

BOOTSTRAP E REGRESSÃO. UMA OPÇÃO AOS
MÍNIMOS QUADRADOS GENERALIZADOS

ANSELMO CHAVES NETO



UNICAMP

UNIVERSIDADE ESTADUAL DE CAMPINAS
INSTITUTO DE MATEMÁTICA, ESTATÍSTICA E CIÊNCIA DA COMPUTAÇÃO

CAMPINAS - SÃO PAULO
BRASIL

BOOTSTRAP E REGRESSÃO. UMA OPÇÃO AOS
MÍNIMOS QUADRADOS GENERALIZADOS

Este exemplar corresponde a redação final da tese devidamente corrigida e defendida pelo Sr. ANSELMO CHAVES NETO e aprovada pela Comissão Julgadora.

Campinas, 29 de *março* de 1985



Prof. Dr. Sebastião de Amorim
Orientador

Dissertação apresentada ao Instituto de Matemática, Estatística e Ciência da Computação, como requisito parcial para obtenção do título de Mestre em Estatística

Março/1985

UNICAMP
BIBLIOTECA CENTRAL

Para Dna. Laura, Sonia, Isabella, Paula e Helena;
que são as mulheres da minha vida

AGRADECIMENTOS

Desejo agradecer, primeiramente, ao meu Professor e Orientador, Prof.Dr. Sebastião de Amorim pelo modo sempre paciente e seguro com que me orientou neste trabalho, bem como pela oportunidade de trabalhar ao seu lado.

Agradeço, também, aos meus Colegas da UNICAMP e Professores do Departamento de Estatística desta Universidade, pela amizade, colaboração e ensinamentos.

E ainda, minha sincera gratidão aos meus Colegas de trabalho da UFPr, aqui simbolizados nas pessoas dos Profs. Jair Mendes Marques e João Carlos Loyola Pires, pelo incentivo de sempre.

ÍNDICE

I -	"BOOTSTRAP" E REGRESSÃO	03
1.1.	O "bootstrap"	03
1.2.	Aplicação do "bootstrap" à Análise de Regressão	07
1.2.1.	As possíveis aplicações	07
1.2.2.	Obtenção da distribuição "bootstrap" do vetor de parâmetros $\underline{\beta}$	07
1.3.	Mínimos Quadrados Generalizados com matriz de co- variâncias estimada	13
1.3.1.	A estimação de $\underline{\beta}$	13
1.3.2.	Aplicação do "bootstrap"	14
II -	EXPERIMENTOS FATORIAIS COM REPLICAÇÕES CORRELACIONADAS	17
2.1.	Delineamento com duas replicações em cada ponto experimental	17
2.1.1.	O modelo	17
2.1.2.	As estimativas dos parâmetros em duas hipó- teses	19
2.1.3.	As variâncias dos estimadores dos parâmetros em duas hipóteses	22
2.1.4.	O fator de correção dos resíduos	24
2.1.5.	Determinação do estimador não viciado para σ^2 no modelo 2.1.1. a partir dos resíduos de um ajuste por MQO	27
2.1.6.	A obtenção das estimativas de σ^2 e ρ usando-se os resíduos de ajustes por MQO para as duas replicações, separadamente	28
2.1.7.	A obtenção de estimativa de $\underline{\beta}$ por meio de ajustes separados com as replicações	37
2.2.	Procedimentos propostos para MQG-1 e "bootstrap"	38
2.2.1.	A abordagem "bootstrap"	38

2.2.2. A abordagem MQG-1 Corrigido.....	39
2.3. Possibilidades de generalização para delineamentos com mais de duas replicações em cada ponto experimental.....	40
III - SIMULAÇÃO EM UM EXPERIMENTO COM O DELINEAMENTO DE BOX PARA MQG-1 E "BOOTSTRAP".....	
3.1. O delineamento de Box(Cubo+Estrela).....	42
3.2. Descrição da Simulação Monte Carlo para MQG-1 e "bootstrap".....	45
3.3. Análise comparativa da performance da abordagem usual por mínimos quadrados generalizados com o "bootstrap", no delineamento composto Cubo+Estrela	47
3.3.1. Resíduos têm distribuição normal.....	47
3.3.2. Situação em que os resíduos provêm de uma mistura de $N(0,1)$ com $N(0,4)$ em igual proporção.....	49
3.4. Conclusão.....	50
IV - APLICAÇÃO A UM CASO REAL.....	
4.1. O problema.....	51
4.2. O ajustamento.....	53
4.3. A ambigüidade entre correlação e desajustamento.....	55
4.4. Estimativas e desvios padrão dos parâmetros.....	57
4.5. Conclusão.....	59
APÊNDICE I.....	60
APÊNDICE II.....	68
REFERÊNCIAS.....	73

INTRODUÇÃO

Este trabalho tem por objetivo trazer alguma contribuição ao estudo das aplicações do "bootstrap", um procedimento estatístico computacionalmente intensivo, de reamostragem, relacionado ao "jackknife".

Um problema que ocorre freqüentemente em experimentação é aquele em que o experimentador, por descuido ou impossibilidade técnica, faz observações de um determinado experimento replicado gerando correlação entre as 2 replicações. Se um dos objetivos do trabalho estatístico for o ajuste do modelo de regressão $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, pode ocorrer do experimentador, desconhecendo a correlação criada, admitir para os resíduos, uma matriz de covariâncias da forma

$$\Sigma = I_{2n} \cdot \sigma^2,$$

em vez de

$$\Sigma = I_n \cdot \sigma^2 \otimes \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix};$$

onde \otimes é o produto de Kronecker; e portanto com sérias consequências na análise posterior dos dados.

Neste contexto desenvolve-se este trabalho em quatro capítulos, onde se faz a exposição teórica; e dois apêndices, nos

quais se descreve os principais programas computacionais necessários ao trabalho. Primeiramente, procura-se apresentar o "bootstrap" no contexto da análise de regressão e faz-se uma ilustração da distribuição "bootstrap" de uma estatística $\hat{\beta}_1$. As limitações do método dos mínimos quadrados generalizados com matriz de variâncias-covariâncias estimada (MQG-1) são enfocadas, bem como a possibilidade de aplicação do "bootstrap". Em seguida, aborda-se o problema da correlação entre replicações em experimento fatorial. Um fator de correção para os resíduos de ajuste clássico por MQG-1 é determinado, assim como estimadores da variância e coeficiente de correlação entre as replicações considerando ajustes separados por mínimos quadrados ordinários (MQO). Procedimento alternativo que traz significativa melhoria na precisão dos resultados por MQG-1 e "bootstrap" é proposto. No terceiro capítulo é feita uma descrição da simulação empregada para efeito de comparação entre a abordagem usual por MQG-1 e o "bootstrap" em um delineamento composto cubo+estrela. A metodologia abordada nos primeiros capítulos é aplicada em um caso real no quarto capítulo. Finalmente, apresenta-se nos dois apêndices os programas em FORTRAN utilizados para a simulação e para resolver o problema real.

1. "BOOTSTRAP" E REGRESSÃO

1.1. O "BOOTSTRAP"

O "bootstrap" é uma técnica estatística relacionada ao "jackknife", introduzida recentemente (Efron [1]), e que tem se mostrado versátil no tratamento de uma variedade de problemas estatísticos. A partir de Efron [1979] tem se verificado um interesse crescente pelo tópic, atestado pelo grande número de artigos publicados que exploram essa técnica. O fato que viabilizou a idéia do "bootstrap" foi o surgimento da computação veloz e barata, de modo que o tempo e custo para a execução de complexas operações computacionais se tornaram reduzidos. O "bootstrap" é um procedimento de reamostragem, computacionalmente intensivo, para estimação de distribuições amostrais. De forma sumária este procedimento pode ser descrito como segue:

- i. Seja $\underline{X} = (X_1, X_2, \dots, X_n)$ uma amostra aleatória com X_i , $i = 1, 2, \dots, n$, iid $\sim F$, sejam $\{\hat{T}_n(\underline{X}), n = 1, 2, \dots\}$ estatísticas reais, e $\{T_n(F), n = 1, 2, \dots\}$, funcionais em F .
- ii. Considere que

$$\sqrt{n}(\hat{T}_n(\underline{X}) - T_n(F)) \xrightarrow{D} N(0, \sigma_T^2)$$

e seja $H_n(x, F)$ a distribuição exata de $\sqrt{n}(\hat{T}_n(\underline{X}) - T_n(F))$.

Note que um dos problemas básicos em estatística consiste em estimar $H_n(x, F)$, uma vez que sua esperança e variância correspondem ao vício e a variância, respectivamente, de \hat{T}_n , como estimador de $T_n(F)$.

iii. Sendo F desconhecida, seja \hat{F} o estimador não paramétrico de mínimos quadrados de F , ou seja, a distribuição que coloca massa igual a n^{-1} em cada ponto da amostra. O método "bootstrap" consiste em estimar $H_n(x, F)$, (ou funcionais de $H_n(x, F)$), por $H_n(x, \hat{F})$, (ou funcionais de $H_n(x, \hat{F})$). Na prática, constrói-se $H_n(x, \hat{F})$ por Monte Carlo, com um número n^* suficientemente grande de repetições. Isto é, retiram-se de x_1, x_2, \dots, x_n , os valores observados da amostra, amostras aleatórias simples, com reposição, de tamanho n , chamadas amostras "bootstrap", ou reamostras, $x_1^*, x_2^*, \dots, x_n^*$. Para cada reamostra $\underline{x}^* = (x_1^*, x_2^*, \dots, x_n^*)$ calcula-se $\hat{T}_n^* = \hat{T}_n(\underline{x}^*)$.

O histograma do conjunto de valores $\hat{T}_n(\underline{x}_1^*), \hat{T}_n(\underline{x}_2^*), \dots, \hat{T}_n(\underline{x}_{n^*}^*)$ dá uma visão da forma da distribuição.

EXEMPLO 1.1.1.

Os dados abaixo (Jerrold H.Zar, [8] págs. 248-249) mostram o soro de colesterol (mg/100 ml) e o depósito arterial de cálcio (mg/100g de tecido) em 12 animais. A estatística real $\hat{\rho}$ que estima o funcional

$$\rho = \frac{\int (x - E(x))(y - E(y)) dF_{xy}(x, y)}{\left[\int (x - E(x))^2 dF_X(x) \cdot \int (y - E(y))^2 dF_Y(y) \right]^{\frac{1}{2}}}$$

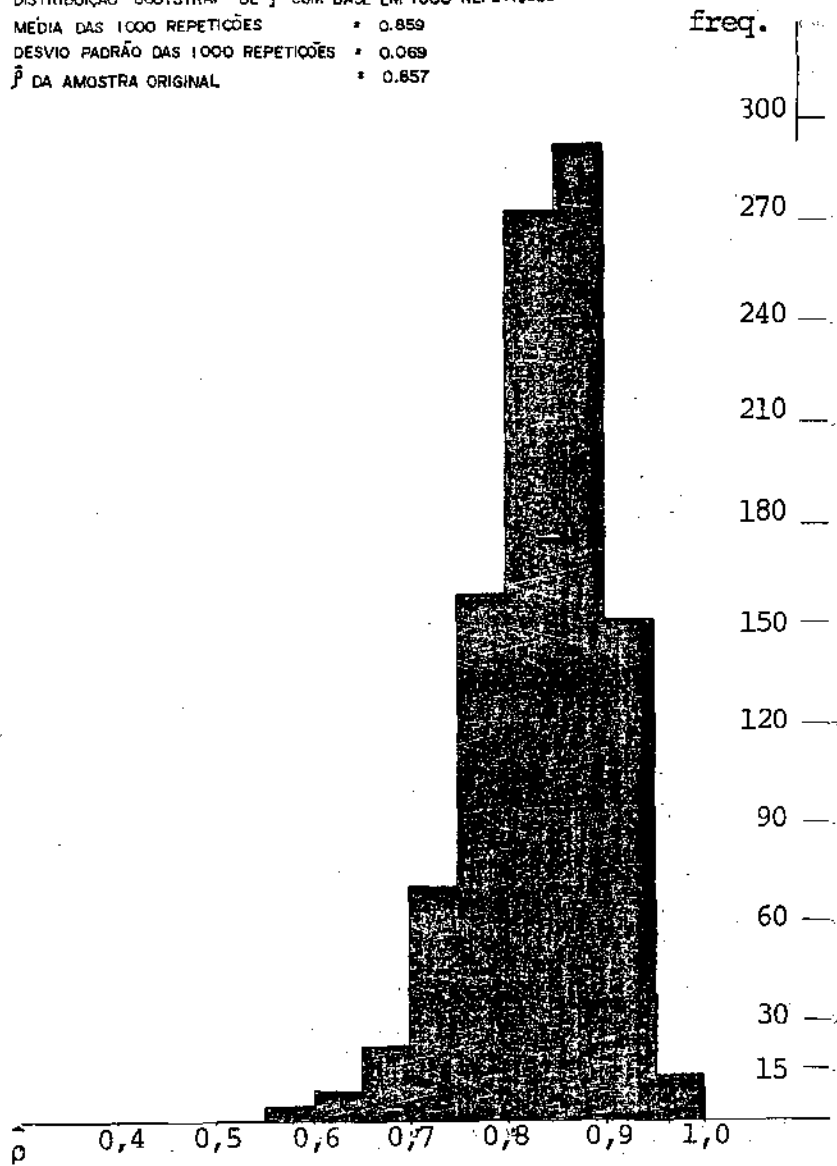
forneceu o seguinte valor

$$\hat{\rho} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\left[\left(\sum_{i=1}^n (x_i - \bar{x})^2 \right) \left(\sum_{i=1}^n (y_i - \bar{y})^2 \right) \right]^{\frac{1}{2}}} = 0,857;$$

a distribuição "bootstrap" da estatística $\hat{\rho}$ com base em $n^*=1000$ repetições "bootstrap" apresentou o valor médio $\hat{\rho}_{\text{médio}}^* = 0,859$ e um desvio padrão de 0,069. O histograma do conjunto de estimativas, $\hat{\rho}^*$ de ρ , é dado na figura nº1 e mostra a forma da distribuição "bootstrap", que estima $H_n(x, F)$, para $\hat{\rho}$.

CÁLCIO (X)	59	52	42	59	24	24	40	32	63	57	36	24
COLESTEROL (Y)	298	303	233	287	236	245	265	233	286	290	264	239

FIGURA Nº1
 COEFICIENTE DE CORRELAÇÃO
 DISTRIBUIÇÃO "BOOTSTRAP" DE $\hat{\rho}$ COM BASE EM 1000 REPETIÇÕES
 MÉDIA DAS 1000 REPETIÇÕES * 0.859
 DESVIO PADRÃO DAS 1000 REPETIÇÕES * 0.069
 $\hat{\rho}$ DA AMOSTRA ORIGINAL * 0.857



1.2. APLICAÇÃO DO "BOOTSTRAP" À ANÁLISE DE REGRESSÃO

1.2.1. AS POSSÍVEIS APLICAÇÕES

A aplicação do "bootstrap" à análise de regressão teve início com Efron [1], seguido de Freedman [2,3]. A aplicação maior do "bootstrap" nesta área parece ser no caso de resultados assintóticos. Assim, mínimos quadrados generalizados com matriz de covariância estimada, modelos com variáveis defazadas, erros não normais, são os problemas de regressão onde o "bootstrap", independentemente da distribuição dos resíduos, poderá ser aplicado para determinação de variâncias, intervalos de confiança, etc.

1.2.2. OBTENÇÃO DA DISTRIBUIÇÃO "BOOTSTRAP" DO VETOR DE PARÂMETROS $\underline{\beta}$

Basicamente a obtenção da distribuição "bootstrap" para o vetor $\underline{\beta}$ no modelo abaixo é feita conforme os passos seguintes:

1º - Seja o modelo $\underline{Y} = X\underline{\beta} + \underline{\varepsilon}$ com o vetor de respostas \underline{Y} e vetor de erros $\underline{\varepsilon}$, ambos com dimensão n , vetor de parâmetros $\underline{\beta}$ com dimensão $p+1$, e assumindo que a matriz de planejamento X de ordem $(n, p+1)$ não é aleatória;

2º - Os erros são iid com distribuição desconhecida F , sendo

que se assume que F está centrada em zero de alguma forma, assim

$$\varepsilon_i \sim \text{iid}^F$$

com $E_F(\varepsilon_i) = 0$ ou $Me_F(\varepsilon_i) = 0$, p.ex.;

3º - Com a matriz de planejamento X especificada e usando-se algum critério, p.ex. mínimos quadrados, faz-se um ajustamento ao modelo obtendo-se $\hat{\beta}$ como estimativa de β ,

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

4º - O vetor de resíduos e é determinado

$$e = Y - X\hat{\beta};$$

5º - Uma distribuição empírica de probabilidade \hat{F} , para os resíduos corrigidos pela média é estabelecida, e deste modo \hat{F} associa a probabilidade $1/n$ para cada e_i ,

$$e_i \sim \text{iid}^{\hat{F}};$$

6º - Uma amostra "bootstrap", e^* , é tomada de \hat{F} ,

$$\underline{e}^* = (e_1^*, e_2^*, e_3^*, \dots, e_n^*);$$

7º - Um vetor de pseudo-dados \underline{y}^* é composto,

$$\underline{y}^* = X\hat{\beta} + \underline{e}^* ;$$

8º - Um ajuste é feito aos pseudo-dados e a estimativa "bootstrap" $\hat{\beta}^*$ de β é determinada, pela mesma técnica de ajuste,

$$\hat{\beta}^* = (X'X)^{-1}X'\underline{y}^*$$

9º - A distribuição "bootstrap" de $\hat{\beta}^*$ é uma aproximação da distribuição amostral de $\hat{\beta}$.

É importante observar que $\hat{\beta}^* = (X'X)^{-1}X'\underline{y}^*$ tem média e variância "bootstrap" dadas por

$$E_*(\hat{\beta}^*) = \hat{\beta} \quad \text{e} \quad \text{Cov}_*(\hat{\beta}^*) = \hat{\sigma}^2 (X'X)^{-1},$$

pois os e_i^* são independentes com média zero e variância $\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2$ e assim

$$\begin{aligned} E_*(\hat{\beta}^*) &= E_*[(X'X)^{-1}X'\underline{y}^*] = (X'X)^{-1}X'E_*(\underline{y}^*) \\ &= (X'X)^{-1}X' \cdot E_*(X\hat{\beta} + \underline{e}^*) \end{aligned}$$

$$= (X'X)^{-1}X'[\underline{\hat{\beta}} + \underline{0}]$$

$$E_*(\underline{\hat{\beta}}^*) = \underline{\hat{\beta}}$$

e

$$\text{Cov}_*(\underline{\hat{\beta}}^*) = E_*[(\underline{\hat{\beta}}^* - \underline{\hat{\beta}})(\underline{\hat{\beta}}^* - \underline{\hat{\beta}})'] =$$

$$= E_*[(X'X)^{-1}X'\underline{\hat{y}}^* - (X'X)^{-1}X'\underline{\hat{y}}](X'X)^{-1}X'\underline{\hat{y}}^* - (X'X)^{-1}X'\underline{\hat{y}})']$$

$$= E_*[(X'X)^{-1}X'(\underline{\hat{\beta}} + \underline{e}^* - X\underline{\beta} - \underline{\varepsilon})(X'X)^{-1}X'(\underline{\hat{\beta}} + \underline{e}^* - X\underline{\beta} - \underline{\varepsilon}))']$$

$$= E_*[(\underline{\hat{\beta}} + (X'X)^{-1}X'\underline{e}^* - \underline{\hat{\beta}} - (X'X)^{-1}X'\underline{\varepsilon})(\underline{\hat{\beta}} + (X'X)^{-1}X'\underline{e}^* - \underline{\hat{\beta}} - (X'X)^{-1}X'\underline{\varepsilon})']$$

$$= E_*[(\underline{\hat{\beta}} + (X'X)^{-1}X'\underline{e}^* - \underline{\hat{\beta}})(\underline{\hat{\beta}} + (X'X)^{-1}X'\underline{e}^* - \underline{\hat{\beta}})']$$

$$= E_*[(X'X)^{-1}X'\underline{e}^* \cdot \underline{e}^{*'}X(X'X)^{-1}]$$

$$= (X'X)^{-1}X'E_*(\underline{e}^*\underline{e}^{*'})X(X'X)^{-1}$$

$$= E_*(\underline{e}^*\underline{e}^{*'}) \cdot (X'X)^{-1}$$

$$\text{Cov}_*(\underline{\hat{\beta}}^*) = \hat{\sigma}^2 (X'X)^{-1}$$

EXEMPLO 1.2.1.

Os dados, pág.12, (Draper & Smith,[6], pág.66) provêm de um experimento onde cristais de gelo são introduzidos em uma câmara cujo interior é mantido em uma temperatura fixa (-5°C) e com um nível fixo de saturação do ar com água. O crescimento dos cristais com o tempo é observado. As 43 medidas apresentadas correspondem ao comprimento axial dos cristais (A) em micrômetros por períodos (T) de 50 até 180 segundos após a introdução dos cristais. Cada medida representa um experimento simples completo; os experimentos são aleatorizados quanto ao tempo observado. A reta de mínimos quadrados ajustada aos dados é

$$\hat{A} = 14,411 + 0,131 T$$

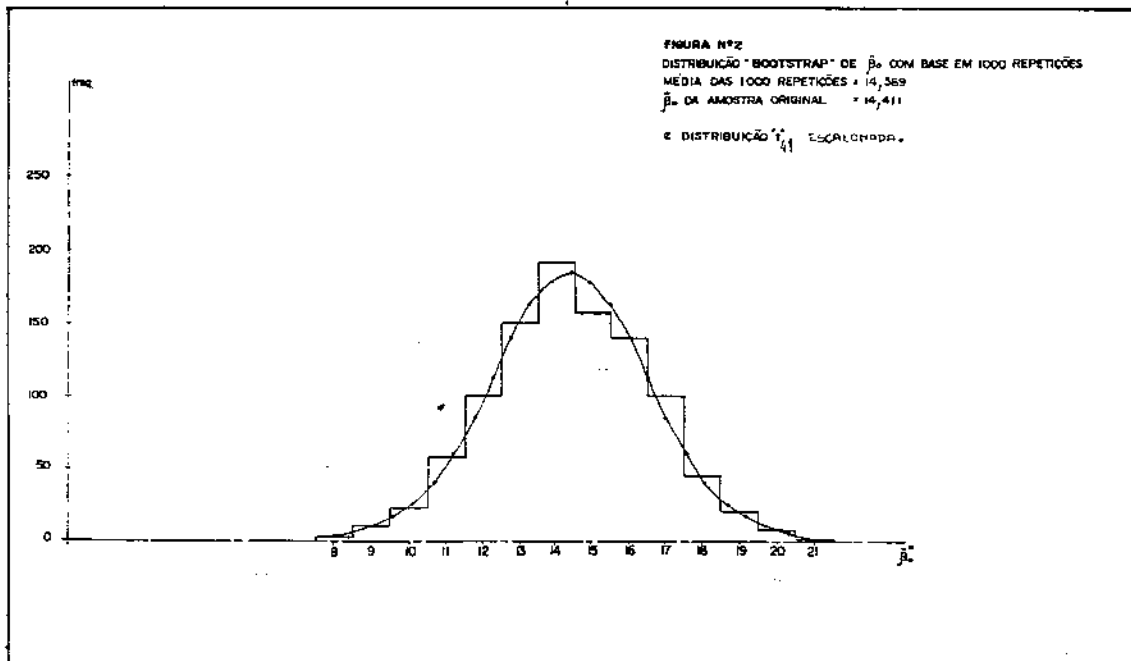
e os desvios padrão das estimativas dos coeficientes são 2,1490 e 0,01760, respectivamente para $\hat{\beta}_0$ e $\hat{\beta}_1$. Considerando-se a distribuição empírica \hat{F} que associa 1/43 para cada ponto observado, aplicou-se o procedimento "bootstrap" conforme descrito em 1.2.2., com um número de repetições $n^* = 1000$, obtendo-se

$$\hat{\beta}_0^* \text{ médio} = 14,389 \text{ com } s_0 = 2,14 \quad e$$

$$\hat{\beta}_1^* \text{ médio} = 0,131 \text{ com } s_1 = 0,0174$$

A figura nº2 mostra o histograma das 1000 estimativas "bootstrap" de β_0 , que formam uma aproximação da distribuição amostral de $\hat{\beta}_0$; e também para efeito de comparação a curva da distribuição "t" de Student. Na comparação observa-se que a aproximação é muito boa.

T	A	T	A
50	19	125	28
60	20,21	130	31,32
70	17,22	135	34,25
80	25,28	140	26,33
90	21,25,31	145	31
95	25	150	36,33
100	30,29,33	155	41,33
105	35,32	160	40,30,37
110	30,28,30	165	32
115	31,36,30	170	35
120	36,25,28	180	38



1.3. MÍNIMOS QUADRADOS GENERALIZADOS, COM MATRIZ DE COVARIÂNCIAS ESTIMADA

1.3.1. A ESTIMAÇÃO DE $\underline{\beta}$

Um problema muito comum na análise de regressão é a estimação do parâmetro $\underline{\beta}$ no modelo $\underline{Y} = \underline{X}\underline{\beta} + \underline{\varepsilon}$, com $E(\underline{\varepsilon}) = \underline{0}$ e $\text{Cov}(\underline{\varepsilon}) = \underline{\Sigma}$. Quando a matriz $\underline{\Sigma}$ é conhecida, o estimador de mínimos quadrados generalizados de $\underline{\beta}$ é dado por

$$\hat{\underline{\beta}}_{\text{mqg}} = (\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}\underline{X}'\underline{\Sigma}^{-1}\underline{Y} \quad (1.3.1)$$

que possui

$$\begin{aligned} E(\hat{\underline{\beta}}_{\text{mqg}}) &= \underline{\beta} \\ \text{Cov}(\hat{\underline{\beta}}_{\text{mqg}}) &= (\underline{X}'\underline{\Sigma}^{-1}\underline{X})^{-1}. \end{aligned} \quad (1.3.2)$$

Se $\underline{\Sigma}$ é desconhecida, como geralmente é o caso, a prática usual é aplicar um procedimento em dois estágios. Primeiro obtém-se um estimador consistente $\hat{\underline{\Sigma}}_0$ de $\underline{\Sigma}$ e a seguir usa-se (1.3.1) e (1.3.2) com $\hat{\underline{\Sigma}}_0$ no lugar de $\underline{\Sigma}$; sendo que $\hat{\underline{\Sigma}}_0$ é construída com os resíduos de um ajuste preliminar de mínimos quadrados ordinários. Este procedimento, contudo, é válido assintoticamente e em um contexto de amostra finita o resultado dependerá de $\hat{\underline{\Sigma}}_0$ ser uma estimativa boa

ou ruim de Σ . Freedman [3] aplicou um procedimento assintótico e o "bootstrap" em um caso de "seemingly unrelated regressions" a fim de comparar a performance dos dois procedimentos e constatou que o assintótico é muito otimista, embora no seu estudo não tenha feito correção nos resíduos para compensar a redução dos graus de liberdade pelo ajuste.

1.3.2. APLICAÇÃO DO "BOOTSTRAP"

A aplicação do "bootstrap" ao problema da regressão por mínimos quadrados generalizados com matriz de covariâncias estimada pode seguir, por exemplo, o seguinte esquema:

1º - de início é feito um ajuste por mínimos quadrados ordinários e se obtém a matriz $\hat{\Sigma}_0$ como estimativa de Σ ;

2º - em um segundo passo é feito um ajuste, agora por mínimos quadrados generalizados, e se obtém o vetor de resíduos \underline{e} dado por $\underline{e} = \underline{Y} - X\hat{\beta}_{mqg}$, onde $\hat{\beta}_{mqg} = (X'\hat{\Sigma}_0^{-1}X)^{-1}X'\hat{\Sigma}_0^{-1}\underline{Y}$;

3º - iniciando o "bootstrap", toma-se da amostra original \underline{e} uma amostra "bootstrap" \underline{e}^* e compõe-se um vetor de pseudo-dados $\underline{y}^* = X\hat{\beta}_{mqg} + \underline{e}^*$;

4º - um novo ajuste por mínimos quadrados generalizados é feito, agora aos pseudo-dados \underline{y}^* , e a estimativa "bootstrap"

$\hat{\beta}^*$ de β é obtida por

$$\hat{\beta}^* = (X' \hat{\Sigma}_0^{-1} X)^{-1} X' \hat{\Sigma}_0^{-1} \underline{y}^* ;$$

5º - nova amostra "bootstrap" é tomada da original, novos pseudo-dados são compostos, é feito ajuste por mínimos quadrados generalizados da mesma forma, e outra estimativa "bootstrap" de β é conseguida. Assim, repetindo o mesmo procedimento, um número, n^* , suficientemente grande, tem-se o conjunto de n^* vetores $\{\hat{\beta}_1^*, \hat{\beta}_2^*, \hat{\beta}_3^*, \dots, \hat{\beta}_{n^*}^*\}$, e, então, é possível determinar variâncias, intervalos de confiança e outras estatísticas "bootstrap" dos parâmetros.

Três observações devem ser feitas no procedimento:

- 1a) a estrutura de correlação existente entre os resíduos deve ser conservada quando se obtêm as amostras "bootstrap";
- 2a) a correção dos resíduos do ajuste de mínimos quadrados generalizados para compensar a perda de graus de liberdade deve ser feita, sempre que possível;
- 3a) quando na matriz de planejamento X não existe a 1a. coluna de 1's (ausência de intercepto no modelo), os resíduos que serão reamostrados devem estar centrados em \bar{u} , para se conseguir a distribuição empírica \hat{F} , adequada

da, para e_i , sendo que $\hat{\mu}$ é a média dos resíduos.

Uma aplicação a um caso real é feita no Capítulo IV.

2. EXPERIMENTOS FATORIAIS COM REPLICAÇÕES CORRELACIONADAS

2.1. DELINEAMENTO COM DUAS REPLICAÇÕES EM CADA PONTO EXPERIMENTAL

2.1.1. O MODELO

Quando se faz um experimento é comum a existência de várias fontes de erro que podem influenciar as variáveis em estudo. Certas fontes de erro interferentes podem ser controladas por alguma técnica específica da área, algumas não podem ser controladas, e outras podem ser controladas por técnicas estatísticas. A aleatoriedade e a replicação das unidades amostrais são princípios básicos do planejamento experimental, e um experimento realizado sem seguir estas premissas pode produzir dados que não cumprem as suposições necessárias à aplicação de certos procedimentos estatísticos de análise de dados. Uma situação que pode ocorrer, é aquela em que se faz um delineamento com duas repetições em cada um dos n tratamentos, mas com as duas unidades amostrais sendo obtidas da mesma unidade experimental. Conquanto as unidades experimentais possam estar aleatorizadas, o mesmo não ocorre com as unidades amostrais, e assim uma correlação ρ entre as duas replicações pode estar sendo gerada, embora mantendo-se independentes os pares de replicações.

Seja o delineamento definido pela matriz de planejam-

to X de ordem $(2n, p)$, onde para cada um dos n pontos experimentais foram feitas duas replicações e seja ρ a correlação gerada entre as duas replicações. O modelo de regressão que se pretende ajustar aos dados é

$$E(\underline{y}) = X\underline{\beta} \quad (2.1.1)$$

$$\text{com } \underline{y} = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_{2n} \end{bmatrix}, \quad \underline{\beta} = \begin{bmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_{p-1} \end{bmatrix},$$

$$\text{e } 2n \times p = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np-1} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix};$$

\underline{y} é o vetor das respostas observadas e pode ser escrito na forma

$$\underline{y} = X\underline{\beta} + \underline{\varepsilon}, \text{ onde } \underline{\varepsilon} = \begin{bmatrix} \varepsilon_{11} \\ \varepsilon_{12} \\ \varepsilon_{21} \\ \varepsilon_{22} \\ \vdots \\ \varepsilon_{n1} \\ \varepsilon_{n2} \end{bmatrix},$$

X é uma matriz fixa, de planejamento; e $\underline{\varepsilon}$ é um vetor aleatório,

dos resíduos, com

$$E(\underline{\varepsilon}) = \underline{0} ;$$

os $\varepsilon_{ij}, i=1, \dots, n \quad j=1, 2$ têm distribuição comum F ; $\varepsilon_{1.}$ são independentes, mas $\varepsilon_{.j}$ têm correlação ρ . A matriz de variâncias covariâncias de $\underline{\varepsilon}$ é dada por:

$$V(\underline{\varepsilon}) = E[\underline{\varepsilon} - E(\underline{\varepsilon})][\underline{\varepsilon} - E(\underline{\varepsilon})]'$$

$$V(\underline{\varepsilon}) = E(\underline{\varepsilon} \underline{\varepsilon}') = \Sigma$$

$$\text{com } {}_{2n}\Sigma_{2n} = \sigma^2 \cdot I_n \otimes \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} \quad (2.1.2)$$

onde \otimes é o produto de Kronecker (ver [7]) e a matriz de correlação é

$$\Sigma_c = I_n \otimes \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}. \quad (2.1.3)$$

2.1.2. AS ESTIMATIVAS DOS PARÂMETROS EM DUAS HIPÓTESES

No modelo (2.1.1) pode ocorrer do experimentador ignorar a existência de correlação entre as replicações, admitir então

$${}_{2n}\Sigma_{2n} = I_{2n} \cdot \sigma^2 \quad (2.1.4)$$

e usar nesta hipótese o estimador de Mínimos Quadrados Ordinários (UMVU e também EMV se $\underline{\varepsilon}$ é normalmente distribuído) de $\underline{\beta}$

$$\hat{\underline{\beta}}_{mqo} = (X'X)^{-1}X'\underline{y} ;$$

quando na realidade deveria usar, na hipótese correta de

$$2n\Sigma 2n = \sigma^2 \cdot I_n \otimes \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix} ,$$

o estimador de Mínimos Quadrados Generalizados (UMVU e também EMV se $\underline{\varepsilon}$ é normalmente distribuído) de $\underline{\beta}$

$$\hat{\underline{\beta}}_{mqg} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{y} ;$$

contudo, nesta situação particular, temos o seguinte resultado:

TEOREMA 2.1. Seja o delineamento com replicação conforme descrito em 2.1.1. . Então,

$$\hat{\underline{\beta}}_{mqo} = \hat{\underline{\beta}}_{mqg} \quad (2.1.5.)$$

PROVA:

$$\hat{\underline{\beta}}_{mqo} = (X'X)^{-1}X'\underline{y} \quad \text{e} \quad \hat{\underline{\beta}}_{mqg} = (X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\underline{y} ,$$

$$\text{mas como } {}_{2n}X_p = \begin{bmatrix} 1 & x_{11} & x_{12} & \dots & x_{1p-1} \\ 1 & x_{21} & x_{22} & \dots & x_{2p-1} \\ \vdots & \vdots & \vdots & & \vdots \\ 1 & x_{n1} & x_{n2} & \dots & x_{np-1} \end{bmatrix} \otimes \begin{bmatrix} 1 \\ 1 \end{bmatrix}$$

tem as linhas iguais 2 a 2 devido as replicações

$$\text{e } {}_{2n}\Sigma_{2n}^{-1} = \frac{1}{\sigma^2} \cdot I_n \otimes \begin{bmatrix} 1/(1-\rho^2) & -\rho/(1-\rho^2) \\ -\rho/(1-\rho^2) & 1/(1-\rho^2) \end{bmatrix}$$

$$\Sigma^{-1} = \frac{1}{(1-\rho^2) \cdot \sigma^2} \cdot I_n \otimes \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix}$$

donde

$$X' \Sigma^{-1} = \frac{1}{(1-\rho^2) \sigma^2} \cdot \begin{bmatrix} 1 & 1 & 1 & \dots & 1 \\ x_{11} & x_{21} & x_{31} & \dots & x_{n1} \\ x_{12} & x_{22} & x_{32} & \dots & x_{n2} \\ \vdots & \vdots & \vdots & & \vdots \\ x_{1p-1} & x_{2p-1} & x_{3p-1} & \dots & x_{np-1} \end{bmatrix} \otimes \begin{bmatrix} 1 & 1 \end{bmatrix} \cdot (I_n \otimes \begin{bmatrix} 1 & -\rho \\ -\rho & 1 \end{bmatrix})$$

e assim

$$X' \Sigma^{-1} = \frac{1}{(1+\rho) \sigma^2} \cdot X' \quad (2.1.6.)$$

portanto

$$\hat{\beta}_{mqg} = (X' \Sigma^{-1} X)^{-1} X' \Sigma^{-1} \underline{y} = \left(\frac{1}{(1+\rho)\sigma^2} \cdot X'X \right)^{-1} \cdot \frac{1}{(1+\rho)\sigma^2} \cdot X' \underline{y}$$

$$\hat{\beta}_{mqg} = (X'X)^{-1} X' \underline{y} = \hat{\beta}_{mqo} \quad \text{c.q.d.}$$

2.1.3. AS VARIÂNCIAS DOS ESTIMADORES DOS PARÂMETROS EM DUAS HIPÓTESES.

Se o experimentador pretender construir intervalos de confiança para as estimativas dos parâmetros ou fazer alguma outra inferência estatística, envolvendo as variâncias, ainda ignorando a correlação entre as replicações, então os resultados produzidos podem ser distorcidos, uma vez que as variâncias estarão sendo subestimadas ou superestimadas. Na hipótese, errônea, de Σ dada por (2.1.4.) a matriz de covariâncias das estimativas dos parâmetros será tomada como

$$E[(\hat{\beta}_{mqo} - \beta)(\hat{\beta}_{mqo} - \beta)'] = \sigma^2 (X'X)^{-1} ;$$

enquanto que na hipótese correta, de existência de correlação com Σ dada por (2.1.2), temos o seguinte resultado:

TEOREMA 2.2. Seja o delineamento com replicação conforme descrito em 2.1.1.. Então

$$E[(\hat{\beta}_{mqg} - \beta)(\hat{\beta}_{mqg} - \beta)'] = (1 + \rho)\sigma^2(X'X)^{-1} \quad (2.1.7.).$$

PROVA

$$\begin{aligned} E[(\hat{\beta}_{mqg} - \beta)(\hat{\beta}_{mqg} - \beta)'] &= E\{[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y - \beta][(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}Y - \beta]'\} \\ &= E\{[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(X\beta + \varepsilon) - \beta][(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}(X\beta + \varepsilon) - \beta]'\} \\ &= E\{[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon][(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon]'\} \\ &= E\{[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon][(X'\Sigma^{-1}\varepsilon)'\Sigma^{-1}(X'\Sigma^{-1}X)^{-1}]\} \\ &= E\{[(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}\varepsilon][\varepsilon'\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}]\} \\ &= [(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1}E(\varepsilon\varepsilon')\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}] \\ &= [(X'\Sigma^{-1}X)^{-1}X'\Sigma^{-1} \cdot \Sigma\Sigma^{-1}X(X'\Sigma^{-1}X)^{-1}] \\ &= (X'\Sigma^{-1}X)^{-1} \end{aligned}$$

mas usando o resultado (2.1.6.), temos

$$E[(\hat{\beta}_{mqg} - \beta)(\hat{\beta}_{mqg} - \beta)'] = \left(\frac{1}{(1 + \rho)\sigma^2} \cdot X'X \right)^{-1}$$

$$= (1 + \rho) \sigma^2 \cdot (X'X)^{-1} \quad \text{c.q.d.}$$

2.1.4. O FATOR DE CORREÇÃO DOS RESÍDUOS

Quando um modelo é ajustado aos dados e se faz estimativa da variância usando-se os resíduos do ajustamento, deve-se pré-multiplicar os resíduos por um fator de correção que compense a perda dos graus de liberdade, de modo a se conseguir estimativa não viciada para a variância. No caso do modelo clássico de regressão

$$\underline{Y} = X\underline{\beta} + \underline{\varepsilon} \quad , \quad E(\underline{\varepsilon}) = 0, \quad V(\underline{\varepsilon}) = I_n \sigma^2$$

com estimação de $\underline{\beta}$ por mínimos quadrados ordinários, o fator de correção $[n/(n-p)]^{\frac{1}{2}}$ é adequado, uma vez que

$$E \left[\sum_{i=1}^n e_i^2 \right] = (n-p) \sigma^2$$

onde n é a dimensão de \underline{Y} e p a de $\underline{\beta}$. Contudo, para situações mais complexas, não é clara a escolha do fator de correção, como nota Freedman [3]. Aqui determinamos o fator de correção para os erros, correlacionados, gerados pelo ajuste do modelo (2.1.1.), bem como um estimador não viciado para a variância σ^2 .

No modelo (2.1.1.), objeto deste estudo, o fator de correção dos resíduos obtidos de um ajuste por MQO, e portanto, não levando em conta a correlação entre replicações, é dado pelo resultado abaixo.

TEOREMA 2.3. Seja o delineamento com replicação conforme descrito em 2.1.1.. O fator de correção dos graus de liberdade que deve ser aplicado aos resíduos quando estes são originados de um ajuste por MQO, a fim de se obter estimativa não viciada de σ^2 é

$$\phi = \sqrt{\frac{2n}{2n - (1 + \rho)p}} \quad (2.1.8)$$

PROVA

Considerando que os resíduos são obtidos admitindo-se (2.1.4.), a SQR é dada por

$$SQR = \underline{\underline{Y}}' \underline{\underline{Y}} - \hat{\underline{\underline{\beta}}} ' \underline{\underline{X}}' \underline{\underline{Y}}$$

e a esperança por

$$\begin{aligned} E(SQR) &= E(\underline{\underline{Y}}' \underline{\underline{Y}} - \hat{\underline{\underline{\beta}}} ' \underline{\underline{X}}' \underline{\underline{Y}}) \\ &= E[\underline{\underline{Y}}' \underline{\underline{Y}} - ((\underline{\underline{X}}' \underline{\underline{X}})^{-1} \underline{\underline{X}}' \underline{\underline{Y}})' \underline{\underline{X}}' \underline{\underline{Y}}] \end{aligned}$$

$$\begin{aligned}
&= E[\tilde{Y}'\tilde{Y} - \tilde{Y}'X(X'X)^{-1}X'\tilde{Y}] \\
&= E\{\tilde{Y}'[I_{2n} - X(X'X)^{-1}X']\tilde{Y}\}
\end{aligned}$$

a expressão entre chaves é uma forma quadrática, logo (ver[4], pág.55)

$$\begin{aligned}
E(SQR) &= \text{tr} \{ \Sigma [I_{2n} - X(X'X)^{-1}X'] + \beta'X'[I_{2n} - X(X'X)^{-1}X']X\beta \} \\
&= \text{tr} [\Sigma - \Sigma X(X'X)^{-1}X'] + 0 \\
&= \text{tr}(\Sigma) - \text{tr}[X(X'X)^{-1}X'\Sigma]
\end{aligned}$$

e de (2.1.6.) $X'\Sigma = (1 + \rho)\sigma^2 X'$, então

$$E(SQR) = \sigma^2 \cdot \text{tr}(\Sigma_C) - (1 + \rho)\sigma^2 \text{tr}[X(X'X)^{-1}X'],$$

onde Σ_C é como definido em (2.1.3.)

$$E(SQR) = [2n - (1 + \rho)p] \sigma^2,$$

portanto

$$E\left(\frac{SQR}{2n}\right) = \frac{[2n - (1 + \rho)p]}{2n} \cdot \sigma^2$$

e

$$\phi = \sqrt{\frac{2n}{2n - (1+\rho)p}}$$

c.q.d.

2.1.5. DETERMINAÇÃO DO ESTIMADOR NÃO VICIADO PARA σ^2 NO MODELO
(2.1.1.) A PARTIR DOS RESÍDUOS DE UM AJUSTE MQO

Uma consequência imediata do Teorema 2.3. é a obtenção de um estimador não viciado para σ^2 , dado pelo resultado abaixo.

COROLÁRIO 2.1. Nas condições do Teorema 2.3., o estimador não viciado de σ^2 é dado pelo resultado

$$s_p^2 = \left[\frac{2n - p}{2n - (1 + \rho)p} \right] \cdot s^2 \quad (2.1.9)$$

onde $s^2 = \sum_{i=1}^n \sum_{j=1}^2 (\bar{y}_{ij} - \hat{\bar{y}}_{ij})^2 / (2n - p)$ para o modelo (2.1.1.)

admitida (2.1.4.).

PROVA

$$\begin{aligned} E(s_p^2) &= \left[\frac{2n - p}{2n - (1 + \rho)p} \right] E(s^2) \\ &= \left[\frac{2n - p}{2n - (1 + \rho)p} \right] \cdot E \left[\frac{SQR}{2n - p} \right] \end{aligned}$$

$$= \frac{2n - p}{2n - (1 + \rho) p} \cdot \frac{1}{2n - p} \cdot [2n - (1 + \rho) p] \sigma^2$$

$$E(s_p^2) = \sigma^2$$

c.q.d.

2.1.6. A OBTENÇÃO DAS ESTIMATIVAS DE σ^2 E ρ USANDO-SE OS RESÍDUOS DE AJUSTES POR MQO PARA AS DUAS REPLICAÇÕES, SEPARADAMENTE.

Embora se tenha disponível um estimador não viciado, s_p^2 , de σ^2 , ocorre que ρ e σ^2 não são separadamente estimáveis no contexto de um único ajuste para as duas replicações na expressão (2.1.9.). Contudo estimativas independentes de σ^2 e ρ são necessárias. Um estimador não-viciado de σ^2 pode ser obtido com os resíduos de um ajuste por MQO com qualquer dos dois conjuntos de replicações. Denotando por \underline{Y}_1 e \underline{Y}_2 , respectivamente, os vetores de dimensão n das respostas na 1a. e 2a. replicação; e por \underline{e}_1 e \underline{e}_2 os vetores de dimensão n dos resíduos do ajuste por MQO com \underline{Y}_1 e \underline{Y}_2 , tem-se

$$\begin{aligned} E\left[\sum_{i=1}^n e_{i1}^2\right] &= E[\underline{e}_1' \underline{e}_1] = E\{\underline{Y}_1' [\underline{I}_n - X(X'X)^{-1}X'] [\underline{I}_n - X(X'X)^{-1}X'] \underline{Y}_1\} \\ &= E\{\underline{Y}_1' [\underline{I}_n - X(X'X)^{-1}X'] \underline{Y}_1\} \end{aligned}$$

a expressão entre chaves é uma forma quadrática e aqui neste item X é a matriz do modelo sem replicação de ordem (n,p) , então (Ver [4], pág.55).

$$E\left[\sum_{i=1}^n e_{i1}^2\right] = E\{\underline{Y}'_1 [I_n - X(X'X)^{-1}X'] \underline{Y}_1\} = \text{tr}\{I_n \sigma^2 [I_n - X(X'X)^{-1}X']\} + \underline{\beta}'X'[I_n - X(X'X)^{-1}X']X\underline{\beta}$$

$$E\left[\sum_{i=1}^n e_{i1}^2\right] = n\sigma^2 - p\sigma^2 + \underline{\beta}'X'X\underline{\beta} - \underline{\beta}'X'X\underline{\beta}$$

$$E\left[\sum_{i=1}^n e_{i1}^2\right] = (n-p)\sigma^2 \quad (2.1.10.)$$

e da mesma forma

$$E\left[\sum_{i=1}^n e_{i2}^2\right] = E[\underline{e}'_2 \underline{e}_2] = (n-p)\sigma^2, \quad (2.1.11.)$$

$$\text{assim} \quad \hat{\sigma}_1^2 = \sum_{i=1}^n e_{i1}^2 / (n-p) \quad (1a.\text{replicação}) \quad (2.1.12.)$$

$$\hat{\sigma}_2^2 = \sum_{i=1}^n e_{i2}^2 / (n-p) \quad (2a.\text{replicação}) \quad (2.1.13.)$$

e

$$E(\hat{\sigma}_1^2) = E(\hat{\sigma}_2^2) = \sigma^2$$

conseqüentemente uma estimativa não viciada de σ^2 é dada pelo resultado seguinte:

TEOREMA 2.4

Sejam \underline{e}_1 e \underline{e}_2 os vetores de dimensão n , dos resíduos oriundos de ajustes em $p-1$ variáveis exógenas, por MQO, respectivamente com as respostas da 1a. e da 2a. replicação do experimento; $\hat{\sigma}_1^2$ e $\hat{\sigma}_2^2$ as estimativas não viciadas da variância populacional σ^2 obtidas respectivamente das somas dos quadrados dos resíduos $\underline{e}_1' \underline{e}_1$ e $\underline{e}_2' \underline{e}_2$. Então uma estimativa não viciada de σ^2 com base nos resíduos dos dois ajustamentos é dada por

$$\hat{\sigma}^2 = \frac{\sum_{i=1}^n e_{i1}^2 + \sum_{i=1}^n e_{i2}^2}{2(n-p)} \quad (2.1.14.)$$

PROVA

$\hat{\sigma}^2$ é a média de duas estimativas não viciadas de σ^2 (2.1.12. e 2.1.13.), portanto

$$E(\hat{\sigma}^2) = \sigma^2 \quad \text{c.q.d.}$$

Considerando que os resíduos do 1º ajuste estão correlacionados com os do 2º, com coeficiente de correlação ρ , e que os erros $\underline{\varepsilon}_1$ e $\underline{\varepsilon}_2$, vetores de dimensão n , são identicamente distribuídos

$$N(\underline{0}, I_n \sigma^2)$$

e com matriz de covariâncias dada por

$$\text{Cov}(\underline{\varepsilon}_1, \underline{\varepsilon}_2) = \mathbf{I}_n \rho \sigma^2 ,$$

temos o seguinte resultado.

TEOREMA 2.5

Sejam \underline{e}_1 e \underline{e}_2 os vetores, de dimensão n , dos resíduos oriundos de ajustes com p variáveis exógenas, por MQO, respectivamente com as respostas \underline{y}_1 e \underline{y}_2 da 1a. e 2a. replicação do experimento. Então, com as considerações do parágrafo anterior, a esperança da soma dos produtos cruzados dos resíduos é dada por

$$E\left(\sum_{i=1}^n e_{i1} e_{i2}\right) = (n - p) \rho \sigma^2 \quad (2.1.15.)$$

PROVA

Os vetores dos resíduos são

$\underline{e}'_1 = (e_{11}, e_{21}, e_{31}, \dots, e_{n1})$ ajuste da 1a. replicação

$\underline{e}'_2 = (e_{12}, e_{22}, e_{32}, \dots, e_{n2})$ ajuste da 2a. replicação

$$\begin{aligned}
E\left[\sum_{i=1}^n e_{i1}e_{i2}\right] &= E(\underline{e}_1'\underline{e}_2) = E\{[(\underline{I}_n - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{y}_1]'[(\underline{I}_n - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}')\underline{y}_2]\} \\
&= E\{\underline{y}_1'[\underline{I}_n - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'][\underline{I}_n - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}']\underline{y}_2\} \\
&= E\{\underline{y}_1'[\underline{I}_n - \underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}']\underline{y}_2\} \\
&= E\{\underline{y}_1'\underline{y}_2\} - E[\underline{y}_1'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}_2] \\
&= E[(\underline{X}\underline{\beta} + \underline{\varepsilon}_1)'(\underline{X}\underline{\beta} + \underline{\varepsilon}_2)] - E[(\underline{X}\underline{\beta} + \underline{\varepsilon}_1)'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'(\underline{X}\underline{\beta} + \underline{\varepsilon}_2)] \\
&= E[\underline{\beta}'\underline{X}'\underline{X}\underline{\beta} + \underline{\beta}'\underline{X}'\underline{\varepsilon}_2 + \underline{\varepsilon}_1'\underline{X}\underline{\beta} + \underline{\varepsilon}_1'\underline{\varepsilon}_2] - E[\underline{\beta}'\underline{X}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{X}\underline{\beta} + \\
&\quad + \underline{\beta}'\underline{X}'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon}_2 + \underline{\varepsilon}_1'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{X}\underline{\beta} + \underline{\varepsilon}_1'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon}_2] \\
&= \underline{\beta}'\underline{X}'\underline{X}\underline{\beta} + 0 + 0 + n\rho\sigma^2 - \underline{\beta}'\underline{X}'\underline{X}\underline{\beta} - 0 - 0 - E[\underline{\varepsilon}_1'\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}'\underline{\varepsilon}_2]
\end{aligned}$$

a expressão entre colchetes é uma forma bilinear portanto (ver [4], pags.64-65).

$$\begin{aligned}
E\left[\sum_{i=1}^n e_{i1}e_{i2}\right] &= n\rho\sigma^2 - \text{tr}[\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}' \cdot \underline{I}_n\rho\sigma^2] - 0[\underline{X}(\underline{X}'\underline{X})^{-1}\underline{X}']0 \\
&= n\rho\sigma^2 - \rho\sigma^2 p \\
&= (n-p)\rho\sigma^2
\end{aligned}$$

c.q.d.

Os pares (e_{i1}, e_{i2}) , $i = 1, 2, 3, \dots, n$, com e_{i1} e e_{i2} obtidos, respectivamente, nos ajustes com as 1as. e 2as. replicações formam uma amostra de uma população

$$N(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$$

onde $\mu_1 = \mu_2 = 0$ e $\sigma_1^2 = \sigma_2^2 = \sigma^2$. O estimador de máximo verossimilhança de ρ pode ser obtido seguindo-se os passos:

1º - a família exponencial com 5 parâmetros é

$$\begin{aligned} p(e_1, e_2) = \{ \exp. [-n \ln(2\pi) - n \ln \sigma_1 \sigma_2 \sqrt{1 - \rho^2} + \left(\frac{\mu_1}{(1 - \rho^2)\sigma_1^2} - \rho \frac{\mu_2}{(1 - \rho^2)\sigma_1 \sigma_2} \right) \sum_{i=1}^n e_{i1} + \\ + \left(\frac{\mu_2}{(1 - \rho^2)\sigma_2^2} - \frac{\rho \mu_1}{(1 - \rho^2)\sigma_1 \sigma_2} \right) \sum_{i=1}^n e_{i2} - \frac{1}{2(1 - \rho^2)\sigma_1^2} \cdot \sum_{i=1}^n e_{i1}^2 - \\ - \frac{1}{2(1 - \rho^2)\sigma_2^2} \cdot \sum_{i=1}^n e_{i2}^2 + \frac{\rho}{(1 - \rho^2)\sigma_1 \sigma_2} \sum_{i=1}^n e_{i1} e_{i2} - \frac{n \mu_1^2}{2(1 - \rho^2)\sigma_1^2} - \\ - \frac{n \mu_2^2}{2(1 - \rho^2)\sigma_2^2} + \frac{n \rho \mu_1 \mu_2}{(1 - \rho^2)\sigma_1 \sigma_2}] \}. \end{aligned}$$

2º - tomando as estatísticas suficientes e considerando que,

C , o interior do campo de variação de $(C_1(\theta), C_2(\theta), C_3(\theta), C_4(\theta), C_5(\theta))$ é não vazio tem-se (ver [9] pág. 106)

$$E\left[\sum_{i=1}^n e_{i1}^2\right] = (n-p)\sigma_1^2 = \sum_{i=1}^n e_{i1}^2 \quad \therefore \hat{\sigma}_1^2 = \frac{1}{n-p} \sum_{i=1}^n e_{i1}^2,$$

$$E\left[\sum_{i=1}^n e_{i2}^2\right] = (n-p)\sigma_2^2 = \sum_{i=1}^n e_{i2}^2 \quad \therefore \hat{\sigma}_2^2 = \frac{1}{n-p} \sum_{i=1}^n e_{i2}^2$$

$$E\left[\sum_{i=1}^n e_{i1}e_{i2}\right] = (n-p)\rho\sigma_1\sigma_2 = \sum_{i=1}^n e_{i1}e_{i2} \quad \therefore \hat{\rho} = \frac{\sum_{i=1}^n e_{i1}e_{i2}}{(n-p)\sigma_1\sigma_2}$$

3º - aplicando o Princípio da Substituição (ver [9], pág.90) e usando-se os dois primeiros resultados do passo anterior obtém-se

$$\hat{\rho} = \frac{\sum_{i=1}^n e_{i1}e_{i2}}{\sqrt{\sum_{i=1}^n e_{i1}^2 \sum_{i=1}^n e_{i2}^2}} \quad (2.1.16.)$$

que é o estimador de máxima verossimilhança de ρ .

Na simulação feita com o delineamento de Box, Cap.III, encontrou-se para 250 repetições Monte Carlo

$$\hat{\sigma}_{\text{médio}}^2 = 0,96370,$$

e

$$\hat{\rho}_{\text{médio}} = 0,87923$$

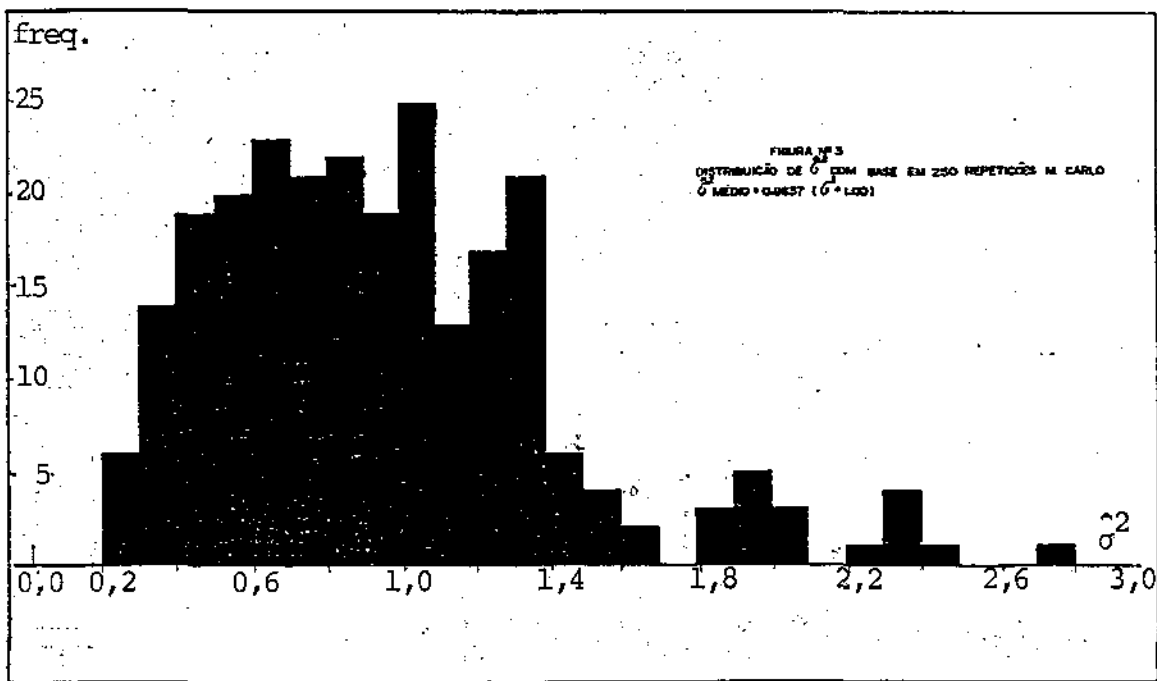
muito próximos dos verdadeiros valores

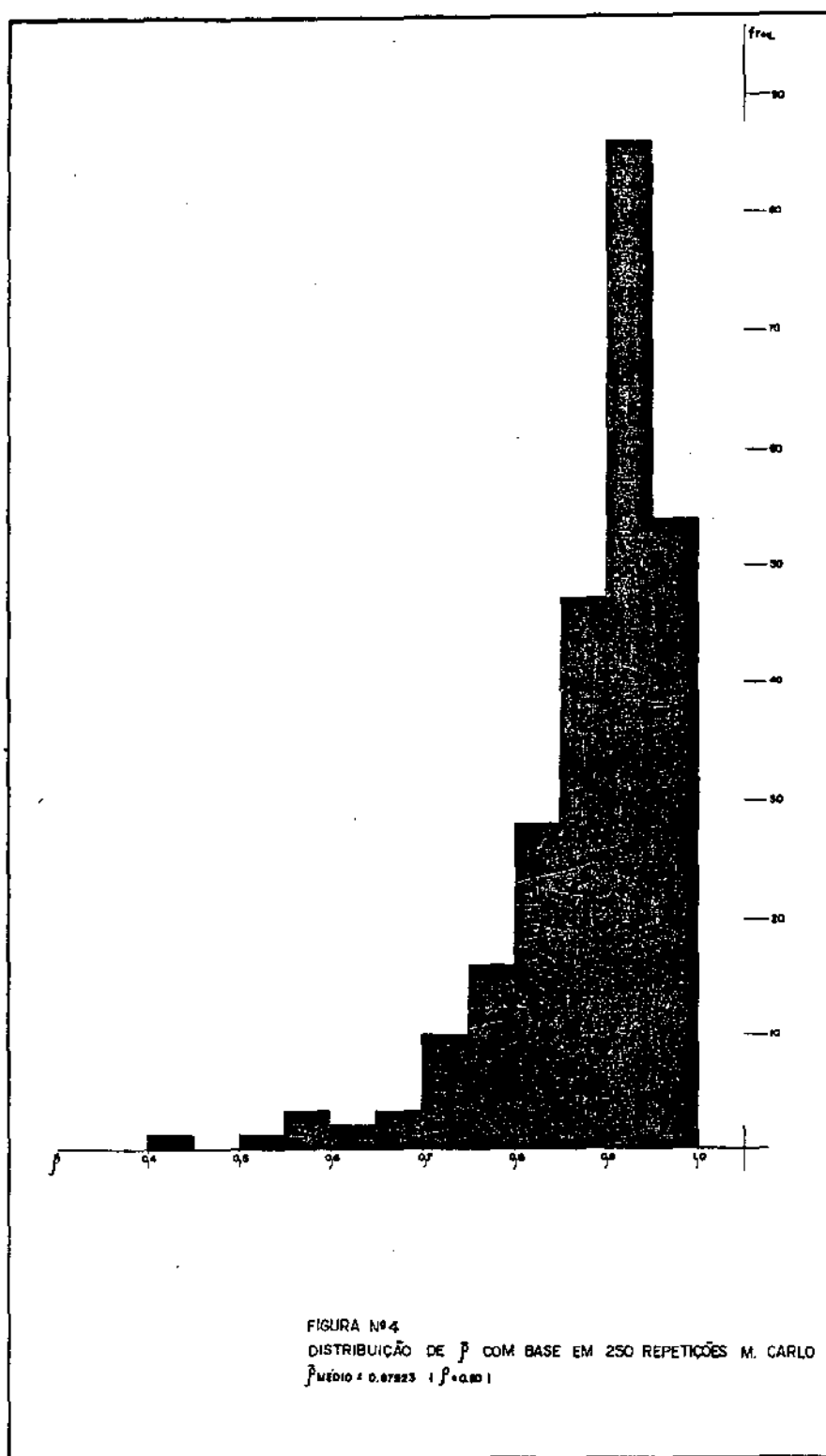
$$\sigma^2 = 1,00$$

e

$$\rho = 0,90;$$

o histograma dos 250 valores observados, para cada uma das estatísticas é dado pelas figuras nº3 e 4, e dá uma visão da forma da distribuição amostral em cada caso.





2.1.7. A OBTENÇÃO DE ESTIMATIVA DE β POR MEIO DE AJUSTES SEPARADOS COM AS REPLICAÇÕES

Quando fazemos os ajustes MQO separadamente com as primeiras e segundas replicações, a estimativa de MQO de β é dada pelo resultado:

TEOREMA 2.6.

Sejam \underline{b}_1 e \underline{b}_2 os vetores, de dimensão $p+1$ das estimativas do vetor de parâmetros β , resultantes de ajustes por MQO feitos separadamente com os vetores de dimensão n das respostas, \underline{y}_1 e \underline{y}_2 , da 1a. e 2a. replicação do experimento. Então, o vetor \underline{b} , de dimensão $p+1$ que se obtém em um ajuste por MQO com o vetor de respostas \underline{y} , de dimensão $2n$, é dado por

$$\underline{b} = \frac{1}{2} (\underline{b}_1 + \underline{b}_2) \quad (2.1.17.)$$

PROVA:

Os ajustes por MQO para as 1a. e 2a. replicações fornecem:

$$\underline{b}_1 = (X'X)^{-1} X' \underline{y}_1$$

$$\underline{b}_2 = (X'X)^{-1} X' \underline{y}_2$$

e o ajuste para todas as $2n$ observações fornece:

$$\underline{b} = (\underline{X}'\underline{X})^{-1}\underline{X}'\underline{y}$$

mas considerando que o ajuste com o vetor \underline{y} determina a mesma estimativa que o ajuste com o vetor $\bar{\underline{y}}$, de dimensão n , cujos elementos são

$$\bar{y}_1 = \frac{y_{11} + y_{12}}{2}, \bar{y}_2 = \frac{y_{21} + y_{22}}{2}, \dots, \bar{y}_n = \frac{y_{n1} + y_{n2}}{2},$$

tem-se

$$\underline{b} = (\underline{X}'\underline{X})^{-1}\underline{X}'\bar{\underline{y}} = (\underline{X}'\underline{X})^{-1}\underline{X}'\left(\frac{\underline{y}_1 + \underline{y}_2}{2}\right)$$

$$= \frac{1}{2}(\underline{X}'\underline{X})^{-1}\underline{X}'(\underline{y}_1 + \underline{y}_2)$$

$$= \frac{1}{2}(\underline{b}_1 + \underline{b}_2) \quad \text{c.q.d.}$$

2.2. PROCEDIMENTOS PROPOSTOS PARA MQG-1 E "BOOTSTRAP"

2.2.1. A ABORDAGEM "BOOTSTRAP"

A abordagem "bootstrap" do modelo 2.1.1. é feita com base em [2] e [3], mas considerando-se as características próprias do caso. Assim, os resíduos que são reamostrados estão centrados

em $\hat{\mu} = \frac{1}{2n} \sum_{i=1}^n \sum_{j=1}^2 \varepsilon_{ij} = 0$, visto que considera-se a existência

de intercepto no modelo. A estrutura de correlação entre as replicações é preservada quando se obtêm as amostras "bootstrap", pois estas são tomadas dos n pares independentes de replicações. Consequentemente, a distribuição empírica, \hat{F} , dos resíduos \underline{e}_i é aquela que associa a probabilidade $1/n$ para cada vetor \underline{e}_i ,

$$\hat{F}(\underline{e}_i) = \begin{cases} 1/n & \text{se } \underline{e}_i \text{ com } i=1,2,\dots,n. \\ 0 & \text{c/c} \end{cases}$$

Os resíduos são corrigidos através da multiplicação pelo fator de correção ϕ dado pelo resultado 2.1.8., com ρ estimado por $\hat{\rho}$, dado por 2.1.16.. Este procedimento produz resultados mais precisos, conforme será visto adiante.

2.2.2. A ABORDAGEM MQG-1 CORRIGIDO

A aplicação do procedimento assintótico MQG-1 ao modelo 2.1.1. pode produzir resultados com precisão significativamente melhorada. Para isto é suficiente se tomar como estimativa de Σ a matriz $\hat{\Sigma}$ dada pelo resultado 2.1.2. com σ^2 e ρ estimados por $\hat{\sigma}^2$ e $\hat{\rho}$, dados respectivamente pelos resultados 2.1.14. e 2.1.16., ou seja estima-se a matriz de covariâncias (2.1.7) por

$$(1 + \hat{\rho})\hat{\sigma}^2.(X'X)^{-1} \quad (2.2.1.)$$

2.3. POSSIBILIDADES DE GENERALIZAÇÃO PARA DELINEAMENTOS COM MAIS DE DUAS REPLICAÇÕES EM CADA PONTO EXPERIMENTAL

Os resultados obtidos no item 2.1. poderão ser generalizados para um número $n > 2$ de replicações, desde que a correlação entre as replicações, ρ , seja constante. Contudo, esta não é uma situação muito fácil de ocorrer na prática. Um problema que merece atenção, no contexto do modelo 2.1.1., é aquele em que a correlação tende a diminuir exponencialmente a medida que as replicações se distanciam na ordem. Assim, para m replicações no ponto experimental i , temos $e_{i1}, e_{i2}, \dots, e_{im}$ e

$$\text{Corr}(e_{ij}, e_{ik}) = \rho^{|k-j|}$$

ou seja, os resíduos num mesmo ponto experimental seguem um processo auto-regressivo de 1ª. ordem. Assim

$$\Sigma = I_n \otimes \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^m \\ & 1 & \rho & \dots & \rho^{m-1} \\ & & 1 & \dots & \vdots \\ & & & \ddots & 1 \end{bmatrix}$$

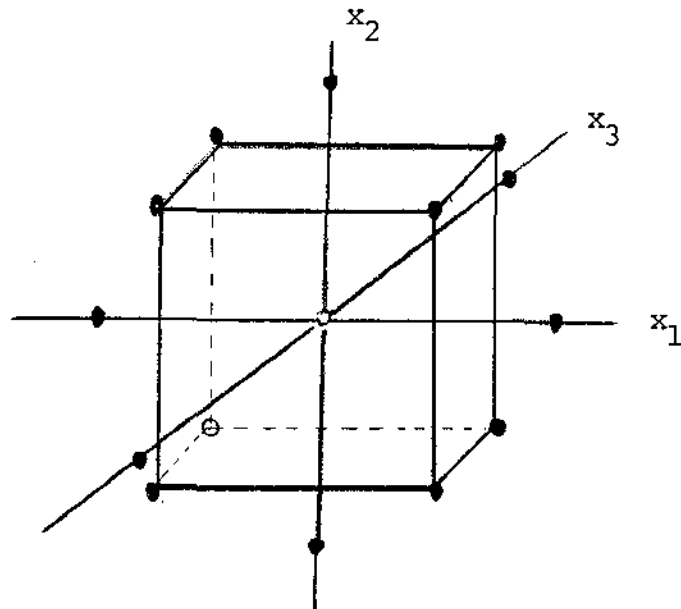
A generalização de resultados para situações em que este problema específico esteja incluído é uma questão aberta para estudos posteriores a este trabalho.

3. SIMULAÇÃO EM UM EXPERIMENTO COM O DELINEAMENTO DE BOX PARA MQG-1 E "BOOTSTRAP".

3.1. O DELINEAMENTO DE BOX (CUBO + ESTRELA)

Para se fazer a simulação envolvendo comparativamente o MQG-1 e o "bootstrap" escolheu-se o delineamento composto cubo+estrela+pontos centrais, para três fatores, pela frequência com que é utilizado no ajustamento de modelos de superfície de resposta de 2a. ordem (ver [5]). Como se sabe, diversos delineamentos foram criados para ajustamento de superfícies de resposta. Quando o modelo é de 1a. ordem os delineamentos fatoriais a dois níveis, do tipo 2^p são comumente empregados por sua simplicidade e parcimônia. No ajustamento de modelos de 2a. ordem um número maior de níveis diferentes para cada fator é necessário. Os delineamentos compostos constituídos por um cubo 2^p mais uma "estrela" de $2p$ pontos axiais e k pontos centrais podem apresentar diversas propriedades que os tornam recomendados, entre elas ortogonalidade e rotacionalidade, além de parcimônia em termos de número de pontos experimentais. Conforme o número de pontos centrais acrescentados, esse delineamento pode ser feito rotacional além de ortogonal ou ainda ter precisão uniforme ([5], pags. 213-215) ou seja, a variância do valor ajustado na origem é igual a variância deste valor à distância $\rho=1$, assim se consegue valores satisfatórios para os vícios nos coeficientes de

regressão devido a possíveis termos de ordem mais alta na verdadeira superfície de resposta. Quando $p=3$ o delineamento composto de 2a. ordem é apresentado na figura abaixo que é clássica na literatura de planejamento de experimentos (ver [10], págs.344-346)



A simulação foi feita assumindo a matriz de planejamento para três fatores de um cubo+estrela, quadro 3.1., porém sem fixação particular de k no sentido de se conseguir ortogonalidade ou precisão uniforme.

QUADRO 3.1. MATRIZ DE PLANEJAMENTO DE UM DELINEAMENTO COMPOSTO
CUBO + ESTRELA + 4 PONTOS CENTRAIS REPLICADOS.

[illegible]

3.2. DESCRIÇÃO DA SIMULAÇÃO MONTE CARLO PARA O MQG-1 E "BOOTS-TRAP"

Seja Y_{ij} $i=1,2,\dots,18$ $j=1,2$ as respostas originárias de um delineamento cubo+estrela com 3 fatores com 4 pontos centrais. Foi feita uma simulação Monte Carlo, ajustando-se a superfície de respostas de 2a. ordem

$$Y_{ij} = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_1 X_2 + \beta_5 X_1 X_3 + \beta_6 X_2 X_3 + \beta_7 X_1^2 + \beta_8 X_2^2 + \beta_9 X_3^2 + \epsilon_{ij}$$

enquadrada no modelo 2.1.1., com a correlação entre as replicações ρ fixada em 0,9. Os parâmetros β_i $i=0,1,2,\dots,9$ foram fixados em zero para maior simplificação, sem perda de generalidade. O número de repetições Monte Carlo e "bootstrap" foram fixadas em 400, cada. Simulou-se primeiramente os procedimentos MQG-1 e "bootstrap" conforme proposto em [3] no qual a estimativa, s^2 , de σ^2 , é dada pelo quadrado médio dos resíduos de um ajuste por MQO e a de ρ pela correlação amostral desses resíduos separados por replicação, abaixo

$$r = \frac{\sum_{i=1}^n e_{i1} e_{i2}}{\sqrt{\sum_{i=1}^n e_{i1}^2 \sum_{i=1}^n e_{i2}^2}} \quad (3.2.1.)$$

Os procedimentos alternativos foram simulados conforme propostos em 2.2. Para cada um dos procedimentos, gerou-se os resíduos segundo uma distribuição $N(0,1)$. Posteriormente foi também empregada uma mistura de $N(0,1)$ e $N(0,4)$ em proporções iguais. Os resultados numéricos obtidos na simulação estão resumidos nos quadros 3.2. e 3.4. para o procedimento sem correção na distribuição $N(0,1)$ e mistura de normais, respectivamente. Os quadros 3.3. e 3.5. mostram os resultados para o procedimento que propomos, que incorpora a correção dos resíduos.

As médias das estimativas de σ^2 e de ρ , dadas por s^2 e r , foram calculadas para as 400 repetições Monte Carlo resultando em 0,643 e 0,763, respectivamente, quando os resíduos provêm de $N(0,1)$. Para a mistura as estimativas são 1,727 e 0,598. No procedimento alternativo onde estima-se σ^2 e ρ por $\hat{\sigma}^2$ e $\hat{\rho}$, os valores médios obtidos na simulação quando os resíduos são oriundos de $N(0,1)$ foram respectivamente de 0,982 e 0,889. Para a mistura obteve-se 2,504 e 0,807, respectivamente. Observa-se por estas estimativas médias que os estimadores s^2 e r são bastante inferiores a $\hat{\sigma}^2$ e $\hat{\rho}$. Os valores verdadeiros destes parâmetros são $\sigma^2 = 1,00$ e $\rho = 0,9$ na situação de resíduos gerados de $N(0,1)$, e $\sigma^2 = 2,50$ e $\rho = 0,9$ na mistura.

Os números citados e os que constam nos quadros em referência foram calculados com emprego do programa CUBO.FOR que é descrito no apêndice 1.

3.3. ANÁLISE COMPARATIVA DA PERFORMANCE DA ABORDAGEM USUAL POR MÍNIMOS QUADRADOS GENERALIZADOS, COM O "BOOTSTRAP", NO DELINEAMENTO COMPOSTO CUBO+ESTRELA.

3.3.1. OS RESÍDUOS TÊM DISTRIBUIÇÃO NORMAL

A situação em que os resíduos têm distribuição normal se constitui na melhor ajustada à abordagem clássica por MQG. Comparamos aqui o desempenho dos procedimentos MQG-1 e o "bootstrap" não corrigido, como em [3], e incorporando as correções propostas na secção 2.2.. Os resultados são apresentados a seguir. O quadro 3.2. apresenta os resultados obtidos por MQG-1 e pelo "bootstrap", com os resíduos de um ajuste por MQO, sem correção. O quadro 3.3. apresenta os resultados obtidos após correção dos resíduos conforme 2.1.8. e com Σ estimada por 2.2.1.. Os números do quadro 3.2. revelam que o "bootstrap" é superior ao MQG-1, embora ambos sejam otimistas. Já no quadro 3.3. se vê que a abordagem assintótica com a estimativa adequada de Σ , e o "bootstrap" com as correções dos resíduos são equivalentes e dão resultados que podem ser considerados bons, ainda que pareçam manter um pequeno vício, sendo o do MQG-1 negativo e do "bootstrap" positivo.

QUADRO 3.2. NÚMERO DE VEZES EM QUE O INTERVALO DE 90% DE CONFIANÇA COBRIU O PARÂMETRO, EM 400 REPETIÇÕES MONTE CARLO E 400 "BOOTSTRAP" NO PROCEDIMENTO SEM CORREÇÃO EM RESÍDUOS $N(0,1)$.

PARÂMETRO	MQO	MQG-1	"BOOTSTRAP"
β_0	261 (65,25%)	315 (78,75%)	335 (83,75%)
β_1	247 (61,75%)	297 (74,25%)	323 (80,75%)
β_2	235 (58,75%)	299 (74,75%)	336 (84,00%)
β_3	260 (65,00%)	306 (76,50%)	345 (86,25%)
β_4	275 (68,75%)	317 (79,25%)	339 (84,75%)
β_5	258 (64,50%)	322 (80,50%)	329 (82,25%)
β_6	276 (69,00%)	321 (80,25%)	341 (85,25%)
β_7	249 (62,25%)	300 (75,00%)	336 (84,00%)
β_8	272 (68,00%)	319 (79,75%)	334 (83,50%)
β_9	258 (64,50%)	307 (76,75%)	336 (84,00%)

QUADRO 3.3. NÚMERO DE VEZES EM QUE O INTERVALO DE 90% DE CONFIANÇA COBRIU O PARÂMETRO, EM 400 REPETIÇÕES MONTE CARLO E 400 "BOOTSTRAP" NO PROCEDIMENTO COM CORREÇÃO EM RESÍDUOS $N(0,1)$.

PARÂMETRO	MQO	MQG-1	"BOOTSTRAP"
β_0	300 (75,00%)	347 (86,75%)	366 (91,50%)
β_1	282 (70,50%)	342 (85,50%)	361 (90,25%)
β_2	286 (71,50%)	342 (85,50%)	369 (92,25%)
β_3	295 (73,75%)	350 (87,50%)	373 (93,25%)
β_4	305 (76,25%)	357 (89,25%)	367 (91,75%)
β_5	302 (75,50%)	359 (89,75%)	368 (92,00%)
β_6	308 (77,00%)	356 (89,00%)	371 (92,75%)
β_7	288 (72,00%)	339 (84,75%)	363 (90,75%)
β_8	305 (76,25%)	356 (89,00%)	367 (91,75%)
β_9	290 (72,50%)	353 (88,25%)	374 (93,50%)

3.3.2. SITUAÇÃO EM QUE OS RESÍDUOS PROVÊM DE UMA MISTURA DE NORMAL $N(0,1)$ COM $N(0,4)$ EM IGUAL PROPORÇÃO.

Para este caso, os resultados são apresentados nos quadros 3.4. e 3.5. adiante e mostram que o MQG-1 e o "bootstrap" determinam praticamente os mesmos resultados que os encontrados em 3.3.1. com resíduos $N(0,1)$. Assim conclui-se que o MQG-1 é um método robusto e quanto ao "bootstrap" esta conclusão já era esperada pois ele independe da distribuição.

QUADRO 3.4. NÚMERO DE VEZES EM QUE O INTERVALO DE 90% DE CONFIANÇA COBRIU O PARÂMETRO, EM 400 REPETIÇÕES MONTE CARLO E 400 "BOOTSTRAP", NO PROCEDIMENTO SEM CORREÇÃO EM RESÍDUOS $N(0,1)$ MISTURADOS COM $N(0,4)$ EM IGUAL PROPORÇÃO.

PARÂMETRO	MQO	MQG-1	"BOOTSTRAP"
β_0	278 (69,50%)	322 (80,50%)	330 (82,50%)
β_1	288 (72,00%)	324 (81,00%)	349 (87,25%)
β_2	294 (73,50%)	328 (82,00%)	349 (87,25%)
β_3	274 (68,50%)	314 (78,50%)	343 (85,75%)
β_4	261 (65,25%)	310 (77,50%)	331 (82,75%)
β_5	277 (69,25%)	318 (79,50%)	340 (85,00%)
β_6	271 (67,75%)	321 (80,25%)	331 (82,75%)
β_7	274 (68,50%)	317 (79,25%)	347 (86,75%)
β_8	290 (72,50%)	323 (80,75%)	351 (87,75%)
β_9	293 (73,25%)	324 (81,00%)	352 (88,00%)

QUADRO 3.5. NÚMERO DE VEZES EM QUE O INTERVALO DE 90% DE CONFIANÇA COBRIU O PARÂMETRO, EM 400 REPETIÇÕES MONTE CARLO E 400 "BOOTSTRAP", NO PROCEDIMENTO COM CORREÇÃO EM RESÍDUOS $N(0,1)$ MISTURADOS COM $N(0,4)$ EM IGUAL PROPORÇÃO.

PARÂMETRO	MQO	MQG-1	"BOOTSTRAP"
β_0	317 (79,25%)	351 (87,75%)	365 (91,25%)
β_1	315 (78,75%)	353 (88,25%)	375 (93,75%)
β_2	317 (79,25%)	358 (89,50%)	370 (92,50%)
β_3	304 (76,00%)	355 (88,75%)	374 (93,50%)
β_4	296 (74,00%)	348 (87,00%)	364 (91,00%)
β_5	307 (76,75%)	341 (85,25%)	363 (90,75%)
β_6	314 (78,50%)	352 (88,00%)	370 (92,50%)
β_7	304 (76,00%)	357 (89,25%)	364 (91,00%)
β_8	316 (79,00%)	355 (88,75%)	377 (94,25%)
β_9	317 (79,25%)	354 (88,50%)	364 (91,00%)

3.4. CONCLUSÃO

Como resultado final desta simulação pode-se concluir que o MQG-1 no contexto do modelo 2.1.1. torna-se significativamente melhor quando é utilizado o estimador da matriz de covariâncias dado por 2.2.1.. Quanto à técnica "bootstrap" pode-se afirmar que no procedimento em que não se faz correção a sua aplicação melhora os resultados obtidos com o MQG-1 usual. Contudo, quando se faz a correção dos resíduos o seu desempenho é equivalente ao MQG-1 corrigido o qual como vimos é bom.

4. APLICAÇÃO A UM CASO REAL

4.1. O PROBLEMA

Um estudo sobre as propriedades do sorgo envolve o descascamento deste cereal em um certo tipo de moinho e considera a influência de três fatores sobre o rendimento, Y . Estes fatores são o tempo de condicionamento T , o teor de umidade U e o tempo de descascamento D . Cada um dos fatores foi delineado com três níveis, 6h, 12h e 24h para T ; 14%, 15% e 16% para U e 2min, 3min e 4min para D . Um experimento fatorial 3^3 foi executado com duas replicações. Entretanto, por impossibilidade técnica as replicações não foram aleatorizadas ou seja uma unidade experimental aleatória (certa quantidade de sorgo) era tomada e um tratamento aplicado às duas unidades amostrais (metade da porção de sorgo) provenientes da mesma unidade experimental. Como se pode observar pela descrição, existe a possibilidade da geração de correlação entre as replicações dentro de cada par, embora matendo-se independência fora desses grupos. Assim este problema enquadra-se no modelo 2.1.1., já descrito. O quadro 4.1. mostra os dados do experimento.

QUADRO 4.1. DADOS ORIGINAIS DO EXPERIMENTO COM OS NÍVEIS DE T, U
e D REPARAMETRIZADOS

	TEMPO DE CONDICIONAMENTO T			TEOR DE UMIDADE U			TEMPO DE DESCASCAMENTO D			RENDIMENTO POR REPLICAÇÃO	
	6h	12h	24h	14%	15%	16%	2min	3min	4min	1a.	2a.
1	-1,333	-	-	-1	-	-	-1	-	-	79,51	79,18
2	-1,333	-	-	-1	-	-	-	0	-	74,70	74,98
3	-1,333	-	-	-1	-	-	-	-	1	71,32	71,23
4	-1,333	-	-	-	0	-	-1	-	-	78,86	80,35
5	-1,333	-	-	-	0	-	-	0	-	77,03	71,66
6	-1,333	-	-	-	0	-	-	-	1	67,68	65,87
7	-1,333	-	-	-	-	1	-1	-	-	83,37	83,78
8	-1,333	-	-	-	-	1	-	0	-	81,19	81,57
9	-1,333	-	-	-	-	1	-	-	1	74,08	74,24
10	-	-0,333	-	-1	-	-	-1	-	-	72,91	72,63
11	-	-0,333	-	-1	-	-	-	0	-	65,35	62,90
12	-	-0,333	-	-1	-	-	-	-	1	36,14	34,56
13	-	-0,333	-	-	0	-	-1	-	-	68,88	70,04
14	-	-0,333	-	-	0	-	-	0	-	57,96	59,53
15	-	-0,333	-	-	0	-	-	-	1	46,01	46,37
16	-	-0,333	-	-	-	1	-1	-	-	78,59	79,22
17	-	-0,333	-	-	-	1	-	0	-	75,07	74,67
18	-	-0,333	-	-	-	1	-	-	1	73,02	72,75
19	-	-	1,667	-1	-	-	-1	-	-	76,60	79,49
20	-	-	1,667	-1	-	-	-	0	-	72,51	72,27
21	-	-	1,667	-1	-	-	-	-	1	62,20	62,10
22	-	-	1,667	-	0	-	-1	-	-	84,36	83,52
23	-	-	1,667	-	0	-	-	0	-	76,16	75,46
24	-	-	1,667	-	0	-	-	-	1	68,78	66,75
25	-	-	1,667	-	-	1	-1	-	-	79,92	80,03
26	-	-	1,667	-	-	1	-	0	-	75,27	76,49
27	-	-	1,667	-	-	1	-	-	1	73,43	75,50

4.2. O AJUSTAMENTO

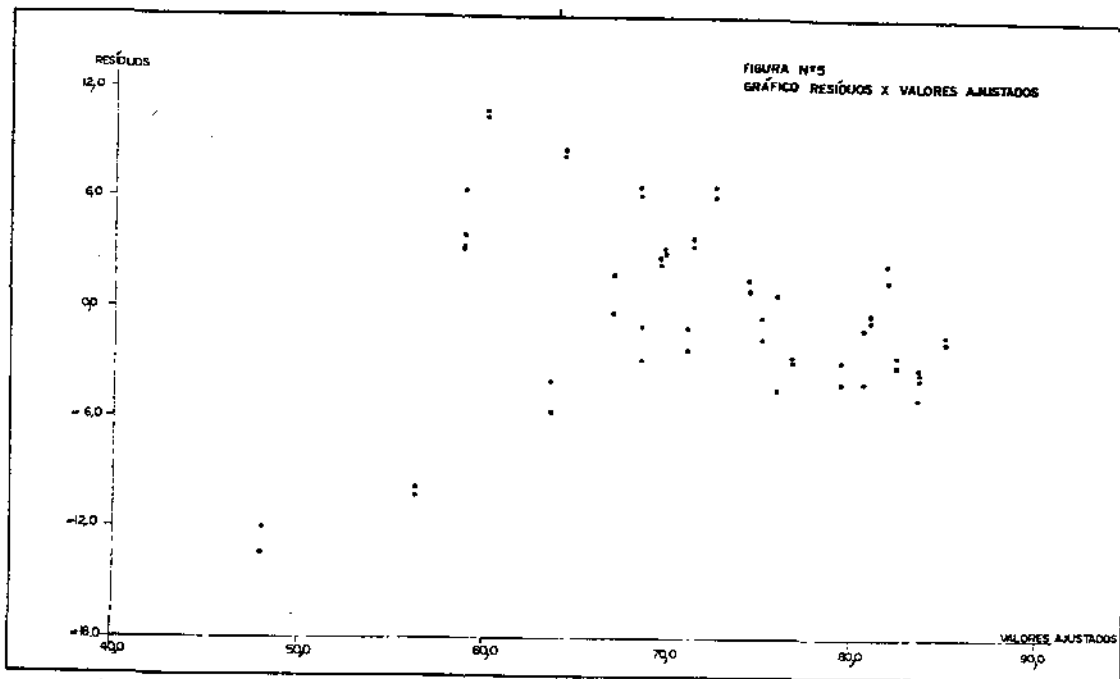
Aos dados do quadro 4.1., utilizando-se o "MINITAB", fez-se uma análise de regressão. De início ajustou-se um polinômio completo do 2º grau com as 9 variáveis $T, U, D, TU, TD, UD, T^2, U^2$ e D^2 verificando-se um bom ajustamento, ainda que várias variáveis não fossem significativas nos níveis estabelecidos. Após a eliminação de TU, TD e D^2 e feito novo ajuste, U^2 mostrou-se significativa a 5% mas não a 1%. Por outro lado, aplicando-se a técnica "bootstrap", neste particular, obtém-se $t_* = \hat{\beta}^*/s = 1,66$ quando se faz $n^* = 1000$ repetições "bootstrap", com $\hat{\beta}^*$ e s respectivamente a média e o desvio padrão do conjunto das 1000 estimativas "bootstrap" do coeficiente verdadeiro de U^2 . Assim, com base neste critério U^2 também foi eliminada.

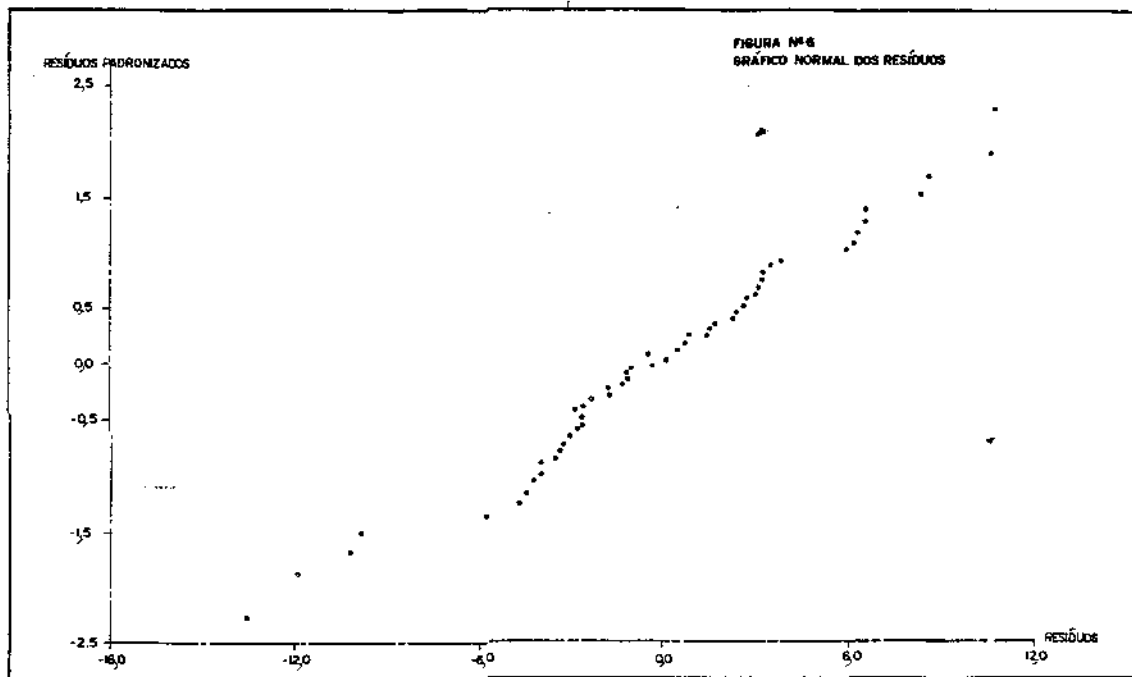
Tem-se, então 5 variáveis significativas a 1% que são T, U, D, UD e T^2 . A equação de regressão, com $R^2 = 0,768$, é

$$\hat{Y} = 62,2 - 2,53T + 4,77U - 7,48D + 3,37UD + 5,95T^2.$$

Examinando-se o gráfico resíduos contra valores ajustados (Fig.5) observa-se uma distribuição satisfatória dos resíduos, que parece não indicar desajuste do modelo adotado. Os resíduos mostram-se aos pares, o que vem revelar uma possível correlação entre as replicações. O gráfico normal dos resíduos (Fig.6) mostra uma

tendência para a linha reta, configurando assim uma situação de normalidade nos resíduos, exceto talvez no extremo inferior. Nos resíduos separados por replicação, tem-se uma correlação entre os dois conjuntos de 0,954, e o gráfico dos resíduos da 1a. replicação contra os da 2a. mostra claramente, pela tendência a uma linha reta, esta forte correlação positiva.





4.3. A AMBIGÜIDADE ENTRE CORRELAÇÃO E DESAJUSTAMENTO

O ajuste de modelos de regressão e posterior análise estatística dos resultados pressupõem, em geral, algumas suposições básicas sobre a adequação do modelo escolhido, e sobre a estrutura estatística dos erros aleatórios. Tais suposições devem, na medida do possível, ser verificadas experimentalmente.

Os testes de desajustamento (lack-of-fit), permitem verificar o ajuste do modelo aos dados, num contexto em que se dispõe de replicações genuínas (ver [6], pág.35).

Contudo, se existe algum tipo de correlação entre os resíduos, a falta de ajuste pode aparecer como significativa no teste embora o modelo ajustado esteja adequado. Assim, o teste de

desajustamento pode indicar falta de ajuste, replicações correlacionadas ou ambos. Frequentemente em experimentos com replicação, por descuido ou impossibilidade técnica criam-se condições para o surgimento de correlação entre as replicações.

No problema por nós abordado, o experimento foi desenvolvido sem aleatorização das replicações; assim, uma estrutura de correlação do tipo descrito abaixo pode ocorrer entre as $2n$ respostas,

$$\text{Corr}(\varepsilon_i, \varepsilon_j) = \begin{cases} 1 & \text{se } i = j \\ \rho & \text{se } i = j - 1 \\ 0 & \text{c/c para } i = 1, 2, 3, \dots, 2n \end{cases}$$

O teste de desajustamento aplicado ao ajuste com as cinco variáveis significativas dado no quadro 4.2., abaixo, resultou significativo.

QUADRO 4.2. ANÁLISE DA VARIÂNCIA (MOSTRANDO O DESAJUSTAMENTO)

FONTE DE VARIAÇÃO	G.L.	S.Q.	Q.M.	F
REGRESSÃO	5	4748,060	949,61	31,72
RESIDUAL	48	1437,290	29,94	
DESAJUSTAMENTO	21	1403,522	66,83	53,47
ERRO PURO	27	33,768	1,25	
TOTAL	53	6185,350		

4.4. ESTIMATIVAS E DESVIOS PADRÃO DOS PARÂMETROS

A partir dos dados do quadro 4.1., as estimativas dos parâmetros da superfície quadrática de respostas

$$Y = \beta_0 + \beta_1 T + \beta_2 U + \beta_3 D + \beta_4 UD + \beta_5 T^2$$

foram obtidas por MQO, sendo que várias variáveis foram eliminadas na análise de regressão. Sabe-se, pelo resultado 2.1.5., que essas estimativas são iguais às obtidas por MQG-1. A matriz de variâncias-covariâncias do ajuste por MQG-1

$$(1 + \hat{\rho}) \hat{\sigma}^2 (X'X)^{-1}$$

foi obtida substituindo-se σ^2 e ρ em 2.1.7. pelas estimativas dadas, respectivamente, pelos resultados 2.1.14. e 2.1.16, ou seja, $\hat{\sigma}^2 = 34,083$ e $\hat{\rho} = 0,961$.

O "bootstrap" foi aplicado aos resíduos do ajuste corrigidos pelo fator

$$\phi = \sqrt{\frac{54}{54 - 6(1 + \hat{\rho})}},$$

conforme resultado 2.1.8., com ρ substituído pela estimativa dada pelo resultado 2.1.16.. O número de repetições foi $n^* = 1000$.

Os resultados são mostrados nos quadros 4.3. e 4.4., respectivamente para MQG-1 e "bootstrap". Contêm estimativas dos parâmetros, desvios padrão das estimativas e intervalos de confiança nos níveis de 0,90 e 0,95, para os dois casos. Os números foram obtidos com o programa MQGAR5.FOR, descrito no apêndice 2.

QUADRO 4.3. RESULTADOS POR MQG-1

PARÂMETROS	$\hat{\beta}_i$	s_{ii}	INT.DE 0,90 DE CONF.	INT.DE 0,95 DE CONF.
β_0	62,194	2,175278	[58,544;65,844]	[57,818;66,571]
β_1	-2,533	1,059770	[-4,311;-0,755]	[-4,665;-0,401]
β_2	4,767	1,362563	[2,480; 7,053]	[2,025; 7,508]
β_3	-7,478	1,362563	[-9,764;-5,191]	[-10,219;-4,736]
β_4	3,370	1,668791	[0,569; 6,170]	[0,012; 6,727]
β_5	5,947	1,201665	[3,930; 7,963]	[3,528; 8,364]

QUADRO 4.4. RESULTADOS POR "BOOTSTRAP" $n^* = 1000$

PARÂMETROS	$\hat{\beta}_i^*$ médio	s_{ii}	INT.DE 0,90 DE CONF.	INT.DE 0,95 DE CONF.
β_0	62,226	2,133121	[58,648;65,766]	[58,013;66,353]
β_1	-2,546	1,039806	[-4,241;-0,883]	[-4,551;-0,531]
β_2	4,752	1,322487	[2,553; 6,986]	[2,085; 7,332]
β_3	-7,500	1,290475	[-9,680;-5,419]	[-9,947;-4,917]
β_4	3,406	1,664950	[0,714; 6,288]	[0,327; 6,883]
β_5	5,922	1,210724	[3,861; 7,838]	[3,491; 8,183]

4.5. CONCLUSÃO

Os números finais mostram que o procedimento "bootstrap", que é independente da distribuição, apresenta uma performance equivalente ao MQG-1.

APÊNDICE I - PROGRAMA CUBO.FOR

Este é o programa para a simulação descrita no Capítulo III. Na sua composição fazem parte sub-rotinas do SSP para operações matriciais, e também outras construídas para gerar misturas de normais e calcular produtos matriciais. Na forma aqui apresentada lê as matrizes de planejamento com replicação em FORØ3.DAT e sem replicação em FORØ1.DAT e faz a simulação em resíduos de $N(0,1)$ com o procedimento corrigido, gerando os números do quadro 3.3.. Para os resíduos oriundos da mistura de normais estes valores são gerados pela sub-rotina NORCON.FOR, construída para este fim, então o laço situado sob o comentário "gerando os resíduos" é ignorado. A sua execução produz os números do quadro 3.5.. No procedimento sem correção, a versão correspondente, calcula a estimativa de σ^2 , s^2 igual ao quadrado médio dos resíduos do ajuste MQO, e a correlação amostral desses resíduos separados por replicação, r . Assim os ajustes separados não são feitos. Obtém-se desta forma os números de 3.2. e finalmente quando se chama NORCON.FOR para gerar a mistura de normais em substituição a geração de resíduos $N(0,1)$ tem-se os resultados do quadro 3.5.. A listagem é a que segue.

```

C
C *****
C *****
C ***** C U B O . F O R *****
C *****
C *****
C
DIMENSION X(36,10),Y(36,1),EBS(10),VXTX(10,10),
* C(10,36),EBS(36),YBS(36),E(36),XI(10,36),SG(36,36)
DIMENSION XIX(10,10),XIXT(10,36),FS(10,36),VXGX(10,10),
* LI(10),MI(10),HCHO(10,1),L(10),M(10),SO(10),SI(10),
* TSO(10),TSI(10),ETO(10),ESO(10),EII(10),ESI(10),
* YCHO(36,1),ECHO(36,1),ICONT0(10),ICONT1(10),ICONT8(10)
DIMENSION REP1(18),REP2(18),AMPO(10),AMP1(10),AMBS(10),
* SAMPO(10),SAMP1(10),SAMBS(10),AMPO3(10),AMP16(10),AMBSB(10),
* AUX(1000),BS(1000,10),ICONTI(10),ICONT5(10)
DIMENSION W(18,10),WT(10,18),WT=(10,10),WTWT(10,18),YR1(18,1),
* YR2(18,1),AR1(10),AR2(10),YCHR1(18,1),YCHR2(18,1),ECHR1(18,1),
* ECHR2(18,2)

C
C *****
C ***** ENTRANDO COM OS PARAMETROS E ARGUMENTOS *****
C *****
C
Type 5
5 FORMAT('***** ENTRE RHO *****')
ACCEPT 10, RHO
10 FORMAT(G)
Type 15
15 FORMAT('***** ENTRE VARIANCIA *****')
ACCEPT 18, SIGMA2
20 FORMAT(G)
B=RHO**2
TRHO = SQRT(1.0 - B)
SIGMA=SQRT(SIGMA2)
PI=3.1415926
N=36
NC=10
MSA=0
MSB=0
N1=1
MN=18
FN=FLOAT(N)
FMN=FLOAT(MN)
FNC=FLOAT(NC)

C
C *****
C ***** LENDO A MATRIZ DE PLANEJAMENTO EM FOR03.DAT *****
C *****
C
DO 50 I=1,36
30 READ(3,30) (X(I,J),J=1,10)
50 FORMAT(F3.0,3F7.3,3F4.0,3F8.5)
CONTINUE

C
C *****
C ***** CALCULO DA MATRIZ (X*Y)TNV *****
C *****
C
CALL MTRA(X,XT,N,NC,MSA)

```

```

CALL MPRD(XT,X,XTX,NC,N,MSA,MSB,NC)
CALL MINV(XTX,NC,D,L,M)

C
C *****
C ***** CALCULO DA MATRIZ (X'X)INV.X' ***** C
C *****
C
CALL PD13(XTX,XT,XIXT,NC,NC,M)

C
C *****
C ***** LENDO A MATRIZ DE PLANEJAMENTO SEM REPLICACAO ***** C
C ***** EM FORO1.0AT PARA OS AJUSTES SEPARADOS DAS ***** C
C ***** 1AS. E 2AS. REPLICACOES ***** C
C *****
C
DD 70 I=1,18
READ(1,60) (W(I,J),J=1,10)
60 FORMAT(F3.0,3F7.3,3F4.0,3F8.5)
70 CONTINUE

C
C *****
C ***** CONCENTRANDO AS OPERACOES MATRICIAIS FORA ***** C
C ***** DO ELO MONTE CARLO ***** C
C *****
C
CALL MTRA(W,WT,MN,NC,MSA)
CALL MPRD(WT,W,WTW,NC,MN,MSA,MSB,NC)
CALL MINV(WTW,NC,D,L,M)
CALL PD15(WTW,WT,WIWT,NC,NC,MN)

C
C *****
C ***** A SIMULACAO MONTE CARLO ***** C
C *****
C
TYPE 80
80 FORMAT(' *** ENTRE NMC (REPS. M CARLO) * NBS (REPS. BOOTS.) *** ')
ACCEPT 82, NMC, NBS
82 FORMAT(2G)
TYPE 86
86 FORMAT(' ***** ENTRE A SEMENTE ***** ')
ACCEPT 88, ISEED
88 FORMAT(I)
CALL SEIRAN(ISEED)

C
C *****
C ***** OS LIMITES DE 0.90 PARA OS INTERVS. BOOTSTRAP ***** C
C *****
C
EIBS=0.05*NBS+1.0
ESBS=0.95*NBS
TYPE 89, EIBS, ESBS
89 FORMAT(' LIMITES SAO',2F6.1)

C
C *****
C ***** ZERANDO OS ACUMULADORES ***** C
C *****
C
DD 90 I=1,10
ICONT0(I)=0
ICONT1(I)=0

```

```

          ICONTB(I)=0
          SAMP0(I)=0.0
          SAMP1(I)=0.0
90      CONTINUE

          *****
          *****      INICIO DO ELG MONTE-CARLO      *****
          *****

          DO 250 IMC=1,NMC

          *****
          *****      ZERANDO OS ACUMULADORES DO CALCULO DAS      *****
          *****      ESTIMATIVAS VARCH E R, DA VARIANCIA E COEF.      *****
          *****      DE CORRELACAO      *****
          *****

          S2ER1=0.0
          S2ER2=0.0
          SER1R2=0.0
          SER1=0.0
          SER2=0.0

          *****
          *****      GERANDO OS RESIDUOS      *****
          *****

          DO 92 I=1,MN
              U1=РАН(ISEED)
              U2=РАН(ISEED)

              Z1=SQRT(-2.0*ALOG(U1))*COS(2.0*PI*"2)
              Z2=SQRT(-2.0*ALOG(U1))*SIN(2.0*PI*"2)

              E(2*I-1)=Z1*SIGMA
              E(2*I)=(RHO*Z1+TRHO*Z2)*SIGMA
92      CONTINUE

          *****
          *****      GERANDO OS Y'S      *****
          *****

          DO 94 I=1,N
              Y(I,1)=E(I)
94      CONTINUE

          *****
          *****      SEPARANDO OS Y'S NAS DUAS REPLICACOES      *****
          *****

          DO 95 I=1,MN
              YR1(I,1)=Y(2*I-1,1)
              YR2(I,1)=Y(2*I,1)
95      CONTINUE

          *****
          *****      O VETOR RCHO DAS ESTIMATIVAS MQD/MQG-1      *****
          *****

```

```

CALL MPRD(XTXT,Y,BCHO,NC,N,MSA,MSB,N1)

*****
***** O VETOR YCHO DAS RESPOSTAS EM MQO/MQG-1 *****
*****

CALL MPRD(X,BCHO,YCHO,N,NC,MSA,MSB,N1)

*****
***** O VETOR ECHO DOS RESIDUOS EM MQO/MQG-1 *****
*****

DO 96 I=1,N
    ECHO(I,1)=Y(I,1)-YCHO(I,1)
CONTINUE

*****
***** FAZENDO OS AJUSTES COM AS REPLICACOES *****
***** SEPARADAS PARA SE OBTER ESTIMATIVA DE *****
***** SIGMA2 E DE RHO *****
*****
***** OS PARAMETROS *****

CALL MPRD(WTWT,YR1,BR1,NC,MN,MSA,MSB,N1)
CALL MPRD(WTWT,YR2,BR2,NC,MN,MSA,MSB,N1)

***** OS Y'S ESTIMADOS *****

CALL MPRD(W,BR1,YCHR1,MN,NC,MSA,MSB,N1)
CALL MPRD(W,BR2,YCHR2,MN,NC,MSA,MSB,N1)

***** OS RESIDUOS *****

DO 98 I=1,MN
    ECHR1(I,1)=YR1(I,1)-YCHR1(I,1)
    ECHR2(I,1)=YR2(I,1)-YCHR2(I,1)
    SER1=SER1+ECHR1(I,1)
    SER2=SER2+ECHR2(I,1)
    S2ER1=S2ER1+ECHR1(I,1)**2
    S2ER2=S2ER2+ECHR2(I,1)**2
    SER1R2=SER1R2+ECHR1(I,1)*ECHR2(I,1)
CONTINUE

*****
***** AS VARIANCIAS OBTIDAS NOS AJUSTES SEPARADOS *****
*****

VARCH1=S2ER1/8.0
VARCH2=S2ER2/8.0

***** A ESTIMATIVA DA VARIANCIA (MEDIA) *****

VARCH=0.5*(VARCH1+VARCH2)
SVARCH=SVARCH +VARCH

***** O COEFICIENTE DE CORRELACAO (AMOSTRAL) *****

R=SER1R2/SQRT(S2ER1*S2ER2)
SR=SR+R

```

```

C
C *****
C ***** CALCULANDO O COEFICIENTE 'A' *****
C *****
C
A=(1.0+R)
C
C *****
C ***** A MATRIZ DE VARIANCIAS-COVARIANCIAS MQO *****
C *****
C
DO 97 I=1,10
  DO 97 J=1,10
    VXTX(I,J)=VARCH*XTX(I,J)
97 CONTINUE
C
C *****
C *** O ERRO DE AMOSTRAGEM, NO NIVEL 0.90, DE C/EST. MQO ***
C *****
C
DO 99 I=1,10
  SO(I)=SQRT(VXTX(I,I))
99 TS0(I)=1.708*SO(I)
C
C *****
C *** AMPLITUDES DOS INTS. E VERIFICACAO DA INCLUSAO ***
C *** DOS PARAMETROS NOS INTERVALOS ***
C *****
C
DO 100 I=1,10
  EIO(I)=BCHO(I,1)-TS0(I)
  ESO(I)=BCHO(I,1)+TS0(I)
  AMP0(I)=2.0*TS0(I)
  SAMPO(I)=5*SAMPO(I) + AMP0(I)
  IF (EIO(I).GT.0.0.OR.ESO(I).LT.0.0) GO TO 100
  ICONT0(I)=ICONT0(I)+1
100 CONTINUE
DO 102 I=1,10
  DO 102 J=1,10
C
C *****
C ***** A MATRIZ DE VARIANCIAS-COVARIANCIAS POR MQG-1 *****
C *****
C
VXGX(I,J)=A*VXTX(I,J)
102 CONTINUE
C
C *****
C *** O ERRO DE AMOSTRAGEM, NO NIVEL 0.90, P/C/EST. MQG-1 ***
C *****
C
DO 104 I=1,10
  S1(I)=SQRT(VXGX(I,I))
104 TS1(I)=1.708*S1(I)
C
C *****
C *** OS LIMITES DE 0.90 DE CONFIANCA P/A EST. MQG-1 ***
C *** E VERIFICACAO DA INCLUSAO DOS PARAMETROS NOS INTS. ***
C *****

```

```

DO 108 I=1,10
  EY1(I)=BCH0(I,1)-TS1(I)
  ES1(I)=BCH0(I,1)+TS1(I)
  AMP1(I)=2.0*TS1(I)
  SAMP1(I)=SAMP1(I)+AMP1(I)
  IF(EY1(I).GT.0.0.OR.ES1(I).LT.0.0) GO TO 108
  ICONT1(I)=ICONT1(I)+1
  CONTINUE
108
C
C *****
C **** O FATOR DE CORRECAO DOS RESIDUOS MQG=1 ***** C
C **** CALCULADO COM BASE NOS GL'S ***** C
C ***** C
C
FC=SQRT(FN/(FN-FNC*(1.0+R)))
C
C ***** C
C **** O VETOR ECHO DOS RESIDUOS MQG/MQG=1 CORRIGIDOS ***** C
C **** PELO FATOR DE CORRECAO DOS GRAUS DE LIBERDADE ***** C
C ***** C
DO 120 I=1,36
  ECHO(I,1)=(Y(I,1)-YCHO(I,1))*FC
  CONTINUE
120
C
C ***** C
C **** O BOOTSTRAP ***** C
C ***** C
C
DO 130 I=1,10
  ICONT1(I)=0.0
  ICONTS(I)=0.0
  CONTINUE
130
C
C ***** INICIO DO ELO DO BOOTSTRAP ***** C
C
DO 200 J=1,NBS
C
C ***** C
C ***** GERANDO OS RESIDUOS BOOTSTRAPS ***** C
C ***** C
C
DO 150 I=1,MN
  U=RAN(ISEED)
  K=2*INT(18*U +1.0)
  EBS(2*I-1)=ECHO(K-1,1)
  EBS(2*I)=ECHO(K,1)
  CONTINUE
150
C
C ***** C
C ***** GERANDO OS PSEUDO-DADOS ***** C
C ***** C
C
DO 160 I=1,36
  YBS(I)=YCHO(I,1)+EBS(I)
160
C
C ***** C
C ***** O VETOR BBS DAS ESTS. BOOTSTRAP DOS PARAMETROS *** C
C ***** C

```



```

CALL MPRD(XTXT,YB5,BB5,NC,N,MSA,MSB,N1)
C
C *****
C *** VERIFICACAO DA INCLUSAO DO PARAMETRO NO INTERVALO ***** C
C ***** C
C
DO 195 I=1,10
IF (BB5(I).LT.0.0) GO TO 190
IF (BB5(I).GT.0.0) GO TO 192
GO TO 195
190 ICONTI(I)=ICONTI(I)+1
GO TO 195
192 ICONTS(I)=ICONTS(I)+1
195 CONTINUE
200 CONTINUE
C
C ***** FIM DO ELO BOOTSTRAP ***** C
C
DO 210 I=1,10
IF (ICONTI(I).LT.EIBS.OR.ICONTS(I).GT.ESBS) GO TO 210
ICONTB(I)=ICONTB(I)+1
210 CONTINUE
250 CONTINUE
C
C ***** FIM DO ELO MONTE-CARLO ***** C
C
FNMC=FLOAT(NMC)
DO 255 I=1,10
C
C ***** AMPLITUDES MEDIAS PARA AS NMC REPS. MONTE CARLO ***** C
C
AMPOB(I)=SAMP0(I)/FNMC
AMP1B(I)=SAMP1(I)/FNMC
255 CONTINUE
C
C *****
C ***** R E S U L T A D O S G E R A I S ***** C
C ***** C
C
DO 270 I=1,10
TYPE 260, ICONT0(I), ICONT1(I), ICONTB(I)
FORMAT(7X, 'MQ0=', I4, 7X, 'MQ0-1=', I4, 7X, 'BTS=', I4)
260 TYPE 265, AMPOB(I), AMP1B(I)
265 FORMAT(' A M MQ0=', F10.6, ' A M MQ0-1=', F10.6)
270 CONTINUE
VARCHM=SVARCH/NMC
RM=SR/NMC
TYPE 268, VARCHM, RM, NMC
268 FORMAT(' VAR MED = ', F10.6, ' RHO MED = ', F10.6, ' NMC= ', I4)
STOP
END

```

APÊNDICE II - PROGRAMA MQGAR5.FOR

Lê os dados do problema real no arquivo MAPL6Y.DAT, onde consta a matriz de planejamento e as respostas observadas, e calcula os resultados que compõem os quadros 4.1. e 4.2.. Da mesma forma que o CUBO.FOR, possui na sua composição sub-rotinas SSP de operações matriciais, além de outras específicas de produto matricial que necessitou-se criar. A listagem é a que segue.

```

*****
*****                                *****
*****                                *****
*****                                *****
*****                                *****
*****                                *****

DIMENSION X(54,7),Y(54,1),YCH1(54,1),PBS(6),BRS2(6),BCH1(6,1),
* ECH1(54,1),C(6,54),EBS(54),YHS(54),SBS(6),SBS2(6),XB(6),VBS(6)
DIMENSION SPBS(6,6),CON(28,28),XI(6,54),TG(6,54),XTX(6,6),L1(6),
* M1(6),SG(54,54),L(54),M(54),VXTX(6,6)

*****
***** ENTRANDO COM OS PARAMETROS E ARGUMENTOS *****
*****

N=54
NC=6
MSA=0
MSB=0
NI=1
MN=27

*****
**** ENTRANDO COM O COEF. DE CORRELACAO E VARIANCIA ****
**** DETERMINADOS FORA DO PROGRAMA, NO MAINITAG, A ****
**** PARTIR DOS RESIDUOS DE AJUSTES SEPARADOS COM AS ****
**** 1AS. E 2AS. REPLICACOES ****
*****

TYPE 5
FORMAT('***** ENTRE COM RHOCH E SIGCHZ *****')
ACCEPT 10,RHOCH,SIGCHZ
FORMAT(2G)

*****
*****          O COEFICIENTE A          *****
*****

A=(1.0+RHOCH)*SIGCHZ

*****
***** LENDO A MATRIZ DE PLANEJAMENTO E O VETOR DE *****
***** RESPOSTAS Y EM MAPL6Y.DAT *****
*****

DO 40 I=1,54
READ(6,30)((X(I,J),J=1,6),Y(I,1))
FORMAT(F3.0,F9.5,3F4.0,F8.5,F6.2)
CONTINUE

*****
*****          CALCULO DA MATRIZ (X'X)INV          *****
*****

CALL MTRA(X,XT,N,NC,MSA)
CALL MPRO(XI,X,XTX,NC,N,MSA,MSB,NC)
CALL MINV(XTX,NC,D1,L1,M1)

```

```

C *****
C ***** A MATRIZ DE VARIANCIAS-COVARIANCIAS MQG-1 ***** C
C *****
C
C DO 47 I=1,6
C   DO 47 J=1,6
C     VXTX(I,J)=A*XTX(I,J)
47 CONTINUE
C   DO 61 I=1,6
C     DO 51 J=I,6
C       TYPE 51, I,J,VXTX(I,J)
51 FORMAT(' COV(',I1,I1,')=',F10.6)
61 CONTINUE
C *****
C ***** A MATRIZ (X'X)INV.X' ***** C
C *****
C
C CALL PD13(XTX,XT,C,NC,NC,N)
C *****
C ***** O VETOR BCH1 DOS PARAMETROS ESTIMADOS POR MQG-1 *** C
C *****
C
C CALL PD11(C,Y,BCH1,NC,N,N1)
C   TYPE 80, BCH1(1,1),BCH1(2,1),BCH1(3,1),BCH1(4,1),BCH1(5,1),BCH1
80 * (6,1)
C   FORMAT(6F10.6)
C *****
C ***** O VETOR YCH1 DOS VALORES AJUSTADOS POR MQG-1 ***** C
C *****
C
C CALL MPRD(X,BCH1,YCH1,N,NC,MSA,MSB,N1)
C *****
C ***** O FATOR DE CORRECAO DOS RESIDUOS MQG-1 ***** C
C *****
C
C FC=SQRT(N/(N-6.0*(1.0+RHOCH)))
C *****
C ***** O VETOR ECH1 DOS RESIDUOS DO AJUSTE MQG-1 ***** C
C ***** COM A CORRECAO DOS GRAUS DE LIBERDADE ***** C
C *****
C
C DO 100 I=1,54
C   ECH1(I,1)=(Y(I,1)-YCH1(I,1))*FC
100 CONTINUE
C *****
C *****
C ***** O B J O T S T R A P ***** C
C *****
C *****
C *****
C
C   TYPE 120
120 FORMAT(' ENTRE NBS -REPETICOES BOOTSTRAPS-')
C   ACCEPT 130,NBS
130 FORMAT(G)
C   TYPE 135

```

```

135  FORMAT(' ENTRE A SEMENTE ')
      ACCEPT 140, ISEED
140  FORMAT(I)
      CALL SETRAN(ISEED)
      DO 196 I=1,6
        SBS(I)=0.0
196  SBS2(I)=0.0
      DO 198 I=1,6
        DO 198 J=1,6
          SPBS(I,J)=0.0
198  CONTINUE
C
C *****
C ***** INICIO DO ELO BOOTSTRAP ***** C
C *****
C
      DO 200 J=1,NBS
C
C *****
C ***** GERANDO OS RESIDUOS BOOTSTRAP ***** C
C *****
C
      DO 150 I=1,MN
        U=RAN(ISEED)
        K=2*INT(27*U +1)
        EBS(2*I-1)=ECH1(K-1,1)
        EBS(2*I)=ECH1(K,1)
150  CONTINUE
C
C *****
C ***** GERANDO OS PSEUDO-DADOS ***** C
C *****
C
      DO 160 I=1,54
160  YBS(I)=YCH1(I,1)+EBS(I)
C
C *****
C ***** O VETOR BBS DOS PARAMETROS ESTIMADOS PELO BOOTSTRAP ***** C
C *****
C
      CALL MPRD(C,YBS,BBS,NC,N,MSA,MSH,N1)
      WRITE(2,170) BBS(1),BBS(2),BBS(3),BBS(4),BBS(5),BBS(6)
170  FORMAT(3F10.6,F11.6,F11.6,F10.6)
C
C *****
C ***** A MEDIA E VARIANCIA DAS ESTIMATIVAS BOOTSTRAP ***** C
C *****
C
      DO 180 I=1,6
        SBS(I)=SBS(I)+BBS(I)
        BBS2(I)=BBS(I)**2
        SBS2(I)=SBS2(I)+BBS2(I)
      DO 180 K=1,6
        SPBS(I,K)=SPBS(I,K)+BBS(I)*BBS(K)
180  CONTINUE
200  FN=FLOAT(NBS)
      DO 220 I=1,6
        XB(I)=SBS(I)/FN
        VBS(I)=(SBS2(I)-FN*XB(I)**2)/(FN-1.0)

```

```

220      CONTINUE
C
C      *****
C      ***** A MATRIZ DE COVARIANCIAS POR BOOTSTRAP ***** C
C      *****
C
      DO 225 I=1,6
        DO 225 J=I,6
          COV(I,J)=(SPBS(I,J) - FN*XB(I)*XB(J))/(FN-1)
225      CONTINUE
C
C      *****
C      ***** A MEDIA DE CADA PARAMETRO PARA AS NBS ***** C
C      ***** REPLICACOES BOOTSTRAP E A SUA VARIANCIA ***** C
C      *****
C
      DO 240 I=1,6
        TYPE 230, XB(I), VBS(I)
230      FORMAT(4X,F10.6,4X,F10.6)
240      CONTINUE
C
C      ***** AS COVARIANCIAS BOOTSTRAP ***** C
C
      DO 250 I=1,6
        DO 250 J=I,6
          TYPE 245, I, J, COV(I, J)
245      FORMAT(' COV(', I1, I1, ') = ', F10.6)
250      CONTINUE
      STOP
      END

```

REFERÊNCIAS

- [1] EFRON, B. Bootstrap methods: another look at the jackknife. Ann. Statist. Jan. 1979, v.7, n.1, p.1-26.
- [2] FREEDMAN, D.A. Bootstrapping Regression Models. Ann. Statist. Nov. 1981, v.9, n.6, p.1218-28.
- [3] FREEDMAN, D.A. e PETERS, S.C. Bootstrapping a Regression Equation: Some Empirical Results. Journal of the American Statistical Association Mar.1984, v.79, n.385, p.97-106.
- [4] SEARLE, S.R. Linear Models. New York, John Wiley & Sons Inc., 1971.
- [5] BOX, G.E.P. and HUNTER, J.S. Multi-factor Experimental Designs for Exploring Response Surfaces. Ann. Math. Stat. 1957, v.28, p.195-241.
- [6] DRAPER, N.R. and SMITH, H. Applied Regression Analysis. 2.ed. New York, John Wiley & Sons, 1981.
- [7] MARDIA, K.V., KENT, J.T. Multivariate Analysis. London, Academic Press Inc., 1979, p.459-60.
- [8] ZAR, J.H. Bioestatistical Analysis, Englewood, N.J., Prentice-Hall Inc., 1974.
- [9] BICKEL, P.J. and DOKSUM, K.A. Mathematical Statistics: Basic Ideas and Selected Topics. São Francisco, Holden-Day, Inc., 1977.

- [10] COCHRAN, W.G. and COX, G.M. Experimental Designs. 2.ed, New York, John Wiley & Sons, Inc., 1957.
- [11] BERAN, R., Estimated Sampling Distributions; The Bootstrap and Competitors. The Annals of Statistics, Mar. 1982, v.10, n.1, p.212-25.
- [12] EFRON, B. and DIACONIS, P. Computer-Intensive Methods in Statistics. Scientific American, Maio 1983, p.96-108.
- [13] BOX, G.E.P., The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples. Biometrics, 1954, n.10, p.16-60.
- [14] BOX, G.E.P., et alii. The Design and Analysis of Industrial Experiments. Oliver and Boyd, 1963.