ANÁLISE DE REGRESSÃO EM
POPULAÇÕES FINITAS

Silvia Nagib Elian

Tese apresentada
ao
Instituto de Matemática e Estatística
da
Universidade de São Paulo
Para Obtenção do Grau de Doutor
em
Estatística

Orientador: Prof. Dr. Josemar Rodrigues

São Paulo, Dezembro de 1991

AGRADECIMENTOS

Estes agradecimentos surgiram gradativamente, a cada etapa vencida, quando, na alegria de mais um passo, se juntando à emoção da proximidade do sonho de atingir o final, vinha a lembrança de alguém que, de alguma forma, contribuiu ou torceu por aquilo.

Agradeço:

Ao Professor Josemar Rodrigues, pela dedicação e incentivo constantes, pela precisão e segurança da orientação e principalmente, pelo aprendizado da convivência com seu trabalho sério, introduzindo-me nos caminhos da pesquisa.

Ao Professor Heleno Bolfarine, por toda sua atenção, propondo o problema do Capítulo 2, estando disponível a discussões e acompanhando meu trabalho, com a paciência e empenho de sempre.

Ao Professor José Galvão Leite, que de forma atenciosa e espontânea, contribuiu com valiosas sugestões e esclarecimentos.

À Professora Marli Mikael da Costa Neves, por seus esclarecimentos, tão importantes num momento fundamental deste trabalho e também pela constante amizade.

Aos Professores Flávio Wagner Rodrigues e Siang Wun Song, pela atenção quando membros da C.P.G.

À Carmen Diva Saldiva de André, a colega e amiga que sempre confiou, me apoiou e defendeu, pela importância de sua amizade e incentivo.

À Elisete da Conceição Quintaneiro Aubin, por ter compartilhado comigo bons e maus momentos.

À Izilda, pelo paciente trabalho de datilografia e pela forma agradável com que sempre me atendeu.

Ao Aires, que teve tanta paciência e compreensão.

Às várias pessoas que auxiliaram em tantas fases difíceis nestes anos, contribuindo assim de forma indireta para este trabalho e me fazendo, com seu carinho e amizade, jamais duvidar da bondade do ser humano.

Por último, mas não com menos intensidade, agradeço a meus pais, que mais do que ninguém, acompanharam, participaram e torceram. A eles dedico este trabalho.

ABSTRACT

The first part of this work is devoted to the optimal prediction of the finite population regression coefficient $\underline{B}_N$, under the superpopulation model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, with a nondiagonal covariance matrix $V = \text{Var}(\underline{y})$. Under this model, three different situations are considered. First, we assume that $\underline{\varepsilon} \sim N_N(\underline{0}, V)$, where $N_N(\underline{0}, V)$ denotes the N -variate normal distribution with mean vector $\underline{0}$ and covariance matrix V . Next, we work with the more general model $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ where no assumption besides that $E(\underline{\varepsilon}) = \underline{0}$ and $\text{Var}(\underline{\varepsilon}) = V$ are made. Finally, the best restricted Bayes least squares predictor of $\underline{B}_N$ is introduced.

In the second part, we propose a predictive likelihood approach in finite populations and use the profile predictive likelihood to get the maximum likelihood predictor of $\underline{\ell}'\underline{y}$ and $\underline{B}_N$.

Asymptotic properties of some predictors of $\underline{B}_N$ are also investigated in the last part of the work.

RESUMO

Na primeira parte deste trabalho, dedicamo-nos à previsão ótima do coeficiente de regressão populacional B_N , no modelo de superpopulação $y = X\beta + \varepsilon$ com matriz de covariância $V = \text{Var}(y)$ não diagonal. Sob este modelo, três diferentes situações são consideradas. Admitimos inicialmente que $\varepsilon \sim N_N(\underline{0}, V)$, onde $N_N(\underline{0}, V)$ representa a distribuição normal N-variada com média $\underline{0}$ e matriz de covariância V . Posteriormente, trabalhamos com um modelo mais geral $y = X\beta + \varepsilon$, onde supomos apenas que $E(\varepsilon) = \underline{0}$ e $\text{Var}(\varepsilon) = V$. Finalmente, apresentamos o melhor previsor de mínimos quadrados Bayesiano restrito de B_N .

Na segunda parte do trabalho, introduzimos uma abordagem de verossimilhança preditiva em populações finitas e utilizamos a função de verossimilhança "profile" preditiva, obtendo o previsor de máxima verossimilhança de $\ell'y$ e B_N .

Finalizando, desenvolvemos algumas propriedades assintóticas do previsor de B_N .

ÍNDICE

CAPÍTULO 1 - INTRODUÇÃO.....	1
CAPÍTULO 2 - PREVISÃO DO COEFICIENTE DE REGRESSÃO $\underline{B}_N$	9
2.1 - Introdução	9
2.2 - O Vetor de Coeficientes de Regressão $\underline{B}_N$	14
2.3 - Previsão Ótima de $\underline{B}_N$ sob o Modelo com Distribuição Normal, V não Diagonal	21
2.4 - Previsão Linear Ótima de $\underline{B}_N$, com Matriz de Covariância V não Diagonal	40
2.5 - Previsão Bayesiana não Paramétrica para $\underline{B}_N$ em Modelos com Matriz de Covariância não Diagonal	45
2.6 - Previsão Ótima da Combinação Linear das Componentes de $\underline{B}_N$	54
2.7 - Exemplos	55
2.8 - Considerações Finais	64
CAPÍTULO 3 - FUNÇÕES DE VEROSSIMILHANÇA PREDITIVAS E APLICAÇÕES A POPULAÇÕES FINITAS	66
3.1 - Introdução	66
3.2 - Previsão em Populações Finitas Através de "Profiles" Preditivas	67
3.3 - Previsão de Máxima Verossimilhança para y_r	70

3.4 - A Função de Verossimilhança Preditiva para y_r	75
3.5 - Previsão de Máxima Verossimilhança para $\underline{\ell}'y$ e B_N	77
3.6 - Exemplos	85
3.7 - Considerações Finais	96
 CAPÍTULO 4 - TEORIA ASSINTÓTICA UNIFICADA	 97
4.1 - Introdução	97
4.2 - Normalidade Assintótica Unificada no Modelo de Regressão	 100
4.3 - Normalidade Assintótica para a Média Populacional	135
4.4 - Aplicações	139
4.5 - Considerações Finais	151
 CAPÍTULO 5 - CONCLUSÕES	 152
 APÊNDICE 1 - INFERÊNCIA EM POPULAÇÕES FINITAS - CONCEITOS BÁSICOS	 156
 APÊNDICE 2 - PREVISÃO DE $\underline{\ell}'y$	 164
A.2.1 - Introdução	164
A.2.2 - Previsão Linear Ótima para $\underline{\ell}'y$	165
A.2.3 - Previsão Ótima para $\underline{\ell}'y$ no Modelo com Distribuição Normal	 172
A.2.4 - Previsão Bayesiana para $\underline{\ell}'y$	177
 APÊNDICE 3 - RESULTADOS DA TEORIA DE MODELOS LINEARES	 189

APÊNDICE 4 - RESULTADOS ASSINTÓTICOS	195
A.4.1 - Introdução	195
A.4.2 - Teorema do Limite Central Unificado no Modelo de Locação	195
A.4.3 - Teorema do Limite Central no Modelo de Regressão ...	206
A.4.4 - Teoremas Auxiliares	221
RERERÊNCIAS BIBLIOGRÁFICAS	225

CAPÍTULO 1 - INTRODUÇÃO

Dada uma população finita $\mathcal{P} = \{1, 2, \dots, N\}$ de N elementos, onde ao i -ésimo elemento está associado y_i , valor de uma quantidade ou variável de interesse, o problema de inferência em populações finitas consiste basicamente em obter informação sobre qualquer função T dos valores $y_1, y_2, \dots, y_N$.

A inferência nesta área difere muito da feita em populações infinitas por inúmeros motivos. O primeiro deles é que, obtida a amostra de tamanho n , o problema da estimação de T se reduz ao da estimação da parte não observada na amostra, função dos $N-n$ elementos restantes. Assim, ao contrário de outros problemas, a amostra selecionada determina a quantidade a ser estimada.

Uma outra característica seria seu caráter não paramétrico, ou seja, na maior parte dos trabalhos, não se adotava modelos probabilísticos, ou quando adotados, a inferência não era baseada nas distribuições de probabilidades definidas por estes modelos. Geralmente toda inferência era feita com base na distribuição gerada pelo planejamento amostral, já que os valores $y_1, y_2, \dots, y_N$ eram considerados fixos. A este enfoque ou abordagem, estamos denominando "abordagem clássica".

Mais ou menos por volta de 1970, começaram a surgir os

primeiros trabalhos introduzindo um diferente enfoque na inferência em populações finitas. Esta nova abordagem, denominada abordagem de modelos de superpopulação, admitia que $Y_1, Y_2 \dots Y_N$ eram valores de N variáveis aleatórias com uma distribuição de probabilidades conjunta. Algumas características desta distribuição eram supostas conhecidas e toda inferência estatística era feita com base na referida distribuição.

O emprego de tal abordagem enriqueceu muito esta área da Estatística, permitindo que muitas técnicas da inferência clássica pudessem também ser aplicadas em populações finitas.

Embora referências anteriores talvez possam ser citadas, os trabalhos que praticamente introduziram a abordagem de modelos de superpopulação foram Royall [1970] e posteriormente e de uma forma mais definitiva, Royall e Herson [1973 a,b].

Estes artigos, apesar de introdutórios, eram muito objetivos e se preocupavam mais com a parte técnica, apresentando apenas breves comentários sobre os "fundamentos" da abordagem à qual se dedicavam.

Preenchendo esta lacuna, Smith [1976] faz um excelente levantamento sobre os princípios da teoria da amostragem, tratando de ambas abordagens, a clássica e a de modelos de superpopulação e colocando alguns argumentos bastante fortes incentivando a utilização da última, visto que um certo receio estava ocorrendo nos estatísticos mais clássicos e também uma pequena divergência entre adeptos de cada uma das abordagens começava a surgir.

Também nesta linha de "fundamentos", um trabalho de especial importância foi Cassel, Särndal e Wretman [1977], tendo o mérito de ser o primeiro livro de inferência em populações finitas a desenvolver a abordagem de modelos de superpopulação, destinando quatro dos seus sete capítulos a este estudo.

Segundo alguns autores, a freqüente utilização da abordagem clássica devia-se ao fato que a qualidade da inferência dependeria apenas do processo de seleção da amostra (planejamento amostral), que é criado e controlado pelo pesquisador.

A maior crítica à abordagem de modelos de superpopulação é que, ao contrário da outra, os resultados dependiam da validade do modelo adotado, validade esta que nunca pode ser perfeitamente comprovada na prática.

Estes fatos geraram alguns trabalhos mais moderados, onde os autores utilizavam uma inferência combinada, dependente do planejamento e também do modelo de superpopulação, com o intuito de "conciliar" as duas teorias.

Outros autores dedicaram ao estudo de procedimentos robustos, ou seja, métodos que assegurassem a qualidade da inferência quando o modelo adotado apresentasse falhas em algumas de suas suposições.

Inúmeros foram os artigos sobre este assunto, dentre os quais destacamos Royall e Herson [1973 a], Scott, Brewer e Ho [1978] e Tallis [1978]. Nestes trabalhos, partindo de particulares modelos de regressão polinomial e definido o

conceito de previsor ótimo, determinava-se o previsor linear ótimo para o total populacional $\sum_{i=1}^N y_i$ e posteriormente estudava-se as chamadas condições de robustez, sob as quais o previsor manteria sua otimalidade num modelo de regressão polinomial de grau superior ao inicialmente adotado.

Pereira e Rodrigues [1983] fazem um estudo semelhante no modelo linear geral $y = X\beta + \varepsilon$, onde $y = [y_1, y_2 \dots y_N]'$, apresentando o previsor linear ótimo para $\ell'y$, qualquer combinação linear dos $y_1, y_2 \dots y_N$, e estudando sua otimalidade sob modelos $y = X\beta^* + \varepsilon^*$, contendo outras variáveis auxiliares além das presentes no modelo inicial. Em ambos os modelos, as únicas suposições feitas são as de que ε e ε^* são vetores aleatórios cuja média é o vetor nulo e matriz de covariância $\sigma^2 V$, V diagonal conhecida.

Admitindo adicionalmente que o vetor aleatório y tem distribuição normal N -variada, isto é, que $y \sim N_N(X\beta, \sigma^2 V)$, Rodrigues, Bolfarine e Rogatko [1985] obtém o previsor ótimo de $\ell'y$ e mostram que este previsor coincide com o linear ótimo quando a distribuição de probabilidades de y não está especificada.

No primeiro modelo, sendo $\text{Var}(y) = \sigma^2 V$ uma matriz diagonal, as variáveis aleatórias $y_1, y_2 \dots y_N$ são duas a duas não correlacionadas. Já, no segundo, as suposições de que y tem distribuição normal multivariada e $\sigma^2 V$ é diagonal garantem que $y_1, y_2 \dots y_N$ são independentes.

Trabalhando também com distribuição normal N -variada e V

diagonal, Royall [1976 b] apresenta um dos poucos artigos sobre funções de verossimilhança em populações finitas, no qual trata da previsão de $\underline{\ell}'\underline{y}$.

Quando $\text{Var}(\underline{y}) = \underline{V}$ é não diagonal, conhecida, Royall [1976 a] obtém o previsor linear ótimo do total populacional no modelo $\underline{y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$ e Tam [1987] mostra que o mesmo previsor é ótimo se admitirmos que $\underline{y}$ tem distribuição normal N-variada. O desenvolvimento feito, embora destinado à previsão do total, pode ser trivialmente estendido para a previsão de $\underline{\ell}'\underline{y}$.

Do que acabamos de expor, podemos notar que a previsão de $\underline{\ell}'\underline{y}$ através da abordagem de modelos de superpopulação já foi exaustivamente estudada.

Acreditamos também que, há bastante tempo, as dúvidas e receios a que nos referimos inicialmente não mais existem e a abordagem de modelos de superpopulação tem sido naturalmente aceita e muito utilizada por inúmeros autores e em vários trabalhos.

Uma referência completa sobre o assunto é Bolfarine e Zacks [1992], que trata do problema da previsão em populações finitas sob a abordagem de modelos de superpopulação, reunindo uma grande quantidade de resultados importantes desenvolvidos nos últimos anos.

Adicionalmente, os autores se dedicaram à previsão do vetor aleatório $\underline{B}_N = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{y}$, o coeficiente de regressão populacional.

Definido o conceito de previsor ótimo para uma quantidade

vetorial, foi determinado o previsor ótimo de $\underline{B}_N$ no modelo $\underline{y} \sim N_N(X\underline{\beta}, \sigma^2 V)$, quando V é uma particular matriz bloco diagonal, conhecida.

O caso em que V é não diagonal não foi abordado pelos autores e desconhecemos qualquer outro trabalho que tenha tratado deste problema.

Quando V é não diagonal, além de uma maior dificuldade algébrica, pois todos os cálculos envolvendo esta matriz se tornam mais complicados, necessitamos também de técnicas estatísticas para variáveis aleatórias correlacionadas.

A contribuição e os objetivos de nosso trabalho consistem em:

- Apresentar uma teoria de previsão para o coeficiente de regressão $\underline{B}_N$, semelhante à existente para $\underline{l}'\underline{y}$. Esta teoria incluirá previsão ótima, previsão "Bayesiana" e previsão de máxima verossimilhança, todas elas admitindo modelos de superpopulação com matriz de covariância qualquer.
- Introduzir uma técnica de previsão em populações finitas através de funções de verossimilhança preditivas e utilizá-la na previsão das principais quantidades de interesse.
- Apresentar uma teoria assintótica unificada, relativa simultaneamente à estimação de $\underline{\beta}$ e à previsão de $\underline{B}_N$.

No capítulo 2, nosso objetivo foi o da previsão de $\underline{B}_N$ quando $\text{Var}(\underline{y}) = V$ é não diagonal, conhecida.

Inicialmente, determinamos o previsor ótimo de $\underline{B}_N$ no modelo em que $\underline{y} \sim N_N(X\underline{\beta}, V)$, V não diagonal, estendendo assim o resultado obtido por Bolfarine e Zacks [1992].

Posteriormente, abandonando a suposição de que $\underline{y}$ tem distribuição normal multivariada, determinamos o previsor linear ótimo de $\underline{B}_N$ no modelo $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$, com $\text{Var}(\underline{y}) = V$, não diagonal.

Finalmente, adotando um modelo "Bayesiano" não paramétrico, que admite que $\underline{\beta}$ é um vetor aleatório com matriz de covariância conhecida, determinamos o melhor previsor de mínimos quadrados Bayesiano restrito de $\underline{B}_N$.

No capítulo 3, propomos uma teoria de verossimilhança preditiva em populações finitas através de "profiles preditivas", no modelo de superpopulação $\underline{y} \sim N_N(X\underline{\beta}, V)$, V uma matriz de covariância qualquer.

A técnica de funções de verossimilhança preditivas, destinada à inferência em populações infinitas, se encontra amplamente discutida em Bjornstad [1990]. A idéia da adaptação em populações finitas deve-se ao fato que, pelo nosso conhecimento, os poucos trabalhos sobre verossimilhança na área, por exemplo, Royall [1976 b], utilizavam outros enfoques e ainda supunham V diagonal.

O desenvolvimento presente no capítulo 3, além do interesse de introduzir a inferência de verossimilhança preditiva em populações finitas, produziu importantes resultados. Na previsão de $\underline{B}_N$, o procedimento proposto apresentou uma conexão com os resultados da seção 2.3,

capítulo 2, em que o modelo de superpopulação era $y \sim N_N(X\beta, V)$. Quando usado na previsão de $\ell'y$, encontramos resultados análogos aos obtidos por Royall [1976 b] através de outra técnica de verossimilhança e com V diagonal.

Finalizando, no capítulo 4, nos dedicamos a um estudo assintótico. Introduzido um conceito de convergência em populações finitas, apresentamos um teorema do limite central envolvendo o vetor aleatório B_N e seu previsor, no modelo $y = X\beta + \varepsilon$, com $\text{Var}(\varepsilon) = \sigma^2 V$, V diagonal.

O teorema obtido, extensão de um resultado de Fuller [1975], possibilitou a inferência de verossimilhança assintótica para B_N , quando a distribuição de probabilidades de y não está especificada.

CAPÍTULO 2 - PREVISÃO DO COEFICIENTE DE REGRESSÃO B_N

2.1 - Introdução

Existem inúmeros trabalhos sobre a estimação e previsão de quantidades lineares associadas a uma população finita.

Com este objetivo, as duas abordagens mais utilizadas são a "clássica" e a abordagem de modelos de superpopulação.

Na primeira, o único mecanismo aleatório no qual a inferência se baseia é a distribuição de probabilidades gerada pelo planejamento amostral, já na segunda, toda inferência é feita através de um modelo probabilístico, denominado modelo de superpopulação.

Os conceitos básicos da inferência em populações finitas, bem como as abordagens clássica e de modelos de superpopulação se encontram descritos brevemente no apêndice 1. Uma referência para a parte de fundamentos de cada uma das abordagens é Elian [1984].

Seja $\mathcal{P} = \{1, 2 \dots N\}$ uma população finita de N elementos, onde cada elemento é rotulado por um único número inteiro entre 1 e N .

Associados ao i -ésimo elemento desta população temos y_i , x_{i1} , $x_{i2} \dots x_{ip}$, onde y_i é o valor da variável de interesse e x_{i1} , $x_{i2} \dots x_{ip}$ são os valores de p variáveis auxiliares, não

aleatórias. Os valores y_i , $i = 1, 2 \dots N$ são desconhecidos e os valores $x_{i1}, x_{i2} \dots x_{ip}$ serão conhecidos para $i = 1, 2 \dots N$.

Utilizaremos ainda a variável x_{i0} constante igual a 1 para todo $i = 1, 2 \dots N$, quando desejarmos adotar modelos de regressão com intercepto.

Definiremos então

$$y' = [y_1, y_2 \dots y_N],$$

o vetor populacional desconhecido,

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & \dots & x_{Np} \end{bmatrix},$$

a matriz de variáveis auxiliares, totalmente conhecida, e

$$\tilde{x}'_i = [1 \quad x_{i1} \quad x_{i2} \quad \dots \quad x_{ip}],$$

o vetor que corresponde à i -ésima linha da matriz X e que contém os valores de todas as variáveis auxiliares para o i -ésimo elemento da população.

Tomada uma amostra s de n elementos, onde $s \subset \mathcal{P}$, e indicando por r ao conjunto dos rótulos das unidades populacionais não pertencentes à amostra ($s \cup r = \mathcal{P}$),

admitiremos até o final do trabalho que a população finita foi reordenada de modo que $y_1, y_2 \dots y_n$ são os valores observados na amostra (e portanto conhecidos) e $y_{n+1}, y_{n+2} \dots y_N$ os valores não observados.

Desta forma, $s = \{1, 2 \dots n\}$, $r = \{n+1, n+2 \dots N\}$ e particionamos y' como

$$y' = \begin{bmatrix} y'_s & y'_r \end{bmatrix},$$

onde

$$y'_s = \begin{bmatrix} y_1 & y_2 & \dots & y_n \end{bmatrix}$$

e

$$y'_r = \begin{bmatrix} y_{n+1} & y_{n+2} & \dots & y_N \end{bmatrix}.$$

A matriz X também será particionada em $X = \begin{bmatrix} X_s \\ X_r \end{bmatrix}$, com

$$X_s = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & \dots & x_{2p} \\ \vdots & & & & \\ 1 & x_{n1} & x_{n2} & \dots & x_{np} \end{bmatrix} \quad \text{e} \quad X_r = \begin{bmatrix} 1 & x_{n+11} & x_{n+12} & \dots & x_{n+1p} \\ 1 & x_{n+21} & x_{n+22} & \dots & x_{n+2p} \\ \vdots & & & & \\ 1 & x_{N1} & x_{N2} & \dots & x_{Np} \end{bmatrix}.$$

Na maior parte deste trabalho adotaremos o modelo de superpopulação

$$y = X \beta + \varepsilon \quad (2.1)$$

onde

$\underline{\varepsilon}' = [\varepsilon_1, \varepsilon_2 \dots \varepsilon_N]$ é um vetor aleatório,

$\underline{\beta}' = [\beta_0, \beta_1, \beta_2 \dots \beta_p]$ é o vetor de parâmetros desconhecidos,

$\underline{y}'$ e X são os vetores e matrizes já definidos, y aleatório e X não.

Além disso, admitiremos que valem as seguintes condições:

a) $E(\underline{\varepsilon}) = \underline{0}$, sendo $\underline{0}' = [0, 0 \dots 0]$ o vetor nulo de ordem N .

b) $\text{Var}(\underline{\varepsilon}) = \text{Var}(\underline{y}) = V$,

Com $\text{Var}(\underline{\varepsilon})$ denotando a matriz de covariância de $\underline{\varepsilon}$ e sendo V uma matriz simétrica, positiva definida, que na maioria dos casos suporemos conhecida.

Particionaremos ainda a matriz V como

$$V = \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix},$$

onde

$$V_s = \text{Var}(\underline{y}_s),$$

$$V_r = \text{Var}(\underline{y}_r),$$

$$V_{sr} = \text{Cov} (Y_s, Y_r) = \begin{bmatrix} \sigma_{1n+1} & \sigma_{1n+2} & \dots & \sigma_{1N} \\ \sigma_{2n+1} & \sigma_{2n+2} & \dots & \sigma_{2N} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{nn+1} & \sigma_{nn+2} & \dots & \sigma_{nN} \end{bmatrix} ,$$

com

$$\sigma_{ij} = \text{Cov}(Y_i, Y_j), \quad i = 1, 2, \dots, n, \quad j = n+1, n+2, \dots, N \quad \text{e}$$

$$V_{rs} = \text{Cov}(Y_r, Y_s) = V'_{sr}.$$

Utilizando a abordagem de modelos de superpopulação, grande parte da literatura em populações finitas trata da previsão de combinações lineares dos $Y_1, Y_2, \dots, Y_N$, que na forma matricial podem ser escritas como $\underline{\ell}'Y$, $\underline{\ell}' = [\ell_1 \ell_2 \dots \ell_N]$ um vetor de constantes conhecidas.

No apêndice 2 apresentamos os fatos relacionados com a previsão de $\underline{\ell}'Y$ que serão úteis no decorrer deste trabalho e dos quais alguns de nossos resultados se constituem em extensões.

Os apêndices 1 e 2 são importantes também para familiarizar o leitor com as técnicas mais comuns de inferência em populações finitas sob a abordagem de modelos de superpopulação.

Nosso objetivo neste capítulo é o da previsão do vetor $(p+1)$ - dimensional (ou p -dimensional, se o modelo for sem intercepto)

$$\underline{B}_N = (X'V^{-1}X)^{-1}X'V^{-1}Y ,$$

onde X , V e Y são os vetores e matrizes definidos no modelo

$$(2.1) \quad \left(X = \begin{bmatrix} X_s \\ X_r \end{bmatrix}, \quad Y = \begin{bmatrix} Y_s \\ Y_r \end{bmatrix} \right).$$

Da teoria de modelos lineares lembramos que

$\hat{\beta} = (X_s' V_s^{-1} X_s)^{-1} X_s' V_s^{-1} Y_s$ é o estimador de mínimos quadrados generalizados de β . Analogamente, B_N será o estimador de mínimos quadrados calculado com base na população toda, ou seja, se $\hat{\beta} = \hat{\beta}(X_s, V_s, Y_s)$, então $B_N = \hat{\beta}(X, V, Y)$.

Sob o modelo (2.1), X e V são conhecidas, mas Y não, e assim, B_N é desconhecido. Como Y é um vetor aleatório, o problema de obter informação sobre B_N segundo esta abordagem é um problema de previsão.

2.2 - O vetor de coeficientes de regressão B_N

Matematicamente, o vetor B_N é uma quantidade bem definida, no entanto, do ponto de vista prático, seu significado pode ser de difícil compreensão.

Para fixar idéias, vamos admitir um modelo de superpopulação

$Y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$, com $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$,
 Y_i, Y_j variáveis aleatórias independentes, $i \neq j$, $i, j = 1, 2, \dots, N$.

Verifica-se neste caso que

$$B_N = \begin{bmatrix} \bar{y} & -b\bar{x} \\ b \end{bmatrix}, \quad \text{onde} \quad b = \frac{\sum_{i=1}^N (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^N (x_i - \bar{x})^2},$$

$$\bar{y} = \frac{\sum_{i=1}^N y_i}{N} \quad \text{e} \quad \bar{x} = \frac{\sum_{i=1}^N x_i}{N}.$$

b corresponde então ao coeficiente angular da reta de mínimos quadrados que poderia ser ajustada se dispuséssemos dos valores dos N pares de observações (x_i, y_i) e $\bar{y} - b\bar{x}$ seria o coeficiente linear desta reta.

O estimador de b ,

$$\hat{b} = \frac{\sum_{i=1}^n (x_i - \bar{x}_s)(y_i - \bar{y}_s)}{\sum_{i=1}^n (x_i - \bar{x}_s)^2}, \quad \bar{x}_s = \frac{\sum_{i=1}^n x_i}{n}, \quad \bar{y}_s = \frac{\sum_{i=1}^n y_i}{n},$$

é freqüentemente determinado na teoria de amostragem clássica para o cálculo do "estimador regressão"

$$\bar{y}_r = \bar{y}_s + \hat{b} (\bar{x} - \bar{x}_s),$$

e o valor b surge no desenvolvimento de algumas propriedades de $\bar{y}_r$.

Considerando agora o modelo

$$Y_i = \beta x_i + \varepsilon_i, \text{ com } E(\varepsilon_i) = 0, \text{ Var}(\varepsilon_i) = \sigma^2 x_i,$$

Y_i e y_j variáveis aleatórias independentes, $i \neq j$, $i, j = 1, 2 \dots N$,

$$\underline{B}_N = (X' V^{-1} X)^{-1} X' V^{-1} Y$$

$$\text{onde } X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, \quad Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad V^{-1} = \text{diag} \left(\frac{1}{x_1}, \frac{1}{x_2}, \dots, \frac{1}{x_N} \right)$$

e assim,

$$\underline{B}_N = B = \frac{\sum_{i=1}^N Y_i}{\sum_{i=1}^N x_i}.$$

Tal valor de B poderia ser encarado como uma medida do tipo "taxa" ou "proporção".

Por exemplo, se a população finita fosse constituída de N cidades, X sendo a variável número de habitantes aptos para trabalho e Y o número de habitantes desempregados, a quantidade B corresponderia à taxa de desemprego populacional. Para variáveis desta natureza é também bastante razoável a hipótese de que $\text{Var}(Y_i) = \text{Var}(\varepsilon_i) = \sigma^2 x_i$.

Hung [1990] trabalha com uma população finita de N unidades, adotando o modelo de superpopulação

$$E(y_i) = f(\underline{x}_i, \beta), \quad i = 1, 2, \dots, N,$$

onde y é a variável dependente, $\underline{x}_i$ é o vetor de variáveis auxiliares e β é o parâmetro do modelo, cujo valor pertence ao espaço paramétrico Θ .

A função $f(\underline{x}_i, \beta)$ é usada para aproximar a relação existente entre y e $\underline{x}_i$ e sua adoção pode ser baseada em conhecimentos prévios ou evidências científicas.

O correspondente de β na população finita é definido como o valor em Θ que minimiza a função de perda

$$Q(\beta) = \sum_{i=1}^N [y_i - f(\underline{x}_i, \beta)]^2.$$

Este parâmetro, indicado por β_N , se constituiria no valor em Θ tal que

$$Q(\beta_N) = \inf_{\beta \in \Theta} Q(\beta).$$

Se $f(\underline{x}_i, \beta)$ for linear em β , isto é, se $E(y_i) = \underline{x}_i' \beta$, segue imediatamente que

$$\beta_N = (X'X)^{-1}X'Y,$$

onde X é a matriz de variáveis auxiliares.

Quando o modelo de superpopulação é tal que $\text{Var}(\underline{y}) = \underline{V}$, $\underline{V} \neq \underline{I}_N$, onde $\underline{I}_N$ é a matriz identidade de ordem N , $Q(\underline{\beta})$ deve ser modificada para

$$Q^p(\underline{\beta}) = (\underline{y} - \underline{X}\underline{\beta})' \underline{V}^{-1} (\underline{y} - \underline{X}\underline{\beta}),$$

a soma de quadrados do resíduo ponderada e neste caso, $\underline{\beta}_N = (\underline{X}' \underline{V}^{-1} \underline{X})^{-1} \underline{X}' \underline{V}^{-1} \underline{y}$.

Restringindo-se a modelos de regressão não linear, o autor constrói o estimador de $\underline{\beta}_N$ minimizando a correspondente soma de quadrados do resíduo ponderada amostral. Numa etapa seguinte, são estabelecidas propriedades assintóticas deste estimador.

Autores como Kish e Frankel [1974], Shah, Holt e Folsom [1977], Särndal [1982], Hartley e Sielken [1975] e Fuller [1975] também têm se dedicado à estimação (ou previsão) de $\underline{B}_N$, utilizando diferentes modelos e enfoques.

Särndal [1982] utiliza simultaneamente a abordagem clássica e de modelos de superpopulação na previsão de $\underline{B}_N$, sob o modelo (2.1), mas com matriz de covariância $\underline{V}$ diagonal.

Hartley e Sielken [1975] também utilizam o modelo (2.1), obtendo intervalos de tolerância para $\underline{C}' \underline{B}_N$, $\underline{C}$ vetor p -dimensional. Os autores apresentam um exemplo no qual as N unidades populacionais estão divididas em p estratos de tamanho N_j , $j = 1, 2, \dots, p$ e as p variáveis auxiliares são do tipo 0, 1, com

$$X_{ij} = \begin{cases} 1, & \text{se o } i\text{-ésimo indivíduo pertence ao } j\text{-ésimo estrato} \\ 0, & \text{caso contrário,} \end{cases}$$

$$i = 1, 2, \dots, N, \quad j = 1, 2, \dots, p.$$

Admite-se inicialmente que $V = I_N$ e que $\bar{y}_j$ é a média da variável Y para o j -ésimo estrato.

Nestas condições,

$$X'X = \text{diag} (N_1, N_2 \dots N_p), \quad X'y = \begin{bmatrix} N_1 \bar{y}_1 \\ N_2 \bar{y}_2 \\ \vdots \\ N_p \bar{y}_p \end{bmatrix}$$

e como consequência, B_N se reduz a $\begin{bmatrix} \bar{y}_1 \\ \bar{y}_2 \\ \vdots \\ \bar{y}_p \end{bmatrix}$, o vetor aleatório

contendo as médias dos p estratos.

Tomando $\underline{c}'$ como o vetor p -dimensional que contém 1 na j -ésima posição e 0 nas demais, $\underline{c}'B_N$ se reduz a $\bar{y}_j$.

Posteriormente, repete-se a previsão para matriz de covariância V diagonal, com

$$\text{Var}(y_{ij}) = \sigma_j^2, \quad j = 1, 2 \dots p, \quad i = 1, 2 \dots N_j,$$

admitindo assim que a variância de Y para elementos do mesmo estrato é constante, mas varia de estrato para estrato.

Em todos os trabalhos a que nos referimos, V é a matriz identidade ou então é uma matriz diagonal.

Bolfarine e Zacks [1992] derivam o previsor ótimo de B_N sob o modelo (2.1), com a suposição de que y tem distribuição normal multivariada e que a matriz de covariância V é bloco

diagonal ($V_{sr} = 0$), de modo que y_s e y_r são independentes.

Os autores introduzem uma definição de previsor ótimo para quantidades vetoriais e, utilizando o conceito de estatística suficiente preditiva e completa e uma versão do teorema de Lehmann-Scheffé para populações finitas, obtêm o previsor ótimo para B_N , que é $\hat{\beta} = (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} y_s$, o estimador de mínimos quadrados generalizados.

Este procedimento permite a determinação de previsores ótimos quando a distribuição de y é conhecida, e depende fortemente do fato que y_s e y_r são vetores aleatórios independentes.

Para o caso em que a distribuição de y não é especificada, ou quando $V_{sr} \neq 0$, até o momento, nenhum trabalho nesta linha é de nosso conhecimento.

Nossa principal contribuição neste capítulo será determinar o previsor ótimo de B_N sob o modelo (2.1), que supõe matriz de covariância qualquer.

Na seção 2.3 obtemos o melhor previsor não viciado de B_N , admitindo que y tem distribuição normal N-variada.

Na seção 2.4 trabalhamos com o mesmo modelo, sem a suposição de normalidade, obtendo o melhor previsor linear não viciado de B_N .

Finalmente, na seção 2.5 determinamos o melhor previsor de mínimos quadrados Bayesiano restrito de B_N .

Encerramos o capítulo com o cálculo do previsor para alguns modelos específicos.

2.3 - Previsão ótima de $\underline{B}_N$ sob o modelo com distribuição normal, V não diagonal

Quando desejamos prever quantidades unidimensionais, o erro quadrático médio usual é a medida de precisão do previsor. Um exemplo disto pode ser encontrado no apêndice 2, quando, na previsão de $\underline{\ell}'\underline{y}$, definia-se o previsor ótimo como aquele que, dentre todos os previsores de $\underline{\ell}'\underline{y}$ pertencentes a uma determinada classe, possuía o menor erro quadrático médio.

Na maioria dos casos, $\underline{B}_N$ é multidimensional e na determinação de seu previsor ótimo, torna-se necessária a utilização de um outro tipo de erro quadrático médio.

A próxima definição introduz o conceito de erro quadrático médio generalizado, proposto por Tam [1986 a] e utilizado por Bolfarine e Zacks [1992].

Definição 2.1 - Se $\hat{\underline{\theta}}$ e $\underline{\theta}$ são vetores aleatórios $(p+1)$ -dimensionais e $\hat{\underline{\theta}}$ é um previsor de $\underline{\theta}$, o erro quadrático médio generalizado de $\hat{\underline{\theta}}$ é definido como

$$EQMG(\hat{\underline{\theta}}) = \underline{\lambda}' E[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})'] \underline{\lambda} \quad ,$$

para cada vetor $(p+1)$ -dimensional $\underline{\lambda}$, $\underline{\lambda} \neq \underline{0}$.

Se $\hat{\underline{\theta}}$ for unidimensional,

$$EQMG(\hat{\underline{\theta}}) = \lambda^2 E[(\hat{\underline{\theta}} - \underline{\theta})^2] = \lambda^2 EQM(\hat{\underline{\theta}}) \quad ,$$

onde $EQM(\hat{\underline{\theta}})$ é o erro quadrático médio usual, apresentado no apêndice 2, definição A.2.3.

Definição 2.2 - $\hat{\underline{\theta}}^*$ é o melhor predictor não viciado de $\underline{\theta}$, ou é o predictor ótimo de $\underline{\theta}$ se for não viciado, ou seja, $E(\hat{\underline{\theta}}^* - \underline{\theta}) = \underline{0}$ e se para todo $\underline{\lambda}$,

$$EQMG(\hat{\underline{\theta}}^*) \leq EQMG(\hat{\underline{\theta}}),$$

qualquer que seja o predictor $\hat{\underline{\theta}}$ não viciado para $\underline{\theta}$.

Verificamos que se $\underline{\theta}$ é unidimensional, a definição 2.2 também se reduz à sua correspondente para o caso unidimensional.

Por outro lado, seja $\hat{\underline{\theta}}^* = [\hat{\theta}_0^* \ \hat{\theta}_1^* \ \dots \ \hat{\theta}_p^*]'$ o predictor ótimo de $\underline{\theta} = [\theta_0 \ \theta_1 \ \dots \ \theta_p]'$, de modo que

$$\underline{\lambda}' E[(\hat{\underline{\theta}}^* - \underline{\theta})(\hat{\underline{\theta}}^* - \underline{\theta})'] \underline{\lambda} \leq \underline{\lambda}' E[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})'] \underline{\lambda}, \quad \forall \underline{\lambda} \neq \underline{0}$$

para todo $\hat{\underline{\theta}} = [\hat{\theta}_0 \ \hat{\theta}_1 \ \dots \ \hat{\theta}_p]'$ não viciado de $\underline{\theta}$.

Tomando $\underline{\lambda}'_j = [0 \ 0 \ \dots \ 1 \ \dots \ 0]$ o vetor contendo 1 na j -ésima posição e 0 nas demais,

$$\begin{aligned} \underline{\lambda}'_j E[(\hat{\underline{\theta}}^* - \underline{\theta})(\hat{\underline{\theta}}^* - \underline{\theta})'] \underline{\lambda}_j &= E[(\hat{\theta}_j^* - \theta_j)^2] \leq \underline{\lambda}'_j E[(\hat{\underline{\theta}} - \underline{\theta})(\hat{\underline{\theta}} - \underline{\theta})'] \underline{\lambda}_j \\ &= E[(\hat{\theta}_j - \theta_j)^2], \end{aligned}$$

mostrando então que

$$E(\hat{\theta}_j^* - \theta_j)^2 = EQM(\hat{\theta}_j^*) \leq E(\hat{\theta}_j - \theta_j)^2 = EQM(\hat{\theta}_j).$$

Logo, fica demonstrado que se $\hat{\theta}^*$ é ótimo para θ , então cada componente de $\hat{\theta}^*$ será o previsor ótimo da respectiva componente de θ .

Salientamos ainda que todas as esperanças mencionadas nas definições 2.1 e 2.2 são calculadas com relação ao modelo de superpopulação adotado.

Definido o critério de otimalidade, passaremos ao objetivo principal desta seção. Dado o modelo

$$y \sim N_N(X\beta, V), \quad (2.2)$$

ou seja, y tem distribuição normal N -variada com média $X\beta$ e matriz de covariância V , onde y , X e β são os vetores e matrizes definidos anteriormente e

$$V = \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix} \text{ é conhecida, com } V_{sr} = V'_{rs} \neq 0,$$

obteremos o melhor previsor não viciado de

$$\underline{B}_N = (X'V^{-1}X)^{-1}X'V^{-1}y.$$

Este previsor, que indicaremos por $\hat{\theta}^*$, é tal que

$$E(\hat{\theta}^* - \underline{B}_N) = 0$$

e que dentre todos os previsores $\hat{\theta}$, não viciados para B_N , minimiza

$$EQMG(\hat{\theta}) = \lambda' E[(\hat{\theta} - B_N)(\hat{\theta} - B_N)'] \lambda,$$

para todo $\lambda \in \mathbb{R}^{p+1}$.

Observamos que B_N é uma quantidade linear em y , pois

$$B_N = (X'V^{-1}X)^{-1}X'V^{-1}y = Ay,$$

com

$$A = (X'V^{-1}X)^{-1}X'V^{-1}$$

e que pode ser decomposto numa parte conhecida após a obtenção da amostra e numa parte desconhecida, que deve ser "prevista". Este fato será demonstrado no próximo lema.

Lema 2.1

Se $y = \begin{bmatrix} y_s \\ y_r \end{bmatrix}$ e $X = \begin{bmatrix} X_s \\ X_r \end{bmatrix}$ então

$$B_N = H^{-1}BC^{-1}y_s + H^{-1}DE^{-1}y_r, \text{ onde}$$

$$B = X'_s - X'_r V_r^{-1} V_{rs}$$

$$C = V_s - V_{sr} V_r^{-1} V_{rs}$$

$$D = X'_r - X'_s V_s^{-1} V_{sr}$$

$$E = V_r - V_{rs} V_s^{-1} V_{sr} \quad e$$

$$H = BC^{-1}X_s + DE^{-1}X_r = X'V^{-1}X.$$

Prova:

Usando resultados para inversa de matrizes na forma particionada (Graybill [1983], página 184), temos que

$$V^{-1} = \begin{bmatrix} (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} & -V_s^{-1} V_{sr} (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} \\ -V_r^{-1} V_{rs} (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} & (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} \end{bmatrix}$$

$$X' V^{-1} = [X'_s (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} \quad -X'_r V_r^{-1} V_{rs} (V_s - V_{sr} V_r^{-1} V_{rs})^{-1}]$$

$$X'_r (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} - X'_s V_s^{-1} V_{sr} (V_r - V_{rs} V_s^{-1} V_{sr})^{-1}]$$

$$= [B \ C^{-1} \quad D \ E^{-1}]$$

$$X' V^{-1} X = X'_s (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} X_s - X'_r V_r^{-1} V_{rs} (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} X_s$$

$$- X'_s V_s^{-1} V_{sr} (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} X_r + X'_r (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} X_r =$$

$$= BC^{-1} X_s + DE^{-1} X_r = H.$$

Então B_N fica

$$B_N = (X' V^{-1} X)^{-1} X' V^{-1} Y = H^{-1} [BC^{-1} \ DE^{-1}] \begin{bmatrix} Y_s \\ Y_r \end{bmatrix}$$

$$= H^{-1} BC^{-1} Y_s + H^{-1} DE^{-1} Y_r.$$

□

É importante notarmos que H , B , C , D e E são matrizes conhecidas, não aleatórias e portanto, o problema de prever $\underline{B}_N$ se reduzirá a prever o vetor aleatório não observado $H^{-1}DE^{-1}\underline{y}_r$.

O próximo lema será utilizado na obtenção do previsor ótimo de $\underline{B}_N$.

Lema 2.2

Seja $\hat{\underline{\theta}}$ um previsor de $\underline{B}_N = A\underline{y}$. Nestas condições, para todo $\underline{\lambda} \in \mathbb{R}^{p+1}$,

$$E[\{\underline{\lambda}'A\underline{y} - \underline{\lambda}'\hat{\underline{\theta}}\}^2] = E[\{\underline{\lambda}'A\underline{y} - E(\underline{\lambda}'A\underline{y}|\underline{y}_s)\}^2] + E[\{\underline{\lambda}'\hat{\underline{\theta}} - E(\underline{\lambda}'A\underline{y}|\underline{y}_s)\}^2].$$

Prova:

Verificamos que $\underline{\lambda}'A = \underline{\lambda}'(X'V^{-1}X)^{-1}X'V^{-1}$ é um vetor linha N -dimensional, $\underline{\lambda}'\underline{B}_N = \underline{\lambda}'A\underline{y}$ é unidimensional e $\underline{\lambda}'\hat{\underline{\theta}}$ é função de $\underline{y}_s$, pois $\hat{\underline{\theta}}$ sendo um previsor, é função do vetor de dados $\underline{y}_s$.

Desta forma, o resultado segue como caso particular do lema A.2.2, apêndice 2, tomando

$$\underline{\ell}' = \underline{\lambda}'A \quad \text{e} \quad f(\underline{y}_s) = \underline{\lambda}'\hat{\underline{\theta}}.$$

□

O previsor ótimo de $\underline{B}_N$ sob o modelo (2.2) é dado no teorema a seguir.

Teorema 2.1

Sob o modelo (2.2), o melhor preditor não viciado de B_N é dado por

$$\hat{\theta}^* = H^{-1}BC^{-1} Y_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}Y_s + (X_r - V_{rs}V_s^{-1}X_s)\hat{\beta}] \quad (2.3)$$

onde

$$\hat{\beta} = (X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}Y_s.$$

Prova:

Seja $\hat{\theta}$ um preditor não viciado de B_N . Devido à decomposição dada no lema 2.1, os preditores de B_N serão da forma $\hat{\theta} = H^{-1}BC^{-1}Y_s + \hat{L}(Y_s)$, onde $\hat{L}(Y_s)$ é um preditor de $H^{-1}DE^{-1}Y_r$.

Como $\hat{\theta}$ é não viciado para B_N ,

$$E(\hat{L}(Y_s) - H^{-1}DE^{-1}Y_r) = 0 = E(\hat{L}(Y_s)) - H^{-1}DE^{-1}X_r\beta,$$

ou seja,

$$E(\hat{L}(Y_s)) = H^{-1}DE^{-1}X_r\beta.$$

$\hat{\theta}^*$ é não viciado para B_N , pois

$$\begin{aligned}
E \{ H^{-1} D E^{-1} [V_{rs} V_s^{-1} Y_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}] \} &= \\
= H^{-1} D E^{-1} [V_{rs} V_s^{-1} X_s \beta + (X_r - V_{rs} V_s^{-1} X_s) \beta] &= \\
= H^{-1} D E^{-1} X_r \beta, \text{ já que } E(\hat{\beta}) = \beta.
\end{aligned}$$

$$\begin{aligned}
EQMG(\hat{\theta}) &= \lambda' E[(\hat{\theta} - B_N)(\hat{\theta} - B_N)'] \lambda = \\
&= E[\lambda' (\hat{\theta} - B_N)(\hat{\theta} - B_N)' \lambda] = E[\{\lambda' (\hat{\theta} - B_N)\}^2]
\end{aligned}$$

sendo a última passagem consequência do fato que $\lambda' (\hat{\theta} - B_N)$ é unidimensional.

Usando o lema 2.2 e posteriormente o lema 2.1,

$$\begin{aligned}
EQMG(\hat{\theta}) &= E[(\lambda' B_N - \lambda' \hat{\theta})^2] = \\
&= E[\{\lambda' B_N - E(\lambda' B_N | Y_s)\}^2] + E[\{\lambda' \hat{\theta} - E(\lambda' B_N | Y_s)\}^2] = \\
&= E[\{\lambda' B_N - E(\lambda' B_N | Y_s)\}^2] + E[\{\lambda' H^{-1} B C^{-1} Y_s + \lambda' \hat{\Gamma}(Y_s) \\
&\quad - \lambda' H^{-1} B C^{-1} Y_s - E(\lambda' H^{-1} D E^{-1} Y_r | Y_s)\}^2].
\end{aligned}$$

$Y \sim N_N(X \beta, V)$ e assim, devido a propriedades da distribuição normal N-variada,

$$E(Y_r | Y_s) = (X_r - V_{rs} V_s^{-1} X_s) \beta + V_{rs} V_s^{-1} Y_s.$$

Portanto,

$$\begin{aligned} \text{EQMG}(\hat{\theta}) &= E[\{\tilde{\lambda}'\tilde{B}_N - E(\tilde{\lambda}'\tilde{B}_N | Y_s)\}^2] + E[\{\tilde{\lambda}'\hat{L}(Y_s) \\ &- \tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\beta - \tilde{\lambda}'H^{-1}DE^{-1}V_{rs}V_s^{-1}Y_s\}^2]. \end{aligned}$$

Analogamente,

$$\begin{aligned} \text{EQMG}(\hat{\theta}^*) &= E[\{\tilde{\lambda}'\tilde{B}_N - E(\tilde{\lambda}'\tilde{B}_N | Y_s)\}^2] + \\ &+ E[\{\tilde{\lambda}'H^{-1}DE^{-1}[V_{rs}V_s^{-1}Y_s + (X_r - V_{rs}V_s^{-1}X_s)\hat{\beta}] - \tilde{\lambda}'H^{-1}DE^{-1}[(X_r - V_{rs}V_s^{-1}X_s)\beta \\ &+ V_{rs}V_s^{-1}Y_s]\}^2] = E[\{\tilde{\lambda}'\tilde{B}_N - E(\tilde{\lambda}'\tilde{B}_N | Y_s)\}^2] + \\ &+ E[\{\tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\hat{\beta} - \tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\beta\}^2]. \end{aligned}$$

Como $Y_s \sim N_n(X_s\beta, V_s)$, $\tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\hat{\beta}$ é o estimador não viciado de variância mínima de $\tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\beta$ (ver apêndice 3) e como $\tilde{\lambda}'\hat{L}(Y_s) - \tilde{\lambda}'H^{-1}DE^{-1}V_{rs}V_s^{-1}Y_s$ é não viciado para $\tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\beta$,

$$\begin{aligned} &E[\{\tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\hat{\beta} - \tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\beta\}^2] \\ &\leq E[\{\tilde{\lambda}'\hat{L}(Y_s) - \tilde{\lambda}'H^{-1}DE^{-1}V_{rs}V_s^{-1}Y_s - \tilde{\lambda}'H^{-1}DE^{-1}(X_r - V_{rs}V_s^{-1}X_s)\beta\}^2], \end{aligned}$$

o que mostra que

$$\text{EQMG}(\hat{\theta}^*) \leq \text{EQMG}(\hat{\theta}) \quad \forall \lambda \in \mathbb{R}^{p+1}.$$

□

Calcularemos a seguir o erro quadrático médio generalizado do previsor ótimo $\hat{\theta}^*$.

Temos que

$$\underline{B}_N = H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}Y_r,$$

$$\begin{aligned} \hat{\theta}^* &= H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}Y_s + (X_r - V_{rs}V_s^{-1}X_s)\hat{\beta}] = \\ &= H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}Y_s + D'(X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}Y_s] \end{aligned}$$

e com isto,

$$\begin{aligned} \hat{\theta}^* - \underline{B}_N &= [H^{-1}DE^{-1}V_{rs}V_s^{-1} + H^{-1}DE^{-1}D'(X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}]Y_s - H^{-1}DE^{-1}Y_r \\ &= M_1 Y_s - M_2 Y_r, \end{aligned}$$

onde

$$M_1 = H^{-1}DE^{-1}V_{rs}V_s^{-1} + H^{-1}DE^{-1}D'(X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}$$

e

$$M_2 = H^{-1}DE^{-1}.$$

Portanto,

$$\begin{aligned}
EQMG(\hat{\theta}^*) &= E[\{\tilde{\lambda}'(\hat{\theta}^* - \underline{B}_N)\}^2] = \text{var}(\tilde{\lambda}'\hat{\theta}^* - \tilde{\lambda}'\underline{B}_N) = \\
&= \text{var}(\tilde{\lambda}'M_1Y_s - \tilde{\lambda}'M_2Y_r) = \tilde{\lambda}'M_1V_sM_1'\tilde{\lambda} + \tilde{\lambda}'M_2V_rM_2'\tilde{\lambda} \\
&- 2\tilde{\lambda}'M_1V_{sr}M_2'\tilde{\lambda} = \tilde{\lambda}'H^{-1}DE^{-1}V_{rs}V_s^{-1}V_{sr}'(E^{-1})'D'(H^{-1})'\tilde{\lambda} \\
&+ \tilde{\lambda}'H^{-1}DE^{-1}D'(X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}V_{sr}'(E^{-1})'D'(H^{-1})'\tilde{\lambda} + \\
&+ \tilde{\lambda}'H^{-1}DE^{-1}V_{rs}V_s^{-1}X_s(X_s'V_s^{-1}X_s)^{-1}D(E^{-1})'D'(H^{-1})'\tilde{\lambda} + \\
&+ \tilde{\lambda}'H^{-1}DE^{-1}D'(X_s'V_s^{-1}X_s)^{-1}D(E^{-1})'D'(H^{-1})'\tilde{\lambda} + \\
&+ \tilde{\lambda}'H^{-1}DE^{-1}V_r(E^{-1})'D'(H^{-1})'\tilde{\lambda} - 2\tilde{\lambda}'H^{-1}DE^{-1}V_{rs}V_s^{-1}V_{sr}'(E^{-1})'D'(H^{-1})'\tilde{\lambda} \\
&- 2\tilde{\lambda}'H^{-1}DE^{-1}D'(X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}V_{sr}'(E^{-1})'D'(H^{-1})'\tilde{\lambda}
\end{aligned}$$

e, após simplificações, obtemos

$$EQMG(\hat{\theta}^*) = \tilde{\lambda}'H^{-1}DE^{-1}[V_r - V_{rs}V_s^{-1}V_{sr}' + D'(X_s'V_s^{-1}X_s)^{-1}D](E^{-1})'D'(H^{-1})'\tilde{\lambda}.$$

Observando que E e H são matrizes simétricas, $EQMG(\hat{\theta}^*)$ pode ainda ser escrito como

$$\begin{aligned}
EQMG(\hat{\theta}^*) &= \tilde{\lambda}'H^{-1}DE^{-1}[V_r - V_{rs}V_s^{-1}V_{sr}' + D'(X_s'V_s^{-1}X_s)^{-1}D]E^{-1}D'H^{-1}\tilde{\lambda} \\
&= \tilde{\lambda}'H^{-1}DE^{-1}[E + D'(X_s'V_s^{-1}X_s)^{-1}D]E^{-1}D'H^{-1}\tilde{\lambda}. \quad (2.4)
\end{aligned}$$

Bolfarine e Zacks [1992] mostraram que $\hat{\beta}$ é o previsor ótimo de β_N segundo a definição 2.2, no modelo $y \sim N_N(X\beta, V)$, com V bloco diagonal. Sendo o modelo (2.2) uma extensão deste, $\hat{\theta}^*$, o previsor que obtivemos, deve se reduzir a $\hat{\beta}$ quando $V_{sr} = 0$. Isto realmente ocorre, conforme demonstra o próximo corolário.

Corolário 2.1

No modelo (2.2), se $V_{sr} = 0$, então

$$\hat{\theta}^* = \hat{\beta} \quad e,$$

$$EQMG(\hat{\theta}^*) = \lambda' (X'V^{-1}X)^{-1} X_r' V_r^{-1} X_r (X_s' V_s^{-1} X_s)^{-1} \lambda.$$

Prova:

$$\hat{\theta}^* = H^{-1} B C^{-1} y_s + H^{-1} D E^{-1} [V_{rs} V_s^{-1} y_s + D' \hat{\beta}]$$

Se $V_{sr} = 0$,

$$D = X_r', \quad B = X_s', \quad C = V_s, \quad E = V_r,$$

$$H = X_s' V_s^{-1} X_s + X_r' V_r^{-1} X_r \quad e$$

$$\hat{\theta}^* = (X_s' V_s^{-1} X_s + X_r' V_r^{-1} X_r)^{-1} X_s' V_s^{-1} y_s +$$

$$\begin{aligned}
& + (X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r)^{-1} X'_r V_r^{-1} X_r (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} Y_s = \\
& = (X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r)^{-1} [I + X'_r V_r^{-1} X_r (X'_s V_s^{-1} X_s)^{-1}] X'_s V_s^{-1} Y_s \\
& = (X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r)^{-1} [X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r] (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} Y_s \\
& = \hat{\beta}.
\end{aligned}$$

Além disso, aplicando (2.4) para $V_{sr} = 0$, o erro quadrático médio generalizado de $\hat{\theta}^* = \hat{\beta}$ fica

$$\begin{aligned}
EQMG(\hat{\theta}^*) & = \lambda' (X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r)^{-1} X'_r V_r^{-1} [V_r + X_r (X'_s V_s^{-1} X_s)^{-1} X'_r] \\
& \quad \times V_r^{-1} X_r (X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r)^{-1} \lambda \\
& = \lambda' (X' V^{-1} X)^{-1} [X'_r V_r^{-1} X_r + X'_r V_r^{-1} X_r (X'_s V_s^{-1} X_s)^{-1} X'_r V_r^{-1} X_r] (X' V^{-1} X)^{-1} \lambda \\
& = \lambda' (X' V^{-1} X)^{-1} X'_r V_r^{-1} X_r (X'_s V_s^{-1} X_s)^{-1} [X'_s V_s^{-1} X_s + X'_r V_r^{-1} X_r] (X' V^{-1} X)^{-1} \lambda \\
& = \lambda' (X' V^{-1} X)^{-1} X'_r V_r^{-1} X_r (X'_s V_s^{-1} X_s)^{-1} \lambda,
\end{aligned}$$

que também coincide com o erro quadrático médio generalizado de $\hat{\beta}$ obtido por Bolfarine e Zacks [1992] quando V é bloco diagonal.

□

O previsor ótimo, apresentado no teorema 2.1, possui uma expressão algébrica bastante complicada, que no entanto se torna natural quando comparada com a do previsor ótimo para $\underline{\ell}'\underline{y}$, derivado por Tam [1987] e presente no apêndice 2.

Temos que

$$\underline{\ell}'\underline{y} = \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r \underline{y}_r,$$

onde

$\underline{\ell}'_s$ e $\underline{\ell}'_r$ contém, respectivamente, os coeficientes da combinação linear dos elementos pertencentes e não pertencentes à amostra (ver apêndice 2, seção A.2.1), e o previsor ótimo de $\underline{\ell}'\underline{y}$ sob o modelo (2.2) é

$$\hat{T}^* = \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r [V_{rs} V_s^{-1} \underline{y}_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}].$$

Analisando os dois previsores, verificamos que em ambos, o vetor não observado $\underline{y}_r$ está sendo previsto por

$$V_{rs} V_s^{-1} \underline{y}_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta},$$

já que $B_N = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}\underline{y}_r$.

Isto demonstra uma coerência entre o previsor que obtivemos e os resultados existentes na literatura.

Apesar desta coerência, seria interessante estudarmos se existem situações onde tal previsor assumiria uma forma mais simples e intuitiva.

Särndal e Wright [1984], no artigo "Cosmetic Form of Estimators in Survey Sampling", comentam que a maioria dos estatísticos se preocupa muito com as propriedades do estimador e pouco com a sua forma e chamam a atenção para a importância de trabalharmos com estimadores simples.

Estas considerações motivaram Bolfarine e Rodrigues [1988] a investigar sob que condições $\hat{T}^*$ se reduziria a $\underline{\ell}'X\hat{\beta}$, onde $\hat{\beta}$ é o estimador de mínimos quadrados generalizados.

$\underline{\ell}'X\hat{\beta}$ é bem mais intuitivo para a previsão de $\underline{\ell}'Y$ pois, sendo $E(\underline{\ell}'Y) = \underline{\ell}'X\beta$, estaríamos prevendo $\underline{\ell}'Y$ através do estimador natural de sua esperança, $\underline{\ell}'X\hat{\beta}$.

Além disso, previsores do tipo $X\hat{\beta}$ têm a vantagem de ser "do tipo projeção" e de possuírem uma interpretação geométrica interessante (a interpretação geométrica usual dos estimadores de mínimos quadrados).

Neste sentido, Bolfarine e Rodrigues [1988] estabelecem uma condição necessária e suficiente para que o previsor ótimo $\hat{T}^*$ seja do tipo projeção.

Os autores verificam que $\hat{T}^* = \underline{\ell}'X\hat{\beta}$ se e somente se $K\underline{\ell} \in M(X_s)$, onde $K = [V_s \ V_{sr}]$ e $M(X_s)$ representa o espaço linear gerado pelas colunas de X_s e apresentam ainda duas outras condições suficientes para que $\hat{T}^* = \underline{\ell}'X\hat{\beta}$.

Estas condições, embora satisfeitas para um número

razoável de modelos, não valem na maior parte dos casos.

O preditor simples e intuitivo para B_N é $\hat{\beta}$, o preditor de mínimos quadrados generalizados, pois como $B_N = (X'V^{-1}X)^{-1}X'V^{-1}Y$, é natural que seja previsto pelo seu correspondente amostral, $\hat{\beta} = (X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}Y_s$.

Quando V é bloco diagonal, conforme verificado no corolário 2.1, o preditor ótimo de B_N coincide com o de mínimos quadrados. O próximo teorema mostra que esta coincidência ocorre também no caso não diagonal.

Teorema 2.2

$$\hat{\theta}^* = \hat{\beta}, \quad \text{onde}$$

$$\hat{\theta}^* = H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}[X_r'\hat{\beta} + V_{rs}V_s^{-1}(Y_s - X_s'\hat{\beta})]$$

é o preditor ótimo de B_N no modelo (2.2) e

$$\hat{\beta} = (X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}Y_s$$

é o preditor de mínimos quadrados.

Prova:

Da demonstração do lema 2.1, temos que

$$A = (X'V^{-1}X)^{-1}X'V^{-1} = \begin{bmatrix} A_s & A_r \end{bmatrix} = \begin{bmatrix} H^{-1}BC^{-1} & H^{-1}DE^{-1} \end{bmatrix},$$

onde A_s e A_r são matrizes respectivamente $(p+1) \times n$ e $(p+1) \times (N-n)$.

Portanto,

$$\begin{aligned} \hat{\theta}^* - \hat{\beta} &= A_s y_s + A_r [X_r \hat{\beta} + V_{rs} V_s^{-1} (y_s - X_s \hat{\beta})] - (X' V^{-1} X)^{-1} X' V^{-1} X \hat{\beta} \\ &= A_s y_s + A_r X_r \hat{\beta} + A_r V_{rs} V_s^{-1} (y_s - X_s \hat{\beta}) - A_s X_s \hat{\beta} - A_r X_r \hat{\beta} \\ &= (A_s + A_r V_{rs} V_s^{-1}) (y_s - X_s \hat{\beta}) \\ &= [A_s \quad A_r] \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} V_s^{-1} (y_s - X_s \hat{\beta}) \\ &= (X' V^{-1} X)^{-1} X' V^{-1} \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} V_s^{-1} (y_s - X_s \hat{\beta}). \end{aligned}$$

Como $V V^{-1} = I_N$, $\begin{bmatrix} V_s & V_{sr} \\ n \times n & n \times (N-n) \end{bmatrix}$ sendo a parte superior de V

é tal que

$$[V_s \quad V_{sr}] V^{-1} = \begin{bmatrix} 1 & 0 & \dots & 0 & 0 & \dots & 0 \\ 0 & 1 & \dots & 0 & 0 & \dots & 0 \\ \vdots & & & & & & \\ 0 & 0 & \dots & 1 & 0 & \dots & 0 \end{bmatrix}_{n \times N} = \begin{bmatrix} I_n & 0_{n \times (N-n)} \end{bmatrix},$$

$\downarrow$ n colunas $\downarrow$ $N-n$ colunas

onde $0_{n \times (N-n)}$ representa a matriz nula com n linhas e $N-n$ colunas e I_n a matriz identidade de ordem n .

Logo,

$$V^{-1} \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} = \begin{bmatrix} I_n \\ 0_{(N-n) \times n} \end{bmatrix},$$

$$X' V^{-1} \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} = \begin{bmatrix} X'_s & X'_r \end{bmatrix} \begin{bmatrix} I_n \\ 0_{(N-n) \times n} \end{bmatrix} = X'_s,$$

e assim,

$$\begin{aligned} \hat{\theta}^* - \hat{\beta} &= (X' V^{-1} X)^{-1} X'_s V_s^{-1} (\chi_s - X_s \hat{\beta}) \\ &= (X' V^{-1} X)^{-1} [X'_s V_s^{-1} \chi_s - X'_s V_s^{-1} X_s \hat{\beta}] \\ &= (X' V^{-1} X)^{-1} [X'_s V_s^{-1} \chi_s - X'_s V_s^{-1} X_s (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} \chi_s] = 0, \end{aligned}$$

ou seja,

$$\hat{\theta}^* = \hat{\beta}.$$

□

O resultado presente neste teorema é muito importante. Ele implica numa considerável simplificação do previsor ótimo de B_N , além de garantir que este previsor é do tipo projeção. Salientamos mais uma vez que, com o previsor ótimo de $\ell' y$, esta simplificação só ocorria quando valiam condições adicionais.

É interessante e até surpreendente que a expressão (2.3) se reduza a $\hat{\beta}$, mas do ponto de vista intuitivo, a coincidência dos previsores talvez devesse até ser esperada.

$\hat{\beta}$, além de ótimo, possui uma forma simples, reunindo assim duas propriedades bastante desejáveis.

Verificamos que

$$\begin{aligned}\text{Var}(\hat{\beta} - \underline{B}_N) &= \text{Var}(\hat{\beta}) + \text{Var}(\underline{B}_N) - 2\text{Cov}(\hat{\beta}, \underline{B}_N) \\&= (X'_s V_s^{-1} X_s)^{-1} + (X' V^{-1} X)^{-1} - 2\text{Cov}((X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} y_s, (X' V^{-1} X)^{-1} X' V^{-1} y) \\&= (X'_s V_s^{-1} X_s)^{-1} + (X' V^{-1} X)^{-1} - 2(X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} [V_s \quad V_{sr}] V^{-1} X (X' V^{-1} X)^{-1} \\&= (X'_s V_s^{-1} X_s)^{-1} + (X' V^{-1} X)^{-1} - 2(X' V^{-1} X)^{-1} = (X'_s V_s^{-1} X_s)^{-1} - (X' V^{-1} X)^{-1},\end{aligned}$$

com a penúltima igualdade sendo devida ao resultado

$$[V_s \quad V_{sr}] V^{-1} X = X_s,$$

já demonstrado na prova do teorema 2.2.

Portanto,

$$\text{EQMG}(\hat{\beta}) = \lambda' [(X'_s V_s^{-1} X_s)^{-1} - (X' V^{-1} X)^{-1}] \lambda \quad (2.5)$$

e, como consequência do teorema 2.2, segue a igualdade dos erros quadráticos médios generalizados (2.4) e (2.5) e também que

$$H^{-1} D E^{-1} [E + D' (X'_s V_s^{-1} X_s)^{-1} D] E^{-1} D' H^{-1} = (X'_s V_s^{-1} X_s)^{-1} - (X' V^{-1} X)^{-1}.$$

Substituídas as matrizes H , D e E por suas respectivas expressões, esta igualdade pode se constituir num importante resultado matricial.

2.4 - Previsão Linear Ótima de $\underline{B}_N$, com Matriz de Covariância V não diagonal

Nesta seção, vamos abandonar a suposição de normalidade para o vetor aleatório $\underline{y}$, trabalhando com o modelo

$$\underline{y} = X \underline{\beta} + \underline{\varepsilon}, \quad (2.6)$$

onde

$$E(\underline{\varepsilon}) = \underline{0},$$

$$\text{Var}(\underline{\varepsilon}) = V = \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix}, \quad V_{sr} \neq 0,$$

$\underline{y}$, X e $\underline{\beta}$ os vetores e matrizes definidos anteriormente.

Sem a suposição de normalidade, nos restringiremos à classe dos previsores lineares de $\underline{B}_N$.

Definição 2.3 - Se T é uma matriz $(p+1) \times n$ de constantes conhecidas, a classe dos previsores lineares de $\underline{B}_N$ é dada por

$$\mathcal{L} = \{ \hat{\underline{\theta}} \mid \hat{\underline{\theta}} = T \underline{y}_s \}$$

e a classe dos previsores lineares não viciados de $\underline{B}_N$ é

$$U_{\mathcal{L}} = \{ \hat{\underline{\theta}} \mid \hat{\underline{\theta}} \in \mathcal{L} \text{ e } E(\hat{\underline{\theta}} - \underline{B}_N) = \underline{0} \}.$$

Definição 2.4 - $\hat{\theta}_{\ell}^*$ é o melhor previsor linear não viciado de B_N , ou é o previsor linear ótimo de B_N se $\hat{\theta}_{\ell}^* \in U_{\ell}$ e, para todo $\hat{\theta}_{\ell}$,

$$EQMG(\hat{\theta}_{\ell}^*) \leq EQMG(\hat{\theta}_{\ell}), \quad \forall \hat{\theta}_{\ell} \in U_{\ell}.$$

O próximo teorema fornece o previsor linear ótimo de B_N sob o modelo (2.6) e mostra que este previsor coincide com $\hat{\theta}_{\ell}^*$, o previsor ótimo dado no teorema 2.1.

Sua demonstração utiliza o resultado que apresentaremos no lema 2.3 e que se encontra provado com detalhes no apêndice 2, seção A.2.2.

Lema 2.3

No modelo (2.6)

$$\begin{aligned} E(\tilde{h}' Y_s - \tilde{\ell}' Y)^2 &= E[(\tilde{h}' - \tilde{\ell}' K' V_s^{-1})(Y_s - X_s \beta)]^2 \\ &+ \tilde{\ell}' V \tilde{\ell} - \tilde{\ell}' K' V_s^{-1} K \tilde{\ell} + [(\tilde{h}' X_s - \tilde{\ell}' X) \beta]^2, \end{aligned}$$

onde $\tilde{h}'$ e $\tilde{\ell}'$ são vetores respectivamente n e N dimensionais e $K = \begin{bmatrix} V_s & V_{sr} \end{bmatrix}$.

□

Teorema 2.3

No modelo (2.6), o previsor linear ótimo de B_N é

$$\hat{\theta}_{\ell}^* = H^{-1} B C^{-1} Y_s + H^{-1} D E^{-1} [V_{rs} V_s^{-1} Y_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}] = \hat{\beta}.$$

Prova:

Sejam $A = (X'V^{-1}X)^{-1}X'V^{-1}$, de modo que $\underline{B}_N = AY$ e T_{Y_s} um previsor linear não viciado de $\underline{B}_N$.

Nestas condições,

$$E(T_{Y_s} - AY) = TX_s\beta - AX\beta = 0 \text{ e}$$

$$EQMG(T_{Y_s}) = E[(\lambda'T_{Y_s} - \lambda'AY)^2].$$

No lema 2.3, tomando

$$\underline{h}' = \lambda'T \text{ e } \underline{\ell}' = \lambda'A, \text{ obtemos}$$

$$EQMG(T_{Y_s}) =$$

$$E\{[(\lambda'T - \lambda'AK'V_s^{-1})(Y_s - X_s\beta)]^2\} + \lambda'AVA'\lambda - \lambda'AK'V_s^{-1}KA'\lambda,$$

$$\text{onde } K = \begin{bmatrix} V_s & V_{sr} \end{bmatrix}.$$

Como $E[(\lambda'T - \lambda'AK'V_s^{-1})Y_s] = (\lambda'T - \lambda'AK'V_s^{-1})X_s\beta$, o previsor linear ótimo de $\underline{B}_N$, $T^*_{Y_s}$, deve minimizar

$$E\{(\lambda'T - \lambda'AK'V_s^{-1})Y_s - E[(\lambda'T - \lambda'AK'V_s^{-1})Y_s]\}^2,$$

ou seja, é tal que $\lambda'T^*_{Y_s} - \lambda'AK'V_s^{-1}Y_s$ é o estimador linear ótimo de

$$\begin{aligned}
E[(\tilde{\lambda}'T - \tilde{\lambda}'AK'V_s^{-1})Y_s] &= (\tilde{\lambda}'T - \tilde{\lambda}'AK'V_s^{-1})X_s\beta \\
&= (\tilde{\lambda}'AX - \tilde{\lambda}'AK'V_s^{-1}X_s)\beta.
\end{aligned}$$

Mas, pelo teorema de Gauss-Markov, o estimador linear ótimo de $(\tilde{\lambda}'AX - \tilde{\lambda}'AK'V_s^{-1}X_s)\beta$ é

$$(\tilde{\lambda}'AX - \tilde{\lambda}'AK'V_s^{-1}X_s)\hat{\beta}$$

e portanto,

$$\tilde{\lambda}'T^*Y_s - \tilde{\lambda}'AK'V_s^{-1}Y_s = (\tilde{\lambda}'AX - \tilde{\lambda}'AK'V_s^{-1}X_s)\hat{\beta}.$$

Como consequência,

$$\tilde{\lambda}'T^*Y_s - \tilde{\lambda}'AK'V_s^{-1}Y_s = \tilde{\lambda}'AX\hat{\beta} - \tilde{\lambda}'AK'V_s^{-1}X_s\hat{\beta},$$

ou ainda,

$$\tilde{\lambda}'T^*Y_s = \tilde{\lambda}'AK'V_s^{-1}Y_s + \tilde{\lambda}'AX\hat{\beta} - \tilde{\lambda}'AK'V_s^{-1}X_s\hat{\beta}.$$

Observando que

$$\tilde{\lambda}'AX = \tilde{\lambda}'H^{-1}[BC^{-1} \quad DE^{-1}] \begin{bmatrix} X_s \\ X_r \end{bmatrix} = \tilde{\lambda}'H^{-1}(BC^{-1}X_s + DE^{-1}X_r)$$

e

$$\tilde{\lambda}'AK'V_s^{-1} = \tilde{\lambda}'H^{-1}(BC^{-1} + DE^{-1}V_{rs}V_s^{-1})$$

obtemos

$$\begin{aligned}
 \tilde{\lambda}' T^* \gamma_s &= \\
 \tilde{\lambda}' H^{-1} (BC^{-1} + DE^{-1} V_{rs} V_s^{-1}) \gamma_s + \tilde{\lambda}' H^{-1} (BC^{-1} X_s + DE^{-1} X_r) \hat{\beta} - \\
 \tilde{\lambda}' H^{-1} (BC^{-1} + DE^{-1} V_{rs} V_s^{-1}) X_s \hat{\beta} &= \\
 = \tilde{\lambda}' H^{-1} BC^{-1} \gamma_s + \tilde{\lambda}' H^{-1} DE^{-1} [V_{rs} V_s^{-1} \gamma_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}] .
 \end{aligned}$$

Esta igualdade vale para todo $\tilde{\lambda} \in \mathbb{R}^{p+1}$, e assim,

$$\hat{\theta}_\ell^* = T^* \gamma_s = H^{-1} BC^{-1} \gamma_s + H^{-1} DE^{-1} [V_{rs} V_s^{-1} \gamma_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}] .$$

Devido ao teorema 2.2, $\hat{\theta}_\ell^* = \hat{\beta}$.

□

O previsor proposto no teorema 2.3 é linear ótimo, ou seja, tem erro quadrático médio generalizado mínimo na classe dos previsores lineares não viciados de $\underline{B}_N$, sob o modelo (2.6). Este modelo não especifica a distribuição de probabilidades de γ , tornando o resultado bastante geral.

Já, o teorema 2.1 estabelece a otimalidade de $\hat{\beta}$ na classe de todos os previsores não viciados de $\underline{B}_N$, uma classe bem maior, que contém inclusive a anterior.

O "preço" que pagamos por esta otimalidade numa classe mais ampla é a exigência da suposição de normalidade para γ .

Na ausência de normalidade, $\hat{\beta}$ é ótimo na classe dos

previsores lineares não viciados, podendo existir um predictor não linear não viciado melhor do que ele. Válida no entanto a normalidade, o melhor predictor não viciado é linear e igual a $\hat{\beta}$.

Fato semelhante ocorreu na previsão de $\ell'y$. Royall [1976a] obtém o predictor linear ótimo para $\ell'y$ sob o modelo (2.6) e Tam [1987] mostra que este predictor é ótimo se vale a suposição adicional de normalidade (ver apêndice 2).

2.5 - Previsão Bayesiana não Paramétrica para B_N em Modelos com Matriz de Covariância não Diagonal

Nesta seção, vamos determinar o melhor predictor de mínimos quadrados Bayesiano restrito para B_N .

A abordagem Bayesiana não paramétrica é livre de distribuição, ou seja, admite que β é um vetor aleatório com média e variância conhecidas, sem fazer nenhuma suposição sobre sua distribuição de probabilidades.

Dada esta abordagem, o modelo (2.6) será tratado como um modelo de superpopulação condicional

$$E(y \mid \beta) = X \beta \quad (2.7)$$

$$\text{Var}(y \mid \beta) = V, \text{ conhecida,}$$

e tal que

$$E(\underline{\beta} \mid \underline{\mu}) = \underline{\mu} \quad e$$

$$\text{Var}(\underline{\beta} \mid \underline{\mu}) = \Sigma ,$$

Σ conhecida.

O melhor preditor de mínimos quadrados Bayesiano restrito para $\underline{\ell}'\underline{y}$ sob o modelo (2.7) foi obtido em Rodrigues [1989] e está apresentado no apêndice 2.

Iniciaremos definindo o correspondente preditor para uma quantidade vetorial.

Se $\underline{w}$ é um vetor $(p+1)$ -dimensional e Q uma matriz $(p+1) \times n$, a classe dos preditores lineares de $\underline{B}_N$ é

$$\mathcal{L} = \{ \hat{\underline{\theta}} \mid \hat{\underline{\theta}} = \underline{w} + Q\underline{y}_s \}$$

e a classe dos preditores de mínimos quadrados Bayesianos restritos de $\underline{B}_N$ é

$$\bar{\mathcal{U}}_{\mathcal{L}} = \{ \hat{\underline{\theta}} \mid \hat{\underline{\theta}} \in \mathcal{L} \quad e \quad E(\hat{\underline{\theta}} - \underline{B}_N) = \underline{0} \}.$$

Definição 2.5 - Sob o modelo (2.7), $\hat{\underline{\theta}}_B^*$ é o melhor preditor de mínimos quadrados Bayesiano restrito de $\underline{B}_N$ se

$$a) \quad \hat{\underline{\theta}}_B^* = \underline{w}^* + Q\underline{y}_s \in \bar{\mathcal{U}}_{\mathcal{L}} \quad e$$

$$b) \quad EQMG(\hat{\underline{\theta}}_B^*) = E[\underline{\lambda}'(\hat{\underline{\theta}}_B^* - \underline{B}_N)(\hat{\underline{\theta}}_B^* - \underline{B}_N)'\underline{\lambda}] \leq EQMG(\hat{\underline{\theta}}), \quad \forall \hat{\underline{\theta}} \in \bar{\mathcal{U}}_{\mathcal{L}} \quad e$$

para todo $\underline{\lambda} \in \mathbb{R}^{p+1}$.

Os próximos lemas serão utilizados na determinação de $\hat{\underline{\theta}}_B^*$.

Lema 2.4

No modelo (2.7), $\hat{\theta} \in \bar{u}_{\mathcal{L}}$ se e somente se

$$E(\tilde{\lambda}' \tilde{W} + \tilde{F}' Y_s - \tilde{P}' \beta) = 0 \quad \forall \tilde{\lambda} \in \mathbb{R}^{p+1},$$

onde

$$\tilde{F}' = \tilde{\lambda}' Q - \tilde{\lambda}' A K' V_s^{-1}, \quad \tilde{P}' = \tilde{\lambda}' A (X - K' V_s^{-1} X_s) \quad \text{e} \quad A = (X' V^{-1} X)^{-1} X' V^{-1}.$$

Prova:

No modelo (2.7), $E(Y) = E(E(Y|\beta)) = E(X\beta) = XE(\beta) = X\mu$

$$\hat{\theta} \in \bar{u}_{\mathcal{L}} \Leftrightarrow E(\tilde{\lambda}' \hat{\theta} - \tilde{\lambda}' B_N) = 0 \quad \forall \tilde{\lambda}, \hat{\theta} \in \mathcal{L} \Leftrightarrow$$

$$E(\tilde{\lambda}' \tilde{W} + \tilde{\lambda}' Q Y_s - \tilde{\lambda}' A Y) = 0 \Leftrightarrow$$

$$E[\tilde{\lambda}' \tilde{W} + (\tilde{\lambda}' Q - \tilde{\lambda}' A K' V_s^{-1}) Y_s - \tilde{\lambda}' A (Y - K' V_s^{-1} X_s)] = 0 =$$

$$= E[\tilde{\lambda}' \tilde{W} + \tilde{F}' Y_s] - \tilde{\lambda}' A [XE(\beta) - K' V_s^{-1} X_s E(\beta)]$$

$$= E[\tilde{\lambda}' \tilde{W} + \tilde{F}' Y_s] - E[\tilde{\lambda}' A (X - K' V_s^{-1} X_s) \beta]$$

$$= E[\tilde{\lambda}' \tilde{W} + \tilde{F}' Y_s - \tilde{P}' \beta].$$

□

Lema 2.5

Para $\hat{\theta} \in \mathcal{L}$,

$$\begin{aligned} E[(\lambda' \hat{\theta} - \lambda' AY)^2] &= \text{Var}(\lambda' W + F' Y_s - P' \beta) + \lambda' AVA' \lambda \\ &\quad - \lambda' AK' V_s^{-1} KA' \lambda + [\lambda' W + (\lambda' QX_s - \lambda' AX) \mu]^2. \end{aligned}$$

□

Este lema, adaptação imediata de um resultado presente em Rodrigues [1989], se constitui numa versão Bayesiana do lema 2.3.

Devido aos lemas 2.4 e 2.5, o problema de obter $\hat{\theta}_B^*$ se reduz a determinar W e F tal que, para todo λ ,

$$E(\lambda' W + F' Y_s - P' \beta) = 0$$

e $\text{Var}(\lambda' W + F' Y_s - P' \beta)$ é mínima.

A determinação destes valores de W e F , que indicaremos respectivamente por W^* e F^* , pode ser feita diretamente através dos resultados desenvolvidos por Rao [1973, pág.234], descritos detalhadamente no apêndice 2, seção A.2.4.

Teorema 2.4

No modelo (2.7) com $\underline{\mu}$ conhecido, o melhor previsor de mínimos quadrados Bayesiano restrito de $\underline{B}_N$ é

$$\hat{\underline{\theta}}_B^* = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}[\underline{x}_r\bar{\underline{\beta}} + \underline{v}_{rs}\underline{v}_s^{-1}(\underline{y}_s - \underline{x}_s\bar{\underline{\beta}})],$$

onde $\bar{\underline{\beta}} = c\hat{\underline{\beta}} + (I - c)\underline{\mu}$,

$$\hat{\underline{\beta}} = (\underline{x}'_s \underline{v}_s^{-1} \underline{x}_s)^{-1} \underline{x}'_s \underline{v}_s^{-1} \underline{y}_s \quad e$$

$$c = [\Sigma^{-1} + \underline{x}'_s \underline{v}_s^{-1} \underline{x}_s]^{-1} (\underline{x}'_s \underline{v}_s^{-1} \underline{x}_s)$$

e, se $\underline{\mu}$ é desconhecido,

$$\hat{\underline{\theta}}_B^* = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}[\underline{x}_r\hat{\underline{\beta}} + \underline{v}_{rs}\underline{v}_s^{-1}(\underline{y}_s - \underline{x}_s\hat{\underline{\beta}})] = \hat{\underline{\theta}}.$$

Prova:

Devido ao lema 2.4,

$\hat{\underline{\theta}} = \underline{w} + Q\underline{y}_s$ é não viciado para $\underline{B}_N \Leftrightarrow$

$$E(\underline{\lambda}'\underline{w} + \underline{f}'\underline{y}_s - \underline{p}'\underline{\beta}) = 0, \quad \forall \underline{\lambda}$$

e como

$$EQMG(\hat{\underline{\theta}}) = E[(\underline{\lambda}'\hat{\underline{\theta}} - \underline{\lambda}'A\underline{y})^2],$$

pelo lema 2.5, o melhor preditor de mínimos quadrados Bayesiano restrito $\hat{\theta}_B^* = \tilde{W}^* + Q^* \tilde{Y}_s$ é tal que, para todo $\tilde{\lambda}$,

$$E(\tilde{\lambda}' \tilde{W}^* + \tilde{F}^* \tilde{Y}_s - \tilde{P}' \tilde{\beta}) = 0$$

e minimize $\text{Var}(\tilde{\lambda}' \tilde{W} + \tilde{F}' \tilde{Y}_s - \tilde{P}' \tilde{\beta})$, onde $\tilde{F}^* = \tilde{\lambda}' Q^* - \tilde{\lambda}' A K' V_s^{-1}$, $\tilde{P}' = \tilde{\lambda}' A (X - K' V_s^{-1} X_s)$ e $A = (X' V^{-1} X)^{-1} X' V^{-1}$.

Rao [1973] fornece os valores de a e $\underline{L}$ de modo que $E(a' + \underline{L}' \tilde{Y}_s - \tilde{P}' \tilde{\beta}) = 0$ e $\text{Var}(a' + \underline{L}' \tilde{Y}_s - \tilde{P}' \tilde{\beta})$ é mínima, $\tilde{P}$ um vetor de constantes $(p+1)$ -dimensional.

Estes valores, indicados respectivamente por a^* e $\underline{L}^*$, se encontram no teorema A.2.2, apêndice 2.

As matrizes W , G , Z e L^* deste teorema correspondem respectivamente a V_s , Σ , X_s e $\tilde{F}^*$, $\tilde{a}^{*'} = \tilde{\lambda}' \tilde{W}^*$ e $\tilde{P} = (X' - X_s' V_s^{-1} K) A' \tilde{\lambda}$ e assim, para μ conhecido,

$$\tilde{F}^* = (V_s + X_s \Sigma X_s')^{-1} X_s \Sigma \tilde{P} = V_s^{-1} X_s (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} \tilde{P}$$

e

$$a^* = \mu' \tilde{P} - \mu' X_s' \tilde{F}^*.$$

Lembrando que $\tilde{F}^{*'} = \tilde{\lambda}' Q^* - \tilde{\lambda}' A K' V_s^{-1}$, segue que

$$\begin{aligned} \tilde{\lambda}' \tilde{W}^* + \tilde{\lambda}' Q^* \tilde{Y}_s &= \tilde{\lambda}' \hat{\theta}_B^* = \\ &= (\mu' \tilde{P} - \mu' X_s' \tilde{F}^*)' + (\tilde{F}^{*'} + \tilde{\lambda}' A K' V_s^{-1}) \tilde{Y}_s, \quad \forall \tilde{\lambda} \end{aligned}$$

e assim

$$\begin{aligned}
\tilde{\lambda}' \hat{\theta}_B^* &= \tilde{\lambda}' A (X - K' V_s^{-1} X_s) \underline{\mu} - \tilde{\lambda}' A (X - K' V_s^{-1} X_s) (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} X_s' V_s^{-1} X_s \underline{\mu} \\
&\quad + \tilde{\lambda}' A K' V_s^{-1} Y_s + \tilde{\lambda}' A (X - K' V_s^{-1} X_s) (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} X_s' V_s^{-1} Y_s \\
&= \tilde{\lambda}' A (X - K' V_s^{-1} X_s) \underline{\mu} - \tilde{\lambda}' A (X - K' V_s^{-1} X_s) C \underline{\mu} + \tilde{\lambda}' A K' V_s^{-1} Y_s \\
&\quad + \tilde{\lambda}' A (X - K' V_s^{-1} X_s) C \hat{\beta}.
\end{aligned}$$

Verificamos que

$$\begin{aligned}
\tilde{\lambda}' A (X - K' V_s^{-1} X_s) &= \tilde{\lambda}' \begin{bmatrix} A_s & A_r \end{bmatrix} \left(\begin{bmatrix} X_s \\ X_r \end{bmatrix} - \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} V_s^{-1} X_s \right) = \\
&= \tilde{\lambda}' A_r (X_r - V_{rs} V_s^{-1} X_s),
\end{aligned}$$

onde $A_s = H^{-1} B C^{-1}$ e $A_r = H^{-1} D E^{-1}$,

$$\tilde{\lambda}' A K' V_s^{-1} Y_s = \tilde{\lambda}' \begin{bmatrix} A_s & A_r \end{bmatrix} \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} V_s^{-1} Y_s = \tilde{\lambda}' A_s Y_s + \tilde{\lambda}' A_r V_{rs} V_s^{-1} Y_s$$

e com isto,

$$\begin{aligned}
\tilde{\lambda}' \hat{\theta}_B^* &= \tilde{\lambda}' A_s Y_s + \tilde{\lambda}' A_r [V_{rs} V_s^{-1} Y_s + (X_r - V_{rs} V_s^{-1} X_s) \{ [I - C] \underline{\mu} + C \hat{\beta} \}] \\
&= \tilde{\lambda}' H^{-1} B C^{-1} Y_s + \tilde{\lambda}' H^{-1} D E^{-1} [X_r \bar{\beta} + V_{rs} V_s^{-1} (Y_s - X_s \bar{\beta})], \quad \forall \tilde{\lambda},
\end{aligned}$$

o que implica em

$$\hat{\theta}_B^* = H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}[X_r\bar{\beta} + V_{rs}V_s^{-1}(Y_s - X_s\bar{\beta})].$$

Para μ desconhecido,

$$F^* = V_s^{-1}X_s(X_s'V_s^{-1}X_s)^{-1}P, \quad a^* = \lambda'W^* = 0 \quad e$$

$$\lambda'\hat{\theta}_B^* = \lambda'Q^*Y_s = P'(X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}Y_s + \lambda'AK'V_s^{-1}Y_s$$

$$= \lambda'A(X - K'V_s^{-1}X_s)\hat{\beta} + \lambda'AK'V_s^{-1}Y_s$$

$$= \lambda'A_r(X_r - V_{rs}V_s^{-1}X_s)\hat{\beta} + \lambda'A_sY_s + \lambda'A_rV_{rs}V_s^{-1}Y_s.$$

Portanto,

$$\hat{\theta}_B^* = H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}Y_s + (X_r - V_{rs}V_s^{-1}X_s)\hat{\beta}] = \hat{\beta}.$$

□

Calcularemos a seguir o erro quadrático médio generalizado de $\hat{\theta}_B^*$. Utilizando o lema 2.5,

$$EQMG(\hat{\theta}_B^*) = E[(\lambda'\hat{\theta}_B^* - \lambda'AY)^2] =$$

$$\text{Var}(\lambda'W^* + F^{*'}Y_s - P'\beta) + \lambda'AVA'\lambda - \lambda'AK'V_s^{-1}KA'\lambda$$

$(\tilde{\lambda}' \tilde{W}^* + [\tilde{\lambda}' \tilde{Q}^* X_s - \tilde{\lambda}' A X] \mu = E(\tilde{\lambda}' \hat{\theta}_B^* - \tilde{\lambda}' B_N) = 0$, porque $\hat{\theta}_B^*$ é não viciado para B_N).

$\text{Var}(\tilde{\lambda}' \tilde{W}^* + \tilde{F}^{*'} Y_s - \tilde{P}' \beta)$, para μ conhecido e desconhecido, está calculada respectivamente nas expressões (A.2.11) e (A.2.12). Para μ conhecido,

$$\text{Var}(\tilde{\lambda}' \tilde{W}^* + \tilde{F}^{*'} Y_s - \tilde{P}' \beta) = \tilde{\lambda}' A (X - K' V_s^{-1} X_s) (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} (X' - X_s' V_s^{-1} K) A' \tilde{\lambda}$$

e para μ desconhecido,

$$\text{Var}(\tilde{\lambda}' \tilde{W}^* + \tilde{F}^{*'} Y_s - \tilde{P}' \beta) = \tilde{\lambda}' A (X - K' V_s^{-1} X_s) (X_s' V_s^{-1} X_s)^{-1} (X' - X_s' V_s^{-1} K) A' \tilde{\lambda}.$$

Após alguns cálculos verificamos que

$$A V A' = A V_s A_s' + A V_r A_{rs} A_s' + A V_s A_{sr} A_r' + A V_r A_r'$$

e

$$A K' V_s^{-1} K A' = A V_s A_s' + A V_s A_{sr} A_r' + A V_r A_{rs} A_s' + A V_r A_{rs} A_s' V_s^{-1} V_{sr} A_r'.$$

Como consequência destes fatos, obtemos então

$$\text{EQMG}(\hat{\theta}_B^*) =$$

$$= \tilde{\lambda}' A_r [(X_r - V_{rs} V_s^{-1} X_s) (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} (X_r' - X_s' V_s^{-1} V_{sr}) + V_r - V_{rs} V_s^{-1} V_{sr}] A_r' \tilde{\lambda}$$

$$= \tilde{\lambda}' H^{-1} D E^{-1} [E + D' (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} D] E^{-1} D' H^{-1} \tilde{\lambda}$$

para $\underline{\mu}$ conhecido e, para $\underline{\mu}$ desconhecido,

$$EQMG(\hat{\underline{\theta}}_B^*) = \underline{\lambda}' H^{-1} D E^{-1} [E + D' (X_s' V_s^{-1} X_s)^{-1} D] E^{-1} D' H^{-1} \underline{\lambda}.$$

Finalizando, observamos que quando $\underline{\mu}$ é desconhecido, o melhor preditor de mínimos quadrados Bayesiano restrito coincide com o preditor linear ótimo de $\underline{B}_N$, o mesmo ocorrendo com os erros quadráticos médios generalizados.

2.6 - Previsão Ótima da Combinação Linear das Componentes de $\underline{B}_N$

Dos resultados apresentados nas seções anteriores, que tratam da previsão ótima do vetor aleatório $\underline{B}_N$, podemos obter o preditor ótimo de qualquer combinação linear das componentes de $\underline{B}_N$.

Assim, se $\underline{h}' = [h_0, h_1, h_2 \dots h_p]$ é um vetor de constantes conhecidas, de modo que $\underline{h}' \underline{B}_N$ é uma combinação linear das componentes de $\underline{B}_N$, temos que

$$E(\hat{\underline{\beta}} - \underline{B}_N) = \underline{0} \implies E(\underline{h}' \hat{\underline{\beta}} - \underline{h}' \underline{B}_N) = 0,$$

que mostra que $\underline{h}' \hat{\underline{\beta}}$ é não viciado para $\underline{h}' \underline{B}_N$.

Além disso, como $\hat{\underline{\beta}}$ é ótimo (ou linear ótimo),

$$EQMG(\hat{\underline{\beta}}) = E[(\underline{\lambda}' \hat{\underline{\beta}} - \underline{\lambda}' \underline{B}_N)^2] \leq E[(\underline{\lambda}' \hat{\underline{\theta}} - \underline{\lambda}' \underline{B}_N)^2] = EQMG(\hat{\underline{\theta}}),$$

$\forall \underline{\lambda} \in \mathbb{R}^{p+1}$, $\forall \hat{\underline{\theta}}$ pertencente à classe considerada (classe dos

lineares não viciados nas seções 2.4 e 2.5 e classe dos não viciados na seção 2.3). Portanto, tomando $\lambda' = \underline{h}'$,

$$E[(\underline{h}'\hat{\underline{\beta}} - \underline{h}'\underline{B}_N)^2] = EQM(\underline{h}'\hat{\underline{\beta}}) \leq E[(\underline{h}'\hat{\underline{\theta}} - \underline{h}'\underline{B}_N)^2] = EQM(\underline{h}'\hat{\underline{\theta}}),$$

$\forall \underline{h}'\hat{\underline{\theta}}$ não viciado para $\underline{h}'\underline{B}_N$ e, nestas condições, fica demonstrado que $\underline{h}'\hat{\underline{\beta}}$ é o previsor ótimo de $\underline{h}'\underline{B}_N$.

Em particular, se quisermos obter o previsor ótimo de uma componente de $\underline{B}_N$, basta tomarmos $\underline{h}'$ como o vetor contendo o número 1 na componente desejada e 0 nas demais posições. Assim, $\hat{\beta}_j$, a j-ésima componente de $\hat{\underline{\beta}}$ é o previsor ótimo de B_j , a j-ésima componente de $\underline{B}_N$.

Evidentemente, as mesmas considerações valem para o melhor previsor de mínimos quadrados Bayesiano restrito de $\underline{B}_N$ quando $\underline{\mu}$ é conhecido, que não coincide com $\hat{\underline{\beta}}$.

2.7 - Exemplos

Nesta seção calcularemos $\underline{B}_N$ e seu previsor linear ótimo $\hat{\underline{\theta}}_\ell^*$ sob dois modelos de superpopulação especificados.

O cálculo de $\hat{\underline{\theta}}_\ell^*$, devido ao resultado do teorema 2.2, foi bastante simplificado. Dado este resultado, nos preocuparemos apenas em obter $\hat{\underline{\beta}}$.

Exemplo 2.1

Consideremos o modelo de superpopulação

$$y = X \beta + \varepsilon,$$

onde

$X = \mathbf{1}_N$ é o vetor coluna de dimensão N contendo 1 em todas as posições,

$$\beta = [\beta] \text{ e}$$

$$V = \text{Var}(y) = (1-\rho) I_N + \rho \mathbf{1}_N \mathbf{1}_N' = \begin{bmatrix} 1 & \rho & \rho & \dots & \rho \\ \rho & 1 & \rho & \dots & \rho \\ \rho & \rho & \rho & \dots & 1 \end{bmatrix},$$

sendo I_N a matriz identidade de ordem N e ρ uma constante conhecida.

Verifica-se (Graybill [1983]) que se

$$M = (a - b) I_n + b J_n,$$

J_n denotando a matriz $n \times n$ contendo o número 1 em todas as posições ($J_n = \mathbf{1}_n \mathbf{1}_n'$), então

$$M^{-1} = \frac{1}{a - b} \left[I_n - \frac{b}{a + (n-1)b} J_n \right].$$

Aplicando este fato para a matriz V , obtemos

$$V^{-1} = \frac{1}{1-\rho} \left[I_N - \frac{\rho}{1+(N-1)\rho} J_N \right],$$

$$X'V^{-1}X = \frac{1}{1-\rho} \mathbf{1}'_N \mathbf{1}_N - \frac{\rho}{(1-\rho)[1+(N-1)\rho]} \mathbf{1}'_N J_N \mathbf{1}_N$$

$$= \frac{N}{1-\rho} - \frac{\rho N^2}{(1-\rho)[1+(N-1)\rho]} = \frac{N}{1+(N-1)\rho}$$

$$X'V^{-1}Y = \frac{1}{1-\rho} \left[\mathbf{1}'_N Y - \frac{\rho}{1+(N-1)\rho} \mathbf{1}'_N J_N Y \right]$$

$$= \frac{1}{1-\rho} \left[N\bar{Y} - \frac{\rho}{1+(N-1)\rho} N^2\bar{Y} \right] = \frac{N\bar{Y}}{1+(N-1)\rho}$$

e

$$\hat{B}_N = (X'V^{-1}X)^{-1}X'V^{-1}Y = \bar{Y},$$

onde $\bar{Y} = \frac{\sum_{i=1}^N Y_i}{N}.$

De forma análoga temos que

$$V_s = (1-\rho)I_n + \rho J_n \implies V_s^{-1} = \frac{1}{1-\rho} \left[I_n - \frac{\rho}{1+(n-1)\rho} J_n \right],$$

$$\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s = \mathbf{1}'_n \mathbf{V}_s^{-1} \mathbf{1}_n = \frac{n}{1+(n-1)\rho} \quad ,$$

$$\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{y}_s = \frac{\sum_{i=1}^n y_i}{1+(n-1)\rho}$$

e assim

$$\hat{\beta} = (\mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{X}_s)^{-1} \mathbf{X}'_s \mathbf{V}_s^{-1} \mathbf{y}_s = \frac{\sum_{i=1}^n y_i}{n} = \bar{y}_s.$$

Desta forma, no modelo considerado, o previsor linear ótimo de $B_N = \bar{y}$, a média populacional, é $\hat{\beta} = \bar{y}_s$, a média amostral.

Neste caso, como $B_N = \frac{\mathbf{1}'_N \mathbf{y}}{N}$ é uma combinação linear de $\mathbf{y}$, seu previsor linear ótimo poderia ser obtido através do teorema A.2.1. No entanto, quando B_N é multidimensional, o previsor dado no teorema A.2.1 não pode ser utilizado.

□

Exemplo 2.2 (Modelo de Amostragem em Dois Estágios)

Vamos admitir que a população de tamanho N está dividida em H conglomerados, onde N_h é o tamanho do h -ésimo conglomerado, de modo que $\sum_{h=1}^H N_h = N$.

Indicando por y_{hj} a variável aleatória correspondendo à j -ésima observação do h -ésimo conglomerado, $j = 1, 2 \dots N_h$, $h = 1, 2 \dots H$, um modelo de superpopulação bastante comum é dado

por

$$E(Y_{hj}) = \mu$$

$$\text{Cov}(Y_{hj}, Y_{\ell m}) = \begin{cases} \sigma_h^2 + \sigma_v^2, & \text{se } h = \ell, \quad j = m \\ \sigma_v^2, & \text{se } h = \ell, \quad j \neq m \\ 0, & \text{se } h \neq \ell \end{cases}$$

$$j = 1, 2, \dots, N_h, \quad m = 1, 2, \dots, N_\ell, \quad h, \ell = 1, 2, \dots, H.$$

Neste modelo, unidades dentro do mesmo conglomerado são correlacionadas e unidades pertencentes a conglomerados distintos são não correlacionadas.

Usando a notação matricial, este modelo pode ser escrito na forma (2.1), com

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_H \end{bmatrix}, \quad X = \begin{bmatrix} 1_{N_1} \\ 1_{N_2} \\ \vdots \\ 1_{N_H} \end{bmatrix}, \quad \beta = [\mu],$$

$$V = \begin{bmatrix} V_1 & 0 & \dots & 0 \\ 0 & V_2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & V_H \end{bmatrix} = \text{diag}(V_1, V_2, \dots, V_H),$$

onde

$$Y_h = \begin{bmatrix} Y_{h1} \\ Y_{h2} \\ \vdots \\ Y_{hN_h} \end{bmatrix},$$

$$V_h = \begin{bmatrix} \sigma_h^2 + \sigma_v^2 & \sigma_v^2 & \dots & \sigma_v^2 \\ \sigma_v^2 & \sigma_h^2 + \sigma_v^2 & & \sigma_v^2 \\ \vdots & & \ddots & \vdots \\ \sigma_v^2 & \sigma_v^2 & & \sigma_h^2 + \sigma_v^2 \end{bmatrix}_{N_h \times N_h}, \quad h = 1, 2, \dots, H,$$

e 0 indica a matriz nula de dimensões apropriadas.

A matriz V_h pode ser escrita como

$$V_h = \sigma_h^2 I_{N_h} + \sigma_v^2 J_{N_h},$$

onde I_{N_h} denota a matriz identidade de ordem N_h e J_{N_h} é a matriz $N_h \times N_h$ contendo o número 1 em todas as posições.

Seleciona-se inicialmente uma amostra de $n \leq H$ conglomerados e indicaremos por s ao conjunto dos índices destes conglomerados selecionados. Num segundo estágio, uma amostra de tamanho m_h é obtida do h -ésimo conglomerado, para todo $h \in s$. O conjunto dos índices dos elementos selecionados no h -ésimo conglomerado será indicado por s_h .

Royall [1976a], utilizando este modelo, obtém através do teorema A.2.1 o previsor linear ótimo do total populacional,

$$T = \sum_{h=1}^H \sum_{j=1}^{N_h} Y_{hj} = \sum_{h \in s} \sum_{j \in s_h} Y_{hj} + \sum_{h \in s} \sum_{j \notin s_h} Y_{hj} + \sum_{h \notin s} \sum_{j=1}^{N_h} Y_{hj}.$$

Nosso objetivo será calcular o previsor linear ótimo de $B_{\sim N}$ no modelo de amostragem em dois estágios.

Sejam

J_k - a matriz $k \times k$, contendo 1 em todas as posições e

$J_{k,\ell}$ - a matriz $k \times \ell$, contendo 1 em todas as posições.

Nestas condições,

$$V_h^{-1} = \frac{1}{\sigma_h^2} \left[I_{N_h} - \frac{\sigma_v^2}{\sigma_h^2 + N_h \sigma_v^2} J_{N_h} \right],$$

$$X'V^{-1}X = \sum_{h=1}^H \mathbf{1}_{N_h}' V_h^{-1} \mathbf{1}_{N_h} = \sum_{h=1}^H \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2}$$

$$X'V^{-1}Y = \sum_{h=1}^H \mathbf{1}_{N_h}' V_h^{-1} Y_h = \sum_{h=1}^H \frac{N_h \bar{Y}_h}{\sigma_h^2 + N_h \sigma_v^2},$$

onde

$$\bar{Y}_h = \frac{1}{N_h} \sum_{j=1}^{N_h} Y_{hj},$$

e assim,

$$\hat{B}_N = (X'V^{-1}X)^{-1}X'V^{-1}Y = \frac{\sum_{h=1}^H \frac{N_h \bar{Y}_h}{\sigma_h^2 + N_h \sigma_v^2}}{\sum_{h=1}^H \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2}}.$$

Para o cálculo de $\hat{\beta}$, o previsor linear ótimo de B_N , observamos que V_s é a matriz bloco diagonal cujas matrizes da diagonal principal são V_{s_h} para $h \in s$ e V_{s_h} é a matriz de covariância do vetor formado pelas unidades amostradas no conglomerado h .

Desta forma, V_{s_h} será a matriz quadrada de ordem m_h , cujos elementos da diagonal principal são $\sigma_h^2 + \sigma_v^2$ e fora desta diagonal são iguais a σ_v^2 , podendo então ser escrita como

$$V_{s_h} = \sigma_h^2 I_{m_h} + \sigma_v^2 J_{m_h}.$$

Indicaremos V_s por

$$V_s = \text{diag}(V_{s_h}, h \in s) \text{ e portanto,}$$

$$V_s^{-1} = \text{diag}(V_{s_h}^{-1}, h \in s).$$

Além disso

$$V_{s_h}^{-1} = \frac{1}{\sigma_h^2} \left[\begin{array}{cc} I_{m_h} & - \frac{\sigma_v^2}{\sigma_h^2 + m_h \sigma_v^2} J_{m_h} \end{array} \right]$$

de modo que

$$X_s' V_s^{-1} X_s = \sum_{h \in s} \frac{1}{\sigma_h^2} \left[m_h - \frac{\sigma_v^2 m_h^2}{\sigma_h^2 + m_h \sigma_v^2} \right] = \sum_{h \in s} \frac{m_h}{\sigma_h^2 + m_h \sigma_v^2},$$

$$X_s' V_s^{-1} Y_s = \sum_{h \in s} \sum_{j \in s_h} \frac{Y_{hj}}{\sigma_h^2 + m_h \sigma_v^2}$$

e

$$\hat{\beta}_s = \frac{\sum_{h \in s} \sum_{j \in s_h} \frac{Y_{hj}}{\sigma_h^2 + m_h \sigma_v^2}}{\sum_{h \in s} \frac{m_h}{\sigma_h^2 + m_h \sigma_v^2}} = \frac{\sum_{h \in s} W_h \bar{Y}_{s_h}}{\sum_{h \in s} W_h},$$

onde $\bar{Y}_{s_h} = \frac{1}{m_h} \sum_{j \in s_h} Y_{hj}$ e $W_h = \frac{m_h \sigma_v^2}{\sigma_h^2 + m_h \sigma_v^2}$.

Aplicando-se a expressão (2.5), obtemos o erro quadrático médio generalizado de $\hat{\beta}$, dado por

$$EQMG(\hat{\beta}) = \lambda^2 \left[\left(\sum_{h \in s} \frac{m_h}{\sigma_h^2 + m_h \sigma_v^2} \right)^{-1} - \left(\sum_{h=1}^H \frac{N_h}{\sigma_h^2 + N_h \sigma_v^2} \right)^{-1} \right].$$

Quando $\sigma_h^2 = 0$, $\underline{B}_N$ se reduz a

$$\underline{B}_N = \frac{\sum_{h=1}^H \bar{y}_h}{H},$$

a média das médias dos H conglomerados, e seu previsor ótimo, $\hat{\underline{\beta}}$, fica

$$\hat{\underline{\beta}} = \frac{\sum_{h \in s} \bar{y}_{s_h}}{n}.$$

□

2.8 - Considerações Finais

Nos dois exemplos considerados, $\underline{B}_N$ era um escalar devido ao tipo de modelo que adotamos.

Na maioria dos casos, no entanto, $\underline{B}_N$ é uma quantidade vetorial. Os resultados presentes neste capítulo fornecem o previsor ótimo em qualquer uma das situações.

Se V , a matriz de covariância do modelo, for tal que $V \underline{1}_N = X \underline{\delta}$, $\underline{\delta}$ um vetor de constantes (esta condição é conhecida na literatura como condição L), então

$$\begin{aligned} \underline{1}_N' Y - \underline{1}_N' X \underline{B}_N &= \underline{1}_N' V V^{-1} [I_N - X (X' V^{-1} X)^{-1} X' V^{-1}] Y = \\ &= \underline{\delta}' X' V^{-1} [I_N - X (X' V^{-1} X)^{-1} X' V^{-1}] Y = \\ &= \underline{\delta}' [X' V^{-1} - X' V^{-1}] Y = 0. \end{aligned}$$

Verificamos então que, satisfeita a condição L, o total populacional $\underline{1}'\underline{y}$ é igual a $\underline{1}'\underline{X}\underline{B}_N$, ou seja, é uma combinação linear das componentes de $\underline{B}_N$.

Traçando um paralelo entre a previsão ótima de $\underline{B}_N$ e a previsão ótima de $\underline{\ell}'\underline{y}$ (descrita no apêndice 2), notamos muitas semelhanças e algumas diferenças.

Constatamos a coincidência entre o previsor linear ótimo de $\underline{B}_N$ no modelo (2.6) com o previsor ótimo sob o modelo (2.2), que admite normalidade, valendo também a coincidência com o previsor ótimo no modelo Bayesiano (2.7), $\underline{\mu}$ desconhecido. Fato idêntico ocorre na previsão de $\underline{\ell}'\underline{y}$.

Por outro lado, o previsor ótimo de $\underline{B}_N$, $\hat{\underline{B}}_N$, tem a vantagem de ser simples, o que nem sempre acontece com o previsor ótimo de $\underline{\ell}'\underline{y}$, que só se simplificava válidas condições adicionais.

Salientamos ainda que, curiosamente, o previsor ótimo de $\underline{B}_N$ independe de $\underline{V}_{sr}$, já que esta matriz, embora suposta conhecida, não é necessária no cálculo de $\hat{\underline{B}}_N$. O conhecimento de $\underline{V}_{sr}$ seria então necessário somente no cálculo do erro quadrático médio generalizado do previsor.

Para um trabalho futuro nesta linha, sugerimos a previsão de $\underline{B}_N$ no caso multivariado, ou seja, quando mais de uma variável dependente $\underline{Y}$ é observada no mesmo indivíduo.

Outra possível pesquisa na área seria um estudo semelhante ao da seção 2.3, obtendo o previsor ótimo de $\underline{B}_N$ sob outras distribuições de probabilidades que não a normal.

CAPÍTULO 3 - FUNÇÕES DE VEROSSIMILHANÇA PREDITIVAS E APLICAÇÕES A POPULAÇÕES FINITAS

3.1 - Introdução

Na inferência estatística clássica, modelos de probabilidade são adotados quando desejamos obter conclusões gerais a partir de um conjunto de dados observados.

Estas conclusões são, na maioria dos casos, referentes a uma ou mais quantidades de interesse, que seriam os parâmetros da distribuição de probabilidades especificada por tais modelos, ou ainda, dizem respeito a valores de variáveis aleatórias, cuja distribuição é definida, a menos de parâmetros desconhecidos, que podem ou não ser de interesse do pesquisador. (Quando tais distribuições de probabilidades não estão especificadas, métodos não paramétricos se fazem necessários).

A primeira situação, correspondente à inferência sobre os parâmetros de interesse, consiste num problema de estimação, e a segunda, a da obtenção de informação sobre valores de particulares variáveis aleatórias, seria enquadrada na área de previsão.

Dentre os vários métodos de estimação existentes, o método de máxima verossimilhança tem sido intensivamente estudado e sua forma intuitiva, aliada a suas propriedades, o tornam bastante atraente.

Muito menos no entanto tem sido estudado sobre o uso de funções de verossimilhança com o objetivo de previsão, mas este tópico vem ganhando substancial ênfase nos últimos anos, sendo abordado em trabalhos relativamente recentes.

Bjornstad [1990] faz um excelente levantamento sobre funções de verossimilhança preditivas, mostrando os diferentes tipos de funções propostas, apresentando exemplos, propriedades e uma grande quantidade de referências sobre trabalhos desenvolvidos no assunto.

No presente capítulo utilizamos esta teoria de verossimilhança preditiva na inferência em populações finitas e apresentamos vários resultados obtidos como consequência de tal utilização.

Estes resultados dizem respeito à previsão de $\underline{y}_r$, $\underline{\ell}'\underline{y}$ e $\underline{B}_N$, quantidades estas já definidas no capítulo anterior.

3.2 - Previsão em Populações Finitas através de "Profiles" Preditivas

Se $(\underline{y}_1, \underline{y}_2)$ é um vetor aleatório com função densidade de probabilidades $f_{\theta}(\underline{y}_1, \underline{y}_2)$, onde θ é um parâmetro desconhecido, $\underline{y}_1 = (y_1, y_2, \dots, y_n)$ corresponde aos dados observados e $\underline{y}_2 = (y'_1, y'_2, \dots, y'_m)$ é um conjunto de variáveis aleatórias não observadas, o problema de previsão consiste em obter informações sobre uma função qualquer de $\underline{y}_2$, $Z = h'(\underline{y}_2)$, uma vez observado o vetor $\underline{y}_1$.

Caso não haja um interesse específico na estimação de θ , este parâmetro é considerado "nuisance".

A teoria de verossimilhança preditiva admite que toda informação sobre (Z, θ) está contida na função de verossimilhança conjunta

$$\ell_{Y_1}(z, \theta) = q_{\theta}(Y_1, z),$$

onde $q_{\theta}(Y_1, z)$ é a densidade conjunta de Y_1 e Z , e a verossimilhança de Z será determinada eliminando-se θ da função ℓ_{Y_1} .

Qualquer função de verossimilhança assim obtida é denominada função de verossimilhança preditiva e diferentes métodos de eliminação do parâmetro "nuisance" dão origem a diferentes funções.

Dentre os vários tipos de funções de verossimilhança preditivas, uma das mais utilizadas é a chamada "profile" preditiva, definida como

$$L_p(z|Y_1) = \sup_{\theta \in \Theta} q_{\theta}(Y_1, z) = q_{\hat{\theta}_z}(Y_1, z),$$

sendo que $\hat{\theta}_z$ é o estimador de máxima verossimilhança de θ , calculado com base em (Y_1, z) , os dados combinados.

Como Y_1 é observado e Z não, Z vai funcionar como variável aleatória, Y_1 como parâmetro e $L_p(z|Y_1)$ pode ser vista como uma distribuição condicional de Z dado Y_1 .

Adicionalmente, verifica-se que funções de verossimilhança preditivas podem ser padronizadas de modo a se tornarem funções

densidade. Para efeito de maximização no entanto, é indiferente utilizarmos a função padronizada ou não.

Definição 3.1 - O previsor de máxima verossimilhança de Z , $\hat{Z}_{ml}$, é definido como o valor de z que maximiza $L_p(z|Y_1)$.

A teoria de verossimilhança preditiva, que acabamos de descrever brevemente, é destinada à previsão de quantidades aleatórias em populações infinitas. Nesta seção, vamos utilizá-la com o objetivo de prever quantidades desconhecidas associadas à população finita.

Existem poucos trabalhos sobre estimação de máxima verossimilhança em populações finitas e todos eles utilizando outros procedimentos. A previsão através de funções de verossimilhança preditivas é a principal contribuição deste capítulo.

O processo vai consistir basicamente em considerar Y_r como o vetor não observado Y_2 , tomar Z como Y_r e Y_1 como Y_s , o vetor observado.

Admitiremos o modelo de superpopulação da seção 2.3,

$$Y \sim N_N(X\beta, V) \quad (3.1),$$

onde $X = \begin{bmatrix} X_s \\ X_r \end{bmatrix}$, $Y = \begin{bmatrix} Y_s \\ Y_r \end{bmatrix}$ e $V = \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix}$ é uma matriz conhecida, não diagonal.

Nestas condições, a densidade conjunta de Y_s e Y_r é dada

por

$$q_{\beta}(Y_s, Y_r) = \frac{1}{(2\pi)^{N/2} |V|^{1/2}} \exp \left\{ -\frac{1}{2} (Y - X\beta)' V^{-1} (Y - X\beta) \right\}$$

e

$$L_p(Y_r | Y_s) = \sup_{\beta} q(Y_s, Y_r) = q_{\hat{\beta}_z}(Y_s, Y_r),$$

onde $\hat{\beta}_z$ é o estimador de máxima verossimilhança de β , calculado como função de (Y_s, Y_r) .

Sob normalidade, o estimador de máxima verossimilhança de β no modelo linear geral é o estimador de mínimos quadrados e portanto,

$$\hat{\beta}_z = (X'V^{-1}X)^{-1}X'V^{-1}Y,$$

coincidindo com o coeficiente de regressão B_N , objeto de nosso estudo no capítulo anterior.

Desta forma, a "profile" preditiva $L_p(Y_r | Y_s)$ fica

$$L_p(Y_r | Y_s) = \frac{1}{(2\pi)^{N/2} |V|^{1/2}} \exp \left\{ -\frac{1}{2} (Y - XB_N)' V^{-1} (Y - XB_N) \right\}. \quad (3.2)$$

3.3 - Previsão de Máxima Verossimilhança para Y_r

Dada a função de verossimilhança preditiva de Y_r , $L_p(Y_r | Y_s)$, o previsor de máxima verossimilhança de Y_r , que indicaremos por $\hat{Y}_{r, mv}$, é o valor de Y_r que maximiza $L_p(Y_r | Y_s)$ em (3.2).

Maximizar $L_p(\underline{y}_r | \underline{y}_s)$ (com respeito a $\underline{y}_r$) é equivalente a minimizar

$$\begin{aligned}
 S &= (\underline{y} - \underline{XB}_{\underline{N}})' V^{-1} (\underline{y} - \underline{XB}_{\underline{N}}) = \\
 &= \underline{y}' V^{-1} \underline{y} - \underline{y}' V^{-1} \underline{XB}_{\underline{N}} - \underline{B}_{\underline{N}}' \underline{X}' V^{-1} \underline{y} + \underline{B}_{\underline{N}}' \underline{X}' V^{-1} \underline{XB}_{\underline{N}} \\
 &= \underline{y}' V^{-1} \underline{y} - 2 \underline{y}' V^{-1} \underline{XB}_{\underline{N}} + \underline{y}' V^{-1} \underline{X} (\underline{X}' V^{-1} \underline{X})^{-1} \underline{X}' V^{-1} \underline{XB}_{\underline{N}} = \\
 &= \underline{y}' V^{-1} \underline{y} - \underline{y}' V^{-1} \underline{XB}_{\underline{N}}.
 \end{aligned}$$

Do lema 2.1, temos que

$$\begin{aligned}
 \underline{B}_{\underline{N}} &= H^{-1} B C^{-1} \underline{y}_s + H^{-1} D E^{-1} \underline{y}_r, \quad \text{onde} \\
 B &= \underline{X}'_s - \underline{X}'_r V_r^{-1} V_{rs}, \quad C = V_s - V_{sr} V_r^{-1} V_{rs}, \\
 D &= \underline{X}'_r - \underline{X}'_s V_s^{-1} V_{sr}, \quad E = V_r - V_{rs} V_s^{-1} V_{sr} \quad \text{e} \\
 H &= B C^{-1} \underline{X}_s + D E^{-1} \underline{X}_r = \underline{X}' V^{-1} \underline{X}.
 \end{aligned}$$

Observamos que C, E e H são matrizes simétricas e que

$$\begin{aligned}
 V^{-1} &= \begin{bmatrix} (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} & -V_s^{-1} V_{sr} (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} \\ -V_r^{-1} V_{rs} (V_s - V_{sr} V_r^{-1} V_{rs})^{-1} & (V_r - V_{rs} V_s^{-1} V_{sr})^{-1} \end{bmatrix} \\
 &= \begin{bmatrix} C^{-1} & -V_s^{-1} V_{sr} E^{-1} \\ -V_r^{-1} V_{rs} C^{-1} & E^{-1} \end{bmatrix}.
 \end{aligned}$$

Além disso, como V é simétrica, V^{-1} é simétrica, e assim,

$$V_s^{-1} V_{sr} E^{-1} = (C^{-1})' V_{sr} V_r^{-1} = C^{-1} V_{sr} V_r^{-1}. \quad (3.3)$$

Portanto,

$$Y' V^{-1} Y = Y_s' C^{-1} Y_s - Y_r' V_r^{-1} V_{rs} C^{-1} Y_s - Y_s' V_s^{-1} V_{sr} E^{-1} Y_r + Y_r' E^{-1} Y_r =$$

$$Y_s' C^{-1} Y_s - 2 Y_s' V_s^{-1} V_{sr} E^{-1} Y_r + Y_r' E^{-1} Y_r,$$

$$Y' V^{-1} X B_N = B_N' X' V^{-1} X B_N =$$

$$= \left[Y_s' (C^{-1})' B' (H^{-1})' + Y_r' (E^{-1})' D' (H^{-1})' \right] H \left[H^{-1} B C^{-1} Y_s + H^{-1} D E^{-1} Y_r \right] =$$

$$= Y_s' C^{-1} B' H^{-1} B C^{-1} Y_s + Y_s' C^{-1} B' H^{-1} D E^{-1} Y_r + Y_r' E^{-1} D' H^{-1} B C^{-1} Y_s +$$

$$+ Y_r' E^{-1} D' H^{-1} D E^{-1} Y_r = Y_s' C^{-1} B H^{-1} B C^{-1} Y_s + 2 Y_s' C^{-1} B H^{-1} D E^{-1} Y_r +$$

$$+ Y_r' E^{-1} D' H^{-1} D E^{-1} Y_r$$

e

$$S = Y_s' C^{-1} Y_s - 2 Y_s' V_s^{-1} V_{sr} E^{-1} Y_r + Y_r' E^{-1} Y_r - Y_s' C^{-1} B' H^{-1} B C^{-1} Y_s$$

$$- 2 Y_s' C^{-1} B' H^{-1} D E^{-1} Y_r - Y_r' E^{-1} D' H^{-1} D E^{-1} Y_r.$$

$$\text{Se } f(\underline{x}) = \underline{a}' \underline{x}, \text{ então } \frac{\partial f(\underline{x})}{\partial \underline{x}} = \underline{a} \text{ e se } f(\underline{x}) = \underline{x}' A \underline{x}, \frac{\partial f(\underline{x})}{\partial \underline{x}} = 2 A \underline{x}.$$

Utilizando estes dois resultados, obtemos

$$\frac{\partial S}{\partial \underline{y}_r} = -2\underline{V}_r^{-1}\underline{V}_{rs}C^{-1}\underline{y}_s - 2E^{-1}D'H^{-1}BC^{-1}\underline{y}_s + 2E^{-1}\underline{y}_r - 2E^{-1}D'H^{-1}DE^{-1}\underline{y}_r.$$

O previsor de máxima verossimilhança de $\underline{y}_r$, $\hat{\underline{y}}_{r,mv}$, é a solução em $\underline{y}_r$ da equação $\frac{\partial S}{\partial \underline{y}_r} = 0$, ou seja, $\hat{\underline{y}}_{r,mv}$ é tal que

$$[\underline{V}_r^{-1}\underline{V}_{rs}C^{-1} + E^{-1}D'H^{-1}BC^{-1}]\underline{y}_s = [E^{-1} - E^{-1}D'H^{-1}DE^{-1}]\hat{\underline{y}}_{r,mv}. \quad (3.4)$$

Do apêndice 2 e capítulo 2, temos que os previsores ótimos para $\underline{\ell}'\underline{y}$ e $\underline{B}_N$ sob o modelo (3.1) são respectivamente

$$\hat{\underline{t}}^* = \underline{\ell}'_s\underline{y}_s + \underline{\ell}'_r[\underline{V}_{rs}\underline{V}_s^{-1}\underline{y}_s + (\underline{X}_r - \underline{V}_{rs}\underline{V}_s^{-1}\underline{X}_s)\hat{\underline{\beta}}]$$

e

$$\hat{\underline{\theta}}^* = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}[\underline{V}_{rs}\underline{V}_s^{-1}\underline{y}_s + (\underline{X}_r - \underline{V}_{rs}\underline{V}_s^{-1}\underline{X}_s)\hat{\underline{\beta}}] = \hat{\underline{\beta}}.$$

Lembrando que $\underline{\ell}'\underline{y} = \underline{\ell}'_s\underline{y}_s + \underline{\ell}'_r\underline{y}_r$ e $\underline{B}_N = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}\underline{y}_r$, observamos que em ambos os casos a parte não observada $\underline{y}_r$ está sendo prevista por

$$\hat{\underline{y}}_r = \underline{V}_{rs}\underline{V}_s^{-1}\underline{y}_s + (\underline{X}_r - \underline{V}_{rs}\underline{V}_s^{-1}\underline{X}_s)\hat{\underline{\beta}} = \underline{V}_{rs}\underline{V}_s^{-1}\underline{y}_s + D'\hat{\underline{\beta}} \quad (3.5)$$

onde

$$\hat{\underline{\beta}} = (\underline{X}'_s\underline{V}_s^{-1}\underline{X}_s)^{-1}\underline{X}'_s\underline{V}_s^{-1}\underline{y}_s.$$

Gostaríamos de, além de resolver a equação (3.4), obtendo o previsor de máxima verossimilhança de $\underline{y}_r$, verificar se esta solução coincide com a expressão (3.5).

O próximo teorema apresenta a solução de (3.4), mostrando que esta solução é realmente da forma (3.5).

Teorema 3.1

O vetor $\hat{y}_r = V_{rs} V_s^{-1} y_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta} = V_{rs} V_s^{-1} y_s + D' \hat{\beta}$ é solução de (3.4).

Prova:

A equação (3.4) pode ser reescrita como

$$V_r^{-1} V_{rs} C^{-1} y_s + E^{-1} D' [H^{-1} B C^{-1} y_s + H^{-1} D E^{-1} \hat{y}_{r,mv}] = E^{-1} \hat{y}_{r,mv}$$

e utilizando (3.3) esta equação fica

$$\begin{aligned} E^{-1} V_{rs} V_s^{-1} y_s + E^{-1} D' [H^{-1} B C^{-1} y_s + H^{-1} D E^{-1} \hat{y}_{r,mv}] &= E^{-1} \hat{y}_{r,mv} \\ \Leftrightarrow V_{rs} V_s^{-1} y_s + D' [H^{-1} B C^{-1} y_s + H^{-1} D E^{-1} \hat{y}_{r,mv}] &= \hat{y}_{r,mv}. \end{aligned} \quad (3.6)$$

Substituindo-se $\hat{y}_{r,mv}$ por $\hat{y}_r = V_{rs} V_s^{-1} y_s + D' \hat{\beta}$ no lado esquerdo de (3.6) obtemos

$$V_{rs} V_s^{-1} y_s + D' [H^{-1} B C^{-1} y_s + H^{-1} D E^{-1} (V_{rs} V_s^{-1} y_s + D' \hat{\beta})] = V_{rs} V_s^{-1} y_s + D' \hat{\beta}$$

e substituindo-se no lado direito fica

$$V_{rs} V_s^{-1} y_s + D' \hat{\beta},$$

mostrando assim que $\hat{y}_r = V_{rs} V_s^{-1} y_s + D' \hat{\beta}$ verifica (3.6), sendo portanto solução de (3.4). □

Salientamos que se $(E^{-1} - E^{-1}D'H^{-1}DE^{-1})$ for não singular, (3.4) terá solução única. O teorema 3.1 mostra que o preditor de máxima verossimilhança de y_r será $V_{rs}V_s^{-1}y_s + D'\hat{\beta}$, o mesmo preditor de y_r que originou os preditores ótimos $\hat{T}^*$ e $\hat{\theta}^*$.

É importante lembrarmos que nem sempre estimadores de máxima verossimilhança coincidem com os estimadores ótimos (se $X \sim N(\mu, \sigma^2)$ com μ e σ^2 desconhecidos, o estimador de máxima verossimilhança de σ^2 , $\hat{\sigma}^2 = \frac{\sum_{i=1}^n (X_i - \bar{X})^2}{n}$, é viciado) e assim, o resultado presente no teorema 3.1 é bem mais do que estávamos esperando.

3.4 - A Função de Verossimilhança Preditiva para y_r

Na seção 3.2 obtivemos a função de verossimilhança preditiva de y_r e na seção 3.3 esta função foi utilizada na determinação de $\hat{y}_{r,mv}$, o preditor de máxima verossimilhança de y_r . No entanto, pouca ênfase foi dada à função propriamente dita.

Nesta seção, vamos escrever $L_p(y_r|y_s)$ numa forma alternativa, de modo a identificá-la com uma densidade conhecida.

Usaremos a notação $L_p(y_2|y_1) \propto N_m(\underline{\mu}, \Sigma)$ para indicar que

$$L_p(y_2|y_1) = K(y_1)h(y_2, \underline{\mu}, \Sigma),$$

onde

$h(y_2, \underline{\mu}, \Sigma)$ é a função densidade da distribuição normal

multivariada com média $\underline{\mu}$ e matriz de covariância Σ e $K(\underline{y}_1)$ é uma constante que não depende de $\underline{y}_2$.

Teorema 3.2

$$L_p(\underline{y}_r | \underline{y}_s) \propto N_{N-n}(\hat{\underline{y}}_r, W^{-1}),$$

onde

$$W = E^{-1} - E^{-1}D'H^{-1}DE^{-1} \quad \text{e} \quad \hat{\underline{y}}_r = V_{rs}V_s^{-1}\underline{y}_s + D'\hat{\underline{\beta}} = \hat{\underline{y}}_{r,mv}.$$

Prova:

Da expressão (3.2)

$$L_p(\underline{y}_r | \underline{y}_s) = \frac{1}{(2\pi)^{N/2} |V|^{1/2}} \exp \left\{ \frac{-1}{2} S \right\},$$

onde $S = (\underline{y} - X\tilde{B}_N)'V^{-1}(\underline{y} - X\tilde{B}_N)$.

$$\begin{aligned} S &= \underline{y}_s' C^{-1} \underline{y}_s - 2\underline{y}_s' V_s^{-1} V_{sr} E^{-1} \underline{y}_r + \underline{y}_r' E^{-1} \underline{y}_r - \underline{y}_s' C^{-1} B' H^{-1} B C^{-1} \underline{y}_s \\ &\quad - 2\underline{y}_s' C^{-1} B' H^{-1} D E^{-1} \underline{y}_r - \underline{y}_r' E^{-1} D' H^{-1} D E^{-1} \underline{y}_r = \\ &= \underline{y}_s' [C^{-1} - C^{-1} B' H^{-1} B C^{-1}] \underline{y}_s - 2\underline{y}_r' [E^{-1} V_{rs} V_s^{-1} + E^{-1} D' H^{-1} B C^{-1}] \underline{y}_s + \\ &\quad \underline{y}_r' [E^{-1} - E^{-1} D' H^{-1} D E^{-1}] \underline{y}_r = \underline{y}_s' [C^{-1} - C^{-1} B' H^{-1} B C^{-1}] \underline{y}_s \\ &\quad - 2\underline{y}_r' [V_{rs}^{-1} V_{rs} C^{-1} + E^{-1} D' H^{-1} B C^{-1}] \underline{y}_s + \underline{y}_r' W \underline{y}_r. \end{aligned}$$

Pelo teorema 3.1

$$[V_r^{-1} V_{rs} C^{-1} + E^{-1} D' H^{-1} B C^{-1}] Y_s = [E^{-1} - E^{-1} D' H^{-1} D E^{-1}] \hat{Y}_r$$

e assim,

$$\begin{aligned} S &= Y_s' [C^{-1} - C^{-1} B' H^{-1} B C^{-1}] Y_s - 2 Y_r' W \hat{Y}_r + Y_r' W Y_r \\ &= (Y_r - \hat{Y}_r)' W (Y_r - \hat{Y}_r) - \hat{Y}_r' W \hat{Y}_r + Y_s' [C^{-1} - C^{-1} B' H^{-1} B C^{-1}] Y_s. \end{aligned}$$

Como a expressão $S_1 = Y_s' [C^{-1} - C^{-1} B' H^{-1} B C^{-1}] Y_s - \hat{Y}_r' W \hat{Y}_r$ não depende de Y_r , segue que

$$\begin{aligned} L_p &= (Y_r | Y_s) = \\ &= \frac{1}{(2\pi)^{N/2} |V|^{1/2}} \exp \left\{ -\frac{1}{2} S_1 \right\} \exp \left\{ -\frac{1}{2} (Y_r - \hat{Y}_r)' W (Y_r - \hat{Y}_r) \right\} \\ &\propto N_{N-n}(\hat{Y}_r, W^{-1}), \text{ com } \hat{Y}_r = V_{rs} V_s^{-1} Y_s + D' \hat{\beta}. \end{aligned}$$

□

Observamos que a moda de $L_p(Y_r | Y_s)$, sendo o previsor de máxima verossimilhança de Y_r , coincide com o valor $V_{rs} V_s^{-1} Y_s + D' \hat{\beta}$, obtido no teorema 3.1.

O teorema 3.2 mostra que a informação sobre Y_r , dado Y_s , está centrada em $\hat{Y}_r$.

3.5 - Previsão de Máxima Verossimilhança para $\tilde{\ell}' Y$ e $\tilde{B}_N$

No apêndice 2 apresentamos o previsor ótimo de $\tilde{\ell}' Y$,

$$\hat{T}^* = \tilde{\ell}'_s Y_s + \tilde{\ell}'_r [V_{rs} V_s^{-1} Y_s + (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}] = \tilde{\ell}'_s Y_s + \tilde{\ell}'_r \hat{Y}_r$$

e no capítulo 2, seção 2.3, obtivemos o previsor ótimo do vetor aleatório $\underline{B}_N$,

$$\hat{\underline{\theta}}^* = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}\underline{y}_s + D'\hat{\underline{\beta}}] = \hat{\underline{\beta}},$$

quando o modelo é $\underline{y} \sim N_N(X\underline{\beta}, V)$.

Nesta seção, sob o mesmo modelo, determinaremos as funções de verossimilhança preditivas e os respectivos previsores de máxima verossimilhança para $\underline{\ell}'\underline{y}$ e $\underline{B}_N$, a partir de $L_p(\underline{y}_r|\underline{y}_s)$.

Através do teorema 3.2, verificamos que $L_p(\underline{y}_r|\underline{y}_s)$ é proporcional a uma densidade de probabilidades.

A partir daí, as funções de verossimilhança preditivas $L_p(\underline{\ell}'\underline{y}|\underline{y}_s)$ e $L_p(\underline{B}_N|\underline{y}_s)$ podem ser obtidas facilmente devido ao princípio da substituição para distribuições condicionais (James [1981]).

Este princípio estabelece que a distribuição condicional de $(\underline{\ell}'\underline{y} = \underline{\ell}'_s\underline{y}_s + \underline{\ell}'_r\underline{y}_r)|\underline{y}_s$ é a mesma que a distribuição de $(\underline{\ell}'_r\underline{y}_r|\underline{y}_s) + \underline{\ell}'_s\underline{y}_s$.

Lembrando que $L_p(\underline{y}_r|\underline{y}_s)$ pode ser encarada como uma distribuição de probabilidades condicional de $\underline{y}_r$ dado $\underline{y}_s$, vamos aplicar o princípio da substituição e daí determinaremos as "profiles" preditivas para $\underline{\ell}'\underline{y}$ e $\underline{B}_N$ nos próximos teoremas.

Teorema 3.3

No modelo $\underline{y} \sim N_N(X\underline{\beta}, V)$,

$$L_p(\underline{\ell}'\underline{y}|\underline{y}_s) \propto N_1(\hat{\underline{\tau}}^*, \text{Var}(\hat{\underline{\tau}}^* - \underline{\ell}'\underline{y})) \quad (3.7)$$

e

$$\text{var}(\hat{T}^* - \underline{\ell}'\underline{y}) = \underline{\ell}'_r W^{-1} \underline{\ell}_r,$$

onde $\hat{T}^* = \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r \hat{\underline{y}}_r$ é o previsor ótimo de $\underline{\ell}'\underline{y}$ e

$$W = E^{-1} - E^{-1}D'H^{-1}DE^{-1}.$$

Prova:

Como $L_p(\underline{y}_r | \underline{y}_s) \propto N_{N-n}(\hat{\underline{y}}_r, W^{-1})$ e $\underline{\ell}'\underline{y} = \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r \underline{y}_r$, devido ao princípio da substituição para distribuições condicionais e às propriedades da distribuição normal multivariada, segue que

$$L_p(\underline{\ell}'\underline{y} | \underline{y}_s) \propto N_1(\underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r \hat{\underline{y}}_r, \underline{\ell}'_r W^{-1} \underline{\ell}_r)$$

Da seção A.2.2 temos que

$$\begin{aligned} \text{EQM}(\hat{T}^*) &= E(\hat{T}^* - \underline{\ell}'\underline{y})^2 = \text{Var}(\hat{T}^* - \underline{\ell}'\underline{y}) = \\ &= \underline{\ell}'_r (V_r - V_{rs} V_s^{-1} V_{sr}) \underline{\ell}_r + \underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) (X_s' V_s^{-1} X_s)^{-1} (X_s' - X_s' V_s^{-1} V_{sr}) \underline{\ell}_r \\ &= \underline{\ell}'_r E \underline{\ell}_r + \underline{\ell}'_r D' (X_s' V_s^{-1} X_s)^{-1} D \underline{\ell}_r \end{aligned}$$

e assim, para provar que $\text{Var}(\hat{T}^* - \underline{\ell}'\underline{y}) = \underline{\ell}'_r W^{-1} \underline{\ell}_r$, é suficiente verificarmos que

$$W^{-1} = E + D' (X_s' V_s^{-1} X_s)^{-1} D.$$

Como $W = E^{-1} - E^{-1}D'H^{-1}DE^{-1}$, aplicando o resultado (A.2.10) do apêndice 2, temos

$$W^{-1} = E - EE^{-1}D'(DE^{-1}EE^{-1}D' - H)^{-1}DE^{-1}E = E + D'(H - DE^{-1}D')^{-1}D.$$

Além disso, $E = V_r - V_{rs} V_s^{-1} V_{sr} \Rightarrow$

$$\begin{aligned} E^{-1} &= V_r^{-1} - V_r^{-1} V_{rs} (V_{sr} V_r^{-1} V_{rs} - V_s)^{-1} V_{sr} V_r^{-1} \\ &= V_r^{-1} + V_r^{-1} V_{rs} C^{-1} V_{sr} V_r^{-1} = V_r^{-1} + V_r^{-1} V_{rs} V_s^{-1} V_{sr} E^{-1}, \end{aligned}$$

sendo a última igualdade decorrente de (3.3).

Segue então que

$$E^{-1} - V_r^{-1} V_{rs} V_s^{-1} V_{sr} E^{-1} = V_r^{-1}. \quad (3.8)$$

Por outro lado, $C = V_s - V_{sr} V_r^{-1} V_{rs} \Rightarrow$

$$C^{-1} = V_s^{-1} - V_s^{-1} V_{sr} (V_{rs} V_s^{-1} V_{sr} - V_r)^{-1} V_{rs} V_s^{-1} = V_s^{-1} + V_s^{-1} V_{sr} E^{-1} V_{rs} V_s^{-1}$$

e desta forma,

$$\begin{aligned} H - DE^{-1}D' &= BC^{-1}X_s + DE^{-1}X_r - DE^{-1}(X_r - V_{rs} V_s^{-1} X_s) = \\ &= BC^{-1}X_s + DE^{-1}V_{rs} V_s^{-1} X_s = \\ &= (X'_s - X'_r V_r^{-1} V_{rs}) (V_s^{-1} + V_s^{-1} V_{sr} E^{-1} V_{rs} V_s^{-1}) X_s + DE^{-1}V_{rs} V_s^{-1} X_s \\ &= X'_s V_s^{-1} X_s + (X'_s V_s^{-1} V_{sr} - X'_r V_r^{-1} V_{rs} V_s^{-1} V_{sr} + D) E^{-1} V_{rs} V_s^{-1} X_s \\ &\quad - X'_r V_r^{-1} V_{rs} V_s^{-1} X_s. \end{aligned}$$

Substituindo D por $X'_r - X'_s V_s^{-1} V_{sr}$ e utilizando (3.8),
obtemos

$$\begin{aligned}
H - DE^{-1}D' &= X_s' V_s^{-1} X_s + X_r' (E^{-1} - V_r^{-1} V_{rs} V_s^{-1} V_{sr} E^{-1}) V_{rs} V_s^{-1} X_s \\
&\quad - X_r' V_r^{-1} V_{rs} V_s^{-1} X_s = \\
&= X_s' V_s^{-1} X_s + X_r' V_r^{-1} V_{rs} V_s^{-1} X_s - X_r' V_r^{-1} V_{rs} V_s^{-1} X_s = X_s' V_s^{-1} X_s,
\end{aligned}$$

o que encerra a demonstração. $\square$

Definição 3.2 - Se $\tilde{y}$ é a quantidade linear a ser prevista, dizemos que V , a matriz de covariância do modelo de superpopulação, satisfaz à condição $\bar{L}$ se

$$K\tilde{\ell} = X_s\tilde{\delta} \quad \text{para algum } \tilde{\delta} \in \mathbb{R}^{p+1}, \text{ onde } K = \begin{bmatrix} V_s & V_{sr} \end{bmatrix}.$$

Corolário 3.3

Se $\bar{L}$ estiver satisfeita, ou se $V\tilde{\ell} = X_s\tilde{\delta}$ (condição L), então

$$L_p(\tilde{\ell}'\tilde{y} | \tilde{y}_s) \propto N_1(\tilde{\ell}'\hat{X}\hat{\beta}, \tilde{\ell}'W^{-1}\tilde{\ell}).$$

Prova:

$$V\tilde{\ell} = X_s\tilde{\delta} \quad \Rightarrow \quad \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix} \tilde{\ell} = \begin{bmatrix} X_s \\ X_r \end{bmatrix} \tilde{\delta} \quad \Rightarrow$$

$$K\tilde{\ell} = X_s\tilde{\delta} \Rightarrow \bar{L} \text{ está satisfeita.}$$

Bolfarine e Rodrigues [1988] demostram que $\hat{T}^* = \tilde{\ell}'\hat{X}\hat{\beta}$ se e somente se $\bar{L}$ está satisfeita.

A partir daí, o resultado sai como consequência direta do teorema 3.3. $\square$

Os resultados presentes no teorema e corolário 3.3 são importantes, porque se constituem numa versão para matriz de covariância não diagonal do teorema apresentado em Royall [1976b].

Utilizando o modelo (3.1) com V diagonal, o autor verifica que, válida a condição L , a função de verossimilhança de $\underline{\ell}'\underline{y}$ dado $\underline{y}_s$ é proporcional a $N(\hat{t}_o, \text{Var}(\hat{t}_o - \underline{\ell}'\underline{y}))$,

onde

$\hat{t}_o = \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r \underline{X}_r \hat{\beta}$ é o previsor ótimo de $\underline{\ell}'\underline{y}$ quando V é diagonal (expressão (A.2.4), apêndice 2) e

$$\begin{aligned} \text{Var}(\hat{t}_o - \underline{\ell}'\underline{y}) &= E(\hat{t}_o - \underline{\ell}'\underline{y})^2 = \text{EQM}(\hat{t}_o) \\ &= \underline{\ell}'_r \underline{V}_r \underline{\ell}_r + \underline{\ell}'_r \underline{X}_r (\underline{X}'_s \underline{V}_s^{-1} \underline{X}_s)^{-1} \underline{X}'_r \underline{\ell}_r \end{aligned}$$

(expressão (A.2.5), apêndice 2).

A função de verossimilhança utilizada, $f(\underline{\ell}'\underline{y}|\underline{y}_s)$, foi obtida diretamente através de condicionamento como

$$f(\underline{\ell}'\underline{y}|\underline{y}_s) \propto f(\underline{y}_s|\underline{\ell}'\underline{y}) \cdot f(\underline{\ell}'\underline{y})$$

onde

$f(\underline{y}_s|\underline{\ell}'\underline{y})$ é a densidade condicional de $\underline{y}_s$ dado $\underline{\ell}'\underline{y}$ e $f(\underline{\ell}'\underline{y})$ é a densidade de $\underline{\ell}'\underline{y}$.

O corolário 3.3 mostra que vale um resultado análogo quando utilizamos função de verossimilhança preditiva e V é não diagonal.

Observamos ainda que se V é diagonal, $L_p(\underline{\ell}'\underline{y}|\underline{y}_s)$ dada em (3.7) coincide com $f(\underline{\ell}'\underline{y}|\underline{y}_s)$.

A coincidência entre estas funções no caso diagonal nos leva a acreditar numa possível coincidência também para o caso não diagonal.

Lembrando que $f(\tilde{\ell}'\tilde{y}|\tilde{y}_s)$ só foi calculada para V diagonal, fica como sugestão para um trabalho futuro o problema da determinação de $f(\tilde{\ell}'\tilde{y}|\tilde{y}_s)$ para V qualquer e ainda o de verificar se a função obtida coincide com $L_p(\tilde{\ell}'\tilde{y}|\tilde{y}_s)$ definida em (3.7).

Também quanto à previsão de $\tilde{B}_N$ através de funções de verossimilhança preditivas, nenhum trabalho anterior é de nosso conhecimento.

O próximo teorema apresenta a "profile" preditiva de $\tilde{B}_N$.

Teorema 3.4

$$L_p(\tilde{B}_N|\tilde{y}_s) \propto N_{p+1}(\hat{\tilde{\beta}}, \text{Var}(\hat{\tilde{\beta}} - \tilde{B}_N))$$

onde

$$\text{Var}(\hat{\tilde{\beta}} - \tilde{B}_N) = H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1} \quad \text{e} \quad W = E^{-1} - E^{-1}D'H^{-1}DE^{-1}.$$

Prova:

$$L_p(\tilde{y}_r|\tilde{y}_s) \propto N_{N-n}(\hat{\tilde{y}}_r, W^{-1})$$

e como $\tilde{B}_N = H^{-1}BC^{-1}\tilde{y}_s + H^{-1}DE^{-1}\tilde{y}_r$, temos que

$$L_p(\tilde{B}_N|\tilde{y}_s) \propto N_{p+1}(H^{-1}BC^{-1}\tilde{y}_s + H^{-1}DE^{-1}\hat{\tilde{y}}_r, H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1}).$$

$$\text{Além disso, } H^{-1}BC^{-1}\tilde{y}_s + H^{-1}DE^{-1}\hat{\tilde{y}}_r = \hat{\tilde{\beta}}.$$

Resta provar que

$$\text{Var}(\hat{\beta} - \underline{B}_N) = H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1}.$$

Da expressão (2.4), seção 2.3,

$$\begin{aligned} \text{EQMG}(\hat{\beta}) &= E[\lambda'(\hat{\beta} - \underline{B}_N)(\hat{\beta} - \underline{B}_N)'\lambda] = \lambda'\text{Var}(\hat{\beta} - \underline{B}_N)\lambda \\ &= \lambda'H^{-1}DE^{-1}[E + D'(X_s'V_s^{-1}X_s)^{-1}D]E^{-1}D'H^{-1}\lambda. \end{aligned}$$

Mas, conforme verificado na prova do teorema 3.3,

$$W^{-1} = E + D'(X_s'V_s^{-1}X_s)^{-1}D,$$

o que implica em

$$\text{Var}(\hat{\beta} - \underline{B}_N) = H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1}.$$

□

A partir de $L_p(\underline{\ell}'Y|Y_s)$ obtemos o previsor de máxima verossimilhança de $\underline{\ell}'Y$, $\hat{T}_{mv}^*$, que é a moda da função densidade expressa em (3.7),

$$\begin{aligned} \hat{T}_{mv}^* &= \underline{\ell}'_s Y_s + \underline{\ell}'_r \hat{Y}_r = \\ &= \underline{\ell}'_s Y_s + \underline{\ell}'_r [V_{rs}V_s^{-1}Y_s + (X_r - V_{rs}V_s^{-1}X_s)\hat{\beta}]. \end{aligned}$$

Poderíamos também obter intervalos de previsão para $\underline{\ell}'Y$ e isto será exemplificado posteriormente.

Como consequência do teorema 3.4, o previsor de máxima verossimilhança de $\underline{B}_N$, que indicaremos por $\hat{\theta}_{mv}^*$, será então

$$\hat{\theta}_{mv}^* = H^{-1}BC^{-1}Y_s + H^{-1}DE^{-1}[V_{rs}V_s^{-1}Y_s + D'\hat{\beta}] = \hat{\beta}.$$

É importante observarmos que nos dois casos foi satisfeita a conhecida propriedade da invariância, ou seja, verificamos que

$$\underline{\ell}'\underline{y} = \underline{\ell}'_s\underline{y}_s + \underline{\ell}'_r\underline{y}_r \Rightarrow \hat{T}_{mv}^* = \underline{\ell}'_s\underline{y}_s + \underline{\ell}'_r\hat{\underline{y}}_{r,mv}$$

e

$$\underline{B}_N = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}\underline{y}_r \Rightarrow \hat{\underline{\theta}}_{mv}^* = H^{-1}BC^{-1}\underline{y}_s + H^{-1}DE^{-1}\hat{\underline{y}}_{r,mv}.$$

Observamos ainda a coincidência entre os previsores de máxima verossimilhança obtidos nesta seção com os correspondentes previsores ótimos para $\underline{\ell}'\underline{y}$ e $\underline{B}_N$.

Novamente salientamos a importância deste fato, dado que nem sempre estimadores de máxima verossimilhança são ótimos.

3.6 - Exemplos

Obtida a função de verossimilhança preditiva para a quantidade de interesse Z e observado $\underline{x}$, Lejeune e Faulkenberry [1982] definem intervalos de previsão de nível $1 - \alpha$ para Z como

$$\left[\hat{z}_{\alpha/2}(\underline{x}), \hat{z}_{1-\alpha/2}(\underline{x}) \right],$$

onde $\hat{z}_{\alpha/2}(\underline{x})$ e $\hat{z}_{1-\alpha/2}(\underline{x})$ são os quantis de ordem $\alpha/2$ e $1 - \frac{\alpha}{2}$ da distribuição $L_p(z|\underline{x})$, desde que padronizada.

Bjornstad [1990] utiliza uma outra definição de intervalo de previsão com base em $L_p(z|\underline{x})$, também padronizada. De posse dessa função, a região de previsão para Z será

$$RP = \{z \mid L_p(z|\underline{x}) \geq k_\alpha\}$$

onde k_α é tal que

$$\int_{RP} L_p(z|\underline{x}) \, dz = 1-\alpha \quad \text{no caso contínuo}$$

ou

$$\sum_{z \in RP} L_p(z|\underline{x}) = 1-\alpha \quad \text{no caso discreto.}$$

Em nossos exemplos, a "profile" preditiva é proporcional à densidade de uma distribuição normal e neste caso, as duas definições de região de previsão ficam equivalentes.

Nesta seção, apresentaremos exemplos que ilustrarão alguns dos resultados obtidos. Nestes exemplos, trabalhando com particulares modelos de superpopulação, obteremos $L_p(\underline{B}_N|\underline{y}_s)$, o preditor de máxima verossimilhança de $\underline{B}_N$ e o correspondente intervalo de previsão para $\underline{B}_N$.

Exemplo 3.1

Consideremos o modelo de superpopulação

$$y_i = \mu + \varepsilon_i, \quad i = 1, 2, \dots, N, \text{ com}$$

$$\varepsilon_i \sim N(0, 1)$$

ε_i e ε_j variáveis aleatórias independentes, $i \neq j$, $i, j = 1, 2, \dots, N$.

Este modelo pode ser escrito na forma matricial como

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \text{ onde}$$

$$X = \underline{1}_N, \quad Y \sim N_N(X\beta, V) \quad \text{e} \quad V = I_N.$$

Verifica-se que

$$B_N = (X'V^{-1}X)^{-1}X'V^{-1}Y = \frac{1}{N} \sum_{i=1}^N y_i = \bar{y}$$

e

$$\hat{\beta} = (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} Y_s = \bar{y}_s = \frac{\sum_{i=1}^n y_i}{n}.$$

Temos

$$H^{-1} = (X'V^{-1}X)^{-1} = \frac{1}{N},$$

$$D = X'_r - X'_s V_s^{-1} V_{sr} = \underline{1}'_{N-n},$$

$$E = V_r - V_{rs} V_s^{-1} V_{sr} = I_{N-n},$$

$$\begin{aligned} W &= E^{-1} - E^{-1} D' H^{-1} D E^{-1} = I_{N-n} - \frac{1}{N} I_{N-n} \underline{1}_{N-n} \underline{1}'_{N-n} I_{N-n} = \\ &= I_{N-n} - \frac{1}{N} J_{N-n} \end{aligned}$$

e

$$W^{-1} = I_{N-n} + \frac{1}{n} J_{N-n}.$$

Com isto,

$$\begin{aligned} H^{-1} D E^{-1} W^{-1} E^{-1} D' H^{-1} &= \frac{1}{N^2} \underline{1}'_{N-n} I_{N-n} \left[I_{N-n} + \frac{1}{n} J_{N-n} \right] I_{N-n} \underline{1}_{N-n} \\ &= \frac{N-n}{N^2} + \frac{(N-n)^2}{N^2 n} = \frac{N-n}{Nn}, \end{aligned}$$

e pelo teorema 3.4,

$$L_p(\bar{y}_s | \bar{y}_s) = L_p(\bar{y} | \bar{y}_s) \propto N_1\left(\bar{y}_s, \frac{N-n}{Nn}\right).$$

Assim, sob este modelo, o previsor de máxima verossimilhança da média populacional $\bar{y}$ é a média amostral $\bar{y}_s$.

O valor $\frac{N-n}{Nn}$, de acordo com o teorema 3.4, é o erro quadrático médio de $\bar{y}_s$. Realmente, temos que

$$\begin{aligned} EQM(\bar{y}_s) &= E(\bar{y}_s - \bar{y})^2 = \text{Var}(\bar{y}_s - \bar{y}) = \\ &= \frac{1}{N^2} \text{Var}[(N-n)\bar{y}_s + (N-n)\bar{y}_r] = \frac{(N-n)^2}{N^2 n} + \frac{(N-n)^2}{N^2 (N-n)} = \\ &= \frac{N-n}{Nn}, \quad \text{onde } \bar{y}_r = \frac{\sum_{i=n+1}^N Y_i}{N-n}. \end{aligned}$$

Obtida a amostra s e observado o valor $\bar{y}_s$, utilizando $L_p(\bar{y} | \bar{y}_s)$, um intervalo de previsão para $\bar{y}$ com coeficiente 0,95 será

$$\bar{y}_s \pm 1,96 \sqrt{\frac{N-n}{Nn}}.$$

□

Exemplo 3.2

Seja o modelo

$$Y_i = \beta x_i + \varepsilon_i, \quad i = 1, 2, \dots, N, \text{ com}$$

$$\varepsilon_i \sim N(0, x_i), \quad x_i > 0,$$

$\varepsilon_i, \varepsilon_j$ variáveis aleatórias independentes, $i \neq j$, $i, j = 1, 2, \dots, N$, e

que escrito na forma matricial fica

$$y = X\beta + \varepsilon,$$

onde

$$X = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix}, \quad y \sim N_N(X\beta, V), \quad V = \text{diag}(x_1, x_2, \dots, x_N).$$

Temos que

$$H = X'V^{-1}X = \sum_{i=1}^N x_i, \quad ,$$

$$A = (X'V^{-1}X)^{-1}X'V^{-1} = \frac{\frac{1}{N}}{\sum_{i=1}^N x_i},$$

$$B_N = \frac{\frac{1}{N}}{\sum_{i=1}^N x_i} y = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} \quad e$$

$$\hat{\beta} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}, \quad \text{que é a média de } L_p(B_N | y_s).$$

Para calcularmos $W = E^{-1} - E^{-1}D'H^{-1}DE^{-1}$, temos

$$E = \text{diag}(x_{n+1} \dots x_N),$$

$$D = X'_r = [x_{n+1} \dots x_N] \quad e$$

$$E^{-1}D'H^{-1}DE^{-1} = \frac{1}{\sum_{i=1}^N x_i} \text{diag}\left(\frac{1}{x_{n+1}}, \dots, \frac{1}{x_N}\right) \begin{bmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_N \end{bmatrix} [x_{n+1} \dots x_N] \text{diag}\left(\frac{1}{x_{n+1}}, \dots, \frac{1}{x_N}\right)$$

$$= \frac{1}{\sum_{i=1}^N x_i} J_{N-n}.$$

Nestas condições,

$$W = \text{diag}\left(\frac{1}{x_{n+1}}, \dots, \frac{1}{x_N}\right) - \frac{1}{\sum_{i=1}^N x_i} J_{N-n}.$$

Se a matriz G , de ordem $k \times k$ é tal que

$$G = Q + \alpha \underline{a}\underline{b}',$$

onde Q é diagonal não singular, $\underline{a}$ e $\underline{b}$ são vetores $k \times 1$ e α é um escalar, então

$$G^{-1} = Q^{-1} + \gamma \underline{a}^* \underline{b}^*,$$

$$\text{onde } \gamma = -\alpha \left(1 + \alpha \sum_{i=1}^k \frac{a_i b_i}{q_{ii}} \right)^{-1},$$

os elementos de $\underline{a}^*$ e $\underline{b}^*$ são

$$a_i^* = \frac{a_i}{q_{ii}}, \quad b_i^* = \frac{b_i}{q_{ii}}$$

a_i , b_i são respectivamente os elementos de $\underline{a}$ e $\underline{b}$ e q_{ii} é o i -ésimo elemento da diagonal de Q .

Utilizando este resultado e após cálculos, obtemos

$$W^{-1} = \text{diag}(x_{n+1} \dots x_N) + \frac{1}{\sum_{i=1}^n x_i} \begin{bmatrix} x_{n+1} \\ x_{n+2} \\ \vdots \\ x_N \end{bmatrix} \begin{bmatrix} x_{n+1} & \dots & x_N \end{bmatrix}$$

e

$$\begin{aligned} H^{-1} D E^{-1} W^{-1} E^{-1} D' H^{-1} &= \frac{1}{\left(\sum_{i=1}^n x_i \right)^2} \mathbf{1}_{N-n}' W^{-1} \mathbf{1}_{N-n} = \\ &= \frac{\sum_{i=n+1}^N x_i}{\sum_{i=1}^n x_i \sum_{i=1}^N x_i} . \end{aligned}$$

Desta forma, pelo teorema 3.4,

$$L_P \left(\frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} \mid y_s \right) \propto N \left(\hat{\beta}, \frac{\sum_{i=n+1}^N x_i}{\sum_{i=1}^n x_i \sum_{i=1}^N x_i} \right)$$

e o previsor de máxima verossimilhança de $B_N = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}$ é

$$\hat{\beta} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} .$$

Verificamos ainda que

$$EQM(\hat{\beta}) = E(\hat{\beta} - B_N)^2 = \text{Var}(\hat{\beta} - B_N) =$$

$$\begin{aligned}
&= \text{Var} \left(\frac{\sum_{i=1}^n y_i \sum_{i=1}^N x_i - \sum_{i=1}^N y_i \sum_{i=1}^n x_i}{\sum_{i=1}^N x_i \sum_{i=1}^n x_i} \right) = \\
&= \text{Var} \left(\frac{\sum_{i=1}^n y_i \sum_{i=n+1}^N x_i - \sum_{i=n+1}^N y_i \sum_{i=1}^n x_i}{\sum_{i=1}^N x_i \sum_{i=1}^n x_i} \right) = \\
&= \frac{1}{\left(\sum_{i=1}^N x_i \sum_{i=1}^n x_i \right)^2} \left[\left(\sum_{i=n+1}^N x_i \right)^2 \sum_{i=1}^n x_i + \left(\sum_{i=1}^n x_i \right)^2 \sum_{i=n+1}^N x_i \right] = \\
&= \frac{\sum_{i=1}^n x_i \sum_{i=n+1}^N x_i}{\left(\sum_{i=1}^N x_i \sum_{i=1}^n x_i \right)^2} \left[\sum_{i=n+1}^N x_i + \sum_{i=1}^n x_i \right] = \frac{\sum_{i=n+1}^N x_i}{\sum_{i=1}^N x_i \sum_{i=1}^n x_i},
\end{aligned}$$

que é a variância da distribuição condicional $L_p(\tilde{B}_N | y_s)$.

Da mesma maneira que no exemplo anterior, um intervalo de previsão com coeficiente 0,95 para

$$\tilde{B}_N = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} \text{ é dado por}$$

$$\hat{\beta} \pm 1,96 \sqrt{\frac{\sum_{i=n+1}^N x_i}{\sum_{i=1}^n x_i \sum_{i=1}^N x_i}},$$

$$\text{onde } \hat{\beta} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}.$$

□

Exemplo 3.3

Seja $y = X\beta + \varepsilon$ um modelo de superpopulação qualquer, com $y \sim N(X\beta, V)$,

$$\text{onde } V = \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix} = \begin{bmatrix} V_s & 0 \\ 0 & V_r \end{bmatrix}, \text{ de modo que } y_s \text{ e } y_r \text{ são}$$

independentes.

Pelo teorema 3.4,

$$L_p(B_N | y_s) \propto N_{p+1}(\hat{\beta}, H^{-1} D E^{-1} W^{-1} E^{-1} D' H^{-1}).$$

Temos que

$$E = V_r, \quad D = X_r', \quad B = X_s', \quad C = V_s,$$

$$W = E^{-1} - E^{-1} D' H^{-1} D E^{-1} = V_r^{-1} - V_r^{-1} X_r (X_r' V_r^{-1} X_r)^{-1} X_r' V_r^{-1}.$$

Usando o resultado (A.2.10) do apêndice 2, após alguns cálculos, obtemos

$$W^{-1} = V_r + X_r (X_s' V_s^{-1} X_s)^{-1} X_r'$$

e

$$H^{-1} D E^{-1} W^{-1} E^{-1} D' H^{-1} = H^{-1} X_r' V_r^{-1} [V_r + X_r (X_s' V_s^{-1} X_s)^{-1} X_r'] V_r^{-1} X_r H^{-1} =$$

$$H^{-1} [X_r' + X_r' V_r^{-1} X_r (X_s' V_s^{-1} X_s)^{-1} X_r'] V_r^{-1} X_r H^{-1} =$$

$$\begin{aligned}
& H^{-1} [X_r' V_r^{-1} X_r + X_r' V_r^{-1} X_r (X_s' V_s^{-1} X_s)^{-1} X_r' V_r^{-1} X_r] H^{-1} = \\
& H^{-1} (X_r' V_r^{-1} X_r) (X_s' V_s^{-1} X_s)^{-1} (X_s' V_s^{-1} X_s + X_r' V_r^{-1} X_r) H^{-1} = \\
& (X' V^{-1} X)^{-1} (X_r' V_r^{-1} X_r) (X_s' V_s^{-1} X_s)^{-1}.
\end{aligned}$$

Lembrando que o erro quadrático médio generalizado de $\hat{\beta}$, $EQMG(\hat{\beta})$, calculado no corolário 2.1 era

$$EQMG(\hat{\beta}) = \lambda' (X' V^{-1} X)^{-1} (X_r' V_r^{-1} X_r) (X_s' V_s^{-1} X_s)^{-1} \lambda,$$

verificamos que $H^{-1} D E^{-1} W^{-1} E^{-1} D' H^{-1}$ corresponde realmente à $Var(\hat{\beta} - \underline{B}_N)$, de modo que

$$EQMG(\hat{\beta}) = E[\lambda' (\hat{\beta} - \underline{B}_N) (\hat{\beta} - \underline{B}_N)' \lambda] = \lambda' Var(\hat{\beta} - \underline{B}_N) \lambda.$$

Assim, quando $V_{sr} = 0$, pelo teorema 3.4

$$L_p(\underline{B}_N | \underline{y}_s) \propto N(\hat{\beta}, Var(\hat{\beta} - \underline{B}_N))$$

e $\hat{\beta}$, o previsor de mínimos quadrados, é o previsor de máxima verossimilhança de $\underline{B}_N$.

□

Exemplo 3.4

Vamos considerar novamente o modelo de superpopulação

$$y = X\beta + \varepsilon,$$

onde

$$X = \underline{1}_N$$

$$V = (1 - \rho) I_N + \rho \underline{1}_N \underline{1}_N'$$

e

$$y \sim N_N(X\beta, V).$$

Do exemplo 2.1, temos que

$$\bar{y}_N = \bar{y} \quad \text{e} \quad \hat{\beta} = \bar{y}_s.$$

Além disso,

$$H = \frac{N}{1+(N-1)\rho} \quad , \quad D = \frac{1-\rho}{1+(n-1)\rho} \mathbf{1}'_{N-n} \quad ,$$

$$E^{-1} = \frac{1}{1-\rho} \left[\mathbf{I}_{N-n} - \frac{\rho}{1+\rho(N-1)} \mathbf{J}_{N-n} \right],$$

de modo que

$$W = E^{-1} - E^{-1}D'H^{-1}DE^{-1} = \frac{1}{1-\rho} \left[\mathbf{I}_{N-n} - \frac{1}{N} \mathbf{J}_{N-n} \right],$$

$$W^{-1} = (1-\rho) \left[\mathbf{I}_{N-n} + \frac{1}{N} \mathbf{J}_{N-n} \right]$$

e

$$H^{-1}DE^{-1}W^{-1}E^{-1}D'H^{-1} = (1-\rho) \frac{(N-n)}{Nn}.$$

Assim,

$$L_p(\bar{y} | \bar{y}_s) \propto N\left(\bar{y}_s, (1-\rho) \frac{(N-n)}{Nn}\right),$$

o previsor de máxima verossimilhança da média populacional $\bar{y}$ é a média amostral $\bar{y}_s$ e

$$EQM(\bar{y}_s) = (1-\rho) \frac{(N-n)}{Nn}.$$

Dada $L_p(\bar{y} | \bar{y}_s)$, um intervalo de previsão para $\bar{y}$, com coeficiente 0,95 é

$$\bar{y}_s \pm 1,96 \sqrt{\frac{(1-\rho)(N-n)}{Nn}}.$$

Se $\rho = 0$, $V = I_N$ e $L_p(\bar{y} | \bar{y}_s) \propto N\left(\bar{y}_s, \frac{N-n}{Nn}\right)$.

Neste caso, o modelo considerado se reduz ao do exemplo 3.1, o mesmo ocorrendo com a função de verossimilhança preditiva $L_p(\bar{y} | y_s)$.

3.7 - Considerações Finais

A motivação principal para propormos uma teoria de verossimilhança preditiva em populações finitas via "profile" preditiva foi o trabalho de Bjornstad [1990].

Através desta teoria foi possível obtermos vários resultados importantes, dentre os quais destacamos o teorema 3.1, que mostra uma conexão entre previsão ótima e de máxima verossimilhança, o corolário 3.3, que apresenta um resultado análogo ao de Royall [1976b] para matriz de covariância qualquer e finalmente o teorema 3.4, que introduz a previsão de máxima verossimilhança para B_N .

Como sugestão para um trabalho futuro, além da já apresentada na seção 3.5, proporíamos um estudo semelhante ao feito neste capítulo, utilizando outras distribuições multivariadas que não a normal.

CAPÍTULO 4 - TEORIA ASSINTÓTICA UNIFICADA

4.1 - Introdução

Na literatura relacionada com populações finitas existe, surpreendentemente até, uma área bastante rica destinada à teoria assintótica.

Pelo fato da população ser finita, tal estudo pode parecer contraditório, no entanto, ele é relativamente comum mesmo em livros clássicos de amostragem. Cochran [1977], sob a abordagem clássica, utiliza a aproximação normal para a média amostral baseada em grandes tamanhos de amostra de populações finitas na construção de intervalos de confiança. Conover [1980] apresenta uma versão do teorema do limite central para populações finitas em amostragem sem reposição.

Dentre os artigos, destacamos Scott e Wu [1981] que, também no contexto clássico, apresentam importantes resultados assintóticos sobre os estimadores razão e regressão.

Sob a abordagem de modelos de superpopulação, inúmeros são os trabalhos na área, dentre os quais citamos Fuller [1975,1987] e Bolfarine e Zacks [1992].

Quando a população é infinita, dispomos apenas do valor n , tamanho da amostra e, dada uma seqüência de variáveis

aleatórias, o que se avalia são os diferentes tipos de convergência (em probabilidades, em distribuição e outros) quando este tamanho de amostra tende a infinito.

No caso de populações finitas, existem dois tamanhos a considerar, o da amostra e o da população, e para que uma teoria assintótica seja possível, deve ficar claro de que forma estes valores se comportam.

Com este objetivo, são estabelecidos conceitos de convergência em populações finitas, que, além de definirem de que forma n e N tendem a infinito, caracterizam também a sequência de populações e de amostras destas populações.

Fuller [1987] e Bolfarine e Zacks [1992] trabalham com uma sequência de populações

$$\{\mathcal{P}_r, r = 1, 2, \dots\},$$

onde $\mathcal{P}_r = \{1, 2, \dots, N_r\}$ corresponde à r -ésima população. Desta r -ésima população é selecionada uma amostra de tamanho n_r e quando $r \rightarrow \infty$ (r tende a infinito), $n_r \rightarrow \infty$, de modo que $f_r = \frac{n_r}{N_r} \rightarrow f$, com $0 \leq f < 1$.

Fuller [1975] utiliza um outro conceito de convergência, onde

$$\{\mathcal{P}_n, n = 1, 2, \dots\}$$

é uma sequência de populações finitas com $\mathcal{P}_n = \{1, 2, \dots, N_n\}$,

$N_n > N_{n-1}$. Da n -ésima população finita é selecionada uma amostra de tamanho n e, se $f_n = \frac{n}{N_n}$, admite-se que $\lim_{n \rightarrow \infty} f_n = f$, $0 \leq f < 1$.

Conforme justificado no apêndice 4, o último conceito apresenta algumas vantagens com relação ao anterior e por este motivo será adotado em nosso trabalho.

Utilizando este conceito de convergência, obtivemos extensões de dois teoremas presentes em Fuller [1987,1975].

Estas extensões, que introduziremos na próxima seção, são referentes à distribuição limite de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B}_N - B_N \end{bmatrix}$ no modelo de superpopulação $y = X\beta + \varepsilon$, propondo uma teoria assintótica unificada, no sentido de simultaneamente estimar o superparâmetro β e prever o coeficiente de regressão B_N .

A principal contribuição deste resultado é teórica, já que o teorema obtido estende o apresentado em Fuller [1975], que determina a distribuição limite das marginais $\sqrt{n}(\hat{\beta} - \beta)$ e $\sqrt{n}(\hat{B}_N - B_N)$. Sua mais importante aplicação, que apresentaremos posteriormente, é a previsão de máxima verossimilhança de B_N quando a distribuição de probabilidades de y não está especificada. Feita esta aplicação, uma interessante conexão com o teorema 3.4 no caso em que V é diagonal vai surgir.

No modelo $y = X\beta + \varepsilon$, supondo que $y_1, y_2, \dots, y_N$ são variáveis aleatórias independentes, duas situações serão consideradas. A primeira, ainda não utilizada neste trabalho, admite que a matriz X é aleatória, correspondendo ao modelo adotado por

Fuller [1975].

Posteriormente, o mesmo teorema é demonstrado para o caso em que X é matriz fixa e então, tomando X aleatória, mas com distribuição degenerada numa constante, mostraremos a analogia entre os resultados.

Na seção 4.3, como consequência do teorema apresentado, provamos a normalidade assintótica da média populacional e finalmente, na seção 4.4 fornecemos aplicações do referido teorema.

No apêndice 4 se encontram os teoremas básicos, dos quais nossos resultados se constituem em extensões e também definições e teoremas auxiliares que utilizaremos em algumas demonstrações.

Todos os teoremas deste capítulo dizem respeito à convergência em distribuição e $\underline{X}_n \xrightarrow[n \rightarrow \infty]{D} \underline{X}$ vai indicar que o vetor aleatório $\underline{X}_n$ converge em distribuição para o vetor $\underline{X}$ quando n tende a infinito.

4.2 - Normalidade Assintótica Unificada no Modelo de Regressão

O teorema estabelecido em Fuller [1987], teorema A.4.3 do apêndice 4, garante que se y_{ri} , $i = 1, 2, \dots, N_r$, $r = 1, 2, \dots$ são variáveis aleatórias independentes, identicamente distribuídas com $E(y_{ri}) = \mu$, $\text{Var}(y_{ri}) = \sigma_{yy}$ e se para $r \rightarrow \infty$, $n_r \rightarrow \infty$ e $\frac{n_r}{N_r} \rightarrow f$, $0 \leq f < 1$, então

$$\sqrt{n_r} \begin{bmatrix} \bar{y}_r - \mu \\ \bar{y}_r - \bar{Y}_r \end{bmatrix} \xrightarrow[r \rightarrow \infty]{D} N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1-f \\ 1-f & 1-f \end{bmatrix} \sigma_{YY} \right) \quad (4.1),$$

onde $\bar{y}_r = \frac{\sum_{i=1}^n y_{ri}}{n_r}$ e $\bar{Y}_r = \frac{\sum_{i=1}^N Y_{ri}}{N_r}$ representam respectivamente as médias amostral e populacional da variável Y , associadas à r -ésima população.

O modelo considerado, conhecido como modelo de locação, pode ser escrito na forma $y = X\beta + \varepsilon$ com $X = \underline{1}_N$, $\beta = [\mu]$ e $\text{Var}(\varepsilon) = V = \sigma_{YY}^2 I_N$, e é o mais simples dos modelos lineares. Observamos ainda que μ é o superparâmetro do modelo e $\bar{Y}_r$ a quantidade aleatória a ser prevista, de modo que o teorema fornece um resultado envolvendo conjuntamente estimação e previsão.

Na presente seção, vamos generalizar este resultado, obtendo a distribuição assintótica de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\beta} - B_N \end{bmatrix}$ no modelo de

regressão $y = X\beta + \varepsilon$ e mostrar que, no caso particular do modelo de locação, o resultado obtido vai se reduzir a (4.1).

De acordo com o que comentamos na introdução, abordaremos os casos em que X é fixa e X é aleatória, que acreditamos, cobrem todos os casos de interesse.

A teoria de regressão usualmente admite que os valores das p variáveis auxiliares são fixos, de modo que a aleatorização ocorre somente na variável Y . Quando estes valores das variáveis

auxiliares não são controlados (isto ocorrerá se, por exemplo, forem selecionadas observações $(y_i, x_{i1}, x_{i2} \dots x_{ip})$ de uma distribuição $(p+1)$ -variada), a matriz de planejamento X será aleatória e o modelo se torna mais complicado.

Quando X é aleatória, o teorema obtido se constituirá também numa extensão de outro resultado, proposto por Fuller [1975] e apresentado no apêndice 4 como teorema A.4.9.

4.2.1 - Teorema do Limite Central Unificado no Modelo $y = X\beta + \varepsilon$, X aleatória

Consideremos o modelo de superpopulação utilizado em Fuller [1975],

$$y = X\beta + \varepsilon \quad (4.2),$$

onde

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_N \end{bmatrix}, \quad X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ \vdots & & & & \vdots \\ 1 & x_{N1} & x_{N2} & & x_{Np} \end{bmatrix} \quad \text{e} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix}$$

são matrizes aleatórias e tal que

i) os vetores aleatórios

$$u'_i = [x_{i1} \quad x_{i2} \quad \dots \quad x_{ip} \quad y_i], \quad i = 1, 2, \dots, N$$

são independentes

ii) os vetores aleatórios

$$\tilde{x}'_i \varepsilon_i = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}] \varepsilon_i = [\varepsilon_i \ x_{i1} \varepsilon_i \ x_{i2} \varepsilon_i \ \dots \ x_{ip} \varepsilon_i],$$

$i = 1, 2, \dots, N$, são independentes e identicamente distribuídos com média 0 e matriz de covariância $G = (g_{rs})$, $g_{rs} = E(x_{ir} x_{is} \varepsilon_i^2)$, $r, s = 0, 1, 2, \dots, p$.

iii) os quartos momentos das variáveis aleatórias x_{ij} são finitos, ou seja

$$E(|x_{ij} x_{ik} x_{il} x_{im}|) < \infty,$$

para todo $i = 1, 2, \dots, N$, $j, k, l, m = 0, 1, 2, \dots, p$, com a variável x_{i0} sendo constante igual a 1.

Dado este modelo, são definidos:

1) O vetor de parâmetros ou superparâmetro,

$$\underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix} = Q^{-1} \underline{H}$$

onde os elementos da matriz Q e do vetor $\underline{H}$ são respectivamente

$$q_{rs} = E(x_{ir} x_{is}), \quad h_s = E(x_{is} y_i), \quad r, s = 0, 1, 2, \dots, p.$$

2) O vetor dos coeficientes de regressão calculados com base nas N unidades populacionais

$$\underline{B} = Q_N^{-1} \underline{H}_N \quad (4.3),$$

onde

Q_N é a matriz cujo elemento que ocupa a posição r,s é dado por

$$q_{N_{rs}} = \frac{1}{N} \sum_{i=1}^N x_{ir} x_{is}, \quad r,s = 0,1,2\dots p,$$

$\underline{H}_N$ é o vetor cujo s -ésimo elemento é dado por

$$h_{N_s} = \frac{1}{N} \sum_{i=1}^N x_{is} y_i, \quad s = 0,1,2\dots p,$$

x_{ir} é o valor observado da r -ésima variável auxiliar para o i -ésimo elemento populacional,

e

y_i é o valor da variável dependente para o i -ésimo elemento populacional.

A notação x_{ir} está sendo usada tanto para representar genericamente a r -ésima variável auxiliar para o i -ésimo elemento populacional quanto para designar o seu valor observado.

O vetor $\underline{B}$ que definimos em (4.3) pode também ser escrito como

$$\underline{B} = (X'X)^{-1}X'Y,$$

correspondendo ao vetor B_N do capítulo 2, mas com a diferença que neste caso, X é aleatória.

3) O estimador de β com base na amostra de tamanho n

$$\hat{\beta} = Q_n^{-1} H_n \quad (4.4),$$

onde os elementos de Q_n e H_n são respectivamente

$$q_{n_{rs}} = \frac{1}{n} \sum_{i=1}^n x_{ir} x_{is}, \quad h_{n_s} = \frac{1}{n} \sum_{i=1}^n x_{is} y_i, \quad r, s = 0, 1, 2, \dots, p.$$

Na notação do capítulo 2, $\hat{\beta} = (X'_s X_s)^{-1} X'_s y_s$, que é o estimador de mínimos quadrados.

Admitiremos então que $\{\mathcal{P}_n, n = 1, 2, \dots\}$ é a sequência de populações finitas, $\mathcal{P}_n = \{1, 2, \dots, N_n\}$, $N_n > N_{n-1}$ e que n é o tamanho de amostra da n -ésima população, de modo que $\lim_{n \rightarrow \infty} f_n = f$, $f_n = \frac{n}{N_n}$, $0 \leq f < 1$. O próximo teorema vai se constituir no principal resultado desta seção.

Devido à abordagem assintótica aqui utilizada, passaremos a trabalhar com o símbolo N_n ao invés de simplesmente N . Seria mais rigoroso utilizarmos também os termos B_{N_n} e $\hat{\beta}_{N_n}$ para nos referirmos às expressões (4.3) e (4.4), mas isto não será feito para simplificar a notação.

Teorema 4.1

No modelo (4.2), para $n \rightarrow \infty$ e $\lim_{n \rightarrow \infty} \frac{n}{N_n} = f$, então

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma),$$

$$\text{com } \Gamma = \begin{bmatrix} Q^{-1} G Q^{-1} & (1-f) Q^{-1} G Q^{-1} \\ (1-f) Q^{-1} G Q^{-1} & (1-f) Q^{-1} G Q^{-1} \end{bmatrix}.$$

Prova:

Os vetores aleatórios $\underline{x}'_i \varepsilon_i = [\varepsilon_i, x_{i1} \varepsilon_i, x_{i2} \varepsilon_i \dots x_{ip} \varepsilon_i]$, $i = 1, 2, \dots, N_n$ são independentes e identicamente distribuídos com média $\underline{0}$ e matriz de covariância G .

Assim, estão satisfeitas as condições do teorema A.4.6 (teorema do limite central multivariado clássico) e com isto,

$$\underline{z}_n = n^{-1/2} \begin{bmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ip} \varepsilon_i \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, G)$$

e

$$\tilde{W}_n = (N_n - n)^{-1/2} \begin{bmatrix} \sum_{i=n+1}^{N_n} \varepsilon_i \\ \vdots \\ \sum_{i=n+1}^N x_{ip} \varepsilon_i \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, G).$$

Como $\tilde{Z}_n$ e $\tilde{W}_n$ são independentes $\forall n$, devido ao teorema

A.4.10

$$\begin{bmatrix} \tilde{Z}_n \\ \tilde{W}_n \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2} \left(\underline{0}, \begin{bmatrix} G & 0 \\ 0 & G \end{bmatrix} \right).$$

Sejam

$$\tilde{R}_n = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ip} \varepsilon_i \end{bmatrix} \quad \text{e} \quad \tilde{R}_{N_n} = \frac{1}{N_n} \begin{bmatrix} \sum_{i=1}^N \varepsilon_i \\ \sum_{i=1}^N x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^N x_{ip} \varepsilon_i \end{bmatrix}.$$

Conforme verificado na prova do teorema A.4.9,

$$\tilde{R}_n = o_p(n^{-1/2}), \quad \hat{\beta} - \beta = Q_n^{-1} \tilde{R}_n = (Q^{-1} + o_p(n^{-1/2})) \tilde{R}_n$$

e assim,

$$\hat{\underline{\beta}} - \underline{\beta} = Q^{-1} \underline{R}_n + o_p(n^{-1}).$$

Além disso, tínhamos que $\hat{\underline{\beta}} - \underline{B} = Q^{-1}(\underline{R}_n - \underline{R}_{N_n}) + o_p(n^{-1})$ e portanto,

$$\begin{aligned} \sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{\beta}} - \underline{B} \end{bmatrix} &= \sqrt{n} \begin{bmatrix} Q^{-1} \underline{R}_n + o_p(n^{-1}) \\ Q^{-1}(\underline{R}_n - \underline{R}_{N_n}) + o_p(n^{-1}) \end{bmatrix} = \\ &= \sqrt{n} \begin{bmatrix} Q^{-1} \underline{R}_n \\ Q^{-1}(\underline{R}_n - \underline{R}_{N_n}) \end{bmatrix} + \sqrt{n} o_p(n^{-1}) = \\ &= \sqrt{n} \begin{bmatrix} Q^{-1} \underline{R}_n \\ Q^{-1}(\underline{R}_n - \underline{R}_{N_n}) \end{bmatrix} + o_p(n^{-1/2}). \end{aligned}$$

Se $\underline{b}_n$ é um vetor aleatório tal que $\underline{b}_n = o_p(n^{-1/2})$, então $\underline{b}_n \xrightarrow{p} \underline{0}$ e como consequência do teorema de Slutsky multivariado, a distribuição limite de

$$\sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{\beta}} - \underline{B} \end{bmatrix} \text{ é a mesma de } \underline{H}_n = \sqrt{n} \begin{bmatrix} Q^{-1} \underline{R}_n \\ Q^{-1}(\underline{R}_n - \underline{R}_{N_n}) \end{bmatrix}.$$

Observamos que

$$\sqrt{n} \underline{R}_n = \underline{Z}_n \quad \text{e, por (A.4.3)}$$

$$\sqrt{n}(\underline{R}_n - \underline{R}_{N_n}) = \begin{bmatrix} n^{-1/2}(1-f_n) \sum_{i=1}^n \varepsilon_i - f_n^{1/2}(1-f_n)^{1/2}(N_n - n)^{-1/2} \sum_{i=n+1}^N \varepsilon_i \\ \vdots \\ n^{-1/2}(1-f_n) \sum_{i=1}^n x_{ip} \varepsilon_i - f_n^{1/2}(1-f_n)^{1/2}(N_n - n)^{-1/2} \sum_{i=n+1}^N x_{ip} \varepsilon_i \end{bmatrix}$$

$$= (1-f_n) \underline{Z}_n - f_n^{1/2}(1-f_n)^{1/2} \underline{W}_n.$$

Desta forma,

$$\underline{H}_n = \sqrt{n} \begin{bmatrix} Q^{-1} \underline{R}_n \\ Q^{-1} (\underline{R}_n - \underline{R}_{N_n}) \end{bmatrix} = \begin{bmatrix} Q^{-1} & 0 \\ (1-f_n) Q^{-1} & -f_n^{1/2}(1-f_n)^{1/2} Q^{-1} \end{bmatrix} \begin{bmatrix} \underline{Z}_n \\ \underline{W}_n \end{bmatrix}$$

$$\text{e como } \begin{bmatrix} \underline{Z}_n \\ \underline{W}_n \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2} \left(\underline{0}, \begin{bmatrix} G & 0 \\ 0 & G \end{bmatrix} \right),$$

segue que

$$\underline{H}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2} (\underline{0}, \Gamma), \text{ onde}$$

$$\Gamma = A \begin{bmatrix} G & 0 \\ 0 & G \end{bmatrix} A' \quad \text{com}$$

$$A = \lim_{n \rightarrow \infty} \begin{bmatrix} Q^{-1} & 0 \\ (1-f_n)Q^{-1} & -f_n^{1/2}(1-f_n)^{1/2}Q^{-1} \end{bmatrix}.$$

Portanto,

$$\begin{aligned} \Gamma &= \begin{bmatrix} Q^{-1} & 0 \\ (1-f)Q^{-1} & -f^{1/2}(1-f)^{1/2}Q^{-1} \end{bmatrix} \begin{bmatrix} G & 0 \\ 0 & G \end{bmatrix} \begin{bmatrix} Q^{-1} & (1-f)Q^{-1} \\ 0 & -f^{1/2}(1-f)^{1/2}Q^{-1} \end{bmatrix} \\ &= \begin{bmatrix} Q^{-1}G & 0 \\ (1-f)Q^{-1}G & -f^{1/2}(1-f)^{1/2}Q^{-1}G \end{bmatrix} \begin{bmatrix} Q^{-1} & (1-f)Q^{-1} \\ 0 & -f^{1/2}(1-f)^{1/2}Q^{-1} \end{bmatrix} = \\ &= \begin{bmatrix} Q^{-1}GQ^{-1} & (1-f)Q^{-1}GQ^{-1} \\ (1-f)Q^{-1}GQ^{-1} & (1-f)Q^{-1}GQ^{-1} \end{bmatrix}. \end{aligned}$$

□

Salientamos que o resultado permanece válido se excluirmos do modelo a variável $X_0 = 1$ (trabalhando assim com o modelo sem intercepto) porque a demonstração utilizada também se aplica neste caso.

Através do teorema 4.1 podemos obter as distribuições assintóticas de $\sqrt{n}(\hat{\beta} - \beta)$ e $\sqrt{n}(\hat{B} - B)$.

Como a distribuição limite de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix}$ é normal

$(2p+2)$ -variada, segue a normalidade de suas marginais, ou seja,

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\beta} - \underline{\beta} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma)$$

$$\Rightarrow \sqrt{n}(\hat{\beta} - \underline{\beta}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, Q^{-1}GQ^{-1}) \quad (4.5)$$

e

$$\sqrt{n}(\hat{\beta} - \underline{\beta}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, (1-f)Q^{-1}GQ^{-1}) \quad (4.6).$$

O resultado (4.6) corresponde ao teorema 1 de Fuller [1975], enunciado e demonstrado no apêndice 4, teorema A.4.9.

Já, (4.5) trata da convergência dos estimadores de mínimos quadrados e também surge em Fuller [1975], embora seja um resultado bastante conhecido na análise de regressão clássica. Este teorema não envolve nenhum conceito de populações finitas e por este motivo, o fator de correção $(1-f)$ não está presente na matriz de covariância da distribuição limite.

No modelo de locação, temos que

$$X = \underline{1}_N,$$

$$\beta = [\mu]$$

$$Q_n^{-1} = [q_{n_{oo}}] = \frac{1}{n} \quad n = 1$$

$$h_{n_0} = n^{-1} \sum_{i=1}^n x_{i0} y_i = \frac{1}{n} \sum_{i=1}^n y_i = \bar{y}_n$$

$$Q_N^{-1} = \frac{1}{N_n} \sum_{i=1}^N x_{i0} x_{i0} = \frac{N_n}{N_n} = 1,$$

$$h_{N_0} = \frac{1}{N_n} \sum_{i=1}^N x_{i0} y_i = \frac{1}{N_n} \sum_{i=1}^N y_i = \bar{y}_{N_n}$$

e como consequência de (4.3) e (4.4)

$$\tilde{B} = Q_N^{-1} \tilde{H}_N = \bar{y}_{N_n} \quad \text{e} \quad \hat{\tilde{\beta}} = Q_n^{-1} \tilde{H}_n = \bar{y}_n.$$

Além disso,

$$Q = E(x_{i0} x_{i0}) = 1, \text{ pois } x_{i0} = 1 \quad \forall i = 1, 2, \dots, N_n,$$

$$G = E(\varepsilon_i^2) = \text{Var}(\varepsilon_i) = \text{Var}(y_i) = \sigma_{yy}$$

e assim, aplicando o teorema 4.1,

$$\sqrt{n} \begin{bmatrix} \hat{\tilde{\beta}} - \tilde{\beta} \\ \hat{\tilde{\beta}} - \tilde{B} \end{bmatrix} = \sqrt{n} \begin{bmatrix} \bar{y}_n - \mu \\ \bar{y}_n - \bar{y}_{N_n} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(Q, \Gamma),$$

onde

$$\Gamma = \begin{bmatrix} Q^{-1} G Q^{-1} & (1-f) Q^{-1} G Q^{-1} \\ (1-f) Q^{-1} G Q^{-1} & (1-f) Q^{-1} G Q^{-1} \end{bmatrix} = \begin{bmatrix} 1 & 1-f \\ 1-f & 1-f \end{bmatrix} \sigma_{yy}.$$

Portanto, o teorema 4.1 estende também o resultado de Fuller [1987], que fornece um teorema do limite central unificado no modelo de locação.

Corolário 4.1

Nas mesmas condições do teorema 4.1,

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{\beta}}{N_n} \\ \frac{\underline{1}' X \hat{\underline{B}}}{N_n} - \frac{\underline{1}' X \underline{B}}{N_n} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(\underline{0}, D \Gamma D'),$$

onde $\underline{1}'$ é o vetor linha N-dimensional contendo 1 em todas as posições.

Prova:

Pelo teorema 4.1, $\sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{B}} - \underline{B} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma)$ e, porque

x_{ij} , $i=1,2,\dots,N_n$ são independentes e identicamente distribuídas para cada $j=1,2,\dots,p$, vale a lei fraca dos grandes números, ou seja,

$$\frac{\underline{1}' X}{N_n} = \begin{bmatrix} 1 & \frac{\sum_{i=1}^N x_{i1}}{N_n} & \frac{\sum_{i=1}^N x_{i2}}{N_n} & \dots & \frac{\sum_{i=1}^N x_{ip}}{N_n} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{P}$$

$$\begin{bmatrix} 1 & E(x_1) & E(x_2) & \dots & E(x_p) \end{bmatrix} = \begin{bmatrix} 1 & q_{10} & q_{20} & \dots & q_{p0} \end{bmatrix}.$$

Se $\begin{bmatrix} 1 & q_{10} & q_{20} & \dots & q_{p0} \end{bmatrix} = q'$, pelo teorema de Slutsky multivariado

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{\beta}}{N_n} \\ \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{B}}{N_n} \end{bmatrix} =$$

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' X}{N_n} & \underline{0} \\ \underline{0} & \frac{\underline{1}' X}{N_n} \end{bmatrix} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{\beta}} - \underline{B} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} \begin{bmatrix} q & \underline{0} \\ \underline{0} - q \end{bmatrix} \underline{U},$$

onde $\underline{U} \sim N_{2p+2}(\underline{0}, \Gamma)$.

Portanto,

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{\beta}}{N_n} \\ \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{B}}{N_n} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(\underline{0}, D\Gamma D') ,$$

$$\text{com } D = \begin{bmatrix} g & 0 \\ 0 & g \end{bmatrix}.$$

□

4.2.2 - Teorema do Limite Central Unificado no modelo $y = X\beta + \varepsilon$, X fixa

Na seção anterior, adotando o modelo de regressão com variáveis auxiliares aleatórias, obtivemos um teorema do limite central unificado.

Nesta seção apresentaremos um resultado semelhante para o modelo de regressão com variáveis auxiliares fixas, que é a situação mais freqüente em nosso trabalho.

Quando as variáveis auxiliares eram aleatórias, necessitávamos de condições sobre os vetores aleatórios

$$X'_i \varepsilon_i = [\varepsilon_i, x_{i1}\varepsilon_i, x_{i2}\varepsilon_i \dots x_{ip}\varepsilon_i], i = 1, 2, \dots, N_n.$$

Agora que X é fixa, um resultado análogo será possível supondo uma condição do tipo Liapunov.

Posteriormente veremos que, graças ao teorema de Lindeberg-Feller multivariado, a condição de Liapunov poderá ser abandonada, produzindo um teorema mais geral e portanto mais forte.

Finalmente, forneceremos um teorema do mesmo tipo para o caso em que $\text{Var}(\varepsilon_i) = \sigma^2 v_i$, v_i conhecida, $i = 1, 2, \dots, N_n$, ou seja, um modelo que admite uma possível heterocedasticidade.

Vamos considerar o modelo de superpopulação

$$Y = X\beta + \varepsilon \quad (4.7),$$

onde

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{N1} & x_{N2} & & x_{NP} \end{bmatrix} \text{ é uma matriz não aleatória de}$$

constantes conhecidas,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \text{ e } \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \text{ são vetores aleatórios tais que}$$

$$i) E(\varepsilon) = 0 \text{ e } \text{Var}(\varepsilon) = \sigma^2 I_N,$$

ii) $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ são independentes e identicamente distribuídas e como consequência, $Y_1, Y_2, \dots, Y_N$ são variáveis aleatórias independentes com $\text{Var}(Y_i) = \sigma^2$.

Sejam

$$X'_i = [1 \ x_{i1} \ x_{i2} \ \dots \ x_{ip}] \text{ a } i\text{-ésima linha da matriz } X,$$

β o parâmetro populacional ou superparâmetro,

$\hat{\beta} = (X'_s X_s)^{-1} X'_s y_s$ o estimador de mínimos quadrados de β

e

$$B = (X'X)^{-1}X'y.$$

Nestas condições, vale o teorema a seguir.

Teorema 4.2

No modelo (4.7), para $n \rightarrow \infty$ e $\lim_{n \rightarrow \infty} \frac{n}{N_n} = f$, se

$$i) |x_{ij}| \leq c, \quad \forall j = 1, 2, \dots, p, \quad i = 1, 2, \dots, N_n, \quad c \in \mathbb{R}, \quad c > 0$$

$$ii) \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n x_i x'_i}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} X'_s X_s = \Sigma_{xx}$$

$$iii) E(|\varepsilon_i|^{2+\delta}) < \infty \quad \text{para } i = 1, 2, \dots, N_n \quad \text{e para algum } \delta > 0$$

(condição de Liapunov)

então

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma),$$

$$\text{onde } \Gamma = \begin{bmatrix} \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \\ (1-f) \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix}.$$

Prova:

Observando que $\frac{1}{n} X'_s X_s$, $\frac{1}{n} X'_s Y_s$, $\frac{1}{N_n} X' X$ e $\frac{1}{N_n} X' Y$ correspondem respectivamente às matrizes Q_n e $\tilde{H}_n$ de (4.4) e Q_N e H_N de (4.3), conforme verificado na prova do teorema A.4.9,

$$\hat{\beta} - \beta = \left(\frac{X'_s X_s}{n} \right)^{-1} R_n \quad \text{e} \quad \hat{\beta} - \beta = \left(\frac{X'_s X_s}{n} \right)^{-1} R_n - \left(\frac{X' X}{N_n} \right)^{-1} R_{N_n}$$

onde

$$R_n = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ip} \varepsilon_i \end{bmatrix} \quad \text{e} \quad R_{N_n} = \frac{1}{N_n} \begin{bmatrix} \sum_{i=1}^N \varepsilon_i \\ \sum_{i=1}^N x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^N x_{ip} \varepsilon_i \end{bmatrix}.$$

Os vetores aleatórios $\tilde{W}'_i = X'_i \varepsilon_i = [\varepsilon_i, x_{i1} \varepsilon_i, x_{i2} \varepsilon_i \dots x_{ip} \varepsilon_i]$, $i = 1, 2, \dots, N_n$ são independentes, com média $E(\tilde{W}_i) = \tilde{X}_i E(\varepsilon_i) = 0$ e matriz de covariância

$$\text{Var}(\underline{W}_i) = \text{Var}(\underline{X}_i \varepsilon_i) = \underline{X}_i \text{Var}(\varepsilon_i) \underline{X}_i' = \underline{X}_i \underline{X}_i' \sigma^2 = \Sigma_i.$$

Como $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_N$ são identicamente distribuídas, $E(|\varepsilon_i|^{2+\delta}) < \infty$ implica que $E(|\varepsilon_i|^{2+\delta}) = d$, $0 < d < \infty$ e assim,

$$E(|\underline{X}_{ij} \varepsilon_i - E(\underline{X}_{ij} \varepsilon_i)|^{2+\delta}) = E(|\underline{X}_{ij} \varepsilon_i|^{2+\delta}) =$$

$$= |\underline{X}_{ij}|^{2+\delta} E(|\varepsilon_i|^{2+\delta}) \leq c^{2+\delta} \cdot d = c^*, \quad j = 1, 2, \dots, p, \quad i = 1, 2, \dots, N_n,$$

que garante que $\max_{1 \leq j \leq k} E(|\underline{X}_{ij} \varepsilon_i - E(\underline{X}_{ij} \varepsilon_i)|^{2+\delta}) \leq c^*, \forall i=1, 2, \dots, N_n$.

Além disso,

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Sigma_i = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \sigma^2 \underline{X}_i \underline{X}_i' = \sigma^2 \Sigma_{xx}.$$

Desta forma, estão satisfeitas as condições do Teorema de Liapunov Multivariado (ver teorema A.4.11) e

$$\underline{Z}_n = n^{-1/2} \begin{bmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n \underline{X}_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n \underline{X}_{ip} \varepsilon_i \end{bmatrix} = \sqrt{n} \underline{R}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n (\underline{W}_i - \underline{Q}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{Q}, \sigma^2 \Sigma_{xx})$$

A hipótese que $\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \tilde{x}_i \tilde{x}'_i}{n} = \Sigma_{xx}$ implica em

$$\lim_{n \rightarrow \infty} \frac{\sum_{i=1}^{N_n} \tilde{x}_i \tilde{x}'_i}{N_n} = \Sigma_{xx} \quad \text{pois, definindo a sequência}$$

$A_n = \frac{1}{n} \sum_{i=1}^n \tilde{x}_i \tilde{x}'_i$, $n \geq 1$ e lembrando que $\frac{1}{N_n} \sum_{i=1}^{N_n} \tilde{x}_i \tilde{x}'_i = A_{N_n}$ é uma subsequência de A_n , pela hipótese ii), $A_n \xrightarrow{n \rightarrow \infty} \Sigma_{xx}$, e portanto

$$A_{N_n} = \frac{1}{N_n} \sum_{i=1}^{N_n} \tilde{x}_i \tilde{x}'_i \xrightarrow{n \rightarrow \infty} \Sigma_{xx}.$$

Nestas condições, pelo mesmo teorema de Liapunov

$$\frac{1}{\sqrt{N_n}} \sum_{i=1}^{N_n} \tilde{w}_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(0, \sigma^2 \Sigma_{xx}).$$

$$\frac{1}{\sqrt{N_n}} \sum_{i=1}^{N_n} \tilde{w}_i = \frac{1}{\sqrt{N_n}} \sum_{i=1}^n \tilde{w}_i + \frac{1}{\sqrt{N_n}} \sum_{i=n+1}^{N_n} \tilde{w}_i =$$

$$\frac{\sqrt{n}}{\sqrt{N_n}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{w}_i + \frac{\sqrt{N_n - n}}{\sqrt{N_n}} \frac{1}{\sqrt{N_n - n}} \sum_{i=n+1}^{N_n} \tilde{w}_i$$

e como

$$\frac{\sqrt{n}}{\sqrt{N_n}} \frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{w}_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(0, f\sigma^2 \Sigma_{xx}), \quad \text{pela independência entre}$$

$\sum_{i=1}^n \tilde{W}_i$ e $\sum_{i=n+1}^N \tilde{W}_i$, segue que

$$\frac{\sqrt{N_n - n}}{\sqrt{N_n}} \frac{1}{\sqrt{N_n - n}} \sum_{i=n+1}^{N_n} \tilde{W}_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \Sigma),$$

$$\Sigma = \sigma^2 \Sigma_{xx} - f \sigma^2 \Sigma_{xx} = (1-f) \sigma^2 \Sigma_{xx}.$$

Com isto,

$$\frac{\sqrt{N_n}}{\sqrt{N_n - n}} \frac{\sqrt{N_n - n}}{\sqrt{N_n}} \frac{1}{\sqrt{N_n - n}} \sum_{i=n+1}^{N_n} \tilde{W}_i = \frac{1}{\sqrt{N_n - n}} \sum_{i=n+1}^{N_n} \tilde{W}_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \sigma^2 \Sigma_{xx}),$$

porque $\sqrt{\frac{N_n}{N_n - n}} \xrightarrow[n \rightarrow \infty]{} \frac{1}{\sqrt{1-f}}.$

Portanto,

$$\underline{U}_n = (N_n - n)^{-1/2} \begin{bmatrix} \sum_{i=n+1}^{N_n} \varepsilon_i \\ \sum_{i=n+1}^n X_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=n+1}^n X_{ip} \varepsilon_i \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \sigma^2 \Sigma_{xx}).$$

$\underline{Z}_n$ e $\underline{U}_n$ são independentes $\forall n$ e então, pelo teorema A.4.10,

$$\begin{bmatrix} \underline{Z}_n \\ \underline{U}_n \end{bmatrix} \xrightarrow{n \rightarrow \infty} N_{2p+2} \left(\underline{0}, \begin{bmatrix} \sigma^2 \underline{\Sigma}_{xx} & \underline{0} \\ \underline{0} & \sigma^2 \underline{\Sigma}_{xx} \end{bmatrix} \right).$$

Temos que

$$\begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{B}} - \underline{B} \end{bmatrix} = \begin{bmatrix} \left(\frac{\underline{X}'_s \underline{X}_s}{n} \right)^{-1} \underline{R}_n \\ \left(\frac{\underline{X}'_s \underline{X}_s}{n} \right)^{-1} \underline{R}_n - \left(\frac{\underline{X}' \underline{X}}{N_n} \right)^{-1} \underline{R}_{N_n} \end{bmatrix} =$$

$$= \begin{bmatrix} \left(\frac{\underline{X}'_s \underline{X}_s}{n} \right)^{-1} & \underline{0} \\ \left(\frac{\underline{X}'_s \underline{X}_s}{n} \right)^{-1} & - \left(\frac{\underline{X}' \underline{X}}{N_n} \right)^{-1} \end{bmatrix} \begin{bmatrix} \underline{R}_n \\ \underline{R}_{N_n} \end{bmatrix}.$$

Como consequência de ii) e do teorema A.4.7,

$$\left(\frac{\underline{X}'_s \underline{X}_s}{n} \right)^{-1} \xrightarrow{n \rightarrow \infty} \underline{\Sigma}_{xx}^{-1} \quad \text{e}$$

$$\left(\frac{\underline{X}' \underline{X}}{N_n} \right)^{-1} = \left(\frac{\sum_{i=1}^N \underline{X}_i \underline{X}'_i}{N_n} \right)^{-1} \xrightarrow{n \rightarrow \infty} \underline{\Sigma}_{xx}^{-1}.$$

Desta forma, a distribuição limite de $\sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{B}} - \underline{B} \end{bmatrix}$ é a mesma de

$$\underline{Y}_n = \sqrt{n} \begin{bmatrix} \underline{\Sigma}_{xx}^{-1} & \underline{0} \\ \underline{\Sigma}_{xx}^{-1} & -\underline{\Sigma}_{xx}^{-1} \end{bmatrix} \begin{bmatrix} \underline{R}_n \\ \underline{R}_{N_n} \end{bmatrix} = \sqrt{n} \begin{bmatrix} \underline{\Sigma}_{xx}^{-1} \underline{R}_n \\ \underline{\Sigma}_{xx}^{-1} (\underline{R}_n - \underline{R}_{N_n}) \end{bmatrix}.$$

Devido a (A.4.3),

$$\sqrt{n} (\tilde{R}_n - \tilde{R}_{N_n}) = (1 - f_n) \tilde{Z}_n - f_n^{1/2} (1 - f_n)^{1/2} \tilde{U}_n$$

e assim,

$$\begin{aligned} \tilde{Y}_n &= \begin{bmatrix} \Sigma_{xx}^{-1} \tilde{Z}_n \\ (1 - f_n) \Sigma_{xx}^{-1} \tilde{Z}_n - f_n^{1/2} (1 - f_n)^{1/2} \Sigma_{xx}^{-1} \tilde{U}_n \end{bmatrix} = \\ &= \begin{bmatrix} \Sigma_{xx}^{-1} & 0 \\ (1 - f_n) \Sigma_{xx}^{-1} & - f_n^{1/2} (1 - f_n)^{1/2} \Sigma_{xx}^{-1} \end{bmatrix} \begin{bmatrix} \tilde{Z}_n \\ \tilde{U}_n \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma), \end{aligned}$$

onde

$$\begin{aligned} \Gamma &= A \begin{bmatrix} \sigma^2 \Sigma_{xx} & 0 \\ 0 & \sigma^2 \Sigma_{xx} \end{bmatrix} A' \quad e \\ A &= \begin{bmatrix} \Sigma_{xx}^{-1} & 0 \\ (1-f) \Sigma_{xx}^{-1} & - f^{1/2} (1-f)^{1/2} \Sigma_{xx}^{-1} \end{bmatrix}. \end{aligned}$$

Efetuada os cálculos, verificamos que

$$\Gamma = \begin{bmatrix} \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \\ (1-f) \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix},$$

o que encerra a demonstração.

□

A prova do teorema 4.1 utilizou o teorema do limite central multivariado clássico porque admitia-se que os vetores

aleatórios $\underline{X}'_i \varepsilon_i = [\varepsilon_i, x_{i1}\varepsilon_i, x_{i2}\varepsilon_i \dots x_{ip}\varepsilon_i]$, $i = 1, 2 \dots N_n$ eram identicamente distribuídos.

Sob o modelo (4.7), $x_{i1}, x_{i2} \dots x_{ip}$ não são aleatórias, mas variam para os diferentes valores de i e como consequência, os vetores $\underline{X}'_i \varepsilon_i$, $i = 1, 2 \dots N_n$ não são mais identicamente distribuídos. Por este motivo, foi necessário utilizar um teorema do limite central multivariado para vetores aleatórios independentes não identicamente distribuídos e estabelecer a condição de Liapunov.

Agora, com o auxílio do teorema de Lindeberg-Feller multivariado, teorema A.4.12, apêndice 4, determinaremos a distribuição assintótica de $\sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{B}} - \underline{B} \end{bmatrix}$, sem exigir a condição

de Liapunov.

Este novo teorema, dada sua maior generalidade, será mais forte que o anterior.

Teorema 4.3

No modelo (4.7), para $n \rightarrow \infty$ e $\lim_{n \rightarrow \infty} \frac{n}{N_n} = f$, se

$$i) |x_{ij}| \leq c, \forall j = 1, 2 \dots p, i = 1, 2 \dots N_n, c \in \mathbb{R}, c > 0$$

$$ii) \lim_{n \rightarrow \infty} \frac{\sum_{i=1}^n \underline{X}_i \underline{X}'_i}{n} = \lim_{n \rightarrow \infty} \frac{1}{n} \underline{X}'_s \underline{X}_s = \underline{\Sigma}_{xx},$$

então

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma),$$

onde

$$\Gamma = \begin{bmatrix} \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \\ (1-f) \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix}.$$

Prova:

Os vetores aleatórios $\underline{W}'_i = \underline{X}'_i \underline{\varepsilon}_i = [\varepsilon_i, x_{i1}\varepsilon_i, x_{i2}\varepsilon_i, \dots, x_{ip}\varepsilon_i]$, $i = 1, 2, \dots, N_n$ são independentes com média $E(\underline{W}_i) = \underline{0}$ e $\text{Var}(\underline{W}_i) = \underline{X}_i \underline{X}'_i \sigma^2$.

Se $F_{\underline{W}_i}$ é a função distribuição de $\underline{W}_i$, $i = 1, 2, \dots, N_n$,

$$\int_{\|\underline{W}\| > \varepsilon \sqrt{n}} \|\underline{W}\|^2 dF_{\underline{W}_i}(\underline{W}) = E\left(\|\underline{W}_i\|^2 1_{(\|\underline{W}_i\| > \varepsilon \sqrt{n})}\right),$$

$$\text{onde } 1_{(\|\underline{W}_i\| > \varepsilon \sqrt{n})} = \begin{cases} 1 & \text{se } \|\underline{W}_i\| > \varepsilon \sqrt{n} \\ 0 & \text{caso contrário} \end{cases}$$

$$\text{e } \|\underline{W}_i\| = \left(\sum_{j=0}^p x_{ij}^2 \varepsilon_i^2 \right)^{1/2}.$$

Nestas condições,

$$\int_{\|\tilde{w}\| > \varepsilon\sqrt{n}} \|\tilde{w}\|^2 dF_{\tilde{w}_i}(\tilde{w}) = E \left(\sum_{j=0}^p x_{ij}^2 \varepsilon_i^2 \cdot 1 \left(\left| \sum_{j=0}^p x_{ij}^2 \varepsilon_i^2 \right|^{1/2} > \varepsilon\sqrt{n} \right) \right)$$

$$\leq (1 + pc^2) E \left(\varepsilon_i^2 \cdot 1 \left(|\varepsilon_i| > \frac{\varepsilon\sqrt{n}}{\left(\sum_{j=0}^p x_{ij}^2 \right)^{1/2}} \right) \right).$$

Temos que

$$\sum_{j=0}^p x_{ij}^2 \leq 1 + pc^2 \implies \left(\sum_{j=0}^p x_{ij}^2 \right)^{1/2} \leq \sqrt{1+pc^2}$$

$$\implies \frac{\varepsilon\sqrt{n}}{\left(\sum_{j=0}^p x_{ij}^2 \right)^{1/2}} \geq \frac{\varepsilon\sqrt{n}}{\sqrt{1+pc^2}}$$

$$\implies \left\{ \varepsilon_i \mid |\varepsilon_i| > \frac{\varepsilon\sqrt{n}}{\left(\sum_{j=0}^p x_{ij}^2 \right)^{1/2}} \right\} \subset \left\{ \varepsilon_i \mid |\varepsilon_i| > \frac{\varepsilon\sqrt{n}}{\sqrt{1+pc^2}} \right\}$$

e assim,

$$\int_{\|\tilde{w}\| > \varepsilon\sqrt{n}} \|\tilde{w}\|^2 dF_{\tilde{w}_i}(\tilde{w}) \leq (1+pc^2) E \left(\varepsilon_i^2 \cdot 1 \left(|\varepsilon_i| > \frac{\varepsilon\sqrt{n}}{\sqrt{1+pc^2}} \right) \right), \text{ para } i=1,2,\dots$$

Como consequência,

$$\begin{aligned} \frac{1}{n} \sum_{i=1}^n \int_{\|\tilde{w}\| > \varepsilon \sqrt{n}} \|\tilde{w}\|^2 dF_{\tilde{w}_i}(\tilde{w}) &\leq \frac{1}{n} \sum_{i=1}^n (1+pc^2) E \left(\varepsilon_i^2 1 \left(|\varepsilon_i| > \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \right) \right) \\ &\leq \frac{1+pc^2}{n} \sum_{i=1}^n E \left(\varepsilon_i^2 1 \left(|\varepsilon_i| > \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \right) \right). \end{aligned}$$

Como as variáveis ε_i , $i = 1, 2, \dots$ são identicamente distribuídas,

$$\sum_{i=1}^n E \left(\varepsilon_i^2 1 \left(|\varepsilon_i| > \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \right) \right) = n E \left(\varepsilon_1^2 1 \left(|\varepsilon_1| > \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \right) \right)$$

e portanto

$$\frac{1}{n} \sum_{i=1}^n \int_{\|\tilde{w}\| > \varepsilon \sqrt{n}} \|\tilde{w}\|^2 dF_{\tilde{w}_i}(\tilde{w}) \leq (1+pc^2) E \left(\varepsilon_1^2 1 \left(|\varepsilon_1| > \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \right) \right) \xrightarrow{n \rightarrow \infty} 0,$$

pois

$$E \left(\varepsilon_1^2 \cdot 1 \left(\left| \varepsilon_1 \right| > \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \right) \right) \leq E(\varepsilon_1^2) = \sigma^2$$

$$\text{e } \frac{\varepsilon \sqrt{n}}{\sqrt{1+pc^2}} \xrightarrow{n \rightarrow \infty} \infty .$$

Verificamos então que é válida a condição (A.4.4) do teorema A.4.12 e com isto,

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n \tilde{W}_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \Sigma),$$

$$\text{onde } \Sigma = \lim_{n \rightarrow \infty} \frac{\sigma^2 \sum_{i=1}^n \tilde{X}_i \tilde{X}_i'}{n} = \sigma^2 \Sigma_{xx} .$$

De maneira análoga, mostraríamos que

$$\frac{1}{\sqrt{N_n}} \sum_{i=1}^{N_n} \tilde{W}_i \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \sigma^2 \Sigma_{xx})$$

e repetindo a demonstração do teorema 4.2 a partir deste ponto, provamos que

$$\sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{B}} - \underline{B} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, \Gamma) ,$$

$$\Gamma = \begin{bmatrix} \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \\ (1-f) \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix}.$$

□

Como consequência do teorema 4.3, segue que

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \sigma^2 \Sigma_{xx}^{-1}) \quad e$$

$$\sqrt{n}(\hat{\beta} - \underline{B}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, (1-f) \sigma^2 \Sigma_{xx}^{-1}).$$

A distribuição assintótica presente nos teoremas 4.2 e 4.3 é análoga à obtida no teorema 4.1, que tratava do caso em que X é aleatória.

No teorema 4.1, os elementos que ocupam as posições r e s das matrizes Q e G são respectivamente

$$q_{rs} = E(x_{ir} x_{is}) \quad e \quad g_{rs} = E(x_{ir} x_{is} \varepsilon_i^2),$$

de modo que, quando X é fixa, estas quantidades ficam

$$q_{rs} = x_{ir} x_{is} \quad e \quad g_{rs} = x_{ir} x_{is} E(\varepsilon_i^2) = x_{ir} x_{is} \sigma^2.$$

Desta forma,

$$Q^{-1}G = \sigma^2 I_{p+1}.$$

Quando X é aleatória, $Q_n \xrightarrow{P} Q$ (ver apêndice 4) e para X fixa, Q_n corresponde a $\frac{1}{n} X'_s X_s \xrightarrow{n \rightarrow \infty} \sum_{xx}$.

Portanto, para X fixa, $Q = \sum_{xx}$ e assim,

$$Q^{-1}GQ^{-1} = \sigma^2 \sum_{xx}^{-1}.$$

Estas considerações mostram que quando X é fixa, as distribuições limites dos teoremas 4.2 e 4.3 poderiam ser obtidas a partir do correspondente teorema para X aleatória.

Finalizando esta seção, obteremos um resultado semelhante para o caso em que $\text{Var}(\varepsilon_i) = v_i \sigma^2$, $i = 1, 2, \dots, N_n$.

Consideraremos o modelo de superpopulação

$$y = X\beta + \varepsilon \quad (4.8),$$

onde

$$X = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p} \\ 1 & x_{21} & x_{22} & & x_{2p} \\ \vdots & & & & \\ 1 & x_{N1} & x_{N2} & & x_{Np} \end{bmatrix} \text{ é a matriz não aleatória de constan}$$

tes conhecidas,

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_N \end{bmatrix} \quad \text{e} \quad \underline{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_N \end{bmatrix} \quad \text{são vetores aleatórios com}$$

$$i) E(\underline{\varepsilon}) = \underline{0}$$

e

$$\text{Var}(\underline{\varepsilon}) = \sigma^2 \begin{bmatrix} v_1 & 0 & \dots & 0 \\ 0 & v_2 & & 0 \\ \vdots & & & \\ 0 & 0 & & v_N \end{bmatrix} = \sigma^2 \text{diag}(v_1, v_2 \dots v_N),$$

$v_1, v_2 \dots v_N$ constantes estritamente positivas, conhecidas

ii) $Y_1, Y_2 \dots Y_N$ são variáveis aleatórias independentes e

$\frac{\varepsilon_1}{\sqrt{v_1}}, \frac{\varepsilon_2}{\sqrt{v_2}} \dots \frac{\varepsilon_N}{\sqrt{v_N}}$ são independentes e identicamente distribuídas.

Além disso, sejam

β o parâmetro populacional ou superparâmetro,

$\hat{\beta} = (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} Y_s$, o estimador de mínimos quadrados ponderados de β , onde V_s é a parte da matriz $V = \text{diag}(v_1, v_2 \dots v_N)$ correspondendo aos elementos da amostra, já definida e utilizada nos capítulos anteriores, e

$$\underline{\underline{B}} = (X' V^{-1} X)^{-1} X' V^{-1} Y .$$

$\hat{\underline{\underline{\beta}}}$ e $\underline{\underline{B}}$ podem ser escritos alternativamente como

$$\hat{\underline{\underline{\beta}}} = \left(\sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \sum_{i=1}^n \underline{\underline{X}}_i \frac{Y_i}{\underline{\underline{V}}_i} \quad e$$

$$\underline{\underline{B}} = \left(\sum_{i=1}^N \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \sum_{i=1}^N \underline{\underline{X}}_i \frac{Y_i}{\underline{\underline{V}}_i} .$$

Observamos ainda que

$$\begin{aligned} \hat{\underline{\underline{\beta}}} - \underline{\underline{\beta}} &= \left(\sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \sum_{i=1}^n \frac{\underline{\underline{X}}_i (\underline{\underline{X}}_i' \underline{\underline{\beta}} + \varepsilon_i)}{\underline{\underline{V}}_i} - \underline{\underline{\beta}} = \\ &= \left(\sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \underline{\underline{\beta}} + \left(\sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \sum_{i=1}^n \frac{\underline{\underline{X}}_i \varepsilon_i}{\underline{\underline{V}}_i} - \underline{\underline{\beta}} \\ &= \left(\sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \sum_{i=1}^n \frac{\underline{\underline{X}}_i \varepsilon_i}{\underline{\underline{V}}_i} = n \left(\sum_{i=1}^n \frac{\underline{\underline{X}}_i \underline{\underline{X}}_i'}{\underline{\underline{V}}_i} \right)^{-1} \underline{\underline{R}}_n, \end{aligned}$$

onde

$$\underline{\underline{R}}_n = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n \frac{\varepsilon_i}{\underline{\underline{V}}_i} \\ \sum_{i=1}^n \frac{\underline{\underline{X}}_{i1} \varepsilon_i}{\underline{\underline{V}}_i} \\ \vdots \\ \sum_{i=1}^n \frac{\underline{\underline{X}}_{ip} \varepsilon_i}{\underline{\underline{V}}_i} \end{bmatrix} .$$

Analogamente,

$$\underline{\hat{B}} - \underline{\beta} = N_n \left(\sum_{i=1}^N \frac{\underline{\hat{X}}_i \underline{\hat{X}}_i'}{V_i} \right)^{-1} \underline{R}_{N_n}, \quad \text{com}$$

$$\underline{R}_{N_n} = \frac{1}{N_n} \begin{bmatrix} \sum_{i=1}^N \frac{\varepsilon_i}{V_i} \\ \sum_{i=1}^N \frac{X_{i1} \varepsilon_i}{V_i} \\ \vdots \\ \sum_{i=1}^N \frac{X_{ip} \varepsilon_i}{V_i} \end{bmatrix}, \quad \text{e assim,}$$

$$\underline{\hat{\beta}} - \underline{\beta} = \underline{\hat{\beta}} - \underline{\beta} - (\underline{\hat{B}} - \underline{\beta}) = n \left(\sum_{i=1}^n \frac{\underline{\hat{X}}_i \underline{\hat{X}}_i'}{V_i} \right)^{-1} \underline{R}_n - N_n \left(\sum_{i=1}^N \frac{\underline{\hat{X}}_i \underline{\hat{X}}_i'}{V_i} \right)^{-1} \underline{R}_{N_n}.$$

Teorema 4.4

No modelo (4.8), para $n \rightarrow \infty$ e $\lim_{n \rightarrow \infty} \frac{n}{N_n} = f$, se

$$\text{i)} \quad \frac{|\underline{X}_{ij}|}{\sqrt{V_i}} \leq c, \quad j = 0, 1, 2, \dots, p, \quad i = 1, 2, \dots, N_n, \quad c \in \mathbb{R}, \quad c > 0$$

$$\text{ii)} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\underline{X}_i \underline{X}_i'}{V_i} = \Sigma_{xx},$$

então

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{2p+2}(\underline{0}, D) ,$$

$$\text{com } D = \begin{bmatrix} \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \\ (1-f) \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix} .$$

□

A prova deste teorema é idêntica à do teorema 4.3 para $\tilde{W}'_i = \tilde{X}'_i \frac{\varepsilon_i}{V_i}$ e por este motivo foi omitida.

Devido ao teorema 4.4, temos que

$$\sqrt{n}(\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, \sigma^2 \Sigma_{xx}^{-1}) \quad \text{e}$$

$$\sqrt{n}(\hat{B} - B) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, (1-f) \sigma^2 \Sigma_{xx}^{-1}) ,$$

resultados obtidos por Bolfarine e Zacks [1992], supondo a condição de Liapunov.

Corolário 4.2

Nas condições do teorema 4.4 e se adicionalmente

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{\underline{1}' X}{N_n} &= \lim_{n \rightarrow \infty} \begin{bmatrix} 1 & \frac{\sum_{i=1}^N x_{i1}}{N_n} & \frac{\sum_{i=1}^N x_{i2}}{N_n} & \dots & \frac{\sum_{i=1}^N x_{ip}}{N_n} \end{bmatrix} \\ &= [1 \quad \bar{X}_1 \quad \bar{X}_2 \quad \dots \quad \bar{X}_p] = g, \end{aligned}$$

então

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{\beta}}{N_n} \\ \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{\beta}}{N_n} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(\underline{0}, D) ,$$

$$D = \begin{bmatrix} \sigma^2 \underline{g} \Sigma_{xx}^{-1} \underline{g}' & (1-f) \sigma^2 \underline{g} \Sigma_{xx}^{-1} \underline{g}' \\ (1-f) \sigma^2 \underline{g} \Sigma_{xx}^{-1} \underline{g}' & (1-f) \sigma^2 \underline{g} \Sigma_{xx}^{-1} \underline{g}' \end{bmatrix} .$$

□

4.3 - Normalidade Assintótica para a Média Populacional

Uma importante aplicação do corolário 4.2 surge quando é válida a condição L, introduzida no capítulo 2. Esta condição está satisfeita se existir um vetor de constantes $\underline{\delta}$, (p+1) dimensional tal que $V \underline{1} = X \underline{\delta}$, onde $V = \text{Var}(\underline{y})$ é a matriz de covariância associada ao modelo de superpopulação adotado e $\underline{1} = \underline{1}_N$ é o vetor N-dimensional contendo 1 em todas as posições.

Conforme verificado anteriormente, $V \underline{1} = X \underline{\delta}$ implica que

$$\underline{1}' X \underline{\beta} = \underline{1}' \underline{y} = \sum_{i=1}^N y_i$$

e

$$\frac{\underline{1}' X \underline{\beta}}{N_n} = \frac{\sum_{i=1}^N Y_i}{N_n} = \bar{Y}_{N_n}, \text{ a média populacional.}$$

No modelo (4.7), $V = \sigma^2 I_N$ e a condição L está satisfeita pois, tomando

$$\underline{\delta} = \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(p+1) \times 1},$$

temos

$$X \underline{\delta} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1p} \\ 1 & x_{21} & & x_{2p} \\ \vdots & & & \\ 1 & x_{N1} & & x_{Np} \end{bmatrix} \begin{bmatrix} \sigma^2 \\ 0 \\ \vdots \\ 0 \end{bmatrix} = \sigma^2 \underline{1} = V \underline{1}.$$

Neste caso, o corolário 4.2 forneceria a distribuição assintótica de

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' X \underline{\beta}}{N_n} \\ \frac{\underline{1}' X \hat{\underline{\beta}}}{N_n} - \bar{Y}_{N_n} \end{bmatrix}.$$

Podem ainda ocorrer várias outras situações sob o modelo (4.8) onde $V = \sigma^2 \text{diag}(v_1, v_2, \dots, v_N)$ satisfaz à condição L. O exemplo a seguir ilustra uma destas situações.

Exemplo 4.1

Consideremos o modelo de superpopulação

$$y_i = \beta x_i + \varepsilon_i,$$

onde

$$x_i > 0, \quad E(\varepsilon_i) = 0, \quad \text{Var}(\varepsilon_i) = \sigma^2 x_i \quad \text{e}$$

$\frac{\varepsilon_i}{\sqrt{x_i}}$, $i = 1, 2, \dots, N_n$ são variáveis aleatórias independentes e identicamente distribuídas.

Sob este modelo,

$$\hat{\beta} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}, \quad B = \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i}, \quad v_i = x_i \quad \text{e} \quad \tilde{x}_i = [x_i].$$

$$\text{Se } \frac{|x_i|}{\sqrt{v_i}} = \sqrt{x_i} \leq c \quad \text{para } i = 1, 2, \dots, N_n \quad \text{e}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \frac{\tilde{x}_i \tilde{x}_i'}{v_i} = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} = \sum_{xx},$$

estão verificadas as hipóteses do teorema 4.4 e, desta forma

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\beta} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(0, D),$$

$$D = \begin{bmatrix} \frac{\sigma^2}{\bar{X}} & \frac{(1-f)\sigma^2}{\bar{X}} \\ \frac{(1-f)\sigma^2}{\bar{X}} & \frac{(1-f)\sigma^2}{\bar{X}} \end{bmatrix}.$$

Observamos ainda que a condição L está satisfeita, pois $V = \text{Var}(y) = \sigma^2 \text{diag}(x_1, x_2, \dots, x_N)$ e

$$V\tilde{1} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = X\tilde{\delta}, \text{ para } \tilde{\delta} = [1].$$

Com isto,

$$\frac{\tilde{1}' X \tilde{B}}{N_n} = \frac{1}{N_n} \sum_{i=1}^{N_n} x_i \frac{\sum_{i=1}^N y_i}{\sum_{i=1}^N x_i} = \frac{\sum_{i=1}^N y_i}{N_n} = \bar{y}_{N_n},$$

$$\frac{\tilde{1}' X \hat{\beta}}{N_n} = \frac{\sum_{i=1}^N x_i}{N_n} \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}, \quad \frac{\tilde{1}' X \beta}{N_n} = \frac{\beta \sum_{i=1}^N x_i}{N_n}$$

e

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n x_i = \bar{X} \text{ implica que } \lim_{n \rightarrow \infty} \frac{1}{N_n} \sum_{i=1}^N x_i = \bar{X}.$$

Portanto, pelo corolário 4.2,

$$\sqrt{n} \begin{bmatrix} \frac{\underline{1}' \underline{X} \hat{\underline{\beta}}}{N_n} - \frac{\underline{1}' \underline{X} \underline{\beta}}{N_n} \\ \frac{\underline{1}' \underline{X} \hat{\underline{\beta}}}{N_n} - \bar{Y}_{N_n} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_2(\underline{0}, D) \quad ,$$

$$D = \begin{bmatrix} \sigma^2 \bar{X} & (1-f) \sigma^2 \bar{X} \\ (1-f) \sigma^2 \bar{X} & (1-f) \sigma^2 \bar{X} \end{bmatrix} .$$

O previsor $\underline{1}' \underline{X} \hat{\underline{\beta}} = \sum_{i=1}^N x_i \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$ é conhecido na literatura

como previsor razão. Através do teorema A.2.1, verifica-se que este é o previsor linear ótimo para o total $T = \sum_{i=1}^N y_i$ no modelo considerado.

□

4.4 - Aplicações

Os teoremas obtidos neste capítulo abordam simultaneamente estimação (de $\underline{\beta}$) e previsão (de $\underline{B}$) e seu principal interesse reside na possibilidade de inferência conjunta sobre $\underline{\beta}$ e $\underline{B}$ quando a distribuição de $\underline{y}$ não está especificada.

Este tipo de inferência conjunta é muito comum em trabalhos como Royall [1976b] e Rodrigues [1990]. Em ambos, os autores

partem de modelos onde y tem distribuição de probabilidades conhecida (normal N-variada), fazendo com que as distribuições conjuntas obtidas sejam exatas.

Nossos resultados permitem uma inferência assintótica conjunta para o caso em que a distribuição de y não está especificada, conforme poderá ser verificado nas aplicações a seguir.

Nestas aplicações, vamos nos fixar no modelo (4.7), onde X é matriz não aleatória e $\text{Var}(\underline{\varepsilon}) = \sigma^2 I_N$ e utilizar o resultado apresentado no teorema 4.3. Satisfeitas as condições deste teorema,

$$\sqrt{n} \begin{bmatrix} \hat{\underline{\beta}} - \underline{\beta} \\ \hat{\underline{B}} - \underline{B} \end{bmatrix} \xrightarrow[n \rightarrow \infty]{D} N_{2p+2}(\underline{0}, \Sigma) \quad (4.9),$$

com

$$\Sigma = \begin{bmatrix} \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \\ (1-f) \sigma^2 \Sigma_{xx}^{-1} & (1-f) \sigma^2 \Sigma_{xx}^{-1} \end{bmatrix}.$$

Aplicação 1: Região de Confiança para $\underline{\beta}$ e $\underline{B}$

Se $\underline{Z}$ é um vetor aleatório m-dimensional tal que $\underline{Z} \sim N_m(\underline{\mu}, T)$, então a variável aleatória $\underline{Z}' T^{-1} \underline{Z}$ tem distribuição qui-quadrado

não central com m graus de liberdade e parâmetro de não centralidade $\nu' T^{-1} \nu$.

Como consequência deste resultado e de (4.9),

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta & \hat{\beta} - B \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\beta} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{D} \chi^2_{2p+2}(0),$$

onde $\chi^2_{2p+2}(0)$ representa uma variável aleatória com distribuição qui-quadrado central com $2p+2$ graus de liberdade.

Assim, obtido $\hat{\beta}$, uma região de confiança simultânea para β e B com coeficiente aproximadamente $1-\alpha$ será

$$\left\{ (\beta, B) \in \mathbb{R}^{2p+2} \mid \sqrt{n} \begin{bmatrix} \hat{\beta} - \beta & \hat{\beta} - B \end{bmatrix} \Sigma^{-1} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{\beta} - B \end{bmatrix} \leq x^2_{2p+2, 1-\alpha}(0) \right\},$$

onde $x^2_{2p+2, 1-\alpha}(0)$ é o quantil de ordem $1-\alpha$ da distribuição qui-quadrado central com $2p+2$ graus de liberdade.

Esta região é um elipsóide de centro $(\hat{\beta}, \hat{\beta})$. Lembramos que aqui $\hat{\beta}$ funciona tanto como estimador de β quanto como previsor de B .

Aplicação 2: Distribuição limite de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ B - \beta \end{bmatrix}$

Temos que

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} = \sqrt{n} \begin{bmatrix} I_{p+1} & 0 \\ I_{p+1} & -I_{p+1} \end{bmatrix} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} \quad e$$

$$\begin{bmatrix} I_{p+1} & 0 \\ I_{p+1} & -I_{p+1} \end{bmatrix} \Sigma \begin{bmatrix} I_{p+1} & I_{p+1} \\ 0 & -I_{p+1} \end{bmatrix} = \sigma^2 \begin{bmatrix} \Sigma_{xx}^{-1} & f\Sigma_{xx}^{-1} \\ f\Sigma_{xx}^{-1} & f\Sigma_{xx}^{-1} \end{bmatrix}$$

e assim,

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix} \xrightarrow[n \rightarrow \infty]{D} N_{2p+2} \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \sigma^2 \begin{bmatrix} \Sigma_{xx}^{-1} & f\Sigma_{xx}^{-1} \\ f\Sigma_{xx}^{-1} & f\Sigma_{xx}^{-1} \end{bmatrix} \right) \quad (4.10).$$

Vamos agora comparar esta distribuição assintótica com a correspondente distribuição exata quando y tem distribuição normal multivariada.

Para $V = \text{Var}(\xi) = \text{Var}(y) = \sigma^2 I_N$,

$$\hat{\beta} = (X'_s X_s)^{-1} X'_s y_s$$

e

$$\begin{aligned} \hat{B} &= (X'X)^{-1} X' y = (X'X)^{-1} \begin{bmatrix} X'_s & X'_r \end{bmatrix} \begin{bmatrix} y_s \\ y_r \end{bmatrix} = \\ &= (X'X)^{-1} X'_s y_s + (X'X)^{-1} X'_r y_r. \end{aligned}$$

Se adicionalmente $y = \begin{bmatrix} y_s \\ y_r \end{bmatrix} \sim N_N(X\beta, \sigma^2 I_N)$, então

$$\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \tilde{B} - \beta \end{bmatrix} = \sqrt{n} \left\{ \begin{bmatrix} (X'_s X_s)^{-1} X'_s & \mathbf{0} \\ (X'X)^{-1} X'_s & (X'X)^{-1} X'_r \end{bmatrix} \begin{bmatrix} Y_s \\ Y_r \end{bmatrix} - \begin{bmatrix} \beta \\ \beta \end{bmatrix} \right\} \sim N_{2p+2}(\underline{\mu}, \tau)$$

onde

$$\underline{\mu} = \sqrt{n} \left\{ \begin{bmatrix} (X'_s X_s)^{-1} X'_s & \mathbf{0} \\ (X'X)^{-1} X'_s & (X'X)^{-1} X'_r \end{bmatrix} \begin{bmatrix} X_s \beta \\ X_r \beta \end{bmatrix} - \begin{bmatrix} \beta \\ \beta \end{bmatrix} \right\} = \begin{bmatrix} \underline{0} \\ \underline{0} \end{bmatrix}$$

pois

$$\begin{aligned} (X'X)^{-1} X'_s X_s \beta + (X'X)^{-1} X'_r X_r \beta &= (X'X)^{-1} (X'_s X_s + X'_r X_r) \beta \\ &= (X'X)^{-1} X'X \beta = \beta \end{aligned}$$

e

$$\tau = n\sigma^2 \begin{bmatrix} (X'_s X_s)^{-1} X'_s & \mathbf{0} \\ (X'X)^{-1} X'_s & (X'X)^{-1} X'_r \end{bmatrix} \begin{bmatrix} X_s (X'_s X_s)^{-1} & X_s (X'X)^{-1} \\ \mathbf{0} & X_r (X'X)^{-1} \end{bmatrix}$$

$$\begin{aligned} &= n\sigma^2 \begin{bmatrix} (X'_s X_s)^{-1} & (X'X)^{-1} \\ (X'X)^{-1} & (X'X)^{-1} (X'_s X_s + X'_r X_r) (X'X)^{-1} \end{bmatrix} = \\ &= n\sigma^2 \begin{bmatrix} (X'_s X_s)^{-1} & (X'X)^{-1} \\ (X'X)^{-1} & (X'X)^{-1} \end{bmatrix}. \end{aligned} \tag{4.11}$$

Através de (4.11), verificamos que

$$\text{Var}(\hat{\beta}) = \sigma^2 (X'_s X_s)^{-1},$$

$$\text{Var}(\underline{B}) = \sigma^2 (X' X)^{-1} \text{ e}$$

$$\text{Cov}(\hat{\beta}, \underline{B}) = \sigma^2 (X' X)^{-1},$$

resultados estes coerentes com os apresentados na seção 2.3 do capítulo 2, para $V = \sigma^2 I_N$ (a seção trabalha com matriz de covariância V qualquer).

Além disso, devido às suposições do teorema 4.3,

$$\lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} \hat{\beta}) = \lim_{n \rightarrow \infty} n (X'_s X_s)^{-1} \sigma^2 = \sigma^2 \Sigma_{xx}^{-1}$$

e

$$\begin{aligned} \lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} \underline{B}) &= \lim_{n \rightarrow \infty} \left(\frac{n}{N_n} N_n (X' X)^{-1} \sigma^2 \right) = f \sigma^2 \Sigma_{xx}^{-1} \\ &= \lim_{n \rightarrow \infty} \text{Cov}(\sqrt{n} \hat{\beta}, \sqrt{n} \underline{B}) \end{aligned}$$

e assim verificamos que a matriz de covariância da

distribuição limite de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \underline{B} - \beta \end{bmatrix}$ é o limite da matriz de

covariância de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \underline{B} - \beta \end{bmatrix}$.

Nota-se então uma perfeita analogia entre a distribuição exata de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix}$ quando y tem distribuição normal N-variada

e a distribuição assintótica quando y tem distribuição qualquer, satisfeitas as condições do teorema 4.3.

Isto sugere que, para grandes tamanhos de amostra (e da população), a inferência assintótica para B é equivalente à exata quando y tem distribuição normal N-variada.

Aplicação 3: Previsão de máxima verossimilhança para B quando a distribuição de probabilidades de y não está especificada

No capítulo 3 apresentamos uma teoria de verossimilhança preditiva que permitiu, entre outras coisas, obtermos o preditor de máxima verossimilhança e a função de verossimilhança preditiva para $B = (X'V^{-1}X)^{-1}X'V^{-1}y$ quando a distribuição de probabilidades de $y = (y_s, y_r)$ era conhecida, normal N-variada com média $X\beta$ e matriz de covariância V qualquer.

Os passos básicos do procedimento eram a construção da densidade conjunta $q_{\beta}(y_s, y_r)$ e a determinação da "profile" preditiva para y_r dado y_s ,

$$L_p(y_r | y_s) = \sup_{\beta} q_{\beta}(y_s, y_r) = q_{\hat{\beta}_z}(y_s, y_r),$$

onde y_s é o vetor de dados observados, y_r o vetor aleatório não observado e $\hat{\beta}_z$ o estimador de máxima verossimilhança de β calculado com base em (y_s, y_r) .

O preditor de máxima verossimilhança do vetor aleatório não observado é o valor de y_r que maximiza $L_p(y_r | y_s)$ e o conhecimento da distribuição de probabilidades de y , $q_{\beta}(y_s, y_r)$ era imprescindível.

Veremos a seguir que o resultado do teorema 4.3 permitirá a previsão de máxima verossimilhança para β quando a distribuição de y não é especificada e as variáveis aleatórias $y_1, y_2 \dots y_N$ são independentes, de modo que $V = \sigma^2 I_N$.

Evidentemente, o enfoque será assintótico, ou seja, a abordagem se destinará a grandes tamanhos de amostra (n) e como consequência, grandes tamanhos da população (N_n).

Consideremos o modelo de superpopulação $y = X\beta + \varepsilon$ com $y_1, y_2 \dots y_N$ variáveis aleatórias independentes, $E(\varepsilon) = 0$ e $\text{Var}(\varepsilon) = \sigma^2 I_N$.

Avaliado $\hat{\beta}$ e não dispondo de $q_{\beta}(y_s, y_r)$, a distribuição limite de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ B - \beta \end{bmatrix}$ substituirá $q_{\beta}(y_s, y_r)$ para efeito de

previsão. B , sendo função dos dados não observados, exercerá o papel de y_r e $\hat{\beta}$, o de y_s . A distribuição assintótica (4.10) será o elo entre a parte observada e a não observada.

Assim, seja

$$q_{\underline{\beta}}^n(\hat{\underline{\beta}}, \underline{B}) = h(\underline{x}_1, \underline{x}_2),$$

onde

$$\underline{x}_1 = \sqrt{n}(\hat{\underline{\beta}} - \underline{\beta}), \quad \underline{x}_2 = \sqrt{n}(\underline{B} - \underline{\beta})$$

e h é a densidade da distribuição normal multivariada com média

$$\begin{bmatrix} 0 \\ 0 \end{bmatrix} \text{ e matriz de covariância } \Sigma = \sigma^2 \begin{bmatrix} \Sigma_{xx}^{-1} & f\Sigma_{xx}^{-1} \\ f\Sigma_{xx}^{-1} & f\Sigma_{xx}^{-1} \end{bmatrix}.$$

Particionando Σ^{-1} , $h(\underline{x}_1, \underline{x}_2)$ pode ser escrita como

$$\frac{1}{(2\pi)^{p+1} |\sigma^2 f \Sigma_{xx}^{-1}|^{1/2} |\sigma^2 (1-f) \Sigma_{xx}^{-1}|^{1/2}} \times \\ \times \exp \left\{ -\frac{1}{2} \left[\underline{x}_1' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} \underline{x}_1 - \underline{x}_2' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} \underline{x}_1 - \underline{x}_1' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} \underline{x}_2 \right. \right. \\ \left. \left. + \underline{x}_2' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} \underline{x}_2 + \underline{x}_2' \frac{\Sigma_{xx}}{\sigma^2 f} \underline{x}_2 \right] \right\},$$

de modo que

$$q_{\underline{\beta}}^n(\hat{\underline{\beta}}, \underline{B}) \propto \exp \left\{ -\frac{1}{2} \left[n(\hat{\underline{\beta}} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} (\hat{\underline{\beta}} - \underline{\beta}) \right. \right. \\ \left. \left. - n(\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} (\hat{\underline{\beta}} - \underline{\beta}) - n(\hat{\underline{\beta}} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2 (1-f)} (\underline{B} - \underline{\beta}) \right] \right\}$$

$$\begin{aligned}
& + n(\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\underline{B} - \underline{\beta}) + n(\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{f\sigma^2} (\underline{B} - \underline{\beta}) \Bigg] \Bigg\} \\
& = \exp \left\{ - \frac{n}{2} \left[(\hat{\underline{\beta}} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\hat{\underline{\beta}} - \underline{\beta}) \right. \right. \\
& \left. \left. - 2(\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\hat{\underline{\beta}} - \underline{\beta}) + (\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2 f(1-f)} (\underline{B} - \underline{\beta}) \right] \right\}.
\end{aligned}$$

Nestas condições, para grandes tamanhos de amostra, definiremos a "profile" preditiva de $\underline{B}$ dado $\hat{\underline{\beta}}$ como

$$L_p^n(\underline{B} \mid \hat{\underline{\beta}}) = \sup_{\underline{\beta}} q_{\underline{\beta}}^n(\hat{\underline{\beta}}, \underline{B}).$$

Maximizar $q_{\underline{\beta}}^n(\hat{\underline{\beta}}, \underline{B})$ com relação a $\underline{\beta}$ é equivalente a minimizar

$$\begin{aligned}
S &= (\hat{\underline{\beta}} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\hat{\underline{\beta}} - \underline{\beta}) - 2(\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\hat{\underline{\beta}} - \underline{\beta}) \\
&+ (\underline{B} - \underline{\beta})' \frac{\Sigma_{xx}}{\sigma^2 f(1-f)} (\underline{B} - \underline{\beta}).
\end{aligned}$$

Temos que

$$\begin{aligned} \frac{\partial S}{\partial \underline{\beta}} = & -2 \frac{\Sigma_{xx}}{\sigma^2(1-f)} \hat{\underline{\beta}} + 2 \frac{\Sigma_{xx}}{\sigma^2(1-f)} \underline{\beta} + 2 \frac{\Sigma_{xx}}{\sigma^2(1-f)} \underline{B} \\ & + 2 \frac{\Sigma_{xx}}{\sigma^2(1-f)} \hat{\underline{\beta}} - 4 \frac{\Sigma_{xx}}{\sigma^2(1-f)} \underline{\beta} - 2 \frac{\Sigma_{xx}}{\sigma^2 f(1-f)} \underline{B} + 2 \frac{\Sigma_{xx}}{\sigma^2 f(1-f)} \underline{\beta} \end{aligned}$$

e

$$\begin{aligned} \frac{\partial S}{\partial \underline{\beta}} = 0 & \implies \frac{2f - 4f + 2}{\sigma^2(1-f)f} \Sigma_{xx} \underline{\beta} = \frac{1 + 1 - 2f}{\sigma^2 f(1-f)} \Sigma_{xx} \underline{B} \\ & \implies \hat{\underline{\beta}}_z = \underline{B}, \end{aligned}$$

onde $\hat{\underline{\beta}}_z$ é o estimador de máxima verossimilhança de $\underline{\beta}$ com base em $(\hat{\underline{\beta}}, \underline{B})$.

Portanto, para n grande, fixo,

$$\begin{aligned} L_p^n(\underline{B} \mid \hat{\underline{\beta}}) &= q_{\hat{\underline{\beta}}_z}^n(\hat{\underline{\beta}}, \underline{B}) \propto \\ & \exp \left\{ -\frac{n}{2} \left[(\hat{\underline{\beta}} - \hat{\underline{\beta}}_z)' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\hat{\underline{\beta}} - \hat{\underline{\beta}}_z) \right. \right. \\ & \quad \left. \left. - 2(\underline{B} - \hat{\underline{\beta}}_z)' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\hat{\underline{\beta}} - \hat{\underline{\beta}}_z) + (\underline{B} - \hat{\underline{\beta}}_z)' \frac{\Sigma_{xx}}{\sigma^2 f(1-f)} (\underline{B} - \hat{\underline{\beta}}_z) \right] \right\} \\ &= \exp \left\{ -\frac{n}{2} \left[(\underline{B} - \hat{\underline{\beta}})' \frac{\Sigma_{xx}}{\sigma^2(1-f)} (\underline{B} - \hat{\underline{\beta}}) \right] \right\} \end{aligned}$$

é a "profile" preditiva para $\underline{B}$ dado $\hat{\underline{\beta}}$.

A partir de $L_p^n(B|\hat{\beta})$ podemos obter o previsor de máxima verossimilhança de B que, sendo a moda desta distribuição, é igual a $\hat{\beta}$.

Verificamos ainda que $L_p^n(B|\hat{\beta})$ é a densidade de uma variável aleatória com distribuição normal $(p+1)$ -variada com média $\hat{\beta}$ e matriz de covariância $\frac{\sigma^2(1-f)}{n} \Sigma_{xx}^{-1}$.

Devido aos cálculos efetuados na aplicação 2,

$$\lim_{n \rightarrow \infty} \text{Var}(\sqrt{n} [\hat{\beta} - B]) = \sigma^2(1-f) \Sigma_{xx}^{-1},$$

de onde se conclui que, para n grande,

$$\text{Var}(\hat{\beta} - B) \approx \sigma^2 \frac{(1-f)}{n} \Sigma_{xx}^{-1}.$$

É importante lembrarmos que algo semelhante ocorreu no resultado do teorema 3.4, quando, no modelo em que $y \sim N_N(X\beta, V)$, V qualquer, $L_p(B | y_s) \propto N_{p+1}(\hat{\beta}, \text{Var}(\hat{\beta} - B))$.

Notamos ainda que, independente do aspecto assintótico, nenhuma perda é constatada quando utilizamos $\hat{\beta}$ ao invés de y_s . Este fato é importante porque, avaliado y_s , $\hat{\beta}$ pode ser obtido, mas o contrário não é verdade.

Salientamos também que, dadas as suposições do teorema 4.3, nesta aplicação, tivemos que nos restringir ao caso em que V é diagonal.

4.5 - Considerações Finais

A principal motivação para determinarmos a distribuição assintótica de $\sqrt{n} \begin{bmatrix} \hat{\beta} - \beta \\ \hat{B} - B \end{bmatrix}$ foi o teorema A.4.3 de Fuller [1987],

que fornece a distribuição assintótica de $\sqrt{n} \begin{bmatrix} \bar{y} - \mu \\ \bar{y} - \bar{Y} \end{bmatrix}$.

Os modelos considerados neste capítulo diferem dos utilizados no restante do trabalho por serem destinados a variáveis aleatórias independentes, de modo que a matriz de covariância $V = \text{Var}(y)$ é diagonal.

Se desejássemos deduzir propriedades assintóticas para $\hat{\beta}$ e $\hat{B}$ no caso não diagonal, teríamos a dificuldade da inexistência de teoremas assintóticos clássicos para modelos com variáveis aleatórias correlacionadas.

Neste caso, resultados assintóticos são possíveis estabelecidas suposições do tipo "mixing", que não serão abordadas neste trabalho, mas que ficam como sugestão para uma pesquisa futura.

CAPÍTULO 5 - CONCLUSÕES

A existência de uma teoria de previsão para $\underline{\ell}'\underline{y}$ sob a abordagem de modelos de superpopulação nos levou a acreditar que estudo semelhante poderia ser desenvolvido para a quantidade vetorial $\underline{B}_N$, recente objeto de alguns artigos com outros enfoques.

No entanto, apesar deste sentimento inicial, não esperávamos todos os resultados encontrados.

Na previsão de $\underline{B}_N$, embora semelhanças com a previsão de $\underline{\ell}'\underline{y}$ tenham sido constatadas, ocorreram também algumas diferenças.

É importante notarmos porém, que todas as vezes que um resultado inesperado ocorreu, ele foi mais forte ou mais simples que seu correspondente (quando existia) para $\underline{\ell}'\underline{y}$, mostrando que embora $\underline{B}_N$ seja uma quantidade mais complexa, sua previsão é mais simples.

Do ponto de vista computacional, devido ao teorema 2.2, o cálculo do previsor ótimo de $\underline{B}_N$ vai se reduzir à mera determinação do estimador de mínimos quadrados generalizados. Este cálculo poderá ser feito:

- 1 - Através de pacotes computacionais que determinam o estimador de mínimos quadrados generalizados. O PC CARP [1987] possui a opção do cálculo deste estimador.
- 2 - Os pacotes computacionais de regressão fornecem o estimador

de mínimos quadrados $(X'X)^{-1}X'y$ nos modelos de regressão $y = X\beta + \varepsilon$, com $\text{Var}(\varepsilon) = \sigma^2 I_n$.

Lembrando que a matriz V_s , sendo positiva definida, pode ser escrita como $V_s = PP'$, onde P é simétrica, não singular e única, segue que

$$V_s^{-1} = P^{-1'}P^{-1},$$

$$\text{Var}(P^{-1}y_s) = P^{-1}V_sP^{-1'} = I_n$$

e

$$\hat{\beta} = (X_s'V_s^{-1}X_s)^{-1}X_s'V_s^{-1}y_s = (Q'Q)^{-1}Q'z,$$

com

$$Q = P^{-1}X_s \text{ e } z = P^{-1}y_s.$$

Assim, o previsor ótimo de B_N poderia ser calculado através destes pacotes de regressão, uma vez determinadas as matrizes P e P^{-1} .

3 - O SOC (Software Científico), SOC[1989], no módulo CM, possui excelentes recursos em cálculos matriciais, permitindo a obtenção direta de $\hat{\beta}$ e de seu erro quadrático médio, dado na expressão (2.5).

Salientamos que os resultados do capítulo 2 e os teoremas do capítulo 3 relativos a B_N podem ser trivialmente estendidos para a previsão do vetor aleatório A^*y , onde A^* é uma matriz qualquer $k \times N$ de constantes conhecidas.

Dependendo da forma como A^* é construída, a quantidade A^*y pode ser de grande interesse.

Se as linhas de A^* se constituírem nos vetores $\underline{\ell}_1, \underline{\ell}_2 \dots \underline{\ell}_k$, a previsão de A^*y se reduzirá à previsão simultânea de k combinações lineares de y , ou seja, de $\underline{\ell}'_1 y, \underline{\ell}'_2 y, \dots \underline{\ell}'_k y$.

Adotando modelos de regressão multivariados, em que h variáveis dependentes estão associadas a cada elemento populacional, e tomando

$$y' = [y'_{1s} \ y'_{2s} \dots y'_{hs} \ y'_{1r} \ y'_{2r} \dots y'_{hr}],$$

onde y_{js} e y_{jr} , $j = 1, 2 \dots h$ contêm, respectivamente, os valores da j -ésima variável dependente para os elementos pertencentes e não pertencentes à amostra, as combinações lineares dos elementos de $y'_j = [y'_{js} \ y'_{jr}]$, $j = 1, 2 \dots h$ podem ser escritas na forma A^*y e o correspondente previsor ótimo será obtido através do mesmo procedimento utilizado no capítulo 2.

Conforme comentamos na introdução, é freqüente o estudo sobre a robustez do previsor ótimo de $\underline{\ell}'y$ sob modelos alternativos com variáveis independentes adicionais.

Observamos que uma análise deste tipo não faria sentido aqui pois diferentes modelos de superpopulação geram diferentes expressões para B_N . Assim, uma mudança no modelo acarretaria uma alteração na quantidade a ser prevista, inviabilizando qualquer estudo de robustez deste tipo.

A continuidade deste trabalho consistirá em obter o

previsor ótimo de B_N relativo à função de perda balanceada

$$\ell(\bar{\beta}, B_N) = \frac{w(\mathbf{y}_s - \mathbf{X}_s \bar{\beta})'(\mathbf{y}_s - \mathbf{X}_s \bar{\beta})}{n} + (1 - w) \left\{ \lambda' [B_N - f(\bar{\beta})] \right\}^2,$$

onde $\bar{\beta}$ é um estimador de β ,

$f(\bar{\beta})$ um previsor de B_N ,

e $0 \leq w \leq 1$.

Esta função, motivada por Zellner [1991], considera simultaneamente a qualidade do ajuste do modelo e a precisão da previsão, com pesos respectivamente w e $1 - w$.

Como sugestões para trabalhos futuros ficam as apresentadas no final de cada capítulo.

Gostaríamos de um prosseguimento na linha e acreditamos que isto seja possível.

Reconhecemos que a literatura na área de modelos de superpopulação foi rica nos últimos anos, mas sabemos também que as eventuais dificuldades na criação de novas pesquisas poderão ser perfeitamente superadas graças à extensão da área e graças ao empenho e imaginação dos pesquisadores.

APENDICE 1 - INFERENCIA EM POPULAÇÕES FINITAS

CONCEITOS BÁSICOS

Geralmente, os procedimentos estatísticos tradicionais são destinados a resolver problemas em populações infinitas, mas na prática, a maioria das populações é finita.

Seja N um número inteiro, positivo e conhecido.

Definição A.1.1 - Uma população finita de tamanho N é uma coleção de N unidades.

Definição A.1.2 - As unidades de uma população finita são ditas identificáveis se cada uma delas puder ser rotulada de maneira única por um número inteiro entre 1 e N . Os rótulos devem ser conhecidos para cada unidade e duas unidades distintas não podem possuir o mesmo rótulo.

Devido a esta correspondência entre as unidades populacionais e seus respectivos rótulos, indicamos por $\mathcal{P} = \{1, 2, \dots, N\}$ indistintamente a população e o conjunto dos rótulos.

Neste trabalho, consideramos uma população finita de tamanho N em que, associado à i -ésima unidade populacional corresponde um número y_i , fixo e desconhecido, que é o valor de uma particular variável de interesse Y e, na maior parte dos

casos, admitimos a existência de p variáveis auxiliares $X_1, X_2, \dots, X_p$, cujos valores são conhecidos para todas as unidades populacionais.

O problema de inferência em populações finitas consiste em estimar qualquer função dos valores desconhecidos $Y_1, Y_2, \dots, Y_N$.

Definição A.1.3 - Chamamos de amostra não ordenada e indicamos por s a qualquer subconjunto não vazio de $\mathcal{P}$.

O conjunto dos rótulos das unidades não pertencentes à amostra não ordenada é indicado por r e assim, $s \cup r = \mathcal{P}$.

Obtida a amostra, os valores da variável Y , inicialmente desconhecidos para todas as unidades da população, serão conhecidos para as unidades amostrais.

Duas abordagens têm sido utilizadas na inferência em populações finitas: a que estamos chamando de abordagem clássica e a abordagem de modelos de superpopulação.

A abordagem clássica é a teoria de amostragem tradicional que tem sido adotada na grande maioria dos problemas de estimação em populações finitas e é descrita em livros como Cochran [1977], Raj [1968] e Kish [1965].

O conceito de planejamento amostral é básico nesta abordagem.

Definição A.1.4 - Seja S a coleção de todos os subconjuntos não vazios de $\mathcal{P}$. Uma função $p: S \rightarrow [0,1]$, tal que $p(s) \geq 0$ para todo

$s \in S$ e $\sum_{s \in S} p(s) = 1$ é chamada planejamento amostral não ordenado.

Definição A.1.5 - Um estimador $\hat{T}$ é uma função qualquer dos valores $y_{b_1}, y_{b_2} \dots y_{b_n}$, onde $\{b_1, b_2 \dots b_n\} = s$.

Definição A.1.6 - A esperança de qualquer estimador $\hat{T}$ sob o planejamento amostral p é definida como:

$$E_p(\hat{T}) = \sum_{s \in S} p(s) \hat{T}(s),$$

onde $\hat{T}(s)$ é o estimador $\hat{T}$ calculado para os elementos da amostra s .

As próximas definições introduzem os conceitos de estimador não viciado para uma função paramétrica, erro quadrático médio e variância de um estimador.

Definição A.1.7 - Se $y' = (y_1, y_2 \dots y_N)$, qualquer função de y é chamada função paramétrica. $\hat{T}$ é um estimador não viciado para a função paramétrica θ sob o planejamento amostral p se

$$E_p(\hat{T}) = \theta, \quad \forall y.$$

Definição A.1.8 - O erro quadrático médio de $\hat{T}$ sob o planejamento amostral p é definido como

$$E_p [(\hat{T} - \theta)^2] = \sum_{s \in S} p(s) (\hat{T}(s) - \theta)^2$$

e a variância de $\hat{T}$ é dada por

$$E_p [(\hat{T} - E_p(\hat{T}))^2] = \sum_{s \in S} p(s) (\hat{T}(s) - E_p(\hat{T}))^2.$$

Todas as esperanças associadas ao estimador $\hat{T}$ são calculadas com relação à distribuição de probabilidades do planejamento amostral p . Para salientar este fato, estas esperanças foram denotadas por $E_p(\cdot)$. Na abordagem clássica, o único mecanismo aleatório que participa da inferência é o planejamento amostral.

A abordagem de modelos de superpopulação admite que $y_1, y_2, \dots, y_N$ são valores de N variáveis aleatórias com uma distribuição de probabilidades conjunta ξ . y' é considerado uma realização do vetor aleatório $y' = (y_1, y_2, \dots, y_N)$ com distribuição ξ . Algumas características da distribuição ξ são supostas conhecidas e toda inferência é feita com base nesta distribuição.

Se o modelo de superpopulação adotado for da forma paramétrica, isto é, indexado por um parâmetro χ , então χ será o parâmetro populacional, que muitas vezes é chamado de superparâmetro.

Devido ao fato que, sob esta abordagem, $y_1, y_2, \dots, y_N$ são considerados valores das variáveis aleatórias $Y_1, Y_2, \dots, Y_N$, passamos a usar o termo "previsor" ao invés de estimador quando

nos referimos a $\hat{T}$. A estimação de γ é fundamental na previsão das quantidades de interesse.

Para um particular modelo de superpopulação ξ , a esperança matemática de qualquer função das variáveis $Y_1, Y_2 \dots Y_N$ é calculada de acordo com a próxima definição.

Definição A.1.9- Seja $Q = Q(Y_1, Y_2 \dots Y_N)$ uma função das variáveis $Y_1, Y_2 \dots Y_N$. A esperança matemática de Q sob o modelo ξ , que denotaremos por $E_\xi(Q)$, é definida como:

$$E_\xi(Q) = \int Q d\xi.$$

Os conceitos de preditor não viciado para qualquer função dos $Y_1, Y_2 \dots Y_N$, T e de erro quadrático médio de um preditor são dados nas próximas definições.

Definição A.1.10 - Dada uma amostra s , um preditor $\hat{T}$ é dito não viciado para T se

$$E_\xi(\hat{T} - T) = 0.$$

Definição A.1.11 - O erro quadrático médio de um preditor de T , $\hat{T}$, sob o modelo ξ , que indicaremos por $EQM_\xi(\hat{T})$ é definido como:

$$EQM_\xi(\hat{T}) = E_\xi(\hat{T} - T)^2.$$

No caso em que T é uma combinação linear dos y_i 's, a

previsão é feita da seguinte maneira:

- Um modelo de superpopulação ξ , com parâmetro γ totalmente ou parcialmente desconhecido, é adotado.

$$T = \sum_{i=1}^N a_i y_i \text{ pode ser escrita como}$$

$$T = \sum_{i \in s} a_i y_i + \sum_{i \in r} a_i y_i \quad (\text{A.1.1})$$

onde $\sum_{i \in s} a_i y_i$ é a combinação linear correspondente aos valores y_i observados na amostra e portanto perfeitamente conhecida. A forma como a amostra s foi obtida (planejamento amostral) é irrelevante, já que a inferência é baseada no modelo ξ . Como a primeira parcela de (A.1.1) é conhecida, o problema de prever T se reduz a prever $\sum_{i \in r} a_i y_i$.

Conforme salientam Pereira e Rodrigues [1983], os valores y_i para $i \in r$ são desconhecidos e portanto indistinguíveis das respectivas variáveis aleatórias Y_i . Para facilitar a notação, passamos então a indicar por y_i tanto a variável aleatória, quanto seu valor.

O próximo exemplo ilustra de forma resumida alguns dos conceitos descritos.

Exemplo A.1.1

Vamos admitir um modelo de superpopulação ξ :

$$y_i = \beta x_i + \varepsilon_i, \quad i = 1, 2 \dots N,$$

onde ε_i é uma variável aleatória com $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = x_i \sigma^2$, ε_i e ε_k são independentes para $i \neq k$, e x_i é uma variável auxiliar, positiva e conhecida para $i=1,2,\dots,N$. O vetor (β, σ^2) é o superparâmetro.

Como consequência, $E(y_i) = \beta x_i$ e $\text{Var}(y_i) = x_i \sigma^2$, $i=1,2,\dots,N$.

Tomada uma amostra de tamanho n desta distribuição, considera-se, sem perda de generalidade, que $y_1, y_2 \dots y_n$ são os valores observados na amostra (e portanto conhecidos) e $y_{n+1}, y_{n+2} \dots y_N$ são os valores não observados.

Se estes valores não observados y_i , $i = n+1, n+2 \dots N$ forem previstos por $\hat{\beta} x_i$, onde $\hat{\beta} = \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$ é o estimador de mínimos

quadrados ponderados de β no modelo de regressão linear pela origem de acordo com a teoria de modelos lineares, o total populacional

$$T = \sum_{i=1}^N y_i = \sum_{i=1}^n y_i + \sum_{i=n+1}^N y_i$$

será previsto por

$$\hat{T} = \sum_{i=1}^n y_i + \sum_{i=n+1}^N x_i \hat{\beta} = \sum_{i=1}^n y_i + \sum_{i=n+1}^N x_i \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i}$$

$$= \frac{\sum_{i=1}^n y_i}{\sum_{i=1}^n x_i} \sum_{i=1}^N x_i,$$

que é o previsor razão.

Além disso,

$$E_{\xi} (\hat{T} - T) = E_{\xi} \left[\sum_{i=1}^n y_i \left(\frac{\sum_{i=1}^N x_i}{\sum_{i=1}^n x_i} - 1 \right) - \sum_{i=n+1}^N y_i \right] =$$

$$= \frac{\sum_{i=n+1}^N x_i}{\sum_{i=1}^n x_i} \beta \sum_{i=1}^n x_i - \beta \sum_{i=n+1}^N x_i = 0.$$

Portanto, o previsor razão $\hat{T}$ é ξ -não viciado para o total populacional.

□

Definição A.1.12 - Seja $\mathcal{U}$ a classe formada por todos os previsores ξ -não viciados para T . Um previsor $\hat{T}^*$ é ξ -ótimo se $\hat{T}^* \in \mathcal{U}$ e

$$E_{\xi} (\hat{T}^* - T)^2 \leq E_{\xi} (\hat{T} - T)^2, \quad \forall \hat{T} \in \mathcal{U}.$$

De acordo com esta definição, o previsor ótimo para T será o previsor não viciado que apresentar o menor erro quadrático médio dentro da classe de todos os previsores ξ -não viciados de T .

APENDICE 2 - PREVISÃO DE $\hat{\underline{z}}'y$

A.2.1 - Introdução

Neste apêndice apresentamos algumas definições e resultados básicos presentes na literatura, relacionados com a previsão da quantidade linear $\underline{z}'y$, quando o modelo de superpopulação é o modelo de regressão linear (2.1) introduzido no capítulo 2.

Este modelo era

$$y = X\underline{\beta} + \underline{\varepsilon} \quad , \quad (A.2.1)$$

onde

$X = \begin{bmatrix} X_s \\ X_r \end{bmatrix}$ é a matriz não aleatória de variáveis auxiliares, totalmente conhecida,

$$y' = [y_1 \quad y_2 \quad \dots \quad y_N] = [y'_s \quad y'_r] \quad e$$

$\underline{\varepsilon}' = [\varepsilon_1 \quad \varepsilon_2 \quad \dots \quad \varepsilon_N]$ são vetores aleatórios, com

$$E(\underline{\varepsilon}) = \underline{0}$$

$$\text{Var}(\underline{\varepsilon}) = V = \begin{bmatrix} V_s & V_{sr} \\ V_{rs} & V_r \end{bmatrix}, \text{ conhecida}$$

e $\underline{\beta}' = [\beta_0 \ \beta_1 \ \beta_2 \ \dots \ \beta_p]$ é o vetor de parâmetros desconhecidos.

A quantidade linear $\underline{\ell}'\underline{y} = \sum_{i=1}^N \ell_i y_i$, onde $\underline{\ell}' = [\ell_1 \ \ell_2 \ \dots \ \ell_N]$ é um vetor de constantes conhecidas, pode ser escrita como $\underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r \underline{y}_r$, sendo que $\underline{\ell}'_s = [\ell_1 \ \ell_2 \ \dots \ \ell_n]$ e $\underline{\ell}'_r = [\ell_{n+1} \ \ell_{n+2} \ \dots \ \ell_N]$ são os vetores contendo respectivamente os coeficientes da combinação linear dos elementos pertencentes e não pertencentes à amostra.

Com esta notação, $\underline{\ell}'_s \underline{y}_s$ é conhecida e o problema de prever $\underline{\ell}'\underline{y}$ se reduz a prever $\underline{\ell}'_r \underline{y}_r$. Por este motivo, restringimo-nos a previsores de $\underline{\ell}'\underline{y}$ da forma

$$\hat{L}(\underline{y}_s) = \underline{\ell}'_s \underline{y}_s + \theta(\underline{y}_s),$$

onde $\theta(\underline{y}_s)$ é um previsor da parte não observada, $\underline{\ell}'_r \underline{y}_r = \underline{\ell}'\underline{y} - \underline{\ell}'_s \underline{y}_s$.

A.2.2 - Previsão Linear Ótima para $\underline{\ell}'\underline{y}$

Nesta seção apresentaremos o previsor linear ótimo de $\underline{\ell}'\underline{y}$ sob o modelo (A.2.1), obtido por Royall [1976a].

Definição A.2.1 - Se $\underline{h}' = [h_1 \quad h_2 \quad \dots \quad h_n]$ é um vetor de constantes conhecidas, então $\underline{h}'\underline{y}_s$ é um previsor linear de $\underline{\ell}'\underline{y}$.

Definição A.2.2 - Um previsor $\underline{h}'\underline{y}_s$ é ξ -não viciado para $\underline{\ell}'\underline{y}$ se e somente se

$$E_{\xi}(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y}) = 0,$$

onde $E_{\xi}(\cdot)$ representa a esperança sob o modelo de superpopulação ξ .

Definição A.2.3 - O erro quadrático médio do previsor $\underline{h}'\underline{y}_s$ sob o modelo ξ é dado por

$$EQM_{\xi}(\underline{h}'\underline{y}_s) = E_{\xi}(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y})^2.$$

As definições A.2.2 e A.2.3 são as correspondentes do apêndice 1, adaptadas agora para a notação que acabamos de introduzir.

A próxima definição estabelece o conceito de previsor linear ξ -ótimo, ou simplesmente linear ótimo para $\underline{\ell}'\underline{y}$.

Definição A.2.4 - Seja $\mathcal{U}_{\ell}$ a classe formada por todos os previsores lineares ξ -não viciados de $\underline{\ell}'\underline{y}$. Diremos que $\underline{h}^{*}'\underline{y}_s$ é o previsor linear ξ -ótimo de $\underline{\ell}'\underline{y}$ se e somente se $\underline{h}^{*}'\underline{y}_s \in \mathcal{U}_{\ell}$ e

$$E_{\xi}(\underline{h}^{*}'\underline{y}_s - \underline{\ell}'\underline{y})^2 \leq E_{\xi}(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y})^2, \quad \forall \underline{h}'\underline{y}_s \in \mathcal{U}_{\ell}.$$

De acordo com esta definição, o previsor linear ξ -ótimo de $\underline{\ell}'\underline{y}$ será o previsor linear ξ -não viciado que apresentar o menor erro quadrático médio dentro da classe de todos os previsores lineares ξ -não viciados de $\underline{\ell}'\underline{y}$.

O lema a seguir permite a obtenção do previsor linear ótimo de $\underline{\ell}'\underline{y}$, que se encontra no teorema A.2.1.

Lema A.2.1 (Tam [1986])

Sejam $\underline{h}'\underline{y}_s$ um previsor linear qualquer de $\underline{\ell}'\underline{y}$ e K a matriz $n \times N$, $[\underline{V}_s \ \underline{V}_{sr}]$. Nestas condições, sob o modelo (A.2.1),

$$\begin{aligned} E(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y})^2 &= E[(\underline{h}' - \underline{\ell}'K'V_s^{-1})(\underline{y}_s - X_s\beta)]^2 + \\ &+ \underline{\ell}'V\underline{\ell} - \underline{\ell}'K'V_s^{-1}K\underline{\ell} + [(\underline{h}'X_s - \underline{\ell}'X)\beta]^2. \end{aligned} \quad (A.2.2)$$

Prova:

$$E(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y})^2 = \text{Var}(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y}) + \{E(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y})\}^2.$$

Como $E(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y}) = \underline{h}'X_s\beta - \underline{\ell}'X\beta = (\underline{h}'X_s - \underline{\ell}'X)\beta$, devemos provar que

$$\text{Var}(\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y}) = E[(\underline{h}' - \underline{\ell}'K'V_s^{-1})(\underline{y}_s - X_s\beta)]^2 + \underline{\ell}'V\underline{\ell} - \underline{\ell}'K'V_s^{-1}K\underline{\ell}.$$

Escrevendo $\underline{h}'\underline{y}_s - \underline{\ell}'\underline{y}$ como

$$(\tilde{h}'\tilde{y}_s - \tilde{\ell}'K'V_s^{-1}\tilde{y}_s) + (\tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}),$$

temos que

$$\begin{aligned} \text{Var}(\tilde{h}'\tilde{y}_s - \tilde{\ell}'\tilde{y}) &= \text{Var}(\tilde{h}'\tilde{y}_s - \tilde{\ell}'K'V_s^{-1}\tilde{y}_s) + \text{Var}(\tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}) \\ &+ 2 \text{Cov}(\tilde{h}'\tilde{y}_s - \tilde{\ell}'K'V_s^{-1}\tilde{y}_s, \tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}) = \\ &= E[(\tilde{h}' - \tilde{\ell}'K'V_s^{-1})(\tilde{y}_s - X_s\beta)]^2 + \text{Var}(\tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}) \\ &+ 2 \text{Cov}(\tilde{h}'\tilde{y}_s - \tilde{\ell}'K'V_s^{-1}\tilde{y}_s, \tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}). \end{aligned}$$

Mas

$$\begin{aligned} \text{Var}(\tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}) &= \text{Var}(\tilde{\ell}'K'V_s^{-1}\tilde{y}_s) + \text{Var}(\tilde{\ell}'\tilde{y}) \\ &- 2 \text{Cov}(\tilde{\ell}'K'V_s^{-1}\tilde{y}_s, \tilde{\ell}'\tilde{y}) = \tilde{\ell}'K'V_s^{-1}V_sV_s^{-1}K\tilde{\ell} + \\ &+ \tilde{\ell}'V\tilde{\ell} - 2\tilde{\ell}'K'V_s^{-1} \text{Cov}(\tilde{y}_s, \tilde{y})\tilde{\ell} = \\ &= \tilde{\ell}'K'V_s^{-1}K\tilde{\ell} + \tilde{\ell}'V\tilde{\ell} - 2\tilde{\ell}'K'V_s^{-1}K\tilde{\ell} = \\ &= \tilde{\ell}'V\tilde{\ell} - \tilde{\ell}'K'V_s^{-1}K\tilde{\ell} \end{aligned}$$

e

$$\begin{aligned} \text{Cov}([\tilde{h}' - \tilde{\ell}'K'V_s^{-1}]\tilde{y}_s, \tilde{\ell}'K'V_s^{-1}\tilde{y}_s - \tilde{\ell}'\tilde{y}) &= \\ \text{Cov}([\tilde{h}' - \tilde{\ell}'K'V_s^{-1}]\tilde{y}_s, \tilde{\ell}'K'V_s^{-1}\tilde{y}_s) &- \text{Cov}([\tilde{h}' - \tilde{\ell}'K'V_s^{-1}]\tilde{y}_s, \tilde{\ell}'\tilde{y}) \\ &= (\tilde{h}' - \tilde{\ell}'K'V_s^{-1})V_sV_s^{-1}K\tilde{\ell} - (\tilde{h}' - \tilde{\ell}'K'V_s^{-1})K\tilde{\ell} = 0. \end{aligned}$$

Portanto,

$$\begin{aligned} \text{Var}(\tilde{h}'\tilde{y}_s - \tilde{\ell}'\tilde{y}) &= E[(\tilde{h}' - \tilde{\ell}'K'V_s^{-1})(\tilde{y}_s - X_s\tilde{\beta})]^2 + \\ &+ \tilde{\ell}'V\tilde{\ell} - \tilde{\ell}'K'V_s^{-1}K\tilde{\ell}, \end{aligned}$$

o que prova o resultado. □

Observamos que $E(\tilde{h}'\tilde{y}_s - \tilde{\ell}'\tilde{y})^2$ corresponde ao erro quadrático médio de $\tilde{h}'\tilde{y}_s$ (o símbolo ξ de $E_\xi(\cdot)$ foi abandonado para simplificar a notação) e, na obtenção do previsor linear ótimo, é a quantidade a ser minimizada.

Se $\tilde{h}'\tilde{y}_s$ é não viciado para $\tilde{\ell}'\tilde{y}$, então

$$E(\tilde{h}'\tilde{y}_s - \tilde{\ell}'\tilde{y}) = \tilde{h}'X_s\tilde{\beta} - \tilde{\ell}'X\tilde{\beta} = (\tilde{h}'X_s - \tilde{\ell}'X)\tilde{\beta} = 0,$$

$\forall \tilde{\beta} \in \mathbb{R}^{p+1}$ (ou $\mathbb{R}^p$ se o modelo for sem intercepto), o que implica em

$$\tilde{h}'X_s = \tilde{\ell}'X,$$

que é a condição para que um previsor linear qualquer $\tilde{h}'\tilde{y}_s$ seja não viciado para $\tilde{\ell}'\tilde{y}$. Válida esta condição, como consequência de (A.2.2), o erro quadrático médio de $\tilde{h}'\tilde{y}_s$ se reduz a

$$E(\tilde{h}'\tilde{y}_s - \tilde{\ell}'\tilde{y})^2 = E[(\tilde{h}' - \tilde{\ell}'K'V_s^{-1})(\tilde{y}_s - X_s\tilde{\beta})]^2 +$$

$$+ \underline{\ell}' V \underline{\ell} - \underline{\ell}' K' V_s^{-1} K \underline{\ell} = \text{Var}([\underline{h}' - \underline{\ell}' K' V_s^{-1}] \underline{y}_s) + \\ + \underline{\ell}' V \underline{\ell} - \underline{\ell}' K' V_s^{-1} K \underline{\ell}. \quad (\text{A.2.3})$$

Teorema A.2.1

No modelo (A.2.1), o previsor linear ótimo de $\underline{\ell}' \underline{y}$ é dado por

$$\hat{T}^* = \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r [X_r \hat{\beta} + V_{rs} V_s^{-1} (\underline{y}_s - X_s \hat{\beta})],$$

onde

$\hat{\beta} = (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} \underline{y}_s$ é o estimador de mínimos quadrados generalizados de β .

Prova:

Seja $\underline{h}' \underline{y}_s$ um previsor linear qualquer. Se $\underline{h}' \underline{y}_s$ é não viciado para $\underline{\ell}' \underline{y}$, então $\underline{h}' X_s = \underline{\ell}' X$.

O previsor linear ξ -ótimo de $\underline{\ell}' \underline{y}$, $\underline{h}^* \underline{y}_s$ é tal que $\underline{h}^* X_s = \underline{\ell}' X$ e, devido a (A.2.3), deve minimizar

$$\text{Var}([\underline{h}' - \underline{\ell}' K' V_s^{-1}] \underline{y}_s) = E\{[\underline{h}' - \underline{\ell}' K' V_s^{-1}] \underline{y}_s - E([\underline{h}' - \underline{\ell}' K' V_s^{-1}] \underline{y}_s)\}^2.$$

Esta variância será minimizada quando $(\tilde{h}' - \tilde{\ell}' K' V_s^{-1}) y_s$ for o estimador linear ótimo de

$$\begin{aligned} E[(\tilde{h}' - \tilde{\ell}' K' V_s^{-1}) y_s] &= \tilde{h}' X_s \beta - \tilde{\ell}' K' V_s^{-1} X_s \beta = \\ &= \tilde{\ell}' X \beta - \tilde{\ell}' K' V_s^{-1} X_s \beta = (\tilde{\ell}' X - \tilde{\ell}' K' V_s^{-1} X_s) \beta \end{aligned}$$

que, pelo teorema de Gauss-Markov, é

$$(\tilde{\ell}' X - \tilde{\ell}' K' V_s^{-1} X_s) \hat{\beta}.$$

Portanto,

$$(\tilde{h}^{*'} - \tilde{\ell}' K' V_s^{-1}) y_s = (\tilde{\ell}' X - \tilde{\ell}' K' V_s^{-1} X_s) \hat{\beta}.$$

Observando que

$$\tilde{\ell}' X = \tilde{\ell}'_s X_s + \tilde{\ell}'_r X_r,$$

$$\tilde{\ell}' K' = \begin{bmatrix} \tilde{\ell}'_s & \tilde{\ell}'_r \end{bmatrix} \begin{bmatrix} V_s \\ V_{rs} \end{bmatrix} = \tilde{\ell}'_s V_s + \tilde{\ell}'_r V_{rs} \quad e$$

$$\tilde{\ell}' K' V_s^{-1} X_s = \tilde{\ell}'_s X_s + \tilde{\ell}'_r V_{rs} V_s^{-1} X_s, \text{ então}$$

$$\tilde{h}^{*'} y_s = (\tilde{\ell}'_r X_r - \tilde{\ell}'_r V_{rs} V_s^{-1} X_s) \hat{\beta} + (\tilde{\ell}'_s + \tilde{\ell}'_r V_{rs} V_s^{-1}) y_s$$

$$= \tilde{\ell}'_s y_s + \tilde{\ell}'_r [X_r \hat{\beta} + V_{rs} V_s^{-1} (y_s - X_s \hat{\beta})] = \hat{\tau}^*.$$

□

Verifica-se ainda que o erro quadrático médio de $\hat{T}^*$ vale

$$EQM(\hat{T}^*) = E (\hat{T}^* - \ell' \underline{y})^2 =$$

$$= \underline{\ell}'_r (V_r - V_{rs} V_s^{-1} V_{sr}) \underline{\ell}_r + \underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) (X'_s V_s^{-1} X_s)^{-1} (X'_r - X'_s V_s^{-1} V_{sr}) \underline{\ell}_r.$$

Quando $V_{rs} = 0$, ou seja, V é bloco diagonal ($\underline{y}_s$ e $\underline{y}_r$ são não correlacionados), $\hat{T}^*$ fica

$$\underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r X_r \hat{\beta} \quad (A.2.4)$$

e seu erro quadrático médio se reduz a

$$\underline{\ell}'_r V_r \underline{\ell}_r + \underline{\ell}'_r X_r (X'_s V_s^{-1} X_s)^{-1} X'_r \underline{\ell}_r. \quad (A.2.5)$$

O resultado presente no teorema A.2.1 foi obtido e apresentado sem demonstração por Royall [1976a], quando $\underline{\ell}' \underline{y}$ era o total populacional $\underline{1}'_N \underline{y} = \sum_{i=1}^N y_i$. A extensão para $\underline{\ell}'$ qualquer, dado o lema A.2.1, é imediata.

A.2.3 - Previsão Ótima para $\underline{\ell}' \underline{y}$ no Modelo com Distribuição Normal

Tam [1987], no artigo "Optimality of Royall's predictor under a Gaussian superpopulation model", mostra que se $\underline{y}$ tem distribuição normal N -variada com média $X\beta$ e matriz de

covariância V conhecida, então $\hat{T}^*$ é o previsor não viciado de variância mínima de $\underline{\ell}'Y$.

O autor trata somente da previsão de $\underline{\ell}'Y$. Apresentaremos seu resultado já com a adaptação imediata para a previsão de $\underline{\ell}'Y$.

Lema A.2.2

Sob o modelo (A.2.1), se $f(Y_s)$ é um previsor qualquer de $\underline{\ell}'Y$, então

$$E\{\underline{\ell}'Y - f(Y_s)\}^2 = E\{\underline{\ell}'Y - E(\underline{\ell}'Y | Y_s)\}^2 + E\{f(Y_s) - E(\underline{\ell}'Y | Y_s)\}^2.$$

Prova:

Como $\underline{\ell}'Y - f(Y_s) = \underline{\ell}'Y - E(\underline{\ell}'Y | Y_s) + E(\underline{\ell}'Y | Y_s) - f(Y_s)$, é suficiente provarmos que

$$E\{[\underline{\ell}'Y - E(\underline{\ell}'Y | Y_s)][E(\underline{\ell}'Y | Y_s) - f(Y_s)]\} = 0.$$

Esta esperança é igual a

$$\begin{aligned} E\{\underline{\ell}'Y E(\underline{\ell}'Y | Y_s)\} - E\{\underline{\ell}'Y f(Y_s)\} - E\{E(\underline{\ell}'Y | Y_s)E(\underline{\ell}'Y | Y_s)\} \\ + E\{E(\underline{\ell}'Y | Y_s)f(Y_s)\}. \end{aligned} \quad (A.2.6)$$

Temos que

$$\begin{aligned}
E\{\tilde{\ell}'Y E(\tilde{\ell}'Y | Y_s)\} &= E\{[\tilde{\ell}'_s Y_s + \tilde{\ell}'_r Y_r] E(\tilde{\ell}'_s Y_s + \tilde{\ell}'_r Y_r | Y_s)\} = \\
&= E\{[\tilde{\ell}'_s Y_s + \tilde{\ell}'_r Y_r] [\tilde{\ell}'_s Y_s + E(\tilde{\ell}'_r Y_r | Y_s)]\} = \\
&= E\{\tilde{\ell}'_s Y_s \tilde{\ell}'_s Y_s + \tilde{\ell}'_s Y_s E(\tilde{\ell}'_r Y_r | Y_s) + \tilde{\ell}'_r Y_r \tilde{\ell}'_s Y_s + \\
&\quad + \tilde{\ell}'_r Y_r E(\tilde{\ell}'_r Y_r | Y_s)\} = E\{\tilde{\ell}'_s Y_s \tilde{\ell}'_s Y_s\} + \\
&\quad + E\{\tilde{\ell}'_s Y_s E(\tilde{\ell}'_r Y_r | Y_s)\} + E\{\tilde{\ell}'_r Y_r \tilde{\ell}'_s Y_s\} + E\{\tilde{\ell}'_r Y_r E(\tilde{\ell}'_r Y_r | Y_s)\} \\
&\hspace{15em} (A.2.7)
\end{aligned}$$

e

$$\begin{aligned}
E\{E(\tilde{\ell}'Y | Y_s) E(\tilde{\ell}'Y | Y_s)\} &= E\{[\tilde{\ell}'_s Y_s + E(\tilde{\ell}'_r Y_r | Y_s)] [\tilde{\ell}'_s Y_s + E(\tilde{\ell}'_r Y_r | Y_s)]\} \\
&= E\{\tilde{\ell}'_s Y_s \tilde{\ell}'_s Y_s\} + E\{\tilde{\ell}'_s Y_s E(\tilde{\ell}'_r Y_r | Y_s)\} + E\{E(\tilde{\ell}'_r Y_r | Y_s) \tilde{\ell}'_s Y_s\} \\
&\quad + E\{E(\tilde{\ell}'_r Y_r | Y_s) E(\tilde{\ell}'_r Y_r | Y_s)\} = \\
&= E\{\tilde{\ell}'_s Y_s \tilde{\ell}'_s Y_s\} + E\{\tilde{\ell}'_s Y_s E(\tilde{\ell}'_r Y_r | Y_s)\} + E\{E(\tilde{\ell}'_s Y_s \tilde{\ell}'_r Y_r | Y_s)\} \\
&\quad + E\{E(\tilde{\ell}'_r Y_r | Y_s) E(\tilde{\ell}'_r Y_r | Y_s)\} = \\
&= E\{\tilde{\ell}'_s Y_s \tilde{\ell}'_s Y_s\} + E\{\tilde{\ell}'_s Y_s E(\tilde{\ell}'_r Y_r | Y_s)\} + E(\tilde{\ell}'_s Y_s \tilde{\ell}'_r Y_r) \\
&\quad + E\{E(\tilde{\ell}'_r Y_r | Y_s) E(\tilde{\ell}'_r Y_r | Y_s)\}. \\
&\hspace{15em} (A.2.8)
\end{aligned}$$

Mas

$$\begin{aligned} E\{\tilde{\ell}'_r Y_r \cdot E(\tilde{\ell}'_r Y_r | Y_s)\} &= E\{E[\tilde{\ell}'_r Y_r E(\tilde{\ell}'_r Y_r | Y_s) | Y_s]\} \\ &= E\{E(\tilde{\ell}'_r Y_r | Y_s) E(\tilde{\ell}'_r Y_r | Y_s)\} \end{aligned}$$

e com isto verificamos que a expressão (A.2.7) é igual à expressão (A.2.8). Deste modo,

$$E\{\tilde{\ell}' Y \cdot E(\tilde{\ell}' Y | Y_s)\} = E\{E(\tilde{\ell}' Y | Y_s) E(\tilde{\ell}' Y | Y_s)\}.$$

Finalizando, temos ainda que

$$\begin{aligned} E\{E(\tilde{\ell}' Y | Y_s) f(Y_s)\} &= E\{E[\tilde{\ell}' Y f(Y_s) | Y_s]\} = \\ &= E\{\tilde{\ell}' Y f(Y_s)\}. \end{aligned}$$

Substituindo-se estes resultados em (A.2.6), fica provado que $E\{[\tilde{\ell}' Y - E(\tilde{\ell}' Y | Y_s)][E(\tilde{\ell}' Y | Y_s) - f(Y_s)]\} = 0$, o que encerra a demonstração.

□

Segue deste lema que o previsor de erro quadrático médio mínimo para $\tilde{\ell}' Y$ é o previsor de erro quadrático médio mínimo para $E(\tilde{\ell}' Y | Y_s)$, que indicaremos por $f^*(Y_s)$.

Como $E(\tilde{\ell}' Y | Y_s) = \tilde{\ell}'_s Y_s + \tilde{\ell}'_r E(Y_r | Y_s)$, se $E(Y_r | Y_s)$ não

dependesse de parâmetros desconhecidos, então $f^*(y_s) = \tilde{z}'_s y_s + \tilde{z}'_r E(y_r | y_s)$.

Sob a hipótese de que y tem distribuição normal N-variada,

$$E(y_r | y_s) = (X_r - V_{rs} V_s^{-1} X_s) \beta + V_{rs} V_s^{-1} y_s,$$

e assim,

$$E(\tilde{z}' y | y_s) = \tilde{z}'_s y_s + \tilde{z}'_r V_{rs} V_s^{-1} y_s + \tilde{z}'_r (X_r - V_{rs} V_s^{-1} X_s) \beta.$$

A única quantidade desconhecida em $E(\tilde{z}' y | y_s)$ é β e, com isto, um previsor para $\tilde{z}' y$ (ou $E(\tilde{z}' y | y_s)$) é dado por

$$f(y_s) = \tilde{z}'_s y_s + \tilde{z}'_r V_{rs} V_s^{-1} y_s + \tilde{z}'_r (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta},$$

onde $\hat{\beta}$ é um estimador de β .

Verifica-se facilmente que $f(y_s)$ é não viciado para $\tilde{z}' y$ se e somente se $\hat{\beta}$ é não viciado para β .

Além disso, como $f^*(y_s)$ que minimiza $E\{f(y_s) - \tilde{z}' y\}^2$ também minimiza

$$\begin{aligned} E\{f(y_s) - E(\tilde{z}' y | y_s)\}^2 &= E\{\tilde{z}'_s y_s + \tilde{z}'_r V_{rs} V_s^{-1} y_s + \\ &+ \tilde{z}'_r (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta} - \tilde{z}'_s y_s - \tilde{z}'_r V_{rs} V_s^{-1} y_s - \tilde{z}'_r (X_r - V_{rs} V_s^{-1} X_s) \beta\}^2 \\ &= E\{\tilde{z}'_r (X_r - V_{rs} V_s^{-1} X_s) (\hat{\beta} - \beta)\}^2, \end{aligned}$$

a escolha ótima para $f(y_s)$ é

$$f^*(y_s) = \underline{\ell}'_s y_s + \underline{\ell}'_r V_{rs} V_s^{-1} y_s + \underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}^*,$$

onde $\hat{\beta}^*$ minimiza

$$E\{\underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) (\hat{\beta} - \beta)\}^2.$$

Se $\hat{\beta}$ é não viciado para β então

$$E\{\underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) (\hat{\beta} - \beta)\}^2 = \text{Var}\{\underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}\}$$

e se $y_s \sim N_n(X_s \beta, V_s)$, $\underline{u}' \hat{\beta}^*$, onde $\hat{\beta}^* = (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} y_s$, é o estimador não viciado de variância mínima de $\underline{u}' \beta$ (ver apêndice 3).

Como consequência, o preditor não viciado de erro quadrático médio mínimo de $\underline{\ell}' y$ é

$$f^*(y_s) = \underline{\ell}'_s y_s + \underline{\ell}'_r V_{rs} V_s^{-1} y_s + \underline{\ell}'_r (X_r - V_{rs} V_s^{-1} X_s) \hat{\beta}^*,$$

com $\hat{\beta}^* = (X'_s V_s^{-1} X_s)^{-1} X'_s V_s^{-1} y_s$, que coincide com $\hat{T}^*$.

A.2.4 - Previsão Bayesiana para $\underline{\ell}' y$

I) Resultados Básicos

Rao [1973, pág.234] trabalha com o modelo destinado a populações infinitas, tal que

$$E(\underline{Y} \mid \underline{\beta}) = Z \underline{\beta}$$

$$D(\underline{Y} \mid \underline{\beta}) = \text{Var}(\underline{Y} \mid \underline{\beta}) = W, \quad (\text{A.2.9})$$

onde

$\underline{Y}$ é um vetor aleatório n-dimensional,

Z é a matriz não aleatória $n \times (p+1)$ de variáveis auxiliares, conhecida,

n é o tamanho da amostra,

$\underline{\beta}$ é um vetor aleatório $(p+1)$ -dimensional com

$$E(\underline{\beta}) = \underline{\mu} \quad \text{e} \quad \text{Var}(\underline{\beta}) = D(\underline{\beta}) = G.$$

Seu objetivo é determinar o previsor linear da variável aleatória $\underline{P}'\underline{\beta}$, $\underline{a} + \underline{L}'\underline{Y}$, tal que $E(\underline{P}'\underline{\beta} - \underline{a} - \underline{L}'\underline{Y}) = 0$ e $\text{Var}(\underline{P}'\underline{\beta} - \underline{a} - \underline{L}'\underline{Y})$ é mínima, que será indicado por $\underline{a}^* + \underline{L}^*\underline{Y}$.

Teorema A.2.2

No modelo (A.2.9), se $\underline{\mu}$ é conhecido, então

$$\underline{L}^* = (W + Z G Z')^{-1} Z G \underline{P} = W^{-1} Z (G^{-1} + Z' W^{-1} Z)^{-1} \underline{P},$$

$$\underline{a}^* = \underline{\mu}' \underline{P} - \underline{\mu}' Z' \underline{L}^*$$

e se $\underline{\mu}$ é desconhecido,

$$\underline{L}^* = W^{-1}Z(Z'W^{-1}Z)^{-1}P,$$

$a^* = 0$, desde que exista $\underline{L}$ tal que $Z'\underline{L} = P$.

Prova:

Sob o modelo (A.2.9),

$$\begin{aligned}\text{Var}(\underline{Y}) &= E(\text{Var}(\underline{Y}|\underline{\beta})) + \text{Var}(E(\underline{Y}|\underline{\beta})) = \\ &= E(W) + \text{Var}(Z\underline{\beta}) = W + Z G Z'\end{aligned}$$

e

$$\text{Cov}(\underline{Y}, \underline{P}'\underline{\beta}) = C(\underline{Y}, \underline{P}'\underline{\beta}) = E(C(\underline{Y}, \underline{P}'\underline{\beta}|\underline{\beta})) + C(E(\underline{Y}|\underline{\beta}), \underline{P}'\underline{\beta}).$$

Dado $\underline{\beta}$, $\underline{P}'\underline{\beta}$ é constante e assim

$$C(\underline{Y}, \underline{P}'\underline{\beta} \mid \underline{\beta}) = 0$$

e

$$C(\underline{Y}, \underline{P}'\underline{\beta}) = C(Z\underline{\beta}, \underline{P}'\underline{\beta}) = Z G \underline{P}.$$

Além disso,

$$E(\underline{Y}) = E(E(\underline{Y} \mid \underline{\beta})) = E(Z\underline{\beta}) = Z \underline{\mu}.$$

a) Para $\underline{\mu}$ conhecido:

$$E(\underline{P}'\underline{\beta} - a^{*'} - \underline{L}^{*'}\underline{Y}) = \underline{P}'\underline{\mu} - \underline{P}'\underline{\mu} + \underline{L}^{*'}Z \underline{\mu} - \underline{L}^{*'}Z \underline{\mu} = 0,$$

$$\text{Var}(\underline{P}'\underline{\beta} - a^{*'} - \underline{L}^{*'}\underline{Y}) = \text{Var}(\underline{P}'\underline{\beta} - a^{*'} - \underline{L}'\underline{Y} - \underline{L}^{*'}\underline{Y} + \underline{L}^{*'}\underline{Y}) =$$

$$\begin{aligned}
&= \text{Var}(\underline{P}'\underline{\beta} - \underline{a}^{*'} - \underline{L}^{*'}\underline{Y}) + \text{Var}([\underline{L}^{*'} - \underline{L}']\underline{Y}) + \\
&\quad + 2C(\underline{P}'\underline{\beta} - \underline{a}^{*'} - \underline{L}^{*'}\underline{Y}, [\underline{L}^{*'} - \underline{L}']\underline{Y}).
\end{aligned}$$

Temos que

$$\begin{aligned}
C(\underline{P}'\underline{\beta}, [\underline{L}^{*'} - \underline{L}']\underline{Y}) &= E\{C([\underline{L}^{*'} - \underline{L}']\underline{Y}, \underline{P}'\underline{\beta} \mid \underline{\beta})\} + \\
&\quad + C(E\{[\underline{L}^{*'} - \underline{L}']\underline{Y} \mid \underline{\beta}\}, E(\underline{P}'\underline{\beta} \mid \underline{\beta})) \\
&= 0 + C([\underline{L}^{*'} - \underline{L}']\underline{Z} \underline{\beta}, \underline{P}'\underline{\beta}) = (\underline{L}^{*'} - \underline{L}')\underline{Z} \underline{G} \underline{P}
\end{aligned}$$

e

$$\begin{aligned}
C(\underline{L}^{*'}\underline{Y}, [\underline{L}^{*'} - \underline{L}']\underline{Y}) &= E\{C(\underline{L}^{*'}\underline{Y}, (\underline{L}^{*'} - \underline{L}')\underline{Y} \mid \underline{\beta})\} \\
&\quad + C(E\{\underline{L}^{*'}\underline{Y} \mid \underline{\beta}\}, E\{[\underline{L}^{*'} - \underline{L}']\underline{Y} \mid \underline{\beta}\}) = \\
&= E\{\underline{L}^{*'}\underline{W}(\underline{L}^{*'} - \underline{L}')\} + C(\underline{L}^{*'}\underline{Z} \underline{\beta}, [\underline{L}^{*'} - \underline{L}']\underline{Z} \underline{\beta}) = \\
&= \underline{L}^{*'}\underline{W}(\underline{L}^{*'} - \underline{L}') + \underline{L}^{*'}\underline{Z} \underline{G} \underline{Z}'(\underline{L}^{*'} - \underline{L}') = \\
&= \underline{L}^{*'}(\underline{W} + \underline{Z}\underline{G}\underline{Z}')(\underline{L}^{*'} - \underline{L}') = \underline{P}'\underline{G}\underline{Z}'(\underline{W} + \underline{Z}\underline{G}\underline{Z}')^{-1}(\underline{W} + \underline{Z}\underline{G}\underline{Z}')(\underline{L}^{*'} - \underline{L}') \\
&= \underline{P}'\underline{G}\underline{Z}'(\underline{L}^{*'} - \underline{L}').
\end{aligned}$$

Com isto verificamos que

$$C(\underline{P}'\underline{\beta} - \underline{a}^{*'} - \underline{L}^{*'}\underline{Y}, [\underline{L}^* - \underline{L}]'\underline{Y}) = 0$$

e, como consequência,

$$\text{Var}(\underline{P}'\underline{\beta} - \underline{a} - \underline{L}'\underline{Y}) \geq \text{Var}(\underline{P}'\underline{\beta} - \underline{a}^{*'} - \underline{L}^{*'}\underline{Y}),$$

o que prova que $\text{Var}(\underline{P}'\underline{\beta} - \underline{a}^{*'} - \underline{L}^{*'}\underline{Y})$ é mínima.

Para provarmos que

$$(\underline{W} + \underline{Z}\underline{G}\underline{Z}')^{-1}\underline{Z}\underline{G}\underline{P} = \underline{W}^{-1}\underline{Z}(\underline{G}^{-1} + \underline{Z}'\underline{W}^{-1}\underline{Z})^{-1}\underline{P}$$

usaremos o resultado (Rao [1973])

$$(\underline{A} + \underline{B}\underline{D}\underline{B}')^{-1} = \underline{A}^{-1} - \underline{A}^{-1}\underline{B}(\underline{B}'\underline{A}^{-1}\underline{B} + \underline{D}^{-1})^{-1}\underline{B}'\underline{A}^{-1} \quad (\text{A.2.10})$$

onde $\underline{A}$ e $\underline{D}$ são matrizes não singulares, respectivamente $m \times m$ e $k \times k$ e $\underline{B}$ é uma matriz $m \times k$.

Nestas condições,

$$\underline{W}^{-1}\underline{Z}(\underline{G}^{-1} + \underline{Z}'\underline{W}^{-1}\underline{Z})^{-1}\underline{P} = \underline{W}^{-1}\underline{Z}[\underline{G} - \underline{G}\underline{Z}'(\underline{Z}\underline{G}\underline{Z}' + \underline{W})^{-1}\underline{Z}\underline{G}]\underline{P}$$

$$= [\underline{W}^{-1} - \underline{W}^{-1}\underline{Z}\underline{G}\underline{Z}'(\underline{Z}\underline{G}\underline{Z}' + \underline{W})^{-1}] \underline{Z}\underline{G}\underline{P} =$$

$$= \{\underline{W}^{-1} - \underline{W}^{-1}[(\underline{Z}\underline{G}\underline{Z}' + \underline{W})(\underline{Z}\underline{G}\underline{Z}')^{-1}]^{-1}\} \underline{Z}\underline{G}\underline{P} =$$

$$\begin{aligned}
& \{W^{-1}[(ZGZ' + W)(ZGZ')^{-1}] - W^{-1}\} [(ZGZ' + W)(ZGZ')^{-1}]^{-1} ZG\underline{P} = \\
& = \{W^{-1} + (ZGZ')^{-1} - W^{-1}\} (ZGZ')(ZGZ' + W)^{-1} ZG\underline{P} = \\
& = (ZGZ' + W)^{-1} ZG\underline{P},
\end{aligned}$$

o que encerra a demonstração.

b) Se $\underline{\mu}$ é desconhecido, devemos provar que

$$\underline{L}^* = W^{-1}Z(Z'W^{-1}Z)^{-1}\underline{P} \quad e$$

$a^* = 0$, desde que exista $\underline{L}$ tal que $Z'\underline{L} = \underline{P}$.

Observamos que, para estes valores de a^* e $\underline{L}^*$, $a^{*'} + \underline{L}^{*'}\underline{Y} = \underline{P}'(Z'W^{-1}Z)^{-1}Z'W^{-1}\underline{Y} = \underline{P}'\hat{\underline{\beta}}$, onde $\hat{\underline{\beta}}$ é o estimador de mínimos quadrados generalizados de $\underline{\beta}$, no caso em que $\underline{\beta}$ é fixo (não aleatório).

A condição $Z'\underline{L} = \underline{P}$ garante que $\underline{P}'\underline{\beta}$ é estimável, ou seja, $E(\underline{P}'\underline{\beta} - \underline{L}'\underline{Y}) = \underline{P}'\underline{\mu} - \underline{L}'Z\underline{\mu} = 0$. Esta condição é necessária quando $Z'W^{-1}Z$ é não inversível, caso em que $\hat{\underline{\beta}}$ não é único.

Temos que

$$E(\underline{P}'\underline{\beta} - \underline{L}^{*'}\underline{Y}) = \underline{P}'\underline{\mu} - \underline{P}'(Z'W^{-1}Z)^{-1}Z'W^{-1}Z\underline{\mu} = 0$$

e

$$\begin{aligned} \text{Var}(\underline{\tilde{P}}'\underline{\beta} - \underline{\tilde{L}}^*\underline{\gamma} - [\underline{\tilde{L}} - \underline{\tilde{L}}^*]'\underline{\gamma}) &= \text{Var}(\underline{\tilde{P}}'\underline{\beta} - \underline{\tilde{L}}^*\underline{\gamma}) + \\ &+ \text{Var}([\underline{\tilde{L}} - \underline{\tilde{L}}^*]'\underline{\gamma}) - 2C(\underline{\tilde{P}}'\underline{\beta} - \underline{\tilde{L}}^*\underline{\gamma}, [\underline{\tilde{L}} - \underline{\tilde{L}}^*]'\underline{\gamma}). \end{aligned}$$

Mas, como $\underline{Z}'\underline{\tilde{L}}' = \underline{\tilde{P}}$ e $\underline{Z}'\underline{\tilde{L}}^* = \underline{Z}'\underline{W}^{-1}\underline{Z}(\underline{Z}'\underline{W}^{-1}\underline{Z})^{-1}\underline{\tilde{P}} = \underline{\tilde{P}}$,

$$\begin{aligned} C(\underline{\tilde{P}}'\underline{\beta} - \underline{\tilde{L}}^*\underline{\gamma}, [\underline{\tilde{L}} - \underline{\tilde{L}}^*]'\underline{\gamma}) &= -(\underline{\tilde{L}}^* - \underline{\tilde{L}})'\underline{ZG}\underline{\tilde{P}} + \underline{\tilde{L}}^*\underline{W}(\underline{\tilde{L}}^* - \underline{\tilde{L}}) \\ &+ \underline{\tilde{L}}^*\underline{ZGZ}'(\underline{\tilde{L}}^* - \underline{\tilde{L}}) = 0 + \underline{\tilde{P}}'(\underline{Z}'\underline{W}^{-1}\underline{Z})^{-1}\underline{Z}'\underline{W}^{-1}\underline{W}(\underline{\tilde{L}}^* - \underline{\tilde{L}}) + 0 \\ &= \underline{\tilde{P}}'(\underline{Z}'\underline{W}^{-1}\underline{Z})^{-1}\underline{Z}'(\underline{\tilde{L}}^* - \underline{\tilde{L}}) = 0, \end{aligned}$$

o que prova o resultado. □

Verifica-se ainda que

$$\text{Var}(\underline{\tilde{P}}'\underline{\beta} - \underline{a}^* - \underline{\tilde{L}}^*\underline{\gamma}) = \underline{\tilde{P}}'\underline{G}\underline{\tilde{P}} + \underline{\tilde{L}}^*(\underline{W} + \underline{ZGZ}')\underline{\tilde{L}}^* - 2\underline{\tilde{P}}'\underline{GZ}'\underline{\tilde{L}}^*$$

e assim, para $\underline{\mu}$ conhecido,

$$\begin{aligned} \text{Var}(\underline{\tilde{P}}'\underline{\beta} - \underline{a}^* - \underline{\tilde{L}}^*\underline{\gamma}) &= \underline{\tilde{P}}'\underline{G}\underline{\tilde{P}} + \underline{\tilde{P}}'\underline{GZ}'\underline{\tilde{L}}^* - 2\underline{\tilde{P}}'\underline{GZ}'\underline{\tilde{L}}^* \\ &= \underline{\tilde{P}}'[\underline{G} - \underline{GZ}'(\underline{W} + \underline{ZGZ}')^{-1}\underline{GZ}]\underline{\tilde{P}} \\ &= \underline{\tilde{P}}'(\underline{G}^{-1} + \underline{Z}'\underline{W}^{-1}\underline{Z})^{-1}\underline{\tilde{P}} \end{aligned} \tag{A.2.11}$$

com a última igualdade sendo consequência de (A.2.10).

Para $\underline{\mu}$ desconhecido, $Z' \underline{L}^* = \underline{P}$ e

$$\begin{aligned} \text{Var}(\underline{P}' \underline{\beta} - \underline{a}^{*'} - \underline{L}^{*'} \underline{y}) &= \underline{L}^{*'} (W + ZGZ') \underline{L}^* - \underline{P}' G \underline{P} \\ &= \underline{P}' (Z' W^{-1} Z)^{-1} Z' W^{-1} (W + ZGZ') W^{-1} Z (Z' W^{-1} Z)^{-1} \underline{P} - \underline{P}' G \underline{P} \\ &= \underline{P}' (Z' W^{-1} Z)^{-1} \underline{P} + \underline{P}' G \underline{P} - \underline{P}' G \underline{P} \\ &= \underline{P}' (Z' W^{-1} Z)^{-1} \underline{P} . \end{aligned} \quad (\text{A.2.12})$$

II) O melhor previsor de mínimos quadrados Bayesiano restrito de $\underline{\ell}' \underline{y}$

Adotando o modelo

$$E(\underline{y} \mid \underline{\beta}) = X \underline{\beta}$$

$$\text{Var}(\underline{y} \mid \underline{\beta}) = V$$

e onde

$$E(\underline{\beta} \mid \underline{\mu}) = \underline{\mu} \quad e$$

$$\text{Var}(\underline{\beta} \mid \underline{\mu}) = \Sigma, \quad (\text{A.2.13})$$

com V e Σ conhecidas, Rodrigues [1989] determina o melhor preditor de mínimos quadrados Bayesiano restrito de $\underline{\ell}'\underline{y}$.

Se $\underline{T}'$ é um vetor de constantes, n -dimensional, a classe dos preditores lineares de $\underline{\ell}'\underline{y}$ é definida como

$$L = \{ \hat{\theta} \mid \hat{\theta} = a + \underline{T}'\underline{y}_s, \quad a \text{ escalar} \}$$

e a classe dos preditores de mínimos quadrados Bayesianos restritos é dada por

$$U_L = \{ \hat{\theta} \mid \hat{\theta} \in L \quad \text{e} \quad E(\hat{\theta} - \underline{\ell}'\underline{y}) = 0 \}.$$

Definição A.2.5 - No modelo (A.2.13), $\hat{\theta}_B^*$ é o melhor preditor de mínimos quadrados Bayesiano restrito de $\underline{\ell}'\underline{y}$ se

$$i) \quad \hat{\theta}_B^* = a^* + \underline{T}'^*\underline{y}_s \in U_L \quad \text{e}$$

$$ii) \quad EQM(\hat{\theta}_B^*) = E(\hat{\theta}_B^* - \underline{\ell}'\underline{y})^2 \leq E(\hat{\theta} - \underline{\ell}'\underline{y})^2 = EQM(\hat{\theta}), \quad \forall \hat{\theta} \in U_L.$$

Lema A.2.3

No modelo (A.2.13), $\hat{\theta} \in U_L$ se e somente se

$$E(a + \underline{F}'\underline{y}_s - \underline{p}'\underline{\beta}) = 0, \text{ onde}$$

$$\tilde{F}' = \tilde{T}' - \tilde{\ell}' K' V_s^{-1}, \quad \tilde{P}' = \tilde{\ell}' (X - K' V_s^{-1} X_s), \quad K = \begin{bmatrix} V_s & V_{sr} \end{bmatrix}.$$

Lema A.2.4

Sob o modelo (A.2.13), para todo a , $\tilde{T}'$ e $\hat{\theta} = a + \tilde{T}' y_s$, vale

$$\begin{aligned} E(\hat{\theta} - \tilde{\ell}' y)^2 &= \text{Var}(a + \tilde{F}' y_s - \tilde{P}' \beta) + \tilde{\ell}' V \tilde{\ell} - \tilde{\ell}' K' V_s^{-1} K \tilde{\ell} \\ &\quad + [a + (\tilde{T}' X_s - \tilde{\ell}' X) \mu]^2, \end{aligned}$$

onde $\tilde{F}'$ e $\tilde{P}'$ são os vetores definidos no lema A.2.3.

Se $\hat{\theta} \in U_L$, $E(a + \tilde{T}' y_s - \tilde{\ell}' y) = a + (\tilde{T}' X_s - \tilde{\ell}' X) \mu = 0$ e, devido aos lemas A.2.3 e A.2.4, obter $\hat{\theta}_B^*$ é equivalente a determinar $a + \tilde{F}' y_s$ tal que

$$E(a + \tilde{F}' y_s - \tilde{P}' \beta) = 0 \quad e$$

$$\text{Var}(a + \tilde{F}' y_s - \tilde{P}' \beta) \quad \text{é mínima.}$$

Estes valores ótimos de a e $\tilde{F}'$, que indicaremos por a^* e $\tilde{F}^*$, são obtidos diretamente do teorema A.2.2.

Para μ conhecido, fazendo as devidas adaptações, temos

$$\tilde{F}^* = (V_s + X_s \Sigma X_s')^{-1} X_s \Sigma \tilde{P} = V_s^{-1} X_s (\Sigma^{-1} + X_s' V_s^{-1} X_s)^{-1} \tilde{P}$$

e

$$a^* = \underline{\mu}' \underline{P} - \underline{\mu}' \underline{X}'_s \underline{F}^*.$$

Lembrando que $\underline{F}^{*'} = \underline{T}^{*'} - \underline{\ell}' \underline{K}' \underline{V}_s^{-1}$ e $\underline{P}' = \underline{\ell}' (\underline{X} - \underline{K}' \underline{V}_s^{-1} \underline{X}_s)$,
após cálculos verifica-se que

$$\hat{\theta}_B^* = a^{*'} + \underline{T}^{*'} \underline{y}_s$$

$$= \underline{\ell}'_s \underline{y}_s + \underline{\ell}'_r [\underline{X}_r \bar{\underline{\beta}} + \underline{V}_{rs} \underline{V}_s^{-1} (\underline{y}_s - \underline{X}_s \bar{\underline{\beta}})],$$

onde

$$\bar{\underline{\beta}} = \underline{C} \hat{\underline{\beta}} + (\underline{I} - \underline{C}) \underline{\mu},$$

$$\hat{\underline{\beta}} = (\underline{X}'_s \underline{V}_s^{-1} \underline{X}_s)^{-1} \underline{X}'_s \underline{V}_s^{-1} \underline{y}_s,$$

$$\underline{C} = (\underline{\Sigma}^{-1} + \underline{X}'_s \underline{V}_s^{-1} \underline{X}_s)^{-1} (\underline{X}'_s \underline{V}_s^{-1} \underline{X}_s)$$

e $\underline{I}$ é a matriz identidade de ordem $p+1$.

Se $\underline{\mu}$ é desconhecido,

$$\underline{F}^* = \underline{V}_s^{-1} \underline{X}_s (\underline{X}'_s \underline{V}_s^{-1} \underline{X}_s)^{-1} \underline{P}, \quad a^* = 0$$

e assim,

$$\begin{aligned}
\hat{\theta}_B^* &= \underline{P}' (X_s' V_s^{-1} X_s)^{-1} X_s' V_s^{-1} y_s + \underline{\ell}' K' V_s^{-1} y_s \\
&= \underline{\ell}' (X - K' V_s^{-1} X_s) \hat{\beta} + \underline{\ell}' K' V_s^{-1} y_s \\
&= \underline{\ell}'_s y_s + \underline{\ell}'_r [X_r \hat{\beta} + V_{rs} V_s^{-1} (y_s - X_s \hat{\beta})].
\end{aligned}$$

Observamos que quando $\underline{\mu}$ é desconhecido, o melhor previsor de mínimos quadrados Bayesiano restrito coincide com o previsor linear ótimo de $\underline{\ell}' y$, dado no teorema A.2.1.

^ APENDICE 3 - RESULTADOS DA TEORIA DE MODELOS LINEARES

Seja o modelo

$$Y_i = \beta_0 x_{i0} + \beta_1 x_{i1} + \dots + \beta_p x_{ip} + \varepsilon_i$$

com $E(\varepsilon_i) = 0$, $\text{Var}(\varepsilon_i) = \sigma^2$, para $i = 1, 2 \dots n$, $n > p+1$, onde n é o tamanho da amostra e as variáveis $y_1, y_2 \dots y_n$ são duas a duas não correlacionadas.

Este modelo pode ser escrito na forma matricial como:

$$y = X \beta + \varepsilon \quad (\text{A.3.1})$$

onde

$$y = \begin{bmatrix} Y_1 \\ Y_2 \\ \vdots \\ Y_n \end{bmatrix}, \quad X = \begin{bmatrix} x_{10} & x_{11} & \dots & x_{1p} \\ x_{20} & x_{21} & & x_{2p} \\ \vdots & \vdots & & \vdots \\ x_{n0} & x_{n1} & \dots & x_{np} \end{bmatrix}, \quad \beta = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_p \end{bmatrix}, \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_n \end{bmatrix}$$

com $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2 I$.

Como consequência, temos que $E(y) = X\beta$ e $\text{Var}(y) = \sigma^2 I$, onde β e σ^2 são os parâmetros do modelo, em geral desconhecidos.

Se tomarmos $x_{i0} = 1$ para $i = 1, 2 \dots n$, obtemos o modelo de regressão linear com intercepto.

Admitiremos ainda que a matriz X é de posto completo.

O estimador de mínimos quadrados de β , que, dentre todos os estimadores $\hat{\beta}$, minimiza a soma de quadrados do resíduo $(y - X\hat{\beta})'(y - X\hat{\beta})$, é dado por

$$\hat{\beta} = (X'X)^{-1}X'y.$$

As propriedades de $\hat{\beta}$ estão resumidas no próximo teorema, conhecido na literatura como Teorema de Gauss-Markov.

Teorema A.3.1

No modelo (A.3.1), se $\underline{\ell}' = [\ell_0, \ell_1 \dots \ell_p]$ é um vetor de constantes e $\hat{\beta} = (X'X)^{-1}X'y$, então $\underline{\ell}'\hat{\beta}$ tem variância mínima na classe de todos os estimadores lineares não viciados de $\underline{\ell}'\beta$.

A prova deste teorema pode ser encontrada em Rao [1973].

No caso mais geral, em que $\text{Var}(y) = \sigma^2 V$, onde V é uma matriz $n \times n$, de posto completo, positiva definida e conhecida, utiliza-se o estimador de mínimos quadrados generalizados

$$\hat{\beta}^* = (X'V^{-1}X)^{-1}X'V^{-1}y,$$

que minimiza a soma de quadrados do resíduo ponderada:

$$(y - X\hat{\beta})'V^{-1}(y - X\hat{\beta}).$$

Teorema A.3.2

No modelo (A.3.1), com $\text{Var}(\underline{y}) = \sigma^2 V$, $\underline{\ell}' \hat{\underline{\beta}}^*$ tem variância mínima na classe de todos os estimadores lineares não viciados de $\underline{\ell}' \underline{\beta}$.

Prova:

Se V é positiva definida e de posto completo, então V pode ser escrita como $P \cdot P' = P^2$, onde P é simétrica, não singular e única.

Multiplicando os dois membros do modelo $\underline{y} = X\underline{\beta} + \underline{\varepsilon}$ por P^{-1} , temos

$$P^{-1} \underline{y} = P^{-1} X \underline{\beta} + P^{-1} \underline{\varepsilon} \quad \text{ou equivalentemente,}$$

$$\underline{z} = Q \underline{\beta} + \underline{\varepsilon}^*, \quad \text{onde } \underline{z} = P^{-1} \underline{y}, \quad Q = P^{-1} X \quad \text{e} \quad \underline{\varepsilon}^* = P^{-1} \underline{\varepsilon}.$$

$$E(\underline{\varepsilon}^*) = P^{-1} E(\underline{\varepsilon}) = \underline{0} \quad \text{e}$$

$$\text{Var}(\underline{\varepsilon}^*) = \text{Var}(P^{-1} \underline{\varepsilon}) = P^{-1} \text{Var}(\underline{\varepsilon}) P^{-1'} = P^{-1} V P^{-1} \sigma^2$$

$= P^{-1} P P P^{-1} \sigma^2 = \sigma^2 I$, o que mostra que o modelo transformado satisfaz às condições do modelo (A.3.1).

Assim sendo, pelo teorema A.1.1, o estimador $\underline{\ell}' \hat{\underline{\beta}} = \underline{\ell}' (Q' Q)^{-1} Q' \underline{z}$ tem variância mínima na classe dos estimadores

lineares não viciados de $\underline{\ell}'\beta$.

Mas $\underline{\ell}'\hat{\beta} = \underline{\ell}'(Q'Q)^{-1}Q'z = \underline{\ell}'(X'P^{-1}P^{-1}X)^{-1}X'P^{-1}P^{-1}y$
 $= \underline{\ell}'(X'V^{-1}X)^{-1}X'V^{-1}y = \underline{\ell}'\hat{\beta}^*$, pois $P^{-1}P^{-1} = V^{-1}$, o que encerra a demonstração.

□

Passaremos agora a estudar o modelo

$y = X\beta + \varepsilon$, com $E(\varepsilon) = 0$, $\text{Var}(\varepsilon) = \sigma^2 I$, supondo adicionalmente que ε e portanto y tem distribuição normal n -variada.

Teorema A.3.3

Se $y \sim N_n(X\beta, \sigma^2 I)$, $\hat{\beta} = (X'X)^{-1}X'y$ e $\underline{\ell}' = [\ell_0, \ell_1 \dots \ell_p]$, então $\underline{\ell}'\hat{\beta}$ é o estimador de variância mínima na classe de todos os estimadores não viciados de $\underline{\ell}'\beta$.

Prova:

Sob este modelo, prova-se que $(\hat{\beta}, \hat{\sigma}^2)$, onde

$$\hat{\sigma}^2 = \frac{(y - X\hat{\beta})'(y - X\hat{\beta})}{n - p - 1}, \text{ é uma estatística suficiente e}$$

completa para (β, σ^2) (Arnold [1981]).

Além disso,

$$E(\underline{\ell}'\hat{\underline{\beta}}) = \underline{\ell}'(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{X}\underline{\beta} = \underline{\ell}'\underline{\beta},$$

ou seja, $\underline{\ell}'\hat{\underline{\beta}}$ é não viciado para $\underline{\ell}'\underline{\beta}$ e, sendo função de uma estatística suficiente e completa é, pelo teorema de Lehmann-Scheffé, o estimador não viciado de variância mínima de $\underline{\ell}'\underline{\beta}$.

□

Para o caso em que $\underline{y} \sim N_n(\underline{X}\underline{\beta}, \sigma^2\underline{V})$ vale o teorema seguinte.

Teorema A.3.4

Se $\underline{y} \sim N_n(\underline{X}\underline{\beta}, \sigma^2\underline{V})$, $\hat{\underline{\beta}}^* = (\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{y}$ e $\underline{\ell}' = [\ell_0, \ell_1 \dots \ell_p]$, $\underline{\ell}'\hat{\underline{\beta}}^*$ é o estimador não viciado de variância mínima de $\underline{\ell}'\underline{\beta}$.

Prova:

$$\underline{y} \sim N_n(\underline{X}\underline{\beta}, \sigma^2\underline{V}) \Leftrightarrow \underline{P}^{-1}\underline{y} \sim N_n(\underline{P}^{-1}\underline{X}\underline{\beta}, \sigma^2\underline{I}),$$

onde $\underline{P}^{-1}$ é tal que $\underline{V} = \underline{P}\underline{P}$.

Assim, de acordo com o teorema A.3.3, o estimador não viciado de variância mínima de $\underline{\ell}'\underline{\beta}$ é dado por

$$\underline{\ell}'(\underline{X}'\underline{P}^{-1}\underline{P}^{-1}\underline{X})^{-1}\underline{X}'\underline{P}^{-1}\underline{P}^{-1}\underline{y} = \underline{\ell}'(\underline{X}'\underline{V}^{-1}\underline{X})^{-1}\underline{X}'\underline{V}^{-1}\underline{y}.$$

□

Neste apêndice, por estarmos apresentando resultados da teoria de modelos lineares usual, que é destinada a populações infinitas, os vetores e matrizes presentes nos teoremas A.3.1 a A.3.4 são n -dimensionais, onde n é o tamanho da amostra.

O vetor y deste apêndice corresponderia a y_s dos capítulos 2 e 3 e apêndice 2 e as matrizes X e V correspondem, respectivamente, a X_s e V_s .

APÊNDICE 4 - RESULTADOS ASSINTÓTICOS

A.4.1 - Introdução

Neste apêndice apresentamos os dois teoremas assintóticos básicos que motivaram as extensões apresentadas no capítulo 4.

As demonstrações destes teoremas estão um pouco mais detalhadas que as presentes nos artigos para que fique claro de que forma os teoremas assintóticos tradicionais são aplicados na convergência em populações finitas e também para facilitar e complementar as demonstrações feitas no capítulo 4.

Introduziremos ainda definições e teoremas auxiliares utilizados na derivação dos resultados daquele capítulo.

A.4.2 - Teorema do Limite Central Unificado no Modelo de Locação

Fuller [1987] utiliza a abordagem de modelos de superpopulação, trabalhando com uma sequência de populações finitas, indexadas por r e admitindo que y_{ri} , $i = 1, 2, \dots, N_r$, $r = 1, 2, \dots$ são variáveis aleatórias independentes, identicamente distribuídas com $E(y_{ri}) = \mu$ e $\text{Var}(y_{ri}) = \sigma_{yy}$. N_r é o tamanho da r -ésima população e $\mathcal{P}_r = \{1, 2, \dots, N_r\}$, $r = 1, 2, \dots$ a sequência de populações finitas.

São definidos

$\bar{Y}_r = \frac{\sum_{i=1}^n Y_{ri}}{n_r}$, a média amostral da variável aleatória Y , com base na amostra de tamanho n_r da r -ésima população e

$\bar{Y}_r = \frac{\sum_{i=1}^N Y_{ri}}{N_r}$, a média populacional dos valores de Y observados para a r -ésima população.

O teorema proposto estabelece a distribuição assintótica

$$\sqrt{n_r} \begin{bmatrix} \bar{Y}_r - \mu \\ \bar{Y}_r - \bar{Y}_r \end{bmatrix} \text{ e sua demonstração utiliza o teorema de}$$

Cramér-Wold (Serfling [1980]) e o de Lindeberg-Feller para arranjos duplos (Billingsley [1968]).

Teorema A.4.1 (Cramér-Wold)

A seqüência de vetores aleatórios k -dimensionais $\underline{X}_n$ converge em distribuição para o vetor aleatório $\underline{X}$ se e somente se cada combinação linear das componentes de $\underline{X}_n$ converge em distribuição para a mesma combinação linear das componentes de $\underline{X}$. □

Como consequência deste teorema, se $\{\underline{X}_n\}_{n \geq 1}$ é uma seqüência de vetores k -dimensionais com média $\underline{\mu}'_n = [\mu_{n_1}, \mu_{n_2}, \dots, \mu_{n_k}]$ e matriz de covariância Σ_n positiva definida, tal que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=1}^n \Sigma_j = \Sigma, \text{ então}$$

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n (\underline{X}_j - \underline{\mu}_j) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_k(\underline{0}, \Sigma)$$

se e somente se, para todo $\underline{t} \in \mathbb{R}^k$,

$$\frac{1}{\sqrt{n}} \sum_{j=1}^n \underline{t}' (\underline{X}_j - \underline{\mu}_j) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_1(0, \underline{t}' \Sigma \underline{t}).$$

A seqüência de variáveis aleatórias $\{X_{nj}, 1 \leq j \leq j_n, n \geq 1\}$ é denominada "arranjo duplo" e pode ser disposta na forma:

$$\begin{array}{c} X_{11}, X_{12} \dots X_{1j_1} \\ X_{21}, X_{22} \dots X_{2j_2} \\ \vdots \\ X_{n1}, X_{n2} \dots X_{nj_n} \end{array}$$

Neste tipo de arranjo, geralmente admite-se que para todo $n \geq 1$, $X_{n1}, X_{n2} \dots X_{nj_n}$ são variáveis aleatórias independentes, ou seja, supõe-se independência dentro das linhas e possível dependência entre as colunas.

Teorema A.4.2 (Teorema de Lindeberg-Feller para arranjos duplos)

Seja $\{X_{nj}, 1 \leq j \leq j_n, n \geq 1\}$ um arranjo duplo de variáveis aleatórias com $X_{n1}, X_{n2} \dots X_{nj_n}$ independentes para todo $n \geq 1$ e tal que $j_n \rightarrow \infty$ para $n \rightarrow \infty$. Se $E(X_{nj}) = 0$, $\text{Var}(X_{nj}) = \sigma_{nj}^2$ e F_{nj} é a

função distribuição de X_{nj} , $n \geq 1$, $1 \leq j \leq n_j$, sejam

$$S_n = X_{n1} + X_{n2} + \dots + X_{nj_n} \quad e$$

$$s_n^2 = \sum_{j=1}^{j_n} \text{Var}(X_{nj}) = \sigma_{n1}^2 + \sigma_{n2}^2 + \dots + \sigma_{nj_n}^2,$$

com $s_n^2 > 0$, $n \geq 1$.

Nestas condições, se

$$\lim_{n \rightarrow \infty} \frac{1}{s_n^2} \sum_{j=1}^{j_n} \int_{\{|x| > \varepsilon s_n\}} x^2 dF_{nj} = 0 \text{ para todo } \varepsilon > 0$$

(Condição de Lindeberg para arranjos duplos),

$$\text{então } \frac{S_n}{s_n} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N(0, 1).$$

□

Apesar da suposição $E(X_{nj})=0$, o teorema também se aplica quando $E(X_{nj}) = \mu_{nj}$, bastando para isto trabalharmos com as variáveis $Z_{nj} = X_{nj} - \mu_{nj}$.

Teorema A.4.3 (Fuller [1987])

Se y_{ri} , $i=1,2,\dots,N_r$, $r=1,2,\dots$ são as variáveis aleatórias definidas anteriormente e se, para $r \rightarrow \infty$, $n_r \rightarrow \infty$ e $N_r \rightarrow \infty$ de

modo que $\frac{n_r}{N_r} \rightarrow f (0 \leq f < 1)$, então

$$\sqrt{n_r} \begin{bmatrix} \bar{y}_r - \mu \\ \bar{y}_r - \bar{y}_r \end{bmatrix} \xrightarrow[r \rightarrow \infty]{\mathcal{D}} N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1-f \\ 1-f & 1-f \end{bmatrix} \sigma_{yy} \right).$$

Prova:

Pelo teorema de Cramér-Wold, é suficiente provarmos que

$$A_r = k_1 \sqrt{n_r} (\bar{y}_r - \mu) + k_2 \sqrt{n_r} (\bar{y}_r - \bar{y}_r) \xrightarrow[r \rightarrow \infty]{\mathcal{D}} N_1,$$

para todo vetor $\begin{bmatrix} k_1 \\ k_2 \end{bmatrix} \in \mathbb{R}^2$, onde N_1 representa uma variável

aleatória com distribuição normal univariada.

Definindo

$$a_{r1i} = \begin{cases} \frac{1}{n_r} & \text{para } i = 1, 2, \dots, n_r \\ 0 & \text{para } i = n_r + 1, \dots, N_r \end{cases}$$

$$a_{r2i} = \begin{cases} \frac{1}{n_r} - \frac{1}{N_r} & \text{para } i = 1, 2, \dots, n_r \\ 0 & \text{para } i = n_r + 1, \dots, N_r, \end{cases}$$

verifica-se facilmente que

$$\bar{Y}_r - \mu = \sum_{i=1}^N a_{r1i} (Y_{ri} - \mu) \quad e$$

$$\bar{Y}_r - \bar{Y}_r = \sum_{i=1}^N a_{r2i} (Y_{ri} - \mu)$$

e assim,

$$A_r = k_1 \sqrt{n_r} \sum_{i=1}^N a_{r1i} (Y_{ri} - \mu) + k_2 \sqrt{n_r} \sum_{i=1}^N a_{r2i} (Y_{ri} - \mu).$$

Podemos, sem perda de generalidade tomar $k_1 = k$ e $k_2 = 1-k$, $0 \leq k \leq 1$, pois isto é equivalente a trabalharmos com

$$\lambda' = \left[\frac{k_1}{k_1 + k_2}, \frac{k_2}{k_1 + k_2} \right].$$

Nestas condições

$$\begin{aligned} A_r &= \sqrt{n_r} \sum_{i=1}^N [ka_{r1i} (Y_{ri} - \mu) + (1-k)a_{r2i} (Y_{ri} - \mu)] = \\ &= \sqrt{n_r} \sum_{i=1}^N (Y_{ri} - \mu) [ka_{r1i} + (1-k)a_{r2i}] = \sqrt{n_r} \sum_{i=1}^N a_{rki} (Y_{ri} - \mu), \end{aligned}$$

onde $a_{rki} = ka_{r1i} + (1-k)a_{r2i}$.

De acordo com a notação do teorema A.4.2, para cada $r = 1, 2, \dots$, sejam

$$X_{ri} = a_{rki} Y_{ri} - E(a_{rki} Y_{ri}) = a_{rki} (Y_{ri} - \mu),$$

$i = 1, 2, \dots, N_r$, variáveis aleatórias independentes,

$$S_r = X_{r1} + X_{r2} + \dots + X_{rN_r} = \sum_{i=1}^{N_r} a_{rki} (y_{ri} - \mu) \quad e$$

$$s_r^2 = \sum_{i=1}^{N_r} \text{Var}(X_{ri}) = \sum_{i=1}^{N_r} a_{rki}^2 \sigma_{yy}$$

Temos que

$$a_{rki}^2 = \begin{cases} \left[\frac{k}{n_r} + \frac{(1-k)(N_r - n_r)}{n_r N_r} \right]^2, & i = 1, 2, \dots, n_r \\ \left[\frac{-(1-k)}{N_r} \right]^2, & i = n_r + 1, \dots, N_r \end{cases}$$

$$\text{Como } \frac{N_r - n_r}{n_r} > 1 \quad e \quad 1-k \geq 0,$$

$$\left[\frac{(1-k)(N_r - n_r)}{n_r N_r} \right]^2 \geq \left[\frac{1-k}{N_r} \right]^2$$

e portanto,

$$\begin{aligned} \sup_{1 \leq i \leq N_r} a_{rki}^2 &= \left[\frac{k}{n_r} + \frac{(1-k)(N_r - n_r)}{n_r N_r} \right]^2 = \\ &= \left[\frac{kN_r + (1-k)(N_r - n_r)}{n_r N_r} \right]^2 = \left(\frac{N_r - n_r + kn_r}{n_r N_r} \right)^2 = \left(\frac{1}{n_r} - \frac{1-k}{N_r} \right)^2. \end{aligned}$$

Além disso

$$\frac{s_r^2}{\sigma_{yy}} = \sum_{i=1}^{N_r} a_{rki}^2 = n_r \left[\frac{k}{n_r} + \frac{(1-k)(N_r - n_r)}{n_r N_r} \right]^2 + (N_r - n_r) \left(\frac{1-k}{N_r} \right)^2 =$$

$$\begin{aligned}
&= n_r \left(\frac{(kn_r + N_r - n_r - kN_r + kn_r)^2}{n_r N_r^2} \right) + \frac{N_r - n_r}{N_r^2} (1 - k)^2 = \\
&= \frac{(N_r - n_r + kn_r)^2}{n_r N_r^2} + \frac{(N_r - n_r)(1 - k)^2}{N_r^2} . \quad (A.4.1)
\end{aligned}$$

Assim,

$$\lim_{r \rightarrow \infty} \frac{s_r^2}{\sup_{1 \leq i \leq N_r} a_{rki}^2} = \lim_{r \rightarrow \infty} \sigma_{yy} \frac{\frac{(N_r - n_r + kn_r)^2}{n_r N_r^2} + \frac{(N_r - n_r)(1 - k)^2}{N_r^2}}{\left(\frac{1}{n_r} - \frac{1 - k}{N_r} \right)^2} =$$

$$\lim_{r \rightarrow \infty} \sigma_{yy} \frac{1 + \frac{(N_r - n_r)(1 - k)^2}{N_r^2} + \frac{N_r^2 n_r}{(N_r - n_r + kn_r)^2}}{\left(\frac{N_r - n_r + kn_r}{n_r N_r} \right)^2 \frac{N_r^2 n_r}{(N_r - n_r + kn_r)^2}} =$$

$$\lim_{r \rightarrow \infty} \sigma_{yy} \frac{1 + \frac{(N_r - n_r)(1 - k)^2 n_r}{(N_r - n_r + kn_r)^2}}{\frac{1}{n_r}} =$$

$$\begin{aligned}
&= \lim_{r \rightarrow \infty} \sigma_{yy} \frac{1 + \frac{\left(\frac{n_r}{f_r} - n_r \right) (1 - k)^2 n_r}{\left(\frac{n_r}{f_r} - n_r + kn_r \right)^2}}{\frac{1}{n_r}} =
\end{aligned}$$

$$\begin{aligned}
&= \lim_{r \rightarrow \infty} \sigma_{yy} \frac{1 + \frac{n_r^2 \left(\frac{1}{f_r} - 1 \right) (1 - k)^2}{n_r^2 \left(\frac{1}{f_r} - 1 + k \right)^2}}{\frac{1}{n_r}} = \lim_{r \rightarrow \infty} \sigma_{yy} \frac{1 + \frac{\left(\frac{1}{f_r} - 1 \right) (1 - k)^2}{\left(\frac{1}{f_r} - 1 + k \right)^2}}{\frac{1}{n_r}} = \\
&= +\infty
\end{aligned}$$

Para verificarmos a condição de Lindeberg dada no teorema A.4.2, devemos verificar se, $\forall \varepsilon > 0$,

$$\lim_{r \rightarrow \infty} \frac{1}{s_r^2} \sum_{i=1}^N \int_{[a_{rki}(y-\mu)]^2 > \varepsilon^2 s_r^2} a_{rki}^2 (y - \mu)^2 dF_{y_{ri}}(y) = 0$$

onde $F_{y_{ri}}$ é a função distribuição da variável aleatória y_{ri} . Lembrando que as variáveis aleatórias y_{ri} , $1 \leq i \leq N_r$, $r = 1, 2, \dots$ são identicamente distribuídas, passaremos a indicar por F_y à sua função distribuição comum.

Observando também que

$$\left\{ y \mid (y - \mu)^2 > \frac{\varepsilon^2 s_r^2}{a_{rki}^2} \right\} \subset \left\{ y \mid (y - \mu)^2 > \frac{\varepsilon^2 s_r^2}{\sup_{1 \leq i \leq N_r} a_{rki}^2} \right\}$$

então

$$\begin{aligned}
& \frac{1}{s_r^2} \sum_{i=1}^N \int_{[a_{rki}(Y-\mu)]^2 > \varepsilon^2 s_r^2} a_{rki}^2 (Y - \mu)^2 dF_y(Y) \leq \frac{1}{s_r^2} \sum_{i=1}^N a_{rki}^2 \int_{(Y-\mu)^2 > \frac{\varepsilon^2 s_r^2}{\sup_{1 \leq i \leq N_r} a_{rki}^2}} (Y - \mu)^2 dF_y(Y) \\
& = \int_{(Y-\mu)^2 > \frac{\varepsilon^2 s_r^2}{\left(\frac{1}{n_r} - \frac{1-k}{N_r}\right)^2}} (Y - \mu)^2 dF_y(Y) \cdot \frac{1}{s_r^2} \sum_{i=1}^N a_{rki}^2 = \sigma_{yy}^{-1} \int_{(Y-\mu)^2 > \frac{\varepsilon^2 s_r^2}{\left(\frac{1}{n_r} - \frac{1-k}{N_r}\right)^2}} (Y - \mu)^2 dF_y(Y) \quad (A.4.2)
\end{aligned}$$

Temos que

$$\lim_{r \rightarrow \infty} \sigma_{yy}^{-1} \int_{(Y-\mu)^2 > \frac{\varepsilon^2 s_r^2}{\left(\frac{1}{n_r} - \frac{1-k}{N_r}\right)^2}} (Y - \mu)^2 dF_y(Y) = 0 \quad \text{pois}$$

$$\int_{-\infty}^{\infty} (Y - \mu)^2 dF_y(Y) = \sigma_{yy} \quad \text{e}$$

$$\frac{s_r^2}{\left(\frac{1}{n_r} - \frac{1-k}{N_r}\right)^2} \rightarrow \infty \quad \text{para } r \rightarrow \infty.$$

Devido a (A.4.2), fica provada a condição de Lindeberg para arranjos duplos e como consequência do teorema A.4.2,

$$\frac{1}{s_r} \sum_{i=1}^N a_{rki} (y_{ri} - \mu) \xrightarrow[r \rightarrow \infty]{\mathcal{D}} N(0, 1).$$

$$\text{Como } \frac{s_r^2}{\sigma_{yy}} n_r = \frac{(N_r - n_r + kn_r)^2}{N_r^2} + \frac{n_r (N_r - n_r) (1 - k)^2}{N_r^2}$$

(ver (A.4.1)), então

$$\begin{aligned} \lim_{r \rightarrow \infty} s_r^2 n_r &= \sigma_{yy} [(1 - f + kf)^2 + f(1 - f)(1 - k)^2] = \\ &= \sigma_{yy} [1 - (1 - k^2)f] \end{aligned}$$

e desta forma

$$\begin{aligned} \sqrt{n_r} \sum_{i=1}^N a_{rki} (y_{ri} - \mu) &= \frac{1}{s_r} \sum_{i=1}^N a_{rki} (y_{ri} - \mu) \sqrt{s_r^2 n_r} \\ &\xrightarrow[r \rightarrow \infty]{\mathcal{D}} N(0, [1 - (1 - k^2)f] \sigma_{yy}). \end{aligned}$$

Notando que

$$\begin{bmatrix} k & 1-k \end{bmatrix} \begin{bmatrix} 1 & 1-f \\ 1-f & 1-f \end{bmatrix} \sigma_{yy} \begin{bmatrix} k \\ 1-k \end{bmatrix} = (1-f + k^2 f) \sigma_{yy},$$

fica provado pelo teorema de Cramér-Wold que

$$\sqrt{n_r} \begin{bmatrix} \bar{y}_r - \mu \\ \bar{y}_r - \bar{Y}_r \end{bmatrix} \xrightarrow[r \rightarrow \infty]{\mathcal{D}} N_2 \left(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} 1 & 1-f \\ 1-f & 1-f \end{bmatrix} \sigma_{yy} \right).$$

□

A.4.3 - Teorema do Limite Central no Modelo de Regressão

Fuller [1975] trabalha com o modelo (4.2), utilizando a abordagem assintótica descrita na seção 4.2.1, que admite que $\{\mathcal{P}_n, n = 1, 2, \dots\}$ é a seqüência de populações finitas, $\mathcal{P}_n = \{1, 2, \dots, N_n\}$, com $N_n > N_{n-1}$, n é o tamanho de amostra da n -ésima população, $f_n = \frac{n}{N_n}$ é tal que $\lim_{n \rightarrow \infty} f_n = f, 0 \leq f < 1$.

Nestas condições, o autor determina a distribuição assintótica de $\sqrt{n}(\hat{\beta} - \underline{B})$, onde $\hat{\beta}$ e $\underline{B}$ são os vetores aleatórios definidos nas expressões (4.4) e (4.3). Na obtenção deste resultado assintótico são utilizadas várias definições e resultados clássicos, que, para uma maior clareza, exporemos preliminarmente.

Definição A.4.1 - Se $\{a_n\}_{n \geq 1}$ e $\{b_n\}_{n \geq 1}$ são seqüências de números reais, $a_n = o(b_n)$ se $\left| \frac{a_n}{b_n} \right|$ é limitada para $n \rightarrow \infty$, ou seja, se existir um número $m > 0$ e um inteiro $n_0 = n_0(m)$ tal que $\left| \frac{a_n}{b_n} \right| \leq m, \quad \forall n \geq n_0$.

Definição A.4.2 - Sejam $\{X_n\}_{n \geq 1}$ uma seqüência de variáveis aleatórias definidas num mesmo espaço de probabilidades e $\{b_n\}_{n \geq 1}$ uma seqüência de números reais. Nestas condições, $X_n = o_p(b_n)$ se, dado $\eta > 0$, existe $n_0(\eta)$ e $\varepsilon(\eta)$ tal que $P(|X_n| > \varepsilon b_n) < \eta$ para todo $n \geq n_0$.

Definição A.4.3 - Se $\underline{X}'_n = [X_{n1}, X_{n2} \dots X_{nk}]$ é um vetor k -dimensional e $\{b_n\}_{n \geq 1}$ uma sequência de números reais,

$$\underline{X}_n = o_p(b_n), \text{ se } \|\underline{X}_n\| = o_p(b_n), \text{ onde } \|\underline{X}_n\| = \sqrt{\sum_{j=1}^k X_{nj}^2}.$$

Verifica-se que $\underline{X}_n = o_p(b_n)$ se e somente se $X_{nj} = o_p(b_n)$, para $j = 1, 2 \dots k$.

Propriedades:

1) Sejam $\{X_n\}_{n \geq 1}$ e $\{Y_n\}_{n \geq 1}$ seqüências de variáveis aleatórias e $\{b_n\}_{n \geq 1}$ e $\{c_n\}_{n \geq 1}$ seqüências de números reais. Se $X_n = o_p(b_n)$ e $Y_n = o_p(c_n)$, então

$$a) X_n Y_n = o_p(b_n) o_p(c_n) = o_p(b_n c_n)$$

$$b) X_n + Y_n = o_p(\max\{|b_n|, |c_n|\})$$

$$c) |X_n|^s = o_p(|b_n|^s), \forall s > 0.$$

2) Se $\{b_n\}_{n \geq 1}$ e $\{c_n\}_{n \geq 1}$ são seqüências de números reais tais que $X_n = o_p(b_n)$ e $b_n = o(c_n)$, então

$$X_n = o_p(o(c_n)) = o_p(c_n).$$

3) Se $X_n = o_p(b_n)$ e $Y_n = o_p(c_n)$, então

$$X_n Y_n = o_p(b_n) o_p(c_n) = o_p(b_n c_n).$$

Teorema A.4.4 (Leite e Singer [1990])

Se $\{X_n\}_{n \geq 1}$ é uma sequência de variáveis aleatórias com $\mu_n = E(X_n)$ e $\text{Var}(X_n) = \sigma_n^2 < \infty$, então

$$X_n - \mu_n = o_p(\sigma_n).$$

Teorema A.4.5

Se $\{X_n\}_{n \geq 1}$ e $\{Y_n\}_{n \geq 1}$ são sequências de variáveis aleatórias tais que $X_n - Y_n \xrightarrow{P} 0$ ($X_n - Y_n$ converge em probabilidade para zero) e $X_n \xrightarrow{D} X$, então $Y_n \xrightarrow{D} X$.

Teorema A.4.6 - (Teorema do Limite Central Multivariado Clássico Serfling [1980])

Seja $\{\underline{X}_n\}_{n \geq 1}$ uma sequência de vetores aleatórios k -dimensionais ($\underline{X}'_n = [X_{n1}, X_{n2} \dots X_{nk}]$), independentes e identicamente distribuídos com vetor de médias $\underline{\mu}' = [\mu_1, \mu_2, \dots, \mu_k]$ e matriz de covariância Σ , positiva definida. Nestas condições,

$$\frac{\sum_{i=1}^n (\underline{X}_i - \underline{\mu})}{\sqrt{n}} \xrightarrow[n \rightarrow \infty]{D} N_k(\underline{0}, \Sigma).$$

Os resultados dos teoremas A.4.7 e A.4.8 foram utilizados nas demonstrações dos teoremas 4.1 e 4.2 e são freqüentemente

utilizados na literatura, sem maiores referências quanto à sua demonstração.

Apresentaremos a seguir os referidos resultados, com os passos básicos de suas demonstrações, que envolvem conceitos de análise e topologia.

Teorema A.4.7

Seja A_n , $n \geq 1$ uma seqüência de matrizes não aleatórias, inversíveis, tais que

$$A_n \xrightarrow{n \rightarrow \infty} A,$$

ou seja, cada elemento de A_n converge para o correspondente elemento de A . Nestas condições,

$$A_n^{-1} \xrightarrow{n \rightarrow \infty} A^{-1}.$$

Prova:

Sejam

$M(p) = \{Y \mid Y \text{ é matriz } p \times p, \text{ com elementos reais}\}$ e a aplicação

$r: G(p) \longrightarrow G(p)$

$$r(Y) = Y^{-1}, \quad \forall Y \in G(p),$$

onde $G(p) = \{W \in M(p) \mid W \text{ é inversível}\}.$

Verifica-se (Lima [1969]) que $G(p)$ é um espaço métrico e que a aplicação r é contínua.

Como consequência, dadas Y_n , $n \geq 1$ e Y , pertencentes a $G(p)$, se $Y_n \xrightarrow{n \rightarrow \infty} Y$ então $r(Y_n) \xrightarrow{n \rightarrow \infty} r(Y)$.

Portanto,

$$A_n \xrightarrow{n \rightarrow \infty} A \Rightarrow A_n^{-1} \xrightarrow{n \rightarrow \infty} A^{-1}.$$

□

Teorema A.4.8

Se D_n , $n \geq 1$ é uma seqüência de matrizes estocásticas inversíveis e D uma matriz inversível tal que

$$D_n - D = o_p(n^{-1/2}),$$

ou seja, $d_{n_{rs}} - d_{rs} = o_p(n^{-1/2}) \quad \forall r, s$, com $d_{n_{rs}}$ e d_{rs} representando os elementos respectivamente de D_n e D , então

$$D_n^{-1} - D^{-1} = o_p(n^{-1/2}).$$

Prova:

Demonstraremos inicialmente que

$D_n - D = o_p(n^{-1/2}) \Rightarrow D_n^{-1} - D^{-1} = o_p(n^{-1/2})$ para D_n , $n \geq 1$ matrizes não estocásticas.

Lima [1970] verifica que, além de contínua, a transformação

$r(Y) = Y^{-1}$ é diferenciável em cada $Y \in G(p)$ e sua derivada é a transformação linear $H \longrightarrow -Y^{-1}HY^{-1}$.

Adicionalmente, tem-se que

$$(Y + H)^{-1} = Y^{-1} - Y^{-1}HY^{-1} + g(H),$$

com $\lim_{H \rightarrow 0} \frac{g(H)}{\|H\|} = 0$, onde $\|H\|$ representa uma norma qualquer associada à matriz H (pois todas as normas são equivalentes).

Seja

$$H = D_n - D, \quad Y = D,$$

então

$$D_n^{-1} = D^{-1} - D^{-1}(D_n - D)D^{-1} + g(H)$$

$$\text{com } \lim_{H \rightarrow 0} \frac{g(H)}{\|H\|} = 0.$$

Tomando $\|H\| = \max\{h_{rs}, h_{rs} \text{ elementos de } H\}$,

$$\lim_{H \rightarrow 0} \frac{g(H)}{\|H\|} = 0 \Leftrightarrow \lim_{H \rightarrow 0} \frac{\|g(H)\|}{\|H\|} = 0.$$

Como $H = D_n - D = O(n^{-1/2})$ por hipótese, $\exists n_0$ tal que $|h_{n_{rs}}| < \varepsilon n^{-1/2}$, $\forall r, s, n \geq n_0 \Rightarrow \|H\| \leq \varepsilon n^{-1/2}$, $\forall n \geq n_0$.

Portanto, para todo $n \geq n_0$,

$$\frac{\|g(H)\|}{\varepsilon n^{-1/2}} \leq \frac{\|g(H)\|}{\|H\|},$$

$$\text{e } \lim_{\substack{H \rightarrow 0 \\ (n \rightarrow \infty)}} \frac{\|g(H)\|}{\|H\|} = 0 \Rightarrow \frac{1}{\varepsilon} \lim_{\substack{H \rightarrow 0 \\ (n \rightarrow \infty)}} \frac{\|g(H)\|}{n^{-1/2}} = 0$$

$$\Rightarrow \frac{\|g(H)\|}{n^{-1/2}} \leq \eta, \quad \forall n \geq k \Rightarrow \|g(H)\| = o(n^{-1/2})$$

$\Rightarrow g(H) = o(n^{-1/2})$, isto é, todos os elementos da matriz $g(H)$ são $o(n^{-1/2})$.

Além disso,

$$D_n - D = o(n^{-1/2}) \Rightarrow D^{-1}(D_n - D)D^{-1} = o(n^{-1/2}),$$

e desta forma

$$D_n^{-1} - D^{-1} = -D^{-1}(D_n - D)D^{-1} + g(H) = o(n^{-1/2}).$$

Fica então provado o resultado para matrizes não estocásticas.

O teorema 5.1.5 de Fuller [1976] mostra que se $\{X_n\}$ é uma sequência de variáveis aleatórias k -dimensionais com

$$X_{jn} = o_p(r_{jn}), \quad j = 1, 2, \dots, k$$

e se $\{a_n\}$ é uma sequência de vetores não aleatórios, k -dimensionais tal que

$$a_{jn} = o(r_{jn}), \quad j = 1, 2, \dots, k \Rightarrow g_n(a_n) = o(s_n), \text{ então}$$

$$g_n(X_n) = o_p(s_n) \quad (g_n \text{ sequência de funções}).$$

Assim, como provamos o resultado para matrizes não estocásticas, o correspondente para matrizes estocásticas sai como consequência direta deste teorema.

□

Teorema A.4.9 (Fuller [1975])

Sob o modelo (4.2), se $f_n = \frac{n}{N_n}$, $n = 1, 2, \dots$, $N_n > N_{n-1}$ e

$\lim_{n \rightarrow \infty} f_n = f$, $0 \leq f < 1$, então

$$\sqrt{n} (\hat{\beta} - B) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(0, (1-f)Q^{-1}GQ^{-1}).$$

Prova:

Sejam

$$\tilde{R}_n = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n (y_i - \sum_{j=0}^p \beta_j x_{ij}) \\ \sum_{i=1}^n x_{i1} (y_i - \sum_{j=0}^p \beta_j x_{ij}) \\ \vdots \\ \sum_{i=1}^n x_{ip} (y_i - \sum_{j=0}^p \beta_j x_{ij}) \end{bmatrix} = \frac{1}{n} \begin{bmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ip} \varepsilon_i \end{bmatrix}$$

e $\tilde{R}_{N_n}$ o vetor obtido substituindo-se n por N_n em $\tilde{R}_n$.

De acordo com a notação introduzida em (4.4),

$$\underline{R}_n = \begin{bmatrix} h_{n_0} \\ h_{n_1} \\ \vdots \\ h_{n_p} \end{bmatrix} - \begin{bmatrix} \sum_{j=0}^p \beta_j q_{n_{j0}} \\ \sum_{j=0}^p \beta_j q_{n_{j1}} \\ \vdots \\ \sum_{j=0}^p \beta_j q_{n_{jp}} \end{bmatrix} = \underline{H}_n - Q_n \underline{\beta}.$$

Portanto, $Q_n^{-1} \underline{R}_n = Q_n^{-1} \underline{H}_n - Q_n^{-1} Q_n \underline{\beta} = \hat{\underline{\beta}} - \underline{\beta}$ e analogamente, verifica-se que $\underline{B} - \underline{\beta} = Q_{N_n}^{-1} \underline{R}_{N_n}$.

Como x_{ir} , $i=1,2,\dots,N_n$ são variáveis aleatórias independentes e identicamente distribuídas, denotando $\text{Var}(x_{ir} x_{is})$ por σ_{rs}^2 , $\forall i = 1,2,\dots,N_n$,

$$\text{Var}(q_{n_{rs}}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_{ir} x_{is}\right) = \frac{1}{n^2} n \sigma_{rs}^2 = \frac{\sigma_{rs}^2}{n}.$$

Devido à suposição $E(|x_{ij} x_{ik} x_{il} x_{im}|) < \infty$, tomando $j = k = r$ e $l = m = s$ segue que $E(|x_{ir}^2 x_{is}^2|) < \infty$ e com isto σ_{rs}^2 é finita e

$$\frac{\text{Var}(q_{n_{rs}})}{\frac{1}{n}} = \sigma_{rs}^2 \text{ é limitada para } n \rightarrow \infty \text{ (pois é constante } \forall n), \text{ o}$$

que implica em $\text{Var}(q_{n_{rs}}) = O(n^{-1})$.

No teorema A.4.4, tomando $X_n = q_{n_{rs}}$, $\mu_n = E(q_{n_{rs}}) = E(x_{ir} x_{is}) = q_{rs}$

e $\sigma_n^2 = \text{Var}(q_{n_{rs}}) = O(n^{-1})$, temos

$$\sigma_n = O(n^{-1/2}) \quad \text{e}$$

$$q_{n_{rs}} - q_{rs} = O_p(\sigma_n) = O_p(O(n^{-1/2})) = O_p(n^{-1/2}).$$

Como este resultado vale para todo r e s , ou seja, para todos os elementos da matriz Q_n , então

$$Q_n - Q = O_p(n^{-1/2}).$$

Da mesma forma, prova-se que $q_{N_{n_{rs}}} - q_{rs} = O_p(N_n^{-1/2})$, ou seja, dado η , existe n_0 e ε tal que

$$P\left(|q_{N_{n_{rs}}} - q_{rs}| > \frac{\varepsilon}{N_n^{1/2}}\right) < \eta \text{ para todo } N_n \geq n_0.$$

$n^{1/2} < N_n^{1/2} \forall n$ e então,

$$\left\{|q_{N_{n_{rs}}} - q_{rs}| > \frac{\varepsilon}{n^{1/2}}\right\} \subset \left\{|q_{N_{n_{rs}}} - q_{rs}| > \frac{\varepsilon}{N_n^{1/2}}\right\} \Rightarrow$$

$$\Rightarrow P\left(|q_{N_{n_{rs}}} - q_{rs}| > \frac{\varepsilon}{n^{1/2}}\right) \leq P\left(|q_{N_{n_{rs}}} - q_{rs}| > \frac{\varepsilon}{N_n^{1/2}}\right) < \eta, \forall N_n \geq n_0.$$

Portanto, $q_{N_{n_{rs}}} - q_{rs} = O_p(n^{-1/2}) \forall r, s$ e assim, verifica-se

que

$$Q_{N_n} - Q = o_p(n^{-1/2}).$$

Pelo teorema A.4.8,

$$\begin{cases} Q_n - Q = o_p(n^{-1/2}) \\ Q_{N_n} - Q = o_p(n^{-1/2}) \end{cases} \Rightarrow \begin{cases} Q_n^{-1} - Q^{-1} = o_p(n^{-1/2}) \\ Q_{N_n}^{-1} - Q^{-1} = o_p(n^{-1/2}) \end{cases}$$

e conforme já verificado,

$$\begin{cases} \hat{\beta} - \beta = Q_n^{-1} R_n \\ \hat{B} - \beta = Q_{N_n}^{-1} R_{N_n} \end{cases} \Rightarrow \hat{\beta} - \hat{B} = Q_n^{-1} R_n - Q_{N_n}^{-1} R_{N_n}.$$

Com isto,

$$\begin{aligned} n^{1/2}(\hat{\beta} - \hat{B}) &= n^{1/2}[Q_n^{-1} R_n - Q_{N_n}^{-1} R_{N_n}] = \\ &= n^{1/2}\left[\left(Q^{-1} + o_p(n^{-1/2})\right)R_n - \left(Q^{-1} + o_p(n^{-1/2})\right)R_{N_n}\right] \\ &= n^{1/2}\left[Q^{-1}(R_n - R_{N_n}) + o_p(n^{-1/2})R_n - o_p(n^{-1/2})R_{N_n}\right] \\ &= n^{1/2}Q^{-1}(R_n - R_{N_n}) + n^{1/2}o_p(n^{-1/2})R_n - n^{1/2}o_p(n^{-1/2})R_{N_n}. \end{aligned}$$

$R_n = o_p(n^{-1/2})$ pois, se r_{n_k} é o k -ésimo elemento de R_n ,

$$\text{Var}(r_{n_k}) = \text{Var}\left(\frac{1}{n} \sum_{i=1}^n x_{ik} \varepsilon_i\right) = \frac{\text{Var}(x_{ik} \varepsilon_i)}{n} = \frac{E(x_{ik}^2 \varepsilon_i^2)}{n} = \frac{g_{kk}}{n}.$$

Desta forma, fica provado que

$$n^{1/2}(\hat{\underline{\beta}} - \underline{B}) = n^{1/2}Q^{-1}(\underline{R}_n - \underline{R}_{N_n}) + o_p(n^{-1/2}).$$

Seja $r_{n_k} - r_{N_{n_k}}$ o k -ésimo elemento de $\underline{R}_n - \underline{R}_{N_n}$, $k = 0, 1, \dots, p$

$$\begin{aligned} r_{n_k} - r_{N_{n_k}} &= \frac{1}{n} \sum_{i=1}^n x_{ik} \varepsilon_i - \frac{1}{N_n} \sum_{i=1}^{N_n} x_{ik} \varepsilon_i = \\ &= \frac{N_n \sum_{i=1}^n x_{ik} \varepsilon_i - n \sum_{i=1}^{N_n} x_{ik} \varepsilon_i}{nN_n} = \frac{(N_n - n) \sum_{i=1}^n x_{ik} \varepsilon_i - n \sum_{i=n+1}^{N_n} x_{ik} \varepsilon_i}{nN_n} \\ &= n^{-1}(1 - f_n) \sum_{i=1}^n x_{ik} \varepsilon_i - \frac{1}{\sqrt{N_n}} \sqrt{\frac{N_n - n}{N_n}} \frac{1}{\sqrt{N_n - n}} \sum_{i=n+1}^{N_n} x_{ik} \varepsilon_i = \\ &= n^{-1}(1 - f_n) \sum_{i=1}^n x_{ik} \varepsilon_i - \frac{1}{\sqrt{N_n}} (1 - f_n)^{1/2} (N_n - n)^{-1/2} \sum_{i=n+1}^{N_n} x_{ik} \varepsilon_i. \end{aligned}$$

Portanto,

$$n^{1/2}(\underline{r}_{n_k} - \underline{r}_{N_{n_k}}) = n^{-1/2}(1 - f_n) \sum_{i=1}^n x_{ik} \varepsilon_i - f_n^{1/2}(1 - f_n)^{1/2} (N_n - n)^{-1/2} \sum_{i=n+1}^{N_n} x_{ik} \varepsilon_i \quad (A.4.3),$$

e

$$n^{1/2}(\hat{\underline{\beta}} - \underline{B}) = Q^{-1} \begin{bmatrix} n^{-1/2}(1 - f_n) \sum_{i=1}^n \varepsilon_i - f_n^{1/2}(1 - f_n)^{1/2} (N_n - n)^{-1/2} \sum_{i=n+1}^{N_n} \varepsilon_i \\ \vdots \\ n^{-1/2}(1 - f_n) \sum_{i=1}^n x_{ip} \varepsilon_i - f_n^{1/2}(1 - f_n)^{1/2} (N_n - n)^{-1/2} \sum_{i=n+1}^{N_n} x_{ip} \varepsilon_i \end{bmatrix} + o_p(n^{1/2}).$$

$X_n = O_p(n^{-1/2}) \Rightarrow X_n \xrightarrow{P} 0$ e assim, devido ao teorema A.4.5, a distribuição limite de $n^{1/2}(\hat{\beta} - \underline{B})$ é a mesma de

$$Q^{-1} \begin{bmatrix} n^{-1/2}(1-f_n) \sum_{i=1}^n \varepsilon_i - f_n^{1/2}(1-f_n)^{1/2}(N_n-n)^{-1/2} \sum_{i=n+1}^N \varepsilon_i \\ \vdots \\ n^{-1/2}(1-f_n) \sum_{i=1}^n x_{ip} \varepsilon_i - f_n^{1/2}(1-f_n)^{1/2}(N_n-n)^{1/2} \sum_{i=n+1}^N x_{ip} \varepsilon_i \end{bmatrix} =$$

$$= Q^{-1} n^{1/2}(\underline{R}_n - \underline{R}_{N_n}).$$

Como $\underline{X}'_i \varepsilon_i = [\varepsilon_i, x_{i1}\varepsilon_i, x_{i2}\varepsilon_i \dots x_{ip}\varepsilon_i]$, $i = 1, 2, \dots, N_n$ são vetores aleatórios independentes e identicamente distribuídos com média 0 e matriz de covariância G, pelo teorema A.4.6,

$$\underline{Z}_n = n^{-1/2} \begin{bmatrix} \sum_{i=1}^n \varepsilon_i \\ \sum_{i=1}^n x_{i1} \varepsilon_i \\ \vdots \\ \sum_{i=1}^n x_{ip} \varepsilon_i \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(\underline{0}, G)$$

e como os vetores $\underline{X}_i \varepsilon_i$, $i=1, 2, \dots, N_n$ são identicamente distribuídos, a distribuição limite de

$$\tilde{W}_n = (N_n - n)^{-1/2} \begin{bmatrix} \sum_{i=n+1}^N \varepsilon_i \\ \vdots \\ \sum_{i=n+1}^N x_{ip} \varepsilon_i \end{bmatrix} \text{ é a mesma de}$$

$$\tilde{W}_n^* = (N_n - n)^{1/2} \begin{bmatrix} \sum_{i=1}^{N-n} \varepsilon_i \\ \vdots \\ \sum_{i=1}^{N-n} x_{ip} \varepsilon_i \end{bmatrix}, \text{ que, pelo teorema A.4.6 é } N_{p+1}(\underline{0}, G).$$

$$\left(\text{observar que } N_n - n = N_n \left(1 - \frac{n}{N_n} \right) = N_n (1 - f_n) \xrightarrow{n \rightarrow \infty} \infty \right)$$

Nestas condições, se $\underline{\lambda}' = [\lambda_0, \lambda_1 \dots \lambda_p]$ é um vetor de números reais, $\underline{\lambda} \neq \underline{0}$, então

$$S_{1n} = \underline{\lambda}' (1-f_n) \underline{Z}_n = \sum_{j=0}^p \lambda_j n^{-1/2} (1-f_n) \sum_{i=1}^n x_{ij} \varepsilon_i$$

$$\xrightarrow[n \rightarrow \infty]{D} N(0, (1-f)^2 \underline{\lambda}' G \underline{\lambda})$$

e

$$S_{2n} = f_n^{1/2} (1-f_n)^{1/2} \underline{\lambda}' \tilde{W}_n = \sum_{j=0}^p \lambda_j f_n^{1/2} (1-f_n)^{1/2} (N_n - n)^{-1/2} \sum_{i=n+1}^N x_{ij} \varepsilon_i$$

$$\xrightarrow[n \rightarrow \infty]{D} N(0, f(1-f) \underline{\lambda}' G \underline{\lambda}).$$

$\underline{Z}_n$ e $\tilde{W}_n$ são vetores aleatórios independentes e desta forma,

S_{1n} e S_{2n} são independentes $\forall n$. Como consequência,

$$S_{1n} - S_{2n} = \tilde{\lambda}' n^{1/2} (\tilde{R}_n - \tilde{R}_{N_n}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N(0, (1-f) \tilde{\lambda}' G \tilde{\lambda}), \forall \tilde{\lambda} \in \mathbb{R}^{p+1}, \tilde{\lambda} \neq 0$$

e, pelo teorema de Cramér - Wold, segue que

$$n^{1/2} (\tilde{R}_n - \tilde{R}_{N_n}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N(0, (1-f)G) \Rightarrow$$

$$n^{1/2} Q^{-1} (\tilde{R}_n - \tilde{R}_{N_n}) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(0, (1-f)Q^{-1}GQ^{-1}),$$

o que encerra a demonstração. □

Sob o mesmo modelo vale ainda o resultado:

$$\sqrt{n} (\hat{\beta} - \beta) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{p+1}(0, Q^{-1}GQ^{-1}).$$

Com relação às notações $r \geq 1$, n_r , N_r , $f_r = \frac{n_r}{N_r}$ e $n \geq 1$, N_n ,

$f_n = \frac{n}{N_n}$, utilizadas respectivamente nos teoremas A.4.3 e A.4.9,

observamos que a primeira exigiu um teorema do limite central para arranjos duplos, enquanto que com a segunda, somente teoremas de convergência usuais foram necessários.

Além de uma maior simplicidade, a notação $n \geq 1$, N_n é também mais geral. Lembrando que convergência em distribuição é convergência pontual da sequência de funções distribuição e que

se uma sequência converge, todas suas subsequências convergem, como $\{n_r, r \geq 1\} = \{n_1, n_2, \dots\}$ é uma subsequência de $\{n, n \geq 1\} = \{1, 2, \dots\}$, provando os resultados para $n \rightarrow \infty$, automaticamente estamos provando o caso $r \rightarrow \infty$, $n_r, N_r \rightarrow \infty$.

Por estes motivos, a notação n, N_n foi adotada em todo capítulo 4.

A.4.4 - Teoremas Auxiliares

O próximo teorema é fundamental em quase todas as demonstrações do capítulo 4.

Teorema A.4.10

Se $\tilde{X}_n$ e $\tilde{Y}_n$ são vetores aleatórios independentes para todo $n = 1, 2, \dots$ e se $\tilde{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{k_1}(\underline{\mu}_1, \Sigma_1)$ e $\tilde{Y}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{k_2}(\underline{\mu}_2, \Sigma_2)$, então

$$\begin{bmatrix} \tilde{X}_n \\ \tilde{Y}_n \end{bmatrix} \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{k_1 + k_2} \left(\begin{bmatrix} \underline{\mu}_1 \\ \underline{\mu}_2 \end{bmatrix}, \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \right).$$

Prova:

Se $\tilde{X}_n \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_{k_1}(\underline{\mu}_1, \Sigma_1)$, então

$$\varphi_{\tilde{X}_n}(\underline{t}_1) \xrightarrow[n \rightarrow \infty]{} \exp \left\{ i \underline{t}_1' \underline{\mu}_1 - \frac{\underline{t}_1' \Sigma_1 \underline{t}_1}{2} \right\}, \quad \forall \underline{t}_1 \in \mathbb{R}^{k_1}$$

onde $\varphi_{\underline{x}_n}$ é a função característica do vetor aleatório $\underline{x}_n$ e $\exp\left\{it'_1\mu_1 - \frac{t'_1\Sigma_1 t_1}{2}\right\}$ é a função característica da variável aleatória com distribuição normal k_1 -variada com média μ_1 e matriz de covariância Σ_1 .

Da mesma forma,

$$\underline{y}_n \xrightarrow{\mathcal{D}} N_{k_2}(\mu_2, \Sigma_2) \Rightarrow \varphi_{\underline{y}_n}(t_2) \xrightarrow{n \rightarrow \infty} \exp\left\{it'_2\mu_2 - \frac{t'_2\Sigma_2 t_2}{2}\right\}, \quad \forall t_2 \in \mathbb{R}^{k_2}.$$

Como $\underline{x}_n$ e $\underline{y}_n$ são independentes,

$$\begin{aligned} \varphi_{(\underline{x}_n, \underline{y}_n)}(t_1, t_2) &= E\left(e^{it'_1 \underline{x}_n + it'_2 \underline{y}_n}\right) = E\left(e^{it'_1 \underline{x}_n}\right) E\left(e^{it'_2 \underline{y}_n}\right) \\ &= \varphi_{\underline{x}_n}(t_1) \varphi_{\underline{y}_n}(t_2) \xrightarrow{n \rightarrow \infty} \exp\left\{it'_1\mu_1 - \frac{t'_1\Sigma_1 t_1}{2}\right\} \exp\left\{it'_2\mu_2 - \frac{t'_2\Sigma_2 t_2}{2}\right\} \\ &= \exp\left\{i[t'_1 \ t'_2] \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} - \frac{1}{2} [t'_1 \ t'_2] \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \begin{bmatrix} t_1 \\ t_2 \end{bmatrix}\right\} \\ &= \exp\left\{it'\mu - \frac{1}{2} t'vt\right\}, \end{aligned}$$

$$\text{onde } t' = [t'_1 \ t'_2] \quad , \quad v = \begin{bmatrix} \Sigma_1 & 0 \\ 0 & \Sigma_2 \end{bmatrix} \quad , \quad \mu = \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} \quad ,$$

e $\varphi_{(\underline{x}_n, \underline{y}_n)}(t_1, t_2)$ é a função característica do vetor aleatório

$$\begin{bmatrix} \underline{x}_n \\ \underline{y}_n \end{bmatrix}.$$

Mas $\exp\left\{it'\underline{\mu} - \frac{1}{2}t'Vt\right\}$ é a função característica do vetor aleatório com distribuição $N_{k_1+k_2}(\underline{\mu}, V)$ e assim, fica provado o resultado.

□

Teorema A.4.11 (Teorema de Liapunov Multivariado)

Seja $\{\underline{X}_n\}_{n \geq 1}$ uma seqüência de vetores aleatórios k-dimensionais independentes com vetor de médias $\underline{\mu}'_n = [\mu_{n_1}, \mu_{n_2} \dots \mu_{n_k}]$ e matrizes de covariância Σ_n , $n \geq 1$ tais que

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Sigma_i = \Sigma.$$

Se para algum $\delta > 0$, $\max_{1 \leq j \leq k} E\left(|X_{nj} - \mu_{nj}|^{2+\delta}\right) \leq c$, $\forall n \geq 1$, então

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\underline{X}_i - \underline{\mu}_i) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_k(\underline{0}, \Sigma).$$

□

O próximo teorema pode ser considerado a versão multivariada do teorema de Lindeberg - Feller.

Teorema A.4.12 (Serfling [1980])

Seja $\{\underline{X}_n\}_{n \geq 1}$ uma seqüência de vetores aleatórios k-dimensionais independentes, com vetor de médias $\underline{\mu}_n$, matriz de

covariância Σ_n e função distribuição $F_{\tilde{x}_n}$, $n \geq 1$. Se

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \Sigma_i = \Sigma \quad \text{e}$$

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_{\|\tilde{x} - \tilde{\mu}_i\| > \varepsilon \sqrt{n}} \|\tilde{x} - \tilde{\mu}_i\|^2 dF_{\tilde{x}_i}(\tilde{x}) = 0 \quad \text{para todo } \varepsilon > 0, \quad (\text{A.4.4})$$

onde $\|\tilde{x}\| = \left(\sum_{j=1}^k x_j^2 \right)^{1/2}$, $\tilde{x}' = [x_1, x_2 \dots x_k]$, então

$$\frac{1}{\sqrt{n}} \sum_{i=1}^n (\tilde{x}_i - \tilde{\mu}_i) \xrightarrow[n \rightarrow \infty]{\mathcal{D}} N_k(0, \Sigma).$$

□

REFERÊNCIAS BIBLIOGRÁFICAS

- ARNOLD, S.F. 1981. The theory of linear models and multivariate analysis. *New York, John Wiley*. 475p.
- BILLINGSLEY, P. 1968. Convergence of probability measures. *New York, John Wiley*. 253p.
- BJORNSTAD, J. F. 1990. Predictive likelihood: a review. *Statistical Science*, 5(1):242-265.
- BOLFARINE, H. & RODRIGUES, J. 1988. On the simple projection predictor in finite populations. *Australian Journal of Statistics*, 30(3):338-341.
- BOLFARINE, H. & ZACKS, S. 1992. Prediction theory for finite populations. *Springer - Verlag, New York*. 205p.
- CASSEL, C.M.; SARNDAL, C.E.; WRETMAN, J.H. 1977. Foundations of inference in survey sampling. *New York, John Wiley*. 192p.
- COCHRAN, W.G. 1977. Sampling techniques. 3. ed. *New York, John Wiley*. 428p.
- CONOVER, W.J. 1980. Practical nonparametric statistics. 2. ed. *New York, John Wiley*. 493p.
- ELIAN, S.N. 1984. Previsão linear em populações finitas. São Paulo. 172p. *Dissertação - Mestrado, IME-USP*.
- FULLER, W. A. 1975. Regression analysis for sample survey. *Sankhya. Series C*, 37:117-132.
- FULLER, W.A. 1976. Introduction to statistical time series. *New York, John Wiley*. 470p.

- FULLER, W.A. 1987. General linear models in complex surveys. *Notas de Aula, Cienes*.
- GRAYBILL, F.A. 1983. Matrices with applications in statistics. 2.ed. Belmont, Wadsworth. 461p.
- HARTLEY, H.O. & SIELKEN, R.L. Jr. 1975. A "super - population viewpoint" for finite population sampling. *Biometrics*, 31: 411-422.
- HUNG, H.M. 1990. Nonlinear regression analysis for complex surveys. *Communication in Statistics, Theory and Methods*, 19 (9):3447-3468.
- JAMES, B.H. 1981. Probabilidade: um curso em nível intermediário. Rio de Janeiro, IMPA. 292p.
- KISH, L. 1965. Survey sampling. New York, John Wiley. 643p.
- KISH, L. & FRANKEL, M.R. 1974. Inference from complex samples. *Journal of the Royal Statistical Society. Series B*, 36(1): 1-37.
- LEITE, J.G. & SINGER, J.M. 1990. Métodos assintóticos em estatística, fundamentos e aplicações. 9^o Simposio Brasileiro de Probabilidade e Estatística. 130p.
- LEJEUNE, M. & FAULKENBERRY, G.D. 1982. A simple predictive density function. *Journal of the American Statistical Association*, 77(379):654-659.
- LIMA, E.L. 1969. Elementos de topologia geral. Rio de Janeiro, IMPA. 192p.
- LIMA, E.L. 1970. Análise no espaço R^n . São Paulo, Edgard Blucher. 97p.

- PC CARP. 1987. Statistical laboratory, Iowa State University.
- PEREIRA, C. A. de B. & RODRIGUES, J. 1983. Robust linear predictions in finite populations. *International Statistical Review*, 51(3):293-300.
- RAJ, D. 1968. Sampling theory. New York, McGraw-Hill. 302p.
- RAO, C.R. 1973. Linear statistical inference and its applications. 2. ed. New York, John Wiley. 625p.
- RODRIGUES, J. 1989. Some results on restricted bayes least squares predictors for finite populations. *Sankhya. Series B*, 51:196-204.
- RODRIGUES, J. 1990. A simple likekihood theory for finite populations. A ser publicado nas Atas do 4º CLAPEM, México 1990.
- RODRIGUES, J.; BOLFARINE, H.; ROGATKO, A. 1985. A general theory of prediction in finite populations. *International Statistical Review*, 53(3):239-254.
- ROYALL, R.M. 1970. On finite population sampling theory under certain linear regression models. *Biometrika*, 57(2):377-387.
- ROYALL, R.M. 1976a. The linear least-squares prediction approach to two - stage sampling. *Journal of the American Statistical Association*, 71(355):657-664.
- ROYALL, R.M. 1976b. Likelihood functions in finite population sampling theory. *Biometrika*, 63(3):605-614.
- ROYALL, R.M. & HERSON, J. 1973a. Robust estimation in finite population I. *Journal of the American Statistical Association*, 68(344):880-889.

- ROYALL, R.M. & HERSON, J. 1973b. Robust estimation in finite populations II. Stratification on a size variable. *Journal of the American Statistical Association*, 68(344):890-893.
- SARNDAL, C.E. 1982. Implications of survey design for generalized regression estimation of linear functions. *Journal of Statistical Planning and Inference* 7:155-170.
- SARNDAL, C.E. & WRIGHT, R.L. 1984. Cosmetic form of estimators in survey sampling. *Scandinavian Journal of Statistics*, 11(3): 146-156.
- SCOTT, A.J.; BREWER, K.R.W.; HO, E.W.H. 1978. Finite population sampling and robust estimation. *Journal of the American Statistical Association*, 73(362):359-361.
- SCOTT, A. & WU, C. 1981. On the asymptotic distribution of ratio and regression estimators. *Journal of the American Statistical Association*, 76(373):98-102.
- SERFLING, R.J. 1980. Approximation theorems of mathematical statistics. New York, John Wiley. 371p.
- SHAH, B.V.; HOLT, M.M.; FOLSOM, R.E. 1977. Inference about regression models from sample survey data. *Bulletin of the International Statistical Institute*, 47(3):43-64.
- SMITH, T.M.F. 1976. The foundations of survey sampling: a review. *Journal of the Royal Statistical Society. Series A*, 146(4):394-403.
- SOC - Software Científico. 1989. Manual do Usuário.
- TALLIS, G.M. 1978. Note of robust estimation in finite populations. *Sankhya. Series C*, 40(2):136-138.

- TAM, S.M. 1986a. Optimal prediction in stochastic regression models with application to the analysis of repeated surveys. *Australian Journal of Statistics*, 28(3):345-353.
- TAM, S.M. 1986b. Characterization of the best model-based predictors in survey sampling. *Biometrika*, 73(1):232-235.
- TAM, S.M. 1987. Optimality of Royall's predictor under a Gaussian superpopulation model. *Biometrika*, 74(3):659-660.
- ZELLNER, A. 1991. Bayesian and Non-Bayesian estimation using balanced loss functions. *Graduate School of Business, University of Chicago*.