Resumo

Neste trabalho apresentamos como a inferência Bayesiana e a relação entre região H.P.D. e os testes de hipóteses, obtida através de algumas propriedades da distribuição t-Student multivariada, podem ser aplicados no estudo de dados de um Modelo de Covariância Linear com e sem erros nas variáveis. Para o nosso trabalho, consideramos um experimento planejado em k tratamentos, sendo cada um deles repetido em n_i unidades experimentais, $i = 1, 2, \ldots, k$.

Abstract

In this work we present as the Bayesian inference and the relation between region H.P.D. and the tests of hypothesis, gotten through some properties of the distribution t-Student multivaried, they can to be applied in the study of data of a Model of Linear Covariance with and without errors in the variables. For our work, we consider an experiment planned in k treatments, being each one of them repeated in n_i experimental units, $i = 1, 2, \ldots, k$.

Universidade Federal de Campina Grande
Centro de Ciências e Teconologia
Programa de Pós-Graduação em Matemática
Curso de Mestrado em Matemática

Sobre Modelo de Covariância: Uma Abordagem Bayesiana

por

Lya Raquel Oliveira de Sousa [1]

sob orientação do

Prof. Dr. Antonio José da Silva

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Campina Grande - PB

Mar/2006

[1]A autora foi bolsista da CAPES durante a elaboção desta Dissertação.

Sobre Modelo de Covariância: Uma Abordagem Bayesiana

por

Lya Raquel Oliveira de Sousa

Dissertação apresentada ao Corpo Docente do Programa de Pós-Graduação em Matemática - CCT - UFCG, como requisito parcial para obtenção do título de Mestre em Matemática.

Área de Concentração: Matemática

Linha de Pesquisa: Probabilidade e Estatística

Aprovada por:

Prof. Dr. André Gustavo Campos Pereira

Prof. Dr. Francisco Antonio Morais de Souza

Prof. Dr. Antonio José da Silva
Orientador

Universidade Federal de Campina Grande
Centro de Ciências e Tecnologia
Programa de Pós-Graduação em Matemática
Curso de Mestrado em Matemática

Mar/2006

Agradecimentos

Primeiramente, agradeço ao Ser que, mesmo com todas as minhas falhas e imperfeições, persiste, e como persiste, em me amar! "Porque mesmo que tua mãe esqueça de ti, Eu jamais te esquecerei." Obrigada, Senhor!

Agradeço à minha família, por acreditar em mim mais do que eu mesma! Especialmente, agradeço à minha mãe, por conseguir forças para superar vários obstáculos visando o bem estar dos seus filhos. Eu a amo, mamãe! Também agradeço a meu pai (*in memorium*) por ter sido um pai maravilhoso enquanto o Senhor permitiu que ele estivesse entre nós. Esteja em paz, papai. Aos meus irmãos, Ana e Rubens, e aos meus sobrinhos, por fazerem da minha vida uma alegria!

Aos meus antigos e eternos amigos "fisicamente distantes", por conseguirem se fazer presente de alguma forma durante esses 2 anos. Vocês são maravilhosos.

Aos meus novos amigos por me fazerem perceber que amigo é um pedaço de nós mesmos, por isso sempre estarão conosco! São eles: Rosângela (minha irmãzinha), Tatiana (a dona dos meus cílios), Marta ("tu é besta, viu!"), Cris (a madrinha mais linda do mundo), Lauriclécio (o comilão), Hallyson (meu amigo toddynho), Grayci (amiga de todas as disciplinas), Lino (meu quase irmão), Juliana (a mãe da Carol), além dos colegas de luta: Jaqueline, Diogo, Moisés, Aluísio, Orlando, Jesualdo, Areli.

Aos meus professores da UFPI, pela minha formação acadêmica, especialmente, aos professores do coração: Barbosa, Benício e Barnabé; e aos que me incentivaram a fazer o mestrado: Marcondes, Xavier, Gilvan, Mário Gomes, além dos já citados.

Aos professores do DME, que fazem deste departamento uma família. Em especial, ao prof. Bráulio, por fazer por nós sempre mais do que esperamos.

Ao prof. Antonio José, pela orientação e amizade.

Aos profs. Francisco Morais, que em muito contribuiu nas correções dessa dissertação, e André Gustavo, ambos, por me avaliarem.

Aos funcionários do DME e da Pós-Graduação em Matemática, em especial, à D. Salete, D. Argentina, Valdir e Marcelino.

Aos que contribuíram de alguma forma para a conclusão deste trabalho.

À CAPES, pelo apoio financeiro, sem ele eu não concluiria o curso.

Dedicatória

À D. Miriam. Mãe melhor do que essa só existe uma: a Mãe de Deus.

Conteúdo

Introdução

Ao longo desse trabalho estudaremos o Modelo de Covariância Linear, com e sem erros nas variáveis, sob uma perspectiva Bayesiana. Para tanto, usaremos uma informação a *priori* do tipo não informativa, proposta por Jeffreys, para que, de posse dessa informação adicional, tenhamos maior precisão nas inferências.

No **Capítulo 1** faremos uma introdução sobre a Inferência Bayesiana, seus princípios e a sua versão para estimação, intervalos de confiança e testes de hipóteses. Veremos também que Jeffreys, no intuito de realizar uma análise neutra, por causa da subjetividade da *priori*, propôs uma classe de *priori* não informativa, baseadas na informação de Fisher. Neste capítulo ainda faremos um breve estudo sobre a análise de covariância que é uma técnica que aumenta a precisão em experimentos aleatórios. Para isso será utilizado um termo chamado de soma de produto total para a análise dos dados.

No **Capítulo 2** estudaremos o artigo da referência [7]. Consideramos um experimento planejado com k tratamentos, onde cada um deles é repetido em n_i unidades experimentais, $i = 1, 2, \ldots, k$, sendo que neste capítulo as variáveis explicativas x_{ij} serão consideradas fixas e medidas sem erros, para $j = 1, 2, \ldots, n_i$, e $i = 1, 2, \ldots, k$. Um modelo estatístico adotado nesta situação é

$$y_{ij} = \tau_i + \beta_i x_{ij} + e_{ij},$$

onde y_{ij} é a observação relativa a variável resposta obtida na j-ésima unidade experimental que recebeu o tratamento i e e_{ij} é o erro correspondente na medição de y_{ij}. Vamos supor que $e_{ij} \sim N(0, \sigma^2)$ e são independentes. Após encontrarmos a função de verossimilhança associada ao modelo, usaremos uma densidade a *priori* proposta por Jeffreys, do tipo não informativa, para encontrarmos a densidadde a *posteriori* de $(\sigma^2, \boldsymbol{\theta})$, onde $\boldsymbol{\theta}$ é o vetor de parâmetros do modelo ou parte deles, assim como as densidades marginais a *posteriori* do vetor do efeito médio dos tratamentos $\boldsymbol{\tau}' = (\tau_1, \tau_2, \ldots, \tau_k)$ e do vetor dos coeficientes da regressão $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$, usando as propriedades da distribuição t-Student multivariada encontradas no **Apêndice B**. Usaremos a definição de região H.P.D., dada na seção 2.8, e sua relação com os testes de hipóteses, vistos na seção 2.8.1, para obtermos inferência sobre o efeito médio e sobre o coeficiente angular da regressão de cada tratamento, assim como inferências sobre combinações lineares destes.

No **Capítulo 3** consideraremos o mesmo modelo proposto no **Capítulo 2**, sendo que, neste caso, as variáveis explicativas x_{ij} são medidas com erros. Já que não existe instrumento de medição perfeito essa suposição está bem fundamentada. Dessa forma, ao invés de observarmos o verdadeiro valor de x_{ij} observamos a variável

$$X_{ij} = x_{ij} + u_{ij},$$

com $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, onde u_{ij} é o erro correspondente na medição de x_{ij}. Esse fato aumentará a quantidade de parâmetros envolvidos no modelo. Para diminuir esse número de parâmetros, iremos supor que a razão entre as variâncias $\lambda_e = \dfrac{\sigma_{ee}}{\sigma_{uu}}$ seja completamente conhecida. E já que o nosso objetivo, em geral, é obtermos inferências sobre os parâmetros de regressão, usaremos as suposições de independência dos erros e_{ij}, u_{ij} e da variável não observada x_{ij}, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, e também a suposição de não correlação entre cada um deles para esse fim. Após encontrarmos a função de verossimilhança do modelo, e usando as suposições acima, estaremos nas hipóteses do Teorema 3.1 e do Corolário 3.1 que nos fornecerão, respectivamente, a densidade a *posteriori* de $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x})$ e a densidade a *posteriori* de $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e)$. Ainda encontramos a densidade preditiva do vetor não observável $\mathbf{x}$.

No **Apêndice A** encontramos alguns resultados referentes a distribuição normal multivariada e alguns resultados de álgebra linear para a inversão e cálculo do determinante de matrizes em blocos usados nos **Capítulo 2** e **Capítulo 3**.

No **Apêndice B** encontramos algumas propriedades da distribuição t-Student multivariada usadas no **Capítulo 2**.

Capítulo 1

Preliminares

Neste capítulo faremos uma breve introdução à inferência Bayesiana, seus princípios e a sua versão para estimação, intervalos de confiança e testes de hipóteses. Faremos também um breve estudo sobre a análise de covariância. As informações deste capítulo são encontradas nas referências [2], [5] e [8].

1.1 Inferência Bayesiana

Sabemos que o objetivo da inferência matemática é reduzir a falta de conhecimento ou a incerteza sobre um determinado valor de interesse θ. Toda informação que temos a respeito de θ pode ser sumarizada, subjetivamente, em $P(\theta|H)$, a distribuição a *priori* de θ, onde H é a informação inicial que dispomos. A distribuição a *priori* é o único elemento novo na abordagem Bayesiana com relação à frequentista, e o seu compromisso é representar todo o conhecimento que se tem sobre θ, antes de se realizar o experimento, isto é, seu compromisso é descrever a incerteza sobre θ, probabilisticamente.

O objetivo da inferência Bayesiana é estudar as formas pelas as quais os resultados experimentais, adicionados a H, alteram $P(\theta|H)$. Por causa da subjetividade de H, diferentes pesquisadores podem possuir informações iniciais H_1 e H_2 fortes, e divergentes, sobre a quantidade desconhecida θ, de forma que é inviável conciliar essas informações. Precisamos, portanto, produzir uma análise neutra gerando-se um referencial. Usaremos a informação a *priori* do tipo não-informativa proposta por Jeffreys que será definida mais adiante. Assim, poderemos realizar a análise nos baseando no mínimo de informação subjetiva a *priori*, mesmo porque, esperamos que o experimento seja mais informativo do que a própria *priori*.

A seguir daremos a definição de parâmetro.

Definição 1.1.1 *Considere a observação X. A quantidade θ é chamada de parâmetro para X se*

$$P(X|H \text{ e } \theta) = P(X|\theta)$$

e nesse caso $P(X|\theta)$ é denominado de seu modelo estatístico.

Ou seja, θ é um parâmetro para X se, no momento em que θ fosse conhecido, sua incerteza sobre X se tornaria independente de qualquer conhecimento anterior H. Isto é,

$$P(X|H) = \sum_{\theta} P(X|\theta)P(\theta|H) = \sum_{\theta} L(\theta; x)P(\theta|H),$$

onde $L(\theta; x)$ é a função de verossimilhança para θ.

A classe de *prioris* não-informativas, propostas por Jeffreys, se baseia na medida de informação de Fisher. Primeiro, vejamos a definição da informação de Fisher e em seguida a definição da *priori* de Jeffreys.

Definição 1.1.2 *Considere a observação X com função de probabilidade $p(X|\theta)$. A medida de informação esperada de Fisher sobre θ através de X é definida por*

$$I(\theta) = E_{X|\theta}\Big[-\frac{\partial^2\, l(\theta; X)}{\partial\, \theta^2}\Big], \tag{1.1}$$

onde $l(\theta; x) = \log\, L(\theta; x)$, na base natural. Se $\boldsymbol{\theta}$ for um vetor, então define-se a matriz de informação esperada de Fisher sobre $\boldsymbol{\theta}$ através de X, $\mathbf{I}(\boldsymbol{\theta})$, com elementos $I_{ij}(\boldsymbol{\theta})$ dados por:

$$I_{ij}(\boldsymbol{\theta}) = E_{X|\theta}\Big[-\frac{\partial^2\, l(\theta; X)}{\partial\, \theta_i \theta_j}\Big]. \tag{1.2}$$

Mostra-se que, sob as condições de regularidade, podemos também calcular a medida de informação esperada de Fisher sobre θ através de X na forma abaixo.

$$I(\theta) = E_{X|\theta}\Big[-\frac{\partial\, l(\theta; X)}{\partial\, \theta}\Big]^2, \tag{1.3}$$

como já é conhecido.

Definição 1.1.3 *Considere a observação X com função de probabilidade $p(X|\theta)$. A priori não-informativa de Jeffreys tem densidade dada por*

$$p(\theta) \ \propto \ [I(\theta)]^{1/2}, \tag{1.4}$$

$\theta \in \Theta$, *onde Θ é o espaço paramétrico de θ. No caso multivariado temos*

$$p(\boldsymbol{\theta}) \ \propto \ |det\ \mathbf{I}(\boldsymbol{\theta})|^{1/2}. \tag{1.5}$$

Seja X uma quantidade aleatória que possui θ como parâmetro. Considere um experimento que observa o valor de X. Antes de observar X temos uma distribuição amostral de X dada por $P(X|\theta \text{ e } H)$. Após observar o valor de $X = x$ a nossa quantidade de informação aumentou: antes só conhecíamos H e depois do experimento passamos a conhecer $H_1 = H \bigcap \{X = x\}$. Agora a informação sobre θ está sumarizada em $P(\theta|x \text{ e } H)$. Podemos passar da distribuição a *priori*, $P(\theta|H)$, para a distribuição a *posteriori*, $P(\theta|x \text{ e } H)$, usando uma versão do Teorema de Bayes

$$P(\theta|x,\ H) = \frac{P(\theta|H)P(X = x|\theta)}{P(X = x|H)} = \frac{P(\theta|H)L(\theta; x)}{P(X = x|H)}.$$

Mas, $P(X = x|H)$ não depende de θ, daí, para a determinação da quantidade de interesse $P(\theta|x,\ H)$ ela é apenas uma constante, logo podemos escrever

$$P(\theta|x,\ H) \ \propto \ P(\theta|H)L(\theta; x). \tag{1.6}$$

Para recuperar a constante, basta observar que $p(\theta|x,\ H) = Wp(\theta|H)L(\theta; x)$, onde

$$W^{-1} = p(x|H) = \int_\theta p(x, \theta|H)d\theta = \int_\theta p(x|\theta)p(\theta)d\theta. \tag{1.7}$$

Note que na última igualdade a dependência em H é removida para facilitar a notação.

1.2 Estimação, Intervalos de Confiança e Testes de Hipóteses Bayesianos

Vejamos agora as versões Bayesianas para a estimação, intervalos de confiança e testes de hipóteses.

Vamos supor que a quantidade desconhecida, θ, seja um número real.

Definição 1.2.1 *O estimador de Bayes, para o parâmetro θ, é o valor esperado de θ dados H e x, isto é,*

$$E(\theta|H \text{ e } X = x) = \sum_{\theta} \theta P(\theta|H \text{ e } x),$$

se Θ for discreto.

Já que a distribuição a *posteriori* de θ é a expressão atualizada da sua incerteza, o valor esperado de θ dado H e x pode ser considerado como um estimador razoável de θ.

Os intervalos de credibilidade são os correspondentes Bayesianos aos intervalos de confiança na abordagem clássica. Segue sua definição.

Definição 1.2.2 *Intervalo ou conjunto de credibilidade de nivel α é um conjunto R_α contido no espaço paramétrico de θ tal que, para $\alpha \in (0,1)$ fixado,*

(i) $P(\theta \in R_\alpha|H \text{ e } x) = 1 - \alpha$ *ou* $P(\theta \in R_\alpha|H \text{ e } x) \geq 1 - \alpha$;

(ii) Para $R_\alpha = (\theta_0, \theta_1)$ *temos que* $p(\theta_0|H \text{ e } x) = p(\theta_1|H \text{ e } x)$.

Na seção (2.8) falamos mais detalhadamente sobre essa definição, onde chamamos R_α de região H.P.D., que é a região em que seus pontos tem densidade a *posteriori* alta. Para encontrarmos os pontos θ_0 e θ_1 das distribuições mais usuais e não simétricas usamos a tabela da referência [1].

Baseados nesse conjunto e fixando um nível α, podemos construir um método de decisão: não rejeitamos que θ_0 é um valor aceitável para θ se $\theta_0 \in R_\alpha$, caso contrário, se $\theta_0 \notin R_\alpha$, então rejeitamos que θ_0 seja um valor aceitável para θ. Isso corresponde, de certa forma, aos testes de hipóteses. Na seção (2.8.1) detalhamos como a região de credibilidade nos ajuda a tomar decisões sobre um determinado valor θ_0.

1.3 Análise de Covariância

A análise de covariância é uma técnica de bastante interesse. Ela combina os aspectos da análise de variância e da análise de regressão. A análise da covariância tem numerosas utilidades, a mais comum é que ela aumenta a precisão em experimentos aleatórios. Aqui supomos que cada unidade experimental também possui as medidas das outras variáveis que estão linearmente relacionada com a variável resposta.

O modelo de um delineamento completamente aleatorizado é

$$y_{ij} = \mu + \beta_j + \gamma(x_{ij} - \bar{x}_{..}) + e_{ij},$$

onde $e_{ij} \sim N(0, \sigma^2)$ e são independentes, $\sum_{i=1}^{k} n_i\beta_i = 0$, $j = 1, \ldots, n_i$, $i = 1, \ldots, k$, $\mu = \frac{1}{N}\sum_{i=1}^{k} n_i\mu_i$, para $N = \sum_{i=1}^{k} n_i$, $\beta_i = \mu_i - \mu$ e $\bar{x}_{..} = \frac{1}{N}\sum_{i=1}^{k}\sum_{j=1}^{n_i} x_{ij}$, com μ_i, β_i e γ assumindo valores reais, $\forall\ i = 1, \ldots, k$.

As suposições acima também podem ser escritas da seguinte forma.

$$\begin{aligned} y_{ij} &\sim N(\mu_{ij}, \sigma^2), \\ \mu_{ij} &= \mu + \beta_j + \gamma(x_{ij} - \bar{x}_{..}), \end{aligned} \tag{1.8}$$

com $\sum_{j=1}^{k} n_j\beta_j = 0$.

Os dados podem ser organizados na seguinte tabela.

Resumo dos Resultados para a Análise de Covariância

num Planejamento em Blocos Aleatórios

	Tratamentos								
	1		2		$\cdots$	k			
	x	y	x	y		x	y		
Observações	x_{11}	y_{11}	x_{12}	y_{12}		x_{1k}	y_{1k}		
	$\vdots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		
	$x_{n_1 1}$	$y_{n_1 1}$	$x_{n_2 2}$	$y_{n_2 2}$		$x_{n_k k}$	$y_{n_k k}$		
Totais	$T_{x.1}$	$T_{y.1}$	$T_{x.2}$	$T_{y.2}$		$T_{x.k}$	$T_{y.k}$	$T_{x..}$	$T_{y..}$
Médias	$\bar{x}_{.1}$	$\bar{y}_{.1}$	$\bar{x}_{.2}$	$\bar{y}_{.2}$		$\bar{x}_{.k}$	$\bar{y}_{.k}$	$\bar{x}_{..}$	$\bar{y}_{..}$

onde $T_{x.i} = \sum_{j=1}^{n_i} x_{ij}$, $T_{y.i} = \sum_{j=1}^{n_i} y_{ij}$, $\bar{x}_{.i} = \frac{T_{x.i}}{n_i}$ e $\bar{y}_{.i} = \frac{T_{y.i}}{n_i}$, $\forall i = 1, \ldots, k$.

Se a variável explicativa estiver estritamente relacionada com a variável resposta, é de se esperar que a reta de regressão desse modelo se ajuste melhor aos valores de y_{ij}, do que a reta do modelo original da análise da variância e os seus resíduos deverão ser menores, como pode ser visto na referência [5].

Notamos que as suposições (1.8), para um delineamento completamente aleatorizado, implica que os x'_{ij}s são constantes, desde que nada seja manifestado sobre eles terem uma distribuição.

O objetivo da análise de covariância é diminuir a variabilidade dos erros, aumentando assim nossas chances de rejeitar uma hipótese falsa. Para isso, usa-se um termo chamado de soma total dos produtos e denotaremos por SP_T. Esse novo termo é dado por

$$SP_T = \sum_{i=1}^{k}\sum_{j=1}^{n_j}(x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}).$$

Ao contrário da soma de quadrados, a soma de produtos não tem que ser necessariamente positiva.

Este novo termo será dividido em duas parcelas, exatamente da mesma maneira como a soma dos quadrados, como segue

$$\begin{aligned}
SP_T &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}[(\bar{x}_{.j} - \bar{x}_{..}) + (x_{ij} - \bar{x}_{.j})][(\bar{y}_{.j} - \bar{y}_{..}) + (y_{ij} - \bar{y}_{.j})] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_{.i} - \bar{x}_{..})(\bar{y}_{.i} - \bar{y}_{..}) + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij} - \bar{x}_{.i})(y_{ij} - \bar{y}_{.i}) \\
&= SP_R + SP_E,
\end{aligned}$$

onde, SP_R é a soma de produto dos tratamentos e SP_E é a soma de produto dos erros.

Já sabemos que a soma de quadrados total pode ser decomposta em

$$SQ_T = SQ_R + SQ_E,$$

onde SQ_R é a soma de quadrados dos tratamentos e SQ_E é a soma de quadrados dos erros. Vamos indexar em x e em y para facilitar a notação. Logo,

$$SQ_{T_x} = SQ_{Rx} + SQ_{Ex}$$

e

$$SQ_{T_y} = SQ_{Ry} + SQ_{Ey},$$

onde cada parcela pode ser escrita de forma a simplificar os cálculos, como segue

$$
\begin{aligned}
SQ_{T_x} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}^2 - \frac{1}{N}\Big(\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}\Big)^2\\
SQ_{Rx} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{x}_{.i}-\bar{x}_{..})^2 = \sum_{i=1}^{k}\frac{T_{x.i}^2}{n_i} - \frac{T_{x..}^2}{N}\\
SQ_{Ex} &= SQ_{T_x} - SQ_{Rx}\\
SQ_{T_y} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(y_{ij}-\bar{y}_{..})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}y_{ij}^2 - \frac{1}{N}\Big(\sum_{i=1}^{k}\sum_{j=1}^{n_i}y_{ij}\Big)^2\\
SQ_{Ry} &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{.i}-\bar{y}_{..})^2 = \sum_{i=1}^{k}\frac{T_{y.i}^2}{n_i} - \frac{T_{y..}^2}{N}\\
SQ_{Ey} &= SQ_{T_y} - SQ_{Ry},
\end{aligned}
$$

como já é conhecido.

Podemos obter formas mais simples de calcular cada soma de produto, da mesma forma como fazemos com a soma de quadrados:

$$
\begin{aligned}
SP_T &= \sum_{i=1}^{k}\sum_{j=1}^{n_i}(x_{ij}-\bar{x}_{..})(y_{ij}-\bar{y}_{..})\\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}y_{ij} - \bar{x}_{..}\sum_{i=1}^{k}\sum_{j=1}^{n_i}y_{ij} - \bar{y}_{..}\sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij} + N\bar{x}_{..}\bar{y}_{..}\\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}y_{ij} - \frac{1}{N}T_{x..}T_{y..} - \frac{1}{N}T_{y..}T_{x..} + N\frac{T_{x..}}{N}\frac{T_{y..}}{N}\\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}x_{ij}y_{ij} - \frac{1}{N}T_{x..}T_{y..}.
\end{aligned}
$$

Analogamente,

$$SP_R = \sum_{i=1}^{k}\frac{1}{n_i}T_{x.i}T_{y.i} - \frac{1}{N}T_{x..}T_{y..}.$$

E, por subtração, obtemos

$$SP_E = SP_T - SP_R.$$

Um dos objetivos da análise é testar $H_0 : \beta_i = 0,\ i = 1, 2, \ldots, k$ contra H_1 : existe ao menos um $\beta_i \neq 0$.

Uma outra formulação equivalente das hipóteses acima, em termos de $\mu_{.i} = \sum_{j=1}^{n_i}\mu_{ij}$, é $H_0 :\ (\mu_{.1} - \gamma\bar{x}_{.1}) = \cdots = (\mu_{.k} - \gamma\bar{x}_{.k})$ contra H_1 : nem todos $\mu_{.i} - \gamma\bar{x}_{.i}$ são iguais.

Para fazermos a análise do delineamento completamente aleatorizado precisaremos das funções abaixo:

$$\begin{aligned} SQ'_{T_y} &= SQ_{T_y} - \frac{(SP_T)^2}{SQ_{T_x}} \\ SQ'_{Ey} &= SQ_{Ey} - \frac{(SP_E)^2}{SQ_{Ex}} \\ SQ'_{Ry} &= SQ'_{T_y} - SQ'_{Ey}. \end{aligned}$$

Assumindo que as suposições (1.8) são verdadeiras, então podemos mostrar que

$$F = \frac{(SQ'_{Ry})/(k-1)}{(SQ'_{Ey})/(N-k-1)} = \frac{QM'_{Ry}}{QM'_{Ey}} \sim F_{k-1,N-k-1}, \tag{1.9}$$

onde $F_{k-1,N-k-1}$ é a distribuição F com $k-1$ graus de liberdade no numerador e $N-k-1$ graus de liberdade no denominador. A razão F é apropriada para testar H_0, que será rejeitada se F for grande, onde QM'_{Ry} e QM'_{Ey} são os respectivos quadrados médios.

Sumarizamos os resultados da análise na tabela abaixo.

Análise da Covariância para um Delineamento Aleatorizado

Fonte	SQ_x	SP	SQ_y	SQ'_y	g.l.	QM'_y	F
Tratamentos	SQ_Rx	SP_R	SQ_{Ry}	SQ'_{Ry}	$k-1$	QM'_{Ry}	$\frac{QM'_{Ry}}{QM'_{Ey}}$
Erro	SQ_{Ex}	SP_E	SQ_{Ey}	SQ'_{Ey}	$N-k-1$	QM'_{Ey}	
Total	SQ_{T_x}	SP_T	SQ_{T_y}	SQ'_{T_y}	$N-2$	-	-

Se H_0 não for rejeitada, então assumimos todos os β_j são iguais a zero. Logo,

$$\mu_{ij} = \mu + \gamma(x_{ij} - \bar{x}_{..}).$$

Existem outros métodos para a análise da covariância que não serão comentados neste trabalho, mas que são encontrados na referência [5].

Capítulo 2

Modelo de Covariância Linear Sem Erros nas Variáveis

Neste capítulo estudamos o artigo dado na referência [7]. Consideramos um experimento planejado com k tratamentos, sendo cada um deles repetido em n_i unidades experimentais, $i = 1, 2, \ldots, k$. Sob informação a *priori* do tipo não informativa e usando a definição de região H.P.D. e as propriedades da distribuição t-Student multivariada, discutidas no **Apêndice B**, faremos inferências sobre o efeito médio e sobre o coeficiente angular de cada tratamento, assim como também inferências sobre contrastes lineares destes.

2.1 O Modelo Estatístico

Um modelo estatístico adequado à este experimento é:

$$y_{ij} = \tau_i + \beta_i x_{ij} + e_{ij} \tag{2.1}$$

$j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k$, onde consideraremos:

(i) y_{ij} = observação relativa à variável resposta obtida na j-ésima unidade experimental que recebeu o tratamento $i, j = 1, 2, \ldots, n_i,\ i = 1, 2, \ldots, k$;

(ii) τ_i = efeito médio do tratamento $i, i = 1, 2, \ldots, k$;

(iii) β_i = coeficiente angular da regressão covariante dentro do tratamento i, $i = 1, 2, \ldots, k$;

(iv) x_{ij} = observação relativa à covariável obtida na j-ésima unidade experimental que recebeu o tratamento i, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, as quais, neste capítulo, serão consideradas fixas e medidas sem erros;

(v) e_{ij} = erro experimental associado à observação y_{ij}, $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, que serão considerados independentes uns dos outros e normalmente distribuídos com média zero e variância comum σ^2.

Matricialmente podemos escrever

$$\mathbf{y} = X\boldsymbol{\theta} + \boldsymbol{\varepsilon} \tag{2.2}$$

onde

- $\mathbf{y} = (y_{11}, \ldots, y_{1n_1}, y_{21}, \ldots, y_{2n_2}, \ldots, \ldots, y_{k1}, \ldots, y_{kn_k})'$ é o vetor de realizações de variáveis aleatórias de dimensão $n \times 1$, com $n = \sum_{i=1}^{k} n_i$;
- X = matriz de planejamento com dimensão $n \times 2k$;
- $\boldsymbol{\theta} = (\tau_1, \tau_2, \ldots, \tau_k, \beta_1, \beta_2, \ldots, \beta_k)'$ é o vetor dos parâmetros com dimensão $2k \times 1$;
- $\boldsymbol{\varepsilon} = (e_{11}, \ldots, e_{1n_1}, e_{21}, \ldots, e_{2n_2}, \ldots, \ldots, e_{k1}, \ldots, e_{kn_k})'$ é o vetor dos erros experimentais de dimensão $n \times 1$.

Desta forma, a equação (2.2) tem o seguinte aspecto:

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix} = \begin{bmatrix} 1 & 0 & \cdots & 0 & x_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & \cdots & 0 & x_{1n_1} & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 & 0 & x_{21} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & \cdots & 0 & 0 & x_{2n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & 1 & 0 & 0 & \cdots & x_{kn_k} \end{bmatrix} \begin{bmatrix} \tau_1 \\ \vdots \\ \tau_k \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \\ \vdots \\ e_{k1} \\ \vdots \\ e_{kn_k} \end{bmatrix}. \tag{2.3}$$

Portanto, segue de (v) que:

$$\mathbf{Y} \sim N_n(X\boldsymbol{\theta}, \sigma^2 I_n). \tag{2.4}$$

O modelo definido ou por (2.1), ou (2.2) ou (2.3) e com as hipóteses (i)-(v) é chamado de Modelo de Covariância Linear Normal.

Vamos considerar ainda que a matriz de planejamento, X, tem posto $2k$. Essa restrição não nos parece séria, pois se X não tem posto máximo, então existem redundâncias na especificação do modelo. Caso essas redundâncias existam, precisamos verificar se o modelo adotado é realmente o modelo adequado ao experimento.

Consideramos também que não existe $i = 1, 2, \ldots, k$, tal que $\sum_{j=1}^{n_i} x_{ij} = 0$, e que $k < (n/2) - 2$ para calcularmos o primeiro e o segundo momento de $\sigma^2|\mathbf{y}$.

2.2 A Função de Verossimilhança

A partir de (2.4) e usando o Teorema A.4 obtemos a função de verossimilhança de Y que é dada por

$$\begin{aligned} L(\boldsymbol{\theta}, \sigma^2; \mathbf{y}) &= (2\pi)^{-n/2}|\sigma^2 I_n|^{-1/2} \exp\left\{-\frac{1}{2}(\mathbf{y} - X\boldsymbol{\theta})'(\sigma^2 I_n)^{-1}(\mathbf{y} - X\boldsymbol{\theta})\right\} \\ &= (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}(\mathbf{y} - X\boldsymbol{\theta})'(\mathbf{y} - X\boldsymbol{\theta})\right\}, \end{aligned}$$

onde $\boldsymbol{\theta} = (\tau_1, \tau_2, \ldots, \tau_k, \beta_1, \beta_2, \ldots, \beta_k)'$ e o espaço paramétrico é

$$\Theta = \{(\boldsymbol{\theta}, \sigma^2); \sigma^2 > 0, -\infty < \tau_i < +\infty, -\infty < \beta_i < +\infty, i = 1, 2, \ldots, k\}.$$

Segue o lema que nos fornecerá o estimador de máxima verossimilhança para $\boldsymbol{\theta}$ e um estimador não viciado para σ^2.

Lema 2.1 *Seja a verossimilhança de* $\mathbf{Y}$ *como dada em (2.5). Então,*

(i) O estimador de máxima verossimilhança para $\boldsymbol{\theta}$ *é*

$$\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\mathbf{y};$$

(ii) Um estimador não viciado para σ^2 *é*

$$s^2 = \frac{n\hat{\sigma}^2}{\nu} = \frac{(\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}})}{\nu},$$

onde $\hat{\sigma}^2$ *é o estimador de máxima verossimilhança de* σ^2 *e* $\nu = n - 2k$.

Prova. Para mostrarmos (i) note que segue de (2.5) que a função de log-verossimilhança é

$$\log L(\boldsymbol{\theta}, \sigma^2; \mathbf{y}) = -\frac{n}{2}\log(2\pi) - \frac{n}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(\mathbf{y}'\mathbf{y} - 2\mathbf{y}'X\boldsymbol{\theta} + \boldsymbol{\theta}'X'X\boldsymbol{\theta}),$$

então:

$$\frac{\partial \log L(\boldsymbol{\theta}, \sigma^2; \mathbf{y})}{\partial \boldsymbol{\theta}}\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0 \quad \Longrightarrow \quad -\frac{1}{2\sigma^2}(-2(\mathbf{y}'X)' + 2X'X\boldsymbol{\theta})\bigg|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}} = 0$$

$$\Longrightarrow \quad \frac{1}{\sigma^2}(X'\mathbf{y} - X'X\hat{\boldsymbol{\theta}}) = 0 \quad \Longrightarrow \quad X'X\hat{\boldsymbol{\theta}} = X'\mathbf{y}.$$

Mas já supomos na seção anterior que posto$(X) = 2k$ e dessa forma, usando o fato de que posto$(X'X) =$ posto$(X) = 2k$, concluimos que $(X'X)_{2k\times 2k}$ tem posto completo e, portanto, sua inversa existe. Então, obtemos

$$\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'\mathbf{y}.$$

Para provarmos (ii) primeiro encontramos o estimador de máxima verossimilhança para σ^2 fazendo

$$\frac{\partial \log L(\hat{\boldsymbol{\theta}}, \sigma^2; \mathbf{y})}{\partial \sigma^2}\bigg|_{\sigma^2=\hat{\sigma}^2} = 0$$

logo,

$$\left[-\frac{n}{2\sigma^2} - \frac{1}{2}\left(-\frac{1}{\sigma^4}\right)(\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}})\right]\bigg|_{\sigma^2=\hat{\sigma}^2} = 0$$

$$\Longrightarrow \quad \frac{1}{2\hat{\sigma}^4}(\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) = \frac{n}{2\hat{\sigma}^2}$$

$$\Longrightarrow \quad \hat{\sigma}^2 = \frac{(\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}})}{n}.$$

Agora, observemos que podemos escrever

$$\begin{aligned} n\hat{\sigma}^2 &= (\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) \\ &= (\mathbf{y}' - \mathbf{y}'X(X'X)^{-1}X')(\mathbf{y} - X(X'X)^{-1}X'\mathbf{y}) \\ &= \mathbf{y}'(I_n - X(X'X)^{-1}X')(I_n - X(X'X)^{-1}X')\mathbf{y} \\ &= \mathbf{y}'(I_n - X(X'X)^{-1}X')\mathbf{y}, \end{aligned}$$

onde a última igualdade ocorre porque

$$\begin{aligned}(I_n - X(X'X)^{-1}X')^2 &= I_n - 2X(X'X)^{-1}X' + X(X'X)^{-1}X'X(X'X)^{-1}X' \\ &= I_n - 2X(X'X)^{-1}X' + X(X'X)^{-1}X' \\ &= I_n - X(X'X)^{-1}X'.\end{aligned}$$

Além disso,

$$(I_n - X(X'X)^{-1}X')' = I_n - X(X'X)^{-1}X'.$$

Ou seja, $(I_n - X(X'X)^{-1}X')$ é uma matriz idempotente e simétrica.

Definindo a matriz $A = (I_n - X(X'X)^{-1}X')\sigma^{-2}$, temos:

$$\begin{aligned} a)\quad A\,Cov(Y) &= (I_n - X(X'X)^{-1}X')\sigma^{-2}(\sigma^2 I_n) \\ &= I_n - X(X'X)^{-1}X',\end{aligned}$$

$$\begin{aligned} b)\quad \lambda &:= \tfrac{1}{2}[E(Y)]'A\,E(Y) \\ &= \tfrac{1}{2}(X\boldsymbol{\theta})'(I_n - X(X'X)^{-1}X')\sigma^{-2}(X\boldsymbol{\theta}) \\ &= \tfrac{1}{2\sigma^2}\,\boldsymbol{\theta}'X'(I_n - X(X'X)^{-1}X')X\boldsymbol{\theta} \\ &= \tfrac{1}{2\sigma^2}\boldsymbol{\theta}'(X'X - X'X(X'X)^{-1}X'X)\boldsymbol{\theta} \\ &= \tfrac{1}{2\sigma^2}\,\boldsymbol{\theta}'(X'X - X'X)\boldsymbol{\theta} \\ &= 0.\end{aligned}$$

Pelo item a, usando a recíproca do Teorema A.7 item i, já que $A\,Cov(Y)$ é uma matriz idempotente e simétrica, então determinamos a distribuição de $\dfrac{n\hat{\sigma}^2}{\sigma^2}$ que é uma qui-quadrado central, já que $\lambda = 0$ por b, e com o número dos graus de liberdade igual ao

$$\begin{aligned}\text{posto}(ACovY) &= \text{posto}(I_n - X(X'X)^{-1}X') \\ &= \text{tr}(I_n - X(X'X)^{-1}X') \\ &= \text{tr}(I_n) - \text{tr}(X(X'X)^{-1}X') \\ &= n - \text{tr}(X'X(X'X)^{-1}) \\ &= n - \text{tr}(I_{2k}) \\ &= n - 2k.\end{aligned}$$

Observe que na primeira igualdade usamos o fato de que o posto de uma matriz idempotente é igual ao seu traço e na quinta igualdade usamos a propriedade comutativa de traço de uma matriz. Em síntese,

$$\frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(\nu)}, \tag{2.5}$$

onde $\nu = n - 2k$.

Porém, note que quando calculamos o valor esperado do estimador $\hat{\sigma}^2$ encontramos um vício:

$$E(\hat{\sigma}^2) = E\Big(\frac{\sigma^2}{n}\,\frac{n}{\sigma^2}\,\hat{\sigma}^2\Big) = \frac{\sigma^2}{n}\,E\Big(\frac{n\hat{\sigma}^2}{\sigma^2}\Big) = \frac{\sigma^2}{n}(n-2k) < \sigma^2$$

levando-nos, assim, a subestimar σ^2.

Precisamos, portanto, encontrar um estimador para σ^2 que não seja viciado. Ora, esse problema é de fácil solução se fizermos a correção dos graus de liberdade do estimador $\hat{\sigma}^2$ definindo

$$s^2 = \frac{n\hat{\sigma}^2}{\nu}.$$

Dessa forma,

$$E(s^2) = \frac{\sigma^2}{\nu}\,E\Big(\frac{n\hat{\sigma}^2}{\sigma^2}\Big) = \sigma^2.$$

Portanto, para estimarmos a variância usaremos

$$s^2 = \frac{n\hat{\sigma}^2}{\nu} = \frac{(\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}})}{\nu},$$

que é um estimador não viciado, provando assim o **Lema 3.1**.

Nas próximas seções iremos encontrar as densidades marginais a *posteriori* de cada parâmetro, $\boldsymbol{\theta}$ e σ^2. Para tornar esse cálculo menos árduo, vamos agora reescrever a função de verossimilhança usando a seguinte decomposição:

$$\begin{aligned}
(\mathbf{y} - X\boldsymbol{\theta})'(\mathbf{y} - X\boldsymbol{\theta}) &= (\mathbf{y} - X\hat{\boldsymbol{\theta}} - X\boldsymbol{\theta} + X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}} - X\boldsymbol{\theta} + X\hat{\boldsymbol{\theta}})\\
&= ((\mathbf{y} - X\hat{\boldsymbol{\theta}}) - X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))'((\mathbf{y} - X\hat{\boldsymbol{\theta}}) - X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))\\
&= ((\mathbf{y} - X\hat{\boldsymbol{\theta}})' - (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X')((\mathbf{y} - X\hat{\boldsymbol{\theta}}) - X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}))\\
&= (\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) - 2(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\\
&= (\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) - 2(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'(X'\mathbf{y} - X'X(X'X)^{-1}X'\mathbf{y})\\
&\quad + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\\
&= (\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) - 2(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'(X'\mathbf{y} - X'\mathbf{y})\\
&\quad + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\\
&= (\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\\
&= \nu s^2 + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}).
\end{aligned}$$

Assim, reescrevemos a função de verossimilhança na forma abaixo.

$$L(\boldsymbol{\theta}, \sigma^2; \mathbf{y}) = (2\pi\sigma^2)^{-n/2} \exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' X' X (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})]\right\}. \tag{2.6}$$

Nas seções (2.8), (2.10)e (2.11) iremos obter intervalos de credibilidade para cada coordenada do vetor paramétrico, $\boldsymbol{\theta}$, faremos inferências para componentes do vetor do efeito médio dos tratamentos, $\boldsymbol{\tau} = (\tau_1, \tau_2, \ldots, \tau_k)'$, e do vetor dos coeficientes da regressão,
$\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$. Para que esses resultados sejam mais facilmente visualizados é interessante encontrarmos a expressão numérica de cada coordenada do estimador $\hat{\boldsymbol{\theta}} = (X'X)^{-1}X'Y$.

Ora,

$$\begin{aligned} X'X &= \begin{bmatrix} n_1 & 0 & \cdots & 0 & \sum_{j=1}^{n_1} x_{1j} & 0 & \cdots & 0 \\ 0 & n_2 & \cdots & 0 & 0 & \sum_{j=1}^{n_2} x_{2j} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & n_k & 0 & 0 & \cdots & \sum_{j=1}^{n_k} x_{kj} \\ \sum_{j=1}^{n_1} x_{1j} & 0 & \cdots & 0 & \sum_{j=1}^{n_1} x_{1j}^2 & 0 & \cdots & 0 \\ 0 & \sum_{j=1}^{n_2} x_{2j} & \cdots & 0 & 0 & \sum_{j=1}^{n_2} x_{2j}^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sum_{j=1}^{n_k} x_{kj} & 0 & 0 & \cdots & \sum_{j=1}^{n_k} x_{kj}^2 \end{bmatrix}_{2k \times 2k} \\ &= \begin{bmatrix} A_{k\times k} & B_{k\times k} \\ B'_{k\times k} & D_{k\times k} \end{bmatrix}_{2k\times 2k} \end{aligned}$$

onde A e D são matrizes simétricas e existe B^{-1}. Então, usando o resultado A.9 teremos:

$$(X'X)^{-1} = \begin{bmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{bmatrix}, \text{ com } \left\{ \begin{array}{l} E = D - B'A^{-1}B \\ F = A^{-1}B \end{array} \right. .$$

Após efetuarmos os cálculos, encontramos a matriz

$$(X'X)^{-1} = \begin{bmatrix} \dfrac{n_1R_{xx_1}+x_{1.}^2}{n_1^2R_{xx_1}} & \cdots & 0 & -\dfrac{x_{1.}}{n_1R_{xx_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \dfrac{n_kR_{xx_k}+x_{k.}^2}{n_k^2R_{xx_k}} & 0 & \cdots & -\dfrac{x_{k.}}{n_kR_{xx_k}} \\ -\dfrac{x_{1.}}{n_1R_{xx_1}} & \cdots & 0 & \dfrac{1}{R_{xx_1}} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & -\dfrac{x_{k.}}{n_kR_{xx_k}} & 0 & \cdots & \dfrac{1}{R_{xx_k}} \end{bmatrix}_{2k\times 2k}$$

onde $x_{i.} = \sum_{j=1}^{n_i} x_{ij}$, $R_{xx_i} = \sum_{j=1}^{n_i} x_{ij}^2 - \dfrac{x_{i.}^2}{n_i}$.

Fazendo $y_{i.} = \sum_{j=1}^{n_i} y_{ij}$, temos também que

$$X'\mathbf{y} = \begin{bmatrix} \sum_{j=1}^{n_1} y_{1j} \\ \vdots \\ \sum_{j=1}^{n_k} y_{kj} \\ \sum_{j=1}^{n_1} y_{1j}x_{1j} \\ \vdots \\ \sum_{j=1}^{n_k} y_{kj}x_{kj} \end{bmatrix} = \begin{bmatrix} y_{1.} \\ \vdots \\ y_{k.} \\ \sum_{j=1}^{n_1} y_{1j}x_{1j} \\ \vdots \\ \sum_{j=1}^{n_k} y_{kj}x_{kj} \end{bmatrix}_{2k\times 1}.$$

Consideremos agora $R_{xy_i} = \sum_{j=1}^{n_i} x_{ij}y_{ij} - \dfrac{x_{i.}y_{i.}}{n_i}$, $\bar{x}_{i.} = \dfrac{x_{i.}}{n_i}$ e $\bar{y}_{i.} = \dfrac{y_{i.}}{n_i}, \forall\, i = 1, \ldots, k$. Com essa notação, a expressão numérica de cada coordenada do estimador de $\boldsymbol{\theta}$ pode ser escrita por

$$\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\tau}_1 \\ \vdots \\ \hat{\tau}_k \\ \hat{\beta}_1 \\ \vdots \\ \hat{\beta}_k \end{bmatrix} = \begin{bmatrix} \dfrac{y_{1.}(n_1 R_{xx_1} + x_{1.}^2)}{n_1^2 R_{xx_1}} - \dfrac{x_{1.} \sum\limits_{j=1}^{n_1} x_{1j} y_{1j}}{n_1 R_{xx_1}} \\ \vdots \\ \dfrac{y_{k.}(n_k R_{xx_k} + x_{k.}^2)}{n_k^2 R_{xx_k}} - \dfrac{x_{k.} \sum\limits_{j=1}^{n_k} x_{kj} y_{kj}}{n_k R_{xx_k}} \\ \dfrac{1}{R_{xx_1}} \Big(\sum\limits_{j=1}^{n_1} x_{1j} y_{1j} - \dfrac{x_{1.} y_{1.}}{n_1} \Big) \\ \vdots \\ \dfrac{1}{R_{xx_k}} \Big(\sum\limits_{j=1}^{n_k} x_{kj} y_{kj} - \dfrac{x_{k.} y_{k.}}{n_k} \Big) \end{bmatrix} = \begin{bmatrix} \bar{y}_{1.} - \bar{x}_{1.} \dfrac{R_{xy_1}}{R_{xx_1}} \\ \vdots \\ \bar{y}_{k.} - \bar{x}_{k.} \dfrac{R_{xy_k}}{R_{xx_k}} \\ \dfrac{R_{xy_1}}{R_{xx_1}} \\ \vdots \\ \dfrac{R_{xy_k}}{R_{xx_k}} \end{bmatrix}_{2k \times 1} .$$

Agora, com esse resultado em mãos nós podemos escrever também a forma numérica de νs^2.

Notemos que

$$X\hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\tau}_1 + \hat{\beta}_1 x_{11} \\ \vdots \\ \hat{\tau}_1 + \hat{\beta}_1 x_{1n_1} \\ \hat{\tau}_2 + \hat{\beta}_2 x_{21} \\ \vdots \\ \hat{\tau}_2 + \hat{\beta}_2 x_{2n_2} \\ \vdots \\ \hat{\tau}_k + \hat{\beta}_k x_{k1} \\ \vdots \\ \hat{\tau}_k + \hat{\beta}_k x_{kn_k} \end{bmatrix}_{n \times 1}$$

Então, dessa maneira, a expressão numérica de νs^2 é

$$
\begin{aligned}
\nu s^2 &= (\mathbf{y} - X\hat{\boldsymbol{\theta}})'(\mathbf{y} - X\hat{\boldsymbol{\theta}}) \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i}\left[y_{ij} - (\hat{\tau}_i + \hat{\beta}_i x_{ij})\right]^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - 2\sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}(\hat{\tau}_i + \hat{\beta}_i x_{ij}) + \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\hat{\tau}_i + \hat{\beta}_i x_{ij})^2 \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - 2\sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i^2 R_{xx_i}\right] + \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i^2 R_{xx_i}\right] \\
&= \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}^2 - \sum_{i=1}^{k}\frac{y_{i.}^2}{n_i} - \sum_{i=1}^{k}\hat{\beta}_i^2 R_{xx_i}.
\end{aligned} \tag{2.7}
$$

A penúltima igualdade ocorre por causa das duas observações abaixo:

$$
\begin{aligned}
a)\quad \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}(\hat{\tau}_i + \hat{\beta}_i x_{ij}) &= \sum_{i=1}^{k}\sum_{j=1}^{n_i} y_{ij}(\bar{y}_{i.} - \hat{\beta}_i \bar{x}_{i.} + \hat{\beta}_i x_{ij}) \\
&= \sum_{i=1}^{k}\left[\sum_{j=1}^{n_i} y_{ij}\bar{y}_{i.} - \hat{\beta}_i \bar{x}_{i.}\sum_{j=1}^{n_i} y_{ij} + \hat{\beta}_i \sum_{j=1}^{n_i} x_{ij} y_{ij}\right] \\
&= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} - \hat{\beta}_i \frac{x_{i.} y_{i.}}{n_i} + \hat{\beta}_i\left(R_{xy_i} + \frac{y_{i.} x_{i.}}{n_i}\right)\right] \\
&= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i(\hat{\beta}_i R_{xx_i})\right] \\
&= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i^{\,2} R_{xx_i}\right]
\end{aligned}
$$

e

$$b)\quad \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\hat{\tau}_i+\hat{\beta}_i x_{ij})^2 = \sum_{i=1}^{k}\sum_{j=1}^{n_i}(\bar{y}_{i.}-\hat{\beta}_i\bar{x}_{i.}+\hat{\beta}_i x_{ij})^2$$

$$= \sum_{i=1}^{k}\left[\sum_{j=1}^{n_i}\frac{y_{i.}^2}{n_i} - 2\sum_{j=1}^{n_i}\bar{y}_{i.}(\hat{\beta}_i\bar{x}_{i.}-\hat{\beta}_i x_{ij}) + \sum_{j=1}^{n_i}(\hat{\beta}_i\bar{x}_{i.}-\hat{\beta}_i x_{ij})^2\right]$$

$$= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i^2} - 2\bar{y}_{i.}\hat{\beta}_i\left(n_i\ \frac{x_{i}.}{n_i} - x_{i.}\right) + \hat{\beta}_i^{\,2}\sum_{j=1}^{n_i}\left(\frac{x_{i.}^2}{n_i^2} - 2\ \frac{x_{i.}}{n_i}x_{ij} + x_{ij}^2\right)\right]$$

$$= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i^{\,2}\left(\frac{x_{i.}^2}{n_i} - 2\frac{x_{i.}^2}{n_i} + R_{xx_i} + \frac{x_{i.}^2}{n_i}\right)\right]$$

$$= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i^{\,2}\left(-\frac{x_{i.}^2}{n_i} + R_{xx_i} + \frac{x_{i.}^2}{n_i}\right)\right]$$

$$= \sum_{i=1}^{k}\left[\frac{y_{i.}^2}{n_i} + \hat{\beta}_i^{\,2}R_{xx_i}\right].$$

2.3 A Distribuição a *Priori* para $(\boldsymbol{\theta}, \sigma^2)$

Supondo que este é um experimento inicial e nada é conhecido em relação ao comportamento probabilístico dos parâmetros, vamos considerar uma *priori* do tipo não informativa para $(\boldsymbol{\theta}, \sigma^2)$, proposta por Jeffreys, associada ao procedimento proposto por Bernardo, encontrado na referência [2], para o caso multivariado.

Precisaremos, inicialmente, de algumas definições sobre modelo de locação, modelo de escala, e modelo de locação-escala.

Definição 2.3.1 *Uma variável aleatória Y tem modelo de locação se existem uma função f e uma quantidade θ tal que a distribuição de Y dado θ satisfaz a*

$$p(y|\theta) = f(y - \theta).$$

Neste caso θ é chamado de parâmetro de locação.

Como exemplo, note que se Y é uma variável aleatória normalmente distribuída com variância conhecida, então sua densidade é dada por

$$p(y|\theta) = (2\pi\sigma^2)^{-1/2} \exp\{-\frac{1}{2\sigma^2}(y-\theta)^2\},$$

que é função de $y - \theta$.

Note que se Y tem modelo de locação, então

$$\frac{\partial \log p(y|\theta)}{\partial \theta} = \frac{\partial \log f(y-\theta)}{\partial \theta} = -\frac{f'(y-\theta)}{f(y-\theta)}.$$

Portanto, por (1.3) temos

$$I(\theta) = E_{Y|\theta}\Big[\Big(-\frac{f'(Y-\theta)}{f(Y-\theta)}\Big)^2\Big] = E_U\Big[\Big(-\frac{f'(U)}{f(U)}\Big)^2\Big],$$

para $u = y - \theta$. Logo, $I(\theta) \propto$ constante, daí

$$p(\theta) \propto \text{constante}.$$

Para um vetor m-dimensional, o resultado é análogo, já que $I_{ij}(\boldsymbol{\theta}) \propto$ constante, $\forall\ i = 1, 2, \ldots, m,\ \forall\ j = 1, 2, \ldots, m$. Então,

$$|\det I(\boldsymbol{\theta})| \propto \text{constante} \implies p(\boldsymbol{\theta}) \propto \text{constante}. \tag{2.8}$$

Definição 2.3.2 *Uma variável aleatória Y tem modelo de escala se existem uma função f e uma quantidade $\sigma > 0$ tal que a distribuição de Y dado σ satisfaz*

$$p(y|\sigma) = \frac{1}{\sigma} f\Big(\frac{y}{\sigma}\Big).$$

Neste caso, σ é chamado de parâmetro de escala.

Como exemplo, note que se Y é uma variável aleatória com distribuição exponencial de parâmetro λ, então sua densidade é dada por

$$p(y|\lambda) = \frac{1}{\lambda} \exp\Big\{-\frac{y}{\lambda}\Big\},$$

portanto, λ é o parâmetro de escala.

Observe que se Y tem modelo de escala, então

$$\begin{aligned} \frac{\partial \log p(y|\sigma)}{\partial \sigma} &= \frac{\partial \log[\sigma^{-1} f(y/\sigma)]}{\partial \sigma} = \frac{\partial}{\partial \sigma}\Big[-\log \sigma + \log f(y/\sigma)\Big] \\ &= -\frac{1}{\sigma} + \frac{f'(y/\sigma)}{f(y/\sigma)}\Big(-\frac{y}{\sigma^2}\Big) = -\frac{1}{\sigma}\Big[1 + \frac{y}{\sigma}\frac{f'(y/\sigma)}{f(y/\sigma)}\Big]. \end{aligned}$$

Portanto, por (1.3) temos

$$I(\sigma) = E_{Y|\sigma}\Big[-\frac{1}{\sigma}\Big(1 + \frac{Y}{\sigma}\frac{f'(Y/\sigma)}{f(Y/\sigma)}\Big)\Big]^2 = \frac{1}{\sigma^2} E_U\Big[\Big(1 + U\frac{f'(U)}{f(U)}\Big)^2\Big],$$

para $u = y/\sigma$. Como a distribuição de U não depende de σ, então $I(\sigma) \propto \sigma^{-2}$, daí

$$p(\sigma) \propto \sigma^{-1}. \tag{2.9}$$

Definição 2.3.3 *Uma variável aleatória Y tem modelo de locação-escala se existem uma função f e uma quantidade θ e $\sigma > 0$ tais que a distribuição de Y dados θ e σ satisfaz a*

$$p(y|\theta, \sigma) = \frac{1}{\sigma} f\Big(\frac{y - \theta}{\sigma}\Big).$$

Neste caso θ é chamado de parâmetro de locação e σ é chamado de parâmetro de escala.

Como exemplo, note que se Y é uma variável aleatória normalmente distribuída com média θ desconhecida e desvio padrão σ desconhecido, então sua densidade é dada por

$$p(y|\theta, \sigma) = (2\pi)^{-1/2}\frac{1}{\sigma} \exp\Big\{-\frac{1}{2}\Big(\frac{y - \theta}{\sigma}\Big)^2\Big\},$$

e portanto, θ é o parâmetro de locação e σ é o parâmetro de escala.

Agora note que se Y tem modelo de locação-escala, podemos determinar a *priori* de referência para (θ, σ) segundo o procedimento proposto por Bernardo, conforme encontrado na referência [2]. Dessa forma, obtemos inicialmente a *priori* de referência para $p(\sigma|\theta)$. Ora, se θ é conhecido, então o modelo é de escala, logo, pelo que já vimos em (2.9) temos que $p(\sigma|\theta) \propto \sigma^{-1}$. Assim, podemos obter a distribuição de $y|\theta$, pois

$$p(y|\theta) = \int p(y, \sigma|\theta)d\sigma = \int p(y|\sigma, \theta)p(\sigma|\theta)d\sigma.$$

Agora podemos ver que como começamos com um modelo de locação-escala, então a dependência em y e θ através da função f continuará sendo na forma $f(y - \theta)$. Ou seja, o modelo é de locação e a *priori* de referência marginal para θ, como já vimos em (2.8), é da forma $p(\theta) \propto$ constante. Portanto, a *priori* de referência total e da forma

$$p(\theta, \sigma) = p(\theta)p(\sigma|\theta) \propto \frac{1}{\sigma}. \tag{2.10}$$

Note, portanto, que, no nosso caso, se queremos determinar uma priori do tipo não informativa para $(\boldsymbol{\theta}, \sigma^2)$, usamos o Teorema de Bayes, daí, escrevemos

$$p\,(\boldsymbol{\theta}, \sigma^2) = p\,(\boldsymbol{\theta})p(\sigma^2|\boldsymbol{\theta}),$$

e a suposição de que $(\boldsymbol{\theta}, \log \sigma^2)$ são aproximadamente independentes e localmente uniformes, para que uma "idéia" a *priori* de $\boldsymbol{\theta}$ não seja influenciada por uma "idéia" a *priori* sobre σ^2). Dessa forma, como $\boldsymbol{\theta}$ é o parâmetro de locação da normal multivariada, então só precisamos determinar $p(\sigma^2|\boldsymbol{\theta}) = p(\sigma^2) = [I(\sigma^2)]^{1/2}$. Logo,

$$\begin{aligned}
\frac{\partial \log L(\boldsymbol{\theta}, \sigma^2; \mathbf{y})}{\partial \sigma^2} &= -\frac{n}{2\sigma^2} + \frac{1}{2\sigma^4}(\mathbf{y} - X\boldsymbol{\theta})'(\mathbf{y} - X\boldsymbol{\theta}) \implies \\
\frac{\partial^2 \log L(\boldsymbol{\theta}, \sigma^2; \mathbf{y})}{\partial (\sigma^2)^2} &= \frac{n}{2\sigma^4} - \frac{1}{\sigma^6}(\mathbf{y} - X\boldsymbol{\theta})'(\mathbf{y} - X\boldsymbol{\theta}) \implies \\
E\Big(\frac{\partial^2 \log L(\boldsymbol{\theta}, \sigma^2; \mathbf{Y})}{\partial (\sigma^2)^2}\Big) &= E\Big[\frac{1}{\sigma^4}\Big(\frac{n}{2} - \frac{1}{\sigma^2}(\mathbf{Y} - X\boldsymbol{\theta})'(\mathbf{Y} - X\boldsymbol{\theta})\Big)\Big] \\
&= \frac{1}{\sigma^4}E\Big[\Big(\frac{n}{2} - \frac{1}{\sigma^2}\sum_{i=1}^{k}\sum_{j=1}^{n_i}[Y_{ij} - (\tau_i + \beta_i x_{ij})]^2\Big)\Big] \\
&= \frac{1}{\sigma^4}\Big(\frac{n}{2} - \frac{n\sigma^2}{\sigma^2}\Big) \propto \frac{1}{\sigma^4}.
\end{aligned}$$

Portanto, vamos considerar para o nosso experimento uma *priori* para $(\boldsymbol{\theta}, \sigma^2)$ do tipo não informativa na forma

$$p\,(\boldsymbol{\theta}, \sigma^2) = p\,(\boldsymbol{\theta})p(\sigma^2|\boldsymbol{\theta}) \;\propto\; \frac{1}{\sigma^2}. \tag{2.11}$$

2.4 A Distribuição a *Posteriori* Conjunta de $(\boldsymbol{\theta}, \sigma^2)$

O Teorema a seguir nos fornecerá a distribuição a *posteriori* conjunta de $(\boldsymbol{\theta}, \sigma^2)$.

Teorema 2.2 *Sob a informação a priori do tipo não informativa para $(\boldsymbol{\theta}, \sigma^2)$, como dada em (2.11), a posteriori para $(\boldsymbol{\theta}, \sigma^2)$ é dada por*

$$p\,(\boldsymbol{\theta}, \sigma^2|\mathbf{y}) = W(2\pi)^{-n/2}(\sigma^2)^{-(\frac{n}{2}+1)}\exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})]\right\}, \quad (2.12)$$

onde W é a constante normalizadora, a qual é dada por

$$W = \frac{|X'X|^{1/2}(2\pi)^{\nu/2}(\nu s^2)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2}},$$

com $|B|$ denotando o determinante da matriz B.

Prova. Para encontrarmos a distribuição a *posteriori* conjunta de $(\boldsymbol{\theta}, \sigma^2)$ precisamos apenas combinar a verossimilhança em (2.6) com a *priori* dada em (2.11) usando o Teorema de Bayes:

$$\begin{aligned} p\,(\boldsymbol{\theta}, \sigma^2|\mathbf{y}) &\propto p\,(\mathbf{y}|\boldsymbol{\theta}, \sigma^2)p\,(\boldsymbol{\theta}, \sigma^2) \\ &\propto (2\pi)^{-n/2}(\sigma^2)^{-(\frac{n}{2}+1)}\exp\left\{-\frac{1}{2\sigma^2}[\nu s^2 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})]\right\} \end{aligned}$$

com $\sigma^2 > 0, -\infty < \theta_i < +\infty,\ i = 1, \ldots, 2k.$

Para obtermos a igualdade precisamos encontrar a constante normalizadora W. Isso é feito efetuando o seguite cálculo:

$$\begin{aligned} W^{-1} &= \int_0^{+\infty}\int_{\boldsymbol{\Theta}}(2\pi)^{-n/2}(\sigma^2)^{-(\frac{n}{2}+1)}\exp\left\{-\frac{1}{2\sigma^2}\left[\nu s^2 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right]\right\}d\boldsymbol{\theta}\,d\sigma^2 \\ &= \int_0^{+\infty}\int_{\boldsymbol{\Theta}}\Big[(2\pi)^{\frac{-n+2k-2k}{2}}(\sigma^2)^{-(\frac{n-2k+2k}{2}+1)}\exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\}\times \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right\}d\boldsymbol{\theta}\Big]\,d\sigma^2 \\ &= \int_0^{+\infty}(2\pi)^{-\frac{\nu}{2}}(\sigma^2)^{-(\frac{\nu}{2}+1)}\exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\}\int_{\boldsymbol{\Theta}}\Big[(2\pi)^{-\frac{2k}{2}}(\sigma^2)^{-(\frac{2k}{2})}\times \\ &\quad \exp\left\{-\frac{1}{2\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right\}d\boldsymbol{\theta}\Big]d\sigma^2 \end{aligned}$$

$$
\begin{aligned}
&= \int_0^{+\infty} (2\pi)^{-\frac{\nu}{2}} (\sigma^2)^{-(\frac{\nu}{2}+1)} \exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\} |X'X|^{-1/2} \int_{\boldsymbol{\Theta}} \Big[(2\pi)^{-\frac{2k}{2}} \sigma^{-2k} |X'X|^{1/2} \times \\
&\quad \exp\left\{-\frac{1}{2\sigma^2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' X'X (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right\} d\boldsymbol{\theta} \Big] d\sigma^2 \\
&= \int_0^{+\infty} (2\pi)^{-\nu/2} \sigma^{-(\nu+2)} \exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\} |X'X|^{-1/2} \times \\
&\quad \left[\int_{\boldsymbol{\Theta}} \underbrace{(2\pi)^{-\frac{2k}{2}} |\sigma^{-2} X'X|^{1/2} \exp\left\{-\frac{1}{2\sigma^2} (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})' X'X (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})\right\}}_{\text{f.d.p. de um vetor aleatório } \boldsymbol{\theta}, \text{ onde } \boldsymbol{\theta} \sim N_{2k}(\hat{\boldsymbol{\theta}}, \sigma^{-2}(X'X))} d\boldsymbol{\theta} \right] d\sigma^2 \\
&= (2\pi)^{-\nu/2} |X'X|^{-1/2} \int_0^{+\infty} \sigma^{-(\nu+2)} \exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\} d\sigma^2 \\
&= (2\pi)^{-\nu/2} |X'X|^{-1/2} \int_{+\infty}^{0} \phi^{(\nu+2)/2} \exp\left\{-\Big(\frac{\nu s^2}{2}\Big)\phi\right\} \Big(-\frac{1}{\phi^2}\Big) d\phi, \text{ para } \phi = 1/\sigma^2 \\
&= (2\pi)^{-\nu/2} |X'X|^{-1/2} \int_0^{+\infty} \phi^{(\nu-2)/2} \exp\left\{-\Big(\frac{\nu s^2}{2}\Big)\phi\right\} d\phi \\
&= (2\pi)^{-\nu/2} |X'X|^{-1/2} \frac{\Gamma(\nu/2)}{(\nu s^2/2)^{\nu/2}} \int_0^{+\infty} \underbrace{\frac{(\nu s^2/2)^{\nu/2}}{\Gamma(\nu/2)} \phi^{(\nu/2)-1} \exp\left\{-\Big(\frac{\nu s^2}{2}\Big)\phi\right\}}_{\text{f.d.p. de uma v. a. } \phi, \text{ onde } \phi \sim G(\nu/2, \nu s^2/2)} d\phi \\
&= \frac{\Gamma(\nu/2) 2^{\nu/2}}{|X'X|^{1/2} (2\pi)^{\nu/2} (\nu s^2)^{\nu/2}}
\end{aligned}
$$

para $\boldsymbol{\theta}$ e $\sigma^2 > 0$.

Logo, $W = \dfrac{|X'X|^{1/2} (2\pi)^{\nu/2} (\nu s^2)^{\nu/2}}{\Gamma(\nu/2) 2^{\nu/2}}$.

2.5 A Distribuição a *Posteriori* Marginal de σ^2

O Teorema a seguir nos fornecerá a distribuição a *posteriori* marginal de σ^2.

Teorema 2.3 *Sob a informação a priori do tipo não informativa para* $(\boldsymbol{\theta}, \sigma^2)$*, como dada em (2.11), a posteriori marginal para* σ^2 *é dada por*

$$p\,(\sigma^2|\mathbf{y}) \;=\; \frac{(\nu s^2)^{-1}}{\Gamma(\nu/2)2^{\nu/2}}\Big(\frac{\sigma^2}{\nu s^2}\Big)^{-(\frac{\nu}{2}+1)}\exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big\}I_{(0,\infty)}(\sigma^2). \tag{2.13}$$

Prova. Vamos obter a *posteriori* marginal de σ^2 integrando a densidade conjunta de $(\boldsymbol{\theta}, \sigma^2)$, dada em (2.12), em relação a $\boldsymbol{\theta}$. Daí,

$$\begin{aligned}
p\,(\sigma^2|\mathbf{y}) &= \int_{\boldsymbol{\Theta}} p\,(\boldsymbol{\theta}, \sigma^2|\mathbf{y})\, d\boldsymbol{\theta} \\
&= \int_{\boldsymbol{\Theta}} \frac{|X'X|^{1/2}(2\pi)^{\nu/2}(\nu s^2)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2}}\,(2\pi)^{-n/2}(\sigma^2)^{-(\frac{n-2k+2k}{2}+1)} \times \\
&\quad \exp\Big\{-\frac{1}{2\sigma^2}\,[\nu s^2 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})]\Big\}\, d\boldsymbol{\theta} \\
&= \frac{(\nu s^2)^{\nu/2}(\sigma^2)^{-(\frac{n-2k}{2}+1)}}{\Gamma(\nu/2)2^{\nu/2}}\exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big\}\times \\
&\quad \int_{\boldsymbol{\Theta}} \underbrace{(2\pi)^{-2k/2}|\sigma^{-2}(X'X)|^{1/2}\exp\Big\{-\frac{1}{2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'(\sigma^{-2}X'X)(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big\}}_{\text{f.d.p. de um vetor aleatório } \boldsymbol{\theta},\ \text{onde } \boldsymbol{\theta}\sim N_{2k}(\hat{\boldsymbol{\theta}},\sigma^{-2}(X'X))}\, d\boldsymbol{\theta} \\
&= \frac{(\nu s^2)^{-1}}{\Gamma(\nu/2)2^{\nu/2}}\Big(\frac{\sigma^2}{\nu s^2}\Big)^{-(\frac{\nu}{2}+1)}\exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big\}I_{(0,\infty)}(\sigma^2).
\end{aligned}$$

Portanto, o resultado está demonstrado.

2.5.1 Propriedades da Distribuição a *Posteriori* Marginal de σ^2

O Lema a seguir apresenta as principais características numéricas da distribuição a *posteriori* marginal de σ^2 que são a média, a variância e a moda.

Lema 2.4 *Seja a distribuição a posteriori de* σ^2 *como dada em (2.13). Então,*

(i) $E(\sigma^2|\mathbf{y}) = \dfrac{\nu s^2}{\nu - 2}$;

(ii) $Var(\sigma^2|\mathbf{y}) = \dfrac{2(\nu s^2)^2}{(\nu-2)(\nu-4)}$;

(iii) A moda é o ponto σ_m^2 *tal que* $\sigma_m^2 = \dfrac{\nu s^2}{\nu+2}$.

Prova. Pela definição de esperança matemática temos

$$
\begin{aligned}
(i)\ E(\sigma^2|\mathbf{y}) &= \int_0^{+\infty} \sigma^2 p\,(\sigma^2|\mathbf{y})\, d\sigma^2 \\
&= \int_0^{+\infty} \sigma^2 \frac{(\nu s^2)^{-1}}{\Gamma(\nu/2)2^{\nu/2}}\Big(\frac{\sigma^2}{\nu s^2}\Big)^{-(\frac{\nu}{2}+1)} \exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big\}\, d\sigma^2 \\
&= \int_0^{+\infty} (\sigma^{-2})^{\nu/2}\Big(\frac{\nu s^2}{2}\Big)^{\nu/2} \frac{\exp\{-(\nu s^2)/(2\sigma^2)\}}{\Gamma(\nu/2)}\, d\sigma^2 \\
&= \int_{+\infty}^0 \phi^{\nu/2}\Big(\frac{\nu s^2}{2}\Big)^{\nu/2} \frac{\exp\Big\{-\phi(\frac{\nu s^2}{2})\Big\}}{\Gamma(\nu/2)}(-\phi^{-2})\, d\,\phi, \text{ para } \phi = 1/\sigma^2 \\
&= \Big(\frac{\nu s^2}{2}\Big)\frac{\Gamma(\frac{\nu}{2}-1)}{\Gamma(\frac{\nu}{2})} \int_0^{+\infty} \underbrace{\frac{(\nu s^2/2)^{\frac{\nu}{2}-1}}{\Gamma(\frac{\nu}{2}-1)}\phi^{(\frac{\nu}{2}-1)-1} \exp\Big\{-\phi\Big(\frac{\nu s^2}{2}\Big)\Big\}}_{\text{f.d.p. de uma v. a. } \phi, \text{ onde } \phi\sim G(\frac{\nu}{2}-1,\,\nu s^2/2)}\ d\phi \\
&= \frac{\nu s^2}{2}\,\frac{\Gamma(\frac{\nu}{2}-1)}{\Gamma(\frac{\nu}{2})} \\
&= \frac{\nu s^2}{2}\,\frac{\Gamma(\frac{\nu}{2}-1)}{(\frac{\nu}{2}-1)\Gamma(\frac{\nu}{2}-1)} \\
&= \frac{\nu s^2}{\nu-2},
\end{aligned}
$$

já que $\nu > 4$;

$$
\begin{aligned}
(ii)\ Var(\sigma^2|\mathbf{y}) &= E((\sigma^2)^2|\mathbf{y}) - [E(\sigma^2|\mathbf{y})]^2 \\
&= \int_0^{+\infty} (\sigma^2)^2 p\,(\sigma^2|\mathbf{y})\, d\sigma^2 - \Big(\frac{\nu s^2}{\nu-2}\Big)^2 \\
&= \int_0^{+\infty} (\sigma^2)^2 \frac{(\nu s^2)^{-1}}{\Gamma(\nu/2)2^{\nu/2}}\Big(\frac{\sigma^2}{\nu s^2}\Big)^{-(\frac{\nu}{2}+1)} \exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big\}\, d\sigma^2 - \Big(\frac{\nu s^2}{\nu-2}\Big)^2 \\
&= \frac{(\nu s^2/2)^{\nu/2}}{\Gamma(\nu/2)} \int_0^{+\infty} (\sigma^2)^{1-\nu/2} \exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big\}\, d\sigma^2 - \Big(\frac{\nu s^2}{\nu-2}\Big)^2 \\
&= \frac{(\nu s^2/2)^{\nu/2}}{\Gamma(\nu/2)} \int_{+\infty}^0 \phi^{\frac{\nu}{2}-1} \exp\Big\{-\phi\,\frac{\nu s^2}{2}\Big\}\Big(-\frac{1}{\phi^2}\Big)\, d\phi - \Big(\frac{\nu s^2}{\nu-2}\Big)^2, \text{ para } \phi = 1/\sigma^2
\end{aligned}
$$

$$= \frac{(\nu s^2/2)^2\Gamma(\frac{\nu}{2}-2)}{\Gamma(\nu/2)} \int_0^{+\infty} \underbrace{\frac{(\nu s^2/2)^{\frac{\nu}{2}-2}}{\Gamma(\frac{\nu}{2}-2)} \phi^{(\frac{\nu}{2}-2)-1} \exp\left\{-\phi\frac{\nu s^2}{2}\right\}}_{\text{f.d.p. da v. a.}\,\phi,\ \text{onde}\ \phi\sim G(\frac{\nu}{2}-2,\,\frac{\nu s^2}{2})} d\,\phi - \left(\frac{\nu s^2}{\nu-2}\right)^2$$

$$= \frac{(\nu s^2)^2\Gamma(\frac{\nu}{2}-2)}{\left(\frac{\nu}{2}-1\right)\left(\frac{\nu}{2}-2\right)\Gamma(\frac{\nu}{2}-2)2^2} - \left(\frac{\nu s^2}{\nu-2}\right)^2$$

$$= \frac{(\nu s^2)^2}{(\nu-2)(\nu-4)} - \frac{(\nu s^2)^2}{(\nu-2)^2}$$

$$= \frac{(\nu s^2)^2(\nu-2-\nu+4)}{(\nu-2)^2(\nu-4)}$$

$$= \frac{2(\nu s^2)^2}{(\nu-2)(\nu-4)},$$

pois $\nu > 4$;

(iii) A moda, o ponto σ^2_m tal que $p\,(\sigma^2|\mathbf{y}) \leq p\,(\sigma^2_m|\mathbf{y})$, $\forall\,\sigma^2 \in (0,+\infty)$, é o ponto que maximiza a densidade de $(\sigma^2|\mathbf{y})$, e assim, também maximiza o $\log p\,(\sigma^2|\mathbf{y})$. Mas,

$$p\,(\sigma^2|\mathbf{y}) = \frac{(\nu s^2)^{-1}}{\Gamma(\nu/2)2^{\nu/2}}\left(\frac{\sigma^2}{\nu s^2}\right)^{-(\frac{\nu}{2}+1)}\exp\left\{-\frac{\nu s^2}{2\sigma^2}\right\}I_{(0,\infty)}(\sigma^2) \Longrightarrow$$

$$\log\,p\,(\sigma^2|\mathbf{y}) = \frac{\nu}{2}\,\log(\nu s^2) - \frac{\nu s^2}{2\sigma^2} - (1+\nu/2)\,\log\sigma^2 - \log\Gamma(\nu/2) - \tfrac{\nu}{2}\,\log 2 \Longrightarrow$$

$$\frac{\partial \log\,p\,(\sigma^2|\mathbf{y})}{\partial\sigma^2}\bigg|_{\sigma^2=\sigma^2_m} = 0 \Longrightarrow \frac{\nu s^2}{2(\sigma^2_m)^2} - \frac{(1+\nu/2)}{\sigma^2_m} = 0 \Longrightarrow$$

$$\frac{\nu s^2}{2(\sigma^2_m)^2} = \frac{(1+\nu/2)}{\sigma^2_m} \qquad \Longrightarrow \sigma^2_m = \frac{\nu s^2}{\nu+2};$$

A mediana é o ponto Md tal que

$$\int_0^{Md} p\,(\sigma^2|\mathbf{y})\,d\sigma^2 = 0,5.$$

Não explicitaremos aqui o valor Md acima pela complexidade do cálculo desta integral.

2.6 A Distribuição a *Posteriori* Marginal de $\boldsymbol{\theta}$

O Teorema a seguir nos fornecerá a distribuição a *posteriori* marginal de $\boldsymbol{\theta}$, que terá a forma de uma densidade t-student multivariada. Aqui $t_{2k}\big(\hat{\boldsymbol{\theta}}, s^2(X'X)^{-1}, \nu\big)$ representa a distribuição t-Student $2k-$variada com parâmetro de locação $\hat{\boldsymbol{\theta}}$, matriz de dispersão $s^2(X'X)^{-1}$ e ν graus de liberdade.

Teorema 2.5 *Sob a informação a priori do tipo não informativa para* $(\boldsymbol{\theta}, \sigma^2)$*, como dada em (2.11), a posteriori marginal para* $\boldsymbol{\theta}$ *é dada por*

$$p\,(\boldsymbol{\theta}|\mathbf{y}) = \frac{\Gamma((\nu+2k)/2)|X'X|^{1/2}s^{-2k}}{(\Gamma(1/2))^{2k}\Gamma(\nu/2)\nu^k}\Big[1 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]^{-\frac{\nu+2k}{2}}, \tag{2.14}$$

ou seja,

$$\boldsymbol{\theta}|\mathbf{y} \sim t_{2k}\big(\hat{\boldsymbol{\theta}}, s^2(X'X)^{-1}, \nu\big).$$

Prova. Integrando a *posteriori* conjunta em (2.12) em relação a σ^2 vamos obter a densidade a *posteriori* marginal de $\boldsymbol{\theta}$. Daí,

$$\begin{aligned}
p\,(\boldsymbol{\theta}|\mathbf{y}) &= \int_0^{+\infty} p\,(\boldsymbol{\theta}, \sigma^2|\mathbf{y})\, d\sigma^2 \\
&= \int_0^{+\infty} \frac{|X'X|^{1/2}(2\pi)^{\nu/2}(\nu s^2)^{\nu/2}}{\Gamma(\nu/2)2^{\nu/2}(2\pi)^{n/2}(\sigma^2)^{(\frac{n}{2}+1)}} \exp\Big\{-\frac{1}{2\sigma^2}\,[\nu s^2 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})]\Big\}\, d\sigma^2 \\
&= \int_0^{+\infty} \frac{|X'X|^{1/2}2^{\nu/2}\pi^{\nu/2}(\nu s^2)^{\nu/2}\exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big[1 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]\Big\}}{\Gamma(\nu/2)2^{\nu/2}\sigma^{n-2k+2k+2}\,(2\pi)^{(n-2k+2k)/2}}\, d\sigma^2 \\
&= \int_0^{+\infty} \frac{|X'X|^{1/2}\pi^{\nu/2}(\nu s^2)^{\nu/2}\exp\Big\{-\frac{\nu s^2}{2\sigma^2}\Big[1 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]\Big\}}{\Gamma(\nu/2)\sigma^{\nu}(\sigma^2)^{k+1}(2\pi)^{\nu/2}(2\pi)^k}\, d\sigma^2 \\
&= \frac{|X'X|^{1/2}\pi^{\nu/2}\pi^{-\nu/2}}{\Gamma(\nu/2)2^{\nu/2}(2\pi)^k}\Bigg[\int_{+\infty}^{0} z^{\nu/2}\Big(\frac{\nu s^2}{z}\Big)^{-(k+1)} \times \\
&\quad \exp\Big\{-\frac{z}{2}\Big[1 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]\Big\}\Big(-\frac{\nu s^2}{z^2}\Big)\, dz\Bigg], \text{ para } z = \frac{\nu s^2}{\sigma^2} \\
&= \frac{|X'X|^{1/2}}{\Gamma(\nu/2)2^{\nu/2+k}(\pi)^k}\Bigg[\int_0^{+\infty} (\nu s^2)^{-k-1+1} z^{(\nu/2+k+1-2)} \times \\
&\quad \exp\Big\{-\frac{z}{2}\Big[1 + (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]\Big\}\, dz\Bigg]
\end{aligned}$$

$$= \frac{|X'X|^{1/2}(\nu s^2)^{-k}}{\Gamma(\nu/2)2^{n/2}(\pi)^k} \frac{\Gamma((\nu+2k)/2)}{\left[\frac{1}{2}\Big(1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big)\right]^{\frac{\nu+2k}{2}}} \times$$

$$\left[\int_0^{+\infty} \frac{\left[\frac{1}{2}\Big(1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big)\right]^{\frac{\nu+2k}{2}}}{\Gamma((\nu+2k)/2)} z^{\frac{\nu+2k}{2}-1} \times \right.$$

$$\left. \exp\left\{-\frac{z}{2}\left[1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right]\right\} dz\right]$$

$$= \frac{|X'X|^{1/2}2^{-n/2}\Gamma((\nu+2k)/2)2^{(\nu+2k)/2}\nu^{-k}s^{-2k}}{\Gamma(\nu/2)(\Gamma(1/2))^{2k}\Big[1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]^{\frac{\nu+2k}{2}}}$$

$$= \frac{\Gamma((\nu+2k)/2)|X'X|^{1/2}s^{-2k}}{(\Gamma(1/2))^{2k}\Gamma(\nu/2)\nu^k}\Big[1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big]^{-\frac{\nu+2k}{2}}.$$

Note que a penúltima igualdade ocorre porque o último integrando é a função densidade de probabilidade de uma variável aleatória Z, tal que

$$Z \sim G\Big(\frac{1}{2}\Big[1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\Big], \frac{\nu+2k}{2}\Big).$$

Assim sendo, mostramos que $\boldsymbol{\theta}|\mathbf{y} \sim t_{2k}\big(\hat{\boldsymbol{\theta}}, s^2(X'X)^{-1}, \nu\big)$. Essa distribuição tem importantes propriedades destacadas no **Apêndice B** que nos serão de grande valia para as inferências realizadas nas seções (2.10) e (2.11) sobre o vetor $\boldsymbol{\theta}$.

2.7 As Distribuições a *Posteriori* Marginais de $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$

A distribuição t-Student multivariada possui propriedades destacadas no **Apêndice B** que nos permitirão determinar as distribuições a *posteriori* marginais dos parâmetros $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$, assim como fazer inferências sobre eles. Suas distribuições são dadas no corolário abaixo do Teorema 2.5.

Corolário 2.1 *Seja $\boldsymbol{\theta}$ um vetor aleatório $2k$-dimensional distribuído a posteriori como em (2.14) e façamos a seguinte partição: $\boldsymbol{\theta}' = (\boldsymbol{\tau}'\,\boldsymbol{\beta}')$, onde $\boldsymbol{\tau}' = (\tau_1, \tau_2, \ldots, \tau_k)$ e $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$. Então $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$ tem distribuições a posteriori marginais na forma de uma t-Student multivariada como segue:*

$$\boldsymbol{\tau}|y \sim t_k\Big(\hat{\boldsymbol{\tau}}, s^2 D_{11}, \nu\Big) \tag{2.15}$$

e

$$\boldsymbol{\beta}|\mathbf{y} \sim t_k\Big(\hat{\boldsymbol{\beta}}, s^2 D_{22}, \nu\Big), \tag{2.16}$$

onde D_{11} e D_{22} são obtidas quando invertemos a matriz $(X'X)$ na seção (2.2).

Prova. Encontramos as densidades a *posteriori* marginais de $\boldsymbol{\tau}$ e de $\boldsymbol{\beta}$ considerando as partições

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\tau} \\ \boldsymbol{\beta} \end{bmatrix} \quad \text{e} \quad (X'X)^{-1} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

com

$$D_{11} = \text{diag}\Big\{\frac{n_1 R_{xx_1} + x_{1.}^2}{n_1^2 R_{xx_1}}, \ldots, \frac{n_k R_{xx_k} + x_{k.}^2}{n_k^2 R_{xx_k}}\Big\}$$

e

$$D_{22} = \text{diag}\Big\{\frac{1}{R_{xx_1}}, \ldots, \frac{1}{R_{xx_k}}\Big\},$$

obtidas quando invertemos a matriz $(X'X)$ na seção (2.2).

A propriedade **P.3** do **Apêndice B** nos garante que

$$\boldsymbol{\tau}|\mathbf{y} \sim t_k\Big(\hat{\boldsymbol{\tau}}, s^2 D_{11}, \nu\Big),$$

ou seja,

$$p\,(\boldsymbol{\tau}|\mathbf{y}) = \frac{\Gamma(\frac{\nu+k}{2})|D_{11}|^{-1/2}s^{-k}}{(\Gamma(1/2))^k\Gamma(\nu/2)\nu^{k/2}}\Big[1 + (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})'\frac{D_{11}^{-1}}{\nu s^2}(\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})\Big]^{-\frac{\nu+k}{2}}$$

e

$$\boldsymbol{\beta}|\mathbf{y} \sim t_k\Big(\hat{\boldsymbol{\beta}}, s^2 D_{22}, \nu\Big),$$

ou seja,

$$p\,(\boldsymbol{\beta}|\mathbf{y}) = \frac{\Gamma(\frac{\nu+k}{2})|D_{22}|^{-1/2}s^{-k}}{(\Gamma(1/2))^k\Gamma(\nu/2)\nu^{k/2}}\Big[1 + (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})'\frac{D_{22}^{-1}}{\nu s^2}(\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})\Big]^{-\frac{\nu+k}{2}}.$$

Também podemos encontrar a distribuição a *posteriori* de cada coordenada do vetor $\boldsymbol{\tau}$ e do vetor $\boldsymbol{\beta}$. O próximo Corolário nos dá esse resultado.

Corolário 2.2 *Para $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$ distribuídos a posteriori como em (2.15) e (2.16), respectivamente, então $\forall\ i = 1, 2, \ldots, k$, teremos*

$$\tau_i|\mathbf{y} \ \sim\ t_1\Big(\hat{\tau}_i, s^2\Big(\frac{1}{n_i} + \frac{\bar{x}_i^{\,2}}{R_{xx_i}}\Big), \nu\Big)$$

e

$$\beta_i|\mathbf{y} \ \sim\ t_1\Big(\hat{\beta}_i, \frac{s^2}{R_{xx_i}}, \nu\Big).$$

Prova. As distribuições marginais a *posteriori* de cada τ_i e de cada β_i, $i = 1, \ldots, k$ são facilmente determinadas se usarmos mais uma vez a propriedade **P.3** do **Apêndice B** para $\boldsymbol{\tau}$ e para $\boldsymbol{\beta}$, ou usando a propriedade **P.4** do mesmo Apêndice para a matriz $C_i = [0 \cdots 0\ 1\ 0 \cdots 0]_{1\times k}$, com 1 somente na i-ésima coluna, $i = 1, \ldots, k$. Portanto,

$$\tau_i|\mathbf{y} \ \sim\ t_1\Big[\hat{\tau}_i, s^2\Big(\frac{1}{n_i} + \frac{\bar{x}_i^{\,2}}{R_{xx_i}}\Big), \nu\Big]$$

e

$$\beta_i|\mathbf{y} \ \sim\ t_1\Big[\hat{\beta}_i, \frac{s^2}{R_{xx_i}}, \nu\Big].$$

2.8 Região H.P.D.

A sigla H.P.D. vem da abreviação da expressão *Highest Posterior Density*.

Depois que entendermos sua definição, perceberemos que cada um dos princípios abaixo são intuitivamente verificados:

(i) a região H.P.D. deve ser tal que a densidade de todo ponto dentro dela seja pelo menos tão grande quanto a de qualquer ponto fora dela;

(ii) e a região H.P.D. deve ser tal que para um dado conteúdo de probabilidade, este ocupa o menor volume possível no espaço paramétrico de $\boldsymbol{\theta}$.

Vejamos agora ssua definição.

Definição 2.8.1 *Sejam* $p\,(\boldsymbol{\theta}|\mathbf{y})$ *a função de densidade a posteriori de* $\boldsymbol{\theta}$ *e* $\alpha \in (0,1)$ *fixado. A região* $R_\alpha \subset \Theta$, *onde* Θ *é o espaço paramétrico de* $\boldsymbol{\theta}$, *é chamada de região H.P.D. de conteúdo* $(1-\alpha)$ *se*

(i) $P(\boldsymbol{\theta} \in R_\alpha|\mathbf{y}) = 1-\alpha$*;*

(ii) $\forall\, \boldsymbol{\theta}_1 \in R_\alpha$ *e* $\forall\, \boldsymbol{\theta}_2 \notin R_\alpha$ *temos* $p\,(\boldsymbol{\theta}_1|\mathbf{y}) \geq p\,(\boldsymbol{\theta}_2|\mathbf{y})$.

Da definição de região H.P.D. podemos verificar que as propriedades abaixo são satisfeitas.

a. A região H.P.D. inclui a moda, pois se $\boldsymbol{\theta}_0$ é a moda então $p\,(\boldsymbol{\theta}|\mathbf{y}) \leq p\,(\boldsymbol{\theta}_0|\mathbf{y})$, $\forall\, \boldsymbol{\theta} \in \Theta$;

b. Segue da definição que, para um dado conteúdo de probabilidade $(1-\alpha)$, a região H.P.D. tem o "menor volume" no espaço paramétrico de $\boldsymbol{\theta}$;

c. Se supusermos que $p\,(\boldsymbol{\theta}|\mathbf{y})$ é não uniforme em toda região do espaço paramétrico de $\boldsymbol{\theta}$, então a região H.P.D. de conteúdo $(1-\alpha)$ é única.

2.8.1 A Região H.P.D. e os Testes de Hipóteses

Nessa seção mostraremos como usar a definição de região H.P.D. e as propriedades da distribuição t-Student multivariada, vistas no **Apêndice B**, para decidir se um determinado ponto paramétrico $\boldsymbol{\theta}_0$ é ou não aceitável para $\boldsymbol{\theta}$.

Pelas propriedades da seção 2.8 vimos que, se R_α é uma região H.P.D. de conteúdo $1-\alpha$, então o evento $\boldsymbol{\theta} \in R_\alpha$ é equivalente ao evento $p\,(\boldsymbol{\theta}|\mathbf{y}) > c$, onde c é uma constante positiva, escolhida adequadamente. Segue-se daí que o ponto $\boldsymbol{\theta}_0 \in R_\alpha$ se, e somente se,

$$P\{p\,(\boldsymbol{\theta}|\mathbf{y}) > p\,(\boldsymbol{\theta}_0|\mathbf{y})\} \leq 1-\alpha.$$

Na expressão acima a função densidade de probabilidade $p\,(\boldsymbol{\theta}|\mathbf{y})$ é tratada como uma variável aleatória. Então, uma vez que pode ser determinada a distribuição a *posteriori* da quantidade $p\,(\boldsymbol{\theta}|\mathbf{y})$, ou alguma função monotônica dela (neste caso, usaremos a propriedade **P.1** do **Apêndice B**), podemos verificar se um determinado ponto paramétrico $\boldsymbol{\theta}_0$ pertence ou não pertence à região R_α.

Na densidade de $\boldsymbol{\theta}|\mathbf{y}$ encontrada em (2.14) já vimos que a relação para uma *priori* não informativa em $\boldsymbol{\theta}$ e σ^2 nos fornece uma distribuição a *posteriori* de $\boldsymbol{\theta}$ na forma de uma t-Student $2k$-variada. Além disso, a quantidade $\dfrac{Q(\boldsymbol{\theta})}{2ks^2}$ é distribuída, a *posteriori*, como uma $F_{(2k,\nu)}$, pela propriedade **P.2** do **Apêndice B**.

Supondo que estamos interessados em decidir se um determinado ponto paramétrico $\boldsymbol{\theta}_0 = (\theta_{10}, \ldots, \theta_{(2k)0})'$ está incluído ou não na região R_α, precisamos então calcular a probabilidade do evento $p\,(\boldsymbol{\theta}|\mathbf{y}) > p\,(\boldsymbol{\theta}_0|\mathbf{y})$.

Sabendo que $p\,(\boldsymbol{\theta}|\mathbf{y})$ é uma função monotonicamente decrescente da quantidade $\dfrac{Q(\boldsymbol{\theta})}{2ks^2}$, pela propriedade **P.1** do **Apêndice B**, e escolhendo $c = F_{(2k,\nu,\alpha)}$, então um particular ponto paramétrico $\boldsymbol{\theta}_0$ está incluído na região R_α se, e somente se,

$$\frac{Q(\boldsymbol{\theta}_0)}{2ks^2} < c = F_{(2k,\nu,\alpha)},$$

ou seja, $\boldsymbol{\theta}_0 \in R_\alpha$ se e somente se a desigualdade abaixo ocorre

$$(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}}) < 2ks^2F_{(2k,\nu,\alpha)}. \tag{2.17}$$

Ou equivalentemente,

$$P\Big\{F_{(2k,\nu)} > (\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})'\frac{X'X}{2ks^2}(\boldsymbol{\theta}_0 - \hat{\boldsymbol{\theta}})\Big\} \tag{2.18}$$

nos dá o conteúdo da região H.P.D. que inclui justamente o ponto $\boldsymbol{\theta}_0$.

Os resultados (2.17) e (2.18) acima proporcionam uma justificativa Bayesiana para a análise da variância, pois:

(i) A região H.P.D. de conteúdo $1-\alpha$ é numericamente identificada como a menor região de confiança $1-\alpha$;

(ii) A desigualdade (2.17) é apropriada para decidir se um dado ponto $\boldsymbol{\theta}_0$ está ou não incluído na região de confiança;

(iii) A propabilidade complementar de (2.18) nos dá o nível de significância associado à hipótese nula $H_0 : \boldsymbol{\theta} = \boldsymbol{\theta}_0$. Daí, rejeitamos H_0, ao nível de significância α, se $\boldsymbol{\theta}_0 \in R_\alpha^c$. Ou seja, se R_α é a região H.P.D. de conteúdo $1 - \alpha$, então rejeitamos H_0, ao nível de significância $100\alpha\%$ se $\boldsymbol{\theta}_0 \notin R_\alpha$.

2.9 Intervalo H.P.D. Para Parâmetros Individuais

Nesta seção encontramos intervalos de credibilidade a *posteriori* de conteúdo $(1-\alpha)$ para σ^2 e para cada coordenada do vetor $\boldsymbol{\theta}$.

2.9.1 Intervalo H.P.D. Para σ^2

Podemos encontrar intervalos de credibilidade a *posteriori* de conteúdo $(1-\alpha)$ para σ^2 definindo a variável aleatória $Z = \dfrac{\nu s^2}{\sigma^2}$, já que a distribuição de Z é dada por

$$Z = \frac{\nu s^2}{\sigma^2} = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_{(\nu)}$$

pelo resultado (2.5).

Portanto, o intervalo de credibilidade a *posteriori* de conteúdo $(1-\alpha)$ para σ^2 será contruido fazendo:

$$\begin{aligned} 1-\alpha &= P(\underline{\chi}^2 < \frac{\nu s^2}{\sigma^2} < \bar{\chi}^2) \\ &= P\Big(\frac{\nu s^2}{\bar{\chi}^2} < \sigma^2 < \frac{\nu s^2}{\underline{\chi}^2}\Big), \end{aligned}$$

para $0 < \underline{\chi}^2 < \bar{\chi}^2 < +\infty$, onde $\underline{\chi}^2$ e $\bar{\chi}^2$ são pontos da distribuição $\chi^2_{(\nu)}$ tais que suas densidades são iguais e a área α é dada por

$$\alpha = \int_0^{\underline{\chi}^2} f(z)dz + \int_{\bar{\chi}^2}^{+\infty} f(z)dz,$$

com $f(z)$ denotando a função densidade de probabilidade da variável aleatória Z. Esses pontos são obtidos através da tabelas da distribuição χ^2_ν encontadas na referência [1].

2.9.2 Intervalos H.P.D. Para Parâmetros Individuais de $\boldsymbol{\theta}$

Já sabemos que cada τ_i e cada β_i, $i = 1, \ldots, k$, é distribuído a *posteriori* segundo uma t-Student. Sendo a distribuição t-Student simétrica, os intervalos de credibilidade para τ_i e β_i coincidem com os intervalos de confiança usuais. Então intervalos H.P.D. para cada τ_i e para cada β_i, $i = 1, \ldots, k$, podem ser construídos, bastando para isso fazer uso das tabelas usuais da distribuição t-Student como segue.

$$
\begin{aligned}
1-\alpha &= P\left\{\left|\frac{\tau_i - \hat{\tau}_i}{\sqrt{s^2\left(\frac{1}{n_i} + \frac{\bar{x}_i{}^2}{R_{xx_i}}\right)}}\right| \leq t_{\frac{\alpha}{2}}\right\} \Longrightarrow \\
1-\alpha &= P\left\{-t_{\frac{\alpha}{2}} \leq \frac{\tau_i - \hat{\tau}_i}{\sqrt{s^2\left(\frac{1}{n_i} + \frac{\bar{x}_i{}^2}{R_{xx_i}}\right)}} \leq t_{\frac{\alpha}{2}}\right\} \Longrightarrow \\
1-\alpha &= P\left\{\hat{\tau}_i - t_{\frac{\alpha}{2}}\sqrt{s^2\left(\frac{1}{n_i} + \frac{\bar{x}_i{}^2}{R_{xx_i}}\right)} \leq\ \tau_i \leq \hat{\tau}_i + t_{\frac{\alpha}{2}}\sqrt{s^2\left(\frac{1}{n_i} + \frac{\bar{x}_i{}^2}{R_{xx_i}}\right)}\right\}.
\end{aligned}
$$

Analogamente,

$$
1-\alpha \;=\; P\left\{\hat{\beta}_i - t_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{R_{xx_i}}} \leq\ \beta_i \leq \hat{\beta}_i + t_{\frac{\alpha}{2}}\sqrt{\frac{s^2}{R_{xx_i}}}\right\},
$$

onde $t_{\frac{\alpha}{2}}$ é o ponto da distribuição t-Student tal que $P(t_\nu \leq t_{\frac{\alpha}{2}}) = 1 - \dfrac{\alpha}{2}$.

2.10 Inferências Conjuntas Sobre $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$

Nessa seção vamos usar toda a análise feita na seção anterior para decidir se determinados pontos paramétricos $\boldsymbol{\tau}_0$ e $\boldsymbol{\beta}_0$ são ou não aceitáveis para $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$, respectivamente.

Empregaremos a análise da seção anterior para testarmos as hipóteses:

- dos efeitos de tratamento serem todos iguais a um determinado valor;
- dos coeficientes da regressão serem todos iguais a um determinado valor.

Do resultado (2.15) e das propriedades da distribuição t-Student multivariada concluimos que:

(i) $\boldsymbol{\tau}|\mathbf{y} \sim t_k\Big(\hat{\boldsymbol{\tau}}, s^2 D_{11}, \nu\Big)$;

(ii) $\dfrac{Q(\boldsymbol{\tau})}{ks^2} \sim F_{(k,\nu)}$;

(iii) $p\,(\boldsymbol{\tau}|\mathbf{y})$ é monotonicamente decrescente da forma quadrática

$$
Q(\boldsymbol{\tau}) = (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}})' D_{11}^{-1} (\boldsymbol{\tau} - \hat{\boldsymbol{\tau}}).
$$

Portanto, testar a hipótese da não existência de efeito do tratamento, ou seja, testar $H_0 : \tau_1 = \cdots = \tau_k = 0$ contra H_1 : existe ao menos um $\tau_i \neq 0$, ao nível de significância α, é equivalente a verificar se o ponto $\boldsymbol{\tau}_0 = (0, \ldots, 0)'$ está incluído na região R_α. Isso ocorrerá se, e somente se,

$$\frac{Q(\boldsymbol{\tau}_0)}{ks^2} = (\boldsymbol{\tau}_0 - \hat{\boldsymbol{\tau}})' \frac{D_{11}^{-1}}{ks^2} (\boldsymbol{\tau}_0 - \hat{\boldsymbol{\tau}}) \leq F_{(k,\nu,\alpha)}.$$

Mas, para $(d_{ij})_{k\times k} = D_{11}^{-1}$, temos

$$\begin{aligned}
Q(\boldsymbol{\tau}_0) &= \hat{\boldsymbol{\tau}}' D_{11}^{-1} \hat{\boldsymbol{\tau}} \\
&= \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{\tau}_i \hat{\tau}_j d_{ij} \\
&= \sum_{i=1}^{k} \left[\left(\bar{y}_{i.} - \bar{x}_{i.} \frac{R_{xy_i}}{R_{xx_i}} \right)^2 \frac{n_i^2 R_{xx_i}}{n_i R_{xx_i} + x_{i.}^2} \right] \\
&= \sum_{i=1}^{k} \frac{(y_{i.} - \hat{\beta}_i x_{i.})^2 R_{xx_i}}{n_i R_{xx_i} + x_{i.}^2}.
\end{aligned}$$

Daí, $\boldsymbol{\tau}_0$ está na região R_α se, e somente se,

$$\left[\sum_{i=1}^{k} \frac{(y_{i.} - \hat{\beta}_i x_{i.})^2 R_{xx_i}}{n_i R_{xx_i} + x_{i.}^2} \right] (ks^2)^{-1} \leq F_{(k,\nu,\alpha)}.$$

Caso a desigualdade acima ocorra, isto é, caso o ponto $\boldsymbol{\tau}_0$ esteja na região R_α, então não rejeitamos a hipótese nula, ou seja, não existe efeito de tratamento. Assim, os dados devem ser analisados via análise de regressão linear passando pela origem. Para tanto, devemos adaptar a matriz de planejamento ao novo modelo:

$$y_{ij} = \beta_i X_{ij} + e_{ij} \quad \text{ou} \quad y_{ij} = \beta X_{ij} + e_{ij},$$

dependendo se rejeitamos ou não a hipótese de que todos os β_i são iguais a um determinado valor conhecido $\beta, i = 1, \ldots, k$. Em cada caso, respectivamente, a matriz de planejamento ficará na forma

$$X = \begin{bmatrix} x_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ x_{1n1} & 0 & \cdots & 0 \\ 0 & x_{21} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & x_{2n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{k1} \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & x_{kn_k} \end{bmatrix}_{n\times k} \quad \text{ou} \quad X = \begin{bmatrix} x_{11} \\ \vdots \\ x_{1n1} \\ x_{21} \\ \vdots \\ x_{2n_2} \\ \vdots \\ x_{k1} \\ \vdots \\ x_{kn_k} \end{bmatrix}_{n\times 1}$$

Se não houver efeito de tratamento, devemos agora fazer testes de hipóteses para o vetor $\boldsymbol{\beta}$.

Assim, do resultado (2.16) e das propriedades da distribuição t-Student multivariada concluimos que:

(i) $\boldsymbol{\beta}|\mathbf{y} \sim t_k\Big(\hat{\boldsymbol{\beta}}, s^2 D_{22}, \nu\Big)$;

(ii) $\dfrac{Q(\boldsymbol{\beta})}{ks^2} \sim F_{(k,\nu)}$;

(iii) $p\,(\boldsymbol{\beta}|\mathbf{y})$ é monotonicamente decrescente da forma quadrática

$$Q(\boldsymbol{\beta}) = (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}})' D_{22}^{-1} (\boldsymbol{\beta} - \hat{\boldsymbol{\beta}}).$$

Então, testar $H_0 : \beta_1 = \cdots = \beta_k = 0$ contra H_1 : existe ao menos um $\beta_i \neq 0$, ao nível de significância α, é equivalente a verificar se o ponto $\boldsymbol{\beta}_0 = (0, \ldots, 0)'$ pertence à região R_α. Isso ocorrerá se, e somente se,

$$\frac{Q(\boldsymbol{\beta_0})}{ks^2} = (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}})' \frac{D_{22}^{-1}}{ks^2} (\boldsymbol{\beta}_0 - \hat{\boldsymbol{\beta}}) \leq F_{(k,\nu,\alpha)}.$$

Mas, para $(d'_{ij})_{k\times k} = D_{22}^{-1}$, temos

$$Q(\boldsymbol{\beta}_0) \;\; = \;\; \hat{\boldsymbol{\beta}}' D_{22}^{-1} \hat{\boldsymbol{\beta}} = \sum_{i=1}^{k} \sum_{j=1}^{k} \hat{\beta}_i \hat{\beta}_j d'_{ij} = \sum_{i=1}^{k} \hat{\beta}_i^2 R_{xx_i}$$

Daí, $\boldsymbol{\beta}_0$ está na região R_α se, e somente se,

$$\left[\sum_{i=1}^{k} \hat{\beta}_i^2 R_{xx_i}\right](ks^2)^{-1} \leq F_{(k,\nu,\alpha)}.$$

Caso o ponto $\boldsymbol{\beta}_0$ pertença à região R_α então não rejeitamos a hipótese nula e os dados deverão ser analisados via ANOVA, sem trabalharmos com a informação a *priori* considerada na seção (2.3).

Caso contrário, se $\boldsymbol{\beta}_0$ não pertencer à região R_α deveremos verificar se os coeficientes de regressão são todos iguais a um determinado valor β conhecido. Isso será feito na seção (2.11), logo após definirmos contrastes lineares.

2.11 Inferência Conjunta e Marginal Sobre Combinações Lineares de $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$

É muito comum o investigador estar interessado e preocupado com a comparação dos valores dos parâmetros em vez dos seus valores absolutos. Nesta seção veremos como podemos efetuar tais comparações usando os resultados das seções anteriores.

2.11.1 Escolha de Contrastes Lineares

Definição 2.11.1 *Diremos que um conjunto de (2k-1) parâmetros* $\boldsymbol{\phi} = (\phi_1, \ldots, \phi_{2k-1})$ *mede o contraste entre os parâmetros* θ_i, $i = 1, \ldots, 2k$ *se eles são funções lineares da forma*

$$\phi_j = \sum_{i=1}^{2k} a_{ij}\theta_i = \mathbf{a}_j'\boldsymbol{\theta},$$

$j = 1, \ldots, 2k-1$, *tais que* $(\phi_1 = 0, \ldots, \phi_{2k-1} = 0)$ *implica necessariamente que* $(\theta_1 = \theta, \ldots, \theta_{2k} = \theta)$ *para algum* θ *conhecido.*

A seguir enunciamos o Lema que facilitará a escolha dos contrastes lineares.

Lema 2.1 *Para que as propriedades acima sejam válidas é necessário e suficiente que os itens a e b abaixo sejam satisfeitos:*

a) os vetores $\mathbf{a}_1, \ldots, \mathbf{a}_{2k-1}$ *sejam linearmente independentes, e*

b) $\mathbf{a}_j'\mathbf{1}_{2k} = 0$, *isto é,* $\sum_{i=1}^{2k} a_{ij} = 0$, $j = 1, \ldots, 2k-1$.

Prova. Verificamos o resultado escrevendo $A\boldsymbol{\theta} = \boldsymbol{\phi}$, onde $A = [\mathbf{a}_1, \ldots, \mathbf{a}_{2k-1}]'$ é uma matriz $(2k-1) \times 2k$. Para $\boldsymbol{\phi} = (0, \ldots, 0)'$ temos que $A\boldsymbol{\theta} = \mathbf{0}$ define um sistema de $(2k-1)$ equações em $2k$ "variáveis" desconhecidas.

Para mostrar a suficiência, suponha que as condições a, b são satisfeitas e que o posto $A = 2k-1$, então, notando que

$$A\mathbf{1}_{2k} = \begin{bmatrix} \mathbf{a}_1' \\ \vdots \\ \mathbf{a}_{2k-1}' \end{bmatrix} \mathbf{1}_{2k} = \Big(\sum_{i=1}^{2k} a_{i1}, \cdots, \sum_{i=1}^{2k} a_{i(2k-1)}\Big)' = (0, \ldots, 0)',$$

segue-se que todas as soluções da equação $A\boldsymbol{\theta} = \mathbf{0}$ são da forma $\boldsymbol{\theta} = \theta\mathbf{1}_{2k}$, isto é, $\theta_1 = \ldots = \theta_{2k} = \theta$.

Supondo que existe uma outra solução para o sistema $A\boldsymbol{\theta} = 0$ diferente de $\boldsymbol{\theta} = \theta\mathbf{1}_{2k}$, ou seja, supondo que existe pelo menos um $i,\ i = 1, \ldots, 2k$, tal que $\theta_i \neq \theta$, então,

$$0 = A\boldsymbol{\theta} \ = \begin{bmatrix} \sum_{i=1}^{2k} a_{i1}\theta_i \\ \sum_{i=1}^{2k} a_{i2}\theta_i \\ \vdots \\ \sum_{i=1}^{2k} a_{i(2k-1)}\theta_i \end{bmatrix} = \theta_1 \begin{bmatrix} a_{11} \\ a_{12} \\ \vdots \\ a_{1(2k-1)} \end{bmatrix} + \cdots + \theta_{2k} \begin{bmatrix} a_{k1} \\ a_{k2} \\ \vdots \\ a_{k(2k-1)} \end{bmatrix}.$$

Como os a_j' são linearmente independentes, então todos os θ_i são iguais a zero. Logo, não existe outra solução para o sistema $A\boldsymbol{\theta} = 0$ além de $\boldsymbol{\theta} = \theta\mathbf{1}_{2k}$.

Agora para mostrar a necessidade, vamos supor que $\boldsymbol{\theta} = \theta\mathbf{1}_{2k}$ e daí,

$$\phi_j = \sum_{i=1}^{2k} a_{ij}\theta_i = \theta \sum_{i=1}^{2k} a_{ij},$$

$j = 1, \ldots, 2k-1$. Ou seja, $A\boldsymbol{\theta} = (\phi_1, \ldots, \phi_{2k-1})'$ é igual ao vetor nulo, quando $(\theta_1, \ldots, \theta_{2k})' = (\theta, \ldots, \theta)'$, isto é, $[\mathbf{a}_1'\boldsymbol{\theta}, \ldots, \mathbf{a}_{2k-1}'\boldsymbol{\theta}] = A\boldsymbol{\theta} = \mathbf{0}$. E assim,

$$\sum_{i=1}^{2k} a_{ij} = 0,$$

$j = 1, \ldots, 2k-1$.

Para ver que os $\mathbf{a}_j$ são linearmente independentes, suponha, por absurdo, que eles não sejam, logo existiria uma outra escolha de $\boldsymbol{\theta}$, digamos,

$$\boldsymbol{\theta} = (\theta^*, \ldots, \theta^*)' = \theta^*\mathbf{1}_{2k},$$

com $\theta^* \neq \theta$, que satisfaz $A\boldsymbol{\theta} = \mathbf{0}$. Mas, isso nos levaria a afirmar que $\theta\mathbf{1}_{2k} = \boldsymbol{\phi} = \theta^*\mathbf{1}_k$, ou seja, $\theta^* = \theta$, que é um absurdo. Logo, os $\mathbf{a}_j$ são linearmente independentes.

Dessa forma, se temos $2k$ parâmetros $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_{2k})$ podemos definir $2k-1$ comparações não redundantes como $2k-1$ funções independentes

$$\phi_j = f_j(\theta_i),$$

$j = 1, \ldots, 2k-1,\ i = 1, \ldots, 2k$, que são todas iguais a zero se, e somente se, $\theta_1 = \cdots = \theta_{2k}$, deixando claro a grande extensão de classes de escolha para tais funções.

De forma análoga, se temos k parâmetros $\boldsymbol{\tau} = (\tau_1, \ldots, \tau_k)'$ ou $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_k)'$ podemos definir $k-1$ comparações não redundantes como $k-1$ funções independentes $\phi_j = f_j(\tau_i)$, $i = 1, \ldots, k$, ou $\phi_j = f_j(\beta_i)$, $i = 1, \ldots, k$, que são todas iguais a zero se, e somente se, $\tau_1 = \ldots = \tau_k$ ou $\beta_1 = \ldots = \beta_k$.

Consideremos agora $\boldsymbol{\psi}|\mathbf{y} = C\boldsymbol{\tau}|\mathbf{y}$ para estudo do efeito comparativo dos k parâmetros, onde C é um conjunto conveniente de $k-1$ coeficientes de contrastes não redundantes de $\boldsymbol{\tau}$. Se os níveis $1, 2, \ldots, n_i$ do fator tratamento i, $i = 1, \ldots, k$, forem quantitativos e se houver interesse em verificar o comportamento da variável resposta em relação aos níveis do fator, então a matriz C será um conjunto conveniente de coeficientes de polinômios ortogonais, para que $\sum_{i=1}^{k} \phi_j(x_i)\phi_k(x_i) = 0$, evitando dificuldades na regressão.

Seja, portanto, $C_{(k-1)\times k} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix}$, logo

$$C\boldsymbol{\tau} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} \tau_1 \\ \tau_2 \\ \vdots \\ \tau_k \end{bmatrix} = \begin{bmatrix} \tau_1 - \tau_2 \\ \tau_2 - \tau_3 \\ \vdots \\ \tau_{k-1} - \tau_k \end{bmatrix}.$$

Assim, testar a hipótese da homogeneidade dos efeitos de tratamento, isto é, testar $H_0 : \tau_1 = \ldots = \tau_k = \tau$, com τ conhecido, contra H_1 : existe pelo menos um $\tau_i \neq \tau$ é equivalente a testar $H_0 : \boldsymbol{\psi}_0 = (0, \ldots, 0)'$ contra H_1 : existe pelo menos um $\psi_j \neq 0$, $j = 1, \ldots, k-1$. Ou seja, precisamos verificar se o ponto $\boldsymbol{\psi}_0 = (0, \ldots, 0)'$ está na região R_α.

Na seção (2.7) mostramos que $\boldsymbol{\tau}|\mathbf{y} \sim t_k\Big(\hat{\boldsymbol{\tau}}, s^2 D_{11}, \nu\Big)$ e fazendo uso dos resultados da distribuição t-Student multivariada vistas no **Apêndice B** temos que

(i) $\boldsymbol{\psi}|\mathbf{y} \sim t_{k-1}\Big(C\hat{\boldsymbol{\theta}}, s^2 C D_{11} C', \nu\Big)$;

(ii) $\dfrac{Q(\boldsymbol{\psi})}{(k-1)s^2} \sim F_{(k-1,\nu)}$;

(iii) $p\,(\boldsymbol{\psi}|\mathbf{y})$ é monotonicamente decrescente da forma quadrática

$$Q(\boldsymbol{\psi}) = (\boldsymbol{\psi} - \hat{\boldsymbol{\psi}})'(CD_{11}C')^{-1}(\boldsymbol{\psi} - \hat{\boldsymbol{\psi}}).$$

Daí, o ponto $\boldsymbol{\psi}_0 = (0, \ldots, 0)'$ pertence à região R_α se, e somente se,

$$\frac{Q(\boldsymbol{\psi}_0)}{(k-1)s^2} = (\boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}})'\frac{(CD_{11}C')^{-1}}{(k-1)s^2}(\boldsymbol{\psi}_0 - \hat{\boldsymbol{\psi}}) \;\leq\; F_{(k-1,\nu,\alpha)}.$$

Caso a desigualdade acima ocorra, então não rejeitamos a hipótese da homogeneidade dos efeitos de tratamento, e assim deveremos analisar os dados via regressão com intercepto $\tau\mathbf{1}_k$ e o modelo fica dado por

$$y_{ij} = \tau + \beta_i X_{ij} + e_{ij},$$

ou visto na forma matricial

$$\begin{bmatrix} y_{11} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{k1} \\ \vdots \\ y_{kn_k} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & 0 & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & x_{1n1} & 0 & \cdots & 0 \\ 1 & 0 & x_{21} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & x_{2n_2} & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & x_{k1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & 0 & 0 & \cdots & x_{kn_k} \end{bmatrix} \begin{bmatrix} \tau \\ \beta_1 \\ \vdots \\ \beta_k \end{bmatrix} + \begin{bmatrix} e_{11} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ \vdots \\ e_{2n_2} \\ \vdots \\ e_{k1} \\ \vdots \\ e_{kn_k} \end{bmatrix}.$$

Caso contrário, consideramos o modelo de covariância com um único coeficiente angular ou com os coeficientes distintos, como já foi comentado na seção 2.11.

Consideremos agora $\boldsymbol{\psi}|\mathbf{y} = C\boldsymbol{\beta}|\mathbf{y}$ para estudo do efeito comparativo dos k coeficientes de regressão, onde C é como dada anteriomente. A partir daí, verificamos se os coeficientes da regressão são todos iguais a um determinado valor β, conhecido.

Temos que,

$$C\boldsymbol{\beta} = \begin{bmatrix} 1 & -1 & 0 & \cdots & 0 & 0 \\ 0 & 1 & -1 & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix} = \begin{bmatrix} \beta_1 - \beta_2 \\ \beta_2 - \beta_3 \\ \vdots \\ \beta_{k-1} - \beta_k \end{bmatrix}.$$

Assim, testar a hipótese nula $H_0 : \beta_1 = \cdots = \beta_k = \beta$, é equivalente a testar a hipótese $H_0 : \phi_1 = \cdots = \phi_{k-1} = 0$. Logo, verificar se $\beta_1 = \cdots = \beta_k = \beta$ é equivalente a verificar se o ponto $\boldsymbol{\phi}_0 = (0, \ldots, 0)'$ está incluido na região R_α. Usaremos mais uma vez os resultados já conhecidos da distribuição t-Student multivariada:

(i) $\boldsymbol{\phi}|\mathbf{y} \sim t_{k-1}\Big(C\hat{\boldsymbol{\beta}}, s^2 C D_{22} C', \nu\Big)$;

(ii) $\dfrac{Q(\boldsymbol{\phi})}{(k-1)s^2} \sim F_{(k-1,\nu)}$;

(iii) $p\left(\boldsymbol{\phi}|\mathbf{y}\right)$ é monotonicamente decrescente da forma quadrática

$$Q(\boldsymbol{\phi}) = (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})'(CD_{22}C')^{-1}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}}).$$

Dessa forma, o ponto $\boldsymbol{\phi}_0 = (0, \ldots, 0)'$ está incluído na região R_α se, e somente se,

$$\frac{Q(\boldsymbol{\phi}_0)}{(k-1)s^2} = (\boldsymbol{\phi}_0 - \hat{\boldsymbol{\phi}})' \frac{(CD_{22}C')^{-1}}{(k-1)s^2} (\boldsymbol{\phi}_0 - \hat{\boldsymbol{\phi}}) \leq F_{(k-1,\nu,\alpha)}.$$

Agora, note que

$$CD_{22}C' = \begin{bmatrix} \frac{1}{R_{xx_1}} + \frac{1}{R_{xx_2}} & -\frac{1}{R_{xx_2}} & 0 & \cdots & 0 & 0 \\ -\frac{1}{R_{xx_2}} & \frac{1}{R_{xx_2}} + \frac{1}{R_{xx_3}} & -\frac{1}{R_{xx_3}} & \cdots & 0 & 0 \\ 0 & -\frac{1}{R_{xx_3}} & \frac{1}{R_{xx_3}} + \frac{1}{R_{xx_4}} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{R_{xx_{k-2}}} + \frac{1}{R_{xx_{k-1}}} & -\frac{1}{R_{xx_{k-1}}} \\ 0 & 0 & 0 & \cdots & -\frac{1}{R_{xx_{k-1}}} & \frac{1}{R_{xx_{k-1}}} + \frac{1}{R_{xx_k}} \end{bmatrix},$$

implicando, pelo Lema A.11, que

$$(CD_{22}C')^{-1} = \begin{bmatrix} \frac{R_{xx_1}(R_{xx_2}+\cdots+R_{xx_k})}{(R_{xx_1}+\cdots+R_{xx_k})} & \frac{R_{xx_1}(R_{xx_3}+\cdots+R_{xx_k})}{(R_{xx_1}+\cdots+R_{xx_k})} & \cdots & \frac{R_{xx_1}R_{xx_k}}{(R_{xx_1}+\cdots+R_{xx_k})} \\ \frac{R_{xx_1}(R_{xx_3}+\cdots+R_{xx_k})}{(R_{xx_1}+\cdots+R_{xx_k})} & \frac{(R_{xx_1}+R_{xx_2})(R_{xx_3}+\cdots+R_{xx_k})}{(R_{xx_1}+\cdots+R_{xx_k})} & \cdots & \frac{(R_{xx_1}+R_{xx_2})R_{xx_k}}{(R_{xx_1}+\cdots+R_{xx_k})} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{R_{xx_1}R_{xx_k}}{(R_{xx_1}+\cdots+R_{xx_k})} & \frac{(R_{xx_1}+R_{xx_2})R_{xx_k}}{(R_{xx_1}+\cdots+R_{xx_k})} & \cdots & \frac{(R_{xx_1}+\cdots+R_{xx_{k-1}})R_{xx_k}}{(R_{xx_1}+\cdots+R_{xx_k})} \end{bmatrix}.$$

Assim,

$\hat{\boldsymbol{\phi}}'(CD_{22}C')^{-1} = (a_{11},\ a_{12},\ \cdots,\ a_{1(k-1)})$, onde

$$a_{11} = \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1} R_{xx_1} \sum_{i=2}^{k} R_{xx_i}(\hat{\beta}_1 - \hat{\beta}_i),$$

$$a_{12} = \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1} \Big[R_{xx_1} \sum_{i=3}^{k} R_{xx_i}(\hat{\beta}_1 - \hat{\beta}_i) + R_{xx_2} \sum_{i=3}^{k} R_{xx_i}(\hat{\beta}_2 - \hat{\beta}_i)\Big],$$

$$a_{1(k-1)} = \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1} \Big[R_{xx_1}R_{xx_k}(\hat{\beta}_1 - \hat{\beta}_k) + \cdots + R_{xx_{k-1}}R_{xx_k}(\hat{\beta}_{k-1} - \hat{\beta}_k)\Big].$$

Logo,

$$\begin{aligned} Q(\boldsymbol{\phi}_0) &= \hat{\boldsymbol{\phi}}'(CD_{22}C')^{-1}\hat{\boldsymbol{\phi}} \\ &= \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1} \Big[\Big(\sum_{i=1}^{k} R_{xx_i}\Big)\Big(\sum_{i=1}^{k} \hat{\beta}_i^2 R_{xx_i}\Big) - \Big(\sum_{i=1}^{k} \hat{\beta}_i R_{xx_i}\Big)^2\Big] \\ &= \Big(\sum_{i=1}^{k} \frac{R_{xy_i}^2}{R_{xx_i}^2} R_{xx_i}\Big) - \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1} \Big(\sum_{i=1}^{k} \frac{R_{xy_i}}{R_{xx_i}} R_{xx_i}\Big)^2 \\ &= \sum_{i=1}^{k} \frac{R_{xy_i}^2}{R_{xx_i}} - \Big[\Big(\sum_{i=1}^{k} R_{xy_i}\Big)^2\Big] \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1} \end{aligned}.$$

Dessa maneira, o ponto $\boldsymbol{\phi}_0 = (0, \ldots, 0)'$ está incluido na região R_α se, e somente se,

$$\Big[\sum_{i=1}^{k} \frac{R_{xy_i}^2}{R_{xx_i}} - \Big[\Big(\sum_{i=1}^{k} R_{xy_i}\Big)^2\Big] \Big(\sum_{i=1}^{k} R_{xx_i}\Big)^{-1}\Big] [(k-1)s^2]^{-1} \leq F_{(k-1,\nu,\alpha)}.$$

Caso $\boldsymbol{\phi}_0 = (0, \ldots, 0)'$ esteja incluido na região R_α então não rejeitamos a hipótese da homogeneidade dos coeficientes angulares e deveremos proceder às análises via modelo de covariância linear com coeficiente angular, β, comum a todos os tratamentos. Dessa forma é preciso adaptar a matriz do modelo como já foi comentado anteriomente.

Caso contrário, nós rejeitamos a hipótese $H_0 : \beta_1 = \cdots = \beta_k = \beta$, e o modelo a ser utilizado é o apresentado em (2.1).

Supondo que queremos fazer inferência para apenas um contraste linear ψ_j de $\boldsymbol{\tau}$, $j = 1, \ldots, k$, a inferência marginal sobre esse contraste de interesse deverá ser feita conforme se segue.

Já sabemos que

$$\psi_j = \sum_{i=1}^{k} c_{ij}\tau_i \sim t_1\Big(\sum_{i=1}^{k} c_{ij}\hat{\tau}_i, s^2 h_{jj}, \nu\Big)$$

onde $i = 1, \ldots, k$, $j = 1, \ldots, n_i$, $\sum_{i=1}^{k} c_{ij} = 0$ e h_{jj} é o j-ésimo elemento da matriz $H = CD_{11}C'$, e C é uma matriz escolhida de forma adequada. Portanto, a função densidade de probabilidade marginal de qualquer contraste j será dada por:

$$p\,(\psi_j|\mathbf{y}) \quad = \quad \frac{\Gamma(\frac{\nu+1}{2})(h_{jj})^{-1/2}s^{-1}}{(\Gamma(1/2))\Gamma(\nu/2)\nu^{1/2}}\Big[1 + \frac{(\psi_j - \hat{\psi}_j)^2}{\nu s^2 h_{jj}}\Big]^{-\frac{\nu+1}{2}}, \tag{2.19}$$

com $-\infty < \psi_j < +\infty$.

Supondo que queremos verificar se esse contraste j é nulo, então basta para isso construir um intervalo com $(1-\alpha)$ de credibilidade a *posteriori* e observar se este contém o ponto zero. Então, pela distribuição de ψ_j dada em (2.19), temos que o intervalo com $(1-\alpha)$ de credibilidade a *posteriori* será dado por

$$\hat{\psi}_j - t_{\frac{\alpha}{2}}\sqrt{s^2 h_{jj}} \leq \psi_j|\mathbf{y} \leq \hat{\psi}_j + t_{\frac{\alpha}{2}}\sqrt{s^2 h_{jj}}.$$

Acabamos de mostrar, portanto, como empregamos a análise Bayesiana, trabalhada com a informação adicional e neutra, a informação a *priori* do tipo não informativa, para conseguirmos fazer inferências para todos os parâmetros do modelo usando a definição de região H.P.D. e suas propriedades.

Convém salientar ainda que o modelo definido em (2.1) generaliza aquele definido em (2.7.7), pag. 114 da referência [1], no sentido de que este último leva em consideração apenas os efeitos médios dos tratamentos, não fazendo nenhuma alusão as variáveis explicativas do modelo.

Capítulo 3

Modelo de Covariância com Erros nas Variáveis

Neste capítulo, assim como no **Capítulo 2**, vamos considerar um experimento planejado com k tratamentos, sendo cada um deles repetido em n_i unidades experimentais, $i = 1, 2, \ldots, k$, sendo que aqui as variáveis explicativas x_{ij} são medidas com erros. Sabemos que não existem instrumentos de medição que sejam perfeitos, então esta suposição está bem fundamentada. Dessa maneira, não observamos o valor verdadeiro da variável x_{ij}, mas sim um valor X_{ij} tal que

$$X_{ij} = x_{ij} + u_{ij},$$

onde u_{ij} representa o erro correspondente, na medição de x_{ij}, com $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$.

3.1 O Modelo Estatístico

O modelo estatístico adequado a esta situação descrita acima é

$$\begin{aligned} Y_{ij} &= \tau_i + \beta_i x_{ij} + e_{ij}, \\ X_{ij} &= x_{ij} + u_{ij}, \end{aligned} \tag{3.1}$$

onde $j = 1, 2, \ldots, n_i, i = 1, 2, \ldots, k$.

Neste caso faremos as seguintes considerações:

(i) $e_{ij}|\boldsymbol{\theta} \sim N(0, \sigma_{ee})$ e são independentes;

(ii) $u_{ij}|\boldsymbol{\theta} \sim N(0, \sigma_{uu})$ e são independentes;

(iii) $x_{ij}|\boldsymbol{\theta} \sim N(\mu_x, \sigma_{xx})$ e são independentes;

(iv) para cada $j = 1, 2, \ldots, n_i$, e para cada $i = 1, 2, \ldots, k$, e_{ij}, u_{ij} e x_{ij} são não correlacionados.

Aqui $\boldsymbol{\theta} = (\boldsymbol{\tau}', \boldsymbol{\beta}', \sigma_{ee}, \sigma_{uu})'$ representa o vetor dos parâmetros envolvidos no modelo, com $\boldsymbol{\tau}' = (\tau_1, \tau_2, \ldots, \tau_k)$ e $\boldsymbol{\beta}' = (\beta_1, \beta_2, \ldots, \beta_k)$.

Com as suposições i-iv o modelo 3.1 é chamado de Modelo de Covariância com Erros nas Variáveis.

Vamos supor ainda que a razão de variâncias $\lambda_e = \dfrac{\sigma_{ee}}{\sigma_{uu}}$ seja conhecida, para que assim possamos reduzir o número de parâmetros envolvidos no modelo.

As conclusões feitas nesse capítulo são baseadas na análise feita no cap.3 da referência [9], que trabalha com modelo de regressão linear simples com erros nas variáveis quando estes erros pertencem a classe das distribuições elípticas. Por isso, sempre que demonstrarmos qualquer resultado para o nosso modelo estaremos fazendo uma analogia ao modelo de regressão linear simples.

3.2 A Função de Verossimilhança

Com o objetivo de dar uma formulação geral da função de verossimilhança, consideramos a f.d.p. de $(Z', \mathbf{x}')'$, dado $\boldsymbol{\theta}$, por:

$$f(Z, \mathbf{x}|\boldsymbol{\theta}) = l(Z|\boldsymbol{\theta}, \mathbf{x})h(\mathbf{x}|\boldsymbol{\theta}), \text{ (ver Zellner, 1971, p. 130)} \tag{3.2}$$

onde $l(Z|\boldsymbol{\theta}, \mathbf{x})$ é a densidade condicional de Z dado os vetores $\boldsymbol{\theta}$ e $\mathbf{x}$, e $h(\mathbf{x}|\boldsymbol{\theta})$ é a densidade condicional de $\mathbf{x}$ dado o vetor $\boldsymbol{\theta}$. O vetor $Z' = (Y', X')$ é o vetor das observações, com

$$Y' = (Y_{11}, \ldots, Y_{1n_1}, Y_{21}, \ldots, Y_{2n_2}, \ldots, \ldots, Y_{k1}, \ldots, Y_{kn_k})$$

e

$$X' = (X_{11}, \ldots, X_{1n_1}, X_{21}, \ldots, X_{2n_2}, \ldots, \ldots, X_{k1}, \ldots, X_{kn_k}),$$

e o vetor $\mathbf{x}$ é o vetor das variáveis não observadas dado por

$$\mathbf{x}' = (x_{11}, \ldots, x_{1n_1}, x_{21}, \ldots, x_{2n_2}, \ldots, \ldots, x_{k1}, \ldots, x_{kn_k})$$

Observemos que $\forall j = 1, 2, \ldots, n_i,$ e $\forall i = 1, 2, \ldots, k,$

$$E(Y_{ij}|\boldsymbol{\theta}, \mathbf{x}) = \tau_i + \beta_i x_{ij} \text{ e } E(X_{ij}|\boldsymbol{\theta}, \mathbf{x}) = x_{ij}.$$

Dessa forma, para I_{n_i} denotando a matriz identidade de ordem $n_i, \ i = 1, \ldots, k,$ se definirmos a matriz

$$B = \begin{bmatrix} \beta_1 I_{n_1} & 0 & \cdots & 0 \\ 0 & \beta_2 I_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \beta_k I_{n_k} \end{bmatrix}_{n \times n}$$

e a matriz

$$\mathbf{1}_n = \begin{bmatrix} \mathbf{i}_{n_1} & 0 & \cdots & 0 \\ 0 & \mathbf{i}_{n_2} & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \mathbf{i}_{n_k} \end{bmatrix}_{n \times k},$$

onde $\mathbf{i}_{n_i}$ representa um vetor n_i-dimensional dado por $\mathbf{i}_{n_i} = (1, \ldots, 1)', \ i = 1, 2, \ldots, k$ e $n = \sum_{i=1}^{k} n_i$, então, podemos escrever

$$E(Z|\boldsymbol{\theta}, \mathbf{x}) = \begin{bmatrix} E(Y|\boldsymbol{\theta}, \mathbf{x}) \\ E(X|\boldsymbol{\theta}, \mathbf{x}) \end{bmatrix} = \begin{bmatrix} \mathbf{1}_n \boldsymbol{\tau} + B\mathbf{x} \\ \mathbf{x} \end{bmatrix}_{2n \times 1} = \boldsymbol{\mu}_{Z|(\boldsymbol{\theta}, \mathbf{x})}.$$

Note ainda que, para $j = 1, 2, \ldots, n_i$, $i = 1, 2, \ldots, k$, e $\forall l \neq i$ ou $\forall m \neq j$ temos:

- $Var(Y_{ij}|\boldsymbol{\theta}, \mathbf{x}) = Var(e_{ij}|\boldsymbol{\theta}, \mathbf{x}) = \sigma_{ee}$;
- $Cov(Y_{ij}, Y_{lm}|\boldsymbol{\theta}, \mathbf{x}) = E(e_{ij}e_{lm}|\boldsymbol{\theta}, \mathbf{x}) = 0$;
- $Var(X_{ij}|\boldsymbol{\theta}, \mathbf{x}) = Var(u_{ij}|\boldsymbol{\theta}, x) = \sigma_{uu}$;
- $Cov(X_{ij}, X_{lm}|\boldsymbol{\theta}, \mathbf{x}) = E(u_{ij}u_{lm}|\boldsymbol{\theta}, \mathbf{x}) = 0$;
- $Cov(X_{ij}, Y_{lm}|\boldsymbol{\theta}, \mathbf{x}) = E(u_{ij}e_{lm}|\boldsymbol{\theta}, \mathbf{x}) = 0$;
- $Cov(X_{ij}, Y_{ij}|\boldsymbol{\theta}, \mathbf{x}) = E(u_{ij}e_{ij}|\boldsymbol{\theta}, \mathbf{x}) = 0$;

Logo, na forma matricial,

$$Cov(Z|\boldsymbol{\theta}, \mathbf{x}) = \begin{bmatrix} Cov(Y|\boldsymbol{\theta}, \mathbf{x}) & 0 \\ 0 & Cov(X|\boldsymbol{\theta}, \mathbf{x}) \end{bmatrix} = \begin{bmatrix} \sigma_{ee}I_n & 0 \\ 0 & \sigma_{uu}I_n \end{bmatrix} = \Sigma_{Z|(\boldsymbol{\theta}, \mathbf{x})}$$

Portanto,

$$Z|(\boldsymbol{\theta}, \mathbf{x}) \sim N_{2n}(\boldsymbol{\mu}_{Z|(\boldsymbol{\theta}, \mathbf{x})}, \Sigma_{Z|(\boldsymbol{\theta}, \mathbf{x})}), \tag{3.3}$$

ou seja, usando o **Teorema A.4** do Apêndice A temos

$$\begin{aligned} l(Z|\boldsymbol{\theta}, \mathbf{x}) &= (2\pi)^{-n}(\sigma_{ee}\sigma_{uu})^{-n/2} \exp\Big\{-\frac{1}{2}\Big[\frac{1}{\sigma_{ee}}(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x})'(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x}) + \\ &\quad + \frac{1}{\sigma_{uu}}(X - \mathbf{x})'(X - \mathbf{x})\Big]\Big\} \\ &= (2\pi)^{-n}\lambda_e^{-n/2}\sigma_{uu}^{-n} \exp\Big\{-\frac{1}{2\sigma_{uu}}\Big[\frac{1}{\lambda_e}(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x})'(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x}) + \\ &\quad + (X - \mathbf{x})'(X - \mathbf{x})\Big]\Big\} \\ &= (2\pi)^{-n}\lambda_e^{-n/2}\sigma_{uu}^{-n} \exp\Big\{-\frac{1}{2\sigma_{uu}}A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)\Big\} \end{aligned} \tag{3.4}$$

onde $\lambda_e = \dfrac{\sigma_{ee}}{\sigma_{uu}}$ e $A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z) = \dfrac{1}{\lambda_e}(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x})'(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x}) + (X - \mathbf{x})'(X - \mathbf{x})$.

Finalmente, $f(Z, \mathbf{x}|\boldsymbol{\theta})$ fica dada por

$$\begin{aligned} f(Z, \mathbf{x}|\boldsymbol{\theta}) &= l(Z|\boldsymbol{\theta}, \mathbf{x})h(\mathbf{x}|\boldsymbol{\theta}) \\ &= (2\pi)^{-n}\lambda_e^{-n/2}\sigma_{uu}^{-n} \exp\Big\{-\frac{1}{2\sigma_{uu}}A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)\Big\}h(\mathbf{x}|\boldsymbol{\theta}). \end{aligned} \tag{3.5}$$

3.3 A *Posteriori* de $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x})$

Vamos agora enunciar o Teorema que nos dará a distribuição a *posteriori* de $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x})$.

Teorema 3.1 *Suponha que* $Z|(\boldsymbol{\theta}, \mathbf{x})$ *tem densidade dada por (3.4), que* $\mathbf{x} \perp \boldsymbol{\theta}$, *isto é,* $h(\mathbf{x}|\boldsymbol{\theta}) = \pi(\mathbf{x})$ *sendo* $\pi(\mathbf{x})$ *qualquer priori arbitrária para* $\mathbf{x}$, *completamente especificada, e que* $\boldsymbol{\theta} = (\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma_{ee}, \sigma_{uu})$ *tem densidade a priori dada na forma*

$$\pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma_{ee}, \sigma_{uu}) \propto \frac{1}{\sigma_{ee}\sigma_{uu}} \pi(\boldsymbol{\tau}, \boldsymbol{\beta}),$$

onde $\pi(\boldsymbol{\tau}, \boldsymbol{\beta})$ *é qualquer densidade a priori para* $(\boldsymbol{\tau}, \boldsymbol{\beta})$. *Então, tem-se que* $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x})$ *tem distribuição a posteriori dada por*

$$\Pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z) \propto \lambda_e^{\frac{n-2}{2}} [(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x})'(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x}) + \lambda_e(X - \mathbf{x})'(X - \mathbf{x})]^{-n} \pi(\mathbf{x})\pi(\boldsymbol{\tau}, \boldsymbol{\beta}).$$

Prova. Sabemos que

$$\begin{aligned}
\Pi(\boldsymbol{\theta}, \mathbf{x}|Z) &\propto l(Z|\boldsymbol{\theta}, \mathbf{x}) h(\mathbf{x}|\boldsymbol{\theta}) \pi(\boldsymbol{\theta}) \\
&= l(Z|\boldsymbol{\theta}, \mathbf{x}) \pi(\mathbf{x}) \pi(\boldsymbol{\theta}) \\
&\propto \lambda_e^{-n/2} \sigma_{uu}^{-n} exp\left\{-\frac{1}{2\sigma_{uu}} A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)\right\} \pi(\mathbf{x}) \frac{1}{\sigma_{ee}\sigma_{uu}} \pi(\boldsymbol{\tau}, \boldsymbol{\beta})
\end{aligned}$$

onde a última expressão segue do resultado (3.4) e das hipóteses para a distribuição a *priori* de $\boldsymbol{\theta}$.

Considere a transformação $(\boldsymbol{\tau}, \boldsymbol{\beta}, \sigma_{ee}, \sigma_{uu}) \longrightarrow (\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \sigma_{uu})$. Logo,

$$\begin{aligned}
\Pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \sigma_{uu}, \mathbf{x}|Z) &\propto \lambda_e^{-n/2} \sigma_{uu}^{-n} \exp\left\{-\frac{1}{2\sigma_{uu}} A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)\right\} \sigma_{uu} \frac{1}{\sigma_{ee}\sigma_{uu}} \pi(\boldsymbol{\tau}, \boldsymbol{\beta}) \pi(\mathbf{x}) \\
&= \lambda_e^{-\frac{n}{2}-1} \sigma_{uu}^{-(n+1)} \exp\left\{-\frac{1}{2\sigma_{uu}} A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)\right\} \pi(\boldsymbol{\tau}, \boldsymbol{\beta}) \pi(\mathbf{x}).
\end{aligned}$$

Dessa forma,

$$\begin{aligned}
\Pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z) &= \int_{\sigma_{uu}} \Pi(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \sigma_{uu}, \mathbf{x}|Z) d\sigma_{uu} \\
&\propto \lambda_e^{-\frac{n+2}{2}} \int_0^{+\infty} \sigma_{uu}^{-(n+1)} \exp\left\{-\frac{1}{2\sigma_{uu}} A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)\right\} d\sigma_{uu} \pi(\boldsymbol{\tau}, \boldsymbol{\beta}) \pi(\mathbf{x}) \\
&= \lambda_e^{-\frac{n+2}{2}} [A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)]^{-n} 2^n \int_0^{+\infty} v^{n-1} \exp\{-v\} dv \, \pi(\boldsymbol{\tau}, \boldsymbol{\beta}) \pi(\mathbf{x}) dv \\
&\propto \lambda_e^{-\frac{n+2}{2}} \big[\frac{1}{\lambda_e}(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x})'(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x}) + (X - \mathbf{x})'(X - \mathbf{x})\big]^{-n} \times \\
&\quad \times \pi(\boldsymbol{\tau}, \boldsymbol{\beta}) \pi(\mathbf{x}) \\
&= \lambda_e^{\frac{n-2}{2}} [(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x})'(Y - \mathbf{1}_n\boldsymbol{\tau} - B\mathbf{x}) + \lambda_e(X - \mathbf{x})'(X - \mathbf{x})]^{-n} \times \\
&\quad \times \pi(\boldsymbol{\tau}, \boldsymbol{\beta}) \pi(\mathbf{x}),
\end{aligned}$$

onde a terceira equação ocorre para $v = \dfrac{A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|Z)}{2\sigma_{uu}}$.

Esse resultado coincide com o resultado apresentado na referência [9].

3.4 A *Posteriori* de $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e)$

Nessa seção enuciaremos o corolário que nos fornecerá a distribuição a *posteriori* de $(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e)$, mas antes provemos o seguinte Lema sobre o estimador de máxima verossimilhança do vetor não observável $\mathbf{x}$.

Lema 3.2 *Sob a verossimilhança de $(Z', \mathbf{x}')'$ como dada em (3.5) o estimador de máxima verossimilhança para o vetor $\mathbf{x}$, denotado por $\hat{\mathbf{x}}$, é dado por*

$$\hat{\mathbf{x}} = (B'B + \lambda_e I_n)^{-1}[B'(Y - \mathbf{1}_n\boldsymbol{\tau}) + \lambda_e X].$$

Prova. Note que com a hipótese $\pi(\mathbf{x}) \propto$ constante teremos

$$\left.\frac{\partial \log f(z, \mathbf{x}|\boldsymbol{\theta})}{\partial \mathbf{x}}\right|_{\mathbf{x}=\hat{\mathbf{x}}} = 0 \iff -\frac{1}{2\sigma_{uu}}\left[-\frac{2}{\lambda_e}B'(Y - \mathbf{1}_n\boldsymbol{\tau}) + \frac{2}{\lambda_e}B'B\hat{\mathbf{x}} - 2X + 2\hat{\mathbf{x}}\right] + \frac{\pi'(\mathbf{x})}{\pi(\mathbf{x})}$$

$$\iff X + \frac{B'}{\lambda_e}(Y - \mathbf{1}_n\boldsymbol{\tau}) = \left(\frac{B'B}{\lambda_e} + I_n\right)\hat{\mathbf{x}}$$

$$\iff \lambda_e X + B'(Y - \mathbf{1}_n\boldsymbol{\tau}) = (B'B + \lambda_e I_n)\hat{\mathbf{x}}$$

$$\iff \hat{\mathbf{x}} = (B'B + \lambda_e I_n)^{-1}[B'(Y - \mathbf{1}_n\boldsymbol{\tau}) + \lambda_e X],$$

como queríamos demonstrar.

Esse resultado generaliza o estimador obtido no modelo de regressão linear simples dado em (3.2), pois nesse caso,

$$\hat{\mathbf{x}} = \frac{\beta(Y - \alpha\mathbf{i_n}) + \lambda_e X}{\beta^2 + \lambda_e}.$$

Note agora que

$$\begin{aligned}
(i)\ I_n - B(B'B + \lambda_e I_n)^{-1}B' &= [(B(B'B + \lambda_e I_n)^{-1}B')^{-1} - I_n][B(B'B + \lambda_e I_n)^{-1}B'] \\
&= [(B')^{-1}(B'B + \lambda_e I_n)B^{-1} - I_n](BB')(B'B + \lambda_e I_n)^{-1} \\
&= [I_n + (B')^{-1}(\lambda_e I_n)B^{-1} - I_n](BB')(B'B + \lambda_e I_n)^{-1} \\
&= \lambda_e(B')^{-1}B^{-1}(BB')(B'B + \lambda_e I_n)^{-1} \\
&= \lambda_e(B'B + \lambda_e I_n)^{-1},
\end{aligned}$$

e

$$
\begin{aligned}
(ii)\quad I_n - \lambda_e(B'B+\lambda_e I_n)^{-1} &= (B'B+\lambda_e I_n)^{-1}[(B'B+\lambda_e I_n)-\lambda_e I_n]\\
&= (B'B+\lambda_e I_n)^{-1}(B'B)\\
&= B'(B'B+\lambda_e I_n)^{-1}B
\end{aligned}
$$

Assim, podemos reescrever a função $A = A(\boldsymbol{\tau}, \boldsymbol{\beta}, \lambda_e, \mathbf{x}|z)$ na forma abaixo:

$$
\begin{aligned}
A &= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau}-B\mathbf{x})'(Y-\mathbf{1}_n\boldsymbol{\tau}-B\mathbf{x}) + (X-\mathbf{x})'(X-\mathbf{x})\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) - 2\mathbf{x}'\lambda_e^{-1}[B'(Y-\mathbf{1}_n\boldsymbol{\tau})+\lambda_e X]+\\
&\quad +\mathbf{x}'\lambda_e^{-1}(B'B+\lambda_e I_n)\mathbf{x} + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) - 2\mathbf{x}'\lambda_e^{-1}B'(Y-\mathbf{1}_n\boldsymbol{\tau})\hat{\mathbf{x}} + \frac{\mathbf{x}'}{\lambda_e}(B'B+\lambda_e I_n)\mathbf{x} + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \mathbf{x}'\lambda_e^{-1}[-2(B'B+\lambda_e I_n)\hat{\mathbf{x}} + (B'B+\lambda_e I_n)\mathbf{x}] + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \mathbf{x}'\lambda_e^{-1}(B'B+\lambda_e I_n)(\mathbf{x}-2\hat{\mathbf{x}}) + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}}+\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}}-\hat{\mathbf{x}}) + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad -\lambda_e^{-1}\hat{\mathbf{x}}'(B'B+\lambda_e I_n)\hat{\mathbf{x}} + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad -\lambda_e^{-1}\{(B'B+\lambda_e I_n)^{-1}[B'(Y-\mathbf{1}_n\boldsymbol{\tau})+\lambda_e\, X]\}'(B'B+\lambda_e I_n)\times\\
&\quad \times\{(B'B+\lambda_e I_n)^{-1}[B'(Y-\mathbf{1}_n\boldsymbol{\tau})+\lambda_e\, X]\} + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad -\lambda_e^{-1}[B'(Y-\mathbf{1}_n\boldsymbol{\tau})+\lambda_e\, X]'(B'B+\lambda_e I_n)^{-1}[B'(Y-\mathbf{1}_n\boldsymbol{\tau})+\lambda_e\, X] + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'(Y-\mathbf{1}_n\boldsymbol{\tau}) + \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad -\lambda_e^{-1}[(Y-\mathbf{1}_n\boldsymbol{\tau})'B(B'B+\lambda_e I_n)^{-1}B'(Y-\mathbf{1}_n\boldsymbol{\tau})+\\
&\quad +2\lambda_e(Y-\mathbf{1}_n\boldsymbol{\tau})B(B'B+\lambda_e I_n)^{-1}X + {\lambda_e}^2 X'(B'B+\lambda_e I_n)^{-1}X] + X'X\\
&= \lambda_e^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})'[I_n - B(B'B+\lambda_e I_n)^{-1}B'](Y-\mathbf{1}_n\boldsymbol{\tau})+\\
&\quad +\lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}}) - 2(Y-\mathbf{1}_n\boldsymbol{\tau})'B(B'B+\lambda_e I_n)^{-1}X+\\
&\quad +X'[I_n-\lambda_e(B'B+\lambda_e I_n)^{-1}]X\\
&= (Y-\mathbf{1}_n\boldsymbol{\tau})'(B'B+\lambda_e I_n)^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau}) + \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad -2(Y-\mathbf{1}_n\boldsymbol{\tau})'(B'B+\lambda_e I_n)^{-1}BX + X'B'(B'B+\lambda_e I_n)^{-1}BX, \text{ por } (i) \text{ e } (ii)
\end{aligned}
$$

$$
\begin{aligned}
&= (Y-\mathbf{1}_n\boldsymbol{\tau})'(B'B+\lambda_e I_n)^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau})+\lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad -2(Y-\mathbf{1}_n\boldsymbol{\tau})'(B'B+\lambda_e I_n)^{-1}(BX)+(BX)'(B'B+\lambda_e I_n)^{-1}(BX)\\
&= \lambda_e^{-1}(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad +(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)'(B'B+\lambda_e I_n)^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)\\
&= \lambda_e^{-1}[(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+\\
&\quad +\lambda_e(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)'(B'B+\lambda_e I_n)^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)].
\end{aligned}
$$

Logo,

$$
\begin{aligned}
\lambda_e A(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x}|z) &= \lambda_e(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)'(B'B+\lambda_e I_n)^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)+\\
&\quad +(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}}). \qquad (3.6)
\end{aligned}
$$

Esta nova forma de escrever a função $A(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x}|z)$ será extremamente útil na demonstração do colorário abaixo.

Corolário 3.1 *Sob as hipóteses do Teorema 3.1, com a suposição de que*

$$\pi(\mathbf{x}) \propto \textit{ constante},$$

e fazendo

$$\hat{\boldsymbol{\eta}} = (\Lambda_X'\Lambda_X)^{-1}\Lambda_X'Y$$

e

$$\nu s^2 = (Y-\Lambda_X\hat{\boldsymbol{\eta}})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\hat{\boldsymbol{\eta}})$$

tem-se que $(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e)$ *tem densidade a posteriori dada por:*

$$
\begin{aligned}
\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e|Z) &\propto \lambda_e^{-1}[\nu s^2+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_X'(B'B+\lambda_e I_n)^{-1}\Lambda_X(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})]^{\frac{n}{2}}\times\\
&\quad \times|B'B+\lambda_e I_n|^{-\frac{1}{2}}\pi(\boldsymbol{\tau},\boldsymbol{\beta}),
\end{aligned}
$$

onde $\boldsymbol{\eta}=(\boldsymbol{\tau}'\ \boldsymbol{\beta}')'$ *e*

$$
\Lambda_X = \begin{bmatrix}
\mathbf{i}_{n_1} & 0 & \cdots & 0 & X_1 & 0 & \cdots & 0\\
0 & \mathbf{i}_{n_2} & \cdots & 0 & 0 & X_2 & \cdots & 0\\
\vdots & \vdots & \ddots & \vdots & \vdots & \vdots & \ddots & \vdots\\
0 & 0 & \cdots & \mathbf{i}_{n_k} & 0 & 0 & \cdots & X_k
\end{bmatrix}
$$

para $X_i=(X_{i1},\ldots,X_{in_i})'$, $\forall\ i=1,\ldots,k$ *e* $\nu=$ *n-2k.*

Prova. Usando a suposição sob $\pi(\mathbf{x})$ e o resultado (3.6), podemos escrever a distribuição a *posteriori* de $(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x})$ na seguinte forma

$$
\begin{aligned}
\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x}|Z) &\propto \lambda_e^{\frac{n-2}{2}}[(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})+ \\
&\quad +\lambda_e(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-n}\pi(\boldsymbol{\tau},\boldsymbol{\beta}) \\
&= \lambda_e^{\frac{n-2}{2}}\left[1+\frac{n(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})}{n\lambda_e(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})}\right]^{-n}\times \\
&\quad \times[\lambda_e(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-n}\pi(\boldsymbol{\tau},\boldsymbol{\beta})
\end{aligned}
$$

pois,

$$
(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)'(B'B+\lambda_e I_n)^{-1}(Y-\mathbf{1}_n\boldsymbol{\tau}-BX)=(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})
$$

Agora podemos encontrar $\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e|Z)$, integrando $\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x}|Z)$ em relação a $\mathbf{x}$. Logo,

$$
\begin{aligned}
\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e|Z) &= \int_{\mathbf{x}}\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x}|Z)d\mathbf{x} \\
&\propto \lambda_e^{\frac{n-2}{2}}[\lambda_e(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-n}\pi(\boldsymbol{\tau},\boldsymbol{\beta})\times \\
&\quad \times\int_{\mathbf{x}}\left[1+\frac{n(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})}{n\lambda_e(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})}\right]^{-n}d\mathbf{x} \\
&= \lambda_e^{-\frac{n+2}{2}}[(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-n}\pi(\boldsymbol{\tau},\boldsymbol{\beta})\times \\
&\quad \times\left|\frac{\lambda_e^{-1}(B'B+\lambda_e I_n)n}{(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})}\right|^{-1/2}\times \\
&\quad \times\int_{\mathbf{x}}\left|\frac{\lambda_e^{-1}(B'B+\lambda_e I_n)n}{(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})}\right|^{1/2}\times \\
&\quad \times\left[1+\frac{n(\mathbf{x}-\hat{\mathbf{x}})'(B'B+\lambda_e I_n)(\mathbf{x}-\hat{\mathbf{x}})}{n\lambda_e(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})}\right]^{-n}d\mathbf{x} \\
&\propto \lambda_e^{-\frac{n+2}{2}+\frac{n}{2}}[(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-n}\times \\
&\quad \times[(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-\frac{n}{2}}\times \\
&\quad \times|B'B+\lambda_e I_n|^{-\frac{1}{2}}\pi(\boldsymbol{\tau},\boldsymbol{\beta}) \\
&= \lambda_e^{-1}[(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-\frac{n}{2}}\times \\
&\quad \times|B'B+\lambda_e I_n|^{-\frac{1}{2}}\pi(\boldsymbol{\tau},\boldsymbol{\beta}).
\end{aligned}
$$

Note que a penúltima equação ocorre porque o integrando é proporcional à função densidade de probabilidade de um vetor aleatório $\mathbf{x}$ tal que

$$\mathbf{x} \sim t_n\left(\hat{\mathbf{x}}, \frac{(B'B+\lambda_e I_n)^{-1}\lambda_e}{n[(Y-\Lambda_X\boldsymbol{\eta})'(B'B+\lambda_e I_n)^{-1}(Y-\Lambda_X\boldsymbol{\eta})]^{-1}}, n\right).$$

Mais uma vez, fazendo uma analogia ao modelo de regressão linear simples teremos

$$\begin{aligned}
\Pi(\alpha,\beta,\lambda_e|Z) &\propto \lambda_e^{-1}[(Y-\alpha\mathbf{1}_n-\beta X)'((\beta^2+\lambda_e)I_n)^{-1}(Y-\alpha\mathbf{1}_n-\beta X)]^{\frac{n}{2}} \times \\
&\quad \times|(\beta^2+\lambda_e)I_n|^{-\frac{1}{2}}\pi(\alpha,\beta) \\
&= \lambda_e^{-1}[(Y-\alpha\mathbf{1}_n-\beta X)'(Y-\alpha\mathbf{1}_n-\beta X)]^{-\frac{n}{2}}\pi(\alpha,\beta).
\end{aligned}$$

Agora note que para $M = B'B+\lambda_e I_n$ temos:

$$\begin{aligned}
(Y-\Lambda_X\boldsymbol{\eta})'M^{-1}(Y-\Lambda_X\boldsymbol{\eta}) &= (Y-\Lambda_X\hat{\boldsymbol{\eta}}+\Lambda_X\hat{\boldsymbol{\eta}}-\Lambda_X\boldsymbol{\eta})'M^{-1}(Y-\Lambda_X\hat{\boldsymbol{\eta}}+\Lambda_X\hat{\boldsymbol{\eta}}-\Lambda_X\boldsymbol{\eta}) \\
&= (Y-\Lambda_X\hat{\boldsymbol{\eta}})'M^{-1}(Y-\Lambda_X\hat{\boldsymbol{\eta}})-2(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_X'M^{-1}(Y-\Lambda_X\hat{\boldsymbol{\eta}})+ \\
&\quad +(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_X'M^{-1}\Lambda_X(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) \\
&= (Y-\Lambda_X\hat{\boldsymbol{\eta}})'M^{-1}(Y-\Lambda_X\hat{\boldsymbol{\eta}})+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_X'M^{-1}\Lambda_X(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}}) \\
&= \nu s^2+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_X'M^{-1}\Lambda_X(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})
\end{aligned}$$

por que,

$$\Lambda_X'M^{-1}(Y-\Lambda_X\hat{\boldsymbol{\eta}}) = \begin{bmatrix}
\sum_{j=1}^{n_1}(\beta_1+\lambda_e)^{-1}(Y_{1j}-\hat{\tau}_1-\hat{\beta}_1X_{1j}) \\
\vdots \\
\sum_{j=1}^{n_k}(\beta_k+\lambda_e)^{-1}(Y_{kj}-\hat{\tau}_k-\hat{\beta}_kX_{kj}) \\
\sum_{j=1}^{n_1}X_{1j}(\beta_1+\lambda_e)^{-1}(Y_{1j}-\hat{\tau}_1-\hat{\beta}_1X_{1j}) \\
\vdots \\
\sum_{j=1}^{n_k}X_{kj}(\beta_k+\lambda_e)^{-1}(Y_{kj}-\hat{\tau}_k-\hat{\beta}_kX_{kj})
\end{bmatrix} = \begin{bmatrix} 0 \\ \vdots \\ 0 \\ 0 \\ \vdots \\ 0 \end{bmatrix}_{(2k\times 1)}$$

já que, $\forall\ i = 1,\ldots,k$,

$$
\begin{aligned}
a)\sum_{j=1}^{n_i}(Y_{ij}-\hat{\tau}_i-\hat{\beta}_iX_{ij}) &= (Y_{i.}-n_i\hat{\tau}_i-\hat{\beta}_iX_{i.})\\
&= \Big(Y_{i.}-\frac{Y_{i.}}{n_i}n_i+n_i\frac{X_{i.}}{n_i}\frac{R_{xy_i}}{R_{xx_i}}-\frac{R_{xy_i}}{R_{xx_i}}X_{.}\Big)\\
&= 0
\end{aligned}
$$

$$
\begin{aligned}
b)\sum_{j=1}^{n_i}X_{ij}(Y_{ij}-\hat{\tau}_i-\hat{\beta}_iX_{ij}) &= \Big(\sum_{j=1}^{n_i}X_{ij}Y_{ij}-\hat{\tau}_iX_{i.}-\hat{\beta}_i\sum_{j=1}^{n_i}X_{ij}^2\Big)\\
&= \Big(\frac{n_iR_{xy_i}+X_{i.}Y_{i.}}{n_i}-\frac{X_{i.}Y_{i.}}{n_i}+\frac{R_{xy_i}}{R_{xx_i}}\frac{X_{i.}^2}{n_i}-\frac{R_{xy_i}}{R_{xx_i}}\frac{n_iR_{xx_i}+X_{i.}^2}{n_i}\Big)\\
&= 0
\end{aligned}
$$

pois a matriz Λ_X é igual a matriz de planejamento X do capítulo 2 e, portanto, $\hat{\boldsymbol{\eta}}=\hat{\boldsymbol{\theta}}$ com esse $\hat{\boldsymbol{\theta}}$ sendo o mesmo do capítulo 2.

Logo,

$$
\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e|Z) \;\propto\; \lambda_e^{-1}[\nu s^2+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_X' M^{-1}\Lambda_X(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})]^{-\frac{n}{2}}|B'B+\lambda_eI_n|^{-1/2}\pi(\boldsymbol{\tau},\boldsymbol{\beta}),
$$

como queríamos demonstrar.

Note que nesse caso, não podemos afirmar que o vetor $\boldsymbol{\eta}$ tem distribuição t-Student $2k$-variada, como ocorre na referência [9].

3.5 A Densidade Preditiva de x

Vejamos agora a densidade preditiva de $\mathbf{x}$.

Corolário 3.2 *Sob as hipóteses do Teorema 3.1, com a suposição de que*

$$\pi(\boldsymbol{\tau},\boldsymbol{\beta}) \propto \textit{ constante},$$

temos que densidade a preditiva de $\mathbf{x}$ *é dada por*

$$
\Pi(\mathbf{x}|Z) \;\propto\; s^{2k}[(X-\mathbf{x})'(X-\mathbf{x})s^2]^{-n/2}\prod_{i=1}^{k}\frac{(\beta_i+\lambda_e)}{(n_iR_{xx_i})^{1/2}}\pi(\mathbf{x}),
$$

onde νs^2 *é como no colorário 3.1, mas com* $\mathbf{x}_1,\ldots,\mathbf{x}_k$ *substituindo* $X_1,\ldots,X_k$ *na matriz* Λ_X*, agora denotada por* $\Lambda_{\mathbf{x}}$.

Prova. Usando o resultado do Teorema 3.1 e a suposição $\pi(\boldsymbol{\tau},\boldsymbol{\beta}) \propto$ constante, podemos reescrever $\Pi(\boldsymbol{\tau},\boldsymbol{\beta},\lambda_e,\mathbf{x}|Z)$ como

$$\Pi(\boldsymbol{\eta},\lambda_e,\mathbf{x}|Z) \propto \lambda_e^{\frac{n-2}{2}}[(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})]^{-n}\Big[1+\frac{\lambda_e(X-\mathbf{x})'(X-\mathbf{x})}{(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})}\Big]^{-n}\pi(\mathbf{x}).$$

Agora podemos encontrar $\Pi(\mathbf{x}|Z)$ integrando $\Pi(\boldsymbol{\eta},\lambda_e,\mathbf{x}|Z)$ em relação a $(\boldsymbol{\eta},\lambda_e)$, portanto,

$$\begin{aligned}\Pi(\mathbf{x}|Z) &\propto \int_{\boldsymbol{\eta}}\int_{\lambda_e}\lambda_e^{\frac{n-2}{2}}[(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})]^{-n}\Big[1+\frac{(X-\mathbf{x})'(X-\mathbf{x})}{(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})}\lambda_e\Big]^{-n}\pi(\mathbf{x})d\lambda_e\, d\boldsymbol{\eta}\\ &= \int_{\boldsymbol{\eta}}[(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})]^{-n}\Big[\frac{(X-\mathbf{x})'(X-\mathbf{x})}{(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})}\Big]^{-n/2}\mathrm{Beta}(n/2,n/2)\pi(\mathbf{x})d\boldsymbol{\eta}\end{aligned}$$

já que

$$\int_0^{+\infty} x^{t-1}(1+bx)^{-s}dx = b^{-t}\mathrm{Beta}(t,s-t).$$

Daí,

$$\begin{aligned}\Pi(\mathbf{x}|Z) &\propto [(X-\mathbf{x})'(X-\mathbf{x})]^{-n/2}\pi(\mathbf{x})\int_{\boldsymbol{\eta}}[(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})'(Y-\Lambda_{\mathbf{x}}\boldsymbol{\eta})]^{-n/2}d\boldsymbol{\eta}\\ &\propto [(X-\mathbf{x})'(X-\mathbf{x})]^{-n/2}\pi(\mathbf{x})\int_{\boldsymbol{\eta}}[\nu s^2+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\Lambda_{\mathbf{x}}'M^{-1}\Lambda_{\mathbf{x}}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})]^{-n/2}d\boldsymbol{\eta}\\ &\propto [(X-\mathbf{x})'(X-\mathbf{x})]^{-n/2}\pi(\mathbf{x})\Big|\frac{\Lambda_{\mathbf{x}}'M^{-1}\Lambda_{\mathbf{x}}}{s^2}\Big|^{-1/2}\times\\ &\quad\times\int_{\boldsymbol{\eta}}(\nu s^2)^{-n/2}\Big[1+(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})'\frac{\Lambda_{\mathbf{x}}'M^{-1}\Lambda_{\mathbf{x}}}{\nu s^2}(\boldsymbol{\eta}-\hat{\boldsymbol{\eta}})\Big]^{-n/2}\Big|\frac{\Lambda_{\mathbf{x}}'M^{-1}\Lambda_{\mathbf{x}}}{s^2}\Big|^{1/2}d\boldsymbol{\eta}\\ &\propto s^{2k}[(X-\mathbf{x})'(X-\mathbf{x})s^2]^{-n/2}\pi(\mathbf{x})|\Lambda_{\mathbf{x}}'M^{-1}\Lambda_{\mathbf{x}}|^{-1/2}. \qquad (3.7)\end{aligned}$$

Note que a última equação ocorre porque o integrando é proporcional a função densidade de probabilidade de um vetor aleatório $\boldsymbol{\eta}$ tal que

$$\boldsymbol{\eta} \sim t_n\big(\hat{\boldsymbol{\eta}},[\Lambda_{\mathbf{x}}'M^{-1}\Lambda_{\mathbf{x}}]^{-1}s^2,\nu\big).$$

Agora observe que $M^{-1} = \begin{bmatrix} (\beta_1+\lambda_e)^{-1}I_{n_1} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & (\beta_k+\lambda_e)^{-1}I_{n_k} \end{bmatrix}_{(n\times n)}$,

então,

$$\Lambda'_{\mathbf{x}} M^{-1} \Lambda_{\mathbf{x}} = \begin{bmatrix} \frac{n_1}{\beta_1+\lambda_e} & \cdots & 0 & \frac{x_{1.}}{\beta_1+\lambda_e} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{n_k}{\beta_k+\lambda_e} & 0 & \cdots & \frac{x_{k.}}{\beta_k+\lambda_e} \\ \frac{n_1}{\beta_1+\lambda_e} & \cdots & 0 & \frac{n_1R_{xx_1}+x_{1.}^2}{n_1(\beta_1+\lambda_e)} & \cdots & 0 \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{x_{k.}}{\beta_k+\lambda_e} & 0 & \cdots & \frac{n_kR_{xx_k}+x_{k.}^2}{n_k(\beta_k+\lambda_e)} \end{bmatrix}.$$

Ou seja,

$$\Lambda'_{\mathbf{x}} M^{-1} \Lambda_{\mathbf{x}} = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

onde cada B_{ij} tem ordem $k \times k,\ i,j = 1,2.$

Como B_{22} é não singular então, pelo Lema A.10 do Apêndice A, temos que

$$|\Lambda'_{\mathbf{x}} M^{-1} \Lambda_{\mathbf{x}}| = |B_{22}||B_{11} - B_{12}B_{22}^{-1}B_{21}|.$$

Ora,

$$B_{12}B_{22}^{-1}B_{21} = \begin{bmatrix} \frac{(\beta_1+\lambda_e)^{-1}n_1x_{1.}^2}{(n_1R_{xx_1}+x_{1.}^2)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{(\beta_k+\lambda_e)^{-1}n_kx_{k.}^2}{(n_kR_{xx_k}+x_{k.}^2)} \end{bmatrix}, \text{ e daí}$$

$$B_{11} - B_{12}B_{22}^{-1}B_{21} = \begin{bmatrix} \frac{(\beta_1+\lambda_e)^{-1}n_1^2R_{xx_1}}{(n_1R_{xx_1}+x_{1.}^2)} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{(\beta_k+\lambda_e)^{-1}n_k^2R_{xx_k}}{(n_kR_{xx_k}+x_{k.}^2)} \end{bmatrix}.$$

Temos também que,

$$|B_{22}| = \prod_{i=1}^{k}(\beta_i+\lambda_e)^{-1}\Big(\frac{n_iR_{xx_i}+x_{i.}^2}{n_i}\Big) = \prod_{i=1}^{k}\frac{n_iR_{xx_i}+x_{i.}^2}{n_i(\beta_i+\lambda_e)}$$

e

$$|B_{11} - B_{12}B_{22}^{-1}B_{21}| = \prod_{i=1}^{k} \frac{n_i^2 R_{xx_i}}{(\beta_i + \lambda_e)(n_i R_{xx_i} + x_{i.}^2)}.$$

Logo,

$$|\Lambda_{\mathbf{x}}' M^{-1} \Lambda_{\mathbf{x}}| = \prod_{i=1}^{k} \frac{(n_i R_{xx_i} + x_{i.}^2)}{n_i(\beta_i + \lambda_e)} \frac{n_i^2 R_{xx_i}}{(\beta_i + \lambda_e)(n_i R_{xx_i} + x_{i.}^2)} = \prod_{i=1}^{k} \frac{n_i R_{xx_i}}{(\beta_i + \lambda_e)^2}$$

Portanto, usando esse resultado em (3.7) temos

$$\Pi(\mathbf{x}|Z) \;\propto\; [(X - \mathbf{x})'(X - \mathbf{x})s^2]^{-n/2} \pi(\mathbf{x}) \tilde{s}^{2k} \prod_{i=1}^{k} \frac{(\beta_i + \lambda_e)}{(n_i R_{xx_i})^{1/2}},$$

como queríamos demonstrar.

Outros resultados podem ser obtidos ao supormos que tanto a verossimilhança de Z como o vetor $\mathbf{x}$ pertencem a classe das distribuições normais bem como supondo que o vetor $W' = (Y', X', x')$ é completamente normal. Resultados mais gerais podem ser obtidos supondo que os erros pertencem a uma classe mais geral de distribuições esféricas, por exemplo, as distribuições elípticas.

Apêndice A

Resultados da Distribuição Normal Multivariada

Neste apêndice enunciaremos alguns resultados que foram utilizados no capítulo 2 deste trabalho. Esses resultados são encontrados nas referências [3], [4] e [6].

Definição A.0.1 *Um vetor aleatório* $\mathbf{X}$ *de dimensão* $m \times 1$ *é dito ter distribuição normal m-variada se, e somente se, toda função linear de* $\mathbf{X}$ *tem distribuição normal univariada, ou seja, se* $\forall\, \boldsymbol{\alpha} \in \mathbb{R}^m$ *tivermos* $\boldsymbol{\alpha}'\mathbf{X}$ *normal univariada.*

Teorema A.1 *Se* $\mathbf{X}$ *é um vetor aleatório m-dimensional, então sua distribuição é unicamente determinada pela distribuição de* $\boldsymbol{\alpha}'\mathbf{X}$, $\forall\, \boldsymbol{\alpha} \in \mathbb{R}^m$.

Teorema A.2 *Seja* $\mathbf{X}$ *um vetor aleatório m-dimensional com distribuição normal m-variada. Então,* $\boldsymbol{\mu} = E\mathbf{X}$ *e* $\Sigma = Cov\mathbf{X}$ *existem e determinam a distribuição de* $\mathbf{X}$.

Lema A.3 *Se* $\mathbf{X} \sim N_m(\mu, \Sigma)$, *então a função característica de X,* φ_X *é dada por:*

$$\varphi_X(t) = exp\{i\mu' t - \frac{1}{2}t'\Sigma t\},\ \forall\, t\ \in \mathbb{R}^m.$$

Teorema A.4 *Se* $\mathbf{X} \sim N_m(\mu, \Sigma)$, *onde* Σ *é uma matriz positiva definida, então a densidade do vetor aleatório X é dada por*

$$(2\pi)^{-m/2}|\Sigma|^{-1/2} \exp\left\{-\frac{1}{2}(x-\mu)'\Sigma^{-1}(x-\mu)\right\}.$$

Teorema A.5 *Sejam* $\mathbf{X} = (X_1, \ldots, X_m)' \sim N_m(\mu, \Sigma)$, *B e b matrizes de ordem* $k \times m$ *e* $k \times 1$, *respectivamente. Então,* $Y = BX + b \sim N_k(B\mu + b, B\Sigma B')$.

Teorema A.6 *Se* $\mathbf{X} \sim N_m(\mathbf{0}, \Sigma)$, *onde* Σ *é positiva definida de posto* m, *então a forma quadrática* $U = X'\Sigma^{-1}X$ *é distribuída como uma* $\chi^2_{(m)}$.

Teorema A.7 *Se* $\mathbf{X} \sim N_m(\mu, \Sigma)$, *onde* Σ *tem posto* m. *Então a forma quadrática* $U = X'AX$ *é distribuído como uma* $\chi^2_{(p,\lambda)}$, *onde* $\lambda = \frac{1}{2}\boldsymbol{\mu}'A\boldsymbol{\mu}$, *se, e somente se, qualquer uma das três condições abaixo for satisfeita:*

(i) $A\Sigma$ *é uma matriz idempotente de posto* p*;*

(ii) ΣA *é uma matriz idempotente de posto* p*;*

(iii) Σ *é inversa condicional de* A, *isto é,* $A\Sigma A = A$, *e* A *tem posto* p.

Teorema A.8 *Se* $\mathbf{X} \sim N_m(\mu, \Sigma)$, *onde* Σ *é uma matriz positiva definida e tem posto* m. *Se* $B\Sigma A = \mathbf{0}$, *então a forma quadrática* $Y'AY$ *é independente da forma linear* BY, *onde* B *é uma matriz de ordem* $q \times m$.

Lema A.9 *Se as matrizes* A *e* D *são simétricas e existe* B^{-1}, *então*

$$\begin{bmatrix} A & B \\ B' & D \end{bmatrix}^{-1} = \begin{bmatrix} A^{-1} + FE^{-1}F' & -FE^{-1} \\ -E^{-1}F' & E^{-1} \end{bmatrix}, \text{ com } \begin{cases} E = D - B'A^{-1}B \\ F = A^{-1}B \end{cases}.$$

Lema A.10 *Seja* B *uma matriz de ordem* $n \times n$ *que é particionada como segue*

$$B = \begin{bmatrix} B_{11} & B_{12} \\ B_{21} & B_{22} \end{bmatrix},$$

onde cada B_{ij} *é de ordem* $n_i \times n_j$, $i, j = 1, 2$ *e com* $n_1 + n_2 = n$. *Se a matriz* B_{22} *é não singular, então o determinante de* B *pode ser calculado por*

$$|B| = |B_{22}||B_{11} - B_{12}B_{22}^{-1}B_{21}|.$$

Lema A.11 *Seja* C *uma matriz na forma*

$$\begin{bmatrix} \frac{1}{a_1} + \frac{1}{a_2} & -\frac{1}{a_2} & 0 & \cdots & 0 & 0 \\ -\frac{1}{a_2} & \frac{1}{a_2} + \frac{1}{a_3} & \cdots & 0 & 0 & \\ 0 & -\frac{1}{a_3} & \frac{1}{a_3} + \frac{1}{a_4} & \cdots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \cdots & \frac{1}{a_{k-2}} + \frac{1}{a_{k-1}} & -\frac{1}{a_{k-1}} \\ 0 & 0 & 0 & \cdots & -\frac{1}{a_{k-1}} & \frac{1}{a_{k-1}} + \frac{1}{a_k} \end{bmatrix},$$

então C^{-1} é dada por

$$\begin{bmatrix} \frac{a_1(a_2+\cdots+a_k)}{a_1+\cdots+a_k} & \frac{a_1(a_3+\cdots+a_k)}{a_1+\cdots+a_k} & \cdots & \frac{a_1 a_k}{a_1+\cdots+a_k} \\ \frac{a_1(a_3+\cdots+a_k)}{a_1+\cdots+a_k} & \frac{(a_1+a_2)(a_3+\cdots+a_k)}{a_1+\cdots+a_k} & \cdots & \frac{(a_1+a_2)a_k}{a_1+\cdots+a_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{a_1 a_k}{a_1+\cdots+a_k} & \frac{(a_1+a_2)a_k}{a_1+\cdots+a_k} & \cdots & \frac{(a_1+\cdots+a_{k-1})a_k}{a_1+\cdots+a_k} \end{bmatrix}.$$

Prova. Para verificarmos esse resultado, basta multiplicarmos as duas matrizes e o resultado segue.

Apêndice B

Propriedades da t-Student Multivariada

A distribuição t-Student multivariada possui propriedades que nos ajudarão a fazer inferências sobre os parâmetros $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$. As propriedades abaixo são consideradas para o vetor aleatório $\boldsymbol{\theta}$, mas seus resultados são análogos para a distribuição t-Student k-variada, como é o caso dos vetores aleatórios $\boldsymbol{\tau}$ e $\boldsymbol{\beta}$, conforme verificaremos na propriedade **P.3** e para combinações lineares desses parâmetros, conforme veremos na proriedade **P.4**.

Vejamos antes a definição da distribuição t-Student multivariada.

Definição B.0.2 *Um vetor aleatório $\boldsymbol{\theta}$ de dimensão $2k \times 1$ é dito ter distribuição a posteriori t-Student $2k$-variada com parâmetro de locação $\hat{\boldsymbol{\theta}}$, de escala $s^2(X'X)^{-1}$ e com ν graus de liberdade, se sua densidade é dada por*

$$p\,(\boldsymbol{\theta}|\mathbf{y}) = \frac{\Gamma((\nu+2k)/2)|X'X|^{1/2}s^{-2k}}{(\Gamma(1/2))^{2k}\Gamma(\nu/2)\nu^k}\left[1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right]^{-\frac{\nu+2k}{2}}.$$

Notação: $\boldsymbol{\theta}|\mathbf{y} \sim t_{2k}\big(\hat{\boldsymbol{\theta}}, s^2(X'X)^{-1}, \nu\big)$.

Consideremos agora que $\boldsymbol{\theta}|\mathbf{y}$ é distribuído segundo uma $t_{2k}\big(\hat{\boldsymbol{\theta}}, s^2(X'X)^{-1}, \nu\big)$.

<u>Propriedade **P.1**</u> A função densidade de probabilidade a *posteriori* de $\boldsymbol{\theta}$ é monotonicamente decrescente da forma quadrática

$$Q(\boldsymbol{\theta}) = (\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'X'X(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}),$$

pois

$$\frac{\Gamma((\nu+2k)/2)|X'X|^{1/2}s^{-2k}}{(\Gamma(1/2))^{2k}\Gamma(\nu/2)\nu^k} > 0$$

e a função

$$\left[1+(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\nu s^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})\right]^{-\frac{\nu+2k}{2}}$$

tem expoente negativo, já que $\nu+2k=n>0$ e é, portanto, monotonicamente decrescente.

Quando fizermos $Q(\boldsymbol{\theta})=c$, $c>0$ estaremos definindo os contornos da distribuição no espaço $2k$-dimensional de $\boldsymbol{\theta}$ onde todo ponto interior a esse contorno tem uma densidade a *posteriori* maior do que qualquer ponto fora do contorno. Particularmente, quando $k=1$, estaremos trabalhando no espaço bi-dimensional e $Q(\boldsymbol{\theta})=c$, $c>0$ define os contornos de uma elipse com centro em $\hat{\boldsymbol{\theta}}=(\hat{\tau}_1,\hat{\beta}_1)$.

Propriedade **P.2** Observemos que:

a. podemos encontrar a distribuição de

$$\frac{Q(\boldsymbol{\theta})}{\sigma^2}=(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})$$

já que

$$\begin{aligned}
E(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) &= E(\boldsymbol{\theta})-E((X'X)^{-1}X'Y)\\
&= \boldsymbol{\theta}-(X'X)^{-1}X'E(Y)\\
&= \boldsymbol{\theta}-(X'X)^{-1}X'(X\boldsymbol{\theta})\\
&= \boldsymbol{\theta}-\boldsymbol{\theta}=\mathbf{0},\\
Cov(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}}) &= Cov(\hat{\boldsymbol{\theta}})=Cov((X'X)^{-1}X'Y)\\
&= (X'X)^{-1}X'Cov(Y)((X'X)^{-1}X')'\\
&= (X'X)^{-1}X'(\sigma^2I)X'(X'X)^{-1}\\
&= \sigma^2(X'X)^{-1},\text{ e}\\
\lambda &:= \frac{1}{2}[E(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})]'Cov(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})[E(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})]\\
&= \frac{1}{2}\mathbf{0}'\sigma^{-2}(X'X)\mathbf{0}=0.
\end{aligned}$$

Então, pelo Teorema A.6 concluimos que

$$(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})'\frac{X'X}{\sigma^2}(\boldsymbol{\theta}-\hat{\boldsymbol{\theta}})=\frac{Q(\boldsymbol{\theta})}{\sigma^2}\sim\chi^2_{2k};$$

b. pelo resultado (2.5) temos que

$$\frac{\nu s^2}{\sigma^2} = \frac{n\hat{\sigma}^2}{\sigma^2} \sim \chi^2_\nu;$$

c. no Teorema A.8 fazendo $B = (X'X)^{-1}X'$, $A = (I - X(X'X)^{-1}X')\sigma^{-2}$ e vendo que

$$\begin{aligned} B\,Cov(Y)A &= (X'X)^{-1}X'(\sigma^2 I)(I - X(X'X)^{-1}X')\sigma^{-2} \\ &= (X'X)^{-1}(X' - X'X(X'X)^{-1}X') \\ &= (X'X)^{-1}(X' - X') = 0 \end{aligned}$$

concluimos que $\hat{\boldsymbol{\theta}} = [(X'X)^{-1}X']Y = BY$ independe de $\dfrac{\nu s^2}{\sigma^2} = Y'AY$. Dessa forma, $\dfrac{\nu s^2}{\sigma^2}$ é independente de $S(\boldsymbol{\theta}) = (\boldsymbol{\theta} - \hat{\boldsymbol{\theta}})'\dfrac{X'X}{\sigma^2}(\boldsymbol{\theta} - \hat{\boldsymbol{\theta}}) = \dfrac{Q(\boldsymbol{\theta})}{\sigma^2}$.

Agora usemos a definição da distribuição F-Snedecor e as observações destacadas em a, b e c para concluirmos que

$$\frac{\left(\frac{Q(\boldsymbol{\theta})}{\sigma^2}\right)/2k}{\left(\frac{\nu s^2}{\sigma^2}\right)/\nu} \sim F_{(2k,\nu)}.$$

Mas, simplificando o quociente acima teremos

$$\frac{\left(\frac{Q(\boldsymbol{\theta})}{\sigma^2}\right)/2k}{\left(\frac{\nu s^2}{\sigma^2}\right)/\nu} = \frac{Q(\boldsymbol{\theta})}{2ks^2},$$

resultando, portanto que

$$\frac{Q(\boldsymbol{\theta})}{2ks^2} \sim F_{(2k,\nu)}.$$

Os contornos elipsoidais de $p\,(\boldsymbol{\theta}|y)$ serão definidos por $\dfrac{Q(\boldsymbol{\theta})}{2ks^2} = F_{(2k,\nu,\alpha)}$, onde $F_{(2k,\nu,\alpha)}$ é o ponto da distribuição $F_{(2k,\nu)}$ tal que $P\{F_{(2k,\nu)} \leq F_{(2k,\nu,\alpha)}\} = 1 - \alpha$.

Propriedade **P.3** Consideremos as partições

$$\boldsymbol{\theta} = \begin{bmatrix} \boldsymbol{\theta}_1 \\ \boldsymbol{\theta}_2 \end{bmatrix}, \quad \hat{\boldsymbol{\theta}} = \begin{bmatrix} \hat{\boldsymbol{\theta}}_1 \\ \hat{\boldsymbol{\theta}}_2 \end{bmatrix}, \quad X'X = \begin{bmatrix} X_1'X_1 & X_1'X_2 \\ X_2'X_1 & X_2'X_2 \end{bmatrix} \quad e \quad (X'X)^{-1} = \begin{bmatrix} D_{11} & D_{12} \\ D_{21} & D_{22} \end{bmatrix}$$

onde $\boldsymbol{\theta}_1$, $\hat{\boldsymbol{\theta}}_1$ têm dimensão $r \times 1$; $X_1'X_1$, D_{11} têm dimensão $r \times r$; $\boldsymbol{\theta}_2$, $\hat{\boldsymbol{\theta}}_2$ têm dimensão $(2k - r) \times 1$; e $X_2'X_2$, D_{22} têm dimensão $(2k - r) \times (2k - r)$.

Se $\boldsymbol{\theta}|y \sim t_{2k}\Big(\hat{\boldsymbol{\theta}}, s^2(X'X)^{-1}, \nu\Big)$, então

$$\boldsymbol{\theta}_1|y \sim t_r\Big(\hat{\boldsymbol{\theta}}_1, s^2 D_{11}, \nu\Big).$$

Propriedade **P.4** Suponha que $\boldsymbol{\phi} = (\phi_1, \cdots, \phi_m)', m < 2k$, seja um conjunto de combinações lineares tais que

$$\boldsymbol{\phi} = C\boldsymbol{\theta}$$

onde C é uma matriz de ordem $m \times 2k$ de posto m. Então,

$$\boldsymbol{\phi}|y \sim t_m\Big(C\hat{\boldsymbol{\theta}}, s^2 C(X'X)^{-1}C', \nu\Big).$$

Para vericarmos a validade dessa afirmação consideremos

$$\boldsymbol{\phi}_o = (\phi_1, \cdots, \phi_m, \phi_{m+1}, \dots, \phi_{2k})' = (\boldsymbol{\phi}', \phi_{m+1}, \dots, \phi_{2k})'$$

um conjunto de combinações lineares de $\boldsymbol{\theta}$ tais que

$$\boldsymbol{\phi}_o = C_o\boldsymbol{\theta},$$

onde $\phi_{m+j} = \theta_j,\ \forall j = 1, \dots, p$, com $m + p = 2k$. A matriz C_o é particionada como

$$C_o = \begin{bmatrix} C_{m\times 2k} \\ C^*_{(2k-m)\times 2k} \end{bmatrix}_{2k\times 2k}$$

onde C^* é de tal forma que a matriz C_o tem posto $2k$.

Daí,

$$\boldsymbol{\phi}_o = C_o\boldsymbol{\theta} = \begin{bmatrix} C\boldsymbol{\theta} \\ C^*\boldsymbol{\theta} \end{bmatrix} = \begin{bmatrix} \boldsymbol{\phi} \\ C^*\boldsymbol{\theta} \end{bmatrix}.$$

Sendo C_o de posto completo, portanto inversível, façamos

$$\boldsymbol{\theta} = C_o^{-1}\boldsymbol{\phi}_o = B\boldsymbol{\phi}_o,$$

com $B = (b_{ij})_{2k\times 2k} = C_o^{-1}$.

Encontramos a densidade da combinação linear $\boldsymbol{\phi}_o$ de $\boldsymbol{\theta}$ usando:

$$f_{\boldsymbol{\phi}_o}(\phi_1, \cdots, \phi_m, \phi_{m+1}, \dots, \phi_{2k}) = f_{\boldsymbol{\theta}}(B\boldsymbol{\phi}_o)|J(\boldsymbol{\theta}, \boldsymbol{\phi}_o)|.$$

A matriz jacobiana de $(\boldsymbol{\theta}, \boldsymbol{\phi}_o)$ é

$$J(\boldsymbol{\theta}, \boldsymbol{\phi}_o) = \begin{bmatrix} \frac{\partial\theta_1}{\partial\phi_1} & \cdots & \frac{\partial\theta_1}{\partial\phi_m} & \frac{\partial\theta_1}{\partial\phi_{m+1}} & \cdots & \frac{\partial\theta_1}{\partial\phi_{2k}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\theta_m}{\partial\phi_1} & \cdots & \frac{\partial\theta_m}{\partial\phi_m} & \frac{\partial\theta_m}{\partial\phi_{m+1}} & \cdots & \frac{\partial\theta_m}{\partial\phi_{2k}} \\ \frac{\partial\theta_{m+1}}{\partial\phi_1} & \cdots & \frac{\partial\theta_{m+1}}{\partial\phi_m} & \frac{\partial\theta_{m+1}}{\partial\phi_{m+1}} & \cdots & \frac{\partial\theta_{m+1}}{\partial\phi_{2k}} \\ \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ \frac{\partial\theta_{2k}}{\partial\phi_1} & \cdots & \frac{\partial\theta_{2k}}{\partial\phi_m} & \frac{\partial\theta_{2k}}{\partial\phi_{m+1}} & \cdots & \frac{\partial\theta_{2k}}{\partial\phi_{2k}} \end{bmatrix},$$

e, já que $\boldsymbol{\theta} = B\boldsymbol{\phi}_o = \begin{bmatrix} \sum_{i=1}^{2k} b_{1i}\phi_i \\ \vdots \\ \sum_{i=1}^{2k} b_{(2k)i}\phi_i \end{bmatrix}$ então, a matriz jacobiana é dada por

$$J(\boldsymbol{\theta}, \boldsymbol{\phi}_o) = \begin{bmatrix} b_{11} & b_{12} & \cdots & b_{1(2k)} \\ b_{21} & b_{22} & \cdots & b_{2(2k)} \\ \vdots & \vdots & \ddots & \vdots \\ b_{(2k)1} & b_{(2k)2} & \cdots & b_{(2k)(2k)} \end{bmatrix} = B.$$

Note também que

$$|B'||B| = |B|^2 \implies 0 < |B| = |B'|^{1/2}|B|^{1/2}.$$

Daí,

$$\begin{aligned} f_{\boldsymbol{\phi}_o}(\phi_1, \ldots, \phi_{2k}) &= \frac{\Gamma(\frac{\nu+2k}{2})|X'X|^{1/2}s^{-2k}}{(\Gamma(1/2))^{2k}\Gamma(\nu/2)\nu^k}\left[1 + (B\boldsymbol{\phi} - B\hat{\boldsymbol{\phi}})'\frac{X'X}{\nu s^2}(B\boldsymbol{\phi} - B\hat{\boldsymbol{\phi}})\right]^{-\frac{\nu+2k}{2}}|B'|^{\frac{1}{2}}|B|^{\frac{1}{2}} \\ &= \frac{\Gamma(\frac{\nu+2k}{2})|B'(X'X)B|^{1/2}s^{-2k}}{(\Gamma(1/2))^{2k}\Gamma(\nu/2)\nu^k}\left[1 + (\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})'\frac{B'X'XB}{\nu s^2}(\boldsymbol{\phi} - \hat{\boldsymbol{\phi}})\right]^{-\frac{\nu+2k}{2}} \end{aligned}$$

$$\therefore \quad \boldsymbol{\phi}_o \sim t_{2k}[\hat{\boldsymbol{\phi}}, C_o(X'X)^{-1}C_o's^2, \nu],$$

pois $(B'X'XB)^{-1} = B^{-1}(X'X)^{-1}(B')^{-1} = C_o(X'X)^{-1}C_o'$.

Fazendo uso da partição

$$C_o(X'X)^{-1}C_o' = \begin{bmatrix} C(X'X)^{-1}C' & C(X'X)^{-1}(C^*)' \\ C^*(X'X)^{-1}C' & C^*(X'X)^{-1}(C^*)' \end{bmatrix}$$

e da propriedade **P.3** concluimos que

$$\boldsymbol{\phi}|y \sim t_m\Big(C\hat{\boldsymbol{\theta}}, s^2C(X'X)^{-1}C', \nu\Big).$$

Bibliografia

[1] Box, G. E. P.; Tiao, G. C. *Bayesian Inference in Statistical Analysis*, Cap. 2. New York: Addison Wesley, 1973.

[2] Gamerman, Dani; Migon, Hélio dos Santos. *Curso de Inferência Estatística*, Caps. 2 e 3, 2001.

[3] Graybill, Franklin A. *Matrices with Application in Statistical*, Cap. 8. Califórnia, 1978.

[4] Graybill, Franklin A. *Theory and Application of the Linear Model*, Caps. 4 e 6, Wadsworth & Brooks Cole Advanced Books & Software, Califórnia, 1976.

[5] Guenther, William C. *Analise of Variance*, Prentice-Hall.

[6] Johnson, Richard A. *Applied Multivariate Statistical Analysis*, Cap 1, 1998.

[7] Luna, João Gil de; Silva, Antonio José da *Modelo de Covariância Linear Normal univariada Com Priori Difusa*, Notas de Matemática N^o 22, 1998.

[8] Pereira, Carlos Alberto de Bragança; Viana, Marlos Augusto Gomes. *Elementos de Inferência Bayesiana*, 5^o SINAPE - Simpósio Nacional de Probabilidade e Estatística, Cap. 1, 1946.

[9] Silva, Antonio José da. *Sobre Condições de Estendibilidade, Simetria Esférica e Inferência Bayesiana em Modelos Elípticos Com Erros nas Variáveis*, Cap. 3. Tese de Doutorado. São Paulo, 1997.

[10] Zellner, A. *An Introduction to Bayesian Inference in Econometrics*, New York: John Wiley, 1971.