COMPONENTES DE VARIÂNCIA

- O PROBLEMA DE ESTIMATIVAS NEGATIVAS -

ROBERTO CLÁUDIO FROTA BEZERRA

DISSERTAÇÃO APRESENTADA

AO

INSTITUTO DE MATEMÁTICA E ESTATÍSTICA

DA

UNIVERSIDADE DE SÃO PAULO

PARA OBTENÇÃO DO GRAU DE MESTRE

EM

ESTATÍSTICA

ORIENTADOR:

PROF.DR. CLÓVIS DE ARAUJO PERES

- SÃO PAULO, JUNHO DE 1976 -

Aos meus pais PRISCO e TACILA

o agradecimento maior

À minha mulher MARIA DAS GRAÇAS

e aos meus filhos

PRISCO e ROBERTO CLÁUDIO

AGRADECIMENTOS

Várias pessoas contribuiram numa etapa de minha formação acadêmica, que culminou com essa dissertação.Destaco,

— o Prof. Dr. Clóvis de Araujo Peres, do Instituto de Matemática e Estatística da Universidade de São Paulo (IME-USP), não só por ser meu orientador e pela pessoa humana que é, mas também pela constante orientação estatística, mesmo fora deste trabalho;

— o Prof. Carlos Alberto de Bragança Pereira, do IME-USP pelo estímulo, ajuda e amizade constante;

— o Prof. Dr. Carlos Alberto Barbosa Dantas, do IME.USP,pioneiro na formação de pós-graduados em Estatística no Brasil, que me convenceu a procurar ser mais um iniciado em Estatística;

— os Profs. Mateus Mosca Vianna e Antonio Clécio Fonteles Thomas, da Universidade Federal do Ceará (UFC), pela execução de toda a parte numérica e computacional deste trabalho;

— o Prof. José Jackson Lima Albuquerque, da UFC, por ter sido meu primeiro mestre e orientador na área de Estatística;

— o Sr. João Baptista Esteves de Oliveira, do IME-USP, responsável final pela forma gráfica deste trabalho.

ÍNDICE

INTRODUÇÃO

Desde a sua introdução por Sir Ronald A. Fisher, a Análise de Variância tem sido largamente usada para testar a significância de efeitos de tratamentos, ou melhor dizendo, para testar a existência de diferenças entre médias de tratamentos, e a maior parte dos usuários da Estatística, a tem como uma técnica específica para este objetivo. No entanto, os problemas resolvidos pela Análise de Variância são um pouco mais gerais. Entre esses problemas destacamos a estimação de *componentes de variância*, que aqui serão caracterizados.

Para uma melhor compreensão do exposto acima, vamos considerar a técnica da Análise de Variância na solução de dois tipos de problemas.

O primeiro é aquele no qual estamos interessados em testar a significância de efeitos de um número fixo de tratamentos, ou então, estimar esses efeitos. Esse é o problema tradicional e é resolvido a partir de um modelo, que na sua forma mais simples, recebe na literatura a denominação de *modelo fixo com um fator*, e é representado por

$$y_{ij} = \mu + \alpha_i + e_{ij},$$

onde y_{ij} é a j-ésima observação do i-ésimo tratamento, μ é a média geral, α_i é o efeito do tratamento i, e_{ij} é o termo (aleatório) do erro correspondente à (ij)-ésima observação.

Estimação e Testes de Hipóteses são efeitos sobre esse modelo, usando além da suposição básica de que os e_{ij} são independentes e identicamente distribuídos (i.i.d) com $N(0,\sigma_e^2)$, a restrição

$$\sum_i \alpha_i = 0,$$

para que os efeitos α_i possam ser funções estimáveis.

No segundo problema, menos conhecido, temos representado no experimento, não mais todos os níveis do fator (tratamento), mas sim apenas uma amostra de níveis do fator. Essa modificação de situação torna α_i um efeito aleatório, e a partir do modelo visto acima teríamos y_{ij} e μ caracterizados como anteriormente e mais $E(\alpha_i) = 0$, $var(\alpha_i) = \sigma_\alpha^2$, $E(e_{ij}) = 0$, $var(e_{ij}) = \sigma_e^2$, $E(\alpha_i\alpha_k) = 0$ $(i \neq k)$ e $E(\alpha_i e_{i'j'}) = 0$ $\forall i, i'$ e j'. Com isto obteríamos:

$$var(y_{ij}) = \sigma_\alpha^2 + \sigma_e^2$$

advindo daí a origem da terminologia *componentes de variância*. Neste segundo problema o nosso interesse é o de estimar as variâncias σ_α^2 e σ_e^2. Com as características aqui discutidas, o modelo visto anteriormente, na literatura recebe a denominação geral de *modelo aleatório com um fator*. As inferências a serem feitas sobre o modelo, ou melhor dizendo, sobre os componentes de variância, exigiriam que a variável aleatória α_i fosse i.i.d. com $N(0,\sigma_\alpha^2)$ e e_{ij} fosse i.i.d. com $N(0,\sigma_e^2)$.

Com o objetivo de motivar o uso de componentes de variância discutiremos a situação abaixo.

Considere a situação em que temos uma amostra de m sacos de lã, selecionados de uma grande população de sacos. De cada um desses sacos de lã, que farão parte do experimen to, selecionamos n amostras que irão constituir as repetições. Suponha ainda que a nossa variável de interesse seja a resistência ao rompimento, codificada de acordo com certos requisitos técnicos. O nosso problema é decompor a variabilidade (medida em termos de variância) da variável observada, em dois componentes, um devido à diferença entre sacos (σ_α^2) e o outro devido à diferença entre as amostras dentro do saco (σ_e^2). Esse problema é resolvido quando estimamos os componentes de variância associados ao modelo aleatório simples.

Uma outra aplicação de componentes de variância é no campo da Genética, quando estamos interessados em decompor a variância de um caráter, medida num ser vivo, em componentes de variância genética e ambiental.

Nesta monografia temos por objetivo precípuo, discutir um *problema* que acompanha a estimação de componentes de variância, que é o de encontrarmos em certas ocasiões, estimativas negativas desses componentes. A discussão desse problema foi dividida em três etapas, correspondendo cada uma a um Capítulo.

No Capítulo 1 fazemos uma abordagem bem elementar sobre a estimação de componentes de variância, levando o leitor a tomar contato, de uma maneira bem informal, com o *problema de estimativas negativas*.

No Capítulo 2 trabalhando com a classe de estimadores quadráticos e não viciados dos componentes de variância, procuramos discutir as condições de estimação não negativa desses componentes, e encerramos com a resolução comentada de um exemplo bem simples, procurando tornar mais inteligíveis as idéias lá discutidas.

No Capítulo 3 abordamos o problema de estimativas negativas de uma maneira diferente, buscando uma solução através da Inferência Bayesiana. Encerramos o Capítulo com a resolução de um exemplo, que no seu desenrolar,por envolver integrais de funções não muito simples, necessita de técnicas de integração numérica, que foram utilizadas,com o auxílio do computador. Essa particularidade, serve para enfatizar, a importância do computador na solução de certos problemas através da Inferência Bayesiana.

Além desses três Capítulos, são apresentados três Apêndices, com o objetivo de clarificar ou complementar algumas passagens existentes no corpo principal dessa monografia.

Uma explicação que devemos aos eventuais leitores desse trabalho, é a de que a introdução de certos neologismos como estimabilidade, bayesiano, bayesianismo, etc., deveu-se à falta de expressões correspondentes na nossa língua.

CAPÍTULO 1

SOBRE A ESTIMAÇÃO DE COMPONENTES DE VARIÂNCIA

Neste capítulo discorreremos de uma maneira tão suscinta quanto possível, sobre os métodos de estimação de componentes de variância, que produzem estimativas negativas desses componentes e, enfecharemos com algumas idéias de como contornar ou resolver este problema. Com o intuito de tornar mais claro o assunto, dividimos o capítulo em três seções; nas duas primeiras falamos sobre a estimação de componentes de variância em dados balanceados e não balanceados e, na terceira abordamos o problema das estimativas negativas.

1.1 - DADOS BALANCEADOS

Suponha a seguinte situação

CLASSES	I	II	III	IV
OBSERVAÇÕES	19	25	20	30
	17	5	18	5
	15	15	14	10
	13	10	10	15

Suponha ainda que os dados acima seguem o modelo abaixo

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i=1,\ldots,4 \quad j=1,\ldots,4 \qquad (1.1.1)$$

onde μ é o efeito médio verdadeiro, α_i é um efeito aleatório com $E(\alpha_i) = 0$, $var(\alpha_i) = \sigma^2_\alpha$ e $covar(\alpha_i, \alpha_{i'}) = 0$ $(i \neq i')$, e_{ij} é o termo usual do erro, que também é aleatório com $E(e_{ij})=0$, $var(e_{ij}) = \sigma^2_e$ e covariâncias entre seus termos e, entre seus termos e α_i, todas supostas iguais a zero. Vale ressaltar que os efeitos aleatórios α_i e e_{ij} tem significados diferentes no modelo, pois α_i é um efeito que aparece pela introdução de um fator. O modelo recebe a denominação de modelo aleatório em classificação simples em dados balanceados.

O nosso principal interesse é estimar os componentes de variâncias σ^2_α e σ^2_e. A maneira usual de estimarmos esses componentes é através do uso de certas quantidades que aparecem no quadro de análise de variância (ANOVA), ou mais especificamente, igualando os quadrados médios observados (Q.M.) aos seus valores esperados [E(QM)]. A partir dos dados vistos anteriormente temos o seguinte quadro da ANOVA,

FONTES DE VARIAÇÃO	GL	SQ	QM	E(QM)
Classes	3	11	3,7	$\sigma^2_e+4\sigma^2_\alpha$
Erro	12	648	54	σ^2_e
TOTAL	15	659	-	-

então,

$$\begin{array}{lcl} \sigma^2_e = 54 & & \hat{\sigma}^2_e = 54 \\ & \Longrightarrow & \\ \sigma^2_e+4\sigma^2_\alpha = 3,7 & & \hat{\sigma}^2_\alpha = (3,7-54)/4 \simeq -12,6 \end{array} \qquad (1.1.2)$$

Obtivemos uma estimativa negativa para um parâmetro sabidamente não negativa, e observe que esse exemplo é uma ilustração bem simples do uso da ANOVA na obtenção de um estimador para um componente de variância. Observe também que os estimadores são as soluções das equações lineares nos parâmetros. Daqui por diante diremos que esses estimadores são obtidos pelo método da ANOVA.Note que esses estimadores são os mais usados.

Neste momento talvez já surja a pergunta:"Afora as facilidades de cálculo (pois já aproveita o quadro da ANOVA), o que esse estimador tem de interessante para ter um uso tão geral e indiscriminado, apesar de produzir em algumas situações, estimativas negativas?" A resposta talvez esteja no fato dos estimadores componentes de variância obtidos pelo método da ANOVA (em dados balanceados) apresentarem algumas propriedades interessantes.

(i) são *não viciados*. Suponha que m seja o vetor dos quadrados médios, σ^2 seja o vetor dos componentes de variância a serem estimados. Então, se $E(m) = P\sigma^2$ para alguma matriz não singular P, nós tomamos $m = P\sigma^2$ como equações para estimar σ^2. Então

$$E(\hat{\sigma}^2) = P^{-1}E(m) = P^{-1}P\sigma^2 = \sigma^2.$$

(ii) são *quadráticos não viciados de mínima variância*. Esta propriedade é mostrada por Graybill & Hultguist [1961]. Ela indica que entre todos os estimadores de σ^2 que são funções quadráticas das observações e que são não viciados, os que possuem variância menor são obtidos pelo método da ANOVA.

Essas duas propriedades não necessitam da suposição de normalidade para os elementos aleatórios dos modelos, apenas que eles tenham média e covariâncias *zero* e variancias *finitas*. Isto porque os valores esperados dos quadrados médios nos quadrados da ANOVA não usam suposições de normalidade, e assim os estimadores dos componentes de variancia obtidos pelo método da ANOVA não vão depender de exigências de normalidade. Entretanto, se fizermos suposições de normalidade, os estimadores dos componentes de variância obtidos pelo método da ANOVA em dados balanceados, são baseados em estatísticas suficientes minimais, resultado encontrado em Graybill & Hultguist [1961].

Todas as considerações feitas até aqui, são válidas para qualquer situação de classificação cruzada ou hierárquica, desde que o modelo seja aleatório e em dados balanceados. Com relação à aplicação do método, podemos dizer que as equações obtidas quando da igualdade entre os quadrados médios observados e os valores esperados, serão sempre consistentes, e portanto as soluções (estimadores) serão sempre encontradas.

O método da ANOVA para estimar componentes de variância em dados balanceados, também é usado em modelos mistos.

Apesar das facilidades de cálculo, e de algumas propriedades ótimas, os estimadores dos componentes de variância pelo método da ANOVA em dados balanceados, apresenta um incoveniente, que é o de produzir *estimativas negativas*. E há possibilidade disto acontecer em qualquer modelo (misto ou aleatório) e em qualquer classificação (cruzada ou hierarquica), e temos que reconhecer que isto é um fato por demais desagradável.

1.2 - DADOS NÃO BALANCEADOS

Abordaremos agora alguns métodos de estimação de componentes de variância em dados não balanceados. Suponha a situação abaixo

FATOR A	A_1		A_2		A_3	
FATOR B	B_1	B_2	B_1	B_2	B_1	B_2
OBSERVAÇÕES	5 25	10	10 15 20	5 30	2 20 30	10 20

Suponha ainda que os dados acima seguem o *modelo aleatório classificação cruzada* em dados não balanceados,

$$y_{ijk} = \mu+\alpha_i+\beta_j+\gamma_{ij}+e_{ijk}, \quad i=1,\dots,3, \; j=1,2, \; k=1,\dots,n_{ij} \quad (1.2.1)$$

onde μ é o efeito médio verdadeiro, α_i é efeito aleatório com $E(\alpha_i) = 0$ e $var(\alpha_i) = \sigma^2_\alpha$, β_j é um efeito aleatório com $E(\beta_j) = 0$ e $var(\beta_j) = \alpha^2_\beta$, a interação γ_{ij} é um efeito aleatório com $E(\gamma_{ij}) = 0$ e $var(\gamma_{ij}) = \sigma^2_\gamma$, e_{ijk} é o termo aleatório do erro com $E(e_{ijk}) = 0$ e $var(e_{ijk}) = \sigma^2_e$ e ainda mais, todas as covariâncias entre os efeitos aleatórios, e entre eles e os termos do erro são supostas iguais a zero. Na linguagem de Planejamento de Experimentos α_i e β_i seriam efeitos principais.

O nosso objetivo a partir do conhecimento dos dados existentes no início desta seção é o de estimar os componentes de variância σ^2_α, σ^2_β, σ^2_γ e σ^2_e. Para isto preencheremos inicialmente o quadro da ANOVA, e a partir deste, estimaremos os componentes de variância.

FONTES DE VARIAÇÃO	GL	SQ	QM	E(QM)
Fator A	2	25,6	12,80	$\sigma^2+n''_0\sigma^2_\gamma+n''_1\sigma^2_\beta+n''_2\sigma^2_\alpha$
Fator B	1	4,8	4,80	$\sigma^2+n'_0\sigma^2_\gamma+n'_1\sigma^2_\beta+n'_2\sigma^2_\alpha$
Interação AB	2	32,7	16,35	$\sigma^2+n_0\sigma^2_\gamma+n_1\sigma^2_\beta+n_2\sigma^2_\alpha$
Erro	7	929,2	132,74	σ^2
TOTAL	12	992,3	-	-

onde

$$n_0 = \left[n_{..} - \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} - \sum_j \frac{\sum_i n_{ij}^2}{n_{.j}} + \frac{\sum_i \sum_j n_{ij}^2}{n_{..}} \right] \Big/ (a-1)(b-1)$$

$$n_0' = \left[\sum_j \frac{\sum_i n_{ij}^2}{n_{.j}} - \frac{\sum_i \sum_j n_{ij}^2}{n_{..}} \right] \Big/ (b-1)$$

$$n_0'' = \left[\sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} - \frac{\sum_i \sum_j n_{ij}^2}{n_{..}} \right] \Big/ (a-1)$$

$$n_1 = \left[\frac{\sum_j n_{.j}^2}{n_{..}} - \sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} \right] \Big/ (a-1)(b-1)$$

$$n_1' = \left[n_{..} - \frac{\sum_j n_{.j}^2}{n_{..}} \right] \Big/ (b-1)$$

$$n_1'' = \left[\sum_i \frac{\sum_j n_{ij}^2}{n_{i.}} - \frac{\sum_j n_{ij}^2}{n_{..}} \right] \Big/ (a-1)$$

$$n_2 = \left[\frac{\sum_i n_{i.}^2}{n_{..}} - \sum_j \frac{\sum_i n_{ij}^2}{n_{.j}} \right] \Big/ (a-1)(b-1)$$

$$n_2' = \left[\sum_j \frac{\sum_i n_{ij}^2}{n_{.j}} - \frac{\sum_i n_{i.}^2}{n_{..}} \right] \Big/ (b-1)$$

$$n''_2 = \left[n_{..} - \frac{\sum_i n_{i.}^2}{n_{..}} \right] \Big/ (a-1) \qquad (1.2.2)$$

Substituindo os respectivos valores nas fórmulas acima,e resolvendo o sistema de equações lineares nos componentes de variância, obtemos

$$\hat{\sigma}^2_e = 132{,}74 \quad \hat{\sigma}^2_\gamma = 59{,}06 \quad \hat{\sigma}^2_\beta = 42{,}58 \quad \hat{\sigma}^2_\alpha = -2{,}58 \qquad (1.2.3)$$

Detalhes de como fórmulas vistas em (1.2.2) podem ser obtidas encontram-se em Searle [1971b].

O método usado acima para estimar componentes de variância é conhecido na literatura como *Método 1 de Henderson* e, é essencialmente o mesmo usado para dados balanceados. Por este motivo é também chamado de método da análise de variância para estimar componentes de variância. De todos os métodos usados para estimar componentes de variância este é o mais utilizado. O *Método 1 de Henderson* quando usado em modelos aleatórios produz estimadores não viciados para todos os componentes de variância. No entanto esta propriedade não é válida quando o método é usado em modelos mistos, pois neste caso produz estimadores viciados. Detalhes destes fatos encontram-se nos artigos de Searle [1968;1971a; 1971b].

Com o objetivo de contornar este problema de tendenciosidade dos estimadores de componentes de variância em modelos mistos, surgiu o método conhecido como *Método 2 de Henderson*. Em síntese, este método consiste em inicialmente estimarmos os efeitos fixos do modelo e, os dados são então ajustados por esses estimadores e os componentes de variância são estimados dos dados assim ajustados. Usando este procedimento, realmente produziremos estimadores não viciados, mas em contrapartida, o procedimento não é definido de maneira única (pois envolve uma inversa generalizada) e também, não pode ser usado quando há interação entre os efeitos fixos e aleatórios. Isto é mostrado em detalhes por Searle [1968; 1971a; 1971b].

Outro método sugerido por Henderson, e que recebeu a denominação de *Método 3 de Henderson*, é baseado no método de ajustamento de constantes (denominação motivada pelo fato de que algumas vezes os efeitos fixos em modelos de efeitos fixos são chamados de constantes), que tem uso tradicional em modelos de efeitos fixos. Este método não usa somas de quadrados da ANOVA como os *métodos 1 e 2*, mas reduções nas somas de quadrados, devido ao ajustamento de constantes, e os componentes de variância são estimados através do ato de igualar cada redução calculada ao seu valor esperado sob o modelo completo. Detalhes deste método, inclusive com aplicações podem ser encontrados em Searle [1968; 1971b].

Parafraseando Searle [1968] podemos dizer que todos esses três métodos envolvem, cálculos de quadrados médios de algum tipo, obtenção de seus valores esperados e resolução das equações lineares nos componentes de variância desconhecidos, obtidos da igualdade entre quadrados médios calculados e seus valores esperados.

Em resumo, poderíamos dizer sobre os três métodos de Henderson: O *Método 1* apresenta como vantagens a sua analogia com o método da ANOVA para dados balanceados e sua relativa facilidade de cálculo, e como desvantagem o fato de produzir estimadores viciados dos componentes de variância, quando usado em modelos mistos; o *Método 2* tem a vantagem de produzir estimadores não viciados em modelos mistos e como desvantagens o fato de não ser unicamente definido e o de não poder ser usado quando existe interação de efeitos fixos e aleatórios; o *Método 3* apresenta como vantagens o fato de produzir estimadores não viciados dos componentes de variância em modelos mistos e o de ser o mais indicado para modelos desse tipo, pois não faz restrições quanto às interações entre os efeitos fixos e aleatórios, e como desvantagem principal o de envolver cálculos excessivamente trabalhosos. Como características comuns todos eles apresentam os fatos de, quando usados em dados balanceados, reduzem-se ao método da ANOVA, e o de produzirem *estimativas negativas dos componentes de variância.*

Outros métodos usados em dados não balanceados e conhecidos na literatura como *método das médias* e *método das somas simétricas* são comentados por Searle [1971a; 1971b]. Esses métodos fornecem estimadores não viciados dos componentes de variância, reduzem-se ao método da ANOVA quando usados em dados balanceados, e produzem estimativas negativas dos componentes de variância.

Após uma rápida enumeração e comentários reduzidos de alguns métodos disponíveis para estimarmos componentes de variância, e que produzem estimativas negativas desses componentes, voltamos ao ponto de partida: que fazer quando as estimativas negativas ocorrem? ou então, que métodos alternativos existem que não produzem estimativas negativas?

Searle [1971a; 1971b] faz um bom comentário, apesar de reduzido, sobre o problema acima. Baseado nisto e complementando com algumas idéias que apareceram em artigos posteriores às publicações acima, discorreremos ligeiramente sobre o problema.

1.3 - O PROBLEMA DE ESTIMATIVAS NEGATIVAS

Devido à desagradabilidade do fato de estimarmos como negativos certos componentes de variância, o problema de estimativas negativas tem merecido por parte dos estatís ticos um tratamento especial. Talvez não seja exagero afirmarmos que este fato tenha sido o responsável pelas recentes contribuições na área de estimação de componentes de variância.

Que devemos fazer quando encontramos uma estimativa negativa de um componente de variância? Searle [1971a] enumera vários procedimentos alternativos, e alguns deles serão aqui transcritos.

(i) Apresentar a estimativa obtida e aprender a conviver com ela, como ervas daninhas no gramado. Ela pode ser tomada como uma evidência de que o verdadeiro valor do componente é zero. Isto não é satisfatório quando queremos utilizar uma estimativa da soma de componentes de variância, um dos quais tendo estimativa negativa. A estimativa da soma incluindo este componente é então menor que a soma sem ele.

(ii) Aceitar a estimativa negativa como uma evidência de que o verdadeiro valor do componente correspondente é *zero*, sugerindo a mudança da estimativa negativa pelo valor zero. Isto pode parecer bastante lógico, mas como parte da estimação foi truncada, isto perturba as propriedades dos estimadores.

(iii) Interpretação da estimativa negativa como indicação de um modelo errôneo. Neste contexto, Anderson [1965] afirma que quando uma estimativa negativa ocorre, há possibilidade de haver problemas nos cálculos ou na especificação do modelo. Então, a primeira etapa seria conferir os cálculos realizados e a segunda, se a primeira estivesse correta, seria talvez a reconsideração do modelo. Mesmo assim, permanecendo o pro blema da estimativa negativa, ele sugere considerar o valor do componente como sendo zero.

No aspecto de reconsideração do modelo, Searle & Fawcett [1970] discutem um procedimento em classificação simples, de considerar o modelo em termos de populações finitas, em vez de populações infinitas, procedimento que nos leva a encontrar estimativas não negativas, onde anteriormente, sob a suposição de populações infinitas elas eram negativas.

Já Nelder [1954] afirma que existem experimentos onde temos possibilidade de encontrarmos estimativas negativas para os componentes de variância. Este fato é mostrado por ele em Blocos Aleatoriazados e em "Split-Plot".

Searle [1971b] dá como conselho final: "Tomar as estimativas negativas como indicação de dados insuficientes, e ir atrás da última esperança do estatístico: coletar mais dados e analisá-los, eles próprios ou combinados com aqueles que produziram estimativas negativas. Se a estima obtida usando os dados combinados é negativa, temos uma evidência adicional de que o componente corresponde é *zero*.

Até a presente etapa dessa monografia, a estimação de componentes de variância foi vista sob a luz dos métodos que produzem estimativas negativas, e discutimos procedimentos vários que tentam contornar o problema. Uma outra abordagem para o problema de estimativas negativas, e considerada mais fundamentada, é sugerir métodos alternativos de estimação que não produzem estimativas negativas para os componentes de variância.

Nesta linha, talvez como primeiro trabalho, surgiu o de Herbach [1959], que usa estimadores de máxima verossimilhança, que são não negativos. No seu trabalho ele descreveu os estimadores dos componentes de variância para o modelo aleatório simples em dados balanceados. Eles apresentam como vantagem o fato de serem não negativos, e como desvantagem o fato de serem truncados e viciados.

Thompson [1962] na tentativa de resolver o problema de estimativas negativas desenvolveu um algoritmo, por ele denominado de "pool-the-minimum-violator". Esse algoritmo resulta da maximização da função de verossimilhança dos quadrados médios, sujeita a um conjunto de restrições, que recebe a denominação de princípio de máxima verossimilhança restritivo. Existem aplicações desse algoritmo no artigo escrito por Thompson e Moore [1963]. Este procedimento provoca um truncamento, e o estimador resultante é viciado.

Supondo um modelo do tipo

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i=1,\ldots,m \text{ e } j=1,\ldots,n$$

com as mesmas suposições dadas em (1.1.1), temos como estimadores de σ^2_α e σ^2_e,

(i) ANOVA

$$\hat{\sigma}^2_\alpha = (QME-QMD)/n$$
$$\hat{\sigma}^2_e = QME$$

(ii) Máxima verossimilhança

$$\hat{\sigma}^2_\alpha = \begin{cases} \left[\frac{(m-1)}{m} QME-QMD\right]/n, & \text{se } [(m-1)/m]\cdot QME \geq QMD \\ \\ 0, & \text{se } [(m-1)/m]\cdot QME < QMD \end{cases}$$

$$\hat{\sigma}_e^2 = \begin{cases} QME, \text{ se } [(m-1)/m]\cdot QME \geq QMD \\ \\ SQT/(mn-1), \text{ se } [(m-1)/m]\cdot QME < QMD \end{cases}$$

(iii) Máxima verossimilhança Restritivo

$$\hat{\sigma}_\alpha^2 = \begin{cases} (QME-QMD)/n, \text{ se } QME \geq QMD \\ \\ 0, \text{ se } QME < QMD \end{cases}$$

$$\hat{\sigma}_e^2 = \begin{cases} QME, \text{ se } QME \geq QMD \\ \\ SQT/mn, \text{ se } QME < QMD \end{cases}$$

onde QME, QMD e SQT significam respectivamente, quadrado médio entre, quadrado médio dentro e soma de quadrados total.

Rao [1972] extendeu um novo método de estimação denominado MINQUE (Minimum Norm Quadratic Unbiased Estimation) ao problema de estimação de componentes de variância e covariância. Os estimadores obtidos, após uma modificação introduzida pelo autor, são não negativos, mas não viciados.

LaMotte [1973] desenvolveu dentro da classe dos estimadores quadráticos não viciados, as condições para obtermos estimativas não negativas de combinações lineares dos componentes de variância. Devido à clara importância das idéias desenvolvidas no artigo acima citado, destinamos o Capítulo 2 dessa monografia, para o detalhamento e discussão dos resultados obtidos por LaMotte.

Todos os trabalhos citados até aqui estudaram o problema de estimação dos componentes de variância dentro do aspecto não bayesiano, isto é, ignorando os métodos bayesianos quando das inferências sobre os componentes de variância. Em meados da década de 60, através de Hill [1965] e Tiao & Tan [1965] surgiram os primeiros trabalhos nesta área, tratando da inferência bayesiana sobre σ^2_α e σ^2_e, num modelo aleatório simples, em dados não balanceados e balanceados respectivamente. A grande vantagem prática do uso dos métodos bayesianos em componentes de variância, é que as "estimativas" obtidas são sempre não negativas. No Capítulo 3 dessa monografia discorreremos sobre o problema de fazermos inferências sobre σ^2_α e σ^2_e num modelo aleatório simples em dados balanceados.

- [] -

CAPÍTULO 2

ESTIMAÇÃO QUADRÁTICA NÃO VICIADA E NÃO NEGATIVA DE COMPONENTES DE VARIÂNCIA

No capítulo 1 fizemos algumas considerações sobre os vários métodos de estimação de componentes de variância. Através de exemplos simples, mostramos a possibilidade de estimar como negativo, um componente de variância (um parâmetro sabidamente não negativo), usando o método da ANOVA. Também ficou caracterizado, que dentro da abordagem não bayesiana, de todos os estimadores que fornecem estimativas não negativas dos componentes de variância, nenhum apresenta a propriedade de ser não viciado.

Neste capítulo restringiremos a classe geral dos estimadores dos componentes de variância à classe dos estimadores quadráticos não viciados e, dentro dela, procuraremos discutir as condições de estimabilidade não negativa desses componentes.

2.1 - O MODELO GERAL

Seja Y uma variável aleatória N-dimensional, com ve tor de médias

$$E(Y) = X\beta \qquad (2.1.1)$$

onde β é um vetor de parâmetros $q{\times}1$, X é uma matriz de constantes conhecidas $N{\times}q$, e Y tem matriz de covariância.

$$\mathrm{cov}(Y) = V(\lambda) = \sum_{i=1}^{K} \lambda_i V_i \qquad (2.1.2)$$

onde λ é um vetor de parâmetros (chamados **componentes de variância**) $K{\times}1$, V_i é uma matriz simétrica $N{\times}N$ de constantes conhecidas. Será suposto que λ é tal que $V(\lambda)$ é *positiva definida*. Onde não existir ambiguidade, $V(\lambda)$ será denotada por V.

O modelo caracterizado por (2.1.1) e (2.1.2) é muito geral, requerendo apenas que os elementos do vetor de médias sejam combinações lineares conhecidas de um conjunto de parâmetros (β), e que os elementos da matriz de covariância sejam combinações lineares de outro conjunto de parâmetros (λ).

O nosso problema central é estimar a função $p'\lambda$, onde p é um vetor de constantes $K{\times}1$. Desejamos estimar $p'\lambda$ por uma forma quadrática das observações. Serão consideradas as formas quadráticas do tipo $Y'AY$, onde A é uma matriz simétrica $N{\times}N$.

Como estamos interessados em estimação quadrática não viciada e não negativa devemos propor um estimador que seja

a. quadrático
b. não viciado
c. não negativo

A condição (a) está satisfeita, já que propomos uma forma quadrática para estimador de $p'\lambda$; a condição (b) ficará satisfeita se tomarmos $E(Y'AY) = p'\lambda$; e finalmente a condição (c) também ficará satisfeita se A for uma *matriz não negativa definida.*

Um resultado (ver Searle [1971b]) que nos será útil é o seguinte:

$$E(Y'AY) = tr(AV) + \beta' X' AX\beta \qquad (2.1.3)$$

onde V é a matriz de covariância de Y e tr(M) é a soma dos elementos da diagonal da matriz M.

PROPOSIÇÃO - Uma condição necessária e suficiente para que $Y'AY$ seja um estimador não viciado de $p'\lambda$ é que $X'AX = 0$ e $tr(AV_i) = p_i$.

PROVA -

(Se) - Se vale $X'AX = 0$ e $tr(AV_i) = p_i$ temos

$$E(Y'AY) = \sum_{i=1}^{K} \lambda_i tr(AV_i) + \beta X'AX\beta = \sum_i \lambda_i p_i = p'\lambda$$

(Só se) - Como a condição é para $\forall\beta$, tomemos $\beta = 0$. Então, $E(Y'AY) = \sum_i \lambda_i tr(AV_i) = T'\lambda$ onde,

$T' = [tr(AV_1), \ldots, tr(AV_k)]$ ou seja

$p_i = tr(AV_i)$ então, $E(Y'AY) = p'\lambda + \beta' X' AX\beta$

e para que seja não viciado

$\beta' X' AX\beta = 0 \;\; \forall\beta \Longrightarrow X'AX = 0.$

2.2 - CONDIÇÕES PARA ESTIMABILIDADE NÃO VICIADA E NÃO NEGATIVA

A forma quadrática proposta $(Y'AY)$ satisfará as condições (a), (b) e (c) se entre todas as formas da matriz A escolhermos:

(i) A não negativa definida, isto é, $x'Ax \geq 0$, $\forall x, N \times 1$

(ii) $X'AX = 0$

(iii) $tr(AV) = p'\lambda$ ou de uma maneira equivalente $tr(AV_i) = p_i$.

Mostraremos agora toda uma formalização capaz de garantir que as condições acima sejam satisfeitas.

LEMA - Uma condição necessária e suficiente para que (i) e (ii) sejam satisfeitas é que exista uma matriz C tal que $A=WCC'W$, onde $W=V^{-1}-V^{-1}X(X'V^{-1}X)^{-1}X'V^{-1}$.

PROVA - (Se) - $A = WCC'W$, então A é não negativa definida e $X'AX = 0$, pois $X'W = 0$

- A é não negativa definida, pois $K'=x'WC$ é um vetor real $\Longrightarrow x'Ax = x'WCC'Wx = KK' \geq 0$
- $X'W = 0$; para isto usaremos dois fatos:

Fato 1 - Se V^{-1} é positiva definida, $A'V^{-1}A = 0 \Longrightarrow A = 0$

Fato 2 - $PX'V^{-1} = PX'V^{-1}X \Longrightarrow PX'V^{-1} = QX'V^{-1}$ fazendo

$P = X'V^{-1}X(X'V^{-1}X)^{-}$ e $Q = I$, onde $MM^{-}M = M$ temos, $X'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}X = X'V^{-1}X \Longrightarrow$ $\Longrightarrow X'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1} = X'V^{-1}$ então,

$X'W = X'V^{-1} - X'V^{-1}X(X'V^{-1}X)^{-}X'V^{-1} =$
$= X'V^{-1} - X'V^{-1} = 0.$

(Só se) - Se A é não negativa definida $\exists B$ real tq $A = BB'$

se $X'AX = 0$ então $X'BB'X = 0 \Longrightarrow X'B = 0$

mas $X'B = 0$ se só se $\exists C$ tal que $B = WC$

(se) - se $B = WC \Longrightarrow X'B = X'WC = 0$

(só se)- se $X'B = 0$, seja $C = VB$, então

$WC = WVB = [V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}]VB =$

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

Então, A é não negativa definida e $X'AX = 0$ se só se

$$A = BB' = WCC'W$$

TEOREMA - Uma condição necessária e suficiente para que (i), (ii) e (iii) sejam satisfeitas é que A obedeça as as seguintes condições: existe C tq $A = WCC'W$ e $tr(C'WV_iWC) = p_i$, onde $W = V^{-1} - V^{-1}X(X'V^{-1}X)^{-}X'V^{-1}$.

PROVA -

Pelo Lema, (i) e (ii) valem se só se $A = WCC'W$

(Se) - Se $p_i = tr(C'WV_iWC) = tr(WCC'WV_i)$ e como $A = WCC'W \Longrightarrow tr(AV_i) = p_i$

(Só se) - Se $tr(AV_i) = p_i$, como A é não negativa definida e $X'AX=0$ pelo Lema, $\exists C$ tq $A = WCC'W$ $\Longrightarrow p_i = tr(AV_i) = tr(WCC'WV_i) = tr(C'WV_iWC)$

COROLÁRIO - Se para algum i $(i=1,\ldots,k)$, V_i é positiva definida e $p_i = 0$, o único vetor p para o qual existe um estimador quadrático não viciado e não negativo para $p'\lambda$ é $p = 0$.

PROVA -

Se $p_i = 0$ então $tr(CWV_iWC) = 0$ e se V_i é positiva definida $\Longrightarrow WC = 0 \Longrightarrow tr(C'WV_jWC) = p_j = 0$, $\forall j=1,\ldots,k$.

2.3 - CONSEQUÊNCIAS E APLICAÇÕES

Usando a notação de Rao [1972] nós podemos escrever a nossa variável N-dimensional Y como uma estrutura linear da forma

$$Y = X\beta + U_1\xi_1 + \ldots + U_k\xi_k \qquad (2.3.1)$$

onde X e β são definidos como em (2.1.1), U_i é uma matriz dada $N \times C_i$ e ξ_i é um vetor $c_i \times 1$ de variáveis não correlacionadas com média zero e matris de dispersão $\sigma_i^2 I_{c_i}$, $i = 1, \ldots, k$, onde os σ_i^2 são desconhecidos e I_{c_i} é uma matriz de identidade de $c_i \times c_i$. Ainda mais, ξ_i e ξ_j são não correlacionados. Em particular, podemos pensar em escrever o modelo da ANOVA de acordo com (2.3.1), e neste caso, $U_k = I$ e $c_k = N$. Então temos que

$$\left.\begin{aligned} E(Y) &= X\beta \\ &\text{e} \\ \operatorname{cov}(Y) &= \sigma_1^2 U_1 U_1' + \ldots + \sigma_k^2 I \end{aligned}\right\} \qquad (2.3.2)$$

isto é, nos modelos da ANOVA temos a seguinte estrutura

$$\left.\begin{aligned} V_i &= U_i U_i' \ , \ i=1,\ldots,k-1 \\ V_k &= I \\ \lambda_i &\geq 0 \ , \ i=1,\ldots,k \end{aligned}\right\} \qquad (2.3.3)$$

Uma sequência imediata do *Corolário* da seção anterior é que, nos modelos da ANOVA, o único componente individual que pode ser estimado isoladamente de uma maneira não viciada por uma forma quadrática não negativa é λ_k, e até mesmo λ_k é assim estimável somente se todo V_i, $i=1,\ldots,k-1$, for singular. Outra consequência do *Teorema* é que nesses modelos da ANOVA, apenas os p's tais que $p_i \geq 0$, $i=1,\ldots,k$, podem fazer com que $p'\lambda$ possa ser estimado de uma maneira não viciada por uma quadrática não negativa.

Uma aplicação de todas as idéias vistas até este ponto será dada através de um exemplo com o modelo aleatório simples em dados balanceados.

Considere o modelo

$$y_{ij} = \mu + \alpha_i + e_{ij} \tag{2.3.4}$$

onde $\alpha_i \sim (0,\sigma_\alpha^2)$ $e_{ij} \sim (0,\sigma_e^2)$, $i=1,\dots,m$, $j=1,\dots,n$ ee ainda $E(\alpha_i\alpha_k)=0$, $i\neq k$, $E(\alpha_i e_{i'j'})=0$, $\forall i,i',j'$. Temos que,

$$\operatorname{cov}(y_{ij},y_{i'j'}) = \begin{cases} \sigma_\alpha^2+\sigma_e^2 & , i=i' \text{ e } j=j' \\ \sigma_\alpha^2 & , i=i' \text{ e } j\neq j' \\ 0 & , i\neq i' \end{cases} \tag{2.3.5}$$

Então, para i = 1 teríamos

$$\operatorname{cov}\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n} \end{bmatrix} = \begin{bmatrix} \sigma_\alpha^2+\sigma_e^2 & \sigma_\alpha^2 \cdots & \sigma_\alpha^2 \\ \sigma_\alpha^2 & \sigma_\alpha^2+\sigma_e^2 \cdots & \sigma_\alpha^2 \\ \vdots & \vdots \ddots & \\ \sigma_\alpha^2 & \sigma_\alpha^2 & \sigma_\alpha^2+\sigma_e^2 \end{bmatrix} = \sigma_e^2 I_n + \sigma_\alpha^2 J_n \tag{2.3.6}$$

e,

$$\operatorname{cov}\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{mn} \end{bmatrix} = \begin{bmatrix} \sigma_e^2 I_n+\sigma_\alpha^2 J_n & & 0 \\ & \ddots & \\ 0 & & \sigma_e^2 I_n+\sigma_\alpha^2 J_n \end{bmatrix} = \sum_{i=1}^{m}{}^{+}(\sigma_e^2 I_n+\sigma_\alpha^2 J_n) \tag{2.3.7}$$

onde $\sum^+$ indica a operação de "soma direta" de matrizes, I_n é uma matriz identidade n×n, J_n é uma matriz n×n, com todos os elementos iguais a 1. Indicaremos "produto de Kronocker" por $\otimes$, e um vetor n×1 que só possui *uns* será denotado por $\mathbf{1}_n$. Então,podemos escrever que

$$\sum_{i=1}^{m+} (\sigma_e^2 I_n + \sigma_\alpha^2 J_n) = I_m \otimes (\sigma_e^2 I_n + \sigma_\alpha^2 J_n) =$$

$$= I_m \otimes \sigma_e^2 {}_n + I_m \otimes \sigma_\alpha^2 J_n =$$

$$= \sigma_e^2 I_{m\cdot n} + \sigma_\alpha^2 (I_m \cdot I_n) \otimes (1_n 1_n') =$$

$$= \sigma_e^2 I_{m\cdot n} + \sigma_\alpha^2 (I_m \otimes 1_n) \cdot (I_m \otimes 1_n)' =$$

$$= \sigma_e^2 I_{m\cdot n} + \sigma_\alpha^2 KK' \qquad (2.3.8)$$

o que nos faz escrever

$$V(\lambda) = \lambda_2 I + \lambda_1 KK' \qquad (2.3.9)$$

onde, $K = I_m \otimes 1_n$.

Seja $\lambda_2 = 1$ e $\lambda_1 = 0$. Então,

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

$$= I_{mn} - 1_{mn}(1_{mn}' 1_{mn})^{-} 1_{mn}' =$$

$$= I_{mn} - \frac{1}{mn} 1_{mn} 1_{mn}' \qquad (2.3.10)$$

de modo que,

$$WV_1W = WKK'W = (I_{mn} - \frac{1}{mn} 1_{mn} 1_{mn}')KK'(I_{mn} - \frac{1}{mn} 1_{mn} 1_{mn}') =$$

$$= KK' - \frac{1}{mn} 1_{mn} 1_{mn}' \qquad (2.3.11)$$

e

$$WV_2W = WIW = WW = W \qquad (2.3.12)$$

observe que se somarmos e subtrairmos $\frac{1}{n}$ KK' de W obtemos

$$W = \frac{1}{n} WV_1W + (I_{mn} - \frac{1}{n} KK') \qquad (2.3.13)$$

onde $y'Wy$, $\frac{1}{n} y'WV_1Wy$, $y'(I_{mn} - \frac{1}{n} KK')y$ funcionam respectivamente como uma soma de quadrados total corrigida, soma de quadrados entre e soma de quadrados dentro. Vejamos,

$$y'Wy = y'[I_{mn} - \frac{1}{mn} 1_{mn}1'_{mn}]y = y'y - \frac{1}{mn} y'1_{mn}1'_{mn} y =$$

$$= y'y - \frac{1}{mn} \left(\sum_i\sum_j y_{ij}\right) \cdot \left(\sum_i\sum_j y_{ij}\right) = \sum_i\sum_j y^2_{ij} - mn\bar{y}^2_{..} =$$

$=$ SQ TOTAL CORRIGIDA

$$y'WV_1Wy = y'[KK' - \frac{1}{m} 1_{mn}1'_{mn}]y = y'KK'y - \frac{1}{m} y'1_{mn}1'_{mn}y =$$

$$= y'(I_m \otimes 1_n)(I_m \otimes 1'_n)y - \frac{1}{m}(\sum_i\sum_j y_{ij})(\sum_i\sum_j y_{ij}) =$$

$$= y'(I_m) \otimes 1_n 1'_n y - mn^2\bar{y}^2_{..} =$$

$$= y'I_m \otimes J_n y - mn^2\bar{y}^2_{..} =$$

$$= (n\bar{y}_{1.})^2 + (n\bar{y}_{2.})^2 + \ldots + (n\bar{y}_{m.})^2 - mn\bar{y}^2_{..} =$$

$$= n^2\sum_i \bar{y}^2_{i.} - mn^2\bar{y}^2_{..}$$

$$\Longrightarrow \frac{1}{n} y'WV_1Wy = n \sum_i \bar{y}^2_{i.} - mn\bar{y}^2_{..} = \text{SQ ENTRE}$$

$$y'[I_m - \frac{1}{n} KK']y = y'y - \frac{1}{n} y'KK'y = \sum_i\sum_j y^2_{ij} - \frac{1}{n} n^2\sum_i \bar{y}^2_{i.} =$$

$$= \sum_i\sum_j y^2_{ij} - n \sum_i \bar{y}^2_{i.} = \text{SQ DENTRO}$$

Em (2.3.13) como os termos da direita são matrizes não negativas definidas, para qualquer vetor α,

$$\alpha' W\alpha \geq \alpha'(\frac{1}{n} WV_1W)\alpha \qquad (2.3.14)$$

Seja $\mathbf{C}$ uma matriz mn×mn com vetores coluna $\mathbf{c}_i$. Por (2.3.14),

$$c_i'Wc_i \geq c_i'(\frac{1}{n} WV_1W)c_i \qquad (2.3.15)$$

de modo que

$$\sum_i c_i'Wc_i \geq \sum_i c_i'(\frac{1}{n} WV_1W)c_i \qquad (2.3.16)$$

o termo do lado esquerdo de (2.3.16) é $tr(c'WV_2WC)$ e o termo do lado direito é $tr(c'WV_1WC)/n$. Se a condição do Teorema está satisfeita deve então ser verdade que

$$p_2 \geq \frac{1}{n} p_1 \geq 0 \qquad (2.3.17)$$

Se $n = 1$ e $m > 1$,

$$I_{mn} - \frac{1}{n} KK' = I_m - KK' = I_m - I_m = 0, \text{ pois } K = I_m \otimes 1_1$$

então

$$tr(C'WV_2WC) = tr(C'WV_1WC)$$

$$\Longrightarrow p_2 = p_1 = p \geq 0$$

resultando,

$$p'\lambda = [pp]\begin{bmatrix}\lambda_1\\ \lambda_2\end{bmatrix} = p\lambda_1 + p\lambda_2 = p(\lambda_1 + \lambda_2)$$

isto é, somente múltiplos positivos de $(\lambda_1 + \lambda_2)$ podem ser estimados de uma maneira não negativa por quadráticas não viciadas.

Se $n > 1$ e $m = 1$,
temos que,

$$tr(C'WV_2WC) = tr(\frac{1}{n} C'WV_1WC) = 0$$

resultando

$$p'\lambda = [0 \quad 0]\begin{bmatrix}\lambda_1 \\ \lambda_2\end{bmatrix} = 0$$

isto é, somente o 0 pode ser estimado de uma maneira não negativa por quadráticas não viciadas.

Se $n > 1$ e $m > 1$, e ainda se (2.3.17) ocorre, então existe uma forma quadrática não viciada e não negativa para $p'\lambda$.

Escolha uma matriz c com apenas a primeira coluna c_1 diferente de zero. Temos que $W = \frac{1}{n} WV_1W + (I - \frac{1}{n} KK')$. Então,

$c_1'Wc_1$ pode ser particionada em

$$c_1' \frac{1}{n} WV_1Wc_1 \quad e \quad c_1'(I - \frac{1}{n} KK')c_1$$

Escolha c_1 tal que

$$c_1'Wc_1 = p_2 \quad e \quad c_1'WV_1Wc_1 = \frac{1}{n}p_1 \quad e \quad p_2 \geq \frac{1}{n}p_1$$

Então se $n > 1$ e $m > 1$, para que $p'\lambda$ seja estimado por uma forma quadrática não viciada e não negativa, é necessário e suficiente que (2.3.17) ocorra. Isto é, sempre podemos estimar $p'\lambda$ por uma forma quadrática não viciada e não negativa desde que

$$p_2 \geq \frac{1}{n} p_1$$

Então,

$$p'\lambda = \lambda_2 = [0 \quad 1] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

satisfaz a condição e

$$p'\lambda = n\lambda_1 + \lambda_2 = [n \quad 1] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

também, mas,

$$p'\lambda = \lambda_1 = [1 \quad 0] \begin{bmatrix} \lambda_1 \\ \lambda_2 \end{bmatrix}$$

não satisfaz a condição.

A conclusão a que se chega, é que o modelo aleatório simples, podem ser estimados por uma forma quadrática não viciada e não negativa, o componente de variância λ_2 e a combinação linear dos componentes de variância dada por $n\lambda_1 + \lambda_2$. Mas, infelizmente o componente de variância λ_1 não pode ser estimado de uma maneira não negativa através de uma forma quadrática não viciada.

Em síntese podemos afirmar que, em qualquer modelo de análise de variância, o único componente de variância que pode ser estimado isoladamente por uma forma quadrática não viciada e não negativa é λ_k (ou seja, a variância residual).

2.4 - ESTIMAÇÃO QUADRÁTICA NÃO VICIADA E NÃO NEGATIVA DE σ_e^2 E σ_α^2 NUM MODELO ALEATÓRIO SIMPLES - UM EXEMPLO

Os estimadores dos componentes de variância, obtidos pelo método de ANOVA em dados balanceados, são sabidamente quadráticos e não viciados, pertencendo logicamente à classe dos estimadores discutidos no artigo de LaMotte [1973], haja visto a seção 2.3. Mesmo assim, e assumindo o ônus da repetição às vezes enfadonha, mas com um objetivo maior de tornar mais simples e mais inteligíveis as idéias já vistas neste Capítulo, procuraremos desenvolver em detalhes o exemplo abaixo,

	GRUPO 1	GRUPO 2
OBSERVAÇÕES	y_{11}	y_{21}
	y_{12}	y_{22}

que tem por modelo matemático

$$y_{ij} = \mu + \alpha_i + e_{ij} \tag{2.4.1}$$

onde $\alpha_i \sim (0,\sigma^2_\alpha)$, $e_{ij} \sim (0,\sigma^2_e)$, $i=1,2$ e $j=1,2$, ainda mais, μ é um parâmetro desconhecido, α_i e e_{ij} são variáveis aleatórias mutuamente independentes com médias zero e variâncias σ^2_α e σ^2_e, respectivamente. De acordo com Rao [1972] este modelo pode ser escrito como

$$Y = X\beta + U_1\xi_1 + U_2\xi_2 \tag{2.4.2}$$

onde Y é um vetor de variáveis aleatórias 4×1, β é um vetor de dimensão 1 de parâmetro desconhecido, X é uma matriz de constantes conhecidas 4×1, ξ_1 é um vetor 2×1 de variáveis aleatórias não correlacionadas com média zero e matriz covariância $\sigma^2_1 I_2$ (como σ^2_1 desconhecido e I_2 uma matriz identidade de 2×2), U_1 é uma matriz de constantes conhecidas 4×2, ξ_2 é um vetor 4×1 dos termos do erro e $U_2 = I_4$. Então,

$$Y = \begin{bmatrix}1\\1\\1\\1\end{bmatrix}\mu + \begin{bmatrix}1 & 0\\1 & 0\\0 & 1\\0 & 1\end{bmatrix}\alpha + \begin{bmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix} e \qquad (2.4.3)$$

E como,

$$\mathrm{cov}(Y) = \sigma_1^2 U_1 U_1' + \sigma_2^2 I \qquad (2.4.4)$$

temos que

$$\mathrm{cov}(Y) = \begin{bmatrix}1 & 1 & 0 & 0\\1 & 1 & 0 & 0\\0 & 0 & 1 & 1\\0 & 0 & 1 & 1\end{bmatrix}\sigma_\alpha^2 + \begin{bmatrix}1 & 0 & 0 & 0\\0 & 1 & 0 & 0\\0 & 0 & 1 & 0\\0 & 0 & 0 & 1\end{bmatrix}\sigma_e^2 \qquad (2.4.5)$$

Até esta altura procuramos apenas, através de um exemplo bem simples (talvez o mais simples possível),mostrar como podemos escrever o nosso modelo, usando uma formulação bem geral para um modelo linear. No entanto, o nosso objetivo primeiro é o de verificar o que ocorre com as estimativas de σ_e^2 e de σ_α^2. Durante todo este Capítulo trabalhamos com a classe dos estimadores quadráticos e não viciados dos componentes de variância, e usando o método da ANOVA em dados balanceados (caso do nosso exemplo) obtemos estimadores para σ_e^2 e σ_α^2 que são funções de formas quadráticas das observações e são não viciados. Os estimadores de σ_e^2 e σ_α^2 são respectivamente,

$$\hat{\sigma}_e^2 = \mathrm{QME} \quad \text{e} \quad \sigma_\alpha^2 = \frac{1}{n}[\mathrm{QMD} - \mathrm{QME}] \qquad (2.4.6)$$

onde, QME representa o quadrado médio "entre" grupos, e QMD representa o quadrado médio "dentro" de grupos, sendo que n é o número de observações por grupo.

Como na ANOVA, quando preenchemos o quadro da análise de variância, obtemos quantidades que representam os quadrados médios das diversas "fontes de variação", e como essas quantidades são as mesmas, quer o modelo seja fixo ou aleatório, desenvolveremos as idéias para o preenchimento do quadro, a partir de um modelo fixo.

Seja um modelo idêntico a (2.4.1), com exceção de α_i ser agora uma constante. Então teremos numa forma matricial, que

$$Y = Ap + e \qquad (2.4.7)$$

onde Y é um vetor de variáveis aleatórias 4×1, p é um vetor de parâmetros desconhecidos 3×1 e e é um vetor dos termos do erro 4×1. Então,

$$Y = \begin{bmatrix} 1 & 1 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 1 & 0 & 1 \end{bmatrix} \begin{bmatrix} \mu \\ \alpha_1 \\ \alpha_2 \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix} e \qquad (2.4.8)$$

E o quadro da análise de variância (ver Searle [1971b]) seria

FONTES DE VARIAÇÃO	GL	SQ	
Entre grupos	$\rho(A)-1$	$Y'\{A(A'A)^{-}A'-N^{-1}1_N1_N'\}Y$	$\frac{SQE}{GLE}$
Dentro de grupos	$N-\rho(A)$	$Y'\{I_N-A(A'A)^{-}A'\}Y$	$\frac{SQD}{GLD}$
T O T A L	N-1	$Y'Y-Y'N^{-1}1_N1_N'Y$	

onde $\rho(A)$ é o posto da matriz A, N é o número total de observações, $(A'A)^{-}$ é a inversa generalizada de $A'A$ e 1_N é um vetor de *uns* de dimensão N.

No nosso exemplo temos que $\rho(A) = 2$, $N = 4$, e mais

$$A(A'A)^{-}A' = \begin{bmatrix} 1/2 & 1/2 & 0 & 0 \\ 1/2 & 1/2 & 0 & 0 \\ 0 & 0 & 1/2 & 1/2 \\ 0 & 0 & 1/2 & 1/2 \end{bmatrix}$$

Então,

$$\hat{\sigma}^2_e = \frac{1}{4} \cdot Y' \cdot \begin{bmatrix} 1 & -1 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & -1 & 1 \end{bmatrix} \cdot Y \qquad (2.4.9)$$

$$\hat{\sigma}^2_\alpha = \frac{1}{8} \cdot Y' \cdot \begin{bmatrix} 0 & -2 & 1 & 1 \\ -2 & 0 & 1 & 1 \\ 1 & 1 & 0 & -2 \\ 1 & 1 & -2 & 0 \end{bmatrix} \cdot Y \qquad (2.4.10)$$

A condição para que $\hat{\alpha}^2_e$ seja uma quantidade não negativa é que a matriz envolvida na forma quadrática (2.4.9) seja *não negativa definida* (n.n.d.), o mesmo acontecendo para $\hat{\sigma}^2_\alpha$. Pelo *Teorema da Redução Canônica* de matrizes (ver Rao [1973]) sabemos que A é uma matriz simétrica, existe uma matriz ortogonal U tal que $A = UD'U$, onde D é uma matriz diagonal com os elementos da diagonal iguais às raízes características de A, e a matriz U tem como colunas vetores característicos (associados às raízes características de A) unitários e mutuamente ortogonais.

No nosso exemplo temos que a matriz associada a (2.4.9) possui quatro raízes características ($\lambda_1 = \lambda_2 = 0$ e $\lambda_3 = \lambda_4 = 2$) todas maiores ou iguais a zero, possibilitando-nos escrever,

$$Y'AY = Y'UDU'Y =$$

$$= Z'DZ =$$

$$= Z'D^{1/2}.D^{1/2}Z \geq 0, \text{ porque é uma SQ de reais}$$

$$\Longrightarrow A \text{ é n.n.d.}$$

Temos que $D^{1/2}$ é uma matriz diagonal com os elementos da diagonal definidos como a raiz quadrada dos elementos correspondentes da diagonal D. Em síntese podemos dizer que, $\hat{\sigma}_e^2$ sempre será estimado de maneira não negativa por uma forma quadrática não viciada, quando usamos o método da ANOVA.

Já a matriz associada a (2.4.10) possui como raízes características ($\lambda_1 = 0$, $\lambda_2 = \lambda_3 = 2$, $\lambda_4 = -4$). Como uma matriz simétrica n.n.d. possui todas as raízes características $\lambda_j \geq 0$ (ver Rao [1973]), podemos dizer que a matriz associada a (2.4.10) não é n.n.d., indicando que nem sempre $Y'AY \geq 0$, isto é, há possibilidade de $\hat{\sigma}_\alpha^2$ ser uma quantidade negativa.

Como já sabemos que podemos estimar σ_α^2 por uma quantidade negativa, uma pergunta que talvez surja, seja: não seria possível calcular a probabilidade de obter uma estimativa negativa para o componente σ_α^2, no modelo aqui discutido? A possibilidade passa a existir a partir do instante em que são satisfeitas as suposições,

$$\alpha_i \sim N(0,\sigma_\alpha^2) \quad \text{e} \quad e_{ij} \sim N(0,\sigma_e^2)$$

no modelo

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i=1,\ldots,m \text{ e } j=1,\ldots,n$$

fazendo com que

$$\frac{SQE}{E(QME)} = \frac{(m-1)QME}{\sigma_e^2+n\sigma_\alpha^2} \sim \chi^2_{(m-1)}$$

$$\frac{SQD}{E(QMD)} = \frac{m(n-1)QMD}{\sigma_e^2} \sim \chi^2_{m(n-1)}$$

e

$$\frac{(QME)/(\sigma_e^2+n\sigma_\alpha^2)}{(QMD)/\sigma_e^2} = \frac{QME}{QMD}\cdot\frac{\alpha_e^2}{\sigma_e^2+n\sigma_\alpha^2} \sim F_{(m-1),m(n-1)}$$

E como,

$$\hat{\sigma}_\alpha^2 = n^{-1}(QME - QMD) < 0 \text{ se } (QME/QMD) < 1$$

temos então que

$$Pr\{\hat{\sigma}_\alpha^2<0\} = Pr\{(QME/QMD)<1\} =$$

$$= Pr\left\{\frac{QME}{QMD}\cdot\frac{\sigma_e^2}{\sigma_e^2+n\sigma_\alpha^2} < \frac{\sigma_e^2}{\sigma_e^2+n\sigma_\alpha^2}\right\} =$$

$$= Pr\left\{F_{(m-1),m(n-1)} < \frac{\sigma_e^2}{\sigma_e^2+n\sigma_\alpha^2}\right\}$$

$$= Pr\left\{F_{(m-1),m(n-1)} < \frac{1}{1+n\rho}\right\}$$

onde $\rho = \sigma_\alpha^2 | \sigma_e^2$.

Isto significa que se conhecermos a relação entre σ_α^2 e σ_e^2, podemos calcular a probabilidade de obtermos uma estimativa negativa para o componente σ_α^2.

Voltando ao exemplo discutido nesta seção, onde m= =n=2, podemos formar o quadro abaixo,

ρ	0	1/20	1/10	1/5	1/2	1	2	10	20
$Pr\{\hat{\sigma}^2_\alpha<0\}$	0,58	0,56	0,54	0,51	0,45	0,38	0,30	0,16	0,10

nos indicando que, num planejamento do tipo visto neste exemplo específico, a probabilidade de obtermos uma estimativa negativa (usando um estimador quadrático e não viciado) para o componente σ^2_α, não é das menores, pelo menos no intervalo $0 \leq \sigma^2_\alpha < \sigma^2_e$.

Usando as mesmas idéias, e apenas a título de ilustração, podemos calcular algumas probabilidades de obtenção de uma estimativa negativa para o componente σ^2_α, a partir de um planejamento do tipo visto no início do Capítulo 1, onde m= n=4; isto nos daria

ρ	0	1/5	1	2	10
$Pr\{\hat{\sigma}^2_\alpha<0\}$	0,57	0,35	0,11	0,05	0,005

Já aqui, poderíamos dizer por exemplo, que a probabilidade de estimarmos como negativo o componente σ^2_α, não é pequena no intervalo $0 \leq \sigma^2_\alpha < (1/5)\sigma^2_e$.

CAPÍTULO 3

INFERÊNCIA BAYESIANA DE COMPONENTES DE VARIÂNCIA

Nos capítulos precedentes discorremos sobre a estimação dos componentes de variância, dando ênfase ao problema de estimativas negativas desses componentes. No entanto, toda a nossa discussão fez-se à luz da teoria clássica.

No presente capítulo nos propomos a estudar o problema de "estimação" dos componentes de variância dentro do contexto bayesiano. Como nosso objetivo primeiro, é o de permitir um acompanhamento "pari-passu" das idéias, que aqui serão desenvolvidas, a nossa abordagem será feita a partir de algumas noções básicas da inferência bayesiana. Em seguida, tendo por base o modelo aleatório simples em dados balanceados, "estimaremos", por métodos bayesianos, os componentes de variância.

3.1 - A IDÉIA DA INFERÊNCIA BAYESIANA

Seja X uma variável aleatória cuja função densidade de probabilidade(*) (f.d.p.) é $f(x/\theta)$, onde $\theta \in \Theta$. Se θ for conhecido, a distribuição de probabilidade estará totalmen-

(*) - A terminologia densidade é aqui usada tanto para variável aleatória discreta como para variável aleatória contínua.

te especificada, e não haverá problemas de *estimação*. Mas, na prática, o comum é θ ser desconhecido, e neste caso, como proceder?

Até o momento, o nosso problema de estimação foi encarado da seguinte maneira: "Retira-se uma amostra

$$\vec{X}' = (X_1, \ldots, X_n)$$

da população que tem f.d.p. $f(x/\theta)$ e, com base nos valores amostrais $\vec{x}' = (x_1, \ldots, x_n)$ estima-se o parâmetro desconhecido θ". Melhor dizendo, o problema é definir uma função $t = t(\vec{x})$, tal que se $\vec{x}' = (x_1, x_2, \ldots, x_n)$ são os valores amostrais observados de $\vec{X}' = (X_1, X_2, \ldots, X_n)$ então $t(\vec{x})$ será a *estimativa* de θ. Tal procedimento faz parte do que será chamado, por nós, de *estimação clássica*.

Na abordagem bayesiana existe uma diferença fundamental, pois θ não mais será suposto um parâmetro, mas sim uma variável aleatória, com f.d.p. $f(\theta)$.

Suponha como anteriormente, que retiramos uma amostra aleatória $\vec{X}' = (X_1, \ldots, X_n)$ com o intuito de obtermos informações sobre θ. Na situação clâssica $f(x/\theta)$ indicava a densidade da variável aleatória X, com θ sendo suposto fixo. Agora que θ é uma variável aleatória, $f(x/\theta)$ representa a a densidade condicional de X, dado θ fixado. Temos então que

$$f(\vec{x}|\theta)\cdot f(\theta) = f(\vec{x},\theta) = f(\theta|\vec{x})\cdot f(\vec{x}) \qquad (3.1.1)$$

Dadas as observações $\vec{x}$, a densidade condicional de θ é

$$f(\theta|\vec{x}) = \frac{f(\vec{x}|\theta)\cdot f(\theta)}{f(\vec{x})} = c\cdot f(\vec{x}|\theta)\cdot f(\theta) \qquad (3.1.2)$$

onde

$$f(\vec{x}) = E\{f(\vec{x}|\theta)\} = c^{-1} = \begin{cases} \int f(\vec{x}|\theta)\cdot f(\theta)\cdot d\theta, & \text{se } \theta \text{ é v.a. contínua} \\ \sum f(\vec{x}|\theta)\cdot f(\theta), & \text{se } \theta \text{ é v.a. discreta} \end{cases}$$

isto porque $E\{g(\theta)\}$ é a esperança matemática de $g(\theta)$ com respeito à distribuição $f(\theta)$.

A relação (3.1.2) ou sua equivalente é usualmente referida na literatura como o *Teorema de Bayes*. Nesta expressão, $f(\theta)$ representa o nosso conhecimento sobre θ, antes de observarmos os dados, e será chamada de *distribuição "a priori" de* θ; correspondentemente, $f(\theta|\vec{x})$ representa o nosso conhecimento sobre θ após observarmos os dados, e será chamada de *distribuição "a posteriori" de* θ dado as observações; $\underline{c}$ é simplesmente uma constante de normalização, necessária para que a integral (ou soma) de $f(\theta|\vec{x})$ sobre θ seja 1. Por uma questão de simplicidade chamaremos em algumas ocasiões $f(\theta)$ de "priori" e $f(\theta|\vec{x})$ de "posteriori".

Como $$L(\theta|\vec{x}) = \prod_{i=1}^{n} f(x_i|\theta) = f(x_1|\theta)\ldots f(x_n|\theta) = f(\vec{x}|\theta) \qquad (3.1.3)$$

podemos escrever a partir de (3.1.2) que,

$$\boxed{\text{distribuição "a posteriori"} \ \alpha \ \text{verossimilhança} \times \text{distribuição "a priori"}} \qquad (3.1.4)$$

Em (3.1.4) temos de uma maneira sintética a idéia da inferência bayesina: a *"priori"* expressa a nossa opinião sobre θ antes do experimento ser realizado, a *verossimilhan*-

ça expressa a informação contida na amostra, e a *"posteriori"* expressa a nossa opinião final sobre θ após a realização do experimento. Melhor dizendo, $f(\theta)$ é um modelo probabilístico representativo do nosso suposto conhecimento sobre a variável aleatória θ, $f(\theta|\vec{x})$ é o modelo probabilístico que representa o nosso conhecimento "calibrado" pela amostra, e $L(\theta|\vec{x})$ é o mescanismo de "calibração". O problema da inferência bayesiana está resolvido no momento em que encontramos a distribuição "a posteriori".

3.2 - INFERÊNCIA BAYESIANA DE COMPONENTES DE VARIÂNCIA NUM MODELO ALEATÓRIO SIMPLES

Nesta seção, que abrange praticamente todo o Capítulo 3 dessa monografia, nos fixaremos no modelo aleatório simples e adotaremos a "abordagem bayesiana" no problema de "estimação" dos componentes de variância. Para isto, será necessário especificarmos detalhadamente o modelo, encontrarmos a função de verossimilhança e determinarmos a distribuição "a priori". Com isso em mãos, poderemos obter a distribuição "a posteriori".

3.2.1 - O MODELO

Como preliminar introduziremos um sumário de notação e suposições que serão utilizadas no modelo que adotaremos, dado por

$$y_{ij} = \mu + \alpha_i + e_{ij}, \quad i=1,\ldots,m, \quad j=1,\ldots,n \qquad (3.2.1.1)$$

onde μ é um parâmetro de locação desconhecida, $\alpha_1,\ldots,\alpha_m$ e $e_{11},\ldots,e_{mn}$ são variáveis aleatórias não observáveis, conjuntamente independentes, com $\alpha_i \sim N(0,\sigma_\alpha^2)$ e $e_{ij} \sim N(0,\sigma_e^2)$. É um fato conhecido (ver Graybill & Multquist [1961]) que

$$(\bar{y}_{..},\ SQE,\ SQD)$$

é uma estatística suficiente minimal para $(\mu,\ \sigma_\alpha^2,\ \sigma_e^2)$, onde

$$\left.\begin{aligned} \bar{y}_{..} &= N^{-1}\sum_{ij} y_{ij} \\ SQE &= \sum_{ij}(\bar{y}_{i.}-\bar{y}_{..})^2 \\ SQD &= \sum_{ij}(y_{ij}-\bar{y}_{i.})^2 \end{aligned}\right\} \qquad (3.2.1.2)$$

com $\bar{y}_{i.} = n^{-1}\sum_j y_{ij}$ e $N = mn$.

Além do mais, as componentes de $(\bar{y}_{..}, SQE, SQD)$ são independentemente distribuidas dado $(\mu, \sigma_\sigma^2, \sigma_e^2)$ como

$$\left.\begin{aligned} &\bar{y}_{..} \sim N(\mu, \bar{\sigma}^2/N) \\ &SQE \sim \bar{\sigma}^2 \chi^2_{\nu_\alpha} \\ &SQD \sim \sigma^2 \chi^2_{\nu_e} \end{aligned}\right\} \qquad (3.2.1.3)$$

onde $\bar{\sigma}^2 = \sigma_e^2 + n\sigma_\alpha^2$, $\nu_\alpha = m-1$ e $\nu_e = m(n-1)$.

Temos também que

$$\left.\begin{aligned} QME &= SQE/\nu_\alpha \\ QMD &= SQD/\nu_e \end{aligned}\right\} \qquad (3.2.1.4)$$

Já é sabido que

$$\hat{\sigma}_\alpha^2 = n^{-1}(QME-QMD)$$

é o estimador não viciado de mínima variância para σ_α^2. Esse é o estimador da ANOVA que foi discutido no Capítulo 1 e que, também pertence à classe dos estimadores que tiveram algumas propriedades de estimabilidade discutidas no Capítulo 2.

3.2.2 - A FUNÇÃO DE VEROSSIMILHANÇA

Para o modelo (3.2.1.1) vimos anteriormente em (2.3.5) que,

$$cov(y_{ij}, y_{i'j'}) = \begin{cases} \sigma_\alpha^2 + \sigma_e^2, & se\ i=i' \quad e \quad j=j' \\ \sigma_\alpha^2 & , se\ i=i' \quad e \quad j \neq j' \\ 0 & , se\ i \neq i' \end{cases}$$

A função de verossimilhança para y_{ij} será

$$L(\vec{\theta}|\vec{y}) = f(y_{11},\ldots,y_{1n},y_{21},\ldots,y_{2n},\ldots,y_{m1},\ldots,y_{mn}) =$$

$$= f(y_{11},\ldots,y_{1n})\ f(y_{21},\ldots,y_{2n})\ldots f(y_{m1},\ldots,y_{mn}) =$$

$$= f(\vec{y}_1)\ f(\vec{y}_2)\ldots f(\vec{y}_m)$$

onde $\vec{\theta}' = (\mu,\sigma_\alpha^2,\sigma_e^2)$, $\vec{y}' = (y_{11},\ldots,y_{mn})$. E mais, $\vec{y}_i$ $(i=1,\ldots,m)$ tem distribuição $N(\mu^*,V^*)$ com $\mu^* = 1_n\mu$ e $V^* = \sigma_e^2 I_n + \sigma_\alpha^2 J_n$ [ver (2.3.6)]. Então,

$$f(\vec{y}_i) = (2\pi)^{-n/2} \cdot |V^*|^{-1/2} \cdot \exp\{-\tfrac{1}{2}(\vec{y}_i-\mu^*)'(V^*)^{-1}(\vec{y}_i-\mu^*)\}$$

o que faz com que

$$L(\vec{\theta}|\vec{y}) = (2\pi)^{-\frac{1}{2}mn} \cdot |V^*|^{-m/2} \cdot \exp\{-\tfrac{1}{2}\sum_{i=1}^{m}(\vec{y}_i-\mu^*)'(V^*)^{-1}(\vec{y}_i-\mu^*)\}$$

Por Graybill [1961] temos que

$$|V^*| = (\sigma_e^2)^{n-1}\cdot(\sigma_e^2+n\sigma_\alpha^2)$$

e também que

$$\sum_{i=1}^{m}(\vec{y}_i-\mu^*)'(V^*)^{-1}(\vec{y}_i-\mu^*) = \frac{1}{\sigma_e^2}\sum_{ij}(y_{ij}-\bar{y}_{i.})^2 +$$

$$+\frac{1}{\sigma_e^2+n\sigma_\alpha^2}\left[\sum_{ij}(\bar{y}_{i.}-\bar{y}_{..})^2+\frac{nm(\bar{y}_{..}-\mu)^2}{\sigma_e^2+n\sigma_\alpha^2}\right]$$

Então, podemos escrever

$$L(\vec{\theta}|\vec{y})\alpha(\sigma_e^2)^{-\nu_e/2}\cdot(\sigma_e^2+n\sigma_\alpha^2)^{-(\nu_e+1)/2}\cdot$$

$$\cdot\exp\left\{-\frac{1}{2}\left[\frac{SQD}{\sigma_e^2}+\frac{SQE}{\sigma_e^2+n\sigma_\alpha^2}+\frac{nm(\bar{y}_{..}-\mu)^2}{\sigma_e^2+n\sigma_\alpha^2}\right]\right\}$$

ou

$$L(\vec{\theta}|\vec{y})\alpha(\sigma_e^2)^{-\nu_e/2}\cdot(\sigma_e^2+n\sigma_\alpha^2)^{-(\nu_\alpha+1)/2}\cdot$$

$$\cdot\exp\left\{-\frac{1}{2}\left[\frac{\nu_e QMD}{\sigma_e^2}+\frac{\nu_\alpha QME}{\sigma_e^2+n\sigma_\alpha^2}+\frac{nm(\bar{y}_{..}-\mu)^2}{\sigma_e^2+n\sigma_\alpha^2}\right]\right\}$$

onde ν_e e ν_α são definidos como em (3.2.1.3).

3.2.3 - A DISTRIBUIÇÃO "A PRIORI" NÃO-INFORMATIVA

A partir desse momento, temos como principal objetivo, a determinação da nossa "priori" $f(\vec{\theta})$ onde

$$\vec{\theta}' = (\mu,\sigma_\alpha^2,\sigma_e^2).$$

Partiremos da suposição de que existe *pouco* ou *nenhum* conhe cimento sobre $\vec{\theta}$, e neste caso a nossa "priori" chamar-se-á *não-informativa* (é aquela que não altera a informação fornecida pela amostra).

De início suponha que nossas opiniões sobre μ são *a priori* idependentes daquelas sobre $(\sigma_\alpha^2, \sigma_e^2)$. Então,

$$f(\mu, \sigma_\alpha^2, \sigma_e^2) = f(\mu) \cdot f(\sigma_\alpha^2, \sigma_e^2) \qquad (3.2.3.1)$$

Como $f(\mu) \alpha$ constante (ver Apêndice A), temos,

$$f(\mu, \sigma_\alpha^2, \sigma_e^2) \alpha \cdot f(\sigma_\alpha^2, \sigma_e^2) \qquad (3.2.3.2)$$

E a *priori não-informativa* para σ_α^2 e σ_e^2 poderá ser obtida através da aplicação da regra de Jeffreys: "a priori não-informativa para o vetor paramétrico $(\sigma_\alpha^2, \sigma_e^2)$ é tomada como sendo proporcional à raiz quadrada do determinante da matriz de informação" (ver Apêndice A). Isto é,

$$f(\sigma_\alpha^2, \sigma_e^2) \alpha \ |Inf(\sigma_\alpha^2, \sigma_e^2)|^{1/2} \qquad (3.2.3.3)$$

onde,

$$Inf(\sigma_\alpha^2, \sigma_e^2) = E\left\{-\frac{\partial^2 \log L(\vec{\theta}|\vec{y})}{\partial\sigma_i^2 \cdot \sigma_j^2}\right\} = \frac{1}{2}\begin{bmatrix} \frac{\nu_\alpha+1}{\bar{\sigma}^4} + \frac{\nu_e}{2\sigma_e^2} & \frac{n(\nu_\alpha+1)}{\bar{\sigma}^4} \\ \frac{n(\nu_\alpha+1)}{\bar{\sigma}^4} & \frac{\underline{n}^2(\nu_\alpha+1)}{\bar{\sigma}^4} \end{bmatrix}$$

e como,

$$|Inf(\sigma_\alpha^2, \sigma_e^2)| = \frac{n^2\nu_e(\nu_\alpha+1)}{2} \cdot \frac{1}{(\sigma_e^2)^2(\sigma_e^2+\sigma_\alpha^2)^2}$$

segue que

$$f(\mu, \sigma_\alpha^2, \sigma_e^2) \alpha \ \frac{1}{\sigma_e^2(\sigma_e^2+n\sigma_\alpha^2)} \qquad (3.2.3.4)$$

Maneiras alternativas de obtermos a priori (3.2.3.4) são encontrados no artigo de Tiao & Tan [1965[e no livro de Box & Tiao [1973].

3.2.4 - AS DISTRIBUIÇÕES "A POSTERIORI"

Por (3.1.4) temos que

$$f(\vec{\theta}|\vec{y}) \propto L(\vec{\theta}|\vec{y}) \times f(\vec{\theta})$$

onde $\vec{\theta} = (\mu, \sigma^2_\alpha, \sigma^2_e)$. Pelos resultados de (3.2.2.1) e (3.2.3.4) podemos escrever

$$f(\mu, \sigma^2_\alpha, \sigma^2_e|\vec{y}) \propto (\sigma^2_e)^{-\nu_e/2} \cdot (\sigma^2_e + n\sigma^2_\alpha)^{-(\nu_\alpha+1)/2} \cdot \exp\left\{-\frac{1}{2}\left[\frac{\nu_e QMD}{\sigma^2_e} + \frac{\nu_\alpha QME}{\sigma^2_e + n\sigma^2_\alpha} + \frac{nm(\bar{y}_{..} - \mu)^2}{\sigma^2_e + n\sigma^2_\alpha}\right]\right\} \times (\sigma^2_e)^{-1} (\sigma^2_e + n\sigma^2_\alpha)^{-1} =$$

$$= (\sigma^2_e)^{-(\frac{1}{2}\nu_e + 1)} \cdot (\sigma^2_e + n\sigma^2_\alpha)^{-\frac{1}{2}(\nu_\alpha + 1) - 1} \times$$

$$\times \exp\left\{-\frac{1}{2}\left[\frac{\nu_e QMD}{\sigma^2_e} + \frac{\nu_\alpha QME}{\sigma^2_e + n\sigma^2_\alpha} + \frac{nm(\bar{y}_{..} - \mu)^2}{\sigma^2_e + n\sigma^2_\alpha}\right]\right\},$$

$$-\infty < \mu < \infty, \quad \sigma^2_e > 0, \quad \sigma^2_\alpha > 0 \qquad (3.2.4.1)$$

Mas como nosso objetivo básico é fazer inferências sobre σ^2_e e σ^2_α, e não sobre $\vec{\theta}$, devemos de início pensar em obter a distribuição conjunta "a posteriori" de $(\sigma^2_e, \sigma^2_\alpha)$ e as distribuições "a posteriori" de σ^2_e e σ^2_α.

A DISTRIBUIÇÃO CONJUNTA "A POSTERIORI" DE $(\sigma^2_\alpha, \sigma^2_e)$

A idéia natural para obtermos a distribuição "a pos teriori" de $(\sigma^2_\alpha,\sigma^2_e)$ é a de integrarmos (3.2.4.1) sobre μ, o que nos dará

$$f(\sigma^2_\alpha,\sigma^2_e|\vec{y}) = K\cdot(\sigma^2_e)^{-(\frac{1}{2}\nu_e+1)}\cdot(\sigma^2_\alpha)^{-\frac{1}{2}(\nu_\alpha+1)}\cdot$$

$$\cdot\exp\left\{-\frac{1}{2}\left[\frac{\nu_e QMD}{\sigma^2_e}+\frac{\nu_\alpha QME}{\sigma^2_e+n\sigma^2_\alpha}\right]\right\},$$

$$\sigma^2_e>0,\ \sigma^2_\alpha>0 \qquad (3.2.4.2)$$

onde

$$K=\frac{n(\nu_e QME)^{\frac{1}{2}\nu_e}\cdot(\nu_\alpha QMD)^{\frac{1}{2}\nu_\alpha}\cdot 2^{-\frac{1}{2}(\nu_e+\nu_\alpha)}}{\Gamma(\frac{1}{2}\nu_e)\cdot\Gamma(\frac{1}{2}\nu_\alpha)\cdot Pr\{F_{\nu_\alpha,\nu_e}<QME/QMD\}}$$

resultado encontrado por Box & Tiao [1973].

Usando a definição da distribuição χ^{-2} (ver Apêndice B), Box & Tiao [1973] escreveram (3.2.4.2) como

$$f(\sigma^2_\alpha,\sigma^2_e|\vec{y}) =$$

$$=\frac{(\nu_e QMD)^{-1}f(\chi^{-2}_{\nu_e}=\sigma^2_e/\nu_e QMD)\,n(\nu_\alpha QME)^{-1}\cdot f[\chi^{-2}_{\nu_\alpha}=(\sigma^2_e+n\sigma^2_\alpha)/\nu_\alpha QME]}{Pr\{F_{\nu_\alpha,\nu_e}<QME/QMD\}},$$

$$\sigma^2_e>0,\ \sigma^2_\alpha>0 \qquad (3.2.4.3)$$

onde $f(\chi^{-2}_\nu=x)$ é a densidade de uma variável χ^{-2} com ν g.l. avaliados em x.

A DISTRIBUIÇÃO "A POSTERIORI" DE σ_e^2

Se integrarmos (3.2.4.3) sobre σ_α^2, obviamente obteremos $f(\sigma_e^2|\vec{y})$. Segundo Box & Tiao [1973] a distribuição "a posteriori" do componente de variância σ_e^2, é dada por

$$f(\sigma_e^2|\vec{y}) = \frac{(\nu_e QMD)^{-1} \cdot f(\chi_{\nu_e}^{-2} = \sigma_e^2/\nu_e QMD) \cdot Pr\{\chi_{\nu_\alpha}^2 < \nu_\alpha QME/\sigma_e^2\}}{Pr\{F_{\nu_\alpha,\nu_e} < QME/QMD\}}$$

$$\sigma_e^2 > 0 \qquad (3.2.4.4)$$

A distribuição "a posteriori" de σ_e^2 é igual ao produto de dois fatores, o primeiro sendo a densidade de uma variável χ^{-2} com ν_e graus de liberdade e o segundo, a razão de uma probabilidade sobre a χ^2 e de uma probabilidade sobre a F.

A DISTRIBUIÇÃO "A POSTERIORI" DE σ_α^2

Chegamos ao ponto que cria mais problemas, quando usamos a abordagem clássica da inferência estatística, que é o de fazermos inferências sobre o componente da variância σ_α^2. Como já vimos anteriormente, na abordagem clássica,o estimador de σ_α^2 que é mais usado e que apresenta a propriedade de ser não viciado e de mínima variância, em algumas situações, produz estimativas negativas.

Na abordagem bayesiana não temos o tipo de problema especificado acima, pois a "posteriori" de σ_α^2 está definida no intervalo $\sigma_\alpha^2 > 0$.

A obtenção da "posteriori" para σ_α^2 segue,em linhas gerais, a mesma idéia até aqui adotada. Então, devemos integrar (3.2.4.3) sobre σ_e^2, o que nos dará,

$$f(\sigma_\alpha^2|\vec{y}) = \omega \cdot \int_0^\infty f\left(\chi_{\nu_\alpha}^{-2} = \frac{\sigma_e^2 + n\sigma_\alpha^2}{\nu_\alpha QME}\right) \cdot f\left(\chi_{\nu_e}^{-2} = \frac{\sigma_e^2}{\nu_e QMD}\right) \cdot d\sigma_e^2,$$

$$\sigma_\alpha^2 > 0 \qquad (3.2.4.5)$$

onde

$$\omega = \frac{n \cdot (\nu_e QMD)^{-1} \cdot (\nu_\alpha QME)^{-1}}{Pr\{F_{\nu_\alpha,\nu_e} < QME/QMD\}}$$

Isto é, usando a "priori não-informativa" vista em (3.2.3.4), a distribuição $f(\sigma_\alpha^2|\vec{y})$ contém todo o conhecimento "a posteriori" sobre σ_α^2. Observe que $f(\sigma_\alpha^2|\vec{y})$ está definida no intervalo $(0, +\infty)$.

Não parece possível expressar $f(\sigma_\alpha^2|\vec{y})$ por funções simples, fazendo com que a densidade para cada valor de σ_α^2 seja obtida por processos numéricos. Em (3.2.4.5) a partir do momento em que temos as observações amostrais, passamos a conhecer as quantidades à esquerda da integral, e as duas funções que constituem o integrando dependem de σ_e^2 e σ_α^2. Se dermos valores a σ_α^2, e usarmos uma técnica adequada de integração numérica, obteremos os correspondentes valores de

$$f(\sigma_\alpha^2|\vec{y}).$$

Numa etapa posterior voltaremos a este problema quando da resolução de um exemplo.

Box & Tiao [1973], discutem algumas aproximações, que podem ser usadas, para expressar $f(\sigma_\alpha^2|\vec{y})$ por funções simples. No entanto, aqui não nos ateremos a esse problema.

Algumas características da "posteriori" de σ_α^2, dada por (3.2.4.5) são estudadas por Box & Tiao [1973]. Quando da resolução do exemplo a ser dado discutiremos algumas delas.

3.2.5 - OS ESTIMADORES DE BAYES

Quando especificamos a distribuição "a posteriori" de θ, temos resolvido o nosso problema de inferência bayesiana. No entanto, às vezes, alguém mais desavisado deseja representar θ de uma maneira mais simples, e a idéia natural é a de procurar uma constante que represente a distribuição "a posteriori". Ou melhor dizendo, desejamos sintetizar a informação contida na "posteriori", através de uma constante. Esta idéia inicial de simplificação, torna-se difícil, no momento em que desejamos saber qual a melhor constante e qual o método geral de escolha.

A essa constante acima comentada, alguns autores deram o nome de *Estimador de Bayes*. Barnett [1973] caracteriza como estimador de Bayes a *moda* da "posteriori", ja Mood, Graybill & Boes [1974] definem a *média* da "posteriori" como sendo o estimador de Bayes. Por essas duas definições,com a primeira acompanhada da justificativa: "... uma interpretação direta do valor mais provável para θ.."; e a segunda, sem maiores detalhes explicativos por parte dos autores,sentimos a arbitrariedade da idéia do "estimador de Bayes".

A idéia desses "estimadores" dentro do contexto da inferência bayesiana, não deixa de ser uma influência da teoria clássica. Esta influência dá margem a uma abordagem dual ao problema, pois especificamente, a "posteriori" é obtida, mas não usada diretamente para fazermos inferências sobre θ, e sim utilizada para sugerir constantes a serem usadas como "estimadores". Uma discussão detalhada desse problema encontra-se em Tiao & Box [1972]. Procurando dar uma melhor visão do problema resolvemos o exemplo abaixo.

3.3 - UM EXEMPLO

A partir da situação discutida em 1.1, obtemos

$$\nu_\alpha = 3, \quad \nu_e = 12, \quad n = 4, \quad QME = 3{,}7, \quad QMD = 54.$$

O nosso objetivo aqui é a obtenção da "posteriori" do componente de variância σ^2_α. Por (3.2.4.5) temos que, após algumas simplificações

$$F(\sigma^2_\alpha|\vec{y}) = \frac{n(\nu_\alpha QME)^{-1}(\nu_e QMD)^{-1}}{\Pr\{F_{\nu_\alpha,\nu_e} < QME/QMD\}} \cdot \frac{1}{\Gamma(\nu_\alpha|2)\cdot 2^{\nu_\alpha/2}\cdot\Gamma(\nu_e|2)\cdot 2^{\nu_e/2}} \cdot$$

$$\cdot\left(\frac{1}{\nu_\alpha QME}\right)^{-\left(\frac{\nu_\alpha}{2}+1\right)} \cdot \left(\frac{1}{\nu_e QMD}\right)^{-\left(\frac{\nu_e}{2}+1\right)} \cdot \int_0^\infty (\sigma^2_e + n\sigma^2_\alpha)^{-\left(\frac{\nu_\alpha}{2}+1\right)} \cdot$$

$$\cdot(\sigma^2_e)^{-\left(\frac{\nu_e}{2}+1\right)} \cdot \exp\left\{\frac{-\nu_\alpha QME}{2\sigma^2_e + 2n\sigma^2_\alpha}\right\} \cdot \exp\left\{\frac{-\nu_e QMD}{2\sigma^2_e}\right\} \cdot d\sigma^2_e.$$

Após a substituição dos valores vistos no início desta seção e notanto σ^2_α por X, σ^2_e por Z e o integrando por $G_X(X,Z)$ temos

$$f(X|\vec{y}) = 2{,}31\times 10^{16} \int_0^\infty G_X(X,Z)\,dZ$$

que é a "posteriori" de σ_α^2.

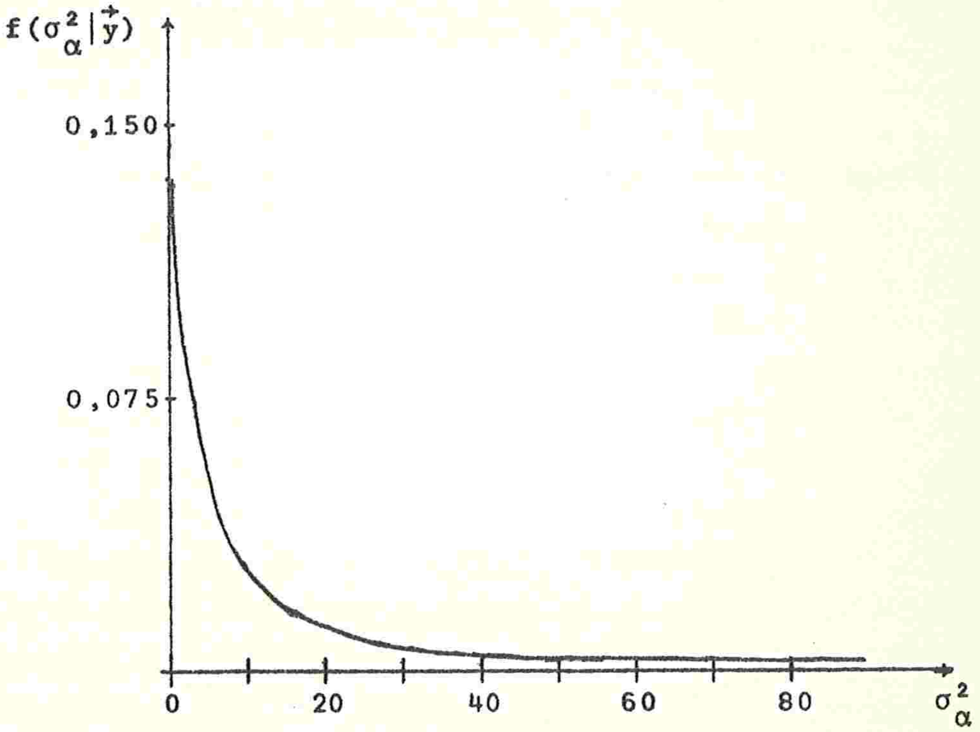
O valor de f para cada valor de X, pode ser obtido através da integração numérica da função G. A técnica de integração numérica usada foi o método de Simpson, ver Hamming [1971]. Após a obtenção desses valores, nos defrontamos com outro tipo de problema: conhecidas algumas ordenadas de

$$f(X|\vec{y})$$

e sabendo que $0 \leq X < \infty$, até que o valor da abcissa devemos ir, para que a integral de f de 0 até esse valor, forneça uma aproximação razoável da área total sob a curva. Obviamente por f se tratar de uma função densidade, a área total deve ser 1. Todos esses procedimentos numéricos foram executados num computador IBM/370, modelo
e no Apêndice C encontra-se o "programa" com base no qual esses cálculos foram executados. Nesse programa temos como saída as ordenadas de f a partir de $X = 0$ com incrementos na abcissa de 0,1, e também a área sob a curva de 0 a 1, 0 a 2, 0 a 3, etc. No primeiro teste com X variando de 0 a 250 obtivemos uma área sob a curva de 0,9781, e na execução final após uma série de testes, trabalhamos com X de 0 a 1.000, o que nos deu uma área total sob a curva igual a 0,9880, e um tempo total de computação de 38 minutos. Resolvemos adotar esse resultado, porque mesmo que quadruplicássemos esse tempo, obteríamos uma pequena melhora nesta área total, além de não afetar substancialmente as conclusões que poderíamos tirar. Abaixo segue um pequeno sumário de algumas ordenadas da "posteriori" de σ_α^2, assim como a idéia gráfica dessa distribuição

σ_α^2	$f(\sigma_\alpha^2 \vert \vec{y}) \times 10^5$
0	12.860
0,1	12.540

σ^2_α	$f(\sigma^2_\alpha \mid \vec{y}) \times 10^5$
0,2	12.260
0,3	11.970
0,4	11.700
0,5	11.430
1,0	10.240
2,0	8,358
3,0	6.950
4,0	5.869
5,0	4.338
10,0	2.356
20,0	981
50,0	216
100,0	52
250,0	4
1.000,00	0,2



DISTRIBUIÇÃO "A POSTERIORI" DO COMPONENTE σ^2_α

Nesse ponto, conseguimos atingir o desejo maior do Estatístico, que "acredita" na Inferência Bayesiana, que é a obtenção da "posteriori". Com base nela serão feitas as inferências necessárias.

É claro que, sob o ponto de vista numérico poderíamos refinar os resultados aqui encontrados, obtendo valores para as ordenadas de f usando incrementos nas abcissas menores que 0,1, e integrando f de 0 a um valor maior que 1.000. No entanto esse refinamento não ofereceria maiores vantagens, quando desejássemos tirar conclusões a partir da "posteriori".

Tentando caracterizar a inconsistência da abordagem clássica em cima da Inferência Bayesiana, voltamos à idéia (não aceita pelos "bayesianos") de tentar sumariar a informação contida na "posteriori", através de algumas características dela (estimadores de Bayes). No nosso exemplo encontramos

Moda = 0,0
Mediana = 7,4
Média = 21,3

E agora, o que fazer?

De acordo com Mood, Graybill & Boes [1974] deveríamos ter 21,3 como "estimador de Bayes" para o componente de variância σ_α^2. Ora, esse valor fica fora do intervalo

$$0 \leq \sigma^2 < 20,$$

cuja probabilidade é 0,75, em outras palavras, é um valor que está à direita do 3º quartil da distribuição "a posteriori". E convenhamos, não é razoável aceitar o fato de que, essa medida (média) represente melhor a nossa "posteriori" do que a moda (0,0) ou a mediana (7,4). Vale notar que, não somente os autores acima recomendam a *média* como estimador de

Bayes, mas também uma certa quantidade de autores de livros clássicos na Estatística.

Longe de nós a idéia de por exemplo, afirmar que a *moda* é um melhor estimador de Bayes, apesar de no presente exemplo, ela ser mais informativa do que a *média*. É lógico que numa outra situação, poderíamos ter a *média* como melhor representante.

A idéia que temos, que aceitamos, e que sentimos, é a de que admitir Estimador de Bayes, é trair a pureza da Inferência Bayesiana.

- 0 -

APÊNDICE A

A PRIORI NÃO-INFORMATIVA

Faremos algumas considerações sobre situações em que desejamos determinar uma "priori" para o parâmetro θ e que temos *pouco* ou *nenhum* conhecimento sobre ele, e nesse caso a nossa "priori" chamar-se-á *não-informativa*. Todo esse apêndice será desenvolvido a partir dos comentários feitos por Zellner [1971] sobre "prioris não-informativas", quando esse autor baseou-se nas idéias de Sir Harold Jeffreys, desenvolvidas no clássico (para os bayesianos, pelo menos) "Theory of Probability". Introduziremos alguns complementos extraídos de Box & Tiao [1973].

Das recomendações de Jeffreys, de como escolher uma distribuição "a priori", que represente ignorância sobre o parâmetro do modelo sob consideração, podemos extrair duas regras,

REGRA 1 - Se o parâmetro pode assumir qualquer valor no intervalo $(-\infty,+\infty)$ sua "priori" deverá ser tomada como sendo uniformemente distribuída.

EXEMPLO - Seja o caso de um parâmetro desconhecido μ digamos uma média, que possa assumir valores de $-\infty$ a $+\infty$. Então, de acordo com a regra,

$$f(\mu)d\mu\alpha d\mu, \quad -\infty<\mu<\infty \qquad (A.A1)$$

ou

$$f(\mu)\alpha \text{ constante.}$$

REGRA 2 - Se o parâmetro pode assumir qualquer valor no intervalo $(0,+\infty)$, a "priori" de seu logaritmo deverá ser tomada como sendo uniformemente distribuída.

EXEMPLO - Seja o caso de um parâmetro desconhecido σ, um desvio padrão. Para tal parâmetro Jeffreys sugere

$$\theta = \log \sigma, \text{ e}$$

$$f(\theta)d\theta\alpha d\theta, \quad -\infty<\theta<\infty \qquad (A.A2)$$

e como $d\theta/d\sigma = 1/\sigma$, segue que

$$F(\sigma)d\sigma\alpha 1/\sigma d\sigma, \quad 0<\sigma<\infty$$

Na situação da *Regra 1* nós vemos que a densidade sugerida, é obviamente imprópria, pois

$$\int_{-\infty}^{\infty} f(\mu)d\mu = \infty.$$

Do ponto de vista de Jeffreys isto é uma vantagem, pois

$$Pr\{a<\mu<b\}/Pr\{c<\mu<d\} = 0/0$$

é indeterminado onde a, b, c e d são quaisquer números finitos, indicando que não podemos fazer qualquer afirmação sobre a chance de que μ pertença a algum par de intervalos finitos. Esta propriedade foi vista por Jeffreys como uma representação formal da ignorância. No entanto, o problema da f.d.p. imprópria persiste. Uma idéia que surge naturalmente é a de considerarmos uma f.d.p. própria da forma

$$f(\mu)d\mu = \frac{1}{2M}\, d\mu\,, \quad -M \leq \mu \leq M \qquad (A.A3)$$

Mas, infelizmente quando assim procedemos, introduzimos uma informação "a priori" sobre a amplitude de variação de μ e então, deixamos de ter ignorância completa sobre μ. No entanto, se fizermos $M \to \infty$, podemos olhar (A.A3) como uma boa aproximação para (A.A1).

Como o objetivo de complementar o raciocínio de Jeffreys, Zellner [1971] introduz uma medida de informação dada por

$$H = \int_{-M}^{M} f(\mu) \log f(\mu) d\mu \qquad (A.A4)$$

e verifica que a f.d.p. própria que minimiza H é (A.A3), e conclui dizendo que a f.d.p. uniforme é uma "priori de informação minimal", e melhor ainda, quando $M \to \infty$ (A.A3) é uma boa aproximação para (A.A.1). Além disso Zellner [1971] mostra através de um exemplo que a combinação de uma f.d.p. "a priori" imprópria (A.A.1) com uma função de verossimilhança, através do Teorema de Bayes, fornece uma f.d.p. "a posteriori" própria.

Com relação à *Regra 2*, a mais importante observação feita por Jeffreys é que (A.A2) é invariante sobre transformações da forma $\phi = \sigma^n$, isto é, $d\phi = n\sigma^{n-1}d\sigma$ e então

$$d\phi/\phi \alpha d\sigma/\sigma.$$

Além disso ele mostra que $Pr\{0<\sigma<a\}/Pr\{a<\sigma<\infty\}$ é indeterminada, pois

$$\int_0^\infty d\sigma/\sigma = \infty, \quad \int_0^a d\sigma/\sigma = \infty \quad \text{e} \quad \int_a^\infty d\sigma/\sigma = \infty,$$

sendo isto uma representação formal da ignorância. Zellner

[1971] mostra que a função H será minimizada tomando $f(\theta)\alpha$ constante, e isto é uma justificativa teórica de informação, para tomarmos $\theta = \log\sigma$ uniformemente distribuída. Pela propriedade de invariância $f(\sigma)\alpha 1/\sigma$ implica $f(\sigma^2)\alpha 1/\sigma^2$, que representam a ignorância sobre σ ou σ^2, respectivamente.

Com respeito ainda às idéias de Jeffreys com relação à "prioris não-informativas" Box & Tiao [1973] dão a seguinte

<u>REGRA DE JEFFREYS</u> - A distribuição a priori para um parâmetro θ é aproximadamente não-informativa se é tomada proporcional à raiz quadrada da medida de informação de Fisher. Isto é,

$$f(\theta)\alpha I^{1/2}(\theta) \qquad \text{(A.A5)}$$

onde *medida de informação de Fisher* sobre θ na amostra $y' = (y_1,\ldots,y_n)$ é definida como

$$I_n(\theta) = E_{\vec{y}|\theta}\left[-\frac{\partial^2 \log f(\vec{y}|\theta)}{\partial\theta^2}\right] = E_{\vec{y}|\theta}\left[\frac{\partial \log f(\vec{y}|\theta)}{\partial\theta}\right]^2$$

e

$$I(\theta) = E_{y|\theta}\left[-\frac{\partial^2 \log f(y|\theta)}{\partial\theta^2}\right]$$

então, quando $\vec{y}$ é uma amostra aleatória $I_n(\theta) = nI(\theta)$. Se aplicarmos esta regra para obtermos "prioris" com relação aos parâmetros (μ,σ) da Normal, nós obtemos (ver Box & Tiao [1973]) quando

— μ desconhecido e σ conhecido.

$$I(\theta) = \text{constante} \Longrightarrow p(\mu)\alpha\ \text{constante}$$

— σ desconhecido e μ conhecido

$$I(\theta) = \frac{1}{\sigma^2}\times\text{constante} \Longrightarrow p(\sigma)\alpha 1/\sigma$$
$$\text{ou } p(\log\sigma)\alpha \text{ constante.}$$

E teremos como uma extensão multiparamétrica a seguinte

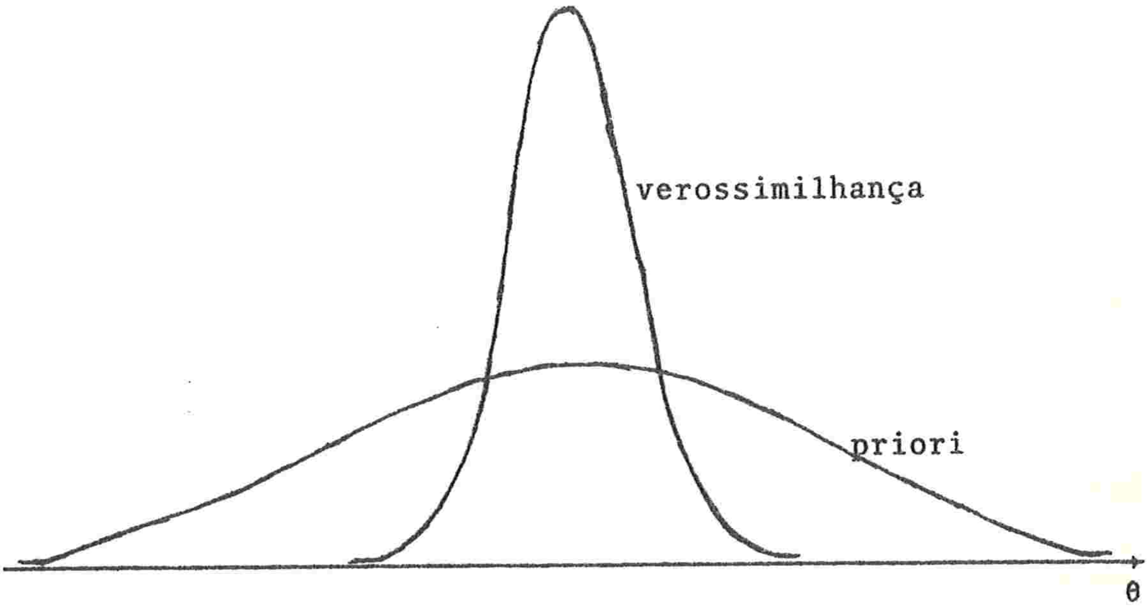
<u>REGRA DE JEFFREYS "MULTIPARAMÉTRICA"</u> - A distribuição a priori para um conjunto de parâmetros é tomada como sendo proporcional à raiz quadrada do determinante da matriz informação. Isto é

$$f(\vec{\theta})\alpha|I_n(\vec{\theta})|^{1/2} \qquad \text{(A.A6)}$$

Jeffreys obteve esta regra, como no caso do parâmetro simples, requerendo *invariância* sobre transformações de parâmetros. Esta regra no entanto, deve ser aplicada com cuidado, principalmente quando parâmetros de locação e escala ocorrem simultaneamente.

Box & Tiao [1973] fazem algumas considerações sobre alguns "cuidados" que devem ser tomados na aplicação da Regra de Jeffreys. É enfatizado por eles a necessidade de investigarmos as implicações das transformações em cada situação à luz de qualquer conhecimento a priori sobre independência. Quando, por exemplo, em termos de um julgamento a priori temos que θ_1 em $\vec{\theta}' = (\theta_1, \theta_2, \ldots, \theta_k)$ é independente de $(\theta_2, \ldots, \theta_k)$ aplicamos a regra separadamente para θ_1 e $(\theta_2, \ldots, \theta_k)$.

Alguns autores hesitam em empregar as f.d.p. impróprias recomendadas por Jeffreys. Em contrapartida eles introduzem "prioris localmente uniformes", que são prioris "razoavelmente achatadas" sobre o intervalo no qual a verossimilhança assume valores apreciáveis, e não assume grandes valores fora desses intervalos (ver figura na página seguinte).



Como a posteriori α priori × verossimilhança, para o caso de uma "priori localmente uniforme", teríamos

posteriori α verossimilhança

A grande vantagem da priori localmente uniforme é que trabalhamos com uma f.d.p. própria. Já como desvantagens temos que a condição de Jeffreys para ignorância completa não está satisfeita, além do fato de termos que conhecer alguma coisa sobre a verossimilhança, o que pode não ser o caso em algumas situações gráficas. No entanto se temos informação sobre a variação de θ e a função de verossimilhança é disponível, ela pode ser usada com boa vantagem. Quando esse tipo de informação não existe, pouca diferença prática existe se nós usamos uma priori "localmente uniforme" ou uma f.d.p. imprópria de Jeffreys, Maiores detalhes sobre as idéias envolvidas na escolha de prioris que representam *pouco* ou *nenhum* conhecimento sobre o parâmetro, podem ser encontrados em Box & Tiao [1973].

- ▯ -

APÊNDICE B

A DISTRIBUIÇÃO DO χ^{-2}

Um dos resultados de importância em Estatística, é que se $X \sim N(\mu,\sigma^2)$, e se uma amostra $\vec{X}' = (X_1,\ldots,X_n)$ dessa população é retirada, a quantidade

$$\chi_n^2 = \sum_{i=1}^{n} \frac{(X_i - \mu)^2}{\sigma^2} = \frac{ns^2}{\sigma^2} \qquad \text{(A.B1)}$$

é

$$f(\chi_n^2) = [\Gamma(n/2)2^{n/2}]^{-1} \cdot (\chi^2)^{(n/2)-1} \cdot \exp\left\{-\frac{1}{2}\chi^2\right\}, \quad \chi^2 > 0 \qquad \text{(A.B2)}$$

que é a distribuição χ^2 com n graus de liberdade.

Na teoria clássica (não-bayesiana), se desejamos fazer inferências sobre σ^2, podemos sem maiores problemas usar a distribuição χ^2.

Na teoria bayesiana, o problema de fazer inferências sobre σ^2, não é tão direto, isto porque deveremos ter a distribuição "a posteriori" de σ^2, e esta poderá não ser uma χ^2.

No caso de termos n observações de $X \sim N(\mu,\sigma^2)$ e, *pouco* ou *nenhum* conhecimento "a priori" sobre σ^2.

$$L(\sigma^2|\vec{x})\alpha(\sigma^2)^{-n/2}\cdot\exp\left\{-\frac{ns^2}{\sigma^2}\right\} \qquad \text{(A.B3)}$$

$$f(\sigma^2)\alpha\sigma^{-2} \quad \text{(ver Apêndice A)} \qquad \text{(A.B4)}$$

e a distribuição "aposteriori" de σ^2, será dada por

$$f(\sigma^2|\vec{x})\alpha f(\sigma^2)\cdot L(\sigma^2|\vec{x}) \quad \text{ou} \qquad \text{(A.B5)}$$

$$f(\sigma^2|\vec{x}) = k\cdot(\sigma^2)^{[(n/2)+1]}\cdot\exp\left\{-\frac{ns^2}{2\sigma^2}\right\},\ \sigma^2>0 \qquad \text{(A.B6)}$$

A constante de normalização de $\underline{k}$ deve garantir que

$$\int_0^\infty f = 1$$

Assim procedendo obteremos

$$f(\sigma^2|\vec{x}) = \frac{(ns^2/2)^{n/2}}{\Gamma(n/2)}\cdot(\sigma^2)^{-[(n/2)+1]}\cdot\exp\left\{-\frac{ns^2}{2\sigma^2}\right\},\ \sigma^2>0 \quad \text{(A.B7)}$$

O resultado acima motivou o "χ^2 invertido" que é obtido de (A.B1), fazendo a tranformação

$$\chi_n^{-2} = \frac{1}{\chi_n^2} \qquad \text{(A.B8)}$$

para produzir

$$f(\chi_n^{-2}) = [\Gamma(n/2)2^{n/2}]^{-1}\cdot(\chi^{-2})^{-[(n/2)+1]}\cdot\exp\left\{-\frac{1}{2\sigma^{-2}}\right\},\ \chi^{-2}>0 \qquad \text{(A.B9)}$$

Se compararmos (A.B9) com (A.B7), veremos que "a posteriori" σ^2/ns^2 tem distribuição χ_n^{-2}.

- 0 -

APÊNDICE C

UM "PROGRAMA" PARA ENCONTRAR A "POSTERIORI" DE σ^2_α

```
DOS FORTRAN IV 360N-FO-479 3-8         MAINPGM          DATE   07/06/76    TIME    09.43.18     PAGE 0001
              C  ***********************************************************************
              C
              C  VARIAVEIS INTEIRAS UTILIZADAS NO PROGRAMA
              C
              C  I - CHAVE PARA DECIDIR QUE FATOR MULTIPLICA A ORDENADA DA G
              C  ICONT - CONTADOR DE ORDENADAS DA F P/ CALCULO DE INTEGRAIS INTERMEDIA
              C    RIAS
              C  J - CHAVE USADA NA MONTAGEM DAS ORDENADAS DA F
              C  KK - CHAVE USADA PARA DECIDIR SOBRE O USO DO TESTE DA MEDIANA OU DO V
              C    ALOR FIN
              C  L - CHAVE PARA CONTROLAR O RETORNO DA ROTINA DE CONTAGEM DE LINHAS
              C  LINHA - CONTADOR DE LINHAS
              C  NR - NUMERO DO EQUIPAMENTO DE ENTRADA
              C  NW - NUMERO DO EQUIPAMENTO DE SAIDA
              C
              C  ***********************************************************************
              C
              C  ***********************************************************************
              C
              C  VARIAVEIS REAIS UTILIZADAS NO PROGRAMA
              C
              C  A,B,C,D,E - PARAMETROS PARA CALCULO DA ORDENADA DA G
              C  DIF1 - DIFERENCA ENTRE 0.500 E A INTEGRAL DA G
              C  DIF2 - DIFERENCA ENTRE 1.000 E A INTEGRAL DA G
              C  DX,DY - INCREMENTOS DE X E Y, RESPECTIVAMENTE
              C  GG - ORDENADA DA G
              C  H1,H2,H3 - ORDENADAS DA F PARA USO NA REGRA DE SIMPSON ITERARIVA
              C  S,S1,S2,S3 - INTEGRAIS DA , USADAS PARA MONTAR AS ORDENADAS DA F
              C  SM - VALOR DA MEDIA DA FUNCAO F
              C  SS - INTEGRAL DA F
              C  X,XN - VALORES INICIAL E FINAL DE X
              C  Y,YN - VALORES INTERMEDIARIO E FINAL DE Y
              C
              C  ***********************************************************************
              C
0001                 EXTERNAL DEXP
0002                 REAL * 8 X,Y,DY,YN,A,B,C,D,S,G,E,SS,DIF1,DIF2,DX,S1,S2,S3
0003                 REAL * 8 H1,H2,H3,SM,XMED
0004                 G(A,B,E)  = 2.31E+16 * A * B *DEXP(E)
              C  INICIALIZACAO DE CAMPOS
0005                 I = 1
0006                 SM = 0.
0007                 SS = 0.
0008                 J = 1
0009                 ICONT = 0
0010                 KK = 1
0011                 NR = 1
0012                 NW = 3
0013                 LINHA = 0
              C  SALTA PARA A PRIMEIRA FOLHA
0014                 WRITE (NW,1)
0015               1 FORMAT ('1',30X,'CALCULO DAS INTEGRAIS DAS FUNCOES DA FORMA'//
                    109X,'G = 2.31E+16 * ((Y+4)*X) ** (-2.5) * Y**(-7) * EXP((-0.5)*',
                    2'(659*Y+2592*X)/(Y+4*X)*Y)'//)
```

```
DOS FORTRAN IV 360N-FO-479 3-8          MAINPGM          DATE  07/        09.43.18        PAGE 0002
      C  LEITURA DO X INICIAL, DOS INCREMENTOS DX E DY
      C  E DOS VALORES FINAIS XN E YN
         READ (NR,5) X,DX,XN,DY,YN
       5 FORMAT (F4.1,F2.1,F6.1,F3.1,F3.0)
      C  INICIO DA ROTINA
       6 Y = 0.
         S = 0.
      C  CALCULO DA ORDENADA DE G
       7 Y = Y + DY
         A=(Y+4*X)**(-2.5)
         B = Y ** (-7)
         C = (659 * Y + 2592 * X) * (-0.5)
         D = (Y + 4 * X) * Y
         E = C / D
         GG= G(A,B,E)
      C  VERIFICA SE JA ATINGIU O ULTIMO VALOR DO INTERVALO DO Y
         IF ( Y - YN ) 9,15,15
      C  DECIDIR QUAL FATOR MULTIPLICA A ORDENADA DA G
       9 GO TO (11,13),I
      11 GG= 4 * GG
         S = S + GG
         I = 2
         GO TO 7
      13 GG= 2 * GG
         S = S + GG
         I = 1
         GO TO 7
      C  ADICIONA ULTIMA ORDENADA, E CALCULA VALOR DA INTEGRAL
      15 S = S + GG
         S = DY * S
         S = S / 3
      C  MONTAGEM DAS ORDENADAS DA F
      42 CONTINUE
         GO TO (22,23,24),J
      22 S1 = S
         H1 = X * S1
         J = 2
         GO TO 25
      23 S2 = 4 * S
         H2 = X * S2
         J = 3
         GO TO 25
      24 S3 = S
         H3 = X * S3
         J = 2
      C  CALCULO ITERATIVO DA INTEGRAL DA F, E DA MEDI
         SS = SS + DX * (S1 + S2 + S3) / 3
         S1 = S3
         SM = SM + DX * (H1 + H2 + H3)/3
         H1 = X * S3
         ICONT = ICONT + 1
         IF(ICONT -5 )39,35,35
      35 WRITE (NW,37) X,SS
      37 FORMAT (/5X,'A AREA SOB A F, ATE ',F7.1,' VALE ',E10.4/,
```

```
      ICONT = 0
      L = 2
      GO TO 49
   39 CONTINUE
      GO TO (30,28),KK
C  TESTE PARA ENCONTRAR A MEDIANA
   30 DIF1 = 0.500 - SS
      IF (0.001 - DIF1 ) 25,26,25
   26 WRITE (NW,27) X,SS
   27 FORMAT (/5X,'SENDO A INTEGRAL DE 0 A ',F5.1,' IGUAL A ',E10.4,
     1', AQUELE E A PROVAVEL MEDIANA'/)
      XMED = X
      L = 3
      GO TO 49
   44 CONTINUE
      KK = 2
      GO TO 25
C  TESTE PARA ENCONTRAR O VALOR FINAL
   28 DIF2 = 1.000 - SS
      IF (0.001 - DIF2) 25,31,31
   31 WRITE (NW,29) X,SS
   29 FORMAT (//5X,'A AREA ATE ',F7.1,' VALE ',E10.4)
      L = 4
      GO TO 49
   46 CONTINUE
      WRITE (NW,41) SM,XMED
   41 FORMAT (//5X,'A MEDIA VALE ',E10.4//5X,'A MEDIANA VALE ',E10.4//)
      GO TO 19
   25 X = X + DX
      IF (X - XN) 6,31,31
C  CONTROLE DE PAGINAS
   49 LINHA = LINHA + 1
      IF (LINHA - 35) 53,51,51
   51 WRITE (NW,1)
      LINHA = 0
   53 CONTINUE
      GO TO (42,39,44,46),L
   19 WRITE(NW,77)
   77 FORMAT(5X,'A MODA VALE 0.0000E 00')
      WRITE(NW,21)
   21 FORMAT (5X,'UFA...  FINALMENTE TERMINEI ESTA DROGA')
      STOP
      END
```

REFERÊNCIAS BIBLIOGRÁFICAS

ANDERSON, R.L. -1965- Negative variance estimates, Tecnometrics, 7: 75-76.

BARNETT, V -1973- Comparative Statistical Inference, N.York: John Wiley.

BOX, G.E.P. and TIAO, G.C. -1970- Bayesian Inference in Statiscal Analysis, Reading, Mass: Addison-Wesley.

GRAYBILL, F.A. -1961- An Introduction to Linear Statistical Models I, New York, McGraw-Hill.

GRAYBILL, F.A. and HULGUIST, R.A. -1961- Theorems concerning Eisenhart's Model II, Ann.Math.Statist. 32:261-269.

HAMMING, R.W. -1971- Introduction to Applied Numerical Analysis, New York: McGraw-Hill.

HERBACH, L.H. -1959- Properties of Model II type analysis of variance tests, Ann.Math.Statist. 30: 939-959.

HILL, B.M. -1965- Inference about variance components in the one-way model. J.Amer.Statis.Ass. 60: 806-825.

LAMOTTE, L.R. -1973- On non-negative quadratic unbiased estimation of variance components. J.Amer.Statist.Ass. 68: 728-730.

MOOD, A.M., GRAYBILL, F.A. and BOES, D.C. -1974- Introduction to the Theory of Statistics, 3d.ed. Tokyo: McGraw-Hill, Kogakusha.

NELDER, J.A. -1954- The interpretation of negative components of variance, Biometrika, 41: 544-548.

RAO, C.R. -1972- Estimation of variance and covariance components in linear models, J.Amer.Statist.Ass. 67:112-115.

RAO, C.R. -1973- Linear Statistical Inference and Applications, 2d. ed. New York, John Wiley

SEARLE, S.R. -1968- Another look at Henderson's methods of estimating variance components. Biometrics, 24: 749-788.

SEARLE, S.R. -1971a- Topics in variance components estimation, Biometrics, 27: 1-76.

SEARLE, S.R. -1971b- Linear Models, New York, John Wiley.

SEARLE, S.R. and FAWCETT, R.F. -1970- Expected mean square in variance components models having finite populations, Biometrics, 26: 243-254.

THOMPSON, W.A. Jr. -1962- The problem of negative estimates of variance components, Ann.Math.Statist.,33: 273-289.

THOMPSON, W.A.Jr. and MOORE, J.R. -1963- Non negative estimates of variance components. Technometrics, 5:441-450.

TIAO, G.C. and Box, G.E.P. -1972- Some Comments on "Bayes" Estimations. Univ.of Wisconsin, Dep.of Statistics, Tech. Report, nº 297.

TIAO, G.C. and TAN, W.Y. -1965- Bayesian analysis of random-effect models in the analysis of variance I.Posterior distribution of variance-components. Biometrika, 52:37-53.

ZELLNER, A. -1971- An Introduction to Bayesian Inferences in Econometrics, New York, John Wiley.

- 0 -