ANÁLISE DE COVARIÂNCIA LINEAR PARA ENSAIOS COM UM FATOR CONSIDERANDO UMA REGRESSÃO LINEAR ESPECÍFICA PARA CADA TRATAMENTO

JOÃO GIL DE LUNA
Estatístico

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo para obtenção do título de Mestre em Agronomia: Área de Concentração Estatística e Experimentação Agronômica.

PIRACICABA

Estado de São Paulo - Brasil

Março, 1987

ANÁLISE DE COVARIÂNCIA LINEAR PARA ENSAIOS COM UM FATOR CONSIDERANDO UMA REGRESSÃO LINEAR ESPECÍFICA PARA CADA TRATAMENTO

JOÃO GIL DE LUNA
Estatístico

Orientador: Prof. Dr. HUMBERTO DE CAMPOS

Dissertação apresentada à Escola Superior de Agricultura "Luiz de Queiroz", da Universidade de São Paulo para obtenção do título de Mestre em Agronomia: Área de Concentração Estatística e Experimentação Agronômica.

PIRACICABA

Estado de São Paulo - Brasil

Março, 1987
def: 10/04/87

À minha esposa Lili,
Aos meus filhos Iviane e Igor,
Aos meus pais

DEDICO.

Em memória de meu irmão
JOSÉ LEANDRO DE LUNA
Um amigo e incentivador

AGRADECIMENTOS

- Ao Prof. Dr. Humberto de Campos, pela amizade e pela orientação segura e valiosa prestada durante todo o curso de pós-graduação.
- Ao Prof. Dr. Antônio Francisco Iemma, pelo incentivo e valiosas sugestões.
- Ao PLANALSUCAR (Programa Nacional de Melhoramento da Cana-de-Açúcar) pela oportunidade, ao Coordenador Regional Norte Engº Agrº Carlos Eduardo Lins e Silva Pires e ao Engº Agrº José Alberto Campelo Marroquim de Souza, pela confiança e ajuda.
- Ao Colega Engº Agrº Lusbi Antonio Herrera Sandoval pelo apoio durante a realização deste trabalho
- Aos Professores do Curso de Pós-Graduação em Estatística e Experimentação Agronômica da ESALQ/USP, na pessoa do Prof. Décio Barbin, por sua acolhida, seus ensinamentos e pela solicitude com a qual nos atenderam dentro e fora das salas de aula.
- Aos Colegas do Curso de Pós-Graduação em Estatística e Experimentação Agronômica, pelo convívio e pela lição que cada um nos ensinou.
- Aos Colegas do PLANALSUCAR, Engº Agrº Norberto Antônio Lavorenti, Geraldo Veríssimo de Sousa Barbosa, Sônia Maria

de Stefano Piedade, Eduardo Fernando Sobral da Costa e Ricardo Otaviano Ribeiro de Lima, pela amizade e colaboração.

À Srta. Maria Sandra da Silva, pela datilografia e impressão.

À minha família, pela paciência e dedicação.

À todos aqueles que contribuíram direta e indiretamente na execução deste trabalho.

ÍNDICE

	Página
RESUMO	viii
SUMMARY	x
1. INTRODUÇÃO	01
1.1. Identificação do Problema	01
1.2. Objetivos	03
2. REVISÃO DE LITERATURA	05
2.1. Introdução	05
2.2. Finalidades	06
2.3. Restrições ao Uso da Análise de Covariância .	07
2.4. Teste de Homogeneidade dos Coeficientes de Regressão Linear	10
2.5. Formulação Geral da Análise de Covariância .	12
2.6. Um Modelo Alternativo	16
3. DESENVOLVIMENTO TEÓRICO	21
3.1. Modelo Estatístico	21
3.2. Sistema de Equações Normais e Estimativas dos Parâmetros,	24
3.3. Média e Variância dos Estimadores dos Parâme- tros	28
3.4. Estimabilidade	32
3.5. Análise de Variância e Teste de Hipótese ...	36
3.6. Comparações Múltiplas	56

	Página
4. ILUSTRAÇÃO DO MÉTODO	57
4.1. Estimativas dos Parâmetros	58
4.2. Análise de Variância para o Ajuste Covarian <u>te</u>	59
4.3. Teste de Homogeneidade dos Coeficientes de Regressão Linear	60
4.4. Análise de Variância para o Ajuste de Trata <u>mentos</u>	61
4.5. Estimativas das Médias dos Tratamentos	63
4.6. Comparações Múltiplas	63
5. CONCLUSÕES	66
5.1. O Modelo Estatístico	66
5.2. O Sistema de Equações Normais	67
5.3. As Funções Lineares Estimáveis	67
5.4. Os Graus de Liberdade para o Resíduo	67
5.5. A Análise de Variância	68
5.6. Teste de Hipótese	68
5.7. Comparações Múltiplas	69
6. BIBLIOGRAFIA	70
7. APÊNDICE	73

ANÁLISE DE COVARIÂNCIA LINEAR PARA ENSAIOS COM UM FATOR CONSIDERANDO UMA REGRESSÃO LINEAR ESPECÍFICA PARA CADA TRATAMENTO

Autor: JOÃO GIL DE LUNA

Orientador: Prof. Dr. HUMBERTO DE CAMPOS

RESUMO

Neste trabalho, considerou-se o caso da análise de covariância linear para ensaios com um fator levando-se em conta uma equação de regressão linear específica para cada tratamento.

Para tanto, adotou-se o modelo estatístico:

$$y_{ij} = \mu + t_i + \beta_i z_{ij} + e_{ij} ;$$

Onde:

y_{ij} é a j-ésima observação da variável dependente no tratamento i, ($i = 1, 2, \dots, v$ e $j = 1, 2, \dots, n_i$);

μ é uma constante inerente a todos os dados;

t_i é o efeito do tratamento i;

β_i é o coeficiente de regressão linear dentro do tratamento i;

z_{ij} é a j-ésima observação covariante correspondente ao tratamento i, fixa e tomada sem erro;

e_{ij} é o erro casual atribuído a observação y_{ij} , com distribui

ção normal de média zero e variância σ^2 , isto é, $e_{ij} \overset{\text{IID}}{\sim} N(0, \sigma^2)$.

Com base neste modelo, foram determinados: O sistema de equações normais; as estimativas dos parâmetros envolvidos no modelo; a matriz de dispersão para os estimadores dos coeficientes de regressão covariante; os componentes de variância para os efeitos considerados fixos do modelo; os critérios para os testes das hipóteses de nulidade para efeito de regressão covariante, homogeneidade dos coeficientes de regressão e efeitos ajustados dos tratamentos; como também, os critérios para comparações múltiplas pelo métodos de TUKEY e teste t com base nas funções lineares estimáveis para efeitos dos tratamentos.

LINEAR COVARIANCE ANALYSIS FOR EXPERIMENTS WITH ONE FACTOR CONSIDERING A SPECIFIC LINEAR REGRESSION FOR EACH TREATMENT

Author: JOÃO GIL DE LUNA

Adviser: Prof. Dr. HUMBERTO DE CAMPOS

SUMMARY

This paper refers to the linear covariance analysis for experiments with one factor when these have different linear regression equations within the treatments.

The following statistic model was applied:

$$y_{ij} = \mu + t_i + \beta_i z_{ij} + e_{ij};$$

Where:

y_{ij} is the j 'th observation of the dependent variable in the i 'th treatment, ($i = 1, 2, \dots, v$; $j = 1, 2, \dots, n_i$);

μ is an inherent constant to all data;

t_i is the effect of the i 'th treatment;

β_i is the linear regression coefficient within the i 'th treatment;

z_{ij} is the j 'th covariant observation correspondent to the i 'th treatment, fixed and taken without error;

e_{ij} is the random error attributed to observation y_{ij} , with n_{ij}

mal distribution of mean zero and variance σ^2 , e.g., $e_{ij} \stackrel{\text{IID}}{\sim} N(0, \sigma^2)$.

Under these circumstances the following items were determined: the normal equations system; the estimates of the involved parameters in the model; the dispersion matrix for the estimates of the covariant regression coefficient; the variance components for the effects considered fixed in the model; the criteria for the null hypothesis test for covariates regression effects, the homogeneity of the regression coefficients and the effects adjusted to the treatments; likewise the criteria for the multiple comparisons by method of TUKEY and t test based on estimable linear functions for the treatment effects.

1. INTRODUÇÃO

1.1. Identificação do Problema

Ao planejar um experimento, o pesquisador deve ter em mira o conhecimento dos fatores que influenciam no resultado final da pesquisa. Esta preocupação é de fundamental importância, pois, não muito raro, eles podem refletir negativamente na qualidade da análise de dados da pesquisa, aumentando o erro experimental e diminuindo a precisão das estimativas.

Para reduzir o erro experimental e para se obter estimativas mais precisas dos efeitos de tratamentos, são utilizados, em geral, dois tipos de controle das fontes de variação: o controle experimental que se caracteriza principalmente pela escolha de um delineamento apropriado, cuidados na instalação e condução do experimento e o controle estatístico. Quando o delineamento utilizado não consegue fornecer um exato controle sobre as variáveis consideradas independentes, mas que se pode me

dí-las através de observações adicionais, a análise de covariância pode ser empregada e o controle estatístico se verifica a través da adoção das restrições acerca do modelo que serão descritas no item 2.3.

A análise de covariância (ANCOVA) é uma técnica que envolve conjuntamente as características das análises de variância (ANAVA) e regressão. A sua finalidade é utilizar uma ou mais variáveis auxiliares que estão linearmente relacionadas à variável dependente y , para proporcionar ao pesquisador uma melhor interpretação dos resultados experimentais, pois as variáveis auxiliares devem se caracterizar pela função de prever de certa forma, os resultados finais das observações colhidas no experimento. Além disso, a ANCOVA ajuda a fazer o controle local e em alguns casos pode até substituí-lo.

O emprego da análise de covariância se verifica nos mais diversos delineamentos experimentais, contudo, a sua finalidade é obter resultados mais precisos e como consequência, resíduos menores.

A análise de covariância, apesar de proporcionar ao pesquisador a eliminação de uma ou mais fontes de variação, difíceis ou impossíveis de serem eliminadas pelo controle experimental, não deve ser empregada indiscriminadamente, pois, ela é baseada numa série de pressuposições, e a violação de uma de

las pode comprometer a sua validade.

Na análise de covariância linear com uma covariável admite-se, além de outras pressuposições, uma regressão linear comum a todos os tratamentos, para ajustar a variável dependente Y, de interesse. Porém, essa condição nem sempre se verifica, uma vez que é comum se encontrar heterocedasticidade dos coeficientes de regressão intra-tratamentos. Quando isto acontece, este tipo de análise não é aconselhável, pois, a sua aplicação nestes casos leva a que as estimativas dos efeitos de tratamentos sejam afetadas, conforme KIRK(1968).

1.2. Objetivos

Os trabalhos constantes na literatura dão muita ênfase ao caso da inviabilidade da análise de covariância linear quando os coeficientes de regressão linear intra-tratamentos são heterogêneos, o que é comum encontrar na prática. Porém, muito pouco se fala sobre alternativas de análise para contornar este problema. Assim, o objetivo principal deste trabalho é apresentar um método de análise alternativa para a tradicional análise de covariância linear com uma covariável, nos ensaios inteiramente ao acaso com um fator, a qual consiste em estimar um coeficiente de regressão linear específico para cada tratamento. Ademais, pretende-se também, desenvolver uma estrutura teórica baseada na teoria de modelos lineares, com o obje

tivo de fornecer subsídios para que pessoas interessadas no assunto apliquem em outros delineamentos. E, ainda, apresentar terminologias e expressões conhecidas e familiares aos profissionais da área de ciências agrárias, com a finalidade de torná-lhes o trabalho acessível.

2. REVISÃO DE LITERATURA

2.1. Introdução

Segundo COX e McCULLAGH (1982), a análise de co variância foi introduzida por Fisher que juntamente com Eden em 1927 apresentaram a decomposição da soma de produtos e os correspondentes coeficientes de correlação. Por sugestão de Fisher em 1930, Sanders, foi o primeiro a usar a análise de covariância com o objetivo de aperfeiçoar a precisão. Tal procedimento foi apresentado em 1932 por Fisher no seu livro Statistical Methods for Research Workers, no qual aplicou-se o método da análise de variância usual aos dados $y - \hat{\beta}z$, onde y é o valor observado da variável de interesse na análise, z é o valor observado da covariável correspondente e $\hat{\beta}$ é a estimativa do coeficiente de regressão linear calculado via análise de covariância. Esta análise foi admitida apenas para levar a uma aproximação do teste F aplicado aos tratamentos. Posteriormente, em trabalhos publicados pela *Royal Statistical Society* em 1934

(Wishart, 1934; Wilsdon, 1934) sobre a análise de covariância, Fisher contribuiu na discussão de ambos, nos quais, já foi admitida a aplicação do teste F exato. Em 1935, Fisher o apresentou detalhadamente em seu livro *Design of Experiment*. A partir daí, vários outros trabalhos contribuíram para o aperfeiçoamento deste tipo de análise, até que em 1957 a *Biometrics* publicou um volume especial sobre análise de covariância, com trabalhos de Cochran, Smith, Federer, etc., tornando-se certamente numa das mais importantes fontes de pesquisa sobre o assunto.

2.2. Finalidades

Como finalidades da análise de covariância, STEEL e TORRIE (1960), a exemplo de COCHRAN (1957), KOCH *et alii* (1982), dentre outros, destacaram como mais importantes as seguintes:

- a) Ajudar na interpretação dos resultados experimentais, especialmente no que se refere à natureza dos efeitos dos tratamentos;
- b) Dividir a covariância total ou soma de produtos em partes componentes;
- c) Controlar o erro experimental e aumentar a precisão das estimativas;
- d) Ajustar as médias dos tratamentos da variável

dependente pelas diferenças dos correspondentes valores das variáveis independentes;

- e) Estimar observações perdidas. As razões básicas desta aplicação, podem ser vistas em MUNIZ (1982), o qual enfoca alguns casos importantes.

Com relação ao item c), COCHRAN (1957) e KIRK (1968), afirmaram que o ganho em precisão ao aplicar a análise de covariância linear depende principalmente do coeficiente de correlação ρ entre z e y nas unidades experimentais que recebem o mesmo tratamento. Se σ_y^2 é a variância do erro experimental da população quando é empregada a análise de variância, o ajustamento através da análise de covariância reduz este erro para $\sigma_y^2(1 - \rho^2)[1 + 1/(fe - 2)]$ aproximadamente, onde: fe é o número de graus de liberdade do resíduo.

2.3. Restrições ao Uso da Análise de Covariância

Segundo HENDERSON (1982), se faz necessário considerar as seguintes pressuposições a respeito do modelo para a análise de covariância:

- a) As pressuposições usuais acerca da análise de variância prevalecem, ou sejam: o modelo é linear nos parâmetros; os erros por hipótese, devem ser independentes e normalmente distribuídas com média zero e variância comum, isto

$$\varepsilon_{ij} \stackrel{\text{IID}}{\sim} N(0, \sigma^2);$$

- b) Os valores das covariáveis são fixados em a mostras aleatórias hipotéticas (covariantes com efeitos aleatórios podem resultar de erro de medida, equações simultâneas ou outros tipos de modelo com variáveis "Latents".);
- c) No modelo usual de covariância, assumem-se que os níveis dos fatores e interações, como também, coeficientes de regressão covariantes são fixados, embora um modelo misto com certos níveis aleatórios, em muitos casos seja mais realístico;
- d) O verdadeiro modelo é conhecido; isto é, todas as variáveis do modelo podem ser completamente especificadas, e as pressuposições distribucionais acerca dos parâmetros no modelo possam ser assumidas, como nos ítems anteriores. Esta pressuposição é necessária para avaliar as propriedades dos estimadores e testes que são usados;
- e) Assume-se também, que para examinar o efeito de um fator, as covariantes não sejam afetadas por ele. Pois, se for usado um modelo no qual as covariantes são afetadas pelos tratamentos, pode se verificar que parte dos efeitos dos tratamentos é devida às covariáveis;

- f) Os coeficientes de regressão covariantes são estimáveis. Isto exige que as colunas da matriz do delineamento das covariantes sejam linearmente independentes das colunas da matriz do delineamento dos fatores fixados e interações, como também sejam linearmente independentes uma da outra;
- g) Assume-se também, na análise tradicional, que as regressões covariantes sejam homogêneas para todos os níveis dos fatores e interações. Esta pressuposição é também restritiva;
- h) A menos que seja feita uma seleção de viés nos elementos aleatórios do modelo, (isto é: seleção na variável independente, no erro ou nos efeitos aleatórios), os estimadores na análise de covariância são viciados. A seleção baseada em valores covariantes fixados, não resulta em viés. Se, entretanto, as covariantes não forem fixadas, mas se houver seleção nas mesmas, método que não considerem esta seleção darão estimadores viciados.

Com relação ao item g), ROBSON e ATKINSON (1960), NETER e WASSERMAN (1974), dentre outros, consideraram que a homogeneidade dos coeficientes de regressão intra-tratamentos na análise tradicional é uma condição estritamente necessária para a validade dos ajustes covariantes. Estes ajus

tes requerem que as funções de regressão sejam diferentes . apenas quanto aos interceptos, ou termo de grau zero, e, comparações entre médias ajustadas dos tratamentos sejam feitas entre interceptos. Quando os tratamentos e o fator representado pela covariável interagem produzindo seus efeitos, resulta em heterogeneidade dos coeficientes de regressão e os efeitos dos tratamentos calculados de maneira usual não estimam os verdadeiros efeitos.

2.4. Teste de Homogeneidade dos Coeficientes de Regressão Linear

Para testar a hipótese de homogeneidade dos coeficientes de regressão linear, isto é, $H_0: \beta_1 = \beta_2 = \dots = \beta_v$ vs H_A : não H_0 , nos ensaios inteiramente ao acaso com um fator, STEEL e TORRIE (1960) e KIRK (1968) propuseram o seguinte teste:

$$F = \frac{S_2 / (v-1)}{S_1 / v(r-2)},$$

o qual tem distribuição de F com (v-1) e v(r-2) graus de liberdade. Segundo KIRK (1968), as razões básicas para este teste foram discutidas por Kendall (1948, 237-242), o qual recomendou um nível de significância elevado ($\alpha = 0,10$ ou $\alpha = 0,25$), a fim de evitar o erro do tipo II, que é aceitar a hipótese da homogeneidade dos coeficientes de regressão quando, de fato, ela é falsa.

As estatísticas S_1 e S_2 são calculadas da seguinte maneira:

$$S_1 = S_{YY} - \sum_i \frac{(S_{zyi})^2}{S_{zzi}};$$

Onde:

$$S_{YY} = \sum_{ij} (y_{ij} - \bar{y}_{i.})^2, \quad S_{zy1} = \sum_j (z_{ij} - \bar{z}_{i.})(y_{ij} - \bar{y}_{i.}) \text{ e } S_{zzi} = \sum_j (z_{ij} - \bar{z}_{i.})^2;$$

e

$$S_2 = \sum_i \frac{(S_{zyi})^2}{S_{zzi}} - \frac{(S_{zy})^2}{S_{zz}},$$

onde:

$$S_{zy} = \sum_{ij} (z_{ij} - \bar{z}_{i.})(y_{ij} - \bar{y}_{i.}) \text{ e } S_{zz} = \sum_{ij} (z_{ij} - \bar{z}_{i.})^2;$$

e ainda, tem-se que v é o número de tratamentos e r é o número de repetições.

Neste caso, S_1 são as variações dos dados individuais em torno da regressão linear dentro de cada tratamento e S_2 corresponde às variações dos v coeficientes de regressão dentro de cada tratamento, em torno do coeficiente conjunto de regressão intra-tratamentos, que é $\sum_i (\beta_i - \beta)^2$. Quanto maior esta causa de variação em relação a S_1 , menos sustentável é a hipótese de homogeneidade dos coeficientes de regressão linear.

SEARLE (1971), apresentou uma maneira de se testar a hipótese de homogeneidade de regressão linear que pode ser empregada em qualquer delineamento experimental. Tal procedimento, consiste em calcular a estatística,

$$F_C = \frac{[R(\beta_i/\mu, \alpha) - R(\beta/\mu, \alpha)]/(v-1)}{QM \text{ Resíduo}},$$

onde F_C tem distribuição de F com $(v-1)$ e $(N-2v)$ graus de liberdade; $R(\beta_i/\mu, \alpha)$, representa a soma dos quadrados de regressão linear dentro dos tratamentos e $R(\beta/\mu, \alpha)$, a soma de quadrados de regressão intra-tratamentos (quando se considera uma única regressão para o ajuste), ambas corrigidas para média, fatores e interações, se houver. Ainda uma outra forma para determinação da estatística F , pode ser vista em SEARLE (1979), baseada em teste de hipótese linear geral, isto é, para testar a hipótese $H_0: K'\beta = \emptyset$, calculou-se,

$$F = \{K'\hat{\beta}\}'[K'(R'R)^{-1}K]^{-1}(K'\hat{\beta})/r(K')QM \text{ Res.},$$

onde K' é uma matriz de posto linha completo.

2.5. Formulação Geral da Análise de Covariância

A forma de apresentação do modelo linear para a análise de covariância linear com uma covariável, encontrada na literatura, para o delineamento a que se propõe este trabalho é igual àquela sugerida por SNEDECOR e COCHRAN (1967), SEARLE (1971, 1979), URQUHART (1982), dentre outros. Assim, para os ensaios inteiramente casualizados tem-se:

$$y_{ij} = \mu + t_i + \beta z_{ij} + e_{ij}, \quad (2.5.1)$$

onde:

y_{ij} é a j -ésima observação, da variável dependente, no tratamento i , ($i = 1, 2, \dots, v$; $j = 1, 2, \dots, n_i$);

μ é a média geral;

t_i é o efeito do tratamento i ;

β é o coeficiente de regressão linear comum a todos os tratamentos;

z_{ij} é a j -ésima observação da covariável no i -ésimo tratamento, fixa e tomada sem erro;

e_{ij} é o erro experimental, normalmente distribuído com média zero e variância σ^2 , isto é, $e_{ij} \stackrel{\text{IID}}{\sim} N(0, \sigma^2)$.

Sob forma matricial este modelo é encontrado na literatura como:

$$Y = X\gamma + \xi, \quad (2.5.2)$$

onde:

Y é o vetor de $N = \sum_i n_i$ realizações de variáveis aleatórias;

X é a matriz característica do delineamento e das observações covariantes;

γ é o vetor dos parâmetros envolvidos no modelo;

ξ é um vetor de erros aleatórios com $\xi \sim N(\emptyset, I\sigma^2)$.

Uma outra forma de apresentação matricial encontrada mais frequentemente na literatura, conforme SEARLE (1971, 1979), DIAS (1981), HENDERSON (1982), COX e McCULLAGH (1982), dentre outros, é:

$$Y = X\theta + Z\beta + \xi, \quad (2.5.3)$$

onde:

Y é o vetor de N realizações de variáveis aleatórias;

X é a matriz que caracteriza o delineamento e θ é o vetor de parâmetros correspondentes;

Z é a matriz das observações covariantes e β é o vetor dos parâmetros da regressão covariante;

ξ é o vetor de erros aleatórios com distribuição normal de média zero e variância $I\sigma^2$, isto é, $\xi \sim N(0, I\sigma^2)$.

A formulação geral para análise de covariância tradicional, discutida por SEARLE (1971) de acordo com o modelo (2.5.3), considerou o seguinte sistema de equações normais:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} \theta^0 \\ \beta^0 \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}, \quad (2.5.4)$$

e a solução de mínimos quadrados para o sistema é:

$$\theta^0 = (X'X)^{-1}X'(Y - Z\beta^0); \quad (2.5.5)$$

e

$$\beta^0 = \hat{\beta} = (R'R)^{-1}R'Y. \quad (2.5.6)$$

onde:

$R = [I - X(X'X)^{-1}X']Z$; $(X'X)^{-1}$ é uma inversa generalizada de $X'X$;

$R'R$ é uma matriz constituída da soma dos quadrados e produtos dos resíduos relativos às covariantes;

$R'Y$ é um vetor constituído da soma dos produtos, das covariantes pela variável dependente, relativos aos resíduos.

Para análise de variância, o autor considerou dois quadros de análise, a saber:

QUADRO 1. Análise de variância para o ajuste covariante (β) a pós fatores e interações (θ) no modelo $Y = X\theta + Z\beta + \xi$

C.VARIAÇÃO	G.L.	SQ
Fatores e Interações	$r(X)$	$R(\theta) = Y'[X(X'X)^{-1}X']Y$
Média	1	$R(\mu) = N\bar{Y}^2$
Fatores e Interações (após média)	$r(X) - 1$	$R(\theta/\mu) = R(\theta) - R(\mu)$
Covariantes (após fatores e interações)	$r(Z)$	$R(\beta/\theta) = Y'[R(R'R)^{-1}R']Y$
Resíduo	$N - r(X) - r(Z)$	$SEE = Y'Y - R(\theta) - R(\beta/\theta)$
TOTAL	N	$Y'Y$

QUADRO 2. Análise de variância para o ajustamento de fatores e interação após covariantes (β) no modelo $Y = X\theta + Z\beta + \xi$

C.VARIAÇÃO	G.L.	SQ
Média	1	$R(\mu) = N\bar{Y}^2_{..}$
Covariantes (após média)	$r(Z)$	$R(\beta/\mu) = Y'[Z(Z'Z)^{-1}Z']Y$
Fatores e Interações (após média e covariantes)	$r(X) - 1$	$R(\theta/\mu, \beta) = R(\theta/\mu) + R(\beta/\theta) - R(\beta/\mu)$
Resíduo	$N - r(X) - r(Z)$	$Y'Y - R(\theta) - R(\beta/\theta)$
TOTAL	N	$Y'Y$

onde:

Z é uma matriz constituída pelos desvios das covariantes em relação às suas médias gerais, isto é, $z = Z - 1_{(N)}\bar{Z}_{..}$.

2.6. Um Modelo Alternativo

Para análise de covariância tradicional, assume-se que os coeficientes de regressão linear dentro dos tratamentos são homogêneos. Porém quando esta hipótese é rejeitada, a análise tradicional não é aconselhável. Portanto, uma maneira de contornar este problema será, sem dúvida, considerar um modelo no qual esta hipótese não seja exigida. Neste caso, deve-se considerar uma regressão específica para cada tratamento. Assim,

conforme SEARLE (1971, 1979) o modelo estatístico para um ensaio inteiramente ao acaso com um fator fica:

$$Y_{ij} = \mu + t_i + \beta_i z_{ij} + e_{ij} \quad (2.6.1)$$

onde:

β_i é o coeficiente de regressão linear dentro do tratamento i e os demais termos do modelo conforme definidos em (2.5.1).

Segundo COX (1958), este tipo de procedimento em pregado num experimento em quadrado latino, onde a resposta do material experimental é uma função contínua de duas variáveis - linhas e colunas - o modelo matemático poderá ser considerado como um modelo que representa uma superfície de respostas, ima ginadas como um conjunto de curvas paralelas para as linhas e um outro conjunto de curvas paralelas para as colunas. A pres suposição mais geral - que no mínimo uma das curvas não seja pa ralela - poderá ser mais realística em algumas circunstâncias, desde que se obtenha maior precisão estatística através de uma análise apropriada. Em razão disso, o autor propôs um tipo de análise na qual se considera uma curvatura individual em uma das direções, cujo modelo estatístico é dado por $Y_{ij} = \mu + k_j + \beta_j a_{ij} + \theta_j + e_{ij}$, onde y_{ij} ($i = j = 1, 2, \dots, n$) é a resposta observada da unidade experimental correspondente ao i -ésimo tratamento na j -ésima coluna; μ é a média geral; k_j e β_j são respectivamente o efeito da j -ésima coluna e o coeficienu

de regressão correspondente; θ_i é o efeito do tratamento i ;
 a_{ij} é o fator fixado da linha para o i -ésimo tratamento na j -ésima coluna. Tomou-se $\sum_j a_{ij}^2 = \sum_i a_{ij}^2 = 1$; $e_{ij} \stackrel{\text{IID}}{\sim} N(0, \sigma^2)$ e as restrições $\sum_j k_j = \sum_i \theta_i = 0$.

O sistema de equações normais, segundo o autor foi:

$$\begin{aligned} n^2 \hat{\mu} &= \text{total geral} = Y_{..} \\ n(\hat{\mu} + \hat{k}_j) &= \text{total da coluna} = K_j \\ \hat{\beta}_j + \sum_i a_{ij} \hat{\theta}_i &= \sum_i a_{ij} Y_{ij} \\ n\hat{\mu} + \sum_j a_{ij} \hat{\beta}_j + n\hat{\theta}_i &= \text{total do trat. } i = T_i \end{aligned}$$

As estimativas dos parâmetros foram:

$$\hat{\mu} = Y_{..}/n^2 \quad \text{e} \quad \hat{k}_j = \frac{1}{n}(K_j - Y_{..}/n^2) = \frac{1}{n}(K_j - \hat{\mu})$$

Sob forma matricial o sistema de equações normais foi:

$$\begin{aligned} \hat{\beta} + A' \hat{\theta} &= R \\ A \hat{\beta} + n \hat{\theta} &= T \end{aligned}$$

onde:

A é a matriz formada pelos elementos a_{ij} , $A = \{a_{ij}\}$;

R é uma matriz coluna de elementos $\sum_i a_{ij} Y_{ij}$;

T é uma matriz coluna de elementos $(T_i - \bar{T})$, assim,

$$\hat{\theta} = [nI - AA']^{-1}[T - AR],$$

e

$$\hat{\beta} = [nI - A'A]^{-1}[nR - A'T]$$

Para a análise de variância, o autor considerou o seguinte quadro:

QUADRO 3. Análise de variância para o ajuste de tratamentos corrigidos pela média e pelas covariantes no modelo de ensaios em quadrados latinos.

C.VARIAÇÃO	G.L	SQ
Colunas	n-1	$\frac{1}{n} \sum_i K_i^2 - Y_{..}^2/n^2$
Reg. ignorando tratamentos	n	$[R'R]$
Tratamentos (aj.) p/regressão	n-1	$\hat{\theta}'[T - AR]$
Resíduo	$n^2 - 3n + 1$	Diferença
TOTAL	$n^2 - 1$	$Y'Y - \frac{Y_{..}^2}{n^2}$

SEARLE (1971, 1979), apresentou uma análise para estes tipos de modelos, nos quais se considera o caso dos ensaios com um fator e com dois fatores com interações. O mesmo

afirmou que o procedimento para análise do modelo citado em (2.6.1), é o mesmo empregado na formulação geral da análise de covariância, apenas com algumas modificações, como por exemplo, a forma de apresentação da matriz das covariantes.

Com base nos trabalhos citados, COX (1958), SEARLE (1971, 1979) e IEMMA (1985), principalmente no que se refere a aplicação de modelos lineares, fundamentou-se a estrutura básica para o desenvolvimento deste trabalho.

3. DESENVOLVIMENTO TEÓRICO

Este trabalho, considera o caso da análise de co variância com uma covariável e uma regressão linear específica para cada tratamento, nos ensaios inteiramente casualizados.

3.1. Modelo Estatístico

Para este caso o modelo considerado é:

$$y_{ij} = \mu + t_i + \beta_i z_{ij} + e_{ij} \quad (3.1.1)$$

onde:

y_{ij} é a j -ésima observação do tratamento i ; ($i = 1, 2, \dots, v$) e
($j = 1, 2, \dots, n_i$);

μ é uma constante inerente à todas as observações;

t_i é o efeito do i -ésimo tratamento;

β_i é o coeficiente de regressão linear dentro do tratamento i ;

z_{ij} é a j -ésima observação covariante correspondente ao trata

mento i , fixa e tomada sem erro;

e_{ij} é o erro casual atribuído a observação y_{ij} , com distribuição normal de média zero e variância σ^2 , isto é, $e_{ij} \stackrel{\text{IID}}{\sim} N(0, \sigma^2)$.

Na forma matricial considerou-se o seguinte modelo:

$$Y = X\theta + Z\beta + \xi \quad (3.1.2)$$

onde:

Y é um vetor de $N = \sum_i n_i$ realizações de variáveis aleatórias;

X é uma matriz de elementos conhecidos que caracteriza o delineamento, de rank v , isto é, $r(X) = v$;

θ é um vetor de $v+1$ parâmetros desconhecidos, correspondente a constante μ e aos efeitos dos tratamentos t_i ;

Z é uma matriz de dimensões $N \times v$ de observações covariantes, consideradas fixas e tomadas sem erro;

β é um vetor de v parâmetros desconhecidos, correspondentes aos coeficientes de regressão dentro de cada tratamento;

ξ é um vetor de dimensões $N \times 1$, de componentes não observáveis e aleatórios, tal que $\xi \sim N(\emptyset; I\sigma^2)$.

Assim, pode-se estruturar:

$$\begin{bmatrix} y_{11} \\ y_{12} \\ \vdots \\ y_{1n_1} \\ y_{21} \\ y_{22} \\ \vdots \\ y_{2n_2} \\ \vdots \\ y_{v1} \\ y_{v2} \\ \vdots \\ y_{vn_v} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \\ 1 & 0 & 0 & \dots & 1 \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ \vdots \\ t_v \\ \vdots \\ t_v \end{bmatrix} + \begin{bmatrix} z_{11} & 0 & \dots & 0 \\ z_{12} & 0 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ z_{1n_1} & 0 & \dots & 0 \\ 0 & z_{21} & \dots & 0 \\ 0 & z_{22} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & z_{2n_2} & \dots & 0 \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & z_{v1} \\ 0 & 0 & \dots & z_{v2} \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & z_{vn_v} \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_v \end{bmatrix} + \begin{bmatrix} e_{11} \\ e_{12} \\ \vdots \\ e_{1n_1} \\ e_{21} \\ e_{22} \\ \vdots \\ e_{2n_2} \\ \vdots \\ e_{v1} \\ e_{v2} \\ \vdots \\ e_{vn_v} \end{bmatrix} \quad (3.1.3)$$

$N \quad 1 \quad N \quad \quad \quad v+1 \quad \quad N \quad \quad \quad v \quad \quad \quad N \quad 1$

Além disso, o modelo assume as pressuposições descritas em 2.3. com exceção da homocedasticidade dos coeficientes de regressão.

Nestas circunstâncias, tem-se:

$$E(Y) = E(X\theta + Z\beta + \xi) = X\theta + Z\beta \quad (3.1.4)$$

e

$$\text{Var}(Y) = \text{Var}(X\theta + Z\beta + \xi) = I\sigma^2 \quad (3.1.5)$$

então,

$$Y \sim N(X\theta + Z\beta; I\sigma^2) \quad (3.1.6)$$

3.2. Sistema de Equações Normais e Estimativas dos Parâmetros

O sistema de equações normais para o modelo ir restrito é:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z \end{bmatrix} \begin{bmatrix} \theta^0 \\ \beta^0 \end{bmatrix} = \begin{bmatrix} X'Y \\ Z'Y \end{bmatrix}$$

ou seja,

$$\begin{cases} X'X\theta^0 + X'Z\beta^0 = X'Y \\ Z'X\theta^0 + Z'Z\beta^0 = Z'Y \end{cases} \quad (3.2.1)$$

$$(3.2.2)$$

uma solução de mínimos quadrados, é obtida do seguinte modo:

Pré-multiplica-se (3.2.1) por $Z'X^+$, onde X^+ é a inversa generalizada de MOORE-PENROSE de X , obtendo-se:

$$Z'X^+X'X\theta^0 + Z'X^+X'Z\beta^0 = Z'X^+X'Y,$$

e desde que $X^+X'X = X$, então tem-se:

$$Z'X\theta^0 + Z'PZ\beta^0 = Z'PY, \quad (3.2.3)$$

onde $P = XX^+ = X^+X' = X(X'X)^{-1}X'$ é o projetor ortogonal de Y no espaço coluna de X , $C(X)$.

Subtraindo (3.2.3.) de (3.2.2), vem:

$$Z'(I - P)Z\beta^0 = Z'(I - P)Y, \quad (3.2.4)$$

que é o sistema de equações normais reduzidas para efeito de regressão, eliminados os efeitos de tratamentos e da média.

Sabe-se que $(I - P)$ é simétrica e idempotente, isto é, $(I - P) = (I - P)' = (I - P)'(I - P)$, então,

$$Z'(I - P)'(I - P)Z\beta^0 = Z'(I - P)'Y$$

e fazendo-se $R = (I - P)Z$, obtem-se:

$$R'R\beta^0 = R'Y \quad (3.2.5)$$

Como $R = (I - P)Z = [I - X(X'X)^{-1}X']Z$ é de posto coluna completo,

então existe $(R'R)^{-1}$, que ao pré-multiplicar a equação (3.2.5) resulta:

$$\beta^0 = \hat{\beta} = (R'R)^{-1}R'Y \quad (3.2.6)$$

tem-se ainda, que $(X'X)^{-}$ é uma inversa condicional de $X'X$, e para este caso tomou-se,

$$(X'X)^{-} = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 0 & 1/n_1 & 0 & \dots & 0 \\ 0 & 0 & 1/n_2 & \dots & 0 \\ \vdots & \vdots & \vdots & & \vdots \\ 0 & 0 & 0 & \dots & 1/n_V \end{bmatrix} \quad (3.2.7)$$

obtendo-se desta forma:

$$R'R = \begin{bmatrix} \sum_j (z_{1j} - \bar{z}_{1.})^2 & 0 & \dots & 0 \\ 0 & \sum_j (z_{2j} - \bar{z}_{2.})^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & \sum_j (z_{Vj} - \bar{z}_{V.})^2 \end{bmatrix} \quad (3.2.8)$$

$$R'Y = \begin{bmatrix} \sum_j (y_{1j} - \bar{y}_{1.})(z_{1j} - \bar{z}_{1.}) \\ \sum_j (y_{2j} - \bar{y}_{2.})(z_{2j} - \bar{z}_{2.}) \\ \vdots \\ \sum_j (y_{vj} - \bar{y}_{v.})(z_{vj} - \bar{z}_{v.}) \end{bmatrix}, \quad (3.2.9)$$

consequentemente,

$$\beta^0 = \hat{\beta} = \begin{bmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \vdots \\ \hat{\beta}_v \end{bmatrix} = \begin{bmatrix} \frac{\sum_j (y_{1j} - \bar{y}_{1.})(z_{1j} - \bar{z}_{1.})}{\sum_j (z_{1j} - \bar{z}_{1.})^2} \\ \frac{\sum_j (y_{2j} - \bar{y}_{2.})(z_{2j} - \bar{z}_{2.})}{\sum_j (z_{2j} - \bar{z}_{2.})^2} \\ \vdots \\ \frac{\sum_j (y_{vj} - \bar{y}_{v.})(z_{vj} - \bar{z}_{v.})}{\sum_j (z_{vj} - \bar{z}_{v.})^2} \end{bmatrix} \quad (3.2.10)$$

por outro lado, uma solução de mínimos quadrados para θ obtida a partir de (3.2.1), é:

$$\theta^0 = (X'X)^{-1}X'(Y - Z\hat{\beta}) \quad (3.2.11)$$

Desenvolvendo-se (3.2.11) chega-se às seguintes estimativas:

$$\theta^0 = \begin{bmatrix} \mu^0 \\ t^0_1 \\ t^0_2 \\ \vdots \\ t^0_v \end{bmatrix} = \begin{bmatrix} 0 \\ \bar{y}_{1.} - \hat{\beta}_1 \bar{z}_{1.} \\ \bar{y}_{2.} - \hat{\beta}_2 \bar{z}_{2.} \\ \vdots \\ \bar{y}_{v.} - \hat{\beta}_v \bar{z}_{v.} \end{bmatrix} \quad (3.2.12)$$

e então, $\mu^0 = 0$ e $t^0_i = \bar{y}_{i.} - \hat{\beta}_i \bar{z}_{i.}$, que corresponde a média do tratamento i ajustada para a regressão correspondente.

3.3. Média e Variância dos Estimadores dos Parâmetros

Comprova-se que $\hat{\beta}$ é um estimador imparcial para β . De fato, considerando-se;

$$\hat{\beta} = (R'R)^{-1}R'Y$$

e substituindo-se (3.1.2.) na expressão, vem:

$$\begin{aligned} \hat{\beta} &= (R'R)^{-1}R'(X\theta + Z\beta + \xi) \\ &= (R'R)^{-1}R'X\theta + (R'R)^{-1}R'Z\beta + (R'R)^{-1}R'\xi \\ &= \quad \quad \quad \emptyset \quad + (R'R)^{-1}Z'(I - P)'Z\beta + (R'R)^{-1}R'\xi \\ &= (R'R)^{-1}R'R\beta + (R'R)^{-1}R'\xi \end{aligned}$$

logo,

$$\hat{\beta} = \beta + (R'R)^{-1}R'\xi$$

ou

$$\hat{\beta} - \beta = (R'R)^{-1}R'\xi \quad (3.3.1)$$

Assim,

$$E(\hat{\beta} - \beta) = (R'R)^{-1}R'E(\xi) = 0$$

e conseqüentemente,

$$E(\hat{\beta}) = \beta \quad (3.3.2)$$

A matriz de dispersão do vetor $\hat{\beta}$, é por definição:

$$D(\hat{\beta}) = E\{[\hat{\beta} - E(\hat{\beta})][\hat{\beta} - E(\hat{\beta})]'\} \quad (3.3.3)$$

levando em conta (3.3.1) e (3.3.2), tem-se:

$$\begin{aligned} D(\hat{\beta}) &= E\{[(R'R)^{-1}R'\xi][(R'R)^{-1}R'\xi]'\} \\ &= E\{(R'R)^{-1}R'\xi\xi'R(R'R)^{-1}\} \\ &= (R'R)^{-1}R'R(R'R)^{-1}I\sigma^2 \end{aligned}$$

ou ainda,

$$D(\hat{\beta}) = (R'R)^{-1}\sigma^2 \quad (3.3.4)$$

ou de acordo com (3.2.8),

$$D(\hat{\beta}) = \begin{bmatrix} 1/\sum_j (z_{1j} - \bar{z}_{1.})^2 & 0 & \dots & 0 \\ 0 & 1/\sum_j (z_{2j} - \bar{z}_{2.})^2 & \dots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \dots & 1/\sum_j (z_{vj} - \bar{z}_{v.})^2 \end{bmatrix} \sigma^2 \quad (3.3.5)$$

ou ainda, $\text{var}(\hat{\beta}_i) = \sigma^2 / \sum_j (z_{ij} - \bar{z}_{i.})^2$; $\text{cov}(\hat{\beta}_i, \hat{\beta}_{i'}) = 0$ para $i \neq i'$.

Como $\hat{\beta}$ é uma combinação linear dos y 's, cuja distribuição de probabilidade é normal, e sabendo-se que combinação linear de variáveis normais tem distribuição normal, então,

$$\hat{\beta} \sim N[\beta; (R'R)^{-1}\sigma^2] \quad (3.3.6)$$

Por outro lado, mostra-se que θ^0 não é um estimador imparcial para θ , mas sim, para combinações lineares dos parâmetros que estão em θ , ou seja,

substituindo-se (3.1.2) em (3.2.11), tem-se:

$$\begin{aligned}
\theta^0 &= (X'X)^{-1}X'[X\theta + X\beta + \xi - Z(R'R)^{-1}R'(X\theta + Z\beta + \xi)] \\
&= (X'X)^{-1}X'X\theta + (X'X)^{-1}X'Z\beta + (X'X)^{-1}X'\xi - (X'X)^{-1}X'Z(R'R)^{-1}R'X\theta - \\
&\quad - (X'X)^{-1}X'Z(R'R)^{-1}R'Z\beta - (X'X)^{-1}X'Z(R'R)^{-1}R'\xi \\
&= (X'X)^{-1}X'X\theta + (X'X)^{-1}X'Z\beta + (X'X)^{-1}X'\xi - \emptyset - (X'X)^{-1}X'Z(R'R)^{-1}Z'(I-P)Z\beta - \\
&\quad - (X'X)^{-1}X'Z(R'R)^{-1}R'\xi \\
&= (X'X)^{-1}X'X\theta + (X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi
\end{aligned}$$

Portanto,

$$\theta^0 = H\theta + (X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \quad (3.3.7)$$

ou

$$\theta^0 - H\theta = (X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \quad (3.3.8)$$

Assim,

$$E(\theta^0 - H\theta) = \emptyset \text{ e conseqüentemente,}$$

$$E(\theta^0) = H\theta \quad (3.3.9)$$

Onde H, é uma forma escalonada de $X'X$. E, para a condicional escolhida em (3.2.7), tem-se:

$$E(\theta^0) = H\theta = \begin{bmatrix} 0 & 0 & 0 & \dots & 0 \\ 1 & 1 & 0 & \dots & 0 \\ 1 & 0 & 1 & \dots & 0 \\ \dots & \dots & \dots & \dots & \dots \\ 1 & 0 & 0 & \dots & 1 \end{bmatrix} \begin{bmatrix} \mu \\ t_1 \\ t_2 \\ \vdots \\ t_v \end{bmatrix} = \begin{bmatrix} 0 \\ \mu + t_1 \\ \mu + t_2 \\ \vdots \\ \mu + t_v \end{bmatrix} \quad (3.3.10)$$

o que confirma a não imparcialidade de θ^0 descrita anteriormente.

Com isto, percebe-se que não faz sentido, no modelo irrestrito, calcular uma matriz de variâncias e covariâncias de θ^0 , mas sim, para funções lineares estimáveis, conforme será exposto a seguir.

3.4. Estimabilidade

Considerando-se os valores esperados de $\hat{\beta}$ em (3.3.2) e de θ^0 em (3.3.9), verifica-se que $\hat{\beta}$ é sempre estimável, e, para uma matriz C' de *rank* linha completo, a função paramétrica $C'\theta$ é estimável sempre que ela for estimável no modelo sem covariantes, isto é, para que $C'\theta$ seja estimável, deve existir uma matriz T , tal que $C' = T'X$.

Além disso, pelo teorema de Gauss-Markov, sabe-se que se $C'\theta$ é estimável, então $\widehat{C'\theta} = C'\theta^0$, de modo único, pa

ra qualquer θ^0 solução das equações normais. Assim, de (3.3.7) tem-se que;

$$\theta^0 = (X'X)^{-1}X'X\theta + (X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi$$

e

$$\begin{aligned} C'\theta^0 &= C'(X'X)^{-1}X'X\theta + C'(X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \\ &= T'X(X'X)^{-1}X'X\theta + C'(X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \\ &= T'X\theta + C(X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \end{aligned}$$

Portanto,

$$C'\theta^0 = C'\theta + C'(X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \quad (3.4.1)$$

ou

$$C'\theta^0 - C'\theta = C'(X'X)^{-1}X'[I - Z(R'R)^{-1}R']\xi \quad (3.4.2)$$

E, considerando-se as pressuposições do modelo conclui-se que,

$$E(C'\theta^0) = C'\theta$$

Por definição, sabe-se que

$$\text{Var}(C'\theta^0) = E\{[C'\theta^0 - E(C'\theta^0)][C'\theta^0 - E(C'\theta^0)]'\} \quad (3.4.3)$$

Substituindo-se (3.4.2) em (3.4.3.), tem-se:

$$\begin{aligned}
 \text{Var}(C'\theta^0) &= E\{C'(X'X)^{-1}X'[I - Z(R'R)^{-1}R'] \xi\xi' [I - R(R'R)^{-1}Z']X(X'X)^{-1}C\} \\
 &= C'\{(X'X)^{-1}X'[I - Z(R'R)^{-1}R'] [I - R(R'R)^{-1}Z']X(X'X)^{-1}\}C\sigma^2 \\
 &= C'\{(X'X)^{-1}X'X(X'X)^{-1} - (X'X)^{-1}X'R(R'R)^{-1}Z'X(X'X)^{-1} - \\
 &\quad - (X'X)^{-1}X'Z(R'R)^{-1}R'X(X'X)^{-1} + \\
 &\quad + (X'X)^{-1}X'Z(R'R)^{-1}R'R(R'R)^{-1}Z'X(X'X)^{-1}\}C\sigma^2 \\
 &= C'\{(X'X)^{-1}X'X(X'X)^{-1} + (X'X)^{-1}X'Z(R'R)^{-1}Z'X(X'X)^{-1}\}C\sigma^2 \\
 &= \{T'X(X'X)^{-1}X'X(X'X)^{-1}X'T - C'(X'X)^{-1}X'Z(R'R)^{-1}Z'X(X'X)^{-1}C\}\sigma^2 \\
 &= \{T'X(X'X)^{-1}X'T + C'(X'X)^{-1}X'Z(R'R)^{-1}Z'X(X'X)^{-1}C\}C\sigma^2
 \end{aligned}$$

Dessa forma, obtem-se,

$$\text{Var}(C'\theta^0) = C'\{(X'X)^{-1} + (X'X)^{-1}X'Z(R'R)^{-1}Z'X(X'X)^{-1}\}C\sigma^2 \quad (3.4.4)$$

Assim,

$$\widehat{C'\theta} = C'\theta^0 \sim N\{C'\theta; C'[(X'X)^{-1} + (X'X)^{-1}X'Z(R'R)^{-1}Z'X(X'X)^{-1}]C\sigma^2\} \quad (3.4.5)$$

As funções paramétricas do tipo $C'\theta = \sum_i C_i \theta_i$, com $\sum_i C_i = 0$, conhecidas como contrastes entre os efeitos ajustados de tratamentos, formam um subconjunto das funções estimáveis, e, por isso, serão aqui enfatizadas por serem de grande interesse prático.

Dado o contraste;

$$y_h = C_h'\theta,$$

sua estimativa é dada por:

$$\hat{y}_h = \hat{C}_h' \hat{\theta} = C_h' \theta^0 = \begin{bmatrix} 0 & C_{h1} & C_{h2} & \dots & C_{hv} \end{bmatrix} \begin{bmatrix} 0 \\ \bar{y}_{1.} - \hat{\beta}_1 \bar{z}_{1.} \\ \bar{y}_{2.} - \hat{\beta}_2 \bar{z}_{2.} \\ \vdots \\ \bar{y}_{v.} - \hat{\beta}_v \bar{z}_{v.} \end{bmatrix}$$

Logo,

$$\hat{y}_h = \sum_{i=1}^v C_{hi} t_i^0 = \sum_{i=1}^v C_{hi} (\bar{y}_{i.} - \hat{\beta}_i \bar{z}_{i.}) \quad (3.4.6)$$

que corresponde à estimativa do contraste entre as médias ajustadas dos tratamentos.

A variância de $\hat{y}_h$ é:

$$\text{Var}(\hat{y}_h) = \text{Var}(C_h' \theta^0) = \{C_h' [(X'X)^{-1} + (X'X)^{-1} X'Z(R'R)^{-1} Z'X(X'X)^{-1}] C_h\} \sigma^2,$$

desenvolvendo-se os cálculos obtém-se a seguinte expressão:

$$\text{Var}(\hat{y}_h) = \text{Var}(C_h' \theta^0) = \sum_{i=1}^v C_{hi}^2 \left[\frac{1}{n_i} + \frac{\bar{z}_{i.}^2}{\sum_j (z_{ij} - \bar{z}_{i.})^2} \right] \sigma^2 \quad (3.4.7)$$

Um outro resultado interessante é aquele relacio

nado com os coeficientes de regressão, isto é, sabe-se que β é sempre estimável, e, para uma matriz C' de *rank* linha completo, as combinações lineares do tipo $C'\beta$, também são estimáveis.

Portanto, pelo teorema de *Gauss-Markov*, tem-se que;

$$\widehat{C'\beta} = C'\hat{\beta},$$

e

$$E(C'\hat{\beta}) = C'E(\hat{\beta}) = C'\beta \quad (3.4.8)$$

Como também,

$$\text{Var}(C'\hat{\beta}) = C'\text{Var}(\hat{\beta})C = C'(R'R)^{-1}C\sigma^2. \quad (3.4.9)$$

Assim,

$$C'\hat{\beta} \sim N[C'\beta; C'(R'R)^{-1}C\sigma^2] \quad (3.4.10)$$

Estes resultados são de fundamental importância na aplicação do teste de hipóteses relacionado à homocedasticidade de dos coeficientes de regressão.

3.5. Análise de Variância e Teste de Hipóteses

Sabe-se que Y pode ser decomposto em dois veto

res de espaços ortogonais, $Y = \hat{Y} + \hat{\xi}$, e, tomando suas normas pode-se interpretar geometricamente esta partição através do teorema de Pitágoras:

$$\|Y\|^2 = \|\hat{Y}\|^2 + \|\hat{\xi}\|^2 = \|X\theta^0 + Z\hat{\beta}\|^2 + \|Y - X\theta^0 - Z\hat{\beta}\|^2,$$

Naturalmente, $\hat{Y}$ será tanto melhor para aproximar Y , quanto menor for o comprimento de $\hat{\xi}$.

Usando-se a definição da norma euclidiana e lembrando que $X'X\theta^0 + X'Z\hat{\beta} = X'Y$ e $Z'X\theta^0 + Z'Z\hat{\beta} = Z'Y$, tem-se:

$$Y'Y = (X\theta^0 + Z\hat{\beta})'(X\theta^0 + Z\hat{\beta}) + (Y - X\theta^0 - Z\hat{\beta})'(Y - X\theta^0 - Z\hat{\beta})$$

$$Y'Y = (\theta^{0'}X'X\theta^0 + \theta^{0'}X'Z\hat{\beta} + \hat{\beta}'Z'X\theta^0 + \hat{\beta}'Z'Z\hat{\beta}) + (Y'Y - Y'X\theta^0 - Y'Z\hat{\beta} - \theta^{0'}X'Y + \theta^{0'}X'X\theta^0 + \theta^{0'}X'Z\hat{\beta} - \hat{\beta}'Z'Y + \hat{\beta}'Z'X\theta^0 + \hat{\beta}'Z'Z\hat{\beta})$$

$$Y'Y = [\theta^{0'}(X'X\theta^0 + X'Z\hat{\beta}) + \hat{\beta}'(Z'X\theta^0 + Z'Z\hat{\beta})] + [Y'Y - 2\theta^{0'}X'Y + \theta^{0'}(X'X\theta^0 + X'Z\hat{\beta}) - 2\hat{\beta}'Z'Y + \hat{\beta}'(Z'X\theta^0 + Z'Z\hat{\beta})]$$

$$Y'Y = \theta^{0'}X'Y + \hat{\beta}'Z'Y + (Y'Y - \theta^{0'}X'Y - \hat{\beta}'Z'Y) \quad (3.5.1)$$

De (3.2.5) e (3.2.11) verifica-se que:

$$\theta^{0'} = Y'X(X'X)^{-1} - Y'R(R'R)^{-1}Z'X(X'X)^{-1} \quad (3.5.2)$$

e

$$\hat{\beta}' = Y'R(R'R)^{-1}, \quad (3.5.3)$$

Substituindo-se (3.5.2) e (3.5.3) em (3.5.1), vem:

$$\begin{aligned}
 Y'Y &= [Y'X(X'X)^{-1}X'Y - Y'R(R'R)^{-1}Z'X(X'X)^{-1}X'Y + Y'R(R'R)^{-1}Z'Y] + \\
 &\quad + [Y'Y - Y'X(X'X)^{-1}X'Y + Y'R(R'R)^{-1}Z'X(X'X)^{-1}X'Y - Y'R(R'R)^{-1}Z'Y] \\
 Y'Y &= \{Y'X(X'X)^{-1}X'Y + Y'R(R'R)^{-1}Z'[I - X(X'X)^{-1}X']Y\} + \{Y'Y - Y'X(X'X)^{-1}X'Y - \\
 &\quad - Y'R(R'R)^{-1}Z'[I - X(X'X)^{-1}X']Y\} \\
 Y'Y &= \{Y'X(X'X)^{-1}X'Y + Y'R(R'R)^{-1}R'Y\} + \{Y'Y - Y'X(X'X)^{-1}X'Y - Y'R(R'R)^{-1}R'Y\} \\
 Y'Y &= Y'[X(X'X)^{-1}X' + R(R'R)^{-1}R']Y + Y'[I - X(X'X)^{-1}X' - R(R'R)^{-1}R']Y \quad (3.5.4)
 \end{aligned}$$

ou

$$SQ \text{ Total} = SQ \text{ Parâmetros} + SQ \text{ Resíduo}$$

Onde a matriz $[X(X'X)^{-1}X' + R(R'R)^{-1}R']$ é o projetor ortogonal de Y no espaço coluna de $[X : Z]$, $C(X : Z)$, o sub-espaço dos parâmetros, e $[I - X(X'X)^{-1}X' - R(R'R)^{-1}R']$ é o projetor ortogonal de Y no espaço ortogonal ao espaço coluna de $[X : Z]$, $C^1(X : Z)$, o espaço resíduo.

A soma de quadrados dos parâmetros pode ser reescrita como:

$$SQ \text{ Par.} = Y'[X(X'X)^{-1}X']Y + Y'[R(R'R)^{-1}R']Y, \text{ e equivale a}$$

$$SQ \text{ Par.} = SQ | \text{Trat.} + \text{Constante} | + SQ \text{ Reg. d. Trat.}$$

Como em geral, é de interesse a soma de quadrados de tratamentos corrigidos pela média, para sua obtenção faz-se:

$$SQ \text{ Trat.} = Y' [X(X'X)^{-1}X' - N^{-1}11'] Y \quad (3.5.5)$$

Onde 1 é o vetor de N elementos unitários.

É claro que para efeito de cálculos, as somas de quadrados não devem ser obtidas através dos projetores ortogonais, pois, para grandes massas de dados, torna-se impraticável. Com este raciocínio, procurou-se generalizar expressões de interesse prático a partir do desenvolvimento das formas quadráticas, isto é,

$$SQ \text{ Total} = Y' I Y = \sum_{i,j} y_{ij}^2 \quad (3.5.6)$$

$$SQ \text{ (Trat. + Constante)} = Y' [X(X'X)^{-1}X'] Y = \sum_i \frac{1}{n_i} y_{i.}^2 \quad (3.5.7)$$

$$SQ \text{ Trat.} = Y' [X(X'X)^{-1}X' - N^{-1}11'] Y = \sum_i \frac{1}{n_i} y_{i.}^2 - (\sum_{i,j} y_{ij})^2 / N \quad (3.5.8)$$

$$SQ \text{ Reg. d. Trat.} = Y' [R(R'R)^{-1}R'] Y = \sum_i \hat{\beta}_i^2 \sum_j (z_{ij} - \bar{z}_{i.})^2 \quad (3.5.9)$$

$$SQ \text{ Resíduo} = Y' [I - X(X'X)^{-1}X' - R(R'R)^{-1}R'] Y = \sum_{i,j} y_{ij}^2 - \sum_i \frac{1}{n_i} y_{i.}^2 - \sum_i \hat{\beta}_i^2 \sum_j (z_{ij} - \bar{z}_{i.})^2 \quad (3.5.10)$$

O número de graus de liberdade das somas de quadrados é o *rank* da matriz da forma quadrática correspondente. Lembrando-se que os núcleos das formas quadráticas são matrizes simétricas e idempotentes, então o *rank* da forma quadrática é i qual ao traço do núcleo, portanto,

$$\text{g.l. SQ Total} = r[I_{(N)}] = N \quad (3.5.11)$$

$$\text{g.l. SQ Trat.} = r[X(X'X)^{-1}X' - N^{-1}11'] = r(X) - 1 = v - 1 \quad (3.5.12)$$

$$\text{g.l. SQ Reg. d. Trat.} = r[R(R'R)^{-1}R'] = r(Z) = v \quad (3.5.13)$$

$$\begin{aligned} \text{g.l. SQ Res.} &= r[I - X(X'X)^{-1}X' - R(R'R)^{-1}R'] = r[I_{(N)}] - r(X) - r(Z) \\ &= N - 2v \end{aligned} \quad (3.5.14)$$

Pelo teorema 1 do apêndice pode se verificar as distribuições de probabilidade das somas de quadrados, ou seja :

$$\frac{1}{\sigma^2} \text{SQ Par.} = \frac{1}{\sigma^2} Y' [X(X'X)^{-1}X' + R(R'R)^{-1}R'] Y \sim \chi^2 [r(X) + r(Z); \lambda_p] \quad (3.5.15)$$

Onde:

$$\begin{aligned} \lambda_p &= \frac{1}{2\sigma^2} [X\theta + Z\beta]' [X(X'X)^{-1}X' + R(R'R)^{-1}R'] [X\theta + Z\beta] \\ &= \frac{1}{2\sigma^2} [\theta'X'X\theta + 2\theta'X'Z\beta + \beta'Z'X(X'X)^{-1}X'Z\beta + \beta'Z'R(R'R)^{-1}R'Z\beta] \\ &= \frac{1}{2\sigma^2} [\theta'X'X\theta + 2\theta'X'Z\beta + \beta'Z'PZ\beta + \beta'R'R\beta] , \end{aligned}$$

ou,

$$\lambda_p = \frac{1}{2\sigma^2} [\theta'X'X\theta + 2\theta'X'Z\beta + \beta'Z'Z\beta] \quad (3.5.16)$$

Fazendo-se os produtos e somas das matrizes, em λ_p obtem-se:

$$\lambda_p = \frac{1}{2\sigma^2} [\sum_i n_i (\mu + t_i)^2 + 2\sum_i \beta_i z_i (\mu + t_i) + \sum_i \beta_i^2 \sum_j z_{ij}^2]$$

$$\lambda_p = \frac{1}{2\sigma^2} \sum_{i,j} (\mu + t_i + \beta_i z_{ij})^2. \quad (3.5.17)$$

Ou seja, $\frac{1}{\sigma^2}$ SQ Parâmetros tem distribuição de probabilidade de qui-quadrado com $r(X) + r(Z) = 2v$ g.l. e parâmetro de não centralidade $\lambda_p = \frac{1}{2\sigma^2} \sum_{i,j} (\mu + t_i + \beta_i z_{ij})^2$.

A soma de quadrados de tratamentos tem a seguinte distribuição de probabilidade, conforme o Teorema 1.

$$\frac{1}{\sigma^2} \text{SQ Trat.} = \frac{1}{\sigma^2} Y' [X(X'X)^{-1}X' - N^{-1}11'] Y \sim \chi^2 [r(X) - 1; \lambda_t] \quad (3.5.18)$$

Onde:

$$\lambda_t = \frac{1}{2\sigma^2} [\theta'X' + \beta'Z'] [X(X'X)^{-1}X' - N^{-1}11'] [X\theta + Z\beta]$$

$$= \frac{1}{2\sigma^2} [\theta'X'X\theta + 2\theta'X'Z\beta + \beta'Z'PZ\beta - N^{-1}(\theta'X'\mathbb{1}\mathbb{1}'X\theta + 2\theta'X'\mathbb{1}\mathbb{1}'Z\beta + \beta'Z'\mathbb{1}\mathbb{1}'Z\beta)]$$

$$= \frac{1}{2\sigma^2} \{ \sum_i n_i (\mu + t_i + \beta_i \bar{z}_i)^2 - N^{-1} [\sum_{i,j} (\mu + t_i + \beta_i z_{ij})]^2 \}$$

Portanto,

$$\lambda_t = \frac{1}{2\sigma^2} \{ \sum_i n_i (t_i + \beta_i \bar{z}_i)^2 - N^{-1} [\sum_{i,j} (t_i + \beta_i z_{ij})]^2 \} \quad (3.5.19)$$

A distribuição de probabilidade da soma de quadrados de regressão dentro de tratamentos é:

$$\frac{1}{\sigma^2} \text{SQ Reg. d. Trat.} = \frac{1}{\sigma^2} Y' [R(R'R)^{-1}R'] Y \sim \chi^2_{[r(Z); \lambda_r]} \quad (3.5.20)$$

Onde:

$$\begin{aligned} \lambda_r &= \frac{1}{2\sigma^2} [\theta'X' + \beta'Z'] [R(R'R)^{-1}R'] [X\theta + Z\beta] \\ &= \frac{1}{2\sigma^2} [\beta'Z'R(R'R)^{-1}R'Z\beta] \\ &= \frac{1}{2\sigma^2} \beta'R'R\beta \end{aligned}$$

Portanto,

$$\lambda_r = \frac{1}{2\sigma^2} \sum_{i,j} \beta_i^2 (z_{ij} - \bar{z}_i.)^2 \quad (3.5.21)$$

Finalmente,

$$\frac{1}{\sigma^2} \text{SQ Res.} = \frac{1}{\sigma^2} Y' [I - X(X'X)^{-1}X' - R(R'R)^{-1}R'] Y \sim \chi^2_{[N - 2v; \lambda_e]} \quad (3.5.22)$$

Onde:

$$\lambda_e = \frac{1}{2\sigma^2} [\theta'X' + \beta'Z'] [I - X(X'X)^{-1}X' - R(R'R)^{-1}R'] [X\theta + Z\beta] = \emptyset$$

Diz-se que $\text{SQ Res.}/\sigma^2$ tem distribuição de qui-quadrado central. E, ainda, através do teorema 1, mostra-se que a soma de quadrados de tratamentos e soma de quadrados de regressão dentro de tratamentos, são independentemente distribuídos da soma de quadrados de resíduo, e entre si, pois,

$$[X(X'X)^{-1}X' - N^{-1}11'] [I - X(X'X)^{-1}X' - R(R'R)^{-1}R] = \emptyset \quad (3.5.23)$$

$$[R(R'R)^{-1}R'] [I - X(X'X)^{-1}X' - R(R'R)^{-1}R] = \emptyset \quad (3.5.24)$$

$$[X(X'X)^{-1}X' - N^{-1}11'] [R(R'R)^{-1}R'] = \emptyset \quad (3.5.25)$$

Por outro lado, através do teorema 2 do apêndice, pode-se determinar as esperanças matemáticas das somas de qua

dados de interesse, isto é:

$$\begin{aligned}
 E(\text{SQ Res.}) &= E\{Y'[I - X(X'X)^{-1}X' - R(R'R)^{-1}R']Y\} = \\
 &= T_r[I - X(X'X)^{-1}X' - R(R'R)^{-1}R']\sigma^2 + [\theta'X' + \beta'Z'] \\
 &\quad [I - X(X'X)^{-1}X' - R(R'R)^{-1}R'] [X\theta + Z\beta] \\
 &= [N - r(X) - r(Z)]\sigma^2 + \emptyset
 \end{aligned}$$

Logo,

$$E(\text{SQ Res.}) = (N - 2v)\sigma^2; \quad (3.5.26)$$

$$\begin{aligned}
 E(\text{SQ Trat.}) &= E\{Y'[X(X'X)^{-1}X' - N^{-1}\mathbf{1}\mathbf{1}']Y\} \\
 &= T_r[X(X'X)^{-1}X' - N^{-1}\mathbf{1}\mathbf{1}']\sigma^2 + [\theta'X' + \beta'Z'] [X(X'X)^{-1}X' - \\
 &\quad - N^{-1}\mathbf{1}\mathbf{1}'] [X\theta + Z\beta] \\
 &= [r(X) - 1]\sigma^2 + \sum_i n_i (t_i + \beta_i \bar{z}_i)^2 - N^{-1} [\sum_i n_i (t_i + \beta_i \bar{z}_i)]^2
 \end{aligned}$$

Portanto,

$$E(\text{SQ Trat.}) = (v - 1)\sigma^2 + f(t, \beta); \quad (3.5.27)$$

$$E(\text{SQ Reg. d. Trat.}) = E\{Y'[R(R'R)^{-1}R']Y\}$$

$$\begin{aligned}
 &= T_r[R(R'R)^{-1}R']\sigma^2 + [\theta'X' + \beta'Z'] [R(R'R)^{-1}R'] \\
 &\quad [X\theta + Z\beta]
 \end{aligned}$$

$$= r(Z)\sigma^2 + \beta'R'R\beta$$

Assim,

$$E(\text{SQ Reg.d.Trat.}) = v\sigma^2 + \sum_{i,j} \beta_i^2 (z_{ij} - \bar{z}_{i.})^2 \quad (3.5.28)$$

Consequentemente, as respectivas esperanças dos quadrados médios ficam,

$$E(\text{QM Res.}) = \frac{1}{N-2v} E(\text{SQ Res.}) = \frac{1}{N-2v} (N-2v)\sigma^2 = \sigma^2; \quad (3.5.29)$$

$$E(\text{QM Trat.}) = \frac{1}{v-1} (v-1)\sigma^2 + \frac{1}{v-1} f(t, \beta) = \sigma^2 + \frac{1}{v-1} f(t, \beta); \quad (3.5.30)$$

$$E(\text{QM Reg.d.Trat.}) = \frac{1}{v} v\sigma^2 + \frac{1}{v} \sum_{i,j} \beta_i^2 (z_{ij} - \bar{z}_{i.})^2 = \sigma^2 + \frac{1}{v} \sum_{i,j} \beta_i^2 (z_{ij} - \bar{z}_{i.})^2; \quad (3.5.31)$$

Com os resultados obtidos anteriormente, pode-se estruturar o seguinte quadro para análise de variância.

QUADRO 4. Análise de variância para ajustamento covariante(β),
corrigido pela média geral e pelos tratamentos no mo-
delo $Y = X\theta + Z\beta + \xi$

C. VARIAÇÃO	G.L.	S.Q.	Q.M.	E (Q.M.)	F
Tratamentos	$v - 1$	$\sum_i \frac{1}{n_i} y_i^2 - N\bar{Y}^2$	V_1	$\sigma^2 + \frac{1}{v-1} f(t, \beta)$	V_1/V_2
Reg. d. Trat.	v	$\sum_i \tilde{\beta}_i^2 \sum_j (z_{ij} - \bar{z}_i)^2$	V_2	$\sigma^2 + \frac{1}{v} \sum_{i,j} \beta_i^2 (z_{ij} - \bar{z}_i)^2$	V_2/V_3
Resíduo	$N - 2v$	Diferença	V_3	σ^2	
TOTAL	$N - 1$	$\sum_{i,j} y_{ij}^2 - (\sum_{i,j} y_{ij})^2 / N$		-	

Sob a hipótese $H_0^{(1)}: \beta_i = \emptyset$, para todo i , $\lambda_r = 0$
e $\frac{1}{\sigma^2} Y' [R(R'R)^{-1}R'] Y \sim \chi^2[r(Z)]$ isto é, tem distribuição de
qui-quadrado central.

Assim,

$$\frac{r(z) \text{QM Reg.d.Trat.}}{\sigma^2} \sim \chi^2[v] \quad (3.5.32)$$

E, de (3.5.26) tira-se que;

$$\frac{[N - r(x) - r(z)] \text{QM res.}}{\sigma^2} \sim \chi^2[N - 2v] \quad (3.5.33)$$

Consequentemente, pelo teorema 3 do apêndice, resulta;

$$\frac{\text{QM Reg.d.Trat.}}{\text{QM Resíduo}} \sim F[v; N - 2v] \quad (3.5.34)$$

Quando a hipótese $H_0^{(1)}: \beta_1 = 0$, é rejeitada a um determinado nível de significância desejado, o próximo passo será, sem dúvida, testar a hipótese de homogeneidade dos coeficientes de regressão, isto é, $H_0^{(2)}: \beta_1 = \beta_2 = \dots = \beta_v$, que pode ser expressa como $H_0^{(2)}: C'\beta = 0$, onde C' é, por exemplo, uma matriz do tipo

$$C' = \begin{bmatrix} 1 & -1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -1 & \dots & 0 & 0 \\ \dots & \dots & \dots & \dots & \dots & \dots \\ 0 & 0 & 0 & \dots & 1 & -1 \end{bmatrix} \quad (3.5.35)$$

$v-1$ v

De (3.4.9), tem-se que:

$$C'\hat{\beta} \sim N[C'\beta; C'(R'R)^{-1}C\sigma^2]$$

e, sabendo-se que a matriz $C'(R'R)^{-1}C$ é positiva definida, e portanto, não singular, pois C' é de rank linha completo. E, seja A uma matriz não singular, tal que,

$$[C'(R'R)^{-1}C]^{-1} = A'A$$

e onde,

$$C'(R'R)^{-1}C = A^{-1}A^{-1'}.$$

Considere a forma quadrática $Q'Q$, onde:

$$Q = \frac{A(C'\hat{\beta} - C'\beta)}{\sigma}$$

Então,

$$E[Q] = \frac{1}{\sigma} E(AC'\hat{\beta} - AC'\beta) = 0$$

e

$$\text{Var}[Q] = \frac{1}{\sigma^2} A \text{Var}(C'\hat{\beta}) A' = \frac{1}{\sigma^2} AA^{-1}A^{-1'}A'\sigma^2 = I.$$

Assim,

$$Q'Q \sim \chi^2[r(C')]. \quad (3.5.36)$$

Dessa forma,

$$Q'Q = \frac{1}{\sigma^2} (C'\hat{\beta} - C'\beta)' [C'(R'R)^{-1}C]^{-1} (C'\hat{\beta} - C'\beta) \sim \chi^2[r(C')]. \quad (3.5.37)$$

Sob $H_0^{(2)}: C'\beta = 0$, tem-se então, que

$$Q'Q = \frac{1}{\sigma^2} (C'\hat{\beta})' [C'(R'R)^{-1}C]^{-1} (C'\hat{\beta}) \sim \chi^2[r(C')] \quad (3.5.38)$$

A forma quadrática dada em (3.5.37) é definida como soma de quadrado devida a hipótese $H_0: C'\beta = w$. Como aqui é considerada a hipótese do tipo $H_0: C'\beta = \emptyset$, então tem-se:

$$SQ(H_0^{(2)}) = (C'\hat{\beta})' [C'(R'R)^{-1}C]^{-1} (C'\hat{\beta}) \quad (3.5.39)$$

que corresponde a:

$$SQ(H_0^{(2)}) = \sum_i \hat{\beta}_i^2 \sum_j (z_{ij} - \bar{z}_{i.})^2 - \left[\sum_{i,j} (y_{ij} - \bar{y}_{i.}) (z_{ij} - \bar{z}_{i.}) \right]^2 / \sum_{i,j} (z_{ij} - \bar{z}_{i.})^2 \quad (3.5.40)$$

Além disso, pelo teorema 3 do apêndice, verifica-se que $SQ(H_0^{(2)})$ e $SQ \text{ Res.}$ são independentes, isto é:

Substituindo-se (3.2.6) em (3.5.39) obtém-se:

$$SQ(H_0^{(2)}) = Y' \{R(R'R)^{-1}C[C'(R'R)^{-1}C]^{-1}C'(R'R)^{-1}R'\}Y \quad (3.5.41)$$

Assim,

$$\{R(R'R)^{-1}C[C'(R'R)^{-1}C]^{-1}C'(R'R)^{-1}R'\} \{I - X(X'X)^{-1}X' - R(R'R)^{-1}R'\} = \emptyset$$

Portanto,

$$\frac{SQ(Ho^{(2)})/r(C')}{QM \text{ Resíduo}} = \frac{SQ(Ho^{(2)})/(v-1)}{QM \text{ Resíduo}} \sim F_{\{(v-1); (N-2v)\}} \quad (3.5.42)$$

Uma vez comprovada a homogeneidade dos coeficientes de regressão, ou seja, a não rejeição da hipótese $Ho^{(2)}$: $C'\beta = \emptyset$ a um nível de significância especificado, deve-se recorrer às técnicas da análise de covariância tradicional para a exploração dos dados, do contrário, caso a hipótese seja rejeitada, o procedimento imediato será averiguar a existência dos efeitos dos tratamentos, através da hipótese $Ho^{(3)}$: $t_1 = t_2 = \dots = t_v$, considerado um β_i específico para cada tratamento i .

Para encontrar a forma quadrática que fornece a soma de quadrados de tratamentos ajustada para média e para covariantes, define-se $Z'Z$, como a matriz da soma de quadrados e produtos dos z 's corrigidos para suas correspondentes médias gerais, ou seja, $Z = (I - N^{-1}E)Z$, onde I , é uma matriz identidade e E é uma matriz quadrada com todos os elementos iguais a 1, ambas as matrizes de ordem N .

Portanto, tem-se:

$$\begin{aligned}
 \mathbf{z} = (\mathbf{I} - \mathbf{N}^{-1}\mathbf{E})\mathbf{z} = & \begin{bmatrix} z_{11} - \frac{1}{N}z_{1.} & -\frac{1}{N}z_{2.} & \dots & -\frac{1}{N}z_{v.} \\ z_{12} - \frac{1}{N}z_{1.} & -\frac{1}{N}z_{2.} & \dots & -\frac{1}{N}z_{v.} \\ \vdots & \vdots & & \vdots \\ z_{1n_1} - \frac{1}{N}z_{1.} & -\frac{1}{N}z_{2.} & \dots & -\frac{1}{N}z_{v.} \\ -\frac{1}{N}z_{1.} & z_{21} - \frac{1}{N}z_{2.} & \dots & -\frac{1}{N}z_{v.} \\ -\frac{1}{N}z_{1.} & z_{22} - \frac{1}{N}z_{2.} & \dots & -\frac{1}{N}z_{v.} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{N}z_{1.} & z_{2n_2} - \frac{1}{N}z_{2.} & \dots & -\frac{1}{N}z_{v.} \\ \vdots & \vdots & & \vdots \\ \vdots & \vdots & & \vdots \\ -\frac{1}{N}z_{1.} & -\frac{1}{N}z_{2.} & \dots & z_{v1} - \frac{1}{N}z_{v.} \\ -\frac{1}{N}z_{1.} & -\frac{1}{N}z_{2.} & \dots & z_{v2} - \frac{1}{N}z_{v.} \\ \vdots & \vdots & & \vdots \\ -\frac{1}{N}z_{1.} & -\frac{1}{N}z_{2.} & \dots & z_{vn_v} - \frac{1}{N}z_{v.} \end{bmatrix} \quad (3.5.43)
 \end{aligned}$$

$\mathbf{N}$ $\mathbf{v}$

$$Z'Z = \begin{bmatrix} \sum_j z_{1j}^2 - \frac{1}{N} z_{1.}^2 & -\frac{1}{N} z_{1.} z_{2.} & \dots & -\frac{1}{N} z_{1.} z_{v.} \\ -\frac{1}{N} z_{2.} z_{1.} & \sum_j z_{2j}^2 - \frac{1}{N} z_{2.}^2 & \dots & -\frac{1}{N} z_{2.} z_{v.} \\ \dots & \dots & \dots & \dots \\ -\frac{1}{N} z_{v.} z_{1.} & -\frac{1}{N} z_{v.} z_{2.} & \dots & \sum_j z_{vj}^2 - \frac{1}{N} z_{v.}^2 \end{bmatrix} \quad (3.5.44)$$

e ainda,

$$Z'Y = \begin{bmatrix} \sum_j y_{1j} z_{1j} - \frac{1}{N} z_{1.} y_{..} \\ \sum_j y_{2j} z_{2j} - \frac{1}{N} z_{2.} y_{..} \\ \vdots \\ \sum_j y_{vj} z_{vj} - \frac{1}{N} z_{v.} y_{..} \end{bmatrix} \quad (3.5.45)$$

A partir do modelo $y_{ij} = \mu + t_i + \beta_i z_{ij} + e_{ij}$ a a té aqui considerado, se for retirado o efeito do tratamento t_i , torna-se um modelo de regressão de y nos z 's, para o qual as es timativas de mínimos quadrados dos parâmetros são obtidos por $\hat{\beta}_i^* = (Z'Z)^{-1}Z'Y$ e a soma de quadrados devida a regressão é dada por:

$$SQ \text{ Reg.} = Y'[Z(Z'Z)^{-1}Z']Y = \hat{\beta}^{*'}Z'Y \quad (3.5.46)$$

ou

$$SQ \text{ Reg.} = \sum_i \hat{\beta}_i^* \left(\sum_j y_{ij} z_{ij} - \frac{1}{N} z_{i.} y_{..} \right) \quad (3.5.47)$$

A soma de quadrados de tratamentos ajustada para média e para covariantes é obtida pela diferença entre a soma de quadrados de parâmetros em (3.5.4) corrigida para a média e a soma de quadrados de regressão em (3.5.46), ou seja:

$$SQ \text{ Trat. (aj.)} = Y' [X(X'X)^{-1}X' - N^{-1}11'] + R(R'R)^{-1}R' - Z(Z'Z)^{-1}Z'Y \quad (3.5.48)$$

ou

$$SQ \text{ Trat. (aj.)} = \sum_i \frac{1}{n_i} y_{i.}^2 - N \bar{y}_{..}^2 + \sum_i \hat{\beta}_i^2 \sum_j (z_{ij} - \bar{z}_{i.})^2 - \sum_i \hat{\beta}_i^* \left(\sum_j y_{ij} z_{ij} - \frac{1}{N} z_{i.} y_{..} \right) \quad (3.5.49)$$

$$= SQ \text{ Trat. (usual)} + SQ \text{ Reg.d.Trat.} - SQ \text{ Reg.}$$

O número de graus de liberdade para esta forma quadrática, é:

$$g.l. \text{ SQ Trat. (aj.)} = r(X) - 1 + r(Z) - r(\bar{Z}) = r(X) - 1 = v - 1 \quad (3.5.50)$$

A distribuição de probabilidade de SQ Trat.(aj.), com a aplicação do teorema 1 do apêndice, é,

$$\frac{1}{\sigma^2} \text{SQ Trat. (aj.)} = \frac{1}{\sigma^2} Y' [X(X'X)^{-1}X' - N^{-1}11'] + R(R'R)^{-1}R' - \\ - Z(Z'Z)^{-1}Z'] Y \sim \chi^2[(v-1); \lambda_a] \quad (3.5.51)$$

onde:

$$\lambda_a = \frac{1}{2\sigma^2} [\theta'X' + \beta'Z'] [X(X'X)^{-1}X' - N^{-1}11'] + R(R'R)^{-1}R' - \\ - Z(Z'Z)^{-1}Z'] [X\theta + Z\beta] \quad (3.5.52)$$

Portanto, sob $H_0^{(3)}$, $\lambda_a = \emptyset$ e,

$$\frac{1}{\sigma^2} \text{SQ Trat. (aj.)} \sim \chi^2[r(X) - 1] \quad (3.5.53)$$

Além disso, a soma de quadrados de tratamentos a justada é independente da soma de quadrados de resíduo. Pois ,

$$[X(X'X)^{-1}X' - N^{-1}11'] + R(R'R)^{-1}R' - Z(Z'Z)^{-1}Z'] [I - X(X'X)^{-1}X' - \\ - R(R'R)^{-1}R'] = \emptyset$$

A esperança da soma de quadrados de tratamentos ajustada, de acordo com o teorema 2 do apêndice, é:

$$E[\text{SQ Trat. (aj.)}] = (v-1)\sigma^2 + [\theta'X' + \beta'Z'] [X(X'X)^{-1}X' - N^{-1}11'] + \\ + R(R'R)^{-1}R' - Z(Z'Z)^{-1}Z'] [X\theta + Z\beta] \quad (3.5.54)$$

ou

$$E[\text{SQ Trat. (aj.)}] = (v - 1)\sigma^2 + f(t); f(t) > 0 \quad (3.5.55)$$

e

$$E[\text{QM Trat. (aj.)}] = \frac{1}{(v-1)} E[\text{SQ Trat. (aj.)}] = \sigma^2 + \frac{1}{(v-1)} f(t) \quad (3.5.56)$$

Com estes resultados, pode-se estruturar um novo quadro de análise:

QUADRO 5. Análise de variância para o ajustamento dos tratamentos pelas covariantes, no modelo $Y = X\theta + Z\beta + \xi$

C. VARIAÇÃO	G.L.	S.Q.	Q.M.	E(Q.M.)
Regressão	v	$Y'[Z(Z'Z)^{-1}Z']Y$	V_1	-
Trat. (aj.)	v - 1	$Y'[X(X'X)^{-1}X' - N^{-1}11']' + R(R'R)^{-1}R' - Z(Z'Z)^{-1}Z']Y$	V_2	$\sigma^2 + \frac{1}{(v-1)} f(t)$
Resíduo	N - 2v	$Y'[I - X(X'X)^{-1}X' - R(R'R)^{-1}R']Y$	V_3	σ^2
TOTAL	N - 1	$Y'[I - N^{-1}11']Y$	-	-

Portanto, sob $H_0^{(3)}$: $t_1 = t_2 = \dots = t_v$,

$$\frac{\text{QM Trat. (aj.)}}{\text{QM Resíduo}} \sim F[(v - 1); (N - 2v)] \quad (3.5.57)$$

Conforme o teorema 3 do apêndice.

3.6. Comparações Múltiplas

Uma vez testada a hipótese $H_0^{(3)}: t_1 = t_2 = \dots = t_v$, a um nível de significância desejado, formulam-se as regras para os métodos de comparação entre os efeitos de tratamentos ajustados para as covariantes.

De (3.4.6), tem-se que a estimativa de um contraste entre médias de tratamentos ajustados é obtida por:

$$\hat{Y}_h = \sum_{i=1}^V C_{hi} (\bar{y}_{i.} - \hat{\beta}_i \bar{z}_{i.}) = \sum_{i=1}^V C_{hi} m_i. \quad (3.6.1)$$

Por outro lado, de (3.4.7) tem-se que a variância da estimativa deste contraste, é dada por:

$$\text{Var}(\hat{Y}_h) = \sum_{i=1}^V C_{hi}^2 \left[\frac{1}{n_i} + \frac{\bar{z}_{i.}^2}{\sum_j (z_{ij} - \bar{z}_{i.})^2} \right] \sigma^2.$$

Consequentemente, a estimativa da variância da estimativa do contraste é,

$$\hat{\text{var}}(\hat{Y}_h) = \sum_{i=1}^V C_{hi}^2 \left[\frac{1}{n_i} + \frac{\bar{z}_{i.}^2}{\sum_j (z_{ij} - \bar{z}_{i.})^2} \right] \text{QM Res.} \quad (3.6.2)$$

Com os resultados obtidos em (3.4.6) e (3.6.2) pode-se aplicar qualquer método de comparações múltiplas que se julgar conveniente.

4. ILUSTRAÇÃO DO MÉTODO

De acordo com o propósito deste trabalho, faz-se conveniente a ilustração dos resultados obtidos através de um exemplo concreto. Para tanto, foram tomados os dados de um experimento instalado em soca de cana-de-açúcar pelo Engº Agrº Antônio Fernando de Sobral do IAA/PLANALSUCAR - Coordenadoria Regional Norte, no qual consideraram-se cinco níveis de N e quatro repetições. O número de colmos por metro linear de sulco e as respectivas produções em t/ha são apresentados na tabela 1.

TABELA 1. Número médio de colmos por metro linear de sulco(z) e produção de cana-de-açúcar em t/ha(y)

Níveis de N(Kg/ha)	R E P E T I Ç Õ E S								TOTAIS	
	1		2		3		4			
	z	y	z	y	z	y	z	y	z _{i.}	y _{i.}
00	17,7	55,0	13,6	50,0	14,9	53,0	15,8	54,8	62,0	212,8
30	12,0	42,3	11,1	43,0	17,6	45,0	17,3	44,5	58,0	174,8
60	14,5	64,0	11,6	60,0	12,2	62,0	9,7	58,0	48,0	244,0
90	8,0	82,0	12,6	100,4	6,0	72,0	9,4	81,6	36,0	336,0
120	12,0	50,0	14,8	48,0	16,0	52,0	17,2	52,4	60,0	202,4

4.1. Estimativas dos Parâmetros

De (3.2.10), tem-se que:

$$\hat{\beta}_i = \frac{\sum_j (y_{ij} - \bar{y}_{i.})(z_{ij} - \bar{z}_{i.})}{\sum_j (z_{ij} - \bar{z}_{i.})^2}$$

Portanto,

$$\hat{\beta}_1 = 10,64 / 8,90 = 1,1955$$

$$\hat{\beta}_2 = 12,15 / 35,26 = 0,3446$$

$$\hat{\beta}_3 = 15,00 / 11,74 = 1,2777$$

$$\hat{\beta}_4 = 96,08 / 23,12 = 4,1557$$

$$\hat{\beta}_5 = 7,68 / 14,88 = 0,5161.$$

e de (3.2.12), tem-se:

$$\mu^0 = 0$$

$$t_i^0 = \bar{y}_{i.} - \hat{\beta}_i \bar{z}_{i.}$$

assim,

$$\mu^0 = 0$$

$$t_1^0 = 34,67$$

$$t_2^0 = 38,70$$

$$t_3^0 = 45,67$$

$$t_4^0 = 46,60$$

$$t_5^0 = 42,86$$

4.2. Análise de Variância para o Ajuste Covariante

Aplicando-se os resultados constantes no quadro 4, vem:

$$SQ \text{ Tratamentos} = \frac{1}{4}[(212,8)^2 + \dots + (202,4)^2] - \frac{1}{20}(1.170)^2 = 3.864,16;$$

$$SQ \text{ Reg.d. Trat.} = (1,1955)^2(8,90) + \dots + (0,5161)^2(14,88) = 439,3151;$$

$$SQ \text{ Total} = (55,0)^2 + \dots + (52,4)^2 - \frac{1}{20}(1.170)^2 = 4.340,06;$$

$$SQ \text{ Resíduo} = 4.340,06 - 3.864,16 - 439,3151 = 36,5849.$$

Com estes resultados, estrutura-se o seguinte quadro de análise:

C.VARIAÇÃO	G.L.	S.Q.	Q.M.	F
Tratamentos	4	3.864,1600	-	-
Reg. d. Trat.	5	439,3151	87,8630	24,02**
Resíduo.	10	36,5849	3,6585	-
TOTAL	19	4.340,0600		

Percebe-se portanto, que se rejeita a hipótese $H_0^{(1)}: \beta_i = \emptyset$, isto é, o efeito de regressão covariante é significativo ao nível de 1% de probabilidade.

4.3. Teste de Homogeneidade dos Coeficientes de Regressão Linear

Aplicando-se o resultado encontrado em (3.5.40), vem:

$$SQ(H_0^{(2)}) = 439,3151 - 213,3802 = 225,9349.$$

Consequentemente,

$$F = \frac{225,9349/4}{3,6585} = 15,44**$$

Logo, rejeita-se a hipótese $H_0^{(2)}: \beta_1 = \beta_2 = \dots = \beta_5$,

isto é, os coeficientes de regressão linear são heterogêneos, e desse modo, a análise dos dados deve ser feita segundo o método proposto por este trabalho.

4.4. Análise de Variância para o Ajuste de Tratamentos

Tomando-se os resultados encontrados em (3.5.44) e (3.5.45) e aplicando-se aos dados, obtém-se:

$$Z'Z = \begin{bmatrix} 777,70 & -179,80 & -148,80 & -111,60 & -186,00 \\ -179,80 & 708,06 & -139,20 & -104,40 & -174,00 \\ -148,80 & -139,20 & 472,54 & -86,40 & -144,00 \\ -111,60 & -104,40 & -86,40 & 282,32 & -108,00 \\ -186,00 & -174,00 & -144,00 & -108,00 & 734,88 \end{bmatrix};$$

$$Z'Y = \begin{bmatrix} - & 317,96 \\ - & 846,25 \\ & 135,00 \\ & 1.014,08 \\ - & 466,32 \end{bmatrix};$$

$$\hat{\beta}^* = \begin{bmatrix} \hat{\beta}_1^* \\ \hat{\beta}_2^* \\ \hat{\beta}_3^* \\ \hat{\beta}_4^* \\ \hat{\beta}_5^* \end{bmatrix} = \begin{bmatrix} 0,6456 \\ 0,0422 \\ 1,4734 \\ 4,5008 \\ 0,4890 \end{bmatrix}$$

consequentemente,

$$SQ \text{ Regressão} = (0,6456)(-317,96) + \dots + (0,4890)(-466,32) = 4.294,0631.$$

$$SQ \text{ Trat. (aj.)} = SQ \text{ Trat.} + SQ \text{ Reg.d.Trat.} - SQ \text{ Reg.}$$

$$= 3.864,16 + 439,3620 - 4.294,0631 = 9,4589.$$

Assim, o seguinte quadro de análise pode ser estruturado:

C.VARIAÇÃO	G.L.	S.Q.	Q.M.	F
Regressão	5	4.294,0631		
Trat. (aj.)	4	9,4589	2.3647	0,65 (n.s.)
Resíduo	10	36,5849	3.6585	
TOTAL	19	4.340,0600		

A hipótese $H_0^{(3)}: t_1 = t_2 = \dots = t_v$, não foi re

jeitada, ou seja, não existe efeito de tratamentos ao nível de 5% de probabilidade.

Caso seja feita uma análise de variância considerando apenas as produções de cana-de-açúcar em t/ha, encontram-se os seguintes resultados: Q.M. Trat. = 966,04; Q.M. Res. = 31,7267 e $F = 30,45$. E, se for empregado a análise de covariância tradicional obtém-se: Q.M. Trat. (aj.) = 744,88; Q.M. Res. (aj.) = 18,75 e $F = 39,73$ o que levaria o pesquisador, em ambos os casos, a concluir erroneamente sobre os efeitos dos tratamentos observados no ensaio.

4.5. Estimativas das Médias dos Tratamentos

De (3.2.12) tira-se que $m_i = t_i^0 = \bar{y}_{i.} - \hat{\beta}_i \bar{z}_{i.}$

Assim,

$$m_1 = 34,67 \text{ t/ha}$$

$$m_2 = 38,70 \text{ "}$$

$$m_3 = 45,67 \text{ "}$$

$$m_4 = 46,60 \text{ "}$$

$$m_5 = 42,86 \text{ "}$$

4.6. Comparações Múltiplas

Considere-se, a título de ilustração, dois contrastes para aplicação do teste de SCHEFFÉ e teste t:

$$a) \hat{Y}_1 = -m_1 + m_4;$$

$$b) \hat{Y}_2 = -3m_1 - 3m_2 + 2m_3 + 2m_4 + 2m_5.$$

Obtendo-se as seguintes estimativas para os contrastes, de acordo com (3.6.1):

$$\hat{Y}_1 = 11,93 \text{ t/ha};$$

$$\hat{Y}_2 = 50,15 \text{ t/ha}.$$

As estimativas das variâncias de $\hat{Y}_1$ e $\hat{Y}_2$, conforme (3.6.2) são:

$$\text{Vâr}(\hat{Y}_1) = 113,2599;$$

$$\text{Vâr}(\hat{Y}_2) = 1.562,6419.$$

Aplicando-se o teste de SCHEFFÉ ao contraste $\hat{Y}_1$,
tem-se:

$$S = \sqrt{(v-1)\text{Vâr}(\hat{Y})F}$$

O valor de F com 4 g.l. para tratamentos e 10 graus de liberdade do resíduo ao nível de 5% de probabilidade é 3,48.

Assim,

$$S = \sqrt{(5-1)(113,2599)(3,48)} \approx 39,71$$

Aplicando-se o teste t ao contraste $\hat{Y}_2$, tem-se:

$$t = \frac{\hat{Y}_2 - 0}{\sqrt{\text{Var}(\hat{Y})}} = \frac{50,15}{\sqrt{1.562,6419}} = 1,27$$

O valor de t para 10 graus de liberdade é 2,23.

Portanto, os dois contrastes não apresentaram significância estatística, o que já era esperado.

5. CONCLUSÕES

Levando em consideração os objetivos deste trabalho, faz-se oportuno que as conclusões sejam feitas, ora comparando os resultados obtidos com aqueles encontrados na literatura e em outros casos apresentando conclusões inerentes às justificativas teóricas aqui desenvolvidas.

5.1. O Modelo Estatístico

O modelo estatístico para análise de covariância de ensaios com um fator e uma covariável, considera uma regressão única para o ajuste de todos os tratamentos. No caso, o modelo proposto por este trabalho, admite uma regressão exclusiva para cada tratamento, não apresentando dificuldades quando se deseja testar as hipóteses básicas de nulidade sobre os parâmetros do modelo.

5.2. O Sistema de Equações Normais

Este, obtido através do método de mínimos quadrados, apresentou-se semelhante ao empregado pela formulação geral de análise de covariância, apenas com uma diferença devida a forma de apresentação da matriz das covariáveis, Z .

5.3. As Funções Lineares Estimáveis

Os coeficientes de regressão, como também qualquer combinação linear destes, são sempre estimáveis. Com relação aos efeitos de tratamentos, observa-se que suas combinações lineares são estimáveis sempre que elas forem estimáveis no modelo sem covariantes.

5.4. Os Graus de Liberdade para o Resíduo

Como em outros casos, deve ser inteiro e positivo o número de graus de liberdade para o resíduo, isto é, $N - 2v > 0 \therefore N > 2v$. Assim, para um experimento balanceado, deve haver, pelo menos, três repetições por tratamento, ou seja, $rv > 2v \therefore r > 2$. Desse modo, é desejável que o número de repetições para cada tratamento, nos ensaios balanceados ou não, seja no mínimo três, uma vez que a estimativa de uma regressão linear dentro de um tratamento com menos de três observações passa a ser um problema matemático e não estatístico.

5.5. A Análise de Variância

Para este tipo de análise, faz-se conveniente a divisão em três análises independentes:

- a) Em primeiro plano, propõe-se testar a existência do efeito de regressão covariante;
- b) numa outra etapa, verifica-se a homocedasticidade dos coeficientes de regressão;
- c) finalmente, testa-se a existência dos efeitos de tratamentos ajustados para regressões covariantes.

Com relação aos itens a) e b), encontraram-se expressões bastante simples para se determinar as somas de quadrados exigidas para se testar as hipóteses desejadas. Contudo, para o item c), a hipótese a ser testada, requer a soma de quadrados de uma regressão múltipla o que não se consegue com expressões muito simples.

5.6. Teste de Hipótese

As três hipóteses básicas testadas, comportam-se de modo usual com relação ao resíduo que é comum às três.

5.7. ~~Comparações Múltiplas~~

De acordo com os resultados obtidos para estima
tivas das variância, das estimativas de contrastes entre duas mé
dias ajustadas de tratamentos, para cada par de médias confron
tadas há um estimador de variância específica para a estimativa
deste contraste.

6. BIBLIOGRAFIA

- COCHRAN, W.H., 1957. Analysis of Covariance: Its Nature and Uses. Biometrics. Raleigh, 13:261-281.
- COX, C.P., 1958. The Analysis of Latin Square Designs With Individual Curvatures in One Direction. Biometrics. Raleigh, 14:193-204.
- COX, D.R. e P. McCULLAGH, 1982. Some Aspects of Analysis of Covariance. Biometrics. Raleigh, 30:541-561.
- DIAS, J.F.S., 1981. Análise de Covariância Intrablocos, com p Variáveis Auxiliares, para Delineamento em Blocos Incompletos Equilibrados. Piracicaba, ESALQ/USP. 96p. (Tese de Doutorado).
- HENDERSON, C.R.Jr., 1982. Analysis of Covariance in the Mixed Model: Higher-level, Nonhomogeneous, and Random Regression. Biometrics. Raleigh, 38:623-640.
- IEMMA, A.F., 1985. Modelos Lineares, uma Introdução para Profissionais da Pesquisa Agropecuária. Piracicaba, ESALQ/USP. 142p. (1º Simpósio de Estatística Aplicada à Experimentação Agronômica. 30a. Reunião Anual da Região Brasileira da Sociedade Internacional da Biometria).

- KIRK, R.E., 1968. Experimental Design: Procedures for the Behavioral Sciences. Belmont, Wadsworth Publishing Company. 577p.
- KOCH, G.G., I.A. AMARA, G.W. DAVIS e D.B. GILLINGS., 1982. A Review of some Statistical Methods for Covariance Analysis of Categorical Data. Biometrics. Raleigh, 38:563-595.
- MUNIZ, J.A., 1982. Análise de Covariância em Experimentos em Blocos Casualizados, com Observações Perdidas. Piracicaba, ESALQ/USP. 71p. (Dissertação de Mestrado).
- NETER, J. e W. WASSERMAN, 1974. Applied Linear Statistical Models. Homewood, Richard D. Irwin. 842p.
- ROBSON, D.S. e G.F. ATKINSON, 1960. Individual Degrees of Freedom for Testing Homogeneity of Regression Coefficients in One-Way Analysis of Covariance. Biometrics. Raleigh, 16:593-605.
- SNEDECOR, W.G. e W.G. COCHRAN, 1967. Statistical Methods. 6a. Ed., Ames, The Iowa State University Press. 593p.
- SEARLE, S.R., 1971. Linear Models. New York, John Wiley e Sons. 532p.

- SEARLE, S.R., 1979. Alternative Covariance Models for the 2-Way Crossed Classifications. Commun. Statist. - Theor. Meth. New York, A8(8):799-818.
- STEEL, R.G.D. e J.H. TORRIE, 1960. Principles and Procedures res of Statistics. New York, McGraw-Hill Book. 481p.
- URQUHART, N.S., 1982. Adjustment in Covariance When one Factor Affects the Covariant. Biometrics. Raleigh, 38:651-660.

7. APÊNDICE

TEOREMA 1 (SEARLE, 1971)

Se $Y \sim N(X\theta; I\sigma^2)$, então:

i) $(1/\sigma^2)Y'AY \sim \chi^2\{k; (1/2\sigma^2)\theta'X'AX\theta\}$ se e somente se A é i dempotente de posto k .

ii) $Y'AY$ e BY são independentemente distribuídas se e só se $BA = \emptyset$.

iii) $Y'AY$ e $Y'CY$ são independentemente distribuídas se e só se $AC = \emptyset$ (ou $CA = \emptyset$).

iv) Se $Y'IY = Y'Y = \sum_{i=1}^p Y'A_iY$, onde $r(A_i) = k_i$, então uma das três asserções:

a) A_i é idempotente para qualquer i ;

b) $A_iA_{i'} = \emptyset$, para $i \neq i'$;

c) $\sum_i k_i = k$, onde k é a dimensão de I ;

é condição necessária e suficiente para que as asserções se

guintes sejam simultaneamente verdadeiras:

$$a) (1/\sigma^2) Y' A_i Y \sim \chi^2 \{k_i; (1/2\sigma^2) \theta' X' A_i X \theta\};$$

$$b) Y' A_i Y \text{ e } Y' A_{i'} Y, \text{ são independentes } (i \neq i').$$

TEOREMA 2 (SEARLE, 1971)

Se $Y \sim D(X\theta; I\sigma^2)$, então,

$$E(Y'AY) = \sigma^2 \text{Tr}(A) + \theta' X' A X \theta$$

TEOREMA 3 (SEARLE, 1971)

Sejam as formas quadráticas $Q_1 = Y'AY$, $Q_2 = Y'BY$, $Q_3 = Y'CY$ e $Q_4 = Y'DY$, independentemente distribuídas e tais que:

$$Q_1 \sim N(\emptyset; I); Q_2 \sim \chi^2 \{k; \lambda\}; Q_3 = \chi^2 \{k\}; Q_4 \sim \chi^2 \{p\}.$$

Então,

$$1) \frac{Q_1}{\sqrt{Q_3}/k} \sim t_{(k)} \quad \text{e} \quad \frac{Q_1}{\sqrt{Q_4}/p} \sim t_{(p)}$$

$$\text{ii) } \frac{Q_2/k}{Q_4/p} \sim F_{\{k;p;\lambda\}} \text{ e } \frac{Q_3/k}{Q_4/p} \sim F_{\{k;p\}}$$

onde:

λ é o parâmetro de não centralidade.