

**AVALIAÇÃO DE MÉTODOS DE ESTIMAÇÃO DE  
PARÂMETROS EM MODELO DE COVARIÂNCIA  
COM ERRO NA COVARIÁVEL**

**TIAGO ALMEIDA DE OLIVEIRA**

**2008**

**Ficha Catalográfica Preparada pela Divisão de Processos Técnicos da  
Biblioteca Central da UFLA**

Oliveira, Tiago Almeida de.

Avaliação de métodos de estimação de parâmetros em modelo de covariância com erro na covariável / Tiago Almeida de Oliveira. – Lavras : UFLA, 2008.

59 p. : il.

Dissertação (Mestrado) Universidade Federal de Lavras, 2008.

Orientador: Augusto Ramalho de Moraes.

Bibliografia.

1. Análise de covariância. 2. Acurácia. 3. Precisão. I. Universidade Federal de Lavras. II. Título.

CDD - 519.538

**TIAGO ALMEIDA DE OLIVEIRA**

**AVALIAÇÃO DE MÉTODOS DE ESTIMAÇÃO DE PARÂMETROS EM  
MODELO DE COVARIÂNCIA COM ERRO NA COVARIÁVEL**

Dissertação apresentada à Universidade Federal de Lavras, como parte das exigências do Programa de Pós-graduação em Estatística e Experimentação Agropecuária, para obtenção do título de “Mestre”.

APROVADA em 15 de fevereiro de 2008.

Prof. Dr. Luiz Alberto Beijo

UNIFAL

Prof. Dr. Rogério de Melo Costa Pinto

UFU

Prof. Dr. Marcelo Angelo Cirillo

UFLA

Prof. Dr. Augusto Ramalho de Moraes  
UFLA  
(Orientador)

LAVRAS  
MINAS GERAIS - BRASIL

*Aos meus pais, Ednaldo e Maria,  
pelo imenso amor que me deram,  
cuidados e apoio.*

## AGRADECIMENTOS

A Deus por estar sempre presente e nunca me deixar faltar forças e ânimo na caminhada que trilho. Agradeço pelo mundo em que vivemos e pelos imensos detalhes da vida;

A minha família, em especial ao meu pai e minha mãe. Por estarem presentes em todas as fases de minha vida, dando conselhos sábios e por sempre me apoiarem e incentivarem. E ao meu irmão por estar sempre disposto a ajudar e ser um exemplo de ser humano;

Ao meu primo/irmão Gabriel, agradeço por toda confiança depositada em mim, amizade insubstituível e exemplo de coragem, mesmo quando estive longe este presente.

Ao professor Dr. Augusto Ramalho, pela orientação, ensinamentos, amizade, conselhos, apoio, confiança e paciência durante os últimos dois anos.

Ao professor Dr. Marcelo Angelo Cirillo, pela co-orientação, preocupação, apoio, incentivo e sugestões tão valiosas para a realização deste trabalho;

À Universidade Federal de Lavras, em especial ao Departamento de Ciências Exatas, pela oportunidade de concretização deste trabalho;

A meus Tios toda a gratidão que posso mensurar especialmente a Tio Nivaldo e Tia Luiza. Graças a vocês hoje sou um estatístico e mestre. Obrigado por toda a paz e confiança que vocês compartilharam comigo nos anos que morei em Campina Grande e por me acolherem com tanta hospitalidade;

Agradeço a meus primos em especial a Tatiane e Thaísinha, Alessandra, Edson, Renato, Erika, Elaine, Marcos e Roberto (primo de consideração) pelos conselhos e por sempre acreditarem em mim e por todos vocês sempre serem meus irmãos;

A Maria e família por me terem como um subrinho, e eu a vocês como tios,  
Também sou grato por toda a ajuda e incentivo que tive;

Ao Geraldo(corintiano) por ser um verdadeiro tio e amigo para comigo e um  
irmão e amigo para com o meu pai.

Aos anjos que encontrei em Lavras, na figura de Dona Magda & família e Ed-  
jane & família. Vocês ao me acolherem foram o eixo motriz da minha caminhada.

A meus amigos de Universidade e convívio nestes dois anos na UFLA, obri-  
gado por nunca me deixarem desanimar e pelas festas na laje maravilhosas que  
compartilhei em família. Agradeço incomensuravelmente a Ricardo (Oliveira) e  
ao Marcelino (popó) pela incessante ajuda neste trabalho. E ao Renato, Fabrício e  
Dorival pelas boas horas de conversas e amizade verdadeira.

Ao quarteto fantástico de Probabilidade (Marcelino, Ricardo e Renato), pelas  
horas de angustia, estudos e felicidade após o resultado do esforço, e pela amizade  
verdadeira, vocês são seres humanos fantásticos;

A todos os professores que contribuíram para o enriquecimento dos conheci-  
mentos indispensáveis a este trabalho.

Aos funcionários do Departamento de Ciências Exatas, em especial à Selmi-  
nha, Edila, Josi e Joyce pela eficiência e amizade.

A Universidade Estadual da Paraíba - UEPB, pelos ensinamentos que serviram  
de base para a realização deste curso. E a todos os professores em especial a João  
Gil de Luna, Antônio Ricardo e a Ruth Silveira pelos ensinamentos, paciência,  
compreensão e encorajamento;

Agradeço a todos que contribuíram direta ou indiretamente neste estado abençoa-  
do por Deus, e por estarem presentes nesta jornada.

Ao CNPq, pelo apoio financeiro e por tornar possível esse sonho.

**MUITO OBRIGADO!!!**

## SUMÁRIO

LISTA DE TABELAS . . . . .	i
LISTA DE FIGURAS . . . . .	ii
RESUMO . . . . .	iv
ABSTRACT . . . . .	v
1 INTRODUÇÃO . . . . .	1
2 REVISÃO BIBLIOGRÁFICA . . . . .	3
2.1 Análise de covariância . . . . .	6
2.1.1 Modelo estatístico . . . . .	9
2.1.2 Equações normais . . . . .	10
2.1.3 Abordagem matricial . . . . .	15
2.2 Erro nas variáveis . . . . .	17
2.2.1 Estimador de James-Stein . . . . .	20
2.2.2 Estimador de Fuller . . . . .	22
2.2.3 Acurácia e precisão . . . . .	23
2.3 Coeficiente de determinação do modelo . . . . .	26
2.4 Simulação Monte Carlo . . . . .	26
3 METODOLOGIA . . . . .	28
4 RESULTADOS E DISCUSSÃO . . . . .	33
5 CONCLUSÕES . . . . .	47
6 ESTUDOS FUTUROS . . . . .	49
REFERÊNCIAS BIBLIOGRÁFICAS . . . . .	50
ANEXOS . . . . .	54

## LISTA DE TABELAS

2.1	Exemplo de tabela para a análise de covariância. . . . .	14
4.1	Viés médio das estimativas dos parâmetros do modelo linear sobre diferentes qualidades de ajuste ( $R^2$ ), tamanhos amostrais e as três distribuições dos erros, nos dois métodos de estimação avaliados. .	44

## LISTA DE FIGURAS

2.1	Ilustração com os conceitos de precisão e acurácia (Fonte: Araújo Neto(2006)) . . . . .	24
4.1	Viés médio do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	34
4.2	Viés médio do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	36
4.3	Viés médio do estimador atenuador de vício considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	37
4.4	Viés médio do estimador atenuador de vício considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	39
4.5	Erro residual do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	40

4.6	Erro residual do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	41
4.7	Erro residual do estimador atenuador de vício, considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	42
4.8	Erro residual do estimador atenuador de vício, considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento. . . . .	43

## RESUMO

OLIVEIRA, Tiago Almeida de. **Avaliação de Métodos de Estimação de Parâmetros em Modelo de covariância com Erro na Covariável**. 2008. 59 p. Dissertação (Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras, MG.\*

Neste trabalho estudou-se o modelo de análise de covariância com um fator e erro de medida na covariável, avaliamos neste modelo por meio de simulação, a acurácia e precisão de dois estimadores, propostos na literatura para estimar parâmetros de um modelo de regressão com erro de medida. A diferença entre os métodos de estimação é que o estimador "Plug-in" estima o valor verdadeiro por meio dos valores observados e a partir disso utiliza a função de estimação usual para estimar o parâmetro de interesse e o estimador conhecido como atenuador de vício apenas realiza uma correção no vício a partir do estimador usual por meio do cálculo de um fator. Sobre diferentes distribuições dos resíduos, coeficientes de determinação e tamanhos amostrais, estudamos o comportamento de ambos os estimadores. As conclusões encontradas sugerem que o estimador "Plug-in" obteve melhor desempenho, tanto na acurácia quanto na precisão em situações de normalidade, nas diferentes configurações analisadas sobre o modelo avaliado. Outro ponto importante que deve ser ressaltado é que no modelo de análise de covariância quanto maior o tamanho amostral e a coeficiente de determinação, melhor se comportam os estimadores avaliados com relação a acurácia e a precisão. Quando os estimadores foram avaliados no modelo de ANCOVA com os resíduos distribuídos pela Gama, os mesmos obtiveram o pior desempenho em relação quando eles eram avaliados pelas demais distribuições.

---

\* **Comite Orientador:** Prof. Dr. Augusto Ramalho de Moraes - UFLA (orientador) e Prof. Dr. Marcelo Angelo Cirillo - UFLA (co-orientador)

## ABSTRACT

OLIVEIRA, Tiago Almeida de. **Evaluation of parameter estimation methods for covariance analysis models with error in covariate.** 2008. 59 p. Dissertation (Master in Statistics and Agricultural Research) - Federal University of Lavras, Lavras, MG.\*

The present work approaches the covariance analysis model with one factor and measurement error in the covariate. Accuracy and precision of two estimators suggested in the literature were evaluated through data simulation, for estimating parameters of a regression model with measurement error. So called “Plug-in” method estimates the real value based on the observed ones and then uses the common function for estimating the desired parameter. The other estimator, known as bias smoother, only performs a bias correction on the usual estimator by computing a factor. Behavior of both estimators was studied under different residual distributions, goodness of fit and sample sizes. Results suggest that the “Plug-in” estimator presented the best performance both for accuracy and precision under normality, for the distinct evaluated situations. It is worth noting that, in covariance analysis model, the high the sample size, the better for accuracy and precision. When the estimators had been evaluated in the model of ANCOVA with the residues distributed for Gamma, the same ones had gotten the worse performance in relation when they were evaluated by the others distributions.

---

\* **Guidance committee** Prof. Dr. Augusto Ramalho de Morais - UFLA (supervisor) and Prof. Dr. Marcelo Angelo Cirillo - UFLA

## 1 INTRODUÇÃO

A análise de covariância (ANCOVA) é comumente descrita para o ajuste de variáveis que não podem ser controladas pelo pesquisador. Ela permite uma correção das médias dos tratamentos após um ajuste com os valores das covariáveis, ou seja, remove a variância associada às observações adicionais (covariáveis) dos resultados da variável dependente, antes de fazer as comparações entre as médias. É muito utilizada em aplicações na experimentação agrícola para o controle do erro experimental, interpretação de resultados experimentais e estimar dados perdidos.

Assume-se que sua maior suposição, em caso de aplicação, é que a covariável não pode ser afetada pelos tratamentos utilizados. Usualmente, os modelos de covariância são ajustados considerando-se a covariável como um efeito fixo e medida sem erro. Porém, em muitas situações, é necessário utilizar a amostragem com repetição para contemplar efeitos aleatórios e deste modo, levar em conta fatores do acaso que interferem no processo de medição, tais como os erros nas variáveis.

Em virtude de que a medição de um resultado está sujeita a pequenas variações aleatórias, nota-se que o enfoque estatístico poderá ser dado em função da modelagem de dados, considerando o erro aleatório nas covariáveis (fatores). Dessa forma, o processo de modelagem, certamente, agregará informações mais precisas, dentro da confiabilidade metrológica, quando comparada à modelagem feita de forma convencional, assumindo efeito fixo nas covariáveis.

Estimadores foram desenvolvidos na literatura para modelar dados com erros nas covariáveis. Dentre eles, existem o estimadores de James & Stein (1961) e Fuller (1987), conhecidos também como estimadores "Plug-in" e atenuador de vício, respectivamente. Ambos os estimadores foram desenvolvidos com a fina-

lidade de minimizar os riscos de incerteza, minimizando também o efeito do erro de medida.

Basicamente, a diferença na formalização desses métodos, citada por Cunha & Colosimo (2003), é que os estimadores “Plug-in” estimam o valor verdadeiro por meio dos valores observados e, de posse dos respectivos valores ajustados, utilizam a função de estimação usual no modelo estudado para estimar o parâmetro de interesse. Já o caso do estimador conhecido como atenuador de vício elucidado anteriormente, este apenas realiza uma correção no vício a partir do estimador usual.

O propósito deste trabalho foi o de avaliar dois métodos de estimação de parâmetros no modelo de ANCOVA com erro nas covariáveis, uma vez que não se conhece o comportamento destes estimadores quando aplicados a este modelo, e sabe-se que no modelo de ANCOVA com erro nas covariáveis as estimativas de mínimos quadrados, para os seus parâmetros, são viciadas. Este trabalho então propõe-se à avaliar a acurácia e a precisão dos estimadores “Plug-in” e atenuador de vício considerando diferentes distribuições nos resíduos de um modelo linear simples de ANCOVA com erro nas covariáveis, valores do coeficiente de determinação deste modelo ( $R^2$ ) e tamanhos amostrais, utilizando-se para isso técnicas Monte Carlo.

## 2 REVISÃO BIBLIOGRÁFICA

A teoria clássica de ANCOVA conforme elucidado em textos clássicos da estatística experimental, como Federer (1955) e Montgomery (1976), entre outros, consideram a covariável  $x$  de efeito fixo e medida sem erro. No entanto, às vezes, esta suposição não é satisfeita. Mesmo que os constantes avanços tecnológicos estejam tornando-se cada vez mais precisos os procedimentos de mensuração das variáveis, não é realista considerar que a covariável  $x$  seja medida sem erro. O mais comum é não se ter acesso aos seus verdadeiros valores (Souza, 2004).

Como o interesse recai, geralmente, em estimar os parâmetros de uma função, a presença de erros de mensuração pode afetar a precisão dos estimadores, como, por exemplo, a comparação entre dois ou mais grupos, em que parte da diferença encontrada (ou mesmo toda) pode ser devida a tais erros (Bolfarine et al., 1992). Também na aplicação de determinada concentração de um produto essa concentração pode não ser bem avaliada, com até pequenos desvios de medida ao redor do valor exato.

Tendo em vista que, geralmente, os modelos de covariância nem sempre possuem efeitos fixos, devido à existência de fatores aleatórios que nem sempre são explicados pelo modelo clássico, faz-se a opção de utilizar a amostragem com repetição da variável auxiliar (covariável) e, desse modo, relacionar os fatores que interferem no processo de medição. Por exemplo, em um experimento para a medição de espessura em fios de cobre com um paquímetro, temperatura, vibrações ou operadores, interpretados como fatores aleatórios, podem introduzir erro nas mensurações. Nesse caso, uma alternativa sugerida pode ser repetir inúmeras vezes a medição, o que permite avaliar a incerteza estatística no resultado final. Neste

contexto, Vuolo (1992) definiu a incerteza como a indicação de quanto o melhor valor pode diferir do valor verdadeiro (mensurado), em termos de probabilidades.

Situações desse tipo se enquadram-se dentro do contexto da engenharia, em que a redução dos “erros”, no sentido mais amplo da palavra, é tratada especificamente pela ciência denominada por metrologia. Em síntese, essa ciência analisa os aspectos relativos às medições e às calibrações de instrumentos, utilizando metodologias estatísticas. Tais medições em conjunto com o uso de uma técnica estatística inadequada, como por exemplo, utilizar um teste que despreza as medidas repetidas, podem resultar em índices de qualidade incoerentes.

A distribuição estatística do erro nem sempre é fácil de ser obtida. Por isso, torna-se complexa a inferência, bem como a formulação de métodos de estimação. Uma alternativa para solucionar esse problema é o uso de técnicas de simulação Monte Carlo e métodos de computação intensiva. Nesse contexto, podem-se citar alguns trabalhos que mostram importantes contribuições proporcionadas por essas técnicas.

Higdon & Schafer (2001) apresentaram métodos computacionais baseados em algoritmos EM (Dempster et al., 1977) e a modificação no método da quadratura Gauss-Hermite (Liu & Pierce, 1994), utilizado na obtenção dos estimadores de máxima verossimilhança de modelos lineares generalizados com erro de medida nas covariáveis. Com o propósito de comparar o desempenho desses métodos, os autores concluíram que a inacurácia observada na quadratura gaussiana, ocasionam problemas de singularidade nesse procedimento numérico.

Em relação ao coeficiente de determinação de modelos com erros de medida, Delicato & Romo (2004) propuseram testes estatísticos paramétricos apresentando suas distribuições assintóticas e intervalos de confiança para os parâmetros utilizando-se a técnica *bootstrap*. Os testes foram validados baseando-se em

resultados referentes ao controle do erro tipo I e poder obtidos por meio de simulação Monte Carlo. Na abordagem *bootstrap*, Delicato & Romo (1999) obtiveram a distribuição empírica sob a hipótese nula de testes baseados nos coeficientes de determinação. As conclusões encontradas pelos autores é que os testes considerados no estudo obtiveram comportamento global similar (considerando juntamente controle do erro tipo I e poder do teste).

Ao se tratar dos métodos de estimação dos parâmetros de um modelo com erro de medida, uma particular atenção é dada para os estimadores baseados em métodos dos momentos, desenvolvidos por James & Stein (1961) e Fuller (1987), sendo estes utilizados para estimar o verdadeiro valor da variável independente. Respectivamente, os métodos citados são conhecidos, na literatura, como estimadores "Plug-in" e de atenuação de vício.

Convém salientar que a construção de intervalos de confiança, para os parâmetros considerando esses métodos de estimação, se depara com o problema de obter suas variâncias assintóticas, uma vez que o processo de estimação é realizado em duas etapas, conforme é relatado por Cunha & Colosimo (2003). Com esses problemas, os autores propuseram intervalos de confiança pela abordagem *bootstrap*, de tal forma que a aplicação desta técnica foi eficiente, destacando-se o intervalo obtido pelo método percentil.

Visto que os erros nas variáveis poderão estar presentes em um experimento, por mais cuidadoso que seja o pesquisador, verifica-se a necessidade de realizar aferições do erro na medida de suas variáveis. Nestas aferições, encontra-se o estudo da acurácia e precisão (Baker & Rose, 1984).

## 2.1 Análise de covariância

Análise de covariância é um procedimento estatístico que permite incorporar uma informação sobre covariáveis dentro da análise da variável resposta. Algumas vezes empregada para reduzir o erro experimental. Outras vezes para compreender melhor o fenômeno que está sendo estudado (Milliken & Johnson, 2002).

Uma outra definição válida para a análise de covariância, pode ser descrita como uma técnica estatística para análise de dados originados de pesquisas, na qual uma variável dependente ( $y$ ) é relacionada com uma variável independente ( $x$ ) para proceder um ajustamento das médias dos tratamentos (Neter et al., 1990). Na terminologia da análise de covariância, cada variável quantitativa independente adicionada no estudo é chamada de variável concomitante, covariável ou variável auxiliar (Neter et al., 1990). Ela examina as variâncias e as covariâncias entre as variáveis selecionadas  $x$  e  $y$ , combinando os conceitos e as técnicas da análise da variância e da regressão, de tal forma que os efeitos dos tratamentos sobre a variável resposta de interesse são melhor caracterizados do que quando se utiliza qualquer uma das técnicas separadamente. A ANCOVA tem sido largamente utilizada em diversas áreas de pesquisa. Nas áreas agrárias, tem ampla aplicação no melhoramento genético de plantas cultivadas.

Segundo Cochran (1957), as primeiras ilustrações da análise de covariância foram feitas por Fisher (1932), em trabalho que considerou uma variável auxiliar ( $x$ ) constituída pela produção de folhas de chá por parcela, obtida em período preliminar à aplicação dos tratamentos e a variável dependente ( $y$ ) produção de folhas de chá por parcela, no final do ensaio. Kempthorne (1952) apresentou os detalhes da metodologia de covariância considerando um modelo linear com dois critérios de classificação, sem interação e com uma observação por parcela. Um estudo

teórico sobre a análise de covariância foi apresentado por Ostle (1954), para um experimento fatorial com dois fatores, no delineamento em blocos casualizados. O mesmo apresentou as expressões dos estimadores para cálculo das soma de quadrados e produtos, teste F para os fatores principais, interação e variâncias das médias ajustadas.

As suposições básicas essenciais à utilização da análise de covariância, de acordo com Neter et al. (1990), são: o modelo linear tem os efeitos aditivos, os erros da variável dependente  $y$  são independentes e normalmente distribuídos com média zero e variância constante, os valores da variável auxiliar (covariável) devem ser fixos e suas medidas tomadas sem erro, a regressão de  $y$  sobre  $x$ , após a retirada dos efeitos de blocos e tratamentos, é linear e independente destes efeitos.

As mais importantes finalidades da análise de covariância, segundo Cochran (1957) são, as seguintes:

- ajudar na interpretação dos resultados experimentais e contribuir para esclarecer os efeitos de tratamentos;
- controlar o erro experimental e aumentar a precisão das estimativas;
- ajustar as médias dos tratamentos da variável dependente pelas diferenças dos correspondentes valores das variáveis independentes;
- remover os efeitos de variáveis perturbadoras;
- estimar observações perdidas.

Com relação ao controle do erro experimental, Cochran (1957) salienta que o ganho em precisão é proporcional ao valor do coeficiente de correlação entre a variável dependente e a covariável nas unidades experimentais que recebem o mesmo tratamento.

Segundo Ramalho et al. (2000), o erro experimental depende de uma série de fatores, como a heterogeneidade na fertilidade do solo, a distribuição de água e os danos mecânicos sobre as plantas, entre outros. Parte do erro experimental pode ser reduzida por meio do controle local. No entanto, mesmo utilizando os delineamentos apropriados, que possibilitam um bom controle local, não é possível eliminar certos tipos de variabilidade entre parcelas, tais como diferenças no número de plantas nas parcelas, pequenas manchas de fertilidade no solo e diferenças na ocorrência de doenças e insetos. Nestas situações, esses tipos de heterogeneidade contribuem para aumentar a variância do erro e a precisão do experimento é reduzida.

Em casos como esse, a análise de covariância pode ser usada com a finalidade de reduzir o erro experimental, por meio do ajuste das médias em função das pequenas diferenças que ocorrem em alguns fatores dentro da parcela.

A análise de covariância tem sido utilizada em várias ocasiões, especialmente visando ao ajustamento nas diferenças de estande na cultura do milho. O uso da covariância com essa finalidade é apresentado, com detalhes, em algumas publicações (Steel & Torrie, 1960; Gomez & Gomez, 1984; Vencovsky & Cruz, 1991).

Para se obter um menor erro experimental possível, o pesquisador deve ficar atento às variações aleatórias não controladas que afetem o tratamento e contribuem para aumentar o erro, usualmente na literatura a análise de covariância é empregada sobre a suposição de que a covariável é medida sem erro. Porém, quando esta suposição não é válida, ainda assim, existe a alternativa de proceder o ajustamento das médias dos tratamentos por meio de uma análise de covariância, em que os valores da variável auxiliar estão sujeitos a erros de medida (Degraacie & Fuller, 1972).

### 2.1.1 Modelo estatístico

O modelo estatístico para a análise de covariância para um experimento no delineamento inteiramente casualizado com um fator e efeito fixo, com uma variável independente, é:

$$y_{ij} = \mu + \tau_i + \gamma x_{ij} + e_{ij} \quad (2.1)$$

em que,

$y_{ij}$  valor da observação ocorrida na j-ésima repetição do i-ésimo tratamento;

$\mu$  é uma constante presente em todas observações;

$\tau_i$  é o efeito fixo do i-ésimo tratamento sujeito à restrição paramétrica  $\sum \tau_i = 0$ ;

$\gamma$  é o coeficiente de regressão para a relação entre y e x;

$x_{ij}$  é o valor da covariável obtida na j-ésima repetição do i-ésimo tratamento;

$e_{ij}$  é o erro experimental associado à observação  $y_{ij}$ , considerados independentes e normalmente distribuídos, com média zero e variância constante.

Outro modo de se escrever o modelo (2.1) é quando se faz a correção da variável auxiliar para a média geral na covariável. Assim, o modelo estatístico é;

$$y_{ij} = \mu + \tau_i + \gamma(x_{ij} - \bar{x}_{..}) + e_{ij} \quad (2.2)$$

sendo:

$y_{ij}$  valor da observação ocorrida na j-ésima repetição do i-ésimo tratamento;

$\mu$  é uma constante presente em todas observações;

$\tau_i$  é o efeito fixo do i-ésimo tratamento sujeito à restrição paramétrica  $\sum \tau_i = 0$ ;

$\gamma$  é o coeficiente de regressão para a relação entre y e x;

$x_{ij}$  é o valor da covariável obtida na j-ésima repetição do i-ésimo tratamento;

$\bar{x}_{..}$  é a média geral da variável auxiliar;

$e_{ij}$  é o erro experimental associado à observação  $y_{ij}$ , considerados independentes e normalmente distribuídos, com média zero e variância constante.

### 2.1.2 Equações normais

As equações normais obtidas pelo método dos quadrados mínimos, de acordo com Searle (1971), entre outros, são:

$$y_{..} = n\hat{\mu} + \sum_{i=1}^I J\hat{\tau}_i \quad (2.3)$$

$$y_{i.} = J(\hat{\mu} + \hat{\tau}_i) + \gamma \sum_{j=1}^J (x_{ij} - \bar{x}_{..}), \quad \text{com } i = 1, \dots, I \quad (2.4)$$

$$\sum_{i=1}^I \sum_{j=1}^J y_{ij}(x_{ij} - \bar{x}_{..}) = \sum_{i=1}^I \sum_{j=1}^J (\hat{\mu} + \hat{\tau}_i)(x_{ij} - \bar{x}_{..}) + \sum_{i=1}^I \sum_{j=1}^J \gamma(x_{ij} - \bar{x}_{..})^2 \quad (2.5)$$

Segundo os mesmos autores, nas equações (2.3), (2.4) e (2.5) existem  $I + 2$  equações e  $I + 2$  parâmetros desconhecidos a serem estimados. Conforme mencionado por Dean & Voss (1999), as equações normais não são linearmente independentes e, assim, as equações não têm uma única solução. Entretanto, há  $I + 1$  equações linearmente independentes, de modo que o sistema pode ser resolvido obtendo-se uma única estimativa de mínimos quadrados para  $\gamma, \mu + \tau_1, \dots, \mu + \tau_I$ .

Os estimadores de quadrados mínimos para os parâmetros do modelo são obtidos resolvendo-se as equações normais:

$$\hat{\gamma} = \frac{\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.})}{\sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.})^2} \quad (2.6)$$

$$\hat{\mu} + \hat{\tau}_i = \bar{y}_{i.} - \hat{\gamma}(\bar{x}_{i.} - \bar{x}_{..}), \quad \text{para } i = 1, \dots, I \quad (2.7)$$

Dean & Voss (1999) afirmam que o parâmetro regressivo é não viciado, de tal forma que  $E[\hat{\gamma}] = \gamma$ , assim como para o efeito de tratamentos tem-se  $E[\hat{\mu} + \hat{\tau}_i] = \mu + \tau_i$

Os estimadores de mínimos quadrados de  $\hat{\mu} + \hat{\tau}_i$ , por esta razão, estimam a resposta média para o  $i$ -ésimo tratamento, quando o valor da covariável é igual a  $\bar{x}_{..}$ .

As estimativas de  $\hat{\mu} + \hat{\tau}_i$  podem ser utilizadas como as médias ajustadas (estimadas), desde que elas sejam ajustadas a resposta média  $\bar{y}_{i.}$  pela quantidade  $\hat{\gamma}(\bar{x}_{i.} - \bar{x}_{..})$ , o que é equivalente a mensurar a respostas sobre os mesmos valores da covariável (Dean & Voss, 1999).

De acordo com Neter et al. (1990), a “chave” da inferência estatística de interesse na análise de covariância é a mesma para os modelos de análise de variância, ou seja, é estimar os efeitos dos tratamentos e contrastar a hipótese de que todos os níveis do fator produzem o mesmo efeito, frente a alternativa de que ao menos 2 tratamentos diferem significamente entre si (Leal & Porras, 1998). Teste para efeitos fixos de tratamentos envolvem as mesmas alternativas dos modelos de análises de variância.

$$H_o : \tau_1 = \tau_2 = \dots = \tau_I$$

$$H_1 : \text{pelo menos um } \tau_i \text{ é diferente de zero}$$

A análise de covariância é usada para testar a hipótese nula  $H_0$  de encontro com a hipótese alternativa  $H_1$ . O teste é baseado sobre as comparações das somas de quadrados do erro do modelo completo e reduzido. As fórmulas para as somas de quadrados e produtos que ajustam as somas de quadrados da análise de covariância, adaptadas de Dean & Voss (1999), são:

$$SPx = \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{..})(y_{ij} - \bar{y}_{..}) \quad Sxx = \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{..})^2$$

$$Syy = Tyy + Eyy = \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{..})^2$$

à soma de quadrados para o modelo reduzido sob  $H_0$  é:

$$Syy_{aj} = \sum_{i=1}^I \sum_{j=1}^J \left[ y_{ij} - \bar{y}_{..} - \frac{SPx(x_{ij} - \bar{x}_{..})}{Sxx} \right]^2$$

$$Syy_{aj} = Syy - \frac{(SPx)^2}{Sxx}$$

Sendo o número de graus de liberdade para o erro igual ao número de observações menos um grau de liberdade para a constante  $\mu + \tau_i$  e um para o parâmetro de inclinação  $\gamma$ , ou seja,  $n - 2$ . Sobre o modelo de análise de covariância completo usando as estimativas de mínimos quadrados obtidas em (2.6 e 2.7), a soma de quadrados para o erro ajustado pela covariável é:

$$Eyy_{aj} = \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \hat{\mu} - \hat{\tau}_i - \hat{\gamma}(x_{ij} - \bar{x}_{..}))^2$$

$$\begin{aligned}
&= \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{i.} + \hat{\gamma}(\bar{x}_{i.} - \bar{x}_{..}) - \hat{\gamma}(\bar{x}_{i.} - \bar{x}_{..}))^2 \\
&= \sum_{i=1}^I \sum_{j=1}^J ((y_{ij} - \bar{y}_{i.}) - \hat{\gamma}(x_{ij} - \bar{x}_{i.}))^2 \\
&= Eyy - \hat{\gamma}(Exy) \\
&= Eyy - \frac{(Exy)^2}{Exx}
\end{aligned}$$

Sendo,

$$\begin{aligned}
Exy &= \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.})(y_{ij} - \bar{y}_{i.}), \\
Exx &= \sum_{i=1}^I \sum_{j=1}^J (x_{ij} - \bar{x}_{i.})^2, \\
Eyy &= \sum_{i=1}^I \sum_{j=1}^J (y_{ij} - \bar{y}_{i.})^2.
\end{aligned}$$

O número de graus de liberdade é  $n - I - 1$  (um a menos que os graus de liberdade para a análise de variância, devido ao parâmetro regressivo  $\gamma$ ).

A soma de quadrados para tratamentos  $S(T|\gamma)$  ou  $Tyy_{aj}$  é a diferença da soma de quadrados do erro sobre o modelo reduzido e o completo (Dean & Voss, 1999).

$$S(T|\gamma) = Tyy_{aj} = Syy_{aj} - Eyy_{aj}$$

A diferença nos graus de liberdade para o modelo reduzido e o completo é:

$$(n - 2) - (n - I - 1) = I - 1$$

Para a análise de covariância baseada no modelo linear,

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

e, com a restrição  $\sum_{i=1}^I \tau_i = 0$ , Reich & Webster (1969) observaram que  $\frac{Tyy_{aj}}{\sigma^2}$

tem distribuição de qui-quadrado não central com  $I-1$  número de graus de liberdade, com parâmetro de não centralidade ( $\lambda$ ) dado por

$$\lambda = \frac{J}{2\sigma^2} \left\{ \sum_{i=1}^I \tau_i^2 - J \left[ \sum_{i=1}^I (\bar{x}_{i.} - \bar{x}_{..}) \tau_i \right]^2 / (Txx + Exx) \right\}$$

sendo  $Tyy_{aj}$  a soma de quadrados de tratamentos, ajustada para a regressão,  $Txx$  a soma de quadrados de tratamentos relativos a covariável,  $Exx$  a soma de quadrado dos resíduos relativos à variável auxiliar e  $J$  o número de repetições.

Do mesmo modo, os autores verificaram que  $\frac{Eyy_{aj}}{\sigma^2}$  tem distribuição qui-quadrado central, com  $ne = I(J-1) - 1$  número de graus de liberdade, sendo  $Eyy_{aj}$  a soma de quadrados dos resíduos, ajustada pela regressão. E, ainda, que a expressão  $[Tyy_{aj}/(t-1)]/[Eyy_{aj}/ne]$  tem distribuição de F não central, com  $(I-1)$  e  $ne$  número de graus de liberdade e parâmetro de não centralidade  $\lambda$ .

A Tabela 2.1 fornece um exemplo para análise de covariância adaptada de Dean & Voss (1999).

TABELA 2.1: Exemplo de tabela para a análise de covariância.

FV	GL	Soma de quadrados	Quadrado médio	F
$\tau \gamma$	$I - 1$	$S(\tau \gamma)$	$\frac{S(\tau \gamma)}{I-1}$	$\frac{S(\tau \gamma)/(I-1)}{Eyy_{aj}/(n-I-1)}$
$\gamma \tau$	1	$S(\gamma \tau)$	$\frac{S(\gamma \tau)/1}{1}$	$\frac{S(\gamma \tau)/1}{Eyy_{aj}/(n-I-1)}$
$Erro_{aj}$	$n - I - 1$	$Eyy_{aj}$	$\frac{Eyy_{aj}}{n-I-1}$	--
$Trat + erro$	$n - 1$	$Syy_{aj}$	--	--

As informações para testar a igualdade dos efeitos de tratamentos estão tipicamente sumariadas na Tabela 2.1, pelas informações desta Tabela, pode-se inferir sobre os efeitos de tratamentos [ $(H_o : \tau_1 = \tau_2 = \dots = \tau_I)$  versus  $(H_1 : \text{Pelo menos um dos efeitos de tratamento são diferentes entre si})$ ], rejeitando  $H_o$  se

$$\frac{S(\tau|\gamma)/(I-1)}{Eyy_{aj}/(n-I-1)} > F_{(I-1, n-I-1, \alpha)}$$

### 2.1.3 Abordagem matricial

Segundo argumento de Searle (1971) a análise de covariância pode ser feita utilizando-se o seguinte modelo linear,

$$y = X\beta + e \quad (2.8)$$

no qual  $y$  é um vetor de variáveis aleatórias;  $X$  é a matriz do delineamento composta dos valores 0 e 1 associados à ausência ou à presença de determinado tratamento e dos valores observados das covariáveis;  $e$  é o vetor dos erros experimentais e  $\beta$  é um vetor de parâmetros.

Entretanto, para melhor distinção entre os dois tipos de parâmetros, Searle (1971) particionou a matriz  $X$  e o vetor  $\beta$  em duas partes, obtendo, desse modo, o modelo:

$$y = Xa + Zb + e \quad (2.9)$$

em que  $X$  é composta pelos valores da variável binária ou matriz de delineamento, também chamada matriz de incidência de tratamentos;  $Z$  é a matriz dos valores da variável auxiliar (covariável), de posto coluna completos e cujas colunas são independentes das colunas da matriz  $X$ ; o vetor  $a$  contém os parâmetros associados à matriz de incidência dos tratamentos,  $b$  é o vetor dos coeficientes de regressão e  $e$  é o vetor dos erros experimentais, quantidades independentes, com distribuição normal e com variância constante.

Os estimadores para os efeitos de tratamentos conforme Searle (1971), obtidos pelo método dos quadrados mínimos, é:

$$\hat{a} = a^* - (X'X)^{-}X'Z\hat{b} \quad (2.10)$$

e, para os efeitos do coeficiente de regressão é

$$\hat{b} = (Z'PZ)^{-1}Z'Py \quad (2.11)$$

sendo  $a^* = (X'X)^{-}X'y$ ,  $P = I - X(X'X)^{-}X'$  e  $(X'X)^{-}$  é a inversa generalizada de  $(X'X)$ .

Nota-se que as estimativas obtidas em  $\hat{b}$  são influenciadas pelas observações e pelos valores da variável auxiliar e que  $Z'PZ$  é uma matriz cujos elementos são as somas de produtos das variáveis auxiliares com a variável dependente, relativos ao resíduo.

As somas de quadrados e de produtos para o cálculo da ANCOVA na abordagem matricial, de acordo com Kshirshgar (1983), são ilustradas a seguir, de forma que as somas para os erros sejam definidas e calculadas da seguinte maneira,  $Eyy = y'(P)y$  é a soma de quadrados do erro da variável dependente, em que,

$Eyz = y'(P)Z$  é a soma de produtos do erro,  $Ezz = Z'(P)Z$  é a soma de quadrados do erro da covariável,  $Eyy_{aj} = Eyy - Eyz'(Ezz)^{-1}Eyz$  é a soma de quadrados do erro ajustado pelo valor da covariável.

As somas de quadrados e produtos para os tratamentos seguem da mesma forma em que  $Hyy = y'(P - P_1)y$ ,  $Hzz = Z'(P - P_1)Z$ ,  $Hyz = y'(P - P_1)Z$ , sendo as soma de quadrados de tratamentos para  $y$ , a soma de quadrados de tratamentos para  $Z$  e a soma de produtos de tratamentos, respectivamente. Desse modo, a soma de quadrados para o total ajustado é  $HEyy_{aj} = (Hyy + Eyy) - (Hyz + Eyz)'(Hzz + Ezz)^{-1}(Hyz + Eyz)$  e a soma de quadrados para os

tratamentos ajustados pela covariável  $Hy_{aj} = HEyy_{aj} - Eyy_{aj}$  sendo  $P = I - X(X'X)^{-1}X'$  o projetor ortogonal,  $I$  a matriz identidade de mesma dimensão que a matriz  $X$ ,  $X_1$  o vetor da primeira coluna da matriz  $X$  e  $P_1$  o projetor da matriz  $X_1$ .

## 2.2 Erro nas variáveis

O pesquisador pode ter motivos para acreditar que o valor observado da covariável não corresponde ao valor exato, ou seja, contém algum erro, o qual pode ser devido à calibração do aparelho escolhido para fazer as aferições (Souza, 2004). Em situações como esta, o erro de medida pode ser incorporado em um modelo linear, como no caso dos modelos (2.2 e 2.9), por meio da relação:

$$z_{ij} = x_{ij} + u_{ij} \quad (2.12)$$

sendo  $x_{ij}$  o valor ideal,  $z_{ij}$  o valor observado e  $u_{ij}$  uma variável aleatória que representa o erro de medida não observável.

A análise de covariância com erro de medida com aplicação na área agrônômica foi abordada, entre outros, por Degraçie & Fuller (1972), Ambos os autores propuseram estimadores para o coeficiente de regressão  $b$ , considerando os modelos com erros nas variáveis, os quais têm viés em menor ordem e menor erro quadrático médio que o estimador usual de mínimos quadrados. Um exemplo dado pelos autores foi extraído de Voss (1969) e trata de um experimento realizado com o objetivo de avaliar dois tipos de solo com relação à produção de milho, em um delineamento inteiramente casualizado, com 11 repetições. Julgou-se necessário coletar os teores de nitrogênio nas parcelas, no intuito de ajustar, por meio da co-

variável (teor de nitrogênio), as médias dos tratamentos. Porém, a determinação da quantidade de nitrogênio envolvendo o tipo de solo e a determinação química usada na análise de laboratório está sujeita a erro de mensuração.

Uma forma de estimar a variância do erro de medida, para, desse modo, ajustar as médias para o valor que deveria ser obtido, é coletando-se várias vezes os pares de amostra (repetindo a coleta de nitrogênio 150 vezes). Por meio disso é possível estimar o erro de medida. Esta análise propicia uma melhora na precisão de experimentos que são afetados por erros de mensuração.

Para Carrol et al. (1985) existem dois subtipos da análise de covariância com erro de medida (os modelos funcional e estrutural) e vários autores abordam esta questão, dentre eles Lord (1960), Cochran (1968) e Degraçie & Fuller (1972). Lord (1960) discutiu o caso em que a variável auxiliar pode ser mensurada com erro. O autor afirma que a ANCOVA usual (mínimos quadrados) tem grande probabilidade de não detectar diferenças estatísticas entre grupos, quando tais diferenças existem e podem ser detectadas por alguma ferramenta apropriada, como, por exemplo, estimadores que levem em conta o vício produzido pelo erro de medida na covariável. Cochran (1968) e Degraçie & Fuller (1972) discutiram sobre dois tipos de análise de covariância, fornecendo uma particular discussão para o caso em que os verdadeiros valores das covariáveis são variáveis aleatórias. Neste caso, usualmente, chama-se, na literatura, de modelo estrutural. Ambos os autores mostraram que, em casos de experimentos não balanceados, que ocorrem em experimentos observacionais com alguma frequência, os erros nas variáveis podem ocasionar estimativas de mínimos quadrados inconsistentes para a verdadeira diferença entre tratamentos. Assintoticamente, quando as covariáveis não são balanceadas, as estimativas de mínimos quadrados para a diferença entre os tratamentos podem ser realizadas, mas, para pequenas amostras não é recomen-

dada.

No caso em que as covariáveis são balanceadas, Cochran (1968) e Degraçie & Fuller (1972) indicaram que a estimativa de mínimos quadrados para a diferença entre os efeitos de tratamentos serão consistentes. Carrol et al. (1985) vão mais além e discutem que em casos balanceados, em estudos completamente aleatorizados com mensuração de erros nas covariáveis, a estimativa de mínimos quadrados para a diferença entre tratamentos pode ser preferida quando comparada a um particular estimador do parâmetro de regressão com erros nas variáveis, sendo que em seu estudo Carrol et al. (1985) não consideraram os estimadores de James & Stein (1961) e Fuller (1987).

Chan et al. (2004) discorrem, em um delineamento inteiramente casualizado na ANCOVA, sobre a divergência encontrada na literatura sobre os erros nas covariáveis. Simulações computacionais são utilizadas para comparar o viés das estimativas dos efeitos dos tratamentos, com e sem o ajuste da mensuração do erro, em diferentes tamanhos amostrais e coeficientes de regressão ( $b$ ). A conclusão do estudo para Chan et al. (2004) é a de que o estimador de mínimos quadrados é não viesado e os estimadores propostos para mensuração do erro em pequenas amostras são viesados. Convém salientar que os estimadores estudados por Chan et al. (2004) não são os mesmos avaliados neste trabalho, além do fato de que, diferentes coeficientes de determinação e distribuições dos resíduos não foram avaliados por esses autores. O estudo em questão torna-se relevante pois admite uma quantidade maior de situações, além de estudar o comportamento dos estimadores “Plug-in” e Atenuador de vício que não foram relatados na literatura para modelos de ANCOVA, assim como, avaliados quanto a acurácia e a precisão.

Souza (2004) apresenta estimadores de máxima verossimilhança para os parâmetros de interesse e testes para a hipótese de inexistência de efeito de trata-

mento nos modelos normais funcional e estrutural. O mesmo assume que a razão entre as variâncias dos erros é conhecida e, assim, contorna o problema de não identificabilidade do modelo estrutural ou de não existência de estimadores de máxima verossimilhança no modelo funcional. Desenvolve também diversos estudos de simulação para avaliar algumas propriedades dos estimadores propostos e o desempenho dos testes de hipóteses apresentados. Concluindo que para os modelos funcional e estrutural normais, o vício e o erro quadrático médio dos estimadores de máxima verossimilhança dos efeitos dos tratamentos são pequenos. Quando a suposição de normalidade dos erros do modelo é violada, os estimadores de máxima verossimilhança dos efeitos de tratamentos são os menos afetados quando comparados aos estimadores dos demais parâmetros de regressão e por fim observa que a suposição de normalidade para os resíduos do modelo de ANCOVA com erros nas covariáveis é importante para que os estimadores dos parâmetros de regressão tenham boas propriedades.

Outra metodologia que pode ser utilizada, e será abordada neste trabalho, é a de estimadores de parâmetros de um modelo com erro de medida na covariável, dos quais destacam-se os estimadores propostos por Fuller (1987) e James & Stein (1961), que têm como a finalidade de diminuir os riscos de incerteza, minimizando assim, também o efeito do erro de medida.

### **2.2.1 Estimador de James-Stein**

O estimador proposto por James & Stein (1961), por ser utilizado para a determinação ou ajuste dos valores  $z_{ij}$  ( $i = 1, \dots, I; j = 1, \dots, J$ ) da covariável observada com o erro de medida, obtida nas  $J$  repetições. Essas, repetições são independentes e identicamente distribuídas por uma distribuição normal com média  $x_i$

e variância  $\sigma_u^2$ .

Estimadores do tipo shrinkage são sempre necessários, independentemente se os efeitos forem tomados como fixos ou aleatórios. O estimador melhorado de James & Stein (1961) não necessita mesmo de qualquer suposição a esse respeito, ou sobre distribuições das médias a serem estimadas. Requer apenas o relaxamento da suposição de não vício. Tal estimador é viciado, mas tem menor erro quadrático médio que o estimador de quadrados mínimos na presença de erros na covariável (Resende & Duarte, 2007).

O estimador proposto pelos autores, parte do pressuposto em ajustar um novo valor para a covariável  $z_{ij}$ . O princípio consiste que, de uma amostra observada com erro de medida, o verdadeiro valor da variável pode ser obtido. As estimativa do parâmetro de regressão podem ser realizada usualmente como se este valor estimado fosse o valor verdadeiro não observado  $x$ . O estimador que substituirá a covariável observada  $z$  é:

$$\theta_{ij}(z) = \hat{B}\bar{z}_{..} + (1 - \hat{B})z_{ij}, \quad i = 1, \dots, I, \quad (2.13)$$

em que,  $z_{ij}$  corresponde às observações com erro de medida,  $\theta_{ij}(z)$  é o novo valor que substituí  $z_{ij}$  e  $\hat{B}$  é dado pela seguinte expressão,

$$\hat{B} = \frac{S_u^2(n-3)}{\hat{S}} \quad (2.14)$$

em que,  $S_u^2$  é o estimador da variância do erro de medida,

$$S_u^2 = \frac{\sum_{j=1}^J (z_{ij} - \bar{z}_{i.})^2}{n(J-1)} \quad (2.15)$$

$$\hat{S} = \sum_i^I (z_i - \bar{z}_{..})^2, \quad \text{em que,} \quad \bar{z} = \frac{\sum_{i=1}^I \sum_{j=1}^J z_{ij}}{IJ}$$

sendo  $\bar{z}_{..}$  a média geral dos valores da covariável  $z_{ij}$ ,  $\hat{S}$  o estimador da distância do valor observado para a média das repetições e  $\bar{z}_{i.}$  a média dos tratamentos das observações  $z_{ij}$ .

$$\bar{z}_{i.} = \frac{\sum_{j=1}^J z_{ij}}{J} \quad (2.16)$$

Este estimador foi usado por Whittermore (1989) para examinar a performance em modelos de regressão não lineares por meio de simulações e para comparar este com o estimador usual de mínimos quadrados que ignora os erros nas variáveis, como também com outros estimadores propostos na literatura. A autora concluí em seu estudo que o estimador de James-Stein obteve melhor desempenho em relação aos demais estimadores inclusive o estimador de Mínimos Quadrados. Porém em situações de não linearidade do modelo de regressão.

### 2.2.2 Estimador de Fuller

O estimador proposto por Fuller (1987) baseou-se na determinação do viés (ou tendência erro/desvio) produzido pelo erro na observação dos dados. O mesmo é conhecido na literatura como estimador atenuador de vício. Para a obtenção do estimador, é necessário determinar, inicialmente, o fator:

$$k_x^{-1} = \frac{\sigma_x^2 + \sigma_u^2}{\sigma_x^2}$$

que pode ser estimado utilizando-se a expressão,

$$k_x^{-1} = \frac{S_z^2}{S_z^2 - S_u^2}, \quad (2.17)$$

em que,  $S_z^2$  corresponde a variância amostral da covariável ajustada com o erro de

medida e  $S_u^2$  é a variância do erro de medida, obtida conforme a expressão (2.15).

Com base nesse fator, o estimador de menor vício é obtido:

$$\hat{\beta}_F = \hat{\gamma} * k_x^{-1}, \quad (2.18)$$

em que,  $\hat{\gamma}$  é o estimador de mínimos quadrados para o parâmetro regressivo do modelo (2.2 e 2.9).

Cunha & Colosimo (2003) avaliaram os intervalos de confiança para os estimador de Fuller (1987) e o de James & Stein (1961) por meio de reamostragens *bootstrap* (Efron & Tibshirani, 1993), os autores concluíram que o estimador de Fuller (1987) obteve desempenho inferior ao estimador de James-Stein, principalmente quando não distribuído pela normal, situação em que foi avaliado por uma distribuição exponencial, o mesmo apresentou intervalos de confiança de grande extensão.

### 2.2.3 Acurácia e precisão

A acurácia é definida como a fidelidade dos dados, um produto da medição de quanto o dado difere daquele tomado, então, como referência. Suas avaliações são feitas por meio de funções estatísticas que estimam o grau de concordância ou discordância dos atributos existentes em relação àqueles tidos como verdadeiros, ou seja, uma medida é dita ser acurada quando ela se aproxima de seu valor verdadeiro.

Segundo Amorin (2004), o termo precisão é o grau de refinamento com o qual uma grandeza é medida. Em outras palavras, significa quanto os valores de uma série de medidas estão próximos uns dos outros. Se uma grandeza é medida várias vezes e os resultados dessa operação forem valores muito próximos uns dos

outros, diz-se que a precisão é alta. Normalmente, a precisão é expressa em termos do desvio padrão ou da variância das medições.

A melhor forma de exemplificar as diferenças entre os conceitos dados é o uso de um alvo de tiro, conforme ilustrado na Figura 2.1.

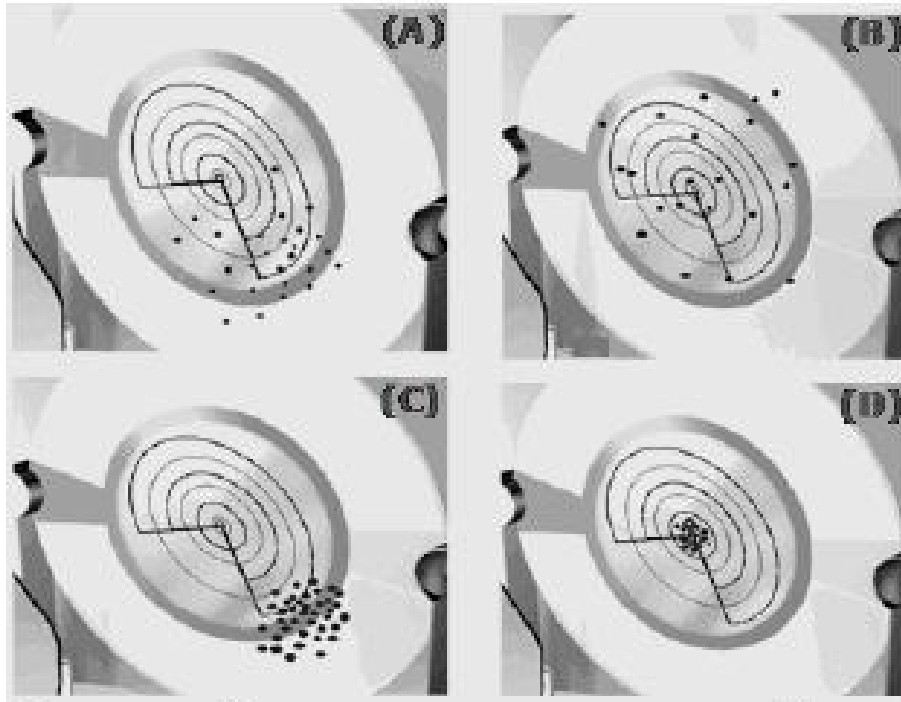


FIGURA 2.1: Ilustração com os conceitos de precisão e acurácia (Fonte: Araújo Neto(2006))

- a) observações com pouca acurácia e imprecisas;
- b) observações com média exata (boa acurácia), porém imprecisas;
- c) observações precisas, porém com pouca acurácia, e
- d) observações precisas e com alta acurácia.

A avaliação dos métodos de estimação "Plug-in" e atenuação de vício é realizada para medir a acurácia e a precisão do modelo (2.9) Bibby & Toutenburg (1977) obtiveram o viés médio, representado por  $V_m$ , no qual, interpretaram a inacurácia média das predições do modelo (2.9).

$$V_m = \frac{\sum_{l=1}^N (\hat{y}_l - y_l)}{N} \quad (2.19)$$

sendo  $N$  o número de simulações Monte Carlo,  $y_l$  é o valor obtido pelo modelo de ANCOVA simulado,  $\hat{y}_l$  é o valor estimado ("Plug-in" ou atenuador de vício) para o modelo de ANCOVA com erro nas variáveis.

A precisão das estimativas é representada pela expressão,

$$Rq_{mep} = \sqrt{\frac{\sum_{l=1}^N (\hat{y}_l - y_l)^2}{N}} \quad (2.20)$$

Em algumas situações, o viés médio ( $V_m$ ) poderá ser relativamente alto, devido à falta de acurácia. Nesses casos, ocorrerá uma super estimativa do  $Rq_{mep}$ , que quantifica o quanto os dados estimados se adequam bem aos valores observados. Em virtude disso, usa-se o erro de predição para avaliar a precisão, sendo este corrigido pela falta de acurácia, conforme a seguir:

$$E_p = \sqrt{Rq_{mep}^2 - (V_m)^2} \quad (2.21)$$

### 2.3 Coeficiente de determinação do modelo

O coeficiente de determinação  $R^2$  pode ser interpretado como a proporção da variabilidade presente nas observações da variável resposta  $y$ , que é explicada indicando o quanto da variação existente foi explicado pelo modelo adotado (2.9) pela variável regressora  $x$  no modelo ajustado aos dados. Uma maneira de condicionar o coeficiente de determinação do modelo é, conforme demonstrado por Veloso (2006):

$$Eyy = \frac{[(1 - R^2) \times S(\gamma|\tau_i)]}{R^2} \quad (2.22)$$

### 2.4 Simulação Monte Carlo

A técnica de resolução de problemas analíticos usando números aleatórios é chamada de Método Monte Carlo (MMC), que explora as propriedades estatísticas dos números aleatórios para assegurar que o resultado correto seja computado, da mesma maneira acontece em um jogo de casino em que a “casa” de apostas se certifica de que sempre tenha lucro, Esta é a razão para o seu nome. O MMC é um método computacional utilizado em simulações estocásticas com diversas aplicações em áreas como a física, a matemática e a biologia e tem sido utilizado como forma de obter aproximações numéricas de função de difícil tratamento analítico. Ele, tipicamente, envolve a geração de observações de alguma distribuição de probabilidades e o uso da amostra obtida para aproximar a função de interesse.

Em estudos dos muitos sistemas que compõem o mundo moderno, usa-se, cada vez mais, a simulação em computador, segundo Bussab & Morettin (2004),

freqüentemente é necessário trabalhar com situações que envolvam aleatoriedade. Nestes casos, mesmo construindo modelos probabilísticos, algumas questões não podem ser resolvidas analiticamente e torna-se necessário recorrer a estudos de simulação para a obtenção de aproximações das quantidades de interesse. Para resolver um problema por meio do Método Monte Carlo, utiliza-se uma série de tentativas aleatórias. A precisão do resultado final depende, em geral, do número de tentativas. Esse equilíbrio entre a precisão do resultado e o tempo de computação é uma característica extremamente útil dos Métodos Monte Carlo. Se o interesse reside somente em uma solução aproximada, então, um Método Monte Carlo pode ser bastante rápido.

### 3 METODOLOGIA

A metodologia proposta neste trabalho visa avaliar a acurácia e a precisão dos estimadores “Plug-in”(James & Stein, 1961) e atenuador de vício (Fuller, 1987), para o parâmetro do coeficiente de regressão, por meio de simulação Monte Carlo, utilizando o modelo (2.9), que trata dos efeitos das verdadeiras diferenças entre tratamentos ajustados pela variável auxiliar. O estudo em questão tem por objetivo obter estimativas dos valores ajustados pela ANCOVA com maior acurácia e precisão, quando as covariáveis estão sujeitas a erro de medida. Para tanto, nas simulações foram consideradas diferentes situações. Entre elas: tamanhos da amostra, coeficiente de determinação e distribuições dos erros aplicados aos diferentes estimadores avaliados.

Supondo que exista uma relação linear entre a variável independente (covariável) e a variável dependente, o modelo de análise de covariância considerado neste trabalho é semelhante ao mencionado por Searle (1971) e foi definido pela seguinte expressão:

$$y = X\beta + Z\gamma + \epsilon$$

em que:

$y$  é o vetor das observações da variável resposta (ou dependente) de dimensões  $n \times 1$ ;

$X$  é a matriz de incidência dos tratamentos nas repetições, compostas pelos valores (0, 1) de dimensão  $n \times 5$ ;

$\beta$  é o vetor de parâmetros relativos à média geral  $\mu$  e aos correspondentes efeitos do fator, de dimensões  $5 \times 1$ ;

$Z$  é o vetor cujos componentes são os valores da covariável ou variável independente, de dimensões  $n \times 1$ ;

$\gamma$  é o vetor do coeficiente de regressão associado à covariável de dimensões  $1 \times 1$ ;

$\epsilon$  é o vetor de erros experimentais do modelo, de dimensão  $n \times 1$ .

Sem perda de generalidades e no intuito de facilitar o processo de simulação e de entendimento, considerou-se como pólo de comparação um experimento conduzido em delineamento inteiramente casualizado com quatro tratamentos.

Para se ter uma idéia geral do comportamento dos estimadores, foram consideradas várias situações experimentais, variando-se o número de repetições, diferentes porcentagens de ajustamento do modelo linear, pelo coeficiente de determinação e diferentes distribuições dos resíduos do modelo.

Inicialmente, a variável auxiliar  $x$  foi considerada sem a presença do erro de medida ( $Z = (x_{11}, \dots, x_{IJ})$ ), em que, o subíndice  $J$  correspondeu ao número total de repetições ( $j = 1, \dots, J$ ), que assumiu os valores 4; 6; 10; 15; 20 e 25. O subíndice  $I$  é o número total de tratamentos ( $i = 1, \dots, 4$ ), previamente fixado em  $I = 4$ .

Por se tratar de um estudo de simulação para casos em modelos lineares aplicados a análise de covariância com pouca complexidade, arbitrariamente considerou-se os valores paramétricos, sendo que, o valor correspondente a média geral ( $\mu$ ) fixado em 1,5. Para os efeitos de tratamentos fixou-se os valores (2, 0; 3, 0; 4, 0; 5, 0). Desta forma definiu-se o vetor  $\beta$  por  $[\beta = (1, 5; 2, 0; 3, 0; 4, 0; 5, 0)]$ . Em relação ao parâmetro de regressão ( $\gamma$ ) fixou-se este por 0,80 em função de verificar o desempenho dos estimadores em uma situação de alta correlação entre a variável dependente ( $y$ ) e a covariável, assim  $\gamma = (0, 8)$ . Portanto, nestas condições os resultados são restritos. Entretanto, a metodologia proposta neste trabalho para a análise de vieses pode e deve ser realizada para modelos mais complexos.

O erro de medida foi incorporado por meio da relação (2.12) conforme equação abaixo:

$$z_{ij} = x_{ij} + u_{ij}$$

em que,  $x_{ij}$  é o valor observado da covariável sem erro de medida relativo ao  $i$ -ésimo tratamento e à  $j$ -ésima repetição, gerada por meio de uma distribuição Uniforme (0,1). O valor do erro de medida ( $u_{ij}$ ), foi calculado por meio de uma distribuição Normal (0,1), conforme discutido por Whitemore (1989), em que a variável aleatória tem, usualmente, distribuição gaussiana, uma vez que  $z_{ij}$  representou a covariável com o efeito do erro de medida, obtida da relação linear entre  $z_{ij}$  e  $u_{ij}$ .

Para avaliar se o coeficiente de determinação pode afetar no processo de estimação, em que, para o mesmo, considerou-se diferentes situações de tamanhos amostrais, distribuições dos erros e dos estimadores avaliados (James & Stein, 1961; Fuller, 1987), considerou-se diferentes valores para o coeficiente de determinação ( $R^2$ ) previamente fixados em 20%; 50%; 70% e 90% sobre a restrição da equação (2.22).

$$Eyy = \frac{[(1 - R^2) \times S(\gamma|\tau_i)]}{R^2}$$

sendo  $Eyy$  a soma de quadrados do erro na variável dependente;  $S(\gamma|\tau_i)$  a soma de quadrados para o parâmetro regressivo e  $R^2$  o coeficiente de determinação do modelo.

A justificativa das diferentes distribuições dos resíduos adotadas tem como objetivo avaliar o comportamento dos estimadores sobre a violação de normalidade, assim como comparará-los quando a normalidade está presente. Para isso, foram

adotadas as seguintes distribuições: a Normal padrão (0,1), a Uniforme (0,1) e a Gama (4,1), tendo esta última sido escolhida por ter um comportamento assimétrico. Desse modo, a avaliação de acurácia e precisão abrange certa quantidade de diferentes configurações. Para se obter uma variável aleatória das distribuições citadas, foram utilizadas rotinas construídas no software estatístico R, versão, 2.5.1. (R Core Development Team, 2007). Foram simuladas diferentes situações, considerando as configurações que a covariável assume nas distintas repetições, as distribuições envolvidas para o modelo em questão e os coeficientes de determinação do modelo.

Para cada situação simulada (coeficiente de determinação, diferentes distribuições do erro e tamanhos amostrais), as estimativas dos parâmetros do modelo (2.9), obtidas por meio das equações dos estimadores de James-Stein e Fuller, respectivamente, foi avaliada a acurácia, por meio do viés médio, pela expressão:

$$Vm = \frac{\sum_{l=1}^N (\hat{y}_l - y_l)}{N}$$

e a precisão, utilizando-se o erro de predição (2.21), dado por,

$$Rqmep = \sqrt{\frac{\sum_{l=1}^N (\hat{y}_l - y_l)^2}{N}} \quad \text{e} \quad Ep = \sqrt{Rqmep^2 - (Vm)^2}$$

Para avaliação do modelo estimado utilizou-se o viés médio e erro residual, por meio de gráficos, os quais consideraram diferentes distribuições do resíduo, diferentes estimadores dos parâmetros do modelo ("Plug-in" e atenuador de vício), diferentes tamanhos amostrais e dois coeficientes de determinação ( $R^2 = 20\%$  e  $90\%$ ). A razão pela qual considerou-se os casos mais extremos ( $R^2 = 20\%$  e  $90\%$ ) é devido ao fato de avaliar a performance dos estimadores na situação de

baixa explicação dos dados e alta explicação dos dados.

A acurácia e a precisão foram calculadas com base nos valores medidos das 5.000 simulações (N) realizadas para cada situação. Para a realização dos cálculos e simulações, foi elaborado um algoritmo construído no software R, versão 2.5.1 (R Core Development Team, 2007).

Para estimativa do parâmetro de regressão, analisou-se apenas o viés médio em três tamanhos amostrais ( $n=16, 60$  e  $100$ ), quatro coeficientes de determinação ( $R^2=20\%, 50\%, 70\%$  e  $90\%$ ) e três distribuições dos resíduos do modelo (Normal, Gama e Uniforme), para os dois estimadores dos parâmetros do modelo de ANCOVA com erro de medida ("Plug-in" e atenuador de vício).

## 4 RESULTADOS E DISCUSSÃO

O critério estatístico utilizado na avaliação dos estimadores deu-se em função da análise do viés médio e do erro residual. Os resultados do viés médio permitiram inferir sobre a acurácia e os erros residuais sobre a precisão dos modelos lineares em estudo. O viés médio e o erro residual foram estimados para diferentes configurações, representadas pelos tamanhos amostrais, qualidades de ajuste e distribuições dos resíduos. Para cada situação, gerou-se uma variável resposta, cujos modelos explicassem diferentes porcentagens da variabilidade total, quantificada pelos valores do coeficiente de determinação  $R^2$ , ilustrados graficamente (Figuras 4.1 - 4.8) neste trabalho para os valores 20% e 90%.

Os resultados referentes à acurácia dos modelos lineares, obtidos por meio do viés médio, indicam que o modelo mais acurado é aquele cujo valor seja o mais próximo do valor zero, enquanto o resultado referente à precisão sugere interpretar o grau de "fineza" com que a variável resposta foi medida. Em outras palavras, a precisão diz respeito à dispersão dos valores estimados pelo modelo. Notoriamente, qualquer estatística correspondente à variabilidade, como, por exemplo, erro padrão e erro residual, poderá ser empregada para estudo da precisão. Evidentemente, quanto menor for o valor, seja do erro residual ou do erro padrão, melhor precisão o modelo proporcionará.

Convém salientar que a acurácia e a precisão, de certa forma, encontram-se relacionadas de forma diretamente proporcionais, ou seja, dado um valor elevado do viés médio, como consequência ocasionará um inflacionamento na estimativa do erro residual, levando a um modelo que apresente deficiência na precisão. Feita esta colocação, justifica-se a discussão dos modelos lineares (ANCOVA) com erro

nas variáveis, tendo em vista que as variáveis dependente e independente possuem componentes aleatórios e são correlacionadas. Os resultados são dados a seguir pelo viés médio (em percentuais) (Figuras 4.1-4.4) e o erro residual (em percentuais) (Figuras 4.5-4.8), comparando-se os estimadores em questão.

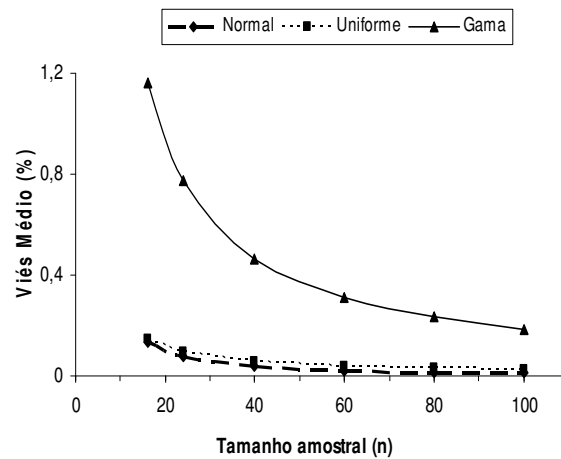


FIGURA 4.1: Viés médio do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Os resultados demonstrados na Figura 4.1 correspondem ao viés médio obtido no modelo linear de ANCOVA com erro de medida expressão (2.19), em que seus parâmetros foram estimados utilizando-se o estimador "Plug-in". Com o propósito de avaliar a influência desse estimador, em um modelo que representasse uma situação de baixa qualidade de ajuste, gerou-se o modelo linear pré-fixando-se o valor do  $R^2 = 20\%$ .

Comparando-se o viés médio obtido no ajuste dos três modelos diferenciados pela distribuição dos resíduos: normal (0,1), uniforme (0,1) e gama (4,1), observou-se que, no caso das amostras menores ( $n=16$ ), os estimadores foram menos acurados, pois apresentaram vieses superiores a 10%. Claramente, essa baixa acurácia foi mais pronunciada (10 vezes maior) ao considerar o modelo com os resíduos dados pela distribuição Gama (4,1). Estes resultados, sobre um baixo coeficiente de determinação, estão de acordo com os encontrados por Chan et al. (2004), tendo em vista que, os autores concluíram que os estimadores que consideram os erros nas covariáveis, são viesados em pequenas amostras, por mais que não sejam os mesmos estimadores, os resultados são parecidos.

Por meio da Figura 4.1 percebeu-se que o estimador "Plug-in" foi viesado para pequenas amostras ( $n=16$ ), fato este esperado, pois, na experimentação ao se utilizar um estimador para um parâmetro de um modelo estatístico, a comparação entre as médias é recomendada apenas para um mínimo de 20 observações por parcela, isto se, verifica empiricamente por meio de experimentos consagrados na literatura, onde o número pequeno de observações tem como consequência um aumento do erro experimental, prejudicando a precisão dos experimentos e consequentemente dificultando a fidelidade nas comparações entre as médias. Por meio da mesma Figura 4.1 pode-se perceber também que o tamanho amostral ( $n=40$ ) é apropriado para este experimento, sobre normalidade dos resíduos, pois a par-

tir deste tamanho amostral o ganho em acurácia é muito pequeno. Outro ponto importante a ser destacado é que, mesmo na melhor situação simulada para esta qualidade de ajuste ( $n=100$ ), o viés médio do estimador "Plug-in" sobre a distribuição gama (4,1) não se aproximou suficientemente do viés médio das distribuições normal e uniforme, estando, neste tamanho amostral, superior à casa dos 18%. Contudo, vale ressaltar que, de modo geral, o aumento do tamanho amostral propiciou uma melhoria na acurácia, em virtude de o viés médio ter sido reduzido nos três casos, chegando perto de 1% para as distribuições normal padrão e uniforme. Whitermore (1989) obteve resultados parecidos, pois conforme o tamanho da amostra aumentava a raiz quadrada do quadrado médio do erro (rqmep) diminuía, indicando uma melhora assintótica para o estimador de James-Stein.

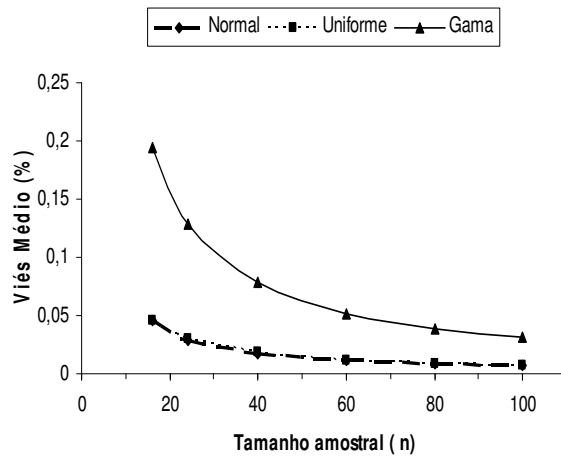


FIGURA 4.2: Viés médio do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Aumentando do coeficiente de determinação,  $R^2 = 90\%$ , os resultados (Figura 4.2) evidenciaram que os modelos com os resíduos distribuídos por uma Normal

(0,1) e Uniforme (0,1) foram mais acurados, tendo um comportamento similar em todos os tamanhos amostrais estudados, incluindo até mesmo as amostras menores nas quais o viés médio não ultrapassou 5% em nenhum dos dois casos. O modelo distribuído pela Gama (4,1), para grandes valores da amostra ( $n=100$ ), teve um ganho considerável (5 vezes) em relação ao mesmo modelo, quando o coeficiente de determinação era pequeno. Mesmo assim, o modelo de ANCOVA, quando distribuído pela Gama, obteve a pior acurácia em relação aos demais.

Este fato não foi verificado quando o modelo com os resíduos dados pelas demais distribuições estudadas e, portanto, há evidências para afirmar que o modelo de ANCOVA com erros de medida, cujas estimativas são obtidas pelos estimadores "Plug-in" é mais acurado nas situações em que este explica alta variabilidade dos dados amostrais com os resíduos dados por uma distribuição simétrica.

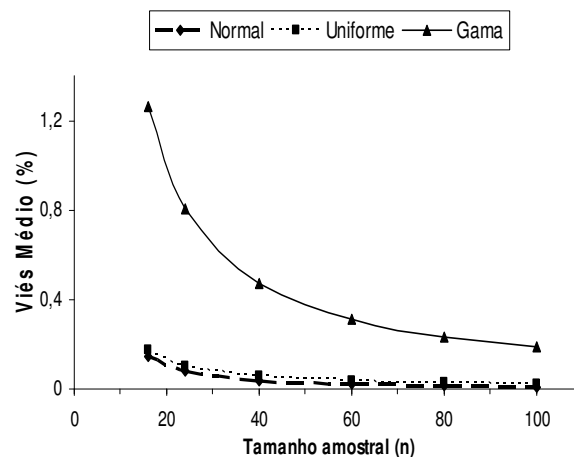


FIGURA 4.3: Viés médio do estimador atenuador de vício considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Em se tratando das estimativas dos parâmetros dos modelos, obtidas pelos estimadores de Fuller, normalmente, denominados como atenuadores de vício, para o caso de o modelo estar com baixo coeficiente de determinação,  $R^2 = 20\%$  (Figura 4.3), o viés médio para este estimador apresentou comportamento muito semelhante ao obtido pelo estimador “Plug-in”(Figura 4.1), confirmando que os vieses são menores quando os erros têm distribuição normal. Porém, existe uma leve vantagem para o estimador “Plug-in” sobre esta distribuição dos resíduos. Embora possa parecer algo relativamente desprezível, se considerarmos aspectos inferências, tais como testes de hipóteses que envolvam estimativas dos parâmetros, supostamente, a confiabilidade dos resultados poderá ser comprometida de acordo com o método de estimação utilizado.

Dado um alto coeficiente de determinação ( $R^2 = 90\%$ ), cujos resultados encontram-se ilustrados na Figura (4.4), pode-se verificar que o viés médio estimado para todos os tamanhos amostrais avaliados foram praticamente os mesmos valores verificados em relação ao estimador “Plug-in”(Figura 4.2). Este fato permitiu inferir que, mediante o alto coeficiente de determinação, ambos os métodos foram precisos, porém, acurados apenas para os modelos com distribuições Normal e Uniforme, com um pequeno destaque para a distribuição simétrica.

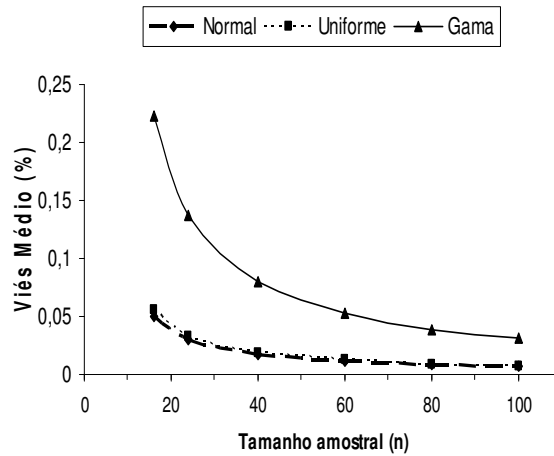


FIGURA 4.4: Viés médio do estimador atenuador de vício considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Pereira (2004) comenta que a confiabilidade das medidas, seja por instrumentos não calibrados ou outra causa de variação controlável, pode ser avaliada por meio do viés. Assim sendo, a diferenciação dos resultados provenientes de diferentes métodos de estimação pode, eventualmente, levantar dúvidas sob essa interpretação, se o estimador com menor viés não for utilizado.

Os resultados encontrados nas Figuras 4.5-4.8 corresponderam à análise do erro residual, uma medida que expressa o grau de precisão. Todavia, vale ressaltar que, pela própria definição do erro, especificada na metodologia (distância média entre os valores preditos e os observados), uma grande variabilidade entre os mesmos, tem, por conseguinte, uma perda na acurácia. Além do mais, excluindo-se o viés médio, o erro residual passa a ser interpretado como erro de predição. Desse modo, tendo em vista as mesmas situações avaliadas na análise da acurácia por meio do viés médio, procedeu-se com a discussão analisando-se o erro de

predição como forma de inferir sobre a precisão do modelo.

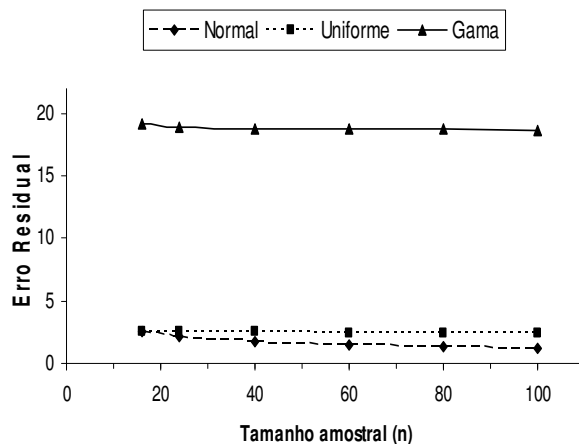


FIGURA 4.5: Erro residual do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Considerando os modelos lineares cujos parâmetros foram estimados utilizando-se os estimadores "Plug-in", verificou-se que o modelo com os resíduos distribuídos pela distribuição Gama (4,1), foi o menos preciso em relação aos demais. Esse comportamento foi verificado para todos os tamanhos amostrais e diferentes coeficientes de determinação, especificados, respectivamente, pelos valores de  $R^2 = 20$  e  $90\%$  (Figura 4.5-4.6). Em relação a outras distribuições dos resíduos Normal (0,1) e Uniforme (0,1), ao confrontar ambos os modelos com diferentes valores de  $R^2$ , notou-se que ambos os modelos apresentaram alta precisão, confirmada pelos baixos valores do erro residual, quando estes apresentam alta explicação da variabilidade amostral (Figura 4.6).

Contudo, respeitando-se os resultados de acurácia verificada para tais modelos (Figura 4.4), uma vez que, dependendo o interesse do pesquisador, ora privilegia

acurácia ora precisão, cabe salientar a sua relação. Nesse sentido, Pereira (2004) afirma que o pesquisador poderá entusiasmar-se com medidas de alta precisão, sem ponderar seus efeitos sobre acurácia. Dessa forma, estudos de baixa validade poderão ser concluídos. Assim sendo, existe respaldo em afirmar que o modelo com alta qualidade de ajuste (Figura 4.6) com os resíduos distribuídos pela uniforme (0,1) e normal (0,1) apresentaram resultados satisfatórios tanto na precisão como também na acurácia (Figura 4.4), com maior pronunciamento para amostras maiores.

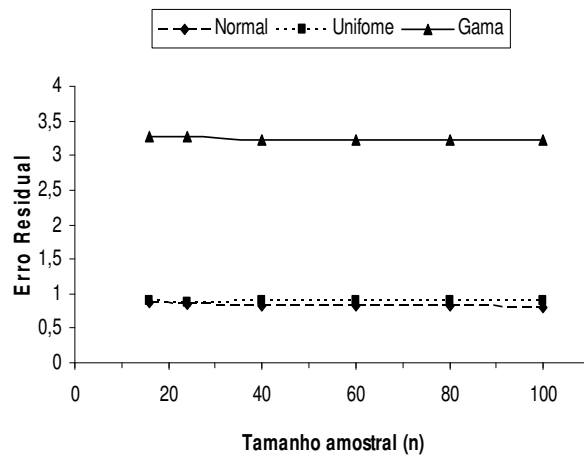


FIGURA 4.6: Erro residual do estimador "Plug-in", considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Em se tratando dos modelos lineares com os estimadores dos parâmetros dados pelo método de Fuller (1987), considerando coeficiente de determinação,  $R^2 = 20\%$  (Figura 4.7) e  $R^2 = 90\%$  (Figura 4.8), observou-se praticamente a mesma precisão, ao comparar-se o erro residual dos modelos, cujas estimativas foram obtidas pelo método "Plug-in"(Figura 4.5) e (Figura 4.6), respectivamente. Um

resultado de destaque entre os estimadores é quando comparam-se os coeficientes de determinação, pois, claramente, percebe-se que o coeficiente de determinação ( $R^2$ ) obteve grande influência na precisão do estimador, principalmente quando este é distribuído pela Gama (4,1), de forma que a precisão foi 6 vezes maior na situação de alto coeficiente de determinação. Porém, entre os estimadores, os valores não foram tão incisivos, pelo fato de que, os métodos de James-Stein (1961) e Fuller (1987) obtiveram valores próximos sobre tamanho amostral alto ( $n=100$ ) e alto coeficiente de determinação ( $R^2 = 90\%$ ), tanto para a precisão quanto para a acurácia, com leve vantagem para o estimador “Plug-in”, em todas as distribuições dos resíduos. Estes resultados estão em conformidade com os obtidos por Cunha & Colosimo (2003), haja vista que, em seu estudo os autores obtiveram intervalos de confiança por meio da técnica *bootstrap* de modo que, o estimador James-Stein obteve melhor desempenho que o estimador de Fuller, pois seus intervalos de confiança apresentavam comprimentos menores.

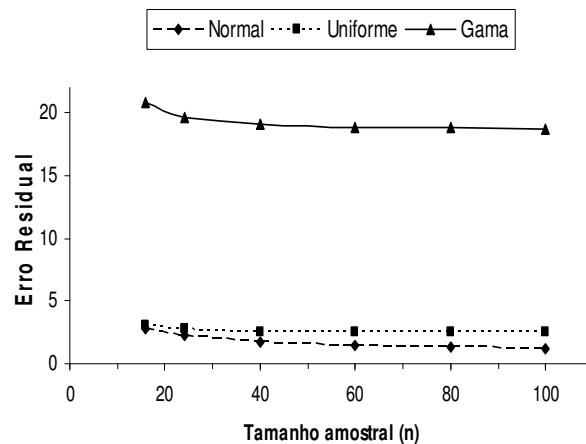


FIGURA 4.7: Erro residual do estimador atenuador de vício, considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 20\%$ , em relação aos diferentes tamanhos amostrais do experimento.

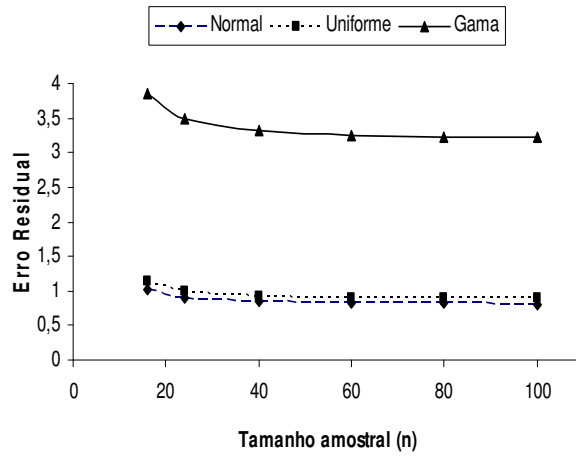


FIGURA 4.8: Erro residual do estimador atenuador de vício, considerando o modelo linear com os resíduos distribuídos por: Normal (0,1), Uniforme (0,1) e Gama (4,1) e coeficiente de determinação  $R^2 = 90\%$ , em relação aos diferentes tamanhos amostrais do experimento.

Em virtude do que foi mencionado, constatou-se que os modelos avaliados, tanto na precisão como na acurácia, apresentaram resultados mais satisfatórios para grandes amostras e foram bastante influenciados pelo coeficiente de determinação. Assim sendo, particularmente considerando os tamanhos amostrais (n) e todos os coeficientes de determinação simulados ( $R^2 = 20\%$ ,  $50\%$ ,  $70\%$  e  $90\%$ ), realizou-se um estudo em relação ao viés médio das estimativas do parâmetro regressivo do modelo para os dois estimadores em questão, “Plug-in” e atenuador de vício, cujos resultados encontram-se na Tabela 4.1.

TABELA 4.1: Viés médio das estimativas dos parâmetros do modelo linear sobre diferentes qualidades de ajuste ( $R^2$ ), tamanhos amostrais e as três distribuições dos erros, nos dois métodos de estimação avaliados.

$R^2$	Estimadores	Normal		
		16	60	100
20	Plug in*	0,054919	0,009547	0,005866
	atenuador**	0,096970	0,006733	0,007415
50	Plug in	0,018292	0,008300	0,001491
	atenuador	0,071672	0,004164	0,045806
70	Plug in	0,014994	0,005593	0,001337
	atenuador	0,068697	0,001444	0,002878
90	Plug in	0,013116	0,001979	0,000906
	atenuador	0,066333	0,001189	0,001971
		Uniforme		
20	Plug in	0,029191	0,002289	0,001735
	atenuador	0,083336	0,006478	0,000198
50	Plug in	0,013597	0,000669	0,001467
	atenuador	0,066704	0,005504	0,000176
70	Plug in	0,013591	0,000627	0,001029
	atenuador	0,056993	0,004805	0,000140
90	Plug in	0,007316	0,000082	0,000614
	atenuador	0,050574	0,004260	0,000129
		Gama		
20	Plug in	0,114882	0,005671	0,004294
	atenuador	0,172881	0,006609	0,017704
50	Plug in	0,022638	0,002170	0,003081
	atenuador	0,088506	0,005356	0,004626
70	Plug in	0,013673	0,001407	0,002064
	atenuador	0,067441	0,003585	0,000527
90	Plug in	0,002830	0,001124	0,001392
	atenuador	0,055516	0,002921	0,000146

\* Estimador de James-Stein

\*\*Estimador de Fuller

Avaliando-se os resultados referentes aos vieses das estimativas do coeficiente de regressão ( $\gamma$ ), pode-se observar que ambos os métodos de estimação foram acurados e precisos, adotou-se com critério para precisão a diferenciação entre o número de casa decimais, onde julgou-se haver diferença quanto a precisão se os resultados entre os métodos diferirem até a 3<sup>a</sup> casa decimal. Sendo assim foi notório a acurácia e a precisão para todas as situações envolvendo os modelos nas diferentes distribuições dos resíduos, coeficientes de determinação e tamanhos amostrais, simuladas conforme os valores do  $R^2$  previamente fixados.

Algo importante a ser comentado é que, nas situações que envolviam a distribuição Normal, o estimador de James-Stein obteve melhor performance em relação à acurácia do parâmetro  $\gamma$  sobre o tamanho amostral ( $n=16$ ), avaliado nas diferentes configurações propostas ( $R^2$ ) e mesma precisão, sobre o tamanho intermediário de amostra ( $n=60$ ), o estimador atenuador de vício foi mais acurado que o seu concorrente, fato este importante pois, o pesquisador ao utilizar este tamanho amostral, claramente pode optar por um dos estimadores. Na maior configuração do tamanho amostral ( $n=100$ ), praticamente em todos os coeficientes de determinação avaliados, o estimador de James-Stein, foi pelo menos duas vezes mais acurado que o seu concorrente, a exceção quando o valor do coeficiente de determinação era ( $R^2 = 20\%$ ), neste coeficiente de determinação a diferenças entre os métodos foi pequena, com leve vantagem para o estimador "Plug-in".

Quando os estimadores foram distribuídos pela uniforme, os seus comportamentos dependeram do tamanho amostral, em outras palavras, para pequenas e médias amostras, houve vantagem para o estimador "Plug-in" quanto a acurácia, enquanto para amostras maiores o estimador de Fuller obteve melhor acurácia ( $n=100$ ). Vale ressaltar que, sobre esta distribuição ambos os métodos foram precisos. Este comportamento pode ser justificado pelo fato de que, na falta de

normalidade, o estimador “Plug-in” perde em desempenho com relação a acurácia se comparado ao desempenho obtido pelo estimador atenuador de vício.

Na distribuição Gama, a maior diferenciação entre a acurácia dos métodos “Plug-in” e atenuação de vício foi restringida para os tamanhos amostrais ( $n=16$  e  $60$ ), onde sobre estes tamanhos amostrais o estimador de James-Stein superou o estimador de Fuller. Quando comparado os estimadores no tamanho amostral ( $n=100$ ), percebeu-se que conforme o coeficiente de determinação ia aumentando, o estimador atenuador de vício ia diminuindo a sua diferença de acurácia para o estimador “Plug-in”. Este fato ocorreu para todos os coeficientes de determinação simulados e sobre o tamanho amostral ( $n=100$ ). Sendo que o estimador de Fuller (1987) superou o estimador de James-Stein à partir do coeficiente ( $R^2 = 70\%$ ). Vale ressaltar que esses resultados retrataram uma pequena vantagem desse método, em situações de falta de normalidade na estimação do parâmetro de regressão  $\gamma$  em grandes amostras.

## 5 CONCLUSÕES

Os resultados encontrados neste trabalho ajudam o pesquisador na escolha do melhor método de estimação, de forma que é possível diferenciar entre os estimadores avaliados, quanto a acurácia e a precisão, no intuito de se obter melhores estimativas para os modelos de ANCOVA, quando estes apresentam erro nas covariáveis, diminuindo assim o erro experimental.

Para a comparação entre os métodos baseado nas equações 2.19 e 2.21 os estudos de simulação demonstraram que o método "Plug-in" obteve melhor desempenho em relação ao atenuador de vício, quanto à acurácia e à precisão em situações de normalidade, nos diferentes tamanhos amostrais e coeficientes de determinação estudados.

O coeficiente de determinação do modelo e o tamanho amostral exerceram grande influência, tanto na precisão quanto na acurácia dos valores estimados pelos modelos analisados. De modo que, quanto maior o tamanho amostral e melhor o coeficiente de ajustamento, maior era o incremento na precisão e na acurácia dos estimadores avaliados, em qualquer configuração da distribuição dos resíduos.

Os métodos, quando avaliados na distribuição Gama (4,1), obtiveram o pior desempenho em relação às demais distribuições analisadas, tanto para precisão quanto para a acurácia, avaliadas por meio das equações 2.19 e 2.21, com vantagem novamente para o estimador "Plug-in". Deste modo, os estimadores foram não eficientes quando os erros eram distribuídos pela distribuição assimétrica.

O viés da estimativa do parâmetro de regressão ( $\gamma$ ) para os estimadores James-Stein e Fuller foram precisos e acurados com vantagem para o estimador de James-Stein quando os erros eram distribuídos pela Normal, e vantagem para o estimador

de Fuller no tamanho amostral ( $n=100$ ) e alto coeficiente de determinação, fora da presença de normalidade dos erros.

## 6 ESTUDOS FUTUROS

1. Implementar o estimador de James-Stein para o modelo de análise de covariância no software R, por meio de um pacote. Para deste modo, facilitar seu uso a um pesquisador;
2. Avaliar os estimadores quanto a acurácia e a precisão no modelo de análise de covariância com erro de medida, no delineamento em blocos casualizados;
3. Estender o uso dos estimadores para a regressão múltipla, devido a necessidade prática e avaliá-los via simulação Monte Carlo.
4. Avaliar a taxa de erro tipo I e poder de testes de comparações múltiplas quando os parâmetros do modelo são estimados por estes métodos.

## REFERÊNCIAS BIBLIOGRÁFICAS

AMORIN, G. P. **Confiabilidade de rede GPS de referência cadastral municipal, estudo de caso: rede do município de Vitória (ES)**. 2004. 149 p. Dissertação (Mestrado em Transportes) - Escola de Engenharia de São Carlos. Universidade de São Paulo, São Carlos.

BARKER, D. J. P.; ROSE, G. **Epidemiology in medical practice**. London: J. Wiley, 1984. 178 p.

BIBBY, J.; TOUTENBURG, H. **Prediction and improved estimation in linear models**. London: J. Wiley, 1977. 582 p.

BOLFARINE, H.; LEITE, J. G.; RODRIGUES, J. On the estimation of the size of a finite and closed population. **Biometrical Journal**, Austria, v. 5, p. 577-593, 1992.

BUSSAB, W. O.; MORETTIN, P. A. **Estatística básica**. São Paulo: Saraiva, 2004. 526 p.

CARROL, R. J.; GALLO, P.; GLEESER, L. J. Comparison of least squares and errors-in-variables regression, with special reference to randomized analysis of covariance. **Journal of the American Statistical Association**, New York, v. 80, n. 392, p. 929-932, Dec. 1985.

CHAN, S. F.; MACASKILL, P.; IRWIG, L. Walter SD Adjustment for baseline measurement error in randomized controlled trials induces bias. **Controlled Clinical Trials**, New York, v. 25, n. 4, p. 408-416, May 2004.

COCHRAN, W. G. Analysis of covariance: its nature and uses. **Biometrics**, Maryland, v. 13, n. 3, p. 261-283, Jan. 1957.

COCHRAN, W. G. Errors of measurement in statistics. **Tecnometrics**. London, v. 10, n. 4, p. 55-83, Nov. 1968.

CUNHA, W. J.; COLOSIMO, E. A. Intervalos de confiança bootstrap para modelos de regressão com erros de medida. **Revista de Matemática e Estatística**. São Paulo, v. 21, n. 2, p. 25-41, 2003.

DEAN, A.; VOSS, D. **Design and analysis of experiments**. New York: Springer, 1999. 457 p.

DEGRACIE, J. S.; FULLER, W. A. Estimation of the slope and analysis of covariance when the concomitant variable is measured with error. **Journal of the American Statistical Association**, Iowa, v. 67, n. 340, p. 930-937, Dec. 1972.

DELICATO, P. F.; ROMO, J. Goodness-of-fit tests in random coefficient regression models. **Annals of the Institute of Statistical Mathematics**, New York, v. 51, p. 125-148, Mar. 1999.

DELICATO, P. F.; ROMO, J. Random coefficient regressions: parametric goodness of-fit-tests. **Journal of Statistical Planning and Inference**, Texas, v. 119, n. 2, p. 377-400, Feb. 2004.

DEMPSTER, A. P.; LAIRD, N. M.; RUBIN, D. B. Maximum likelihood from incomplete data via the EM algorithm (with discussion). **Journal Royal Statistics Society**. New York, v. 39, n. 1, p. 1-38, 1977.

EFRON, B.; TIBSHIRANI, R. **An Introduction to the bootstrap**. New York: Chapman & Hall, 1993. 436 p.

FEDERER, W. T. **Experimental design**. New York: Macmillan, 1955. 326 p.

FISHER, R. A. **Statistical methods for research workers**. London: Edinburgh, 1932. 332 p.

FULLER, W. A. **Measurement error models**, New York: J. Wiley, 1987. 512 p.

GOMEZ, K. A.; GOMEZ, A. A. **Statistical procedures for agricultural research**. New York: J. Wiley, 1984. 680 p.

HIGDON, R.; SCHAFER, D. W. Maximum likelihood computations for regression with measurement error. **Computational Statistics & Data Analysis**, Amsterdam, v. 35, n. 3, p. 283-299, Jan. 2001.

JAMES, W.; STEIN, C. Estimation with quadratic loss. In: SYMPOSIUM ON MATHEMATICS, STATISTICS AND PROBABILITY, 1961, Berkeley. **Proceedings...** University of California, 1961.

KEMPTHORNE, O. **The design and analysis of experiments**. New York: J. Wiley, 1952. 631 p.

KSHIRSHGAR, A. M. A course in linear models. **Biometrika**, New York, v. 40, n. 3, p. 871-872, Sept. 1983.

LEAL, J. G.; PORRAS, A. M. L. **Diseño estadístico de experimentos análisis de la varianza**, Granada: G. E. Universitario, 1998, 357 p.

LIU, Q.; PIERCE, D. A. A note on gauss hermite quadrature. **Biometrika**, Oxford, v. 81, p. 624-629, May 1994.

LORD, F. M. Large-sample covariance analysis when the control variable is fallible. **Journal of the American Statistical Association**, London, v. 2, p. 307-621, 1960.

MILLIKEN, G. A.; JONHSON, D. E. **Analysis of messy data**. London: Champ & Hall, 2000. 605 p.

MONTGOMERY, D. C. **Design and analysis of experiments**. New York: J. Wiley, 1976. 327 p.

NETER, J.; WASSERMAN, W.; KUTNER, M. H. **Applied linear statistical models**. Sidney: R. D. Irwin, 1990. 413 p.

ARAÚJO NETO, J. O. **Análise de precisão e acurácia de pontos georreferenciados com a técnica do código suavizado pela fase da portadora utilizando GPS de simples frequência**. 2006. 207 p. Dissertação (Mestrado em Engenharia) - Escola de Engenharia de São Carlos. Universidade de São Paulo, São Carlos.

OSTLE, B. **Statistics in research basic concepts and techniques for research workers**. Iowa State College: Ames, 1954. 365 p.

PEREIRA, J. C. R. **Análise de dados qualitativos: Estratégias metodológicas para as ciências da saúde, humanas e sociais**. São Paulo: Edusp, 2004. 156 p.

R DEVELOPMENT CORE TEAM. **R: a language and environment for statistical computing.** Vienna, Austria: R Foundation for Statistical Computing. Disponível em: <<http://www.R-project.org>>. Acesso em: 20 dez. 2007.

RAMALHO, M. A. P.; FERREIRA, D. F.; OLIVEIRA, A. C. **Experimentação em genética e melhoramento de plantas.** Lavras: UFLA, 2000. 326 p.

RESENDE, M. D. V; DUARTE, J. B. Precisão e controle de qualidade em experimentos de avaliação de cultivares. **Pesquisa Agropecuária Tropical.** São Paulo, v. 3, p. 182-194, Sept. 2007

SEARLE, S. R. **Linear models.** New York: J. Wiley, 1971. 532 p.

SOUZA, F. G. **Análise de covariância com um fator e erro de medida na covariável.** 2004. 54 p. Dissertação (Mestrado em Estatística) - Instituto Matemática e Estatística. Universidade de São Paulo, São Paulo.

STEEL, R. G. D.; TORRIE, J. H. **Principles and procedures of statistics with special reference to the biological sciences.** New York: M. Hill, 1960. 481 p.

VELOSO, R. B. **Verificação por simulação monte carlo da eficiência da aplicação do teste de razão de verossimilhança para identidade de modelos não lineares.** 2006. 61 p. Dissertação(Mestrado em Estatística e Experimentação Agropecuária) - Universidade Federal de Lavras, Lavras.

VENCOVSKY, R.; CRUZ, C. D. Comparação de métodos de correção de rendimento de parcelas experimentais com estandes variados. **Pesquisa Agropecuária Brasileira,** Brasília, v. 26, n. 5, p. 647-657, maio 1991.

VOSS, R. E. **Response by corn to NPK fertilization on marshal and monona soils as influenced by management and meteorological factors.** Iowa, 1969. 156 p.

VUOLO, J. H. **Fundamentos da teoria dos erros.** São Paulo: E. Blücher, 1992. 236 p.

WHITTEMORE, A. S. Errors-in-variables regression using Stein estimates. **American Statistician.** New York, v. 43, n. 4, p. 226-228, June 1989.

## ANEXOS

ANEXO A	Página
PROGRAMA 1A Rotina para a análise dos dados. ....	55

### PROGRAMA 1A: Rotina para a análise dos dados.

```
##### Inicio do programa #####
library(MASS)
### Declaração das variáveis
nrep=4 ### Número de repetições
n=5 ### Número de parâmetros a serem estimados
nmatrix=nrep*(n-1) ###Tamanho de linhas
x=matrix(0,nmatrix,n)###Matriz de delineamento
z=matrix(0,nmatrix,1) ##covariável do modelo
zz=matrix(0,nmatrix,1) ##covariável do modelo sem erro
e=matrix(0,nmatrix,1)
B=c(1.5,2.0,3.0,4.0,5.0)
G=c(0.8)
I=diag(nmatrix)
nsim=5000
nsim=5000
r2=0.2
err=matrix(0,nmatrix,1)
#####Parametros iniciais#####
i=1
j=1
l=1
#### Vetor de médias das repetições ####
zmb=as.vector(matrix(0,nmatrix,1)) # Medias dos grupos
mest=matrix(0,nsim,6) # Parâmetros Estimados
dados<-matrix(0,nmatrix,2) # Dados (x,y)
pred=matrix(0,nmatrix,3)
ss=matrix(0,nsim,3)
mvies=matrix(0,2,3,dimnames=list(c(''Inter'', ''Incl''),
c(''QM'', ''James'', ''Fuller'')))
d=matrix(0,2,3,dimnames=list(c(''Vies Medio'', ''Erro
Residual''),c(''QM'', ''James'', ''Fuller'')))
### INICIO DA SIMULAÇÃO ###
sb=0 ; sf=0; sj=0;
somab=0 ; somaj=0 ; somaf=0
somaeb=0; somaef=0; somaej=0
for (q in 1:nsim)
{
### Identificação das medidas repetidas ###
```

```

c=1
aux2=0
conrta=0
auxind=0
cont=0
for (i2 in 1:nrep)
{
ind=rep(i2,nrep)
auxind=c(auxind,ind)
}
maux=as.matrix(auxind)
ind=maux[1:nmatrix+1,]
###Programa ANCOVA###
for(i in 1:nmatrix){
x[i,1]=1
}
for(j in 2:n){
for(i in 1:nrep){
x[c,j]=1
c=c+1
}
}
for(l in 1:nmatrix){
xis=runif(1,0,1)
u=rnorm(1,0,1)
z[l,]=xis+u
zz[l,]=xis
}
ob=x**B+z**G ##com erro
obs=x**B+zz**G ##sem erro
sy=as.real(sum(ob))
my=as.real(sy/nmatrix)
sigmay=(my*(1-r2))/r2
sigma2y=5*sigmay
for(k in 1:nmatrix){
e[k,]=rnorm(1,0,sqrt(sigmay))
#e[k]=runif(1,0,sqrt(sigmay))
#e[k]=rgamma(1,4,1/(sqrt(sigmay)))
}

```

```

dados[,2]=ob+e
obs+e
dados[,1]=z
auxdados=cbind(ind,dados)
### Medidas utilizadas no ajuste dos parâmetros ###
zb=mean(auxdados[1:nrow(auxdados),2])
li=1 ; lif=nrep
k=n-1
for (i2 in 1:k)
{
zm=mean(auxdados[li:lif,2:2])
aux1=t(rep(zm,nrep))
li=li+nrep
lif=lif+nrep
aux2=cbind(aux2,aux1)
}
aux3=aux2[1:1,2:ncol(aux2)]
auxdados=cbind(auxdados,aux3)
aux4=auxdados[,2]-auxdados[,4]
sigma2=(t(aux4)%*%aux4)/(nmatrix*nrep*(nrep-1))
s=t((auxdados[,2]-zb))%*%(auxdados[,2]-zb)
bhat=((sigma2*(nmatrix-3))/s)*(1/nrep)
ez=bhat*(mean(auxdados[,2]))+(1-bhat)*auxdados[,2]
xez=ez
auxdados=cbind(auxdados,xez)
###Estimativa dos parâmetros###
P=x%*%ginv(t(x)%*%x)%*%t(x)
###1° passo achar a estimativa do parametro de
regressão
ghat=ginv(t(z)%*%(I-x%*%ginv(t(x)%*%x)%*%t(x))%*%z)
%*%t(z)%*%(I-x%*%ginv(t(x)%*%x)%*%t(x))%*%auxdados[,3])
###abaixo estimativa do parametro de regressão
gamahat=ginv(t(z)%*%(I-P)%*%z)%*%t(z)%*%(I-P)
%*%auxdados[,3]###
###2° Passo achar a estimativa do parâmetros,
média, efeito de tratamentos
bh=ginv(t(x)%*%x)%*%t(x)%*%auxdados[,3]
betahat=bh-ginv(t(x)%*%x)%*%t(x)%*%z%*%gamahat
###Fim do programa ANCOVA###

```

```

### Estimativa do Modelo ###
#xest=cbind(rep(1,nmatrix),auxdados[,2])
# best=t(solve(t(xest)%*%xest)%*%t(xest)
%*%auxdados[,3])
### Estimativa James-Stein ###
estj=ginv(t(auxdados[,5])%*%(I-P)%*%auxdados[,5])
%*%t(auxdados[,5])%*%(I-P)%*%auxdados[,3]
bestj=ginv(t(x)%*%x)%*%t(x)%*%auxdados[,3]
betaestj=bestj-ginv(t(x)%*%x)%*%t(x)%*%auxdados[,5]
%*%estj
estj=t(solve(t(xez)%*%xez)%*%t(xez)%*%auxdados[,3])
# estj
###fuller###
aux5=(auxdados[,2]-zb)
l=mean(auxdados[,2])
sigmaz=sum(((l)-auxdados[,2]))^2)/(nmatrix-1)
kx=sigmaz/(sigmaz-sigma2)
pred[,1]=x%*%betahat+auxdados[,2]%*%ghat
pred[,2]=x%*%betaestj+auxdados[,5]%*%estj
pred[,3]=x%*%(betahat%*%kx)+(auxdados[,2]%*%ghat%*%kx)
difm=abs((pred[,1]-obs))
difj=abs((pred[,2]-obs))
diff=abs((pred[,3]-obs))
vb=sum(difm[1])/nmatrix
vj=sum(difm[1])/nmatrix
vf=sum(difm[1])/nmatrix
errob=sqrt((sum(((pred[,1]-obs)^2))/nmatrix))
erroj=sqrt((sum(((pred[,2]-obs)^2))/nmatrix))
errof=sqrt((sum(((pred[,3]-obs)^2))/nmatrix))
somab=somab+vb
somaeb=somaeb+errob
somaj=somaj+vj
somaej=somaej+erroj
somaf=somaf+vf
somaef=somaef+errof
mest[q,1]=betahat[1,1]+betahat[2,1]+betahat[3,1]
+betahat[4,1]+betahat[5,1]
mest[q,2]=ghat
mest[q,3]=betaestj[1,1]+betaestj[2,1]+betaestj[3,1]

```

```

+betaestj[4,1]+betaestj[5,1]
mest[q,4]=estj
mest[q,5]=(betahat[1,1]+betahat[2,1]+betahat[3,1]
+betahat[4,1]+betahat[5,1])*kx
mest[q,6]=ghat*kx
vpm=0 ; vpj=0; vpf=0
vpm=somab/nsim
vpf=somaf/nsim
vpj=somaj/nsim
errovm=somaeb/nsim
errovf=somaef/nsim
errovj=somaej/nsim
viesbo=abs((mean(mest[,1])-(B[1]+B[2]+B[3]+B[4]+B[5])
/(B[1]+B[2]+B[3]+B[4]+B[5])))
viesb1=abs((mean(mest[,2]-G)/G)
viesboj=abs((mean(mest[,3])-(B[1]+B[2]+B[3]+B[4]+B[5])
/(B[1]+B[2]+B[3]+B[4]+B[5])))
viesb1j=abs((mean(mest[,4]-G)/G)
viesbof=abs((mean(mest[,5])-(B[1]+B[2]+B[3]+B[4]+B[5])
/(B[1]+B[2]+B[3]+B[4]+B[5])))
viesb1f=abs((mean(mest[,6]-G)/G)
viespred[1,1]=vpm
viespred[2,1]=errovm
viespred[1,2]=vpj
viespred[2,2]=errovj
viespred[1,3]=vpf
viespred[2,3]=errovf
mvies[1,1]=viesbo
mvies[2,1]=viesb1
mvies[1,2]=viesboj
mvies[2,2]=viesb1j
mvies[1,3]=viesbof
mvies[2,3]=viesb1f
}
mvies ####saida do programa viés médio dos parâmetros
viespred ####saida do progama viés do modelo ajustado
write.table(mvies, '' mviesnrep4R0,2.txt'')
write.table(viespred, ''viesprednrep4R0,2.txt'')
#####

```