UNIVERSIDADE DE SÃO PAULO
Instituto de Ciências Matemáticas de São Carlos

**INFERÊNCIA BAYESIANA
PARA REGRESSÃO LINEAR
SIMPLES COM ERRO NAS
VARIÁVEIS**

MARTA YUKIE BABA

Orientação: Prof. Dr. Josemar Rodrigues

Dissertação apresentada ao Instituto de Ciências Matemáticas de São Carlos, da Universidade de São Paulo, como parte dos requisitos para obtenção do título de mestre em Ciência - Área: Ciências de Computação e Matemática Computacional.

São Carlos

1994

Aos meus pais, Julio e Yoshiko,
que não mediram esforços para que
eu sempre buscasse meu horizonte.

Ao Anderson e Julinho,
pelo infinito amor (maior que a distância
daqui até o papai do céu) e, principalmente,
pela compreensão da minha ausência.

Aos meus irmãos, Alice e Amauri,
pelo incentivo e por sempre acreditarem em mim.

A Sra. Cecília,
pela sua constante força espiritual.

AGRADECIMENTOS

Ao Prof. Dr. Josemar Rodrigues, que não se limitou a oferecer somente a sua orientação para este trabalho, colaborando na realização do mesmo e incentivando a nossa formação em Estatística; atuando muito mais como amigo do que como mestre.

Ao Fernando, pelo amparo nesta trajetória, enfrentando as dificuldades, auxiliando para que eu nunca desistisse e atingisse minha meta.

Ao Prof. Dr. Heleno Bolfarine, ao Prof. Dr. Jorge Alberto Achcar e ao Prof. Dr. Emanuel P. Barbosa, pelos seus valiosos comentários e sugestões na participação do exame de qualificação.

Ao Prof. Silvio José Govone, que orientou e incentivou minhas primeiras pesquisas na área de Estatística.

Ao Carlos, Rosana, Claudia, Gaby, Vitória e a todos os colegas do curso de Pós-graduação pelo convívio amigo.

Aos professores e funcionários da Pós-graduação, do ICMSC, pela sua dedicação profissional e colaboração.

Ao Prof. Dr. Pedro Nowosad pela sua assistência ampla e irrestrita, e ao Aparecido Doniseti da Costa pela sua dedicada orientação na área computacional.

À Cristina e Bernadete, funcionárias da Biblioteca Central do Campus de Ribeirão Preto, que gentilmente esclareceram a utilização das normas que regem a apresentação de referências bibliográficas.

Aos amigos e funcionários da Faculdade de Filosofia Ciências e Letras de Ribeirão Preto, em especial do Departamento de Geologia, Física e Matemática, que propiciaram um ambiente agradável.

À CAPES pelos auxílios concedidos.

Enfim, meus sinceros agradecimentos a todos que direta ou indiretamente colaboraram na execução deste trabalho.

RESUMO

O objetivo deste trabalho é realizar uma Inferência Bayesiana para o coeficiente de inclinação do modelo de regressão linear simples com erros nas variáveis. Em consequência dos erros de mensuração os estimadores não-Bayesianos dos parâmetros envolvidos são viciados. O modelo com erros nas variáveis pode ser classificado de duas formas distintas: Modelo Funcional e Modelo Estrutural. Para o Modelo Funcional estudamos o parâmetro de inclinação sujeito a priori não-informativa e colocamos seu estimador de mínimos quadrados sob o modelo funcional com réplicas. Para o modelo estrutural analisamos o parâmetro de inclinação utilizando primeiramente a priori não-informativa e depois é submetido a uma priori informativa seguindo a estrutura hierárquica que Lindley propôs em um artigo publicado em 1972. Utilizando esta estrutura hierárquica, a influência da priori sobre a média a posteriori é analisada no sentido de Robustez introduzido por Pericchi & Smith (1992). O método de Laplace é usado como uma alternativa para aproximar a densidade a posteriori do parâmetro de interesse quando não conseguimos determinar explicitamente suas soluções analíticas.

ABSTRACT

This work deals with Bayesian Inference for the slope coefficient of the simple linear regression model error in the variables. Due to measurement errors the non-Bayesian parameter estimators are biased. The model with error in the variables can be classified in two distinct ways: as the Functional Model or as the Structural Model. For the first one we study the slope parameter under noninformative prior analyzing its least square estimator via the functional model applied repeatedly. For the structural model we analyze the slope parameter using firstly a noninformative prior and then an informative one according to the hierarchical structure proposed by Lindley in 1972. Using this structure, the effect of the prior on the posterior average is analyzed via the robustness concept introduced by Pericchi & Smith (1992). The Laplace's method is used as an alternative to approximate the posterior density of the slope parameter when analytic solutions are not explicitly known.

SUMÁRIO

INTRODUÇÃO	01
1. FUNDAMENTOS	04
1.1 Primeiros Desenvolvimentos	07
1.2 Modelo Funcional e Modelo Estrutural	08
1.3 Um modelo Particular	09
1.4 O Modelo de Regressão Linear Usual	10
2. INFERÊNCIA BAYESIANA NO MODELO FUNCIONAL	11
2.1 Estimador de Mínimos Quadrados Generalizados	11
2.2 Inferência Bayesiana para β	17
2.3 Modelo Funcional com Réplicas	20
3. INFERÊNCIA BAYESIANA PARA O MODELO ESTRUTURAL	25
3.1 Fator de Atenuação Conhecido	27
3.1.1 Intervalo HPD para β	30
3.1.2 Exemplo Numérico	31
3.2 Fator de Atenuação Desconhecido	35
3.2.1 Exemplo Numérico	39

4. ESTRUTURA HIERÁRQUICA PARA A PRIORI NO MODELO

ESTRUTURAL	41
4.1 Introdução.....	41
4.1.1 Função de Verossimilhança	43
4.2 Inferência Bayesiana com Priori Normal.....	43
4.3 Inferência Bayesiana com Priori t de Student.....	48
4.4 Exemplo Numérico	55
5. Propostas Futuras	61
Bibliografia	63

APÊNDICE A Ortogonalidade de Parâmetros

APÊNDICE B Método de Laplace para Aproximação de Integrais

APÊNDICE C Região Highest Posterior Density

APÊNDICE D Programas Utilizados

INTRODUÇÃO

A análise de Regressão é um dos métodos estatísticos que trata da formulação de modelos que descrevem as relações entre as variáveis e os parâmetros populacionais. Estes modelos são utilizados com o propósito de predição, controle estatístico e outras inferências estatísticas.

Tem sido reconhecido que as variáveis muitas vezes contêm erros de medida e que a presença desses erros de medida pode afetar essencialmente os resultados da análise. Atualmente, os modelos com erros de mensuração são alvos para muitos estudos nesta área, pois este modelo pode ser visto como uma generalização do modelo de regressão linear usual.

Um estudo sério de métodos estatísticos baseados nas variáveis sujeitas a erros de mensuração começou pouco mais do que um século atrás e vagarosos progressos foram feitos antes da Segunda Guerra Mundial. Nestes últimos anos, esta área vem atraindo a atenção dos estatísticos clássicos e Bayesianos através de importantes trabalhos como Fuller (1987), Lindley & Sayad (1968), etc. Neste trabalho nos restringimos a analisar os parâmetros de interesse sob a abordagem da Inferência Bayesiana de um modelo particular com erros nas variáveis.

No Capítulo 1 definimos o modelo de regressão linear com erro nas variáveis e colocamos, logo a seguir, os primeiros desenvolvimentos na história deste modelo onde é mencionada a distinção entre a relação Funcional e Estrutural que foi formalizado por Maurice G. Kendall em 1951 conduzindo, assim, para um melhor entendimento de muitos

dos problemas subordinados. Ao final do Capítulo colocamos o modelo de regressão linear clássica e seu estimador de mínimos quadrados para que possam ser realizadas futuras comparações.

Vale ressaltar que, em qualquer parte deste trabalho, o objetivo principal é analisar sob o ponto de vista Bayesiano o parâmetro de inclinação de um modelo de Regressão Linear Simples com Erros de Mensuração.

No Capítulo 2, todo o estudo ficará sujeito ao modelo Funcional. Analisaremos, sob o ponto de vista da Estatística clássica, um estimador proposto por Sprent (1966), colocando quais as vantagens e sob quais suposições podemos utilizá-lo. Sequencialmente, utilizamos o estudo de Rodrigues (1990), onde determina uma densidade a priori não-informativa justificando os resultados obtidos por Sprent. Abordamos, também, um caso particular do modelo funcional que é o modelo funcional com réplicas.

No Capítulo 3, utilizando uma parametrização ortogonal no modelo de regressão linear estrutural condicional, com priori localmente uniforme, determinamos a posteriori do parâmetro de inclinação. No Capítulo 4 abordaremos o modelo estrutural analisando a densidade a posteriori sujeita a uma priori informativa seguindo a estrutura hierárquica que Lindley coloca em 1972. A análise de sensibilidade da densidade a priori é considerada, seguindo a definição de Robustez de Pericchi & Smith (1992).

Ao final do trabalho estão colocados quatro Apêndices que tratam de métodos, teoria e programas que foram utilizados no decorrer deste trabalho. No Apêndice A é colocada a definição de ortogonalidade e um método, dado por Cox & Reid (1987), para reparametrizar os parâmetros de tal forma a conseguir, sob certas condições, a ortogonalidade dos parâmetros. No Apêndice B descrevemos o método de Laplace para aproximação de integrais que é utilizado em nosso trabalho para determinar a densidade a posteriori através de uma expressão explícita. No Apêndice C formalizamos a definição da região Highest Posterior Density que é empregada para fornecer o estimador intervalar do parâmetro de interesse. No Apêndice D são colocados os principais programas empregados

para a execução de simulações, gráficos e aproximações; usufruindo os recursos computacionais de softwares tais como MINITAB, EUREKA, DERIVE e MAPLE.

Capítulo 1

FUNDAMENTOS

Em muitas áreas de investigação científica, a variável da unidade experimental pode estar relacionada com outras variáveis que poderão estar sujeitas a erro de mensuração.

O conceito de relação entre variáveis pode ser descrito da seguinte forma:

- (1) Relação funcional - uma relação funcional entre variáveis é expressa por uma função matemática precisa. Por exemplo, se X é a variável independente e Y é a variável dependente, a relação funcional é da forma $Y = f(X)$, isto é, para um valor particular de X , a função f indica exatamente o valor correspondente de Y .
- (2) Relação Estatística - uma relação estatística é uma relação funcional imprecisa, isto é, existe um componente aleatório que torna a relação funcional não exata como no caso da relação funcional, descrita acima.

Um modelo de Regressão é definido como uma relação estatística entre a variável dependente (resposta) e as variáveis independentes (explicativas) envolvendo dois componentes essenciais:

- i) componente funcional: $f(X)$;
- ii) componente aleatório.

Suponhamos que o interesse esteja somente na relação linear de duas variáveis, definida da seguinte forma:

$$y_i = \alpha + \beta x_i \quad i = 1, 2, \dots, n.$$

Se considerarmos que estas variáveis, y_i e x_i , são os verdadeiros valores e que os valores observáveis estão sujeitos a erros de mensuração, isto é,

$$\begin{aligned} Y_i &= y_i + e_i \\ X_i &= x_i + u_i \quad i = 1, 2, \dots, n \end{aligned}$$

então, condensando estas considerações acima, podemos definir um modelo, que denominamos modelo de regressão linear simples com erros nas variáveis, da seguinte forma:

$$\begin{aligned} Y_i &= y_i + e_i \\ X_i &= x_i + u_i \\ y_i &= \alpha + \beta x_i \quad i = 1, 2, \dots, n \end{aligned} \tag{1.1}$$

onde os valores y_i e x_i não podem ser observados diretamente, ou seja, eles são os verdadeiros valores das variáveis, e e_i e u_i é o erro cometido ao observar Y_i e X_i , respectivamente.

Ao tratarmos com os modelos de regressão algumas hipóteses sobre a regularidade das populações devem ser levadas em consideração:

para todo $i = 1, 2, \dots, n$;

i) $E(e_i) = 0$ e $E(u_i) = 0$;

ii) $Var(e_i) = \sigma_e^2$ e $Var(u_i) = \sigma_u^2$;

iii) $Cov(e_i, u_i) = \sigma_{eu}^2$;

iv) para todo $i \neq j$, Y_i e Y_j são independentes e, também, X_i e X_j são independentes.

Definimos Σ como a matriz de covariâncias dos erros, de tamanho 2×2 , da seguinte forma:

$$\Sigma = \begin{bmatrix} \sigma_e^2 & \sigma_{eu}^2 \\ \sigma_{eu}^2 & \sigma_u^2 \end{bmatrix} \tag{1.3}$$

Usualmente a teoria de análise de regressão linear clássica assume que a variável independente (explicativa) não contém erro de mensuração, por exemplo, quando a variável independente é o preço de um produto em um país de economia estável, número de variáveis em um problema de otimização, sexo, idade (sem restrição de faixa etária), etc. Assim, quando não consideramos erro na variável independente, isto é, $X_i = x_i$, o modelo (1.1) reduz-se a:

$$Y_i = \alpha + \beta x_i + e_i \quad i = 1, 2, \dots, n. \quad (1.4)$$

Note que o modelo de regressão linear com erros nas variáveis é um modelo mais genérico, tornando o modelo de regressão linear usual, simplesmente, como um caso particular.

Na prática, porém, tem sido reconhecido que os dados, muitas vezes, são contaminados com erros de mensuração e que a presença desses erros pode afetar drasticamente os resultados da análise. Os erros de mensuração podem introduzir-se no valor observado para a variável independente, por exemplo, quando é a posição social, taxa de desemprego, velocidade da linha de produção, teste de cognição (teste de Q.I.), etc. Usando uma ilustração, suponha que os operários estão voltando de um trabalho por empreitada e que o salário é pago sobre um período pré-determinado. Se x_i denota o verdadeiro período do i -ésimo operário e X_i o período fornecido por esse operário em seu registro de trabalho então podemos, obviamente, deduzir que os dois nem sempre são o mesmo.

Esta é a filosofia do modelo com erros nas variáveis: “ *WHAT YOU SEE IS NOT WHAT YOU GET!* ”

1.1 Primeiros Desenvolvimentos

Em um trabalho feito em 1990, Sprent coloca algumas histórias sobre o desenvolvimento da relação funcional e estrutural.

Usualmente R.J. Adcock (1877, 1878) é considerado como a primeira pessoa a ter ponderado seriamente sobre o problema de ajustamento de uma relação linear quando ambas as variáveis estão sujeitas a erros de mensuração; embora Gauss, provavelmente, tivesse erros de medida em mente, quando propôs o método de mínimos quadrados. Em essência, Adcock obteve a solução de mínimos quadrados para um caso especial quando assumiu que $\sigma_e^2 = \sigma_u^2$.

Kummel (1879) estendeu o resultado para o caso onde a razão das variâncias dos erros é suposta ser conhecida. Pearson (1901), estendeu a solução de Adcock (varância dos erros iguais) para a situação de p -variáveis. Essas soluções para o caso bivariado ou p -variado, com erros independentes, identicamente e normalmente distribuídos com média zero, foi redescoberto por muitos pesquisadores nas décadas de 1920 e 1930.

Durante este período uma diversidade de terminologias foram utilizadas, tais como: Regressão com erros em x , modelos com erros nas variáveis, modelo de regressão com erros nas variáveis e modelos lineares com erros de medida. O termo modelos com erros de mensuração estão sempre em uso comum.

O ajustamento de uma relação linear com erros nas variáveis não é usual na teoria estatística e nem mesmo na prática, onde a suposição de normalidade, geralmente, não é de fácil comprovação. Logo após a II Guerra Mundial, os artigos publicados por Lindley (1947), Geary (1949), Reiersøl (1950), Neyman & Scott (1948, 1951), e Kiefer & Wolfowitz (1956) esclareceram muitos dos problemas de identificabilidade e consistência que persistiram no problema de estimação para um ou ambos os modelos que Kendall (1951, 1952) distinguiu como relação funcional ou estrutural, onde x_i é uma variável fixa ou aleatória, respectivamente.

O artigo de Lindley (1947) trata, principalmente de um problema de regressão onde existe uma relação linear entre os verdadeiros valores (não observáveis), isto é,

$$E(y|x) = \alpha + \beta x$$

e admitiu que esta linearidade permanece para os valores observados, ou seja,

$$E(Y|X) = \phi_0 + \phi_1 X.$$

O resultado de Lindley (1947) foi dado em termos de cumulantes e está indiretamente ligado com o trabalho de Geary (1949). Geary mostrou que, na ausência do conhecimento sobre a variância do erro, a estimação em uma relação estrutural, que será definida posteriormente, é possível somente se cumulantes de ordem maior do que dois não for zero. O método cumulante de Geary é análogo ao método de momentos, ele não pode ser usado para a distribuição Normal, porque cumulantes de ordem superior a dois desaparece.

Madansky (1959) comparou alguns métodos de estimação da inclinação de uma relação bivariada para um conjunto de dados reais, e o método de cumulantes foi o que pior se comportou entre todos aqueles considerados.

1.2 Modelo Funcional e Modelo Estrutural

A primeira distinção clara entre a relação funcional e estrutural foi feita por Kendall (1951, 1952), na tentativa de remover algumas obscuridades que apareciam a respeito da ligação entre regressão, relação funcional e equações estruturais em modelos teóricos contendo um elemento estocástico.

Baseado no modelo de regressão com erros nas variáveis, dado por (1.1), e utilizando a mesma nomenclatura que Kendall introduziu em 1951, vamos definir a seguir o modelo funcional e o modelo estrutural.

Definição 1.1 No MODELO FUNCIONAL considere o modelo dado por (1.1) onde as quantidades não-observáveis x_i são constantes ou variáveis matemáticas (sem propriedades específicas de distribuição).

Definição 1.2 No MODELO ESTRUTURAL considere o modelo dado por (1.1) onde as quantidades não-observáveis x_i são variáveis aleatórias e têm uma distribuição específica.

Em geral é difícil decidir se um modelo funcional ou estrutural é apropriado para um particular conjunto de dados, em outros pode ser muito claro que somente um modelo é aceitável, por exemplo, no teste de Q.I. o verdadeiro valor é uma variável aleatória, pois muda a cada indivíduo que é aplicado o teste, e, assim, é aceitável usarmos o modelo estrutural.

Os problemas que aparecem na estimação dos parâmetros tendem a manifestar-se de diferentes formas para os dois modelos; particularmente, a teoria assintótica é muito diferente para cada caso.

1.3 Um modelo Particular

Salientamos, novamente, que o nosso interesse está em estudar o parâmetro de inclinação β , e por este fato, ao fazermos uma simples transformação dos dados podemos analisar, sem perda de generalidade, o modelo de regressão linear com erro nas variáveis onde o parâmetro de intercepto é nulo, ou seja,

$$\begin{aligned} Y_i &= y_i + e_i \\ X_i &= x_i + u_i \\ y_i &= \beta x_i \end{aligned} \quad i = 1, 2, \dots, n \quad (1.5)$$

sobre o qual iremos considerar em toda análise deste trabalho, com as suposições consideradas acima.

1.4 O Modelo de Regressão Linear Usual

Para que as comparações, sobre o efeito dos erros de mensuração, possam ser analisadas sem que haja confusões de notações, considere o modelo de regressão simples usual, com o parâmetro de intercepto nulo:

$$Y_i = \beta X_i + w_i \quad i = 1, 2, \dots, n, \quad (1.6)$$

onde o estimador de mínimos quadrados ordinários, para o parâmetro de inclinação, é dado por:

$$\hat{\beta}^* = \frac{\sum_{i=1}^n X_i Y_i}{\sum_{i=1}^n X_i^2} \quad (1.7)$$

e as suposições consideradas para este modelo são:

- i) o erro w_i , para um específico X_i (fixo), tem média zero e variância constante, isto é, $E(e_i) = 0$ e $Var(e_i) = \sigma_e^2$, para todo $i = 1, 2, \dots, n$. Consequentemente, w_i e X_i são independentes o que não se verifica.
- ii) duas observações distintas Y_i e Y_j , para todo $i \neq j$, possuem erros não correlacionados, isto é, $Cov(e_i, e_j) = 0$.

Capítulo 2

INFERÊNCIA BAYESIANA NO MODELO FUNCIONAL

Suponhamos que os dados (Y_i, X_i) sejam provenientes do modelo funcional, definido no capítulo anterior, dado pelas equações:

$$\begin{aligned} Y_i &= y_i + e_i \\ X_i &= x_i + u_i \\ y_i &= \beta x_i \end{aligned} \quad i = 1, 2, \dots, n \quad (2.1)$$

onde o verdadeiro valor não-observável x_i é considerado fixo. Considere as suposições (1.2) para o modelo com erros nas variáveis, feitas no capítulo anterior.

2.1 Estimador de Mínimos Quadrados Generalizados

Para este modelo funcional, Sprent (1966) propõe a estimação do parâmetro β pelo método de mínimos quadrados generalizados, que ao contrário do modelo clássico de regressão leva em conta não só as distâncias verticais do ponto do modelo ajustado, mas também, as distâncias horizontais, dado que há erros nas duas direções.

Das equações (2.1), dadas pelo modelo funcional, definimos

$$v_i = e_i - \beta u_i. \quad (2.1.1)$$

Não é difícil verificar que a variância residual de (2.1.1) é dada por:

$$Var(v_i) = \sigma_e^2 + \beta^2 \sigma_u^2 - 2\beta\sigma_{eu}^2. \quad (2.1.2)$$

A matriz de covariância amostral, para as n observações, pode ser escrita da seguinte maneira:

$$\mathbf{B} = \begin{bmatrix} b_{XX} & b_{XY} \\ b_{XY} & b_{YY} \end{bmatrix} \quad (2.1.3)$$

$$\text{onde } b_{XX} = \sum_{i=1}^n X_i^2, \quad b_{XY} = \sum_{i=1}^n X_i Y_i \quad \text{e} \quad b_{YY} = \sum_{i=1}^n Y_i^2$$

Levando em conta a suposição em que os erros e_i e e_j são não-correlacionados, e também os erros u_i e u_j são não-correlacionados, para todo $i \neq j$, um procedimento utilizado para estimar β é a minimização da soma de quadrados dos resíduos, com pesos inversamente proporcionais às suas variâncias. Isto significa que o estimador $\hat{\beta}^\#$ de β é o valor que minimiza

$$U = \sum_{i=1}^n \frac{(Y_i - \beta X_i)^2}{Var(Y_i - \beta X_i)},$$

ou seja,

$$U = \frac{\beta^2 b_{XX} - 2\beta b_{XY} + b_{YY}}{\beta^2 \sigma_u^2 - 2\beta \sigma_{eu}^2 + \sigma_e^2}. \quad (2.1.4)$$

A equação normal pode ser obtida pela diferenciação de U com respeito a β e igualando a derivada a zero. Alternativamente, a minimização pode ser feita diretamente da seguinte forma: para uma constante ζ fixa, podemos escrever U como

$$\frac{Q_1}{Q_2} = \zeta + \frac{(Q_1 - \zeta Q_2)}{Q_2}, \quad (2.1.5)$$

$$\text{onde } Q_1 = \beta^2 b_{XX} - 2\beta b_{XY} + b_{YY}$$

e

$$Q_2 = \beta^2 \sigma_u^2 - 2\beta \sigma_{eu}^2 + \sigma_e^2 \quad (2.1.6)$$

Resultado 2.1

Sob o modelo funcional temos que:

$$U = \frac{Q_1}{Q_2} = \zeta_1 + \frac{(b_{xx} - \zeta_1 \sigma_u^2)(\beta - \hat{\beta}^\#)^2}{Q_2} \quad (2.1.7)$$

é mínimo quando

$$\hat{\beta}^\# = \frac{b_{xy} - \zeta_1 \sigma_{eu}^2}{b_{xx} - \zeta_1 \sigma_u^2} \quad (2.1.8)$$

onde

$$\zeta_1 = \frac{f(X, Y, e, u) - \sqrt{[f(X, Y, e, u)]^2 - 4|\Sigma||\mathbf{B}|}}{2|\Sigma|}, \quad (2.1.9)$$

$$f(X, Y, e, u) = b_{xx} \sigma_e^2 + b_{yy} \sigma_u^2 - 2b_{xy} \sigma_{eu}^2, \quad (2.1.10)$$

$$\Sigma \text{ é uma matriz não singular conhecida, dada por (1.3).} \quad (2.1.11)$$

Prova:

Para algum ζ fixo, temos que:

$$U = \frac{Q_1}{Q_2} = \zeta + \frac{(Q_1 - \zeta Q_2)}{Q_2} \quad (2.1.12)$$

onde as variâncias σ_e^2 , σ_u^2 e σ_{eu}^2 são fixas.

Supondo que $\beta = \beta_0$ é o valor que minimiza U , ζ é o menor valor consistente que satisfaz a igualdade $(Q_1 - \zeta Q_2) = (b_{xx} - \zeta \sigma_u^2)(\beta - \beta_0)^2$.

Das equações dadas em (2.1.6), temos que:

$$\begin{aligned} Q_1 - \zeta Q_2 &= \beta^2 b_{xx} - 2\beta b_{xy} + b_{yy} - \zeta(\beta^2 \sigma_u^2 - 2\beta \sigma_{eu}^2 + \sigma_e^2) \\ &= \beta^2(b_{xx} - \zeta \sigma_u^2) - 2\beta(b_{xy} - \zeta \sigma_{eu}^2) + (b_{yy} - \zeta \sigma_e^2) \\ &= (b_{xx} - \zeta \sigma_u^2) \left[\beta^2 - 2\beta \frac{(b_{xy} - \zeta \sigma_{eu}^2)}{(b_{xx} - \zeta \sigma_u^2)} + \frac{(b_{yy} - \zeta \sigma_e^2)}{(b_{xx} - \zeta \sigma_u^2)} \right] \end{aligned} \quad (2.1.13)$$

É desejável para obtermos (2.1.7) que

$$Q_1 - \zeta Q_2 = (b_{xx} - \zeta \sigma_u^2)(\beta - \beta_0)^2 \quad (2.1.14)$$

para algum ζ e β_0 .

Portanto, a igualdade das equações (2.1.12) e (2.1.13) serão satisfeitas se, e somente se, para todo β ,

$$\beta^2 - 2\beta\beta_0 + \beta_0^2 = \beta^2 - 2\beta \frac{(b_{XY} - \zeta\sigma_{eu}^2)}{(b_{XX} - \zeta\sigma_u^2)} + \frac{(b_{YY} - \zeta\sigma_e^2)}{(b_{XX} - \zeta\sigma_u^2)}$$

se, e somente se,

$$\beta_0 = \frac{(b_{XY} - \zeta\sigma_{eu}^2)}{(b_{XX} - \zeta\sigma_u^2)} \quad (2.1.15)$$

e

$$\beta_0^2 = \frac{(b_{YY} - \zeta\sigma_e^2)}{(b_{XX} - \zeta\sigma_u^2)} \quad (2.1.16)$$

consequentemente, de (2.1.15) e (2.1.16), temos que:

$$\left[\frac{(b_{XY} - \zeta\sigma_{eu}^2)}{(b_{XX} - \zeta\sigma_u^2)} \right] = \frac{(b_{YY} - \zeta\sigma_e^2)}{(b_{XX} - \zeta\sigma_u^2)} \quad (2.1.17)$$

se, e somente se,

$$(b_{XY} - \zeta\sigma_{eu}^2)^2 - (b_{XX} - \zeta\sigma_u^2)(b_{YY} - \zeta\sigma_e^2) = 0$$

se, e somente se,

$$\zeta^2 [\sigma_u^2 \sigma_e^2 - \sigma_{eu}^4] - \zeta [b_{XX}\sigma_e^2 + b_{YY}\sigma_u^2 - 2b_{XY}\sigma_{eu}^2] + [b_{XX}b_{YY} - (b_{XY})^2] \quad (2.1.18)$$

se, e somente se,

$$\zeta^2 |\Sigma| - \zeta f(X, Y, e, u) + |\mathbf{B}| = 0, \quad (2.1.19)$$

onde $f(X, Y, e, u) = b_{XX} \sigma_e^2 + b_{YY} \sigma_u^2 - 2b_{XY} \sigma_{eu}^2$.

Desse modo, as raízes que satisfazem (2.1.19) são:

$$\zeta = \frac{f(X, Y, e, u) \pm \sqrt{[f(X, Y, e, u)]^2 - 4|\Sigma||\mathbf{B}|}}{2|\Sigma|}, \quad (2.1.20)$$

porém, como mencionamos anteriormente ζ deve ter o valor menor, para minimizar U .

Assim, se denotarmos por ζ_1 como a menor raiz determinada em (2.1.20), ou seja,

$$\zeta_1 = \frac{f(X, Y, e, u) - \sqrt{[f(X, Y, e, u)]^2 - 4|\Sigma||\mathbf{B}|}}{2|\Sigma|} \quad (2.1.21)$$

teremos determinado ζ explicitamente, com a devida condição satisfeita para minimizar U .

Portanto, baseado no procedimento acima, o estimador de mínimos quadrados generalizados, proposto por Sprent é dado por:

$$\hat{\beta}^\# = \frac{b_{XY} - \zeta_1 \sigma_{eu}^2}{b_{XX} - \zeta_1 \sigma_u^2}, \quad (2.1.22)$$

onde ζ_1 é dado por (2.1.21) e $f(X, Y, e, u) = b_{XX} \sigma_e^2 + b_{YY} \sigma_u^2 - 2b_{XY} \sigma_{eu}$.

□

Nota: De (2.1.19), como $|\mathbf{B} - \zeta \Sigma| = \zeta^2 |\Sigma| - \zeta f(X, Y, e, u) + |\mathbf{B}|$, uma forma alternativa para determinar ζ_1 seria calculando a menor raiz característica da equação:

$$|\mathbf{B} - \zeta \Sigma| = 0 \quad (2.1.23)$$

A seguir, ressaltaremos algumas particularidades do estimador de Sprent quando está sujeito a algumas restrições. Os resultados que podemos verificar são:

(i) quando $\sigma_u^2 = 0$.

Dizer que $\sigma_u^2 = 0$, é equivalente dizer que $X_i = x_i$. Portanto, $Cov(e_i, u_i) = \sigma_{eu}^2 = 0$ e isto implica que Σ é singular.

Consequentemente, a equação (2.1.8) fica reduzida a:

$$\hat{\beta}^\# = \frac{b_{XY}}{b_{XX}}.$$

Podemos notar que o estimador de mínimos quadrados ordinários $\hat{\beta}^*$ do modelo de regressão usual (1.6) e o estimador de mínimos quadrados generalizados de Sprent, dado pelo modelo funcional, coincidem. Intuitivamente é natural esperar que esse fato acontecesse, pois observe que se $\sigma_u^2 = 0$ significa que a variável independente x_i não contém erro de mensuração, que é a suposição feita no modelo de regressão usual.

(ii) quando $\sigma_e^2 = \sigma_u^2$ e $\sigma_{eu}^2 = 0$.

Neste caso, os erros têm variâncias iguais e são não correlacionados. Logo, pela equação (2.1.8), temos que

$$\hat{\beta}^\# = \frac{b_{XY}}{b_{XX} - \zeta_2},$$

onde ζ_2 é obtido determinando simplesmente a menor raiz característica de

$$|\mathbf{B} - \zeta \mathbf{I}_2| = 0,$$

pois, $\Sigma = \sigma_u^2 \mathbf{I}_2$ e consequentemente $\zeta_2 = \sigma_u^2 \zeta_1$.

Note que o estimador $\hat{\beta}^\#$ torna-se independente de σ_u^2 (variância do erro da variável observável X). Portanto, desde que tenhamos a variância dos erros iguais, o estimador de Sprent para β não se altera quando não conhecemos individualmente σ_e^2 ou σ_u^2 .

(ii) quando $\sigma_e^2 \neq \sigma_u^2$ e $\sigma_{eu}^2 = 0$ (e_i e u_i não correlacionados).

Seja ϕ a razão entre as variâncias dos erros σ_e^2 e σ_u^2 , ou seja, $\phi = \frac{\sigma_e^2}{\sigma_u^2}$. Desse modo, a

equação (2.1.8) é dada por:

$$\hat{\beta}^\# = \frac{(b_{YY} + \phi b_{XX}) + \sqrt{(b_{YY} + \phi b_{XX})^2 - 4\phi(b_{XY})^2}}{2b_{XY}}.$$

Neste caso específico, para se determinar o estimador de Sprent, basta que, somente, a razão das variâncias dos erros seja conhecida e não seus valores individualmente.

Nota:

- (1) Nos casos acima, quando os elementos de Σ não são conhecidos, eles podem ser substituídos pelos seus estimadores, que são obtidos pela réplica das observações de X e Y correspondendo aos valores fixos x_i, y_i de x, y .
- (2) O estimador de Sprent não exige a normalidade dos resíduos.

- (3) Uma indagação importante que pode ser levantada, em relação ao resultado de Sprent é a seguinte: Que tipo de priori não-informativa justifica os resultados obtidos por Sprent? A resposta para esta pergunta é dada por Rodrigues (1990) e trabalhada na próxima seção.

2.2 Inferência Bayesiana para β

Para podermos realizar a inferência sobre o parâmetro de inclinação, considere as suposições adicionais a respeito da distribuição dos erros, da seguinte maneira:

- i) $e_i \sim N(0, \sigma_e^2)$ com σ_e^2 conhecida;
- ii) $u_i \sim N(0, \sigma_u^2)$ com σ_u^2 conhecida;
- iii) e_i e u_i não correlacionados.

Nosso objetivo nesta seção é obter informação sobre o parâmetro β , em presença do parâmetro desconhecido $x = (x_1, x_2, \dots, x_n)$.

O logaritmo da função de verossimilhança é dado por:

$$\ell(\beta, x) = \log(L) \propto -\frac{1}{2} \left\{ \frac{\sum_{i=1}^n (Y_i - \beta x_i)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_i)^2}{\sigma_u^2} \right\}, \quad (2.2.1)$$

e o logaritmo da função de verossimilhança “profile” é dado por:

$$\ell_p(\beta) = \ell(\beta, \hat{x}_\beta) \propto \sum_{i=1}^n \frac{(Y_i - \beta X_i)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} \quad (2.2.2)$$

onde $\hat{x}_\beta$ é o estimador de máxima verossimilhança de x dado β .

Utilizando o resultado 2.1 podemos reescrever o logaritmo da função de verossimilhança “profile” como:

$$\ell_p(\beta) \propto -\frac{1}{2}(b_{xx} - \zeta \sigma_u^2) \frac{(\beta - \hat{\beta}^*)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} \quad (2.2.3)$$

onde $\hat{\beta}^*$, para este caso, restringe-se a:

$$\hat{\beta}^* = \frac{b_{xy}}{b_{xx} - \zeta_1 \sigma_u^2}$$

e de (2.1.21) temos que $\zeta_1 = \frac{f(X, Y, e, u) - \sqrt{[f(X, Y, e, u)]^2 - 4|\Sigma||\mathbf{B}|}}{2|\Sigma|}$,

também $f(X, Y, e, u) = b_{xx} \sigma_e^2 + b_{yy} \sigma_u^2$.

Lindley em sua discussão do artigo de Sprent (1966) formulou algumas densidades a priori não informativa para justificar os resultados obtidos por Sprent através do método dos mínimos quadrados generalizados, porém, ele não conseguiu justificar exatamente os resultados por Sprent. Analisando este fato, Rodrigues (1990) formulou uma densidade a priori não informativa baseada na regra de Jeffreys com a qual conseguiu uma justificativa exata.

A regra de Jeffreys é definida como a raiz quadrada do determinante da matriz de Informação de Fisher (Box & Tiao, 1973) e a informação de Fisher é obtida através da derivada segunda do logaritmo da função de verossimilhança. Analiticamente, a matriz de informação de Fisher é dada por:

$$I(\beta, x) = \begin{bmatrix} \frac{\sum_{i=1}^n x_i^2}{\sigma_e^2} & \frac{\beta x_1}{\sigma_e^2} & \frac{\beta x_2}{\sigma_e^2} & \dots & \frac{\beta x_n}{\sigma_e^2} \\ \frac{\beta x_1}{\sigma_e^2} & \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ \frac{\beta x_n}{\sigma_e^2} & 0 & \dots & \frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \end{bmatrix} \quad (2.2.4)$$

Desse modo, a densidade a priori baseada na regra de Jeffreys é formulada por:

$$p(\beta, x) \propto \left(\frac{\beta}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{\frac{n}{2}} \quad (2.2.5)$$

Portanto a densidade a posteriori conjunta para β e x é obtida pela combinação de (2.2.5) e a função de verossimilhança que utiliza o resultado 2.1. Consequentemente, integrando a posteriori conjunta em relação a x , determinamos a densidade a posteriori marginal de β que é dada por:

$$p(\beta|\text{dados}) \propto \exp \left\{ -\frac{1}{2} (b_{xx} - \zeta \sigma_u^2) \frac{(\beta - \hat{\beta}^\#)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} \right\} \quad (2.2.6)$$

A densidade a posteriori “profile” de β é dada por:

$$p_p(\beta|\text{dados}) \propto \left(\frac{\beta^2}{\sigma_e^2} + \frac{1}{\sigma_u^2} \right)^{\frac{n}{2}} \exp \left\{ -\frac{1}{2} (b_{xx} - \zeta \sigma_u^2) \frac{(\beta - \hat{\beta}^\#)^2}{\sigma_u^2 \beta^2 + \sigma_e^2} \right\} \quad (2.2.7)$$

Note que por (2.2.6) podemos verificar que a posteriori marginal coincide com a função de verossimilhança profile dada por (2.2.3), o que não ocorre com a densidade a posteriori profile, resgatando a mesma situação que Lindley apresentada na discussão do artigo de Sprent.

2.3 Modelo Funcional com Réplicas

Nesta seção analisaremos um caso particular do modelo funcional que na literatura é denominado por modelo funcional com réplicas, onde o valor real da variável independente x_i é fixado para todas as repetições, isto é, $x_i = x_0$ para $i = 1, 2, \dots, n$.

$$\begin{aligned} Y_i &= y_0 + e_i \\ X_i &= x_0 + u_i \\ y_0 &= \beta x_0 \quad i = 1, 2, \dots, n \end{aligned} \tag{2.3.1}$$

onde o verdadeiro valor não-observável x_0 é considerado fixo. Ilustrando este modelo com um exemplo, suponha que estamos interessados em testar a eficácia do instrumento de mensuração e para isso mensuramos repetidas vezes o mesmo objeto; mais especificamente considere o quilo padrão mensurado por várias máquinas da empresa Bem Pesado e nosso interesse é verificar a eficiência desta marca de balança.

Definição 2.1

Dado o modelo funcional com réplicas (2.3.1), $(\hat{\beta}, \hat{x}_0)$ é um estimador de mínimos quadrados generalizados para (β, x_0) se

$$\text{Min } Q(\beta, x_0) = Q(\hat{\beta}, \hat{x}_0)$$

$$\text{onde } Q(\beta, x_0) = \frac{\sum_{i=1}^n (Y_i - \beta x_0)^2}{\sigma_e^2} + \frac{\sum_{i=1}^n (X_i - x_0)^2}{\sigma_u^2}. \quad (2.3.2)$$

RESULTADO 2.2

Sob o modelo funcional com réplicas (2.2.1), o estimador de mínimos quadrados generalizados para (β, x_0) é dado por:

$$\hat{\beta} = \frac{\sigma_e^2 \sigma_u^2 \left(\sum_{i=1}^n Y_i \right) \left(\sum_{i=1}^n X_i \right)}{n \sigma_u^2 \left[\sigma_u^2 \left(\sum_{i=1}^n Y_i^2 \right) + \sigma_e^2 \left(\sum_{i=1}^n X_i^2 \right) - \Lambda \sigma_e^2 \sigma_u^2 \right] - \sigma_u^4 \left(\sum_{i=1}^n Y_i \right)^2} \quad (2.3.3)$$

$$\text{e } \hat{x}_0 = \frac{\hat{\beta} \sigma_u^2 \left[\sum_{i=1}^n Y_i \right] + \sigma_e^2 \left[\sum_{i=1}^n X_i \right]}{n \left(\hat{\beta}^2 \sigma_u^2 + \sigma_e^2 \right)} \quad (2.3.4)$$

onde

$$\Lambda = \frac{(Aa + Bb) - \sqrt{(Aa + Bb)^2 - 4(AB - C^2)(ab)}}{2ab} \quad (2.3.5)$$

e, também, a, b, A, B, C definidos por:

$$a = n \sigma_e^2 \sigma_u^4$$

$$b = n \sigma_e^4 \sigma_u^2$$

$$A = n \sigma_u^2 \left(\sum_{i=1}^n Y_i^2 \right) + n \sigma_e^2 \sigma_u^2 \left(\sum_{i=1}^n X_i^2 \right) - \sigma_u^4 \left(\sum_{i=1}^n Y_i \right)^2$$

$$B = n \sigma_e^2 \left[\sigma_u^2 \left(\sum_{i=1}^n Y_i^2 \right) + \sigma_e^2 \left(\sum_{i=1}^n Y_i^2 \right) \right] - \sigma_e^4 \left(\sum_{i=1}^n X_i \right)^2$$

$$C = \sigma_u^2 \sigma_e^2 \left(\sum_{i=1}^n Y_i \right) \left(\sum_{i=1}^n X_i \right)$$

Prova:

A determinação dos estimadores de β e x_0 , pelo método dos mínimos quadrados generalizados, visa minimizar a soma de quadrados dos resíduos ponderados Q .

Desenvolvendo a expressão (2.3.2), temos que:

$$Q = \frac{\sum_{i=1}^n Y_i^2}{\sigma_e^2} - 2x_0 \left[\frac{\beta \sum_{i=1}^n Y_i}{\sigma_e^2} + \frac{\sum_{i=1}^n X_i}{\sigma_u^2} \right] + x_0^2 \left[\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} \right] + \frac{\sum_{i=1}^n X_i^2}{\sigma_u^2}$$

ou após uma operação algébrica trivial obtemos que:

$$Q = \frac{\sum_{i=1}^n Y_i^2}{\sigma_e^2} + \frac{\sum_{i=1}^n X_i^2}{\sigma_u^2} + \left[\frac{n\beta}{\sigma_e^2} + \frac{n}{\sigma_u^2} \right] \left[x_0^2 - 2 \frac{\frac{\beta \sum_{i=1}^n Y_i}{\sigma_e^2} + \frac{\sum_{i=1}^n X_i}{\sigma_u^2}}{\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2}} \right]. \quad (2.3.6)$$

Se denotarmos por

$$\hat{x}_0(\beta) = \frac{\frac{\beta \sum_{i=1}^n Y_i}{\sigma_e^2} + \frac{\sum_{i=1}^n X_i}{\sigma_u^2}}{\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2}} \quad (2.3.7)$$

e, em (2.3.6), completando os quadrados em $\hat{x}_0$, podemos reescrever esta expressão (2.3.6) da seguinte forma:

$$Q = \frac{\sigma_u^2 \sum_{i=1}^n Y_i^2 + \sigma_e^2 \sum_{i=1}^n X_i^2}{\sigma_e^2 \sigma_u^2} + \left[\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} \right] \left[(x_0(\beta) - \hat{x}_0(\beta))^2 - \hat{x}_0^2(\beta) \right].$$

Equivalentemente, temos

$$Q = \frac{\sigma_u^2 \sum_{i=1}^n Y_i^2 + \sigma_e^2 \sum_{i=1}^n X_i^2}{\sigma_e^2 \sigma_u^2} - \frac{\left[\frac{\beta \sum_{i=1}^n Y_i}{\sigma_e^2} + \frac{\sum_{i=1}^n X_i}{\sigma_u^2} \right]^2}{\left[\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} \right]} + \left[\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} \right] (x_0 - \hat{x}_0(\beta))^2 \quad (2.3.8)$$

Podemos verificar que, de (2.3.8) acima, $\hat{x}_0(\beta)$ em (2.3.7) é o valor de x_0 que minimiza Q . Queremos determinar agora, qual é o valor de β que minimiza Q . É fácil notar que em (2.3.8), queremos determinar qual o β que minimiza a seguinte expressão:

$$Q = \frac{\sigma_u^2 \sum_{i=1}^n Y_i^2 + \sigma_e^2 \sum_{i=1}^n X_i^2}{\sigma_e^2 \sigma_u^2} - \frac{\left[\frac{\beta \sum_{i=1}^n Y_i}{\sigma_e^2} + \frac{\sum_{i=1}^n X_i}{\sigma_u^2} \right]^2}{\left[\frac{n\beta^2}{\sigma_e^2} + \frac{n}{\sigma_u^2} \right]} \quad (2.3.9)$$

Supondo que $x_0 = \hat{x}_0(\beta)$, (2.3.8) ou (2.3.9) pode ser escrito da seguinte maneira:

$$Q = \frac{n(\beta^2 \sigma_u^2 + \sigma_e^2) \left[\sigma_u^2 \sum_{i=1}^n Y_i^2 + \sigma_e^2 \sum_{i=1}^n X_i^2 \right] - \left[\sigma_u^2 \beta \left(\sum_{i=1}^n Y_i \right) + \sigma_e^2 \left(\sum_{i=1}^n X_i \right) \right]^2}{n \sigma_e^2 \sigma_u^2 (\beta^2 \sigma_u^2 + \sigma_e^2)} \quad (2.3.10)$$

Usando o Resultado 2.1, onde

$$Q_1 = n(\beta^2 \sigma_u^2 + \sigma_e^2) \left[\sigma_u^2 \left(\sum_{i=1}^n Y_i^2 \right) + \sigma_e^2 \left(\sum_{i=1}^n X_i^2 \right) \right] - \left[\sigma_u^2 \beta \left(\sum_{i=1}^n Y_i \right) + \sigma_e^2 \left(\sum_{i=1}^n X_i \right) \right]^2$$

$$Q_2 = n \sigma_e^2 \sigma_u^2 (\beta^2 \sigma_u^2 + \sigma_e^2)$$

temos, finalmente, que:

$$\hat{\beta}^P = \frac{\sigma_e^2 \sigma_u^2 \left(\sum_{i=1}^n Y_i \right) \left(\sum_{i=1}^n X_i \right)}{n \sigma_u^2 \left[\sigma_u^2 \left(\sum_{i=1}^n Y_i^2 \right) + \sigma_e^2 \left(\sum_{i=1}^n X_i^2 \right) - \Lambda \sigma_e^2 \sigma_u^2 \right] - \sigma_u^4 \left(\sum_{i=1}^n Y_i \right)^2}$$

e

$$\hat{x}_0 = \hat{x}_0(\hat{\beta}) = \frac{\frac{\hat{\beta}^P \left[\sum_{i=1}^n Y_i \right]}{\sigma_e^2} + \frac{\left[\sum_{i=1}^n X_i \right]}{\sigma_u^2}}{\frac{n \hat{\beta}^{P^2}}{\sigma_e^2} + \frac{n}{\sigma_u^2}}.$$

Capítulo 3

INFERÊNCIA BAYESIANA PARA O MODELO ESTRUTURAL

Suponhamos que os dados (Y_i, X_i) sejam provenientes do modelo estrutural, definido anteriormente pelas equações:

$$\begin{aligned} Y_i &= y_i + e_i \\ X_i &= x_i + u_i \\ y_i &= \beta x_i \end{aligned} \quad i = 1, 2, \dots, n \quad (3.1)$$

onde o verdadeiro valor não-observável x_i é considerado uma variável aleatória, ou seja, assumimos que existe uma certa distribuição sobre esta variável.

Considere as seguintes suposições adicionais, as suposições (1.2), para este modelo: i) $e_i \sim N(0, \sigma_e^2)$;

ii) $u_i \sim N(0, \sigma_u^2)$;

iii) $x_i \sim N(\mu_x, \sigma_x^2)$, com μ_x conhecido;

iv) e_i, u_i, x_i não correlacionados.

Utilizando os resultados conhecidos da teoria de distribuição Normal Multivariada (por exemplo, consulte Seber, 1977) verificamos que o vetor $(X_i, Y_i)'$ tem distribuição Normal bivariada com vetor de médias

$$E \begin{bmatrix} Y \\ X \end{bmatrix} = \begin{pmatrix} \mu_Y \\ \mu_X \end{pmatrix} = \begin{pmatrix} \beta \mu_x \\ \mu_x \end{pmatrix} \quad (3.2)$$

e matriz de covariância

$$\Sigma = \begin{bmatrix} \sigma_Y^2 & \sigma_{XY}^2 \\ \sigma_{XY}^2 & \sigma_X^2 \end{bmatrix} = \begin{bmatrix} \beta^2 \sigma_x^2 + \sigma_e^2 & \beta \sigma_x^2 \\ \beta \sigma_x^2 & \sigma_x^2 + \sigma_u^2 \end{bmatrix} \quad (3.3)$$

Nota: Pelas propriedades da distribuição Normal bivariada podemos verificar que o estimador de mínimos quadrados ordinário, utilizado pela análise de regressão linear usual, é um estimador viciado, ou seja,

$$E[\hat{\beta}] = \frac{\sigma_{XY}^2}{\sigma_{XX}^2} = \beta \left[\frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} \right] \quad (3.4)$$

Ressaltando, novamente, que como nosso interesse principal está em estudar o parâmetro de inclinação vamos supor, sem perda de generalidade, que a média da variável não observável é nula, ou seja, $\mu_x = 0$.

Uma forma é analisar o parâmetro sujeito ao modelo bidimensional, porém, Rodrigues & Cordani (1990) visualizam sob outra perspectiva, trabalhando com a distribuição condicional de Y_i dado X_i . Uma indagação pode surgir na formulação deste método: será que é razoável este procedimento? A resposta para esta pergunta parece ser bastante simples, pois, uma vez que X_i é a variável controle. Em acréscimo, pode-se observar uma grande simplificação do modelo, pois, passar-se-á de um modelo bidimensional para um modelo unidimensional.

Verifica-se facilmente de (3.2), usando os resultados conhecidos da distribuição Normal bivariada, que a distribuição condicional de Y_i dado X_i é Normal com média $K_X \beta X_i$ e variância $K_X \beta^2 \sigma_u^2 + \sigma_e^2$, isto é,

$$Y_i | X_i \sim N[K_X \beta X_i, K_X \beta^2 \sigma_u^2 + \sigma_e^2] \quad (3.5)$$

$$\text{onde } K_X = \frac{\sigma_x^2}{\sigma_x^2 + \sigma_u^2} = \frac{\sigma_x^2}{\sigma_X^2}, \quad (3.6)$$

que define o grau de atenuação do parâmetro β , isto é, explica realmente qual é a parte de β que estamos estimando, pois, de (3.4) temos que $E[\hat{\beta}] = \beta K_X$.

Existe um número significativo de nomes para o quociente K_X , que é chamada de confiabilidade de X_i na literatura de Ciências Sociais, é chamada herdabilidade em genética, etc. Sugerimos consultar Fuller (1987) para exemplos ilustrativos sobre o efeito do fator de atenuação na estimativa de β .

De (3.6), temos a interpretação de K_X como sendo a proporção da variabilidade da variável não observável x (σ_x^2) correspondente a variabilidade da variável observável X (σ_X^2).

A determinação do intervalo Highest Posterior Density (HPD) para β pode ser bastante simplificada trabalhando-se com uma parametrização ortogonal (definido no apêndice A), onde os parâmetros ortogonais são determinados seguindo a técnica de Cox & Reid (1987).

Nas próximas seções analisaremos a inferência para β quando o fator de atenuação K_X é conhecido e desconhecido, via parâmetros ortogonais.

3.1 Fator de Atenuação Conhecido

Existe um número de situações, particularmente em psicologia, sociologia e estudo de amostragem, onde a informação sobre o fator K_X é acessível, por exemplo, se o interesse está em uma característica particular do indivíduo, tal como a inteligência, lealdade da comunidade, consciência social, espontaneidade para adotar novas práticas, capacidade administrativa, a probabilidade que o indivíduo votará em uma eleição em particular, etc. O teste de Q.I. é, talvez, o melhor exemplo conhecido; de onde é impossível observar o valor

diretamente da característica e a resposta X_i é obtida como uma pontuação construída das respostas a um certo número de questões. Neste particular exemplo, geralmente, K_X é conhecido.

A parametrização ortogonal, como mencionado acima, facilita-nos em muitos cálculos para a estimação de β . Assim, seguindo o procedimento dado por Cox & Reid (1987), temos a seguinte reparametrização:

$$\begin{aligned}\beta &= \beta \\ \sigma^2 &= K_X \beta^2 \sigma_u^2 + \sigma_e^2\end{aligned}\tag{3.1.1}$$

onde o parâmetro de interesse é β .

Pergunta: Como o parâmetro σ^2 depende de β em (3.1.1), será que comprometemos a inferência sobre β ?

Cox & Reid (1987) fornece a resposta para esta pergunta, onde afirmam que a variância de uma distribuição Normal é ortogonal à sua média.

Reescrevendo a distribuição condicional (3.5) em termos dos parâmetros ortogonais β e σ^2 , tem-se que:

$$Y_i | X_i \sim N[K_X \beta X_i, \sigma^2]\tag{3.1.2}$$

Assim, a função de verossimilhança para (β, σ^2) é dada por:

$$L(\beta, \sigma) \propto \frac{1}{\sigma^n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - K_X \beta X_i)^2\right\}\tag{3.1.3}$$

e os estimadores de máxima verossimilhança para β e σ^2 são:

$$\hat{\beta} = K_X^{-1} \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2} = K_X^{-1} \hat{\beta}^*\tag{3.1.4}$$

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta} K_X X_i)^2 \quad (3.1.5)$$

Note que o papel de K_X será o de corrigir o estimador de mínimos quadrados $\hat{\beta}^*$, quando a variável X é considerada sem erro de mensuração. Além disso, pode-se constatar em (3.14) que o estimador $\hat{\beta}$, realmente, não depende de σ^2 , como consequência da ortogonalidade entre os parâmetros β e σ^2 .

A densidade a priori conjunta localmente uniforme para β e $\log \sigma$ (consultar Box & Tiao, 1973 para explicações mais detalhadas) é dada por:

$$p(\beta, \log \sigma^2) \propto \text{constante}$$

Consequentemente, a priori conjunta para β e σ^2 é:

$$p(\beta, \sigma^2) \propto \frac{1}{\sigma^2} \quad (3.1.6)$$

Logo, podemos expressar a distribuição conjunta a posteriori como:

$$p(\beta, \sigma^2 | \text{dados}) \propto \frac{1}{\sigma^{\frac{n+1}{2}}} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta K_X X_i)^2 \right\} \quad (3.1.7)$$

Para obtermos informações somente a respeito de β , é necessário determinar sua distribuição marginal, a qual será obtida integrando-se a posteriori conjunta (3.1.7) em relação a σ^2 , que resultará numa distribuição Gama Invertida (ver Zellner, 1971). Deduzindo-se, também, a constante normalizadora a distribuição marginal posteriori de β é dada por:

$$p(\beta | \text{dados}) = \frac{\Gamma\left(\frac{n+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n}{2}\right)} \left[\frac{K_X^2 \sum_{i=1}^n X_i^2}{\hat{\sigma}^2} \right]^{\frac{1}{2}} \left[1 + \frac{(\beta - \hat{\beta})^2 K_X^2 \sum_{i=1}^n X_i^2}{n \hat{\sigma}^2} \right]^{-\frac{n+1}{2}} \quad (3.1.8)$$

onde $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta} K_X X_i)^2$

A equação (3.1.8) mostra claramente que a distribuição a posteriori de β segue uma distribuição t de Student com n graus de liberdade e centrada em $\hat{\beta}$.

3.1.1 Intervalo HPD para β

Pelo fato da densidade a posteriori de β ter uma distribuição t de Student e não estar centrada na origem, usaremos, primeiramente, a transformação dada a seguir com o propósito de se determinar explicitamente o intervalo HPD para β :

$$T = \frac{(\beta - \hat{\beta}) K_X \left[\sum_{i=1}^n X_i^2 \right]^{\frac{1}{2}}}{\hat{\sigma}} \quad (3.1.1.1)$$

Após utilizarmos a transformação (3.1.1.1) implica que a distribuição da variável aleatória T é a distribuição t de Student padronizada dada por:

$$p\left(T \mid \underline{\tilde{Y}}, \underline{\tilde{X}}\right) \propto \left[1 + \frac{T^2}{n} \right]^{-\frac{n+1}{2}}, \quad (3.1.1.2)$$

ou seja,

$$T \mid \underline{\tilde{Y}}, \underline{\tilde{X}} \sim t_n \quad (3.1.1.3)$$

onde t_n é a distribuição t de Student com n graus de liberdade e centrada na origem, que pode ser encontrada em tabelas existentes em diversos livros da área.

A partir da transformação dada por (3.1.1.1) e de (3.1.1.2), supondo que:

$$P\left[|t| \leq t_{\alpha/2}\right] = 1 - \alpha,$$

não é difícil verificar que o intervalo HPD, com nível de credibilidade $100(1 - \alpha)\%$ é dado pela seguinte expressão:

$$\hat{\beta} \pm \frac{\hat{\sigma} t_{\alpha/2}}{1} \quad (3.1.1.4)$$

Um detalhe que vale ressaltar na expressão (3.1.1.4) é que este intervalo HPD para β , determinado pela abordagem Bayesiana, coincide com o intervalo de confiança clássico (veja em Rodrigues & Cordani, 1990).

Observe que quando o fator de atenuação K_X decresce o comprimento do intervalo cresce, isto mostra que o erro de mensuração interfere efetivamente no intervalo HPD. Em consequência disto o pesquisador tem uma visão errônea do intervalo a posteriori quando despreza o erro de mensuração na variável independente x , ou seja, julga ter um determinado intervalo para o parâmetro de interesse, porém seu verdadeiro intervalo é mais amplo.

3.1.2 Exemplo Numérico

Apresentamos um exemplo numérico, usando o método de Monte Carlo, com a finalidade de comparar o estimador de mínimos quadrados ordinários $\hat{\beta}$ dado por (1.7) e o estimador de máxima verossimilhança $\hat{\beta}$ dado por (3.1.4). Para realizarmos a simulação, utilizamos o software Minitab, onde foram geradas 50 amostras, fixando $\beta = 2$ e as variâncias σ_e^2 , σ_u^2 e σ_x^2 com seu respectivo fator de atenuação K_X são colocados na tabela 3.1, logo abaixo.

Tabela 3.1 : As variâncias e seu respectivo fator de atenuação K_X .

K_X	0.5	0.8	1.0
σ_e^2	1	1	1
σ_u^2	1	0.25	0
σ_x^2	1	1	1

Os tamanhos da amostra, que estão denotados por n , e os valores de K_X são colocados na tabela 3.2, também, para cada estimador é colocado o valor médio e o valor entre parênteses é seu correspondente erro médio quadrático.

Tabela 3.2 : Estimador de máxima verossimilhança e o estimador de mínimos quadrados.

K_X	$n = 10$	$n = 30$	$n = 50$
0.5	$\hat{\beta}^* = 1.038$ (1.089)	$\hat{\beta}^* = 1.002$ (1.039)	$\hat{\beta}^* = 0.978$ (1.084)
	$\hat{\beta} = 2.076$ (0.660)	$\hat{\beta} = 2.004$ (0.173)	$\hat{\beta} = 1.957$ (0.164)
0.8	$\hat{\beta}^* = 1.609$ (0.282)	$\hat{\beta}^* = 1.572$ (0.220)	$\hat{\beta}^* = 1.607$ (0.174)
	$\hat{\beta} = 2.011$ (0.202)	$\hat{\beta} = 1.966$ (0.060)	$\hat{\beta} = 2.009$ (0.031)
1.0	$\hat{\beta}^* = 1.968$ (0.128)	$\hat{\beta}^* = 1.987$ (0.038)	$\hat{\beta}^* = 2.037$ (0.018)
	$\hat{\beta} = 1.968$ (0.128)	$\hat{\beta} = 1.987$ (0.038)	$\hat{\beta} = 2.037$ (0.018)

A tabela 3.2 mostra o efeito do erro de mensuração e a eficiência do estimador de máxima verossimilhança $\hat{\beta}$. Note que o estimador de mínimos quadrados $\hat{\beta}^*$ é fortemente afetado pelo erro de mensuração e, assim, podemos constatar que $\hat{\beta}^*$ realmente estima somente uma parte de β . Note, também, que, para todos os tamanhos de amostra, o erro quadrático médio aumenta quando o fator de atenuação diminui.

Pudemos notar, pelos resultados da simulação acima, as consequências nos estimadores ao desprezarmos os erros de mensuração na variável X . Vamos verificar agora qual a influência dos erros nas variáveis sobre a densidade a posteriori de β e para este propósito, considerando o parâmetro de inclinação igual a 2, foi gerada uma amostra de tamanho 50 para três níveis do fator de atenuação e os dados estão organizados na tabela logo abaixo.

Tabela 3.3: Dados simulados e níveis para o fator de Atenuação

Dados	Fator de Atenuação K_x		
	0.5	0.8	1.0
$\sum_{i=1}^{50} Y_i^2$	247.643	270.616	304.140
$\sum_{i=1}^{50} X_i^2$	89.6292	65.8563	59.6736
$\sum_{i=1}^{50} X_i Y_i$	92.1172	108.273	121.737

Assim, apresentamos abaixo os gráficos das densidades a posteriori de β versus β , para os três níveis do fator de atenuação.

Figura 3.1 : Densidade a Posteriori de β , para o caso onde K_x é conhecido.

Conclusão : Dos gráficos acima, ao compararmos a posteriori de β com $K_x = 1$ com a posteriori de β com $K_x = 0.5$, verificamos que a diferença da forma com que se apresentam

é altamente significativa, ou seja, a posteriori de β é mais "fechada", quando $K_x = 1$, significando que tem menor variabilidade (mais informativa). Consequentemente, ao considerarmos que um determinado conjunto de dados não contém erros nas variáveis, estaremos cometendo um gravíssimo engano; pois ao interpretar a posteriori estaremos fornecendo um intervalo de credibilidade menor que seu real intervalo.

3.2 Fator de Atenuação Desconhecido

Existem casos onde as informações a respeito do fator de atenuação K_x nem sempre está disponível, isto acarreta em fazer algumas suposições referentes a alguns parâmetros envolvidos no problema em questão. De estudo anterior é reconhecido que o modelo estrutural (3.1) não é identificável, quando nada conhecemos a respeito das variâncias σ_e^2 e σ_u^2 . Contudo, bastaria que soubéssemos informações a respeito da razão $\delta = \sigma_e^2 / \sigma_u^2$. O modelo estrutural (3.1) com a razão δ conhecida é definido como o MODELO COM ERRO NAS VARIÁVEIS.

Para que o modelo estrutural (3.1) seja identificável, no nosso caso, vamos supor que $\sigma_e^2 = \sigma_u^2 = \sigma_0^2$ (conhecido). Esta situação é definida na literatura como o caso "overidentifiable".

De (3.5), temos que:

$$Y_i | X_i \sim N[K_x \beta X_i, K_x \beta^2 \sigma_u^2 + \sigma_e^2],$$

analogamente a seção anterior, pela simplificação dos cálculos, faremos uma reparametrização ortogonal, da seguinte maneira:

$$\theta = K_x \beta$$

$$\varphi = K_x \beta^2 \sigma_u^2 + \sigma_e^2 = \sigma_0^2 (K_x \beta^2 + 1)$$

Reescrevendo a distribuição condicional para os parâmetros ortogonais θ e φ , tem-se que:

$$Y_i | X_i \sim N[\theta X_i, \varphi] \quad (3.2.1)$$

A função de verossimilhança em termos dos parâmetros θ e φ , é dada por:

$$L(\theta, \varphi) \propto \frac{1}{\varphi^{n/2}} \exp \left\{ -\frac{1}{2\varphi} \sum_{i=1}^n (Y_i - \theta X_i)^2 \right\} \quad (3.2.2)$$

A densidade a priori conjunta que é localmente uniforme expressa-se por:

$$p(\theta, \log \varphi) \propto \text{constante}, \quad (3.2.3)$$

dessa forma a densidade a priori conjunta não informativa, para θ e φ , é:

$$p(\theta, \varphi) \propto \frac{1}{\varphi^{1/2}}. \quad (3.2.4)$$

Portanto, conjugando a densidade a priori com a função de verossimilhança, a densidade a posteriori conjunta resulta em:

$$p(\theta, \varphi \mid \text{dados}) \propto \frac{1}{\varphi^{(n+1)/2}} \exp \left\{ -\frac{1}{2\varphi} \sum_{i=1}^n (Y_i - \theta X_i)^2 \right\}. \quad (3.2.5)$$

Como nosso interesse é estudar o parâmetro de inclinação β , e não o parâmetro θ (usado somente para ortogonalização), considere a seguinte reparametrização:

$$\beta = \frac{\frac{\varphi}{\sigma_0^2} - 1}{\theta} \quad (3.2.6)$$

$$\varphi = \varphi$$

Utilizando a densidade (3.2.5) e as equações (3.2.6), e aplicando a técnica de transformação de variáveis, onde o Jacobiano da transformação é dado por:

$$\mathbf{J} = \begin{bmatrix} \frac{1 - \frac{\varphi}{\sigma_0^2}}{\beta^2} & \frac{1}{\beta \sigma_0^2} \\ 0 & 1 \end{bmatrix}, \text{ desde que } \beta \neq 0,$$

podemos expressar a densidade a posteriori conjunta para β e φ como:

$$p(\beta, \varphi \mid \text{dados}) \propto \frac{\frac{\varphi}{\sigma_0^2} - 1}{\beta^2 \varphi^{(n+1)/2}} \exp \left\{ -\frac{1}{2\varphi} \sum_{i=1}^n \left[Y_i - \left(\frac{\frac{\varphi}{\sigma_0^2} - 1}{\beta} \right) X_i \right]^2 \right\} \quad (3.2.7)$$

RESULTADO 3.1

Da distribuição a posteriori conjunta de β e φ dada por (3.2.7), temos que a distribuição a posteriori marginal de β é expressa por:

$$p(\beta|\text{dados}) \propto \frac{\hat{\varphi}_\beta^{\frac{\varphi}{2}-1}}{\beta^2 \hat{\varphi}_\beta^{(n+1)/2}} \exp\left\{-\frac{1}{2\hat{\varphi}_\beta} \sum_{i=1}^n \left[Y_i + \frac{X_i}{\beta} - \frac{\hat{\varphi}_\beta X_i}{\beta \sigma_0^2}\right]^2\right\} \quad (3.2.8)$$

onde

$$\hat{\varphi}_\beta = \beta \sigma_0^2 \left[\frac{\sum_{i=1}^n \left(Y_i + \frac{X_i}{\beta}\right)^2}{\sum_{i=1}^n X_i^2} \right]^{1/2}; \quad \beta > 0 \quad (3.2.9)$$

Prova

Para obter a distribuição marginal de β , é necessário fazer a integração da distribuição conjunta (3.2.7) em relação a φ . Como não podemos obter uma distribuição exata, pois não é similar a alguma distribuição conhecida, utilizaremos o método de Laplace (comentado no Apêndice B).

Portanto, temos que:

$$\begin{aligned} p(\beta|\text{dados}) &\propto \int_0^\infty \frac{\varphi^{\frac{\varphi}{2}-1}}{\beta^2 \varphi^{(n+1)/2}} \exp\left\{-\frac{1}{2\varphi} \sum_{i=1}^n \left[Y_i - \left(\frac{\varphi}{\sigma_0^2} - 1\right) \frac{X_i}{\beta}\right]^2\right\} d\varphi \\ &= \frac{1}{\beta^2} \int_0^\infty \frac{\left|\frac{\varphi}{\sigma_0^2} - 1\right|}{\varphi^{(n+1)/2}} \exp\left\{-\frac{1}{2\varphi} \sum_{i=1}^n \left[Y_i - \left(\frac{\varphi}{\sigma_0^2} - 1\right) \frac{X_i}{\beta}\right]^2\right\} d\varphi \end{aligned} \quad (3.2.10)$$

Fazendo uma simples identificação com a notação do método de Laplace, consideramos

$$f(\varphi) = \frac{\varphi}{\sigma_0^2} - 1 \quad e$$

$$h(\varphi) = \frac{1}{2n\varphi} \sum_{i=1}^n \left[Y_i - \left(\frac{\varphi}{\sigma_0^2} - 1 \right) \frac{X_i}{\beta} \right]^2 = \frac{1}{2n\varphi} \sum_{i=1}^n \left[Y_i + \frac{X_i}{\beta} - \frac{\varphi X_i}{\sigma_0^2} \right]^2 \quad (3.2.11)$$

Calculando o ponto que maximiza a função $-h(\varphi)$, denotado por $\hat{\varphi}_\beta$ encontraremos sua expressão dada por:

$$\hat{\varphi}_\beta^2 = \beta^2 \sigma_0^4 \frac{\sum_{i=1}^n \left[Y_i + \frac{X_i}{\beta} \right]^2}{\sum_{i=1}^n X_i^2},$$

consequentemente, extraindo a raiz quadrada, temos que:

$$\hat{\varphi}_\beta = \beta \sigma_0^2 \sqrt{\frac{\sum_{i=1}^n \left[Y_i + \frac{X_i}{\beta} \right]^2}{\sum_{i=1}^n X_i^2}}, \quad \beta > 0 \quad (3.2.12)$$

Dessa forma, a segunda derivada de $h(\varphi)$, aplicada no ponto $\hat{\varphi}_\beta$, é da seguinte forma:

$$h^{(2)}(\hat{\varphi}_\beta) = \frac{\sum_{i=1}^n X_i^2}{n \hat{\varphi}_\beta^2 \sigma_0^2}. \quad (3.2.13)$$

Assim, ao utilizar o método de aproximação de Laplace, resulta que a densidade a posteriori de β é:

$$p(\beta|\text{dados}) \propto \frac{\hat{\varphi}_\beta}{\beta^2 \hat{\varphi}_\beta^{(n+1)/2} \sigma_0^2} \exp \left\{ -\frac{1}{2\hat{\varphi}_\beta} \sum_{i=1}^n \left[Y_i + \frac{X_i}{\beta} - \frac{\hat{\varphi}_\beta X_i}{\sigma_0^2} \right]^2 \right\}$$

□

3.2.1 Exemplo Numérico

Com o intuito de analisar o comportamento da posteriori de β , assim como foi feito na seção anterior, realizamos várias simulações utilizando o método de Monte Carlo.

Observe a seguir o que ocorre com a função densidade a posteriori de β para um particular conjunto de dados, gerado pelo Minitab, de tamanho 50. Supondo que $\beta = 2$ e $K_X = 0.5$, onde os dados são colocados logo abaixo.

$$\text{Dados: } \sum_{i=1}^{30} X_i^2 = 402.181, \sum_{i=1}^{30} Y_i^2 = 112.492 \text{ e } \sum_{i=1}^{30} Y_i X_i = 145.207.$$

Utilizando o método de integração numérica, dado pelo software Eureka, determinamos a constante normalizadora. Assim, o gráfico da posteriori pode ser colocado da seguinte maneira:

Figura 3.2 : Densidade a posteriori de β , para K_X desconhecido.

Conclusão: Após realizar inúmeras simulações, as densidades a posteriori em geral não forneceram boas estimativas do parâmetro β . Algumas suposições são levantadas, como por exemplo, o conjunto de dados gerado não segue uma distribuição Normal, ou a aproximação de Laplace não está sendo precisa. Esses problemas serão abordados em uma análise futura utilizando a técnica de Gibbs Sampler (uma referência Casella, G. & George, E.I., 1992).

O que podemos garantir é que a distribuição é convergente e com um único ponto de máximo.

Capítulo 4

ESTRUTURA HIERÁRQUICA PARA A PRIORI NO MODELO ESTRUTURAL

4.1 Introdução

Toda análise feita nos capítulos anteriores foi sob o enfoque de priori não-informativa. Neste capítulo analisaremos qual será o comportamento da posteriori, do parâmetro de inclinação, quando é submetida a uma priori informativa, definida através da modelagem hierárquica (Lindley, 1972).

Inicialmente, vamos fazer uma breve introdução à modelagem hierárquica da seguinte forma: existem determinados casos em que a distribuição da priori de um parâmetro pode ser dividida em estágios para facilitar sua especificação ou para melhor descrever uma situação experimental.

A incorporação de informação pode ser desdobrada em duas fases:

(i) Estrutural : para divisão dos estágios;

(ii) Subjetiva : para especificação de cada estágio.

Exemplo: Suponha que $Y_1, Y_2, \dots, Y_n$ são tais que $Y_i \sim N(\theta_i, \sigma^2)$, com σ^2 conhecido. A especificação da priori de $\theta = (\theta_1, \theta_2, \dots, \theta_n)$ pode ser baseada nas seguintes hipóteses: os θ_i 's são independentes, isto é, a distribuição de θ é dada por $p(\theta) = \prod_{i=1}^n p(\theta_i)$ ou os θ_i 's

constituem uma amostra aleatória de uma população $p(\theta|\lambda)$, onde $\lambda = (\mu, \tau^2)$ contém os parâmetros que descrevem a população. Assim,

$$p(\theta|\lambda) = \prod_{i=1}^n p(\theta_i|\lambda).$$

Esta especificação constitui o primeiro estágio. Para se completar a priori é necessário especificar o segundo estágio: a distribuição de λ , que denotaremos por $p(\lambda)$. Observe que $p(\lambda)$ não depende do primeiro estágio.

Portanto, a distribuição marginal a priori de θ é :

$$p(\theta) = \int_{\Lambda} p(\theta, \lambda) d\lambda = \int_{\Lambda} p(\theta|\lambda) p(\lambda) d\lambda = \int_{\Lambda} \prod_{i=1}^n p(\theta_i|\lambda) p(\lambda) d\lambda$$

Esta divisão em estágios é um artifício probabilístico que auxilia a identificação e a especificação coerente da priori. A hierarquia pode ter mais de dois estágios, por exemplo, a distribuição de λ pode depender de ϕ . Dessa maneira, a priori marginal de θ é redefinida por :

$$p(\theta) = \int_{\Phi} \int_{\Lambda} p(\theta|\lambda) p(\lambda|\phi) p(\phi) d\lambda d\phi.$$

Os parâmetros λ e ϕ , que auxiliam a especificação da priori, são chamados de *hiperparâmetros*.

Teoricamente, a divisão em estágios pode prosseguir indefinidamente, porém, na prática torna-se difícil raciocinar e argumentar algo sobre os parâmetros do terceiro estágio. Consequentemente é comum assumir uma priori não-informativa para este estágio e todos os estágios subsequentes.

De uma forma geral, adotando-se a construção da estrutura hierárquica descrita acima e dentro do contexto de modelos lineares, para especificar o modelo com sua priori pode-se reescrevê-lo como:

$$Y|\beta_1, \phi \sim N(X_1\beta_1, \phi^{-1}I_n) \quad (4.1.1)$$

$$\beta_1|\beta_2, \phi \sim N(X_2\beta_2, \phi^{-1}C_1^{-1}) \quad (4.1.2)$$

$$\beta_2|\phi \sim N(\mu, \phi^{-1}C_2^{-1}) \quad (4.1.3)$$

$$n_0\sigma_0^2\phi \sim \chi_{n_0}^2 \quad (4.1.4)$$

Utilizando o modelo estrutural, que foi definido anteriormente em (3.1), e supondo o fator de atenuação conhecido, nosso interesse volta-se, nesta etapa, para analisarmos a robustez do estimador do parâmetro de inclinação, seguindo a definição de Robustez que Pericchi e Smith (1992) colocam em seu trabalho.

4.1.1 Função de Verossimilhança

A função de verossimilhança de β e σ^2 é expressa por:

$$L(\beta, \sigma^2) \propto \sigma^{-n} \exp\left\{-\frac{1}{2\sigma^2} \sum_{i=1}^n (Y_i - \beta K_x X_i)^2\right\} \quad (4.1.1.4)$$

e os estimadores de máxima verossimilhança de β e σ^2 são:

$$\hat{\beta} = K_x^{-1} \frac{\sum_{i=1}^n Y_i X_i}{\sum_{i=1}^n X_i^2} \quad \text{e} \quad \hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{\beta} K_x X_i)^2$$

Fazendo a transformação do parâmetro $\phi = 1/\sigma^2$, que denota a precisão, a função de verossimilhança de β e ϕ é dada por:

$$L(\beta, \phi) \propto \phi^{n/2} \exp\left\{-\frac{\phi}{2} \sum_{i=1}^n (Y_i - \beta K_x X_i)^2\right\} \quad (4.1.5)$$

4.2 Inferência Bayesiana com Priori Normal

Nesta seção, faremos uma inferência Bayesiana sobre o parâmetro de inclinação β sujeito a uma priori informativa seguindo a estrutura hierárquica definida logo acima, ou seja, a análise será submetida ao modelo de Regressão Hierárquico com dois estágios, para a priori

$$Y|Z, \beta, \phi \sim N[Z\beta, \phi^{-1}I] \quad (4.2.1)$$

$$\beta|\phi \sim N[\mu, \phi^{-1}] \quad (4.2.2)$$

$$n_0\sigma_0^2\phi \sim \chi^2_{(n_0)} \quad (4.2.3)$$

onde os parâmetros μ, n_0 e σ_0^2 são supostamente conhecidos; e denotamos por $Z' = K_X(X_1, X_2, \dots, X_n)$ e I a matriz Identidade.

De (4.2.2) temos que

$$p(\beta|\phi) \propto \phi^{1/2} \exp\left\{-\frac{\phi}{2}(\beta - \mu)^2\right\}. \quad (4.2.4)$$

De (4.2.3) temos que

$$p(\phi) \propto \phi^{\frac{n_0}{2}-1} \exp\left\{-\frac{n_0\sigma_0^2\phi}{2}\right\}. \quad (4.2.5)$$

Logo, podemos escrever a distribuição conjunta a priori de β e ϕ como :

$$p(\beta, \phi) \propto \phi^{\frac{n_0+1}{2}-1} \exp\left\{-\frac{\phi}{2}\left[n_0\sigma_0^2 + (\beta - \mu)^2\right]\right\}. \quad (4.2.6)$$

Acrescentando essas informações da priori à função de verossimilhança, obtemos a distribuição conjunta a posteriori dada por :

$$p(\beta, \phi|\text{dados}) \propto \phi^{\frac{n_0+n+1}{2}-1} \exp\left\{-\frac{\phi}{2}\left[n_0\sigma_0^2 + (\beta - \mu)^2 + n\hat{\sigma}^2 + (\beta - \hat{\beta})^2 \sum_{i=1}^n Z_i^2\right]\right\}$$

$$= \phi^{\frac{n_0+n+1}{2}-1} \exp \left\{ -\frac{\phi}{2} \left[n_0 \sigma_0^2 + n \hat{\sigma}^2 + \frac{\sum_{i=1}^n Z_i^2}{1 + \sum_{i=1}^n Z_i^2} (\mu - \hat{\beta})^2 + \left(1 + \sum_{i=1}^n Z_i^2 \right) (\beta - \hat{\beta}_1) \right] \right\}, \quad (4.2.7)$$

$$\text{onde } \hat{\beta}_1 = \frac{1}{1 + \sum_{i=1}^n Z_i^2} \left[\mu + \hat{\beta} \sum_{i=1}^n Z_i^2 \right], \quad (4.2.8)$$

Consequentemente, integrando (4.2.7) em relação a ϕ , a distribuição marginal a posteriori de β é formulada por :

$$p(\beta | \text{dados}) \propto \left[1 + \frac{h}{v} (\beta - \hat{\beta}_1)^2 \right]^{-\frac{v+1}{2}}, \quad (4.2.9)$$

$$\text{onde } h = \frac{v \left(1 + \sum_{i=1}^n Z_i^2 \right)}{n_0 \sigma_0^2 + n \hat{\sigma}^2 + \frac{\sum_{i=1}^n Z_i^2}{1 + \sum_{i=1}^n Z_i^2} (\mu - \hat{\beta})^2} \quad (4.2.10)$$

$$\text{e } v = n + n_0$$

Da expressão (4.2.9) podemos notar que a densidade a posteriori marginal de β tem a distribuição t de Student com parâmetro de escala h , parâmetro de forma v (graus de liberdade) e o parâmetro de locação é dado por:

$$E[\beta | \text{dados}] = \hat{\beta}_1 = \frac{1}{1 + \sum_{i=1}^n Z_i^2} \mu + \frac{\sum_{i=1}^n Z_i^2}{1 + \sum_{i=1}^n Z_i^2} \hat{\beta} = \frac{1\mu + \sum_{i=1}^n Z_i^2 \hat{\beta}}{1 + \sum_{i=1}^n Z_i^2} \quad (4.2.11)$$

e variância a posteriori é dada por:

$$\text{Var}[\beta | \text{dados}] = \frac{v}{h(n + n_0 - 2)} \quad (4.2.12)$$

Observe que a média a posteriori, dada pela equação (4.2.11), é uma média ponderada da média conhecida a priori μ e o estimador de máxima verossimilhança $\hat{\beta}$.

Para melhor podermos interpretar qual é a influência da média a priori sobre a média a posteriori, reformulamos a expressão (4.2.11) da seguinte forma:

$$\hat{\beta}_1 - \hat{\beta} = \frac{1}{1 + K_x^2 \sum_{i=1}^n X_i^2} (\mu - \hat{\beta})$$

e, de uma forma genérica, descrevemos o gráfico logo abaixo.

Figura 4.1: Discrepância entre a média a priori e o estimador de máxima verossimilhança contra a discrepância entre a média a posteriori e o estimador de máxima verossimilhança

É interessante notar que, como a diferença $(\mu - \hat{\beta})$ não é limitada, para todas as possíveis escolhas de μ , a diferença $E[\beta|\text{dados}] - \hat{\beta}$ também não é limitada. Assim, as discrepâncias entre a média priori μ e o estimador de máxima verossimilhança $\hat{\beta}$ pode levar a média a posteriori $E[\beta|\text{dados}]$ a afastamentos ilimitados de $\hat{\beta}$, violando, desta forma, a robustez, segundo Pericchi e Smith.

Observe que, para o caso em que o fator de atenuação $K_x = 1$ (isto é, sem erro de mensuração) e $X_i = 1$, $i = 1, 2, \dots, n$, o resultado é coincidente com o resultado obtido por Pericchi e Smith (1992). Observe, também, que o acréscimo no erro de mensuração aumenta a falta de robustez.

Podemos concluir que a propriedade técnica atrativa da priori Normal no modelo, dadas pelas formas (4.2.1), (4.2.2) e (4.2.3) é somente de importância secundária comparada à influência da média a priori sobre a média e variância a posteriori. Dessa forma, podemos nos indagar sobre qual tipo de priori devemos substituir a priori Normal, de tal maneira que a posteriori fosse robusta, no sentido que a priori não tenha tanta influência sobre a posteriori. Buscando a resposta para esta pergunta, isto será discutido em detalhes na próxima seção.

4.3 Inferência Bayesiana com Priori t de Student

Com o objetivo de obter uma densidade a posteriori robusta para o parâmetro de inclinação, de tal forma que a discrepância entre a média a posteriori e o estimador de máxima verossimilhança não possa ser tão fortemente influenciada pela discrepância entre a média a priori e o estimador de máxima verossimilhança, Pericchi sugere em seu artigo que seja utilizada a priori t de Student ao invés da priori Normal.

$$Y|Z, \beta, \phi \sim N[Z\beta, \phi^{-1}I] \quad (4.3.1)$$

$$\beta|\phi \sim t[\mu, \phi^{-1}] \quad (4.3.2)$$

$$n_0\sigma_0^2\phi \sim \chi^2_{(n_0)} \quad (4.3.3)$$

onde os parâmetros μ , n_0 e σ_0^2 são supostamente conhecidos.

Do modelo acima definido, é fácil verificar, calculando de modo análogo à seção anterior, que a priori conjunta para β e ϕ e sua respectiva posteriori são dadas por:

$$p(\beta, \phi) \propto \phi^{\frac{n_0+1}{2}-1} \exp\left\{-\frac{n_0\sigma_0^2\phi}{2}\right\} \left[1 + \frac{\phi}{n}(\beta - \mu)^2\right]^{-\frac{n+1}{2}} \quad (4.3.4)$$

$$p(\beta, \phi|\text{dados}) \propto \phi^{\frac{n+n_0+1}{2}-1} \exp\left\{-\frac{\phi}{2}\left[n_0\sigma_0^2 + \sum_{i=1}^n (Y_i - \beta Z_i)^2\right]\right\} \left[1 + \frac{\phi}{n}(\beta - \mu)^2\right]^{-\frac{n+1}{2}} \quad (4.3.5)$$

Para se obter a densidade a posteriori marginal de β basta integrar (4.3.5) com relação a ϕ , porém, isto não é tão simples; assim, para auxiliar o cálculo da densidade a posteriori utilizaremos um resultado bem conhecido de misturas de distribuições que colocamos a seguir em um Lema.

Lema 4.3.1 (Berger 1985) Se β tem uma distribuição t de Student, com parâmetro de localização μ , parâmetro de escala ϕ e n graus de liberdade, definido por:

$$t(\beta|\mu, \phi, n) = \frac{\sqrt{\phi} \Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left[1 + \phi \frac{(\beta - \mu)^2}{n}\right]^{-\frac{n+1}{2}} \quad (4.3.6)$$

$$\text{então } t(\beta|\mu, \phi, n) = \int_0^\infty N\left(\beta|\mu, \frac{1}{\phi\lambda}\right) G\left(\lambda|\frac{n}{2}, \frac{2}{n}\right) d\lambda, \quad (4.3.7)$$

onde $N(\cdot)$ e $G(\cdot)$ representam a função densidade de probabilidade Normal e Gamma, respectivamente.

$$\text{Prova : Seja } N\left(\beta|\mu, \frac{1}{\phi\lambda}\right) = \frac{1}{\sqrt{2\pi\phi\lambda}} \exp\left\{-\frac{1}{2\phi\lambda}(\beta - \mu)^2\right\}$$

$$\text{e } G\left(\lambda|\frac{n}{2}, \frac{2}{n}\right) = \frac{\lambda^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\left(\frac{2}{n}\right)^{\frac{n}{2}}} \exp\left\{-\frac{\lambda n}{2}\right\}.$$

Logo,

$$\begin{aligned} \int_0^\infty N\left(\beta|\mu, \frac{1}{\phi\lambda}\right) G\left(\lambda|\frac{n}{2}, \frac{2}{n}\right) d\lambda &= \int_0^\infty \frac{\sqrt{\phi\lambda}}{\sqrt{2\pi}} \exp\left\{-\frac{\phi\lambda(\beta - \mu)^2}{2}\right\} \frac{\lambda^{\frac{n}{2}-1}}{\Gamma\left(\frac{n}{2}\right)\left(\frac{2}{n}\right)^{\frac{n}{2}}} \exp\left\{-\frac{\lambda n}{2}\right\} d\lambda \\ &= \frac{\sqrt{\phi} n^{\frac{n}{2}}}{\sqrt{2\pi} 2^{\frac{n}{2}} \Gamma\left(\frac{n}{2}\right)} \underbrace{\int_0^\infty \lambda^{\frac{n+1}{2}-1} \exp\left\{-\frac{\lambda}{2} [\phi(\beta - \mu)^2 + n]\right\} d\lambda}_{G\left(\lambda|\frac{n+1}{2}, \frac{[\phi(\beta - \mu)^2 + n]}{2}\right)} \\ &= \frac{\sqrt{\phi} n^{\frac{n}{2}}}{\sqrt{2\pi} 2^{\frac{n}{2}} \Gamma(n)} \Gamma\left(\frac{n+1}{2}\right) \left\{\frac{[\phi(\beta - \mu)^2 + n]}{2}\right\}^{-\frac{n+1}{2}} = \frac{\sqrt{\phi} \left(\frac{n}{2}\right)^{\frac{n}{2}} \Gamma\left(\frac{n+1}{2}\right)}{\sqrt{2\pi} \Gamma(n)} \left\{\frac{n}{2} \left[1 + \frac{\phi(\beta - \mu)^2}{n}\right]\right\}^{-\frac{n+1}{2}} \end{aligned}$$

$$= \frac{\sqrt{\phi} \Gamma\left(\frac{n+1}{2}\right)}{\sqrt{\pi n} \Gamma\left(\frac{n}{2}\right)} \left\{ 1 + \frac{\phi(\beta - \mu)^2}{2} \right\}^{-\frac{n+1}{2}}$$

□

Dessa forma a priori conjunta para β e ϕ , utilizando o Lema 4.3.1 e as densidades a priori dadas por (4.3.2) e (4.3.3), é:

$$p(\beta, \phi) \propto \phi^{\frac{n_0}{2}-1} \exp\left\{-\frac{n_0 \sigma_0^2 \phi}{2}\right\} \int_0^\infty N\left(\beta \mid \mu, \frac{1}{\phi \lambda}\right) G\left(\lambda \mid \frac{n}{2}, \frac{2}{n}\right) d\lambda \quad (4.3.8)$$

Juntando a informação proveniente da função de verossimilhança temos a densidade a posteriori de β e ϕ dada por:

$$p(\beta, \phi \mid \text{dados}) \propto \phi^{\frac{n_0+n}{2}-1} \left[\int N\left(\beta \mid \mu, \frac{1}{\phi \lambda}\right) G\left(\lambda \mid \frac{n}{2}, \frac{2}{n}\right) d\lambda \right] \exp\left\{-\frac{\phi}{2} \left[n_0 \sigma_0^2 + n \hat{\sigma}^2 + \sum_{i=1}^n Z_i^2 (\beta - \hat{\beta})^2 \right]\right\} \quad (4.3.9)$$

Integrando (4.3.9) com respeito a ϕ e trocando a ordem de integração, a densidade a posteriori marginal de β resultante é:

$$p(\beta \mid \text{dados}) \propto \int_0^\infty \lambda^{\frac{1}{2}} G\left(\lambda \mid \frac{n}{2}, \frac{2}{n}\right) \left[m_\lambda + \left(\lambda + \sum_{i=1}^n Z_i^2 \right) (\beta - \hat{\beta}_1(\lambda))^2 \right]^{-\frac{n_0+n+1}{2}} d\lambda \quad (4.3.10)$$

onde

$$m_\lambda = n_0 \sigma_0^2 + n \hat{\sigma}^2 + \frac{\lambda \sum_{i=1}^n Z_i^2}{\lambda + \sum_{i=1}^n Z_i^2} (\mu - \hat{\beta})^2$$

e

$$\hat{\beta}_1(\lambda) = \frac{1}{\lambda + \sum_{i=1}^n Z_i^2} \left[\lambda \mu + \sum_{i=1}^n Z_i^2 \hat{\beta} \right]$$

Na tentativa de calcular a posteriori marginal de β , não dispomos de recursos que nos permita achar uma expressão exata; assim utilizaremos as aproximações que serão fornecidos pelos próximos dois lemas.

Lema 4.3.2 Se $\sigma_0^2 \geq \frac{(\mu - \beta)^2 \sum_{i=1}^n Z_i^2 - n \hat{\sigma}^2}{n_0}$

então

$$m_\lambda^{-\frac{n_0+n+1}{2}} \cong [n_0 \sigma_0^2 + n \hat{\sigma}^2]^{-\frac{n_0+n+1}{2}} \exp \left\{ -\frac{(\mu - \hat{\beta})^2}{2 \left(\frac{b_n}{\lambda} + a_n \right)} \right\} \quad (4.3.11)$$

onde $b_n = \frac{n_0 \sigma_0^2 + n \hat{\sigma}^2}{n_0 + n + 1}$ e

$$a_n = \frac{b_n}{\sum_{i=1}^n Z_i^2}$$

Prova: Podemos verificar que a desigualdade $\sigma_0^2 \geq \frac{(\mu - \beta)^2 \sum_{i=1}^n Z_i^2 - n \hat{\sigma}^2}{n_0}$ é equi-

valente a

$$(\mu - \beta)^2 \leq \frac{n_0 \sigma_0^2 + n \hat{\sigma}^2}{\sum_{i=1}^n Z_i^2}. \quad (4.3.12)$$

Como a razão $\frac{\left(\lambda + \sum_{i=1}^n Z_i^2\right)}{\lambda} \geq 1$ para quaisquer valores de λ e $\sum_{i=1}^n Z_i^2$, então (4.3.12) é

equivalente a

$$(\mu - \beta)^2 \leq \frac{\left(\lambda + \sum_{i=1}^n Z_i^2\right)}{\lambda} \frac{n_0 \sigma_0^2 + n \hat{\sigma}^2}{\sum_{i=1}^n Z_i^2}. \quad (4.3.13)$$

Usando o fato que a aproximação $\ln(1+x) \cong x$ é válida, para $|x| < 1$, temos que o resultado (4.3.11) é imediato.

□

Para utilizarmos o resultado do próximo lema, é conveniente colocarmos a seguinte notação:

$$r\left(\frac{b_n}{\lambda} + z\right) = \exp\left\{-\frac{(\mu - \hat{\beta})^2}{2\left(\frac{b_n}{\lambda} + z\right)}\right\} \quad (4.3.14)$$

onde $z = a_n$.

$$\textbf{Lema 4.3.3} \quad r\left(\frac{b_n}{\lambda} + a_n\right) = r\left(\frac{b_n}{\lambda}\right) \left[1 + \frac{\lambda^2(\mu - \hat{\beta})^2}{2 b_n \sum_{i=1}^n Z_i^2} + O(n^{-2})\right] \quad (4.3.15)$$

Prova:

Usando a aproximação de séries de Taylor, isto é, expandindo $r\left(\frac{b_n}{\lambda} + z\right)$ em

(4.3.14) em torno de $z = 0$, obtendo assim

$$r\left(\frac{b_n}{\lambda} + a_n\right) = r\left(\frac{b_n}{\lambda}\right) + r'\left(\frac{b_n}{\lambda}\right) a_n + O(n^{-2}). \quad (4.3.16)$$

É fácil verificar, a partir de (4.3.14) e aplicando a aproximação de Taylor (4.3.16), o resultado (4.3.15). □

Utilizando os Lemas 4.3.2 e 4.3.3, obtemos uma aproximação para a densidade a posteriori marginal (4.3.10), assim, em decorrência disto temos:

$$p(\beta|\text{dados}) \propto \int_0^\infty G\left(\lambda \left| \frac{n}{2}, \frac{2}{n} \right.\right) \lambda^{\frac{1}{2}} \exp\left\{-\frac{\lambda}{2b_n}(\mu - \hat{\beta})^2\right\} \left[1 + \frac{\left(\lambda + \sum_{i=1}^n Z_i^2\right)(\beta - \hat{\beta}_1(\lambda))^2}{m_\lambda}\right]^{-\frac{n_0+n+1}{2}} \left[1 + \frac{\lambda^2(\mu - \hat{\beta})^2}{2b_n \sum_{i=1}^n Z_i^2}\right] d\lambda \quad (4.3.17)$$

Para eliminarmos o termo λ na densidade a posteriori marginal de β a integral acima (4.3.17) pode ser aproximada pelo método de Laplace, no qual

$$p(\beta|\text{dados}) = \int_0^\infty \exp\{-nh(\lambda)\} f(\lambda, \beta) d\lambda \cong \left(\frac{2\pi\sigma^2}{n}\right)^{\frac{1}{2}} f(\hat{\lambda}, \beta) \exp\{-nh(\hat{\lambda})\}$$

$$\text{onde } -nh(\lambda) = \frac{(n-1)}{2} \ln(\lambda) - \frac{n\lambda}{2} - \frac{\lambda(\mu - \hat{\beta})^2}{2b_n}; \quad (4.3.18)$$

$$f(\lambda, \beta) = \left[1 + \frac{\lambda^2(\mu - \hat{\beta})^2}{2b_n \sum_{i=1}^n Z_i^2}\right] \left[1 + \frac{\left(\lambda + \sum_{i=1}^n Z_i^2\right)(\beta - \hat{\beta}_1(\lambda))^2}{m_\lambda}\right]^{-\frac{n_0+n+1}{2}} \quad (4.3.19)$$

$$\text{e } \sigma^2 = h''(\hat{\lambda})^{-1}$$

Dessa forma, a densidade a posteriori marginal de β resulta em:

$$p(\beta|dados) \propto \left[1 + \frac{\left(\hat{\lambda} + \sum_{i=1}^n Z_i^2 \right)}{m_{\hat{\lambda}}} (\beta - \hat{\beta}_1(\hat{\lambda}))^2 \right]^{-\frac{n_0+n+1}{2}} \quad (4.3.20)$$

onde $\hat{\lambda}$ é o ponto de mínimo de $h(\lambda)$, determinado como solução de

$$\left. \frac{\partial [-nh(\lambda)]}{\partial \lambda} \right|_{\lambda = \hat{\lambda}} = 0$$

e pode ser colocado na forma explícita

$$\hat{\lambda} = \frac{(n-1)b_n}{(\mu - \hat{\beta})^2 + nb_n} \quad (4.3.21)$$

Note que, da equação (4.3.20), a distribuição a posteriori de β é t de Student, com parâmetro de locação $\hat{\beta}_1(\hat{\lambda})$, parâmetro de escala ϕ^* e $(n_0 + n)$ graus de liberdade,

$$\text{onde } \phi^* = \frac{(n_0 + n) \left(\hat{\lambda} + \sum_{i=1}^n Z_i^2 \right)}{m_{\hat{\lambda}}} \quad (4.3.22)$$

A média a posteriori de β pode ser colocada da seguinte forma:

$$\hat{E}[\beta|dados] = \hat{\beta}_1(\hat{\lambda}) = \frac{1}{\hat{\lambda} + \sum_{i=1}^n Z_i^2} \left[\hat{\lambda} \mu + \hat{\beta} \sum_{i=1}^n Z_i^2 \right] \quad (4.3.23)$$

e a variância a posteriori de β

$$\hat{Var}[\beta|dados] = \frac{n_0 + n}{(n_0 + n - 2)\phi^*} \quad (4.3.24)$$

4.4 Exemplo Numérico

Uma ilustração numérica é importante e necessária para verificarmos o quão eficiente está sendo a aproximação utilizada para determinar a densidade a posteriori (4.3.20).

Na tabela 4.1 apresentamos os valores da média a posteriori exata seguindo (4.3.17) e os valores da média a posteriori utilizando as aproximações seguindo (4.3.23), para $n = n_0 = 10$, fixando alguns valores para a média a priori μ e para o coeficiente de atenuação K_X . Sob certas particularidades, uma vez que o modelo que abordamos é mais genérico, estes valores foram estudados por Pericchi & Smith (1992) e Fuller (1987).

Ao determinar a densidade a posteriori (4.3.20) usufruímos de um resultado dado pelo lema 4.3.2 e necessitamos atender a condição deste lema, assim para obtermos uma boa aproximação para grandes divergências entre μ e $\hat{\beta}$, escolhemos $\sigma_0^2 = 50$.

Tabela 4.1: Média a posteriori exata e média a posteriori aproximada para $n = n_0 = 10$, $\sigma_0^2 = \hat{\sigma}^2 = 50$, $\hat{\beta} = 0$ e $X_i = 1, i = 1, 2, \dots, 10$.

μ	$K_X = 1.0$		$K_X = 0.8$		$K_X = 0.5$	
	$E(\beta \text{dados})$	$\hat{E}(\beta \text{dados})$	$E(\beta \text{dados})$	$\hat{E}(\beta \text{dados})$	$E(\beta \text{dados})$	$\hat{E}(\beta \text{dados})$
0.5	0.1903	0.1887	0.1563	0.1544	0.1078	0.1054
5.0	0.1869	0.1864	0.1533	0.1524	0.1058	0.1038
10.0	0.1789	0.1809	0.1462	0.1477	0.1002	0.1000

Com as informações colhidas da tabela acima, podemos concluir que a média a posteriori $\hat{E}(\beta | \text{dados})$ dada por (4.3.23) é uma boa aproximação para a média a posteriori exata $E(\beta | \text{dados})$.

Para que possamos analisar graficamente o comportamento da média a posteriori a respeito da robustez que estamos interessados, reescrevemos (4.3.23) da seguinte maneira:

$$\beta_1(\hat{\lambda}) - \hat{\beta} = \frac{1}{\hat{\lambda} + \sum_{i=1}^n Z_i^2} \left[\hat{\lambda} \mu + \hat{\beta} \sum_{i=1}^n Z_i^2 \right] - \hat{\beta} = \frac{\hat{\lambda} (\mu - \hat{\beta})}{\hat{\lambda} + \sum_{i=1}^n Z_i^2} \text{ e substituindo (4.3.21) na}$$

expressão acima, temos:

$$\beta_1(\hat{\lambda}) - \hat{\beta} = \frac{(n-1) b_n (\mu - \hat{\beta})}{b_n \left[(n-1) + n K_X^2 \sum_{i=1}^n X_i^2 \right] + K_X^2 \sum_{i=1}^n X_i^2 (\mu - \hat{\beta})^2}. \quad (4.3.25)$$

Também, substituindo (4.3.21) e (4.3.22) em (4.3.24) temos a variância a posteriori

$$\begin{aligned} \text{Var}[\beta|\text{dados}] = & \frac{(n_0 \sigma_0^2 + n \hat{\sigma}^2) \left[(\mu - \hat{\beta})^2 + n b_n \right] + \frac{(n-1) b_n K_X^2 \sum_{i=1}^n X_i^2 (\mu - \hat{\beta})^2 \left[(\mu - \hat{\beta})^2 + n b_n \right]}{(n-1) b_n + K_X^2 \sum_{i=1}^n X_i^2 \left[(\mu - \hat{\beta})^2 + n b_n \right]} \\ & (n_0 + n - 2) \left\{ b_n (n-1) + K_X^2 \sum_{i=1}^n X_i^2 \left[(\mu - \hat{\beta})^2 + n b_n \right] \right\} \end{aligned} \quad (4.3.26)$$

Figura 4.2: Divergência da média a posteriori versus divergência da média a priori em relação ao estimador de máxima verossimilhança, para fator de atenuação fixado em 0.5, 0.8 e 1.0.

Figura 4.3: Variância a posteriori versus divergência da variância a priori em relação ao estimador de máxima verossimilhança, para fator de atenuação fixado em 0.5, 0.8 e 1.0

Figura 4.4: Densidade a posteriori exata dada por (4.3.15) e aproximada pelo método de Laplace dada por (4.3.18), para o fator de atenuação fixado em 1.0.

Figura 4.5: Densidade a posteriori exata dada por (4.3.15) e aproximada pelo método de Laplace dada por (4.3.18), para o fator de atenuação fixado em 0.5.

Conclusão

Para usar a abordagem Bayesiana para um determinado problema, é importante conhecer as implicações de uma particular escolha da distribuição a priori. Se estas implicações são inaceitáveis então a escolha para a utilização da priori é injustificável. Nas discussões do artigo de Lindley (1968), Beale e Lindley concordam que a discrepância entre a média a posteriori $E[\beta|\text{dados}]$ e o estimador de máxima verossimilhança $\hat{\beta}$ ser proporcional à discrepância entre a média a priori μ o estimador de máxima verossimilhança $\hat{\beta}$, muitas vezes é uma implicação indesejável. Esta propriedade indesejável da distribuição a priori surge porque existe uma sensibilidade que a discrepância entre a média a priori e os dados pode, de certo modo, “duvidar” do valor a priori μ . Neste caso, o pesquisador Bayesiano deve encontrar uma distribuição a priori apropriada tal que a discrepância entre a média a posteriori e o estimador de máxima verossimilhança é aproximadamente zero. Lindley realizou uma análise aproximada que indica que o comportamento indesejável seria evitado se a distribuição a priori tem a forma da t de Student, e isto é confirmado pelos resultados que estão logo a seguir.

As figuras 4.2 e 4.3 mostram as formas qualitativas de dependência da média a posteriori e variância a posteriori sobre a divergência entre a média a priori do parâmetro de inclinação e o estimador de mínimos quadrados $\hat{\beta}$, para valores do fator de atenuação $K_x = 1.0, 0.8$ e 0.5 ; $b_n = 1$; $n = n_0 = 10$; $X_i = 1$ ($i = 1, 2, \dots, 10$) e $\nu = 20$.

O uso da priori t de Student produz uma análise robusta para o parâmetro de locação como foi sugerido por muitos autores; veja, por exemplo, Pericchi & Smith (1992), Dawid (1973), West (1981), Lindley (1968) e Smith (1983).

Para $K_x = 1$, ou seja, sem erro de mensuração, a dependência é similar à dependência encontrado por Pericchi & Smith (1992). Porém, próximo a zero, a influência da média a priori cresce sobre a média a posteriori quando o fator de atenuação K_x decresce. As figuras 4.2 e 4.3 mostram que é importante ter em mente os erros de

medida quando estudamos a análise de robustez Bayesiana. É interessante observar que, pela forma da figura 4.2, quando a diferença absoluta $|\mu - \hat{\beta}|$ cresce, a priori t de Student produz uma redução na diferença entre a média a posteriori e o estimador de máxima verossimilhança, ou seja, quanto mais a média a priori diverge do que os dados informam (e.m.v.), a média a posteriori e o e.m.v. tendem a convergir para um mesmo valor, isto significa que usando a priori t de Student, a informação proveniente dos dados é enfatizada em relação à informação a priori; concordando, assim, com o Princípio da Verossimilhança. Conclusão análoga foi obtida por Pericchi & Smith (1992) supondo um modelo Normal utilizando variância conhecida. A variância a posteriori, ilustrada pela figura 4.3, é monotônica em $|\mu - \hat{\beta}|$. Esses aspectos qualitativos da média e variância a posteriori oferece alguma direção em Análise Bayesiana Robusta.

Pelas figuras 4.4 e 4.5 podemos notar que a aproximação de Laplace está sendo bastante eficiente, no cálculo da densidade a posteriori do parâmetro de inclinação, e que o aumento no erro de mensuração afeta a aproximação de Laplace, porém de forma insignificante.

Capítulo 5

Propostas Futuras

Como continuação do nosso trabalho pretendemos abordar os seguintes tópicos , surgidos no desenvolvimento desta pesquisa:

- (1) No Capítulo 3, quando trabalhamos com o modelo estrutural, onde o fator de atenuação K_x é desconhecido, nos deparamos com um problema o qual a densidade a posteriori não fornece boas estimativas para o nosso parâmetro de interesse. Portanto um procedimento a ser feito é verificar a possibilidade de utilizar o método de Gibb Sampling.
- (2) Desenvolver a Inferência Bayesiana para um Modelo Funcional mais geral, como formulado por Reilly & Patino-Leal (1981).
- (3) Analisar sob o ponto de vista clássico o modelo estrutural via parâmetros ortogonais. A partir disso, realizar a Inferência Bayesiana e, conseqüentemente, comparar com os resultados obtidos por Wong (1989).

- (4) Desenvolver a Inferência Fiducial para o Modelo Estrutural quando o fator de atenuação é desconhecido, partindo dos resultados assintóticos que foram obtidos por Rodrigues & Cordani (1990).

Bibliografia

ACHCAR, J.A. *Alguns Aspectos de Inferência Bayesiana Aproximada*. São Carlos , 1988.

Tese (Livre Docência) - Instituto de Ciências Matemáticas de São Carlos, Universidade de São Paulo.

_____ Reparametrization and Laplace Approximations for Posterior Moments in Binomial Models. *Revista de Matemática e Estatística*, São paulo, v.7, p.73-86, 1989.

_____ Reparametrization and accuracy of Laplace approximations for posterior moments. *Revista de Matemática Estatística*, São Paulo, v.8, p.23-30, 1990.

_____; SMITH, A.F.M. Aspects of Reparametrization in Approximate Bayesian Inference. *Bayesian and Likelihood Methods in Statistics and Econometrics*, North-Holland, Elsevier Science Publishers B.V., 1990.

ADCOCK, R.J. Note on the method of least squares. *The Analyst*, n.4, p.183-4, 1877.

_____ A problem in least squares. *The Analyst*, n.5, p.53-4, 1878.

ASSOCIAÇÃO BRASILEIRA DE NORMAS TÉCNICAS. Comissão de Estudos de Documentação. Rio de Janeiro. *Referências Bibliográficas*. Rio de Janeiro, ago. 1989.

- BERGER, J.O. *Statistical Decision Theory and Bayesian Analysis*. New York, Springer Verlag, 1985.
- BOLFARINE, H.; NASCIMENTO, J.A.; RODRIGUES, J. *On the estimation of the true x values in a simple regression model when both variables have error*. São Paulo, RT-MAE-9117/USP, 1991. (Relatório Técnico RT-MAE/USP. AMS62J99)
- BOLFARINE, H.; RODRIGUES, J.; CORDANI, L. *O modelo de Regressão com erros nas variáveis*. In: SIMPÓSIO NACIONAL DE PROBABILIDADE E ESTATÍSTICA, 10., Universidade do Rio de Janeiro, 1992. Minicurso. Rio de Janeiro, Associação Brasileira de Estatística, 1992. 90p.
- BOX, G.E.P.; TIAO, G.C. *Bayesian Inference in Statistical Analysis*. Massachusetts, Addison-Wesley Publishing Company, 1973. 588p.
- CASELLA, G.; GEORGE, E.I. Explaining the Gibbs Sampler. *The American Statistician*, v.46, n.3, p.167-74. Aug, 1992.
- COX, D.R.; REID, N. Parameter orthogonality and approximate conditional inference. *Journal of the Royal Statistical Society*, v.49, p.1-39, 1987. (Série B)
- DAWID, A.P. Posterior expectations for large observations. *Biometrika*, Great Britain, v.60, p.664-6, 1973.
- EDWARDS, W.; LINDMAN, H.; SAVAGE, L.J. Bayesian Statistical Inference for Psychological Research. *Psychological Rev.* 70,193, 1963.
- FULLER, W.A. *Measurement Error Models*. New York, J. Wiley, 1987. 439p.

- GEARY, R.C. Determinations of linear relations between systematic parts of variables with errors of observations the variances of which are Unknown. *Econometrica*, 17, p.30-58, 1949.
- KALBFLEISCH, J.D. *Probability and Statistical Inference*. New York, Springer Verlag, v.2, 1985. 360p.
- KASS, R.E.; TIERNEY, L.; KADANE, J.B. *Laplace method in Bayesian analysis*. Unpublished report, 1990.
- KENDALL, M.G. Regression, Structure and Functional Relationship I. *Biometrika*, Great Britain, v.38, p.11-25, 1951.
- _____ Regression, Structure and Functional Relationship II. *Biometrika*, Great Britain, v.39, p.96-108, 1952.
- _____; STUART, A. *The Advanced Theory of Statistics*. Griffin, v.2, 1966.
- KIEFER, J.; WOLFOWITZ, J. Consistency of the maximum likelihood estimator in the presence of infinitely many incidental parameters. *Ann. Math. Statistics*, n.27, p.887-906, 1956.
- KUMMEL, C.H. Reduction of observed equations which contain more than one observed quantity. *The Analyst*, n.6, p.97-105, 1879.
- LINDLEY, D.V. Regression lines and the functional relationship. *J. Roy. Statist. Soc.*, n.2, p.218-44, 1947. Supplement 9.
- _____ *Introduction to Probability and Statistics from a Bayesian Viewpoint*, part 2, Inference, Cambridge, Cambridge University Press, 1965.

- _____ The choice of variables in multiple regression. *J. Roy. Statist. Soc.*, 1968.
- _____ *Bayesian Statistics, a Review*. Pennsylvania, Society for Industrial and Applied Mathematics, 1972. 83p.
- _____ Bayes Estimates for the Linear Model. *J. Roy. Statist. Soc.*, v.34, n.1, p.1-41, 1972.
- _____; EL-SAYAD, G.M. The Bayesian estimation of a linear functional relationship. *J. Roy. Statist. Soc.*, v.30, n.1, p.190-202, 1968. (Série B).
- MADANSKY, A. The fitting of straight lines when both variables are subject to error. *J. Amer. Statist. Assoc.*, v.54, p.173-205, March, 1959.
- NEYMAN, J.; SCOTT, E.L. Consistent estimates based on partiality consistent observations. *Econometrica*, n.16, p.1-32, 1948.
- _____ On certain methods of estimating the linear structural relation. *Ann. Math. Statist.*, n.22, p.352-61, 1951.
- NETER, J.; WASSERMAN, W. *Applied Linear Statistical Models*. Richard D. Irwin, Inc., 1974.
- O' HAGGAN, A. On outlier rejection phenomena in Bayes inference. *J. Roy. Statist. Soc.*, v.41, p.358-67, 1979.
- PEARSON, K. On lines and planes of closest fit to systems of points in space. *Philosophical Magazine*, 2, p.559-72, 1901.
- PERICCHI, L.R.; SMITH, A.F.M. Exact and approximate posterior moments for a normal location parameter. *J. Roy. Statist. Soc.*, v.54, n.3, p.793-804, 1992. (Série B)

- _____ ; WALLEY, P. Robust Bayesian credible intervals and prior ignorance. *Int. Statist. Rev.*, 59, p.1-24, 1991.
- REILLY, P.M.; PATINO-LEAL, H. A Bayesian Study of the Error-in-Variables Model. *Technometrics*, v.23, n.3, p.221-31, Aug., 1981.
- REIERSØL, O. Identifiability of a linear relation between variables wich are subject to error. *Econometrica*, 18, p.375-89, 1950.
- RODRIGUES, J. A Bayesian analysis of the generalized least-squares procedure for functional relationships. 1990. (no prelo).
- _____ ; BOLFARINE, H. *Bayesian Estimation of the slope of a linear functional regression model*. São Carlos, n.72, ISSN/USP, 1990. (Relatório Técnico ISSN 0103-2577)
- _____ ; CORDANI, L.K. A Note on likelihood estimation of a simple regression model with measurement error, via the orthogonal parametrization. *South African Stat. J.*, n.24, p.177-83, 1990.
- SEBER, G.A.F. *Linear Regression Analysis*. New York, J. Wiley, 1977. 465p.
- SMITH, A.F.M. Bayesian approaches to outliers and robustness. In *Specifying Statistical Models* (eds J.P. Florens, M. Mouchart, J.P. Raoult, L. Simar and A.F.M. Smith), New York, Springer, p.13-55, 1983.
- SPRENT, P. A generalized least-squares approach to linear functional relationships. *J. Roy. Statist. Soc.*, v.28, p.278-97, 1966. (Série B).

- _____ Some historical of functional and structural relationships. *Comtemporary Mathematics*, v.112, p.3-15, 1990.
- TIBSHIRANI, R. Noninformative priors for one parameter of many. *Biometrika*, Great Britain, v.76, n.3, p.604-8, 1989.
- TIERNEY, L.; KADANE, J.B. Accurate approximations for posterior moments and marginal densities. *Journal of the American Statistical Association*, v.81, p.82-6, 1986.
- UNIVERSIDADE DE SÃO PAULO. Prefeitura do Campus Administrativo de Ribeirão Preto. Serviço de Biblioteca e Documentação. *Referências e Citações Bibliográficas: Manual de Orientação*. Ribeirão Preto, 1994.
- WEST, M. Robust sequential approximate Bayesian estimation. *J. Roy. Statist. Soc.*, v.43, p.157-66, 1981. (Série B).
- WONG, M.Y. Likelihood estimation of a siple linear regression model when both variables have error. *Biometrika*, Great Britain, v.76, n.1, p.141-8, 1989.
- ZELLNER, A. *An Introduction to Bayesian Inference in Econometrics*. New York, J. Wiley, 1971. 431p.

APÊNDICE A

Ortogonalidade de parâmetros

No nosso trabalho utilizamos o conceito da ortogonalização de parâmetros e a reparametrização ortogonal de parâmetros que foi baseado em Cox & Reid (1987), onde estudam a ortogonalização de parâmetros de uma forma genérica, explorando a conexão entre ortogonalidade de parâmetros e a teoria assintótica da inferência condicional.

Considere uma variável aleatória X com função densidade de probabilidade $f(x|\phi)$ onde $\phi = (\phi_1, \dots, \phi_p)$ é um vetor de parâmetros desconhecidos e $L(\phi|x)$ a função de verossimilhança. Considere, também, $\phi = (\underline{\delta}, \underline{\lambda})$ uma partição onde $\underline{\delta} = (\delta_1, \delta_2, \dots, \delta_k)$ e $\underline{\lambda} = (\lambda_1, \lambda_2, \dots, \lambda_{p-k})$.

Definição: O vetor de parâmetros $\underline{\delta}$ é ORTOGONAL ao vetor de parâmetros $\underline{\lambda}$ se os elementos da matriz de Informação de Fisher $I(\phi)$ satisfazem a seguinte condição:

$$i_{\delta_s, \lambda_t} = E_{\phi} \left[- \frac{\partial^2 \log L(\phi|x)}{\partial \delta_s \partial \lambda_t} \right] = 0 \quad (A1)$$

para $s = 1, 2, \dots, k$ e $t = 1, 2, \dots, p - k$.

Levando em conta a partição feita no parâmetro $\underline{\delta}$, a matriz de Informação de Fisher $I(\phi)$ é dada por:

$$I(\underline{\delta}, \underline{\lambda}) = \begin{bmatrix} I(\underline{\delta}, \underline{\delta}) & 0_1 \\ 0_2 & I(\underline{\lambda}, \underline{\lambda}) \end{bmatrix} \quad (\text{A2})$$

onde $I(\underline{\delta}, \underline{\delta})$ e $I(\underline{\lambda}, \underline{\lambda})$ são matrizes de Informação de Fisher de tamanho $k \times k$ e $(p-k) \times (p-k)$ respectivamente e 0_1 e 0_2 são matrizes nulas de tamanho $(p-k) \times (p-k)$ e $k \times k$ respectivamente.

Quando a condição (A1) for verificada para todo vetor $\underline{\phi}$ no espaço paramétrico, é caracterizada a ortogonalidade Global ou Total. Quando a condição (A1) for verificada somente para $\underline{\phi} = \underline{\phi}_0$ então os parâmetros $\underline{\delta}$ e $\underline{\lambda}$ são Localmente Ortogonais em $\underline{\phi}$.

Observação:

1. Ortogonalidade Local implica em matriz de Informação de Fisher bloco diagonal, dada pela matriz em (A2).
2. Ortogonalidade Global implica em matriz de Informação de Fisher diagonal, dada por:

$$I(\underline{\delta}, \underline{\lambda}) = \begin{bmatrix} i_{\delta_1, \delta_1} & & 0 \\ & \ddots & \\ 0 & & i_{\lambda_p, \lambda_p} \end{bmatrix}$$

Caso particular: Se $\underline{\phi} = (\underline{\delta}, \underline{\lambda})$, então $\underline{\delta}$ é ortogonal a $\underline{\lambda}$ se, e somente se, $I(\underline{\delta}, \underline{\lambda})$ for diagonal.

Um exemplo bastante utilizado em nosso trabalho e de fácil verificação, é que os parâmetros de locação e escala da distribuição Normal são ortogonais.

Propriedades da ortogonalidade

Suponha que os parâmetros $\underline{\delta}$ e $\underline{\lambda}$ sejam ortogonais.

- P1. Os estimadores de máxima verossimilhança $\hat{\underline{\delta}}$ e $\hat{\underline{\lambda}}$ são assintoticamente independentes;
- P2. O desvio padrão assintótico de $\hat{\underline{\delta}}$ não depende de $\underline{\lambda}$ ser (ou não) fixado;
- P3. A resolução numérica para $(\underline{\delta}, \underline{\lambda})$ é facilitada, por exemplo, em regressão não linear (para maiores detalhes consulte Cox & Reid, 1987);
- P4. O estimador de máxima verossimilhança de $\underline{\delta}$, quando $\underline{\lambda}$ é fixado, não varia muito com $\underline{\lambda}$;
- P5. Se $\hat{\underline{\delta}}(\underline{\lambda}) = \hat{\underline{\delta}}$, para todo $\underline{\lambda}$ fixado, então $\underline{\delta}$ e $\underline{\lambda}$ são ortogonais.

Reparametrização ortogonal

Se os parâmetros não forem ortogonais é possível, em alguns casos, obter uma reparametrização que produza a ortogonalidade desejada. O procedimento que Cox & Reid abordam utiliza os conceitos de equações diferenciais, discutindo a solução para a ortogonalização quando o parâmetro de interesse é um escalar, isto é, quando $k = 1$ e $\underline{\lambda}$ é o vetor de parâmetros “nuisance”.

Suponhamos que o vetor $\underline{\theta} = (\theta_1, \underline{\theta}_2)$, onde θ_1 é um escalar e o parâmetro de interesse, e $\underline{\theta}_2 = (\theta_{21}, \theta_{22}, \dots, \theta_{2q})$ é o parâmetro “nuisance”, com $q = p - 1$. Considere, originalmente, que a função de verossimilhança é especificada em termos dos parâmetros $\theta_1, \theta_{21}, \theta_{22}, \dots, \theta_{2q}$. O objetivo é reparametrizar o modelo obtendo o vetor de parâmetros $(\delta, \lambda_1, \lambda_2, \dots, \lambda_q)$, com $\delta = \theta_1$ ortogonal a λ_t , $t = 1, 2, \dots, q$. O procedimento consiste de um sistema de q equações atendendo a seguinte condição:

$$\sum_{r=1}^q i_{\theta_{2r} \theta_{2s}} \frac{\partial \theta_{2r}}{\partial \theta_1} = -i_{\theta_1 \theta_{2s}}, \quad s = 1, 2, \dots, q. \quad (\text{A3})$$

onde i^* são os elementos da matriz de informação de Fisher submetidos à nova parametrização (θ_1, θ_2) e a suposição de que a transformação de $(\theta_1, \theta_2) \rightarrow (\delta, \lambda_1, \dots, \lambda_q)$ tenha o Jacobiano não nulo.

Desta forma, o conjunto original $(\theta_1, \theta_{21}, \theta_{22}, \dots, \theta_{2q})$ é reparametrizado para $(\delta, \lambda_1, \lambda_2, \dots, \lambda_q)$. Quando o sistema de equações (A3) admite solução esta é arbitrária, sendo interessante escolher λ com uma interpretação conveniente à situação particular do problema que está sendo estudado.

APÊNDICE B

Método de Laplace para Aproximação de Integrais

Em Inferência Bayesiana as integrais são ostensivamente utilizadas, porém em determinados casos não conseguimos obter a solução explícita dessas integrais, como por exemplo no cálculo da constante normalizadora, densidades a posteriori marginais, momentos a posteriori, etc.

Uma técnica muito utilizada para obtenção de uma expressão explícita aproximada para integrais é o Método de Laplace proposto por Tierney & Kadane (1986).

O propósito deste Apêndice é apresentar resumidamente esse método (para mais detalhes veja, Tierney & Kadane, 1986; Kass et al, 1990, Achar & Smith, 1990).

Considere, inicialmente, $h: \mathcal{R}^m \rightarrow \mathcal{R}$ uma função suave (diferenciável), de um parâmetro $\underline{\theta} = (\theta_1, \theta_2, \dots, \theta_m)$, onde $-h$ tem um único ponto de máximo, que denotamos por $\hat{\theta}$.

O método de Laplace aproxima uma integral da forma

$$\int_{-\infty}^{\infty} \dots \int_{-\infty}^{\infty} f(\underline{\theta}) \exp[-nh(\underline{\theta})] d\theta_1 d\theta_2 \dots d\theta_m. \quad (\text{B1})$$

Fazendo a expansão de Taylor de 3ª ordem nas funções f e h sobre $\hat{\theta}$, o fator resultante no integrando é proporcional à densidade de uma distribuição Normal Multivariada. Assim,

$$\int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} f(\underline{\theta}) \exp[-nh(\underline{\theta})] d\theta_1 d\theta_2 \dots d\theta_m \cong \left[\frac{2\pi}{n} \right]^{m/2} |\Sigma|^{1/2} f(\hat{\underline{\theta}}) \exp[-nh(\hat{\underline{\theta}})] \quad (\text{B2})$$

onde $\Sigma^{-1} = D^2 h(\hat{\underline{\theta}})$ é a matriz Hessiana de f para $\hat{\underline{\theta}}$.

Considerando para o caso unidimensional, temos que:

$$\int_{-\infty}^{\infty} f(\theta) \exp\{-nh(\theta)\} d\theta \cong \left[\frac{2\pi h^{(2)}(\hat{\theta})}{n} \right]^{1/2} f(\hat{\theta}) \exp\{-nh(\hat{\theta})\} \quad (\text{B3})$$

onde $h^{(k)}(\hat{\theta})$ é k-ésima derivada da função h , aplicada no ponto $\hat{\theta}$.

Kass et al argumentam em seu artigo que

$$I = \hat{I} [1 + O(n^{-1})] \quad (\text{B4})$$

$$\text{onde } \hat{I} = \left[\frac{2\pi h^{(2)}(\hat{\theta})}{n} \right]^{1/2} f(\hat{\theta}) \exp\{-nh(\hat{\theta})\}$$

Podemos, então, notar que o Método de Laplace transforma um problema de integração em um problema de maximização de uma função.

Um fator importante a salientar, também, é que pelo fato do Método de Laplace basear-se na expansão de Taylor, a aproximação será tanto mais precisa quanto mais próxima a função estiver da densidade Normal. Estudos referentes à precisão deste método podem ser encontrados em Tierney & Kadane (1986), Kass et al (1990), Achcar (1988, 1989) e Achcar & Smith (1990).

Precisão para um Caso Particular

Se a integral é dada por:

$$I = \int_{-\infty}^{\infty} \exp\{-nh(\theta)\} d\theta \quad (\text{B5})$$

onde θ é unidimensional, e fazendo a expansão de Taylor de 4ª ordem sobre $\hat{\theta}$, a precisão da Aproximação de Laplace pode ser verificada pela seguinte expressão:

$$\frac{5[h^{(3)}(\hat{\theta})]^2}{24 n [h^{(2)}(\hat{\theta})]^3} \approx 0 \quad (\text{B6})$$

APÊNDICE C

Região Highest Posterior Density (HPD)

Quando o interesse é fazer Inferência sobre um vetor de parâmetros a teoria clássica difere totalmente da teoria Bayesiana. Do ponto de vista Bayesiano a inferência é baseada na região a posteriori de maior probabilidade, que colocamos logo abaixo a definição da região HPD, dado por Box & Tiao (1973).

Definição: Seja $p(\underline{\theta}|\underline{Y})$ uma função densidade a posteriori. Uma região $\mathbf{R}$ no espaço paramétrico de $\underline{\theta}$, é chamada uma região Highest Posterior Density (HPD) com credibilidade $(1 - \alpha)$ se satisfaz as seguintes condições:

a) $P[\underline{\theta} \in \mathbf{R} | \underline{Y}] = 1 - \alpha;$

b) para $\underline{\theta}_1 \in \mathbf{R}$ e $\underline{\theta}_2 \notin \mathbf{R},$

$$p(\underline{\theta}_1 | \underline{Y}) \geq p(\underline{\theta}_2 | \underline{Y}).$$

No caso uniparamétrico, o intervalo HPD para θ , dado por (A,B), pode ser ilustrado pelo seguinte gráfico:

Figura a1: Intervalo HPD para θ com nível de credibilidade $(1 - \alpha)$

Algumas propriedades da Região HPD

1. Segue da definição que, para uma dada probabilidade $(1 - \alpha)$, a região HPD tem o menor volume possível no espaço paramétrico de θ .
2. Se fizermos a suposição que $p(\theta|Y)$ é não uniforme sobre toda a região no espaço de θ , então a região HPD de conteúdo $(1 - \alpha)$ é única. Porém, se θ_1 e θ_2 são dois pontos tal que $p(\theta_1|Y) = p(\theta_2|Y)$, então estes dois pontos são simultaneamente incluídos (ou excluídos) em uma região HPD $(1 - \alpha)$. A inversa também é verdadeira, isto é, se $p(\theta_1|Y) \neq p(\theta_2|Y)$, então existe uma região HPD $(1 - \alpha)$ que inclui um ponto mas não o outro.

Nota: Os intervalos a posteriori baseado em priori não informativa são chamados “intervalos de credibilidade” por Edwards, Lindman & Savage (1963), e “intervalos de confiança Bayesiano” por Lindley (1965).

Em alguns casos os intervalos HPD e os intervalos de confiança para um determinado parâmetro podem ser numericamente idênticos, porém é importante enfatizar suas interpretações são totalmente diferentes. Na formulação Bayesiana os parâmetros são considerados como variáveis aleatórias e um intervalo $(1 - \alpha)$ é aquele que baseado em um dado estimador, calculado dos dados observados, contém valores do parâmetro cujo volume de probabilidade é proporcional a $(1 - \alpha)$ da probabilidade a posteriori.

APÊNDICE D

Programas Utilizados

O propósito deste Apêndice é deixar à disposição, para as pessoas interessadas, alguns dos programas que foram utilizados para a simulação de dados, integração numérica e gerar gráficos.

Programa 1

Este programa utiliza o software EUREKA (versão 1.0), para determinar os pontos, através de integração numérica, da equação (4.3.17) com a finalidade de construir os gráficos das figuras 4.4 e 4.5.

$$p(\beta|\text{dados}) = \frac{\int_0^\infty \lambda^{\frac{r-1}{2}} \exp\left\{-\frac{\lambda}{2}\left[\frac{(\mu-\hat{\beta})^2}{b_n} + r\right]\right\} \left[1 + \frac{\lambda^2(\mu-\hat{\beta})^2}{2b_n \sum Z_i^2}\right] \left[1 + \frac{(\lambda + \sum Z_i^2)}{m_\lambda} (\beta - \hat{\beta}_1(\lambda))^2\right]^{\frac{n+n_0+1}{2}} d\lambda}{\Gamma\left(\frac{1}{2}\right) \Gamma\left(\frac{n+n_0}{2}\right) \int_0^\infty \lambda^{\frac{r-1}{2}} \exp\left\{-\frac{\lambda}{2}\left[\frac{(\mu-\hat{\beta})^2}{b_n} + r\right]\right\} \left[1 + \frac{\lambda^2(\mu-\hat{\beta})^2}{2b_n \sum Z_i^2}\right] \left[\frac{m_\lambda}{\lambda + \sum Z_i^2}\right]^{1/2} d\lambda}$$

; CALCULO DA DENSIDADE A POSTERIORI EXATA (aproximacao numerica de Simpson)

n=10:n0=10:bn=50:S0=50:bc=0
 SX2=n:S2=50
 mi=0.25:kx=0.8
 b=-4

```

gam1=sqrt(pi());gam2=fact(9)
gam3=9.5*8.5*7.5*6.5*5.5*4.5*3.5*2.5*1.5*0.5*gam1

dl(l):=n0*S0+(n)*S2+(l*(kx^2)*SX2*((mi-bc)^2))/(l+(kx^2)*SX2)
b1(l):=(l*mi+bc*(kx^2)*SX2)/(l+(kx^2)*SX2)
p1(l):=l^((n-1)/2)
p2(l):=exp((-l/2)*(((mi-bc)^2)/bn)+n))
p4(l):=(1+((l+(kx^2)*SX2)*((b-b1(l))^2))/dl(l))^(-(n+n0+1)/2)
p5(l):=1+(l^2)*((mi-bc)^2)/(2*bn*(kx^2)*SX2)
p6(l):=p1(l)*p2(l)*p4(l)*p5(l)
p7(l):=sqrt(dl(l)/(l+(kx^2)*SX2))
p8(l):=p1(l)*p2(l)*p5(l)*p7(l)
num=integ(p6(l),l,0,10)
den=integ(p8(l),l,0,10)
constN=gam3/(gam1*gam2*den)
POSTBETA=constN*num

```

Programa 2

Este programa utiliza o software EUREKA para determinar os pontos da equação (4.3.20) com a finalidade de construir os gráficos das figuras 4.4 e 4.5.

; DENSIDADE A POSTERIORI PELA APROXIMACAO DE LAPLACE

;Dados

n=10:n0=10:bn=50:S0=50:bc=0

SX2=n:S2=50

mi=0.25:kx=1

```

gam1=sqrt(pi());gam2=fact(9)
gam3=9.5*8.5*7.5*6.5*5.5*4.5*3.5*2.5*1.5*0.5*gam1
lc=(n-1)*bn/(((mi-bc)^2)+(n*bn)) ;lambda chapeu
dlc=n0*S0+(n)*S2+(lc*(kx^2)*SX2*((mi-bc)^2))/(lc+(kx^2)*SX2) ;dlambda chapeu
b1lc=(lc*mi+bc*(kx^2)*SX2)/(lc+(kx^2)*SX2) ;beta1 de lambda chapeu
p1=((lc+(kx^2)*SX2)/dlc)^(1/2)
p2(b):=(1+((lc+(kx^2)*SX2)*((b-b1lc)^2))/dlc)^(-(n+n0+1)/2)
const=gam3/(gam1*gam2)
PLAPLACE(b):=const*p1*p2(b)

```

Programa 3

Este programa utiliza o software MAPLE (versão 5.2 - Windows), para construir os gráficos da densidade a posteriori da Figura 4.2

```
y1:=(9*x)/(109+10*x^2);
y2:=(9*x)/(73+6.4*x^2);
y3:=(9*x)/(34+2.5*x^2);
plot({y1,y2,y3},x=-50..50);
```

Programa 4

Este programa foi feito utilizando o software MAPLE, para construir os gráficos da variância a posteriori da Figura 4.3

```
y1:=((1000*x^2)+10000+((90*x^2)/(109+10*x^2)))/(18*(109+10*x^2));
y2:=((1000*x^2)+10000+((57.6*x^2)/(73+6.4*x^2)))/(18*(73+6.4*x^2));
y3:=((1000*x^2)+10000+((22.5*x^2)/(34+2.5*x^2)))/(18*(34+2.5*x^2));
plot({y1,y2,y3},x=-20..20,0..25,xtickmarks=4,ytickmarks=6);
```

Programa 5

Este programa utiliza o software EUREKA, para determinar o inverso da constante normalizadora da densidade a posteriori a fim de construir o gráfico da Figura 3.2. O cálculo da constante normalizadora é feito através de integração numérica que é dado pelo método de Simpson.


```
; sigma zero = sigma e = sigma u = 1
```

```
X2=402.181; Y2=112.492; XY=145.207; S02=1; n=50
```

```
fi(x):=x*S02*sqrt((Y2+(2/x)*XY+(1/x^2)*X2)/(X2))
```

```
p(x):=(fi(x)-S02)/(x^2*S02*(fi(x)^((n+1)/2))) ;troca de n por n+1
```

```
somae(x):=Y2+XY*(2/x)*(1-(fi(x)/S02))+(X2/x^2)*((1-(fi(x)/S02)))^2
```

```
po(x):=p(x)*exp((somae(x))/(-2*fi(x)))
```

```
constante=integ(po(x),x,0.001,1000)
```

Programa 6

Este programa utiliza o software MAPLE para construir o gráfico da Figura 3.2

```
#sigma zero = sigma e = sigma u = 1
```

```
n:=50; #tamanho da amostra
```

```
Y2:=402.181; X2:=112.492; XY:=145.207; S02:=1;
```

```
f:=x*S02*sqrt((Y2+(2/x)*XY+(1/x^2)*X2)/(X2));
```

```
p:=(f-S02)/(x^2*S02*(f^((n+1)/2)));
```

```
s:=Y2+XY*(2/x)*(1-(f/S02))+(X2/x^2)*((1-(f/S02)))^2;
```

```
po:=p*exp((s)/(-2*f));
```

```
constante:=1.1716560e-28;
```

```
post:=po/constante; #densidade a posteriori de beta
```

```
plot (post,x=0.001..5);
```

Programa 7

Programa 7

Este programa utiliza o software MAPLE, para construir os gráficos da densidade a posteriori da Figura 3.1

```

n:=50; # tamanho da amostra
y1:=247.643; x1:=89.6292; xy1:=92.1172;
c:=GAMMA((n+1)/2)/(GAMMA(0.5)*GAMMA(n/2));
kx1:=0.5; # fator de atenuação
b1:=xy1/(kx1*x1); # EMV de beta
s1:=(y1-2*b1*kx1*xy1+(kx1*b1)^2*x1)/n; #EMV da variância
pa:=sqrt((kx1^2*x1)/s1);
p1:=c*pa*(1+((x-b1)^2*kx1^2*x1)/(n*s1))^(-(n+1)/2);

y2:=270.616; x2:=65.8563; xy2:=108.273;
kx2:=0.8; # fator de atenuação
b2:=xy2/(kx2*x2);
s2:=(y2-2*b2*kx2*xy2+(kx2*b2)^2*x2)/n;
pb:=sqrt((kx2^2*x2)/s2);
p2:=c*pb*(1+((x-b2)^2*kx2^2*x2)/(n*s2))^(-(n+1)/2);

y3:=304.140; x3:=59.6736; xy3:=121.737;
kx3:=1; # fator de atenuação
kx3:=1; # fator de atenuação
s3:=(y3-2*b3*kx3*xy3+(kx3*b3)^2*x3)/n;
pc:=sqrt((kx3^2*x3)/s3);
p3:=c*pc*(1+((x-b3)^2*kx3^2*x3)/(n*s3))^(-(n+1)/2);

plot ({p1,p2,p3},x=0..4);

```

Programa 8:

Este programa foi gravado pelo editor do Northon (arquivo não documento) para realizar a Simulação onde os resultados estão colocados na tabela 3.2. Para isto, utilizamos o MINITAB, versão 2.0 (Student).

```

LET K49=50                                # indica quantas vezes ira simular
LET K50=0.5 # var(e)=1 var(x)=1 var(u)=1
STORE 'SIMULE'
NOECHO
NOPRINT                                # EMQO Estimador de Minimos Quadrados Ordinarios
NRAN 10 0 1 C1                          # gera os erros e
NRAN 10 0 1 C2                          # gera os erros u
NRAN 10 0 1 C3                          # gera os valores de x
LET C4=2*C3+C1                          # valores de Y com beta=2
LET C5=C2+C3                            # valores de X
LET k2=SUM(C5**2)                        # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2)            # EMQO quando a va X esta sem erro
LET K4=K3*(K50**(-1))                   # Beta estimado
LET K5=SUM((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49 # Media dos estimadores de beta1
LET K7=SUM((C7-2)**2)/K49 # EQM de beta1
LET K8=SUM (C8)/K49 # Media dos estimadores de beta
LET K9=SUM((C8-2)**2)/K49 # EQM de beta
outfile'prn' # EQM=erro quadratico medio
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 30 0 1 C1 # gera os erros e
NRAN 30 0 1 C2 # gera os erros u
NRAN 30 0 1 C3 # gera os valores de x
LET C4=2*C3+C1 # valores de Y com beta=2
LET C5=C2+C3 # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49 # Media dos estimadores de beta1
LET K7=SUM((C7-2)**2)/K49 # EQM de beta1

```

```

LET K8=SUM (C8)/K49      # Media dos estimadores de beta
LET K9=SUM((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 50 0 1 C1  # gera os erros e
NRAN 50 0 1 C2  # gera os erros u
NRAN 50 0 1 C3  # gera os valores de x
LET C4=2*C3+C1  # valores de Y com beta=2
LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49      # Media dos estimadores de beta l
LET K7=SUM((C7-2)**2)/K49 # EQM de beta l
LET K8=SUM (C8)/K49      # Media dos estimadores de beta
LET K9=SUM((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile
#####

LET K50=0.8 # var(e)=1 var(x)=1 var(u)=0.25
STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 10 0 1 C1  # gera os erros e
NRAN 10 0 0.5 C2 # gera os erros u
NRAN 10 0 1 C3  # gera os valores de x
LET C4=2*C3+C1  # valores de Y com beta=2
LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1

```

```

ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49      # Media dos estimadores de beta1
LET K7=SUM(((C7-2)**2)/K49 # EQM de beta1
LET K8=SUM (C8)/K49      # Media dos estimadores de beta
LET K9=SUM(((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 30 0 1 C1  # gera os erros e
NRAN 30 0 0.5 C2 # gera os erros u
NRAN 30 0 1 C3  # gera os valores de x
LET C4=2*C3+C1  # valores de Y com beta=2
LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM(((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49      # Media dos estimadores de beta1
LET K7=SUM(((C7-2)**2)/K49 # EQM de beta1
LET K8=SUM (C8)/K49      # Media dos estimadores de beta
LET K9=SUM(((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 50 0 1 C1  # gera os erros e
NRAN 50 0 0.5 C2 # gera os erros u
NRAN 50 0 1 C3  # gera os valores de x
LET C4=2*C3+C1  # valores de Y com beta=2
LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM(((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2

```

```

LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49      # Media dos estimadores de beta l
LET K7=SUM(((C7-2)**2)/K49 # EQM de beta l
LET K8=SUM (C8)/K49      # Media dos estimadores de beta
LET K9=SUM(((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile
#####

LET K50=1  # var(e)=1 var(x)=1 var(u)=0
STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 10 0 1 C1  # gera os erros e
NRAN 10 0 0 C2  # gera os erros u
NRAN 10 0 1 C3  # gera os valores de x
LET C4=2*C3+C1  # valores de Y com beta=2
LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM(((C4-K4*K50*C5)**2)/9 # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49      # Media dos estimadores de beta l
LET K7=SUM(((C7-2)**2)/K49 # EQM de beta l
LET K8=SUM (C8)/K49      # Media dos estimadores de beta
LET K9=SUM(((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 30 0 1 C1  # gera os erros e
NRAN 30 0 0 C2  # gera os erros u
NRAN 30 0 1 C3  # gera os valores de x
LET C4=2*C3+C1  # valores de Y com beta=2

```

```

LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM(((C4-K4*K50*C5)**2)/9) # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49    # Media dos estimadores de beta l
LET K7=SUM(((C7-2)**2)/K49 # EQM de beta l
LET K8=SUM (C8)/K49    # Media dos estimadores de beta
LET K9=SUM(((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

STORE 'SIMULE'
NOECHO
NOPRINT
NRAN 50 0 1 C1 # gera os erros e
NRAN 50 0 0 C2 # gera os erros u
NRAN 50 0 1 C3 # gera os valores de x
LET C4=2*C3+C1 # valores de Y com beta=2
LET C5=C2+C3    # valores de X
LET k2=SUM(C5**2) # soma de X ao quadrado
LET k3=SUM(C4*C5)/SUM(C5**2) # EMV quando a va X esta sem erro
LET K4=K3*(K50**(-1)) # Beta estimado
LET K5=SUM(((C4-K4*K50*C5)**2)/9) # S2
LET C6(K1)=K2
LET C7(K1)=K3
LET C8(K1)=K4
LET C9(K1)=K5
LET K1=K1+1
ECHO
END
LET K1=1
EXEC 'SIMULE' K49
LET K6=SUM (C7)/K49    # Media dos estimadores de beta l
LET K7=SUM(((C7-2)**2)/K49 # EQM de beta l
LET K8=SUM (C8)/K49    # Media dos estimadores de beta
LET K9=SUM(((C8-2)**2)/K49 # EQM de beta
outfile'prn'
PRINT C7 C8 C6 C9
PRINT K6-K9
outfile

```