porque não temos disponíveis esses valores mesmo por que, quando aplicamos o tratamento i à unidade k, de alguma forma alteramos a unida de experimental em sua capacidade de resposta a outro tratamento distinto j, quando não destruimos a unidade experimental.

Denotaremos por y_{ij} a observação resultante da j-ésima replicação do i-ésimo tratamento.

Usaremos as variáveis aleatórias amostrais δ_{ij}^{k} para associar os resultados experimentais com as respostas conceituais Y_{ik}.

Definiremos as variáveis δ_{ij}^{k} da seguinte forma:

$\delta_{ij}^{k} = 1$ se a j-ésima replicação do i-ésimo tratamento está aplicado na k-ésima unidade experimental.

$\delta_{ij}^{k} = 0$ caso contrário.

Temos $n.t.r = n^2$ dessas variáveis e

$$P(\delta_{ij}^{k}) = \frac{1}{n}.$$

Como na seção anterior, obtemos facilmente algumas propriedades dessas variáveis.

1. $E(\delta_{ij}^{k}) = E(\delta_{ij}^{k})^2 = P(\delta_{ij}^{k} = 1) = \dfrac{1}{n}$

2. $E(\delta_{ij}^{k}\ \delta_{i'j'}^{k'}) = P(\delta_{ij}^{k}\ \delta_{i'j'}^{k'} = 1) =$

$$= P(\delta^{k'}_{i'j'}=1/\delta^{k}_{ij}=1)P(\delta^{k}_{ij}=1) = \frac{1}{n(n-1)} \qquad (i,j) \neq (i',j') \quad k \neq k'$$

3. $$E(\delta^{k}_{ij}\ \delta^{k}_{i'j'}) = P(\delta^{k}_{ij}\ \delta^{k}_{i'j'} = 1) = 0 \qquad (i,j) \neq (i',j')$$

4. $$E(\delta^{k}_{ij}\ \delta^{k'}_{ij}) = P(\delta^{k}_{ij}\ \delta^{k'}_{ij} = 1) = 0 \qquad k \neq k'$$

5. $$\sum_{k=1}^{n} \delta^{k}_{ij} = 1 = \sum_{i=1}^{t} \sum_{j=1}^{r} \delta^{k}_{ij}\ .$$

As observações y_{ij} são associadas aos valores Y_{ik} da seguinte forma:

$$y_{ij} = \sum_{k=1}^{n} \delta^{k}_{ij}\ Y_{ik} \qquad (4.2.1)$$

Cada resposta y_{ij} é uma função linear das variáveis aleatórias amostrais.

Escreveremos Y_{ik} da seguinte forma:

$$Y_{ik} = Y_{..} + (Y_{i.}-Y_{..}) + (Y_{.k}-Y_{..}) + (Y_{ik}-Y_{i.}-Y_{.k}+Y_{..}) \qquad (4.2.2)$$

sendo

$$Y_{..} = \frac{1}{nt} \sum_{i=1}^{t} \sum_{k=1}^{n} Y_{ik}$$

$$Y_{i.} = \frac{1}{n} \sum_{k=1}^{n} Y_{ik}$$

$$Y_{.k} = \frac{1}{t} \sum_{i=1}^{t} Y_{ik} .$$

Quando dizemos que a resposta da unidade k ao tratamento i, Y_{ik}, é um valor fixo, estamos supondo que temos controle absoluto sobre quaisquer outros fatores externos que podem influir na resposta da unidade k. Então a resposta conceitual Y_{ik} será devida a características da unidade k, a propriedades do tratamento i e a uma possível interação entre a unidade k a o tratamento i. Suporemos aqui que os efeitos da unidade e do tratamento são aditivos e que não existe interação entre a unidade e o tratamento, ou seja, que Y_{ik} pode ser escrito como soma de duas parcelas x_i e w_k, sendo x_i o efeito do tratamento i e w_k o efeito da unidade k. Então temos:

$$Y_{ik} = x_i + w_k .$$

Como consequência dessa suposição temos:

$$Y_{ik} - Y_{i'k} = x_i + w_k - x_i' - w_k = x_i - x_i' \quad , \text{ ou seja,}$$

se pudéssemos aplicar ambos os tratamento i e i' a uma mesma unidade experimental, então a diferença das respostas seria constan te para cada unidade e dependeria apenas dos tratamentos aplicados.

Definiremos que tratamento e unidade são aditivos se e somente Y_{ik} é expressável como soma de duas partes, uma devido ao tratamento e outra devido à unidade.

Valendo a aditividade entre tratamento e unidade temos:

$$Y_{..} = x_{.}+w_{.} = \mu$$

$$Y_{i.} = x_i+w_{.}$$

$$Y_{.k} = x_{.}+w_k$$

e a identidade (4.2.2) torna-se

$$Y_{ik} = \mu+(x_i-x_{.}) + (Z_k-Z_{.})$$

$$= \mu+t_i + \varepsilon_k \qquad (4.2.3)$$

De (4.2.1) e (4.2.3) temos

$$Y_{ij} = \sum_{k=1}^{n} \delta_{ij}^{k} Y_{ik} = \sum_{k=1}^{n} \delta_{ij}^{k} (\mu+t_i+\varepsilon_k) =$$

$$= \mu+t_i+\sum_{k=1}^{n} \delta_{ij}^{k} \varepsilon_k = \mu+t_i+e_{ij}$$

sendo que

$$\sum_{i=1}^{t} t_i = \sum_{i=1}^{t} (x_i - x_{.}) = 0$$

$$\sum_{i=1}^{t} \sum_{j=1}^{r} e_{ij} = \sum_{i=1}^{t} \sum_{j=1}^{r} \sum_{k=1}^{n} \delta_{ij}^{k} (Z_k - Z_{.}) =$$

$$= \sum_{k=1}^{n} (Z_k - Z_{.}) \sum_{i=1}^{t} \sum_{j=1}^{r} \delta_{ij}^{k} = 0 .$$

O modelo

$Y_{ij} = \mu + t_i + e_{ij}$, $i = 1,2,\dots,t$; $j = 1,2,\dots,r$, é o modelo linear deduzido para o experimento completamente ao acaso.

Colocando o modelo na forma matricial $y = X\beta + e$ temos:

$$y = \begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1r} \\ \vdots \\ y_{t1} \\ y_{t2} \\ \vdots \\ y_{tr} \end{bmatrix} , \quad \beta = \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ \vdots \\ t_t \end{bmatrix} , \quad e = \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1r} \\ \vdots \\ e_{t1} \\ e_{t2} \\ \vdots \\ e_{tr} \end{bmatrix}$$

$$X = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & & & & \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & & & & \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & & & & \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix}$$

$$E\,(e_{ij}) = E(\sum_{k=1}^{n} \delta_{ij}^{k} \varepsilon_k) = \sum_{k=1}^{n} \varepsilon_k (\delta_{ij}^{k}) = \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k = 0$$

$$\mathrm{Var}(e_{ij}) = E(\sum_{k=1}^{n} \delta_{ij}^{k} \varepsilon_k)^2 = \sum_{k=1}^{n} \varepsilon_k{}^2 E(\delta_{ij}^{k})^2 + \sum_{\substack{k=1 \\ k \neq p}}^{n} \sum_{p=1}^{n} \varepsilon_k \varepsilon_p E(\delta_{ij}^{k}\ \delta_{ij}^{p}) =$$

$$= \frac{1}{n} \sum_{k=1}^{n} \varepsilon_k{}^2 = \sigma^2$$

$$\mathrm{Cov}(e_{ij},\ e_{i'j'}) = E(e_{ij} e_{i'j'}) = E(\sum_{k=1}^{n} \delta_{ij}^{k} \varepsilon_k \sum_{p=1}^{n} \delta_{ij}^{p} \varepsilon_p) = -\frac{1}{n-1} \sigma^2 .$$

Então

$$V = \text{cov}(e) = E(ee') = \sigma^2 \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \vdots & & & & \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix} , \rho = -\frac{1}{n-1}$$

que é análoga à forma da matriz de covariância dos erros deduzida na seção anterior com a ressalva que agora, como temos $\sum_{i=1}^{t} \sum_{j=1}^{r} e_{ij} = 0$, a matriz V é singular. Temos também que a matriz de planejamento X se escreve como $\left[\begin{array}{c|c} 1 & \\ 1 & \\ \vdots & \psi \\ 1 & \end{array}\right]$.

Portanto V pode ser escrita como

$XT_1X' + (I-XB)T_2(I-XB) + kI$ e obtemos o BLUE de qualquer função estimável, no planejamento completamente ao acaso, como se estivéssemos trabalhando com o modelo

$$Y_{ij} = \mu + t_i + e_{ij}$$

satisfazendo as hipóteses do teorema de Gauss-Markoff, ou seja, $E(e) = \emptyset$ e $E(ee') = \sigma^2 I$.

Os resultados acima podem ser estendidos para o planejamento completamente ao acaso, na situação em que temos N unidades na população e sorteamos $n = rt$ $(n < N)$ unidades experimentais a serem submetidas aos tratamentos. A matriz de covariância dos erros tem a mesma estrutura deduzida acima com $\rho = -\dfrac{1}{N-1}$ e $\sigma^2 = \dfrac{1}{N}\sum_{k=1}^{N}(Z_k - Z_.)^2$. Nesse caso $\sum_{i=1}^{t}\sum_{j=1}^{r} e_{ij} \neq 0$ e a matriz de covariância é não singular.

4.3 BLOCOS COMPLETAMENTE AO ACASO

No planejamento de blocos completamente ao acaso temos ainda $n = rt$ unidades experimentais, as quais são agrupadas em r blocos com t unidades cada um; os t tratamentos são aplicados ao acaso dentro de cada bloco e independentemente entre cada bloco. Esse procedimento produz bons resultados quando é possível agrupar as unidades experimentais em blocos de unidades que apresentam um alto grau de homogeneidade (dentro de cada bloco) em relação a alguns fatores que sabemos que influenciam na resposta da unidade ao tratamento. Os outros possíveis fatores influentes, desconhecidos do pesquisador, são controlados pelo processo de aleatorização dentro de cada bloco.

À j-ésima unidade experimental dentro de cada bloco chamaremos de j-ésima parcela.

Denotaremos por Y_{ijk} à resposta conceitual devido ao i-ésimo tratamento aplicado a j-ésima parcela do k-ésimo bloco, $i = 1,2,\ldots t$; $j = 1,2,\ldots,t$; $k = 1,2,\ldots,r$.

Podemos escrever Y_{ijk} da seguinte forma

$$Y_{ijk} = Y_{...} + (Y_{..k} - Y_{...}) + (Y_{ijk} - Y_{.jk}) + (Y_{.jk} - Y_{..k}) \quad (4.3.1)$$

sendo

$$Y_{...} = \frac{1}{t^2 r} \sum_{i=1}^{t} \sum_{j=1}^{t} \sum_{k=1}^{r} Y_{ijk}$$

$$Y_{..k} = \frac{1}{t^2} \sum_{i=1}^{t} \sum_{j=1}^{t} Y_{ijk}$$

$$Y_{.jk} = \frac{1}{t} \sum_{i=1}^{t} Y_{ijk} \quad .$$

Suponhamos que temos aditividade dos efeitos de tratamento e unidade, ou seja

$$Y_{ijk} = x_i + w_{jk}.$$

Então temos

$$Y_{...} = x_{.} + w_{..} = \mu$$

$$(Y_{..k} - Y_{...}) = (x_{.} + w_{.k} - x_{.} - w_{..}) = w_{.k} - w_{..} = b_k$$

$$(Y_{ijk} - Y_{.jk}) = (x_i + w_{jk} - x_{.} - w_{jk}) = (x_i - x_{.}) = t_i$$

$$(Y_{.jk} - Y_{..k}) = (x_{.} + w_{jk} - x_{.} - w_{.k}) = w_{jk} - w_{.k} = \varepsilon_{jk} \quad .$$

Substituindo as expressões acima em (4.3.1) temos:

$$Y_{ijk} = \mu + b_k + t_i + \varepsilon_{jk} \qquad (4.3.2)$$

Cabe aqui uma observação. Uma razão para decompormos o valor conceitual numa soma de parcelas, como fizemos em (4.2.2) e (4.3.1), é que essa decomposição goza de uma propriedade interessante: a soma das quadrados dos termos à esquerda da identidade é igual à soma dos quadrados dos termos à direita, e são essas somas de quadrados que são utilizadas na análise de variância do modelo. Essa propriedade vale porque os experimentos que tratamos até aqui são experimentos com estrutura balanceada. Além disso, sob a hipótese de aditividade entre unidade e tratamento, o valor conceitual é decomposto em efeitos aditivos de blocos e tratamento, como podemos observar em (4.3.2).

Denotaremos a resposta observada devido ao i-ésimo tratamento no k-ésimo bloco por y_{ik}.

Sejas as variáveis aleatórias amostrais definidas como:

$\delta^{i}_{jk} = 1$ se o i-ésimo tratamento é aplicado na j-ésima parcela do k-ésimo bloco.

$\delta^{i}_{jk} = 0$ caso contrário.

As propriedades das variáveis δ^{i}_{jk} que nos interessam são:

1) $E(\delta^{i}_{jk}) = E(\delta^{i}_{jk})^2 = P(\delta^{i}_{jk}=1) = \dfrac{1}{t}$

2) $E(\delta^{i}_{jk}\ \delta^{i'}_{j'k}) = P(\delta^{i}_{jk}\ \delta^{i'}_{j'k} = 1) = \dfrac{1}{t(t-1)}$ $\quad i \neq i'$, $j \neq j'$

3) $E(\delta^{i}_{jk}\ \delta^{i'}_{j'k'}) = \frac{1}{t^2}$ $\qquad k \neq k'$

As observações y_{ik} podem ser associadas aos valores conceituais Y_{ijk} da seguinte forma:

$$y_{ik} = \sum_{j=1}^{t} \delta^{i}_{jk} Y_{ijk} = \sum_{j=1}^{t} \delta^{k}_{ij} (\mu+b_k+t_i+\varepsilon_{jk}) =$$

$$= \mu + b_k + t_i + \sum_{j=1}^{t} \delta^{i}_{jk}\ \varepsilon_{jk} =$$

$$= \mu + b_k + Z_i + e_{ik}$$

sendo que

$$\sum_{k=1}^{r} b_k = \sum_{i=1}^{t} t_i = \sum_{i} \sum_{k} e_{ik} = 0 .$$

O modelo $y_{ik} = \mu + b_k + t_i + e_{ik}$ é o modelo deduzido para o planejamento de blocos completamente ao acaso. Os parâmetros do modelo são:

$$E(e_{ik}) = E(\sum_{j=1}^{t} \delta^{i}_{jk}\varepsilon_{jk}) = \sum_{j=1}^{t} \varepsilon_{jk} E(\delta^{i}_{jk}) = \frac{1}{t} \sum_{j=1}^{t} \varepsilon_{jk} = 0$$

$$Var(e_{ik}) = E(\sum_{j=1}^{t} \delta^{i}_{jk}\varepsilon_{jk})^2 =$$

$$= \sum_{j=1}^{t} \varepsilon_{jk}^{2} E(\delta_{jk}^{i})^{2} + \sum_{j=1}^{t} \sum_{\substack{p=1 \\ j\neq p}}^{t} \varepsilon_{jk}\varepsilon_{pk} E(\delta_{jk}^{i}\ \delta_{pk}^{i}) =$$

$$= \frac{1}{t} \sum_{j=1}^{t} \varepsilon_{jk}^{2} = \sigma_{k}^{2} \qquad k = 1,2,\ldots,r \,.$$

Vamos supor que $\sigma_k^2 = \sigma^2$, $k = 1,2,\ldots,r$.

Dentro de um mesmo bloco k a situação é análoga à de um planejamento completamente ao acaso. Então, pelos mesmos argumentos da seção 4.2,

$$\text{cov}(e_{ik}, e_{i'k}) = \sigma^2\left(-\frac{1}{t-1}\right) .$$

Como a aplicação dos tratamentos é independente entre blocos temos que

$$\text{cov}(e_{ik}, e_{i'k'}) = E(e_{ik}e_{i'k'}) = E(e_{ik})E(e_{i'k'}) = 0$$

$$k \neq k'$$

A matriz de covariância dos erros tem então a seguinte forma:

$$V = \text{Cov}(e) = E(ee') = \begin{bmatrix} V_1 & \emptyset & \ldots & \emptyset \\ \emptyset & V_2 & \ldots & \emptyset \\ \vdots & & & \\ \emptyset & \emptyset & \ldots & V_r \end{bmatrix}$$

$$X_i X_j' = \begin{cases} J_t & \text{se } i = j \\ \emptyset & \text{se } i \neq j \end{cases} .$$

Seja T uma matriz quadrada de dimensão (r+t+1) definida como:

$$T = \begin{bmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{bmatrix} .$$

Então

$$XTX' = \begin{bmatrix} X_1 & \psi_1 \\ X_2 & \psi_2 \\ \vdots & \vdots \\ X_r & \psi \end{bmatrix} \begin{bmatrix} I_r & \emptyset \\ \emptyset & \emptyset \end{bmatrix} \begin{bmatrix} X_1' & X_2' & \dots & X_r' \\ \psi_1' & \psi_2' & & \psi_r' \end{bmatrix} =$$

$$= \begin{bmatrix} X_1 & \emptyset \\ X_2 & \emptyset \\ \vdots & \vdots \\ X_r & \emptyset \end{bmatrix} \begin{bmatrix} X_1' & X_2' & \dots & X_r' \\ \Psi_1' & \Psi_2' & \dots & \Psi_r' \end{bmatrix} =$$

$$= \begin{bmatrix} J_t & \emptyset & \dots & \emptyset \\ \emptyset & J_t & \dots & \emptyset \\ \emptyset & \emptyset & \dots & J_t \end{bmatrix}$$

Fazendo $T_1 = \rho\sigma^2 T$, $T_2 = \emptyset$ e $k = (1-\rho)\sigma^2$ temos

$$XT_1X' + (I-XB)T_2(I-XB) + kI = \rho\sigma^2 \begin{bmatrix} J_t & \emptyset & \dots & \emptyset \\ \emptyset & J_t & \dots & \emptyset \\ \vdots & & & \\ \emptyset & \emptyset & \dots & J_t \end{bmatrix} + (1-\rho)\sigma^2 I =$$

$$= V .$$

Portanto, a "condição dos autovetores" é válida para blocos completamente aleatorizados quando a variância intra-blocos é igual qualquer que seja o bloco. Então, nessa condição, os estimadores simples de mínimos quadrados são BLUE's.

REFERENCIA BIBLIOGRAFICA

ALBERT,A. (1973). The Gauss-Markoff Theorem for Regression Models with possibly singular covariances. SIAM Jour. Appl. Math. 24, 182 - 187.

HOFFMAN, K. e KUNZE, R. (1971). Álgebra Linear. Editora Polígono, São Paulo.

KEMPTHORNE, O. (1952). The Design and Analysis of Experiments. Wiley, New York,

KEMPTHORNE, O. (1955). The Randomization Theory of Experimental Inference. Journal of the American Statistical Association 50, 946 - 967.

KEMPTHORNE, O. (197?). Best linear unbiased estimation with arbitrary variance matrix. Notas não publicadas.

KRUSKAL, W. (1968). When are Gauss-Markoff and Least Squares estimators identical? A coordinate-free approach. Ann. Math. Stat. 39, 70 - 75.

RAO, C. R. (1971). Unified Theory of Linear Estimation. Sankhya, Series A 33, 371 - 394.

RAO, C.R. (1976). Estimation of parameters in a linear model. The Annals of Statistics. Vol. 4 Nº 6, 1023 - 1037.

RAO, C.R. (1978). Choice of Best Linear Estimators in the Gauss-Markoff model with a singular dispersion matrix. Communications in Statistics - Theory and Methods 13, 1199 - 1208.

SCHEFFÉ, H. (1959). <u>The Analysis of Variance</u>. Wiley, New York.

SEARLE, S. R. (1971). <u>Linear Models</u>. Wiley, New York.

SEBER, G. A. F. (1977). <u>Linear Regression Analysis</u>. Wiley, New York.

SEELY, J. (1970). (1) Linear spaces and unbiased estimation.
(2) Linear spaces and unbiased estimation-Application to the mixed linear model. Ann. Math. Stat. 41, 1725 - 1748.

SEELY, J. and ZYSKIND, G. (1971). Linear spaces and minimum variance estimation. Ann. Math. Stat. 42, 691 - 703.

WATSON, G. S. (1967). Linear least squares regression. Ann. Math. Stat. 38, 1679 - 1699.

WATSON, G. S. (1972). Prediction and the efficiency of least squares. Biometrika 59, 91 - 98.

ZYSKIND, G. and KEMPTHORNE, O. (1962). Covariance structures of randomized experiments under Additivity and their relation to least squares estimation. ARL Technical Report 149, Wright-Patterson Air Force Base, Ohio.

ZYSKIND, G. (1964). Topics in general linear models theory. ARL Technical Report 64 - 193, 93 - 168. Wright-Patterson Air Force Base, Dayton.

ZYSKIND, G. (1967). On canonical forms, non-negative covariance matrices and best and simple linear estimators in linear models. Ann. Math. Stat. 38, 1092 - 1109.

ZYSKIND, G. and MARTIN, F. G. (1969). On best linear estimation and a general Gauss-Markoff theorem in linear models with arbitrary non-negative covariance structure. SIAM Jour..Appl. Math. 17, 1190 - 1202.