

ASPECTOS DA ANÁLISE DE RESÍDUOS

ROSANA CRISTINA PEREIRA PARENTE

Estatística

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 em "Estatística e Experimentação Agronômica".

PIRACICABA
Estado de São Paulo - Brasil
Novembro, 1984

ASPECTOS DA ANÁLISE DE RESÍDUOS

ROSANA CRISTINA PEREIRA PARENTE

Estatística

Orientador: Prof. Dr. ANTÔNIO FRANCISCO IEMMA

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 em "Estatística e Experimentação Agrônômica".

PIRACICABA
Estado de São Paulo - Brasil
Novembro, 1984

*Aos meus pais,
José Maria e Iolanda,*

*Aos meus irmãos,
Ricardo,
Ronan,
Ranieri,
Roseani,*

*Ao meu filho,
João Paulo,*

À família Ferreira Alves,

DEDICO.

AGRADECIMENTOS

- Ao Prof. Dr. Antonio Francisco Iemma, Prof. Assistente Doutor do Deptº de Matemática e Estatística da ESALQ/USP, pelo estímulo, pela amizade, e pelo minucioso cuidado com que foi orientador deste trabalho.
- Ao Prof. Dr. Humberto de Campos, Coordenador do Curso de Estatística e Experimentação Agronômica, pela atenção e apoio durante o Curso de Pós-Graduação.
- Ao Corpo Docente do Departamento de Matemática e Estatística da ESALQ/USP, pelos ensinamentos recebidos.
- À Universidade do Amazonas, pela oportunidade para realização do Curso.
- Ao Prof. Afonso Celso Maranhão Nina, da Universidade do Amazonas, pela ajuda e estima.
- Ao Prof. Luis Irapuan Pinheiro, da Universidade do Amazonas, por ter acreditado em meu trabalho, pelo apoio recebido e pela amizade.
- À Maria Izalina Ferreira Alves, pela grande amizade, pelo carinho e dedicação nos momentos mais difíceis, e pelo excelente trabalho datilográfico.
- À Rosa Maria Alves, pela amizade e carinho, e pela versão do Resumo.

- Ao Mário Luiz Venturini, pelo auxílio na confecção das figuras.
- Ao Sr. Octávio Frassetto, Sra. Maria Alayde Penteado de Souza, Srta. Nilse Maria de Camargo, e Sra. Djanira Ortolan Forti, funcionários do Departamento de Matemática e Estatística, pelos serviços prestados e pelo convívio.
- Ao colega José Carlos Barbosa, pelas críticas e sugestões recebidas.
- Aos colegas do Curso de Pós-Graduação, pelo companheirismo e amizade.
- Aos amigos Luis Carlos e José Cardoso, pelo carinho dispensado, apesar da distância.
- A Eng^a-Agr^a Maria Elizabeth da Costa Vasconcellos, pesquisadora do Centro Nacional de Pesquisa da Seringueira e Dendê - EMBRAPA, Manaus, AM, pela cessão dos dados.
- A todos que, direta ou indiretamente, colaboraram para a realização deste trabalho.

Í N D I C E

	Pág.
LISTA DE TABELAS	ix
LISTA DE FIGURAS	xi
RESUMO	xiii
SUMMARY	xv
1. INTRODUÇÃO	1
1.1 - Localização do Problema	1
1.2 - Objetivos	4
2. REVISÃO BIBLIOGRÁFICA	6
2.1 - Análise Gráfica dos Resíduos	7
2.1.1 - Condição ideal para análise de dados	8
2.1.2 - Heterogeneidade de variâncias	9
2.1.3 - Não independência dos erros	11
2.1.4 - Presença de dados discrepantes ("outliers")	12
2.1.5 - Ausência de normalidade	14
2.2 - Inferência sobre Resíduos	17
2.2.1 - Teste para não-aditividade removível	17
2.2.2 - Testes para verificação de heterocedasticidade das variâncias	21
2.2.2.1 - Teste para verificação da homocedasticidade, proposto por ANSCOMBE e TUKEY (1963)	23
2.2.2.2 - Teste de BARTLETT (1937)	25
2.2.2.3 - Teste de HARTLEY (1950)	27
2.2.2.4 - Teste de COCHRAN (1947)	29

	Pág.
2.2.2.5 - Uma aproximação do teste de Hartley (HAN, 1969)	30
2.2.3 - Teste para não independência dos erros	32
2.2.4 - Testes para verificação da presença de dados discrepantes ("outliers")	34
2.2.4.1 - Uma regra de rejeição para dados discrepantes (ANSCOMBE e TUKEY, 1963)	36
2.2.5 - Testes para não normalidade dos erros	38
2.2.5.1 - Teste de LILLIEFORS (1967)	39
2.3 - Procedimentos Alternativos para Amenizar Transgressões no Modelo de Gauss-Markov	41
2.3.1 - Heterogeneidade de Variâncias	42
2.3.1.1 - Transformações que estabilizam a variância	42
2.3.2 - Não independência dos erros	47
2.3.3 - Presença de dados discrepantes ("outliers")..	48
2.3.4 - Ausência de normalidade	53
3. MATERIAL E MÉTODOS	55
3.1 - Material	55
3.2 - Métodos	61
3.2.1 - Obtenção dos resíduos	61
3.2.1.1 - Delineamento Inteiramente ao Acaso	61
3.2.1.2 - Delineamento em Blocos Casualizados	63
3.2.1.3 - Delineamento em Quadrado Latino ..	65
3.2.1.4 - Esquema em Látice	67

	Pág.
3.2.2 - Análise gráfica dos resíduos	69
3.2.2.1 - Obtenção do gráfico dos resíduos contra a variável em estudo	69
3.2.2.2 - Obtenção do gráfico no "papel normal de probabilidade"	69
3.2.3 - Inferências sobre os resíduos	70
3.2.3.1 - Teste F para não-aditividade	71
3.2.3.2 - Teste de Cochran	74
3.2.3.3 - Modificação de dados discrepantes ("outliers")	75
3.2.3.4 - Teste de Lilliefors	77
3.2.3.5 - Teste de Kolmogorov-Smirnov	80
3.2.4 - Transformação de dados	81
4. APRESENTAÇÃO DE UM EXEMPLO.....	85
4.1 - Introdução	85
4.2 - Análise de Resíduos Considerando o Modelo de Látice	86
4.3 - Análise de Resíduos Considerando o Modelo de Blocos Casualizados	94
5. CONCLUSÕES	108
6. REFERÊNCIAS BIBLIOGRÁFICAS	110
7. APÊNDICES	114

LISTA DE TABELAS

TABELA		Pág.
1	Nove diferentes escolhas para o parâmetro c para o cálculo das probabilidades para o "papel normal de probabilidade"	15
2	Valores do coeficiente de regressão linear simples b , com a transformação correspondente	46
3	Clones estudados e seus respectivos parentais e espécies	57
4	Dados de produção de borracha seca (P.B.S.)	58
5	Valores de Y_{ij} estimados e valores ordenados dos resíduos padronizados de um experimento em Látice, fornecidos pelo programa ROSAN	87
6	Resultados dos testes de não-aditividade e teste de Lilliefors, fornecidos pelo programa ROSAN	91
7	Dados de produção de borracha seca (P.B.S.), dispostos em Blocos Casualizados	94
8	Valores de Y_{ij} estimados e valores ordenados dos resíduos padronizados dos dados dispostos em Blocos ao Acaso, fornecidos pelo programa ROSAN	96
9	Resultados dos testes de não-aditividade e teste de Lilliefors, fornecidos pelo programa ROSAN	100

TABELA		Pág.
10	Valores de Y_{ij} transformados, obtidos através do programa ROSA1	103
11	Valores de Y_{ij} estimados e valores ordenados dos resíduos padronizados, obtidos a partir dos Y_{ij} transformados, fornecidos pelo programa ROSAN	104
12	Resultados do teste de não-aditividade e teste de Lilliefors, fornecidos pelo programa ROSAN	106

LISTA DE FIGURAS

FIGURA		Pág.
1	Representação gráfica dos resíduos padronizados, quando existem as condições ideais para análise dos dados	8
2	Representação gráfica dos resíduos padronizados, indicando aumento da variância à medida em que \bar{Y}_{ij} aumenta	9
3	Representação gráfica dos resíduos padronizados, indicando diminuição da variância à medida em que \bar{Y}_{ij} aumenta	10
4	Representação gráfica dos resíduos padronizados, indicando aumento de variância para valores médios de \bar{Y}_{ij}	10
5	Representação gráfica dos resíduos padronizados, indicando diminuição da variância para valores médios de \bar{Y}_{ij}	10
6	Representação gráfica dos resíduos padronizados, indicando correlação positiva entre eles	11
7	Representação gráfica dos resíduos padronizados, indicando correlação negativa entre eles	11
8	Representação gráfica dos resíduos padronizados, indicando presença de dados discrepantes ("outliers")	13
9	Comportamento gráfico dos resíduos padronizados num papel normal de probabilidade"	16

FIGURA		Pág.
10	Representação gráfica dos resíduos padronizados versus Y_{ij} estimados, de dados de um experimento em Látice ..	88
11	Representação gráfica dos resíduos padronizados de um experimento em Látice, no "papel normal de probabilidade"	90
12	Representação gráfica dos resíduos padronizados versus Y_{ij} estimados, dos dados dispostos em Blocos ao Acaso	97
13	Representação gráfica dos resíduos padronizados dos dados dispostos em Blocos ao Acaso, no "papel normal de probabilidade"	99
14	Representação gráfica dos resíduos padronizados versus Y_{ij} estimados, depois da modificação de dados discrepantes	105

ASPECTOS DA ANÁLISE DE RESÍDUOS

Autora: Rosana Cristina Pereira Parente
Orientador: Prof. Dr. Antonio Francisco Iemma

R E S U M O

O modelo linear de Gauss-Markov tem sido de grande valia para os pesquisadores das Ciências Agropecuárias. No entanto, sua utilização requer uma análise exploratória dos dados, com o objetivo de se constatarem possíveis violações das suas pressuposições básicas.

Uma alternativa interessante para a execução da análise exploratória é, sem dúvidas, a análise de resíduos.

Neste estudo apresenta-se uma compilação das técnicas mais significativas sobre a análise de resíduos, visando atingir a pesquisadores das Ciências Agropecuárias que eventualmente não conheçam o assunto, com o objetivo de alertá-los sobre o problema e de divulgar os procedimentos mais importantes.

Apresenta-se, ainda, a discussão resumida de um exemplo numérico, e a listagem de um programa em linguagem FORTRAN, para computadores IBM 1130, elaborado para aplicação de testes sobre os resíduos.

Na discussão do exemplo numérico evidencia-se a vantagem da análise gráfica de resíduos, que independe da estrutura dos dados, em relação à inferência sobre resíduos, que traz consigo toda uma gama de pressuposições sobre os aspectos distribucionais dos dados. Evidenciou-se, também, a vantagem do processo de modificação de dados discrepantes sobre os métodos de descarte, em experimentos que envolvem mais de um fator.

ASPECTS OF RESIDUALS ANALYSIS

Author: Rosana Cristina Pereira Parente
Adviser: Prof. Dr. Antonio Francisco Iemma

S U M M A R Y

The linear model of Gauss-Markov has been helped the sciences of agropecuary seachers very much. Although its utilization requires an exploratory analysis of the data, with the objective to detect possibles violations in its basic pressuppositions.

An interesting alternative to execute the exploratory analysis is certainly the residual analysis.

In this study we present a compilation of the most significative technics about the residual analysis, objecting to attain sciences agricultural searchers, that sometimes do not know the subject, to stir up them about the problem and divulgating the most important proceedings.

We also present a summary discussion of one numerical example and the roll of a FORTRAN program for IBM-1130 computer, elaborated to apply on residuals tests.

In the discussion of the numerical example was evident the advantage of the graphical analysis of the residue, that is not dependent of the structure of the data, in relation to the Inference on the Residuals, that bring with itself, all the structure of presuppositions about the distributional aspect of the data.

Was evident, as well, the advantage of the process of modification of outliers on the outcast methods, in experiment that involving more than one factor.

1. INTRODUÇÃO

1.1 - Localização do Problema

Quando elege um modelo linear para a análise de um conjunto de dados, o pesquisador está aceitando o fato de que os dados em questão satisfazem às pressuposições inerentes ao modelo adotado. Desse modo, é importante que se analise a adequação modelo \leftrightarrow dados, para que os resultados, em termos de inferências, sejam confiáveis.

Uma das técnicas usadas para essa verificação é a análise de resíduos.

De modo geral, os autores têm usado a análise de resíduos para detectar os seguintes problemas em relação ao modelo matemático adotado:

- se o modelo matemático proposto não é apropriado;
- se os erros não têm variância constante;
- se os erros não são independentes;
- se existem dados discrepantes ("outliers");
- se os erros não têm distribuição normal.

Quando, para a análise e interpretação de um conjunto de dados provenientes de algum tipo de experimento, usa-se a técnica da análise de variância através do modelo de Gauss-Markov, pressupõe-se um modelo linear e a aceitação de algumas hipóteses básicas, a saber:

- os diversos efeitos são aditivos. Essa condição é imposta pelo modelo, que varia de acordo com o delineamento usado;
- os erros são não correlacionados, isto é, a probabilidade de que o erro de uma observação qualquer tenha um determinado valor não deve depender dos valores dos outros erros;
- os erros têm, todos, a mesma variância;
- os erros são normalmente distribuídos.

NETER e WASSERMAN (1974) argumentam que para a realização de uma análise de variância é conveniente examinar primeiramente se o modelo adotado é apropriado

para o conjunto de dados. Se o modelo não for o apropriado, então devem ser usadas medidas amenizadoras, que vão desde uma simples transformação dos dados até a escolha de um novo modelo. Somente depois de se verificar a adequacidade do modelo, é que a análise dos dados deve ser emprelendida.

SCHEFFÉ (1959) e NETER e WASSERMAN (1974), dentre outros, mostram que não é absolutamente necessário, nem usualmente possível, que as pressuposições do modelo sejam, todas, perfeitamente satisfeitas. Segundo esses autores, a análise de variância é razoavelmente robusta contra certos tipos de violações do modelo, como, por exemplo, a falta de normalidade dos erros, desde que sua distribuição seja ao menos simétrica.

A literatura tem mostrado que, se os dados não seguem a estrutura de um modelo de Gauss-Markov, $\epsilon \sim N(\phi, I\sigma^2)$, existem outros tipos de análises, através dos quais pode-se estudar o comportamento dos dados. Assim, por exemplo, se as variâncias dos tratamentos não forem iguais, pode-se adotar o método dos Mínimos Quadrados Ponderados; se existir correlação o método dos Mínimos Quadrados Generalizados e, sob ausência de normalidade, uma Análise Não-Paramétrica aos dados pode ser uma alternativva.

No entanto, é muito mais cômodo, e geralmente desejável, principalmente para pesquisadores que fazem dos procedimentos estatísticos apenas ferramentas de trabalho, que o modelo adotado tenha uma estrutura de Gauss-Markov. Assim sendo, as análises são obtidas com grande simplicidade, e a adoção das restrições tradicionais nas soluções e/ou nos parâmetros, permite que a estimação e os testes sejam obtidos sem grandes recursos computacionais.

Nesse contexto, a Análise de Resíduos afigura-se como uma técnica fundamentalmente importante, pois permite verificar se o conjunto de dados em estudos satisfaz as exigências do modelo de Gauss-Markov. Ademais, dada sua simplicidade e consistência, mormente na versão gráfica, a Análise de Resíduos praticamente não tem competidores, constituindo-se numa poderosa aliada dos pesquisadores ao longo do tempo.

1.2 - Objetivos

Apesar da importância incontestável da Análise de Resíduos, suas técnicas têm sido pouco divulgadas, exceção feita a grandes centros de pesquisa que dispõem de recursos computacionais com capacidade para "pacotes" mais elaborados, como por exemplo o S.A.S. ("Statistical Analy

sis Systems). A falta de divulgação entre os pesquisadores da área agronômica, crê-se, deve-se principalmente ao fato de que as publicações nessa linha são feitas geralmente em revistas de Estatística e correlatas, que nem sempre são de fácil acesso ou despertam interesse em tais pesquisadores.

Desse modo, emergiu como um dos objetivos desse estudo, a apresentação de uma compilação ordenada, suscinta e acessível a esses pesquisadores, dos artigos mais significativos existentes na bibliografia da Análise de Resíduos, visando, no mínimo, alertá-los para o problema e, se possível, apresentar subsídios de reflexão para aqueles já iniciados na área.

Além disso, a discussão de um exemplo, e a elaboração de um programa em linguagem FORTRAN para computadores de pequeno porte como o IBM 1130, são apresentados, com o objetivo de divulgar o uso da Análise de Resíduos.

2. REVISÃO BIBLIOGRÁFICA

ANSCOMBE e TUKEY (1963) definem resíduo como a diferença entre o valor observado e o valor ajustado. Como tal, o resíduo pode ser identificado como o valor observado do erro (em distinção ao verdadeiro e desconhecido erro).

Para COX e SNELL (1968), no contexto da teoria normal dos modelos lineares de Gauss-Markov, o vetor das variáveis aleatórias $Y_{n \times 1}$ tem a forma;

$$Y = X\beta + \varepsilon, \quad (1)$$

onde: X é uma matriz conhecida;

β é o vetor de parâmetros desconhecidos; e

ϵ e o vetor de variáveis aleatórias não observadas
 $\epsilon \sim N(\phi, I\sigma^2)$.

Se β é o vetor solução de mínimos quadrados para β , os resíduos e podem ser definidos na equação:

$$Y - X\beta = e \quad (2)$$

Para COX e SNELL (1968), desde que o número de parâmetros é pequeno comparado com n , muitas das propriedades de e são próximas daquelas de ϵ , isto é, e deve ter aproximadamente as propriedades de uma amostra aleatória de distribuição normal.

O procedimento usual, após a obtenção dos resíduos, consiste na execução da análise gráfica de resíduos.

2.1 - Análise Gráfica dos Resíduos

Para DRAPER e SMITH (1966), SEBER (1977) e BOX *et alii* (1978), as principais maneiras de plotar os resíduos são:

- total, fazendo um gráfico de pontos com todos os resíduos;
- em sequência de tempo, se a ordem é conhecida;
- contra valores ajustados \hat{Y}_{ij} ;

- contra a variável independente X_{ij} ;
- em qualquer direção que seja sensível ao problema particular considerado.

DRAPER E SMITH (1966), NETER e WASSERMAN (1974), SEBER (1977) e BOX *et alii* (1978), dentre outros, aconselham fazer a análise gráfica dos resíduos com os resíduos considerados na forma $d_{ij} = \frac{e_{ij}}{\sqrt{QMRes.}}$, que são chamados resíduos padronizados. Desta maneira sua dispersão variará em torno de (-3; 3).

2.1.1 - Condição ideal para análise de dados

Se o modelo linear escolhido para o estudo dos dados através da análise de variância for o apropriado, é de se esperar que as hipóteses básicas impostas por ele sejam atendidas. Quando isso acontece, o gráfico dos resíduos padronizados contra o valor ajustado (\hat{Y}_{ij}) apresentará o comportamento mostrado na Figura 1.

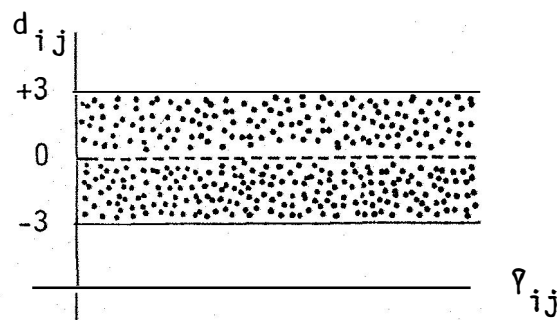


Figura 1 - Representação gráfica dos resíduos padronizados, quando existem as condições ideais para análise dos dados.
 FONTE: NETER e WASSERMAN (1974).

A literatura é escassa em mostrar gráficos onde se possa visualizar, por exemplo, a presença da interação entre dois fatores no modelo. Para tanto, um teste de não-aditividade proposto por TUKEY (1949) é apresentado em 2.2.1.

2.1.2 - Heterogeneidade de variâncias

A não constância da variância pode ser detectada em um gráfico dos resíduos padronizados contra o valor ajustado \hat{Y}_{ij} , ou contra a variável independente X_{ij} , apresentando os comportamentos mostrados nas Figuras 2, 3, 4 e 5.

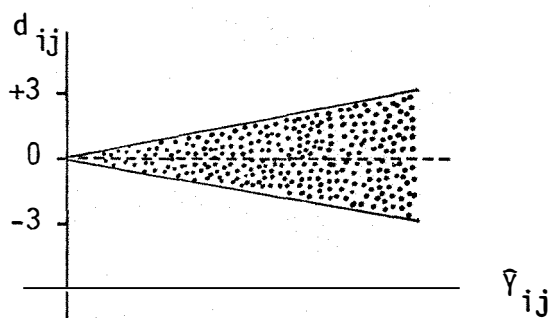


Figura 2 - Representação gráfica dos resíduos padronizados, indicando aumento da variância à medida em que \hat{Y}_{ij} aumenta.

FONTE: NETER e WASSERMAN (1974).

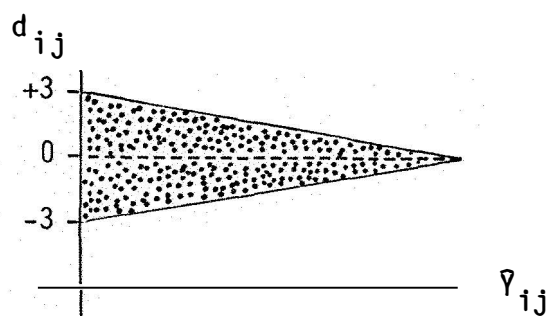


Figura 3 - Representação gráfica dos resíduos padronizados, indicando diminuição de variância à medida em que \bar{Y}_{ij} aumenta.

FONTE: DRAPER e SMITH (1966).

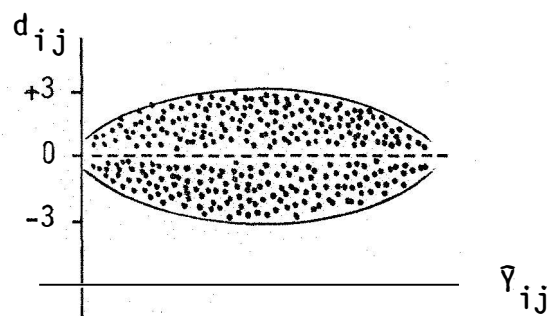


Figura 4 - Representação gráfica dos resíduos padronizados, indicando aumento de variância para valores médios de \bar{Y}_{ij} .

FONTE: DRAPER e SMITH (1966).

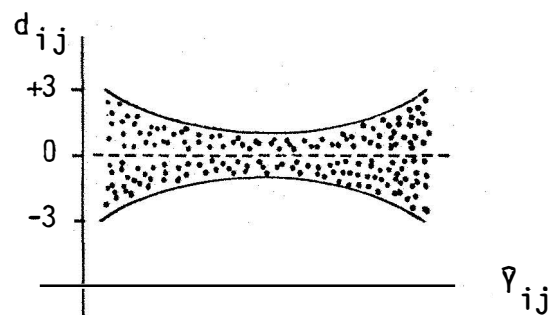


Figura 5 - Representação gráfica dos resíduos padronizados, indicando diminuição da variância para valores médios de \bar{Y}_{ij} .

FONTE: DRAPER e SMITH (1966).

2.1.3 - Não independência dos erros

Para DRAPER e SMITH (1966), BOX *et alii* (1978), NETER e WASSERMAN (1974), ASCOMBE e TUKEY (1963), dentre outros autores, sempre que os dados são obtidos em uma sequência de tempo, ainda que este fator não seja incorporado explicitamente no modelo, é aconselhável se fazer o gráfico dos resíduos contra o tempo, para verificar se existe alguma correlação dos erros no tempo. O comportamento dos resíduos, indicando não independência dos erros, pode ter os aspectos mostrados nas Figuras 6 e 7.

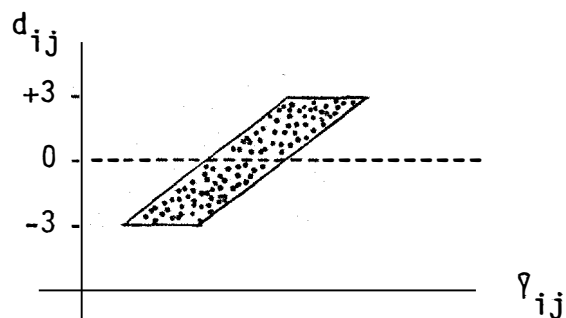


Figura 6 - Representação gráfica dos resíduos padronizados, indicando correlação positiva entre eles.
 FONTE: NETER e WASSERMAN (1974).

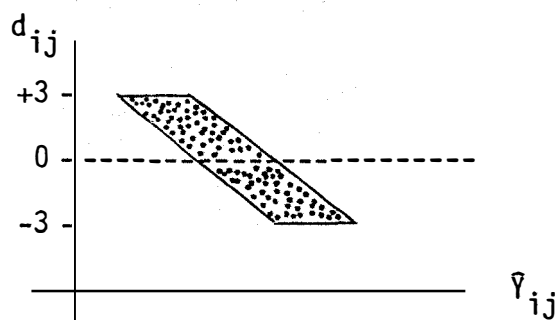


Figura 7 - Representação gráfica dos resíduos padronizados, indicando correlação negativa entre eles.
 FONTE: NETER e WASSERMAN (1974).

2.1.4 - Presença de dados discrepantes ("outliers")

"Outliers" são observações discrepantes, ou seja, que possuem um resíduo muito grande em comparação com a maioria dos resíduos.

Para ANSCOMBE e TUKEY (1963), a razão mais importante para o cálculo de resíduos é detectar um dado discrepante. Segundo esses autores, uma das causas da ocorrência de dados discrepantes pode ser a existência de erros na leitura da escala (como quando o último dígito é lido incorretamente), ou um erro de cálculo ao transformar o valor da escala para o valor de Y a ser utilizado.

SEBER (1977) recomenda um estudo cuidadoso de um dado discrepante, que por vezes pode ser simples erro de transcrição, mas que pode ser, também, um indicativo de uso incorreto do modelo.

Para DRAPER e SMITH (1966), uma observação é considerada um dado discrepante se ela é maior que o restante, em valor absoluto, pelo menos três vezes o desvio padrão da média dos resíduos.

Na análise gráfica dos resíduos uma observação será considerada um dado discrepante quando apresentar o comportamento mostrado na Figura 8.

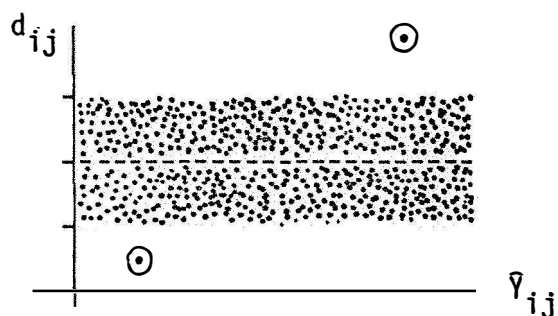


Figura 8 - Representação gráfica dos resíduos padronizados, indicando presença de dados discrepantes ("outliers").
 FONTE: NETER e WASSERMAN (1974).

Quando se verifica a presença de um dado discrepante numa análise gráfica de resíduos, este fato pode levar à ausência de normalidade dos erros visivelmente percebida quando usa-se o papel normal de probabilidade, e fatalmente à heterogeneidade de variâncias, uma vez que um dado discrepante altera substancialmente a variância do grupo onde ele é detectado.

Com respeito à falta de normalidade, ANSCOMBE e TUKEY (1963) propõem a modificação do dado discrepante, mas a grande dificuldade disso reside no fato de se desconhecer qual a distribuição dos resíduos. Este procedimento está descrito em (2.3.4.1), apenas para o caso em que os resíduos podem ser estudados através de uma distribuição t de Student.

2.1.5 - Ausência de normalidade

A literatura tem mostrado que a normalidade pode ser estudada através de várias formas de gráficos.

BOX *et alii* (1978), NETER e WASSERMAN (1974), entre outros, para a verificação da normalidade fazem uso da construção de histogramas ou de gráficos de folhas, com todos os resíduos, aos quais dão o nome de "plot" total, e observam se eles se comportam como uma curva normal centrada em zero.

Uma outra forma de verificação gráfica de normalidade pode ser obtida através do uso do papel normal de probabilidade.

Para SEBER (1977), se $d_{(1)} < d_{(2)} < \dots < d_{(n)}$ representam n resíduos padronizados ordenados, e para n moderadamente grande, um gráfico de $d_{(i)}$ contra $[i-(1/2)]/n$ no papel probabilístico detectará uma acentuada ausência de normalidade. Quando os dados são normais, esses pontos devem estar dispersos em torno de uma reta $y = x$.

Kimbal (1960), citado por MAGE (1982) revisa o problema da escolha de posição para gráficos das observações ordenadas. Comparou diversos tipos de gráficos convencionais e sugeriu que a escolha de uma fórmula para gráfico depende, em parte, do propósito da análise. Em ge

ral, as fórmulas para a construção do gráfico têm sido descritas como $P_i = (i-c)/(n+1-2c)$, onde c é função de n , e $0 \leq c \leq 1$.

Na prática, muitos autores simplificam o uso dessa fórmula, assumindo ser c uma constante.

A Tabela 1, tirada de MAGE (1982, p. 117) apresenta nove diferentes escolhas para o parâmetro c .

Tabela 1 - Nove diferentes escolhas para o parâmetro c para o cálculo das probabilidades para o papel normal de probabilidade.

FÓRMULAS	Valores de c	Posição de plot (P_i)	Referências
1	0,0000	$(i)/(n+1)$	GUMBEL (1954)
2	0,3000	$(i-0,3)/(n+0,4)$	BERNARD e BOS-LEVENBACH(1953)
3	0,3175	$(i-0,3175)/(n+0,365)$	FILLIBEN (1975)
4	0,3333	$(i-0,33)/(n+0,33)$	BIOMEDICAL (1979)
5	0,3750	$(i-0,375)/(n+0,25)$	BLOM (1958)
6	0,4000	$(i-0,4)/(n+0,2)$	LARSEN, CURRAN e HUNT (1980)
7	0,4400	$(i-0,44)/(n+0,12)$	GRINGORTEN (1963)
8	0,5000	$(i-0,5)/(n)$	HAHN e SHAPIRO (1967); WILK e GNANADESIKAN (1966); BOWKER e LIEBERMAN (1972)
9	0,5670	$(i-0,567)/(n-0,134)$	LARSEN, CURRAN e HUNT (1980)

FONTE: MAGE (1982).

Para a verificação da normalidade, a literatura tem recomendado o uso do papel normal de probabilidade, construindo-se o gráfico dos resíduos padronizados contra os valores obtidos através da fórmula 8 ($c=0,500$) da Tabela 1.

A Figura 9 mostra o comportamento gráfico dos resíduos com distribuição normal num "papel normal de probabilidade".

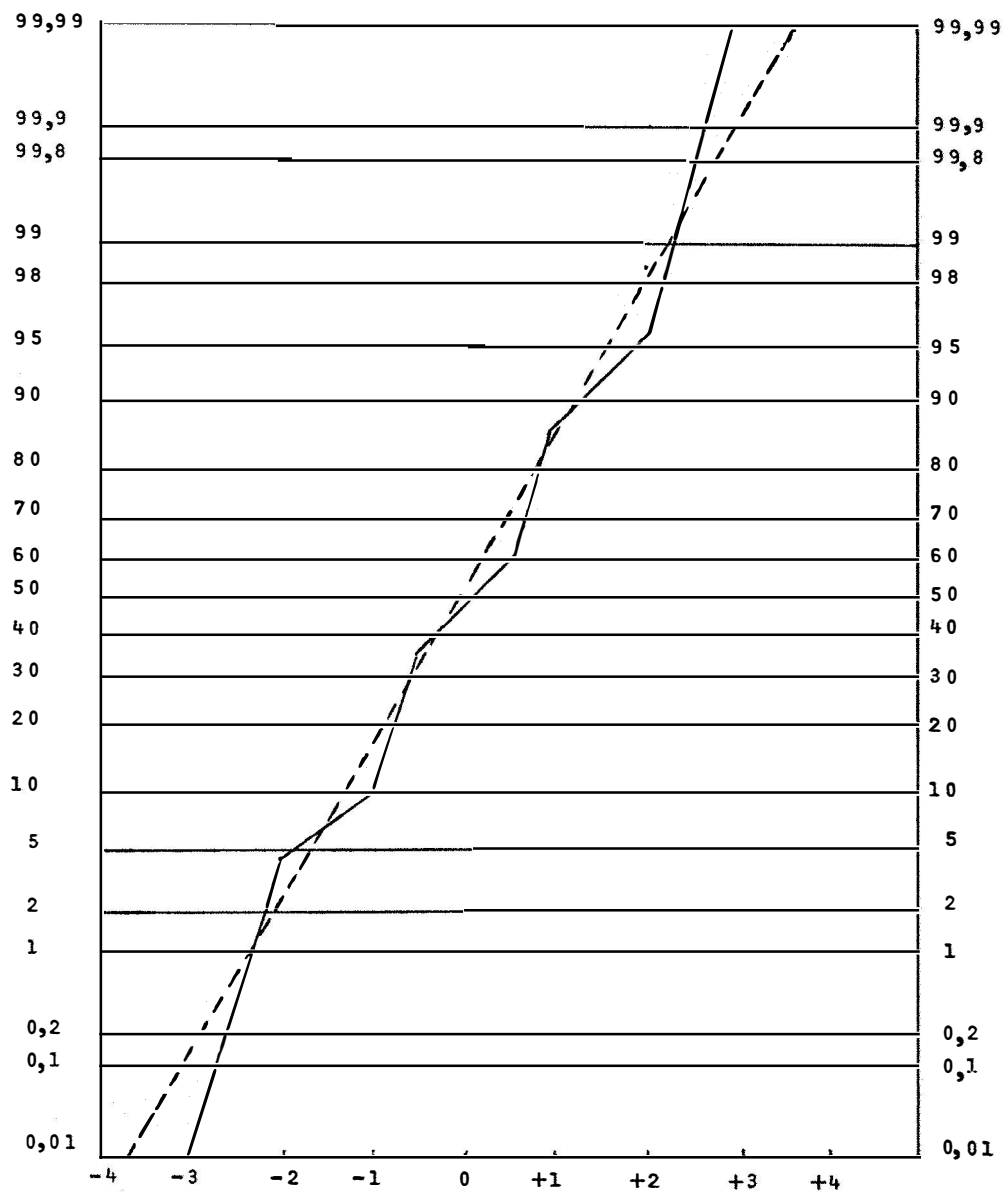


Figura 9 - Comportamento gráfico dos resíduos padronizados num "papel normal de probabilidade".
FONTE: NETER e WASSERMAN (1974).

2.2 - Inferência sobre Resíduos

Para a confirmação ou não da análise gráfica dos resíduos podem ser aplicados testes estatísticos.

2.2.1 - Teste para não-aditividade removível

Para a verificação da suposição da aditividade, quando se adota um modelo para estudo de um conjunto de dados através da análise de variância, um teste apropriado é proposto por TUKEY (1949).

O teste de TUKEY (1949) para não aditividade consiste basicamente em testar as hipóteses:

H_0 : o modelo é aditivo;

H_a : o modelo não é aditivo.

Para o procedimento do teste, calcula-se

$$K = B^2/A,$$

onde,

A é a soma de quadrados do resíduo obtida aplicando-se o procedimento convencional da análise de variância, ao quadrado dos valores ajustados (\hat{Y}_{ij}), em lugar das observações originais (Y_{ij}). Antes de se elevar ao quadrado os \hat{Y}_{ij} , pode-se subtrair alguma constante de \hat{Y}_{ij} , tal como a média \bar{Y} , sem que se afetem os valores de A ;

B é uma soma de produtos:

$$B = \sum_{i,j} (Y_{ij} - \bar{Y}_{ij})(\bar{Y}_{ij})^2$$

ou

$$B = \sum_{i,j} (Y_{ij} - \bar{Y}_{ij})(\bar{Y}_{ij} - \bar{Y})^2$$

Uma vez obtido $K = B^2/A$, este termo será subtraído da soma de quadrados do resíduo da tabela da análise de variância original, e será então chamado de SQ Restante, onde:

$$\text{SQ Restante} = \text{SQ Resíduo} - K$$

e os graus de liberdade correspondentes a essa nova soma de quadrados serão:

$$\text{g.l. restante} = \text{g.l. resíduo} - 1$$

Em seguida, calcula-se a estatística QMNA/QMNR que, sob a veracidade de H_0 , tem distribuição F com 1 e g.l. restante graus de liberdade.

Na tabela da análise de variância, se eventualmente o delineamento usado for o de Blocos ao Acaso, tem-se o seguinte esquema.

Causas de Variação	G.L.	S.Q.	Q.M.	F
Blocos	J-1	SQB		
Tratamentos	I-1	SQT		
Não-Aditividade	1	$SQ\bar{N}A = K$	$QM\bar{N}A$	$QM\bar{N}A/QMNR$
Novo Resíduo	$(I-1)(J-1)-1$	$SQRes-SQ\bar{N}A$	$QMNR$	
Resíduo	$(I-1)(J-1)$	SQR		
TOTAL	IJ-1			

TUKEY (1949) avalia fórmulas especiais para delineamentos e esquemas particulares, como Blocos Casualizados, Fatorial e Quadrado Latino.

ANSCOMBE e TUKEY (1963), para dar expressão numérica à aparência de não-aditividade removível, recorrem ao coeficiente de regressão dos resíduos sobre os valores ajustados (\bar{Y}_{ij}), estes elevados ao quadrado.

O coeficiente de regressão é:

$$K = B/A ,$$

com B e A calculados como anteriormente descrito.

Segundo ANSCOMBE e TUKEY (1963), um procedimento seguro consiste em:

- 1º) obter todos os resíduos, através da decomposição da nova variável $X_{ij} = Y_{ij}^2$ ou $X_{ij} = (Y_{ij} - \bar{Y})^2$,

onde \bar{Y}_{ij} é o valor ajustado da observação original Y_{ij} ;

2º) uma vez calculado esse novo resíduo, o qual é definido por $z_{ij} = (X_{ij} - \bar{X}_{ij})$, procede-se ao cálculo do valor de A e B, respectivamente, da seguinte maneira:

$$B = \sum_{i,j} e_{ij} z_{ij} \quad ; \quad A = \sum_{i,j} z_{ij}^2 \quad ,$$

onde e_{ij} é o resíduo obtido da variável original Y_{ij} .

Para ANSCOMBE e TUKEY (1963), uma importante vantagem em se calcular todos os resíduos z_{ij} reside no fato de que se pode fazer um novo diagrama de dispersão, no qual cada ponto tem o conhecido resíduo e_{ij} como ordenada e o novo resíduo z_{ij} como abscissa. Se a não-aditividade removível estiver presente nos dados, o diagrama mostrará uma regressão linear de e_{ij} sobre z_{ij} , sem dúvida mais facilmente perceptível que uma regressão quadrática de e_{ij} sobre \bar{Y}_{ij} . Tendo-se obtido K substancialmente diferente de zero, uma alternativa para remover a não-aditividade seria considerar uma transformação das observações, o que será visto no item 2.3.

Para a realização desses cálculos, ANSCOMBE e TUKEY (1963) recomendam o uso de resíduos sem peso,

e na sua forma normal, isto é, sem a padronização dos mesmos.

BOX *et alii* (1978) apresentam o teste de não-aditividade da seguinte forma:

$$SQ \bar{N}A = \frac{\sum_{i,j} (Y_{ij} - \bar{Y}_{ij})(q_{ij} - \hat{q}_{ij})}{\sum_{i,j} (q_{ij} - \hat{q}_{ij})^2}$$

onde:

$(Y_{ij} - \bar{Y}_{ij})$ é o resíduo dos dados originais;

$q_{ij} = \bar{Y}_{ij}^2$;

\hat{q}_{ij} = estimativa de q_{ij} ;

$(q_{ij} - \hat{q}_{ij})$ = novo resíduo.

Pode-se verificar que este procedimento pode ser usado em qualquer tipo de delineamento.

2.2.2 - Testes para verificação de heterocedasticidade das variâncias

Para se efetuar uma análise de variância, uma das pressuposições é que as variâncias devem ser homogêneas. Entretanto, pode-se usualmente encontrar casos em que a variância não é a mesma nos vários grupos de dados. A isso chama-se de Heterocedasticidade ou Heterogeneidade de suas variâncias.

Cochran (1947), citado por DEMÉTRIO (1978), mostra que a heterogeneidade dos erros pode ser classificada como irregular e regular.

Segundo DEMÉTRIO (1978), Cochran (1947) diz que a heterogeneidade é irregular quando aparentemente não existe uma relação entre médias e variâncias. Neste caso, o procedimento empregado é omitir certos tratamentos ou subdividi-los de tal forma que, com os tratamentos restantes, ou dentro de cada subdivisão, tenha-se a homocedasticidade. Como alternativa, sugere decompor o quadrado médio do resíduo em componentes apropriados às várias comparações de interesse.

Ainda segundo o autor, citado por DEMÉTRIO (1978), a heterogeneidade do tipo regular usualmente decorre da não normalidade dos dados, existindo uma certa relação entre a média e a variância dos vários tratamentos. Sendo conhecida a distribuição da qual são provenientes os dados, a relação entre a média e a variância dos tratamentos é também conhecida e, nestes casos, os dados podem ser transformados de modo que passem a ter distribuição aproximadamente normal e as médias e variâncias se tornem independentes, resultando também em variâncias homogêneas.

2.2.2.1 - Teste para verificação da homocedasticidade, proposto por ANSCOMBE e TUKEY (1963)

ANSCOMBE e TUKEY (1963) propõem, para a verificação de heterogeneidade, um teste que consiste basicamente no ajuste de uma equação de regressão linear simples de e_{ij}^2 sobre \bar{Y}_{ij} . Segundo eles, a homocedasticidade é aceita se, pelo teste t, a soma dos desvios do coeficiente de regressão linear for estatisticamente igual a zero.

Para a aplicação do teste, os autores utilizam o coeficiente de regressão linear, dividido pelo quadrado médio do resíduo, indicado por h , da seguinte forma:

$$h = \frac{\sum_{i,j} e_{ij}^2 (\bar{Y}_{ij} - \bar{Y})}{s^2 H}$$

onde H , em alguns casos, pode ser facilmente calculado.

Em estudo anterior, ANSCOMBE (1961) obteve a variância h , dada por

$$\text{Var}(h) = \frac{2v}{(v+2)H} \quad 1/$$

onde v é o número de graus de liberdade do resíduo, e H , para o caso de tabelas de dupla entrada, é da forma:

1/ A demonstração é dada em ANSCOMBE (1961), assim como a obtenção de uma fórmula geral de H .

$$H = \frac{(r-2)(c-1)}{rc} [\text{SQ Linhas}] + \frac{(r-1)(c-2)}{rc} [\text{SQ Colunas}],$$

onde:

r e o numero de linhas;
 c e o numero de colunas.

Para uma classificação simples, com o mesmo número de observações em cada coluna,

$$H = \frac{r-1}{r} [\text{SQ Colunas}] ,$$

ou

$$H = \frac{r-1}{r} [\text{SQ Tratamentos}] .$$

Para um Quadrado Latino $r \times r$, com k observações por parcela,

$$H = \frac{(r-2)(r-3)}{r^2} [\text{SQ Linhas} + \text{SQ Colunas} + \text{SQ Tratamentos}].$$

Fazendo considerações a respeito de h , ANS COMBE (1961) fala das dificuldades para sua obtenção e suas restrições de uso, indicando como uma alternativa para a verificação de homocedasticidade, a análise gráfica dos resíduos.

2.2.2.2 - Teste de BARTLETT (1937)

Sejam S_1^2, \dots, S_k^2 variâncias amostrais de k populações, definidas por:

$$S_i^2 = \frac{\sum_{j=1}^{n_i} (Y_{ij} - \bar{Y})^2}{n_i - 1} .$$

Sejam também:

$A = \frac{1}{n-k} \sum_{i=1}^k (n_i - 1) S_i^2$, a média aritmética ponderada das variâncias amostrais; e

$G = \left[(S_1^2)^{n_1-1} (S_2^2)^{n_2-1} \dots (S_k^2)^{n_k-1} \right]^{1/(n-k)}$, a média geométrica ponderada das variâncias amostrais.

Mostra-se que $G \leq A$ é sempre verdadeira, e a igualdade ocorre quando as variâncias forem iguais, ou seja:

$$A = G \quad \text{se} \quad S_1^2 = S_2^2 = \dots = S_k^2 .$$

Se A/G é próximo de 1 (um), as variâncias são consideradas iguais. Se A/G for grande indicará que as variâncias são desiguais. As mesmas conclusões são vã-

lidas se considerar-se

$$\text{Log } \frac{A}{G} = \text{Log } A - \text{Log } G .$$

Bartlett (1937), citado por BARBOSA (1983) mostrou que, para grandes amostras, uma função de $\text{Log } A - \text{Log } G$ tem distribuição aproximadamente $\chi^2_{(k-1)}$. Esta função é:

$$E = \frac{2,302585}{L} [(r-k) \text{Log } A - \text{Log } G] ,$$

ou, ainda:

$$E = \frac{2,302585}{L} \left\{ (r-k) \text{Log} \left[\frac{1}{r-k} \sum_{i=1}^k (r_i-1) S_i^2 \right] - \sum_{i=1}^k (r_i-1) \text{Log } S_i^2 \right\} ,$$

onde:

$$L = 1 + \frac{1}{3(k-1)} \left[\sum_{i=1}^k \frac{1}{r_i-1} - \frac{1}{r-k} \right] .$$

Para se testar a hipótese $H_0: \sigma_1^2 = \dots = \sigma_k^2$, calcula-se E e compara-se com o valor de $\chi^2_{(k-1)}$. Se $E \geq \chi^2_{(k-1)}$ rejeita-se a hipótese de nulidade.

Quando as estimativas de variâncias possuem todas o mesmo número de graus de liberdade $(r-1)$, as fórmulas tornam-se mais simples:

$$E = \frac{2,302585}{L} (r-1) \left[k \operatorname{Log} \frac{\sum_{i=1}^k S_i^2}{k} - \sum_{i=1}^k \operatorname{Log} S_i^2 \right]$$

e

$$L = 1 + \frac{k+1}{3k(r-1)} .$$

NETER e WASSERMAN (1974) argumentam que o teste de Bartlett é muito sensível à ausência de normalidade, não sendo recomendada a sua aplicação quando os dados não seguem distribuição normal.

2.2.2.3 - Teste de HARTLEY (1950)

Se as amostras simples n_i são todas iguais ($r_i=r$), um simples teste para decidir entre:

$$H_0: \sigma_1^2 = \dots = \sigma_k^2 ,$$

H_1 : nem todas as σ_i^2 são iguais,

pode ser obtido em HARTLEY (1950).

É baseado unicamente na maior variância amostral, denotada por $\bar{m}ax (S_i^2)$, e a menor variância amostral, denotada por $\bar{m}in (S_i^2)$. O teste estatístico é:

$$H_c = \frac{\bar{m}ax (S_i^2)}{\bar{m}in (S_i^2)} ,$$

e compara o resultado com o valor $H_{(k,r-1)}$ da Tabela 31 de PEARSON e HARTLEY (1956), onde:

$k = n^\circ$ de grupos;

$r = n^\circ$ de dados em cada grupo.

Se $H_c > H_{(k,r-1)}$, rejeita-se H_0 , caso contrário, aceita-se H_0 .

Se os dados não são balanceados, toma-se

$$r = \frac{r(\text{mín}) + r(\text{máx})}{2} .$$

Se a variação do tamanho de cada grupo é moderada, toma-se:

$$r = \frac{\sum_{i=1}^k r_i}{k}$$

NETER e WASSERMAN (1974) comentam que o teste de Hartley, assim como o teste de Bartlett, é sensível à ausência da suposição de normalidade dos dados.

A eficiência deste teste é relativamente alta quando os tamanhos das amostras são iguais ($r_i=r$) e ocorre k < 12.

2.2.2.4 - Teste de COCHRAN (1947)

Este teste consiste em calcular o quociente

$$C_c = \frac{S^2(\max)}{\frac{k}{\sum_{i=1}^k S_i^2}},$$

e comparar com valores tabelados $C_{(k,r-1)}$, dados, por exemplo, na Tabela A-17 de DIXON e MASSEY (1969), para os níveis de 5% e 1% de probabilidade.

A hipótese de nulidade $H_0: \sigma_1^2 = \dots = \sigma_k^2$ será rejeitada se $C_c > C_{(k,r-1)}$, caso contrário não se rejeita H_0 .

Se os dados forem não balanceados, toma-se $r = \frac{k}{\sum_{i=1}^k r_i/k}$, e neste caso o teste é aproximado.

DEMÉTRIO (1984) diz que estes testes podem ser utilizados para o caso do delineamento Inteiramente ao Acaso, pois as variâncias amostrais obtidas são independentes. Além disso, adverte quanto ao uso dos mesmos em ensaios em Blocos Casualizados, uma vez que as variâncias amostrais não são independentes.

2.2.2.5 - Uma aproximação do teste de Hartley (HAN, 1969)

HAN (1969) sugere uma aproximação do teste H_C de HARTLEY (1950), quando se considera o seguinte modelo matemático:

$$Y_{ij} = \mu + \alpha_i + \beta_j + e_{ij}, \quad \begin{matrix} i=1,2,\dots,t \\ j=1,2,\dots,n, n>t \end{matrix}$$

onde:

μ é a média geral;

α_i é o efeito do i -ésimo tratamento;

β_j é o efeito do j -ésimo bloco;

e_{ij} são os erros com distribuição $N(0, \sigma_i^2)$.

Se S_i^2 é a variância amostral para a i -ésima coluna de z_{ij} , onde $z_{ij} = Y_{ij} - \bar{Y}_{.j}$, $i=1,2,\dots,t>3$ e $j=1,2,\dots,n$, então:

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (z_{ij} - \bar{z}_{i.})^2,$$

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} - \bar{Y}_{..})^2,$$

resultando:

$$S_i^2 = \frac{1}{n-1} \sum_{j=1}^n (e_{ij})^2.$$

Sabe-se que:

$$QM \text{ Res.} = \frac{1}{(t-1)(n-1)} \sum_{i,j} (e_{ij})^2$$

$$QM \text{ Res.} = \frac{1}{(t-1)} \sum_i \frac{\sum_j e_{ij}^2}{(n-1)}$$

$$QM \text{ Res.} = \frac{1}{(t-1)} \sum_i S_i^2, \quad ,$$

onde S_i^2 é a variância amostral dentro do tratamento i . Observa-se que a variância residual não é a média aritmética das variâncias amostrais. Segundo HAN (1969), isto decorre do fato das S_i^2 estarem correlacionadas com

$$COV(S_i^2, S_{i'}^2) = \frac{2(1-\rho)^2}{t^2(n-1)} \sigma^4, \quad ,$$

onde ρ é a correlação entre e_{ij} e e_{km} , se $j=m$; esta correlação será zero se $j \neq m$. O autor considera, no seu trabalho, que os erros são igualmente correlacionados dentro de linhas, mas independente entre linhas.

Quando $\rho = 0$, o modelo reduz-se ao modelo aditivo completamente independente, e diz-se que, para n suficientemente grande, S_i^2 tem uma distribuição normal assintótica, com média $\epsilon = \frac{(t-1)(1-\rho)}{t} \sigma^2$ e variância $\frac{2(t-1)^2(1-\rho)^2}{t^2(n-1)} \sigma^4$.

Trabalhando com

$$\text{Log } F_{\text{m}\bar{\text{a}}\text{x}} = \text{Log } s_{\text{m}\bar{\text{a}}\text{x}}^2 - \text{Log } s_{\text{m}\bar{\text{i}}\text{n}}^2 ,$$

HAN (1969) determina a distribuição assintótica de $F_{\text{m}\bar{\text{a}}\text{x}}$ para uma classificação com dois fatores, podendo ser aplicada no caso de Blocos Casualizados. Ele mostra que o valor crítico, ao nível α de significância de $F_{\text{m}\bar{\text{a}}\text{x}}$, sob a hipótese de homogeneidade de variância, fica:

$$F_{\text{m}\bar{\text{a}}\text{x}}(\alpha) \cong \exp \left\{ W_t(\alpha) \sqrt{\frac{2}{n-1} \left[1 - \frac{1}{(t-1)^2} \right]} \right\}$$

onde $W_t(\alpha)$ é o valor tabelado da distribuição da amplitude obtida por PEARSON e HARTLEY (1956) para uma amostra de tamanho t e com um coeficiente de confiança $100(1-\alpha)\%$.

Observa-se que a grande restrição deste teste é a de que exige $n > t$, o que nem sempre se encontra em experimentos de campo.

2.2.3 - Teste para não independência dos erros

ANSCOMBE e TUKEY (1963) mostram que os resíduos não são independentes, mas são correlacionados em vários graus. Dizem que os resíduos correspondentes a duas observações na mesma linha têm coeficiente de correlação

$-1/(c-1)$, onde c é o número de colunas, e tendo-se numa mesma coluna coeficiente de correlação $-1/(r-1)$, com r igual ao número de linhas. Se os resíduos correspondentes a duas observações se encontram entre linhas e colunas diferentes, seu coeficiente de correlação é $1/v$, onde $v = (r-1)(c-1)$. Com isso os resíduos estão sujeitos às restrições: têm que somar zero através de todas as linhas e ao longo de todas as colunas.

Quando $r=2$ e $c=2$, os pares de resíduos são idênticos, exceto pelo sinal, e muitas técnicas de exames de resíduos falham completamente. Quando $r=3$ e $c=3$, as trinças de resíduos têm intercorrelação de $-0,5$, e isto afeta substancialmente as propriedades simplistas amostrais de alguma combinação numérica de resíduos. Mas, segundo ANSCOMBE e TUKEY (1963), quando r e c são, no mínimo, iguais a quatro (4), os efeitos dessas correlações podem ser negligenciados para a análise gráfica de resíduos.

NETER e WASSERMAN (1974), quando examinam o problema de não independência dos erros na análise de variância, exemplificam apenas o caso em que os dados são obtidos em sequência de tempo, não mencionando nenhum tipo de teste estatístico. Apenas para o gráfico dos resíduos versus tempo, em que for encontrada uma das situações apresentadas nas Figuras 6 e 7, eles aconselham a inclusão no modelo

de um termo linear de efeito do tempo, que pode ser suficiente para assegurar a independência dos erros no modelo corrigido.

Pela literatura revista, DRAPER e SMITH (1966) e NETER e WASSERMAN (1974) citam os testes do Sinal e o de Durbin-Watson para a verificação da não-independência dos erros, mas somente nos casos de uso de regressão em dados obtidos em sequência de tempo.

Para NETER e WASSERMAN (1974), testes estatísticos que requeiram observações independentes não são exatamente apropriados para resíduos. Se, contudo, o número de resíduos para cada nível de fator (tratamentos) é grande, o efeito de correlação será pequeno. Num gráfico, os resíduos estão menos sujeitos ao efeito de correlação do que em testes estatísticos, porque os gráficos contêm os resíduos individuais e não simplesmente funções deles.

2.2.4 - Testes para verificação da presença de dados discrepantes ("outliers")

Para ANSCOMBE e TUKEY (1963), quando se suspeitar que a observação discrepante ("outlier") foi causada por um grosseiro erro de leitura, causa menos problemas no cálculo ou na condução da observação, descartá-la que utilizá-la.

Portanto, um procedimento muito simples e bastante utilizado é descartar a observação. Entretanto, algumas vezes é conveniente utilizar o dado discrepante na obtenção do valor modificado para a observação.

Quando se suspeita que os dados discrepantes são observações corretas de algum fenômeno pouco comum, então eles podem ser tomados como um peso reduzido, ou, ainda, eliminados da análise da variância e discutidos isoladamente. Um outro procedimento seria transformar todos os dados e aplicar a análise de variância usual, ou, então, usar testes não-paramétricos.

ANSCOMBE e TUKEY (1963) mostram, ainda, que quase todos os procedimentos para análise de resíduos são sensíveis à presença de dados discrepantes, em particular, os testes para não normalidade, não-aditividade removível, e não constância de variância. Então:

- Qualquer desses procedimentos poderia ser usado para testar dados discrepantes;
- O uso destes testes para dados discrepantes poderia gerar confusões, uma vez que um resultado positivo em algum deles poderia ser devido a dados discrepantes, ou ao verdadeiro propósito do teste;
- É geralmente perigoso aplicar um procedimento numérico aos resíduos antes de se estudarem os dados

discrepantes. Sem dúvida, o procedimento gráfico é o menos afetado pela presença de dados discrepantes.

2.2.4.1 - Uma regra de rejeição para dados discrepantes (ANSCOMBE e TUKEY, 1963)

Para ANSCOMBE e TUKEY (1963), em algumas situações, o bom senso nos leva a usar uma regra definida de rejeição para dados discrepantes. Quando se rejeita qualquer observação, ela é tratada como se fosse perdida, e aplicando-se a análise convencional às observações restantes, isto equivale a modificar o método dos mínimos quadrados, dando peso zero para as observações rejeitadas. Esse procedimento dá uma proteção contra erros grosseiros.

Por outro lado, mesmo quando todas as observações forem boas e as condições ideais forem plenamente satisfeitas, tal regra poderia casualmente levar a rejeição do dado discrepante, e a variância residual média dos parâmetros estimados, tal como efeitos de tratamentos, poderia então aumentar.

Pode-se, então, considerar a porcentagem de acréscimo na variância residual como um "custo" taxado pela regra de rejeição, que é uma espécie de garantia con-

tra observações ruins.

Uma possível regra de rejeição apresentada por ANSCOMBE e TUKEY (1963) seria: rejeitar a observação cujo valor absoluto do resíduo excede em magnitude a \underline{Cs} , onde \underline{s} é a raiz quadrada do quadrado médio do resíduo, e \underline{C} é uma constante dada. Se uma observação é rejeitada, recalculam-se todos os resíduos e os \underline{s} , e aplica-se a regra novamente, e assim sucessivamente, até que não existam mais rejeições.

Supondo-se:

n = nº total de resíduos,

v = nº de graus de liberdade do resíduo,

para calcular o valor de C correspondente a um dado "custo" (taxa de erro), procede-se como a seguir:

Multiplica-se o "custo" por v/n e encontra-se o desvio normal equivalente, N , tal que:

$$\text{"Custo"} = 100 \frac{n}{v} \phi(-N)\% ,$$

onde $\phi(\mu)$ se refere à distribuição acumulada $N(0,1)$.

Então, se $K = 1,40 + 0,85 N$, o valor desejado de \underline{C} é dado por:

$$C = K \left\{ 1 - \frac{K^2 - 2}{4v} \right\} \sqrt{\frac{v}{n}} .$$

Esta fórmula é uma aproximação de uma outra fórmula envolvendo uma função beta incompleta, dada em ANSCOMBE (1961).

Em muitos casos, um custo de 2,5% poderia ser considerado pequeno para que a observação fosse rejeitada. A correspondente regra de rejeição poderia dar uma considerável proteção se v não for também pequeno.

Se n é grande e y próximo de \bar{u} , um custo de 2,5% poderia levar a um N de cerca de 1,96, e então:

$$K = 1,40 + 0,85 \times 1,96 = 3,066$$

$$C = 3,066 \left\{ 1 - \frac{1,85}{v} \right\}$$

Desde que cada resíduo tem uma influência grande sobre s^2 quando y é pequeno, o valor de C para um custo fixado é menor para menores valores de n .

2.2.5 - Testes para não normalidade dos erros

Como se sabe, a hipótese de normalidade faz-se particularmente necessária a partir do momento em que são feitas as inferências estatísticas, como o teste F para diferença entre médias de tratamentos. Conforme SCHEFFÉ (1959), dentre outros, este teste é pouco afetado pela falta de normalidade, tanto em termos de nível de sig

nificância como de poder do teste. Portanto, o teste F é um teste robusto contra a ausência de normalidade, desde que a distribuição seja simétrica (SCHEFFÉ, 1959).

Para modelos de efeitos aleatórios, segundo NETER e WASSERMAN (1974), a falta de normalidade tem mais sérias complicações. Os estimadores dos componentes de variância são, porém, não tendenciosos, mas o coeficiente de confiança para estimativas por intervalo pode ser substancialmente diferente de um nível de significância estabelecido "a priori".

2.2.5.1 - Teste de LILLIEFORS (1967)

O teste de Lilliefors (1967), conforme CAMPOS (1983), é uma modificação do teste de Kolmogorov-Smirnov, introduzido em 1933, para verificar se uma série de dados pertence a uma determinada distribuição com média e variância conhecidas. Para testar a normalidade, Lilliefors (1967) introduziu uma modificação neste teste. Essa modificação é feita quando não se conhece a média e a variância da população, e estas são estimadas através dos dados da amostra.

Inicialmente calculam-se a média e a variância dos dados, em seguida

$$z_i = \frac{X_i - \bar{m}}{s}$$

os ordenam (z_i^*) .

Consideram-se:

$F(z_i^*)$ = proporção de valores esperados $\leq z_i^*$; valor obtido a partir da distribuição normal;

$S(z_i^*)$ = proporção de valores obtidos $\leq z_i^*$, isto é, $S(z_i^*) = k/n$, onde k é o número de valores obtidos a partir dos observados $\leq z_i^*$.

Define-se:

$$|F(z_i^*) - S(z_{i-1}^*)|]$$

onde $\text{Sup}_{z_i^*}$ = supremo em relação a z_i^* , isto é, a máxima distância vertical entre $F(z_i^*)$ e $S(z_i^*)$.

Para determinação de D , consideram-se em cada ponto z_i as diferenças $|F(z_i^*) - S(z_i^*)|$ e $|F(z_i^*) - S(z_{i-1}^*)|$, e toma-se apenas a maior delas.

O teste é bilateral e as hipóteses são:

H_0 : os dados podem ser estudados através da distribuição normal;

H_1 : os dados não podem ser estudados através da distribuição normal.

Rejeita-se a hipótese de nulidade, a um nível α de probabilidade, quando $D \geq d$. O valor de d é encontrado, na Tabela 6 de CAMPOS (1983), entre outros.

2.3 - Procedimentos Alternativos para Amenizar Transgressões no Modelo de Gauss-Markov

Quando, através da Análise de Resíduos, observam-se violações das pressuposições do modelo com estrutura de Gauss-Markov, $\varepsilon \sim N(\phi, I\sigma^2)$, diversas podem ser as alternativas para amenizar essas transgressões. A literatura recomenda, na maioria dos casos de violação das pressuposições, transformações de dados como possível solução. Um outro procedimento pode ser a modificação do modelo adotado, seja pela inclusão ou pela exclusão de uma variável, ou pela troca completa do modelo. Pode-se ainda adotar outros tipos de análises para os dados, como: quando as variâncias dos tratamentos não forem iguais, adotar-se o Método dos Mínimos Quadrados Ponderados; se existir correlação, o Método dos Mínimos Quadrados Generalizados; para existência de não normalidade, uma Análise Não-Paramétrica.

NETER e WASSERMAN (1974) chamam a atenção quanto ao uso da transformação de dados quando, ao se aplicar o teste de não-aditividade, se detectar a presença de interação no modelo. Segundo esses autores, a utilização de uma transformação, nesse caso, pode reduzir, às vezes, uma interação importante numa interação nula. Este fato

após a análise exploratória dos dados transformados, pode sugerir um modelo sem interação, quando na verdade a interação pode ser fundamental no contexto da interpretação dos resultados do experimento.

2.3.1 - Heterogeneidade de variâncias

2.3.1.1 - Transformações que estabilizam a variância

Quando a análise gráfica dos resíduos, ou um teste para heterogeneidade de variâncias, detecta a existência de heterocedasticidade, e ela for do tipo regular (classificação de Cochran, 1940, citado por DEMÉTRIO, 1978), e ainda, a relação entre média e variância dos tratamentos for conhecida, os dados podem ser transformados, de tal modo que passem a ter distribuição aproximadamente normal.

Uma fórmula geral para isto é dada pelo teorema:

Teorema (RAO, 1965): Seja $\{T_n\}$, $n=1,2,\dots$, uma sequência de estatísticas, tais que:

$$\sqrt{n} (T_n - \theta) \rightarrow X \sim N[0, \sigma^2(\theta)]$$

Seja g uma função de uma única variável e

admita-se que existe a derivada de primeira ordem g' , $g'(\theta) \neq 0$. Então:

$$\sqrt{n} [g(T_n) - g(\theta)] \rightarrow X \sim N\{0, [g'(\theta) \sigma(\theta)]^2\} .$$

Prova: Usando-se a expansão de Taylor (aproximação até 2º termo, numa vizinhança de θ , com as pressuposições sobre g' , tem-se:

$$g(T_n) - g(\theta) = (T_n - \theta) [g'(\theta) + \varepsilon_n] ,$$

onde $\varepsilon_n \rightarrow 0$ quando $T_n \rightarrow \theta$.

Isto implica que, para qualquer \underline{e} arbitrário,

$$|\varepsilon_n| < \phi \text{ e sempre que } |T_n - \theta| < \delta .$$

Então, $P(|\varepsilon_n| < e) \geq P(|T_n - \theta| < \delta) \rightarrow 1$, quando $n \rightarrow \infty$ por suposição. Desde que \underline{e} é arbitrário, $\varepsilon_n \xrightarrow{P} 0$. Logo,

$$\sqrt{n} [g(T_n) - g(\theta)] - \sqrt{n} (T_n - \theta) g'(\theta) = \sqrt{n} (T_n - \theta) \varepsilon_n \xrightarrow{P} 0 ,$$

desde que $\sqrt{n} (T_n - \theta)$ tem uma distribuição assintótica e $\varepsilon_n \xrightarrow{P} 0$. Então, a distribuição assintótica de $\sqrt{n} [g(T_n) - g(\theta)]$ é a mesma de $\sqrt{n} (T_n - \theta) g'(\theta)$, que é $N\{0, [g'(\theta) \sigma(\theta)]^2\}$ completando a prova.

Toma-se agora: $g'(\theta) \sigma(\theta) = C$ (constante, independente de θ), a variância assintótica de estatísti-

ca transformada $g(T_n)$ será independente de θ .

Então, $g(T_n)$ pode ser obtida da seguinte forma:

$$\frac{dg}{d\theta} = \frac{C}{\sigma(\theta)} .$$

Logo,

$$g = \int \frac{C}{\sigma(\theta)} d\theta .$$

Assim sendo, se em uma série de observações a variância amostral é função da média amostral, e esta função pode ser expressa por

$$S_x^2 = f(\bar{x}) ,$$

então, a função $T(X)$, dada por

$$T(X) = \int \frac{C}{\sqrt{f(\bar{x})}} d\bar{x} ,$$

tomando-se o valor de C (constante) adequado, define uma nova variável $Y = T(X)$, que possui variância constante.

DEMÉTRIO (1978), em seu trabalho, descreve certas transformações para estabilização da variância:

- a) arc sen \sqrt{x} : variáveis binomiais (porcentagem);
- b) raiz quadrada: variância proporcional à média - variável Poisson;

- c) logarítmica: variância proporcional ao quadrado da média;
- d) arco seno hiperbólico: variáveis binomiais negativas.

BARBOSA (1983) mostra um caso muito comum em análise de dados de contagens, em que a relação entre a média e a variância é dada pela expressão:

$$S_x^2 = a\bar{x}^b,$$

ou ainda:

$$\text{Log } S_x^2 = \text{Log } a + b \text{ Log } \bar{x}.$$

Aplicando-se a fórmula geral da transformação têm-se:

$$T(X) = \int \frac{c}{\sqrt{ax^b}} dx$$

$$T(X) = \frac{c}{\sqrt{a}} \int \frac{dx}{x^{b/2}}$$

$$T(X) = \begin{cases} \frac{c}{\sqrt{a} [1-(b/2)]} x^{[1-(b/2)]}, & \text{se } b \neq 2 \\ \frac{c}{\sqrt{a}} \log x, & \text{se } b = 2 \end{cases}$$

onde \underline{b} é obtido, plotando-se $\text{Log } S_x^2$ contra $\text{Log } \bar{x}$. As transformações apropriadas, tomando-se o valor de \underline{c} adequado, são:

$$Y = \text{Log } X \quad , \quad \text{se } b=2 \quad ,$$

$$Y = X^{[1-(b/2)]} \quad , \quad \text{se } b \neq 2 \quad .$$

Usando-se estes resultados, chega-se à seguinte tabela:

Tabela 2 - Valores do coeficiente de regressão linear simples, \underline{b} , com a transformação correspondente.

RELAÇÃO ENTRE S_x^2 e \bar{x}	VALORES DE \underline{b}	$1 - \frac{b}{2}$	TRANSFORMAÇÃO $Y = X^{[1-(b/2)]}$
$S_x^2 \propto c^{te}$	0	1	$Y = X$
$S_x^2 \propto \bar{x}$	1	1/2	$Y = \sqrt{X}$
$S_x^2 \propto \bar{x}^2$	2	0	$Y = \text{Log } X$
$S_x^2 \propto \bar{x}^3$	3	-1/2	$Y = 1/\sqrt{X}$
$S_x^2 \propto \bar{x}^4$	4	-1	$Y = 1/X$

FONTE: BARBOSA (1983).

Ainda BARBOSA (1983) diz que quando se analisam dados experimentais utilizando-se uma transformação da variável, normalmente se está interessado em apresen-

tar os resultados em termos da variável original. Cita diversos autores (Kempthorne, 1952; Snedecor, 1956; Steel e Torrie, 1960) que recomendam a aplicação da transformação inversa às médias da variável transformada.

Esses autores, segundo BARBOSA (1983), justificam este procedimento dizendo que a variável de interesse é aquela que tem as propriedades de normalidade, homogeneidade de variância, linearidade, etc. O interesse é apenas o de encontrar estimativas (de médias, de variâncias, etc.) em termos da variável original Y_{ij} , e para apresentar os resultados, estas estimativas (Y^*) são substituídas pelas correspondentes, aplicando-se a transformação inversa $X^* = T^{-1}(Y^*)$.

2.3.2 - Não independência dos erros

Quando os dados forem obtidos em sequência de tempo, e se verificar a não independência dos erros, uma alternativa sugerida na literatura é a inclusão desta variável no modelo. Uma outra alternativa sugerida quando a variável tempo não está presente, é a análise dos dados pelo método dos Mínimos Quadrados Generalizados, ou ainda uma transformação de dados apropriada. RAO (1965) indica a transformação arco tangente hiperbólica:

$$\tanh^{-1} r = \frac{1}{2} \log \left(\frac{1+r}{1-r} \right) ,$$

onde r é o coeficiente de correlação, para o caso de amostras correlacionadas.

2.3.3 - Presença de "outliers"

ANSCOMBE e TUKEY (1963) citam Jeffreys (1939, 1948), que diz que na presença de "outliers" é mais vantajoso usar um peso que decresça mais rapidamente com o aumento do valor absoluto do resíduo, do que usar qualquer procedimento de rejeição, cujo peso caia bruscamente de um para zero, para uma determinada magnitude dos resíduos. Os autores também citam que De Finetti (1961) apresenta a mesma solução.

Suponha-se que os desvios $y_{ij} - \mu_{ij}$ são identicamente distribuídos de acordo com uma distribuição com função de densidade $f(y_{ij} - \mu_{ij})$, e suponha-se que se queiram estimar os parâmetros no modelo linear μ_{ij} pelo método da máxima verossimilhança.

Se θ_k representa um parâmetro qualquer do modelo, a equação de máxima verossimilhança será:

$$L(\theta) = \prod_{i,j} f(y_{ij} - \mu_{ij}) ,$$

$$\text{Log } L(\theta) = \sum_i \sum_j \text{Log } f(y_{ij} - \mu_{ij})$$

$$\frac{\partial L(\theta)}{\partial \theta_k} = 0 \implies \sum_i \sum_j \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \frac{\partial \mu_{ij}}{\partial \theta_k} = 0$$

Agora, se a distribuição dos desvios é normal, então:

$$f(y_{ij} - \mu_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{y_{ij} - \mu_{ij}}{\sigma} \right)^2 \right\}$$

pois, $y \sim N(0, \sigma^2)$, então:

$$f'(y_{ij} - \mu_{ij}) = \frac{1}{\sigma\sqrt{2\pi}} \exp \left\{ -\frac{1}{2} \left(\frac{y_{ij} - \mu_{ij}}{\sigma} \right)^2 \right\} \cdot \left[\frac{-2(y_{ij} - \mu_{ij})}{2\sigma^2} \right]$$

ou seja:

$$\frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} = - \frac{(y_{ij} - \mu_{ij})}{\sigma^2}$$

E, portanto, a equação de máxima verossimi

lhança será:

$$\sum_{i,j} \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \frac{\partial \mu_{ij}}{\partial \theta_k} = 0$$

$$\implies \sum_{i,j} \frac{(y_{ij} - \mu_{ij})}{\sigma^2} \frac{\partial \mu_{ij}}{\partial \theta_k} = 0$$

$$\implies \sum_{i,j} (y_{ij} - \mu_{ij}) \frac{\partial \mu_{ij}}{\partial \theta_k} = 0 ,$$

que coincide com as equações de mínimos quadrados ordinários.

Se a distribuição não é normal, consideram-se as equações de máxima verossimilhança como um procedimento de mínimos quadrados modificados. Neste caso, o método da máxima verossimilhança é equivalente ao método de mínimos quadrados ordinários, aplicados não às observações originais, mas aos valores transformados:

$$- \frac{f'(y_{ij} - \hat{y}_{ij})}{f(y_{ij} - \hat{y}_{ij})} ,$$

onde \hat{y}_{ij} são os valores ajustados, estimando os parâmetros de μ iterativamente. (Qualquer outro parâmetro desconhecido em f , também pode ser estimado iterativamente). De fato, aplicando-se o método dos mínimos quadrados à variável

$$- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} ,$$

tem-se:

$$y_{ij} = \mu_{ij} + \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right]$$

$$y_{ij} - \mu_{ij} = \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right]$$

$$\sum_{i,j} (y_{ij} - \mu_{ij})^2 = \sum_{i,j} \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right]^2 = Q$$

$$\frac{\partial Q}{\partial \theta_k} = 2 \sum_{i,j} \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right] \frac{\partial \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right]}{\partial \theta_k}$$

Mas,

$$\frac{\partial \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right]}{\partial \theta_k} = \frac{\partial (y_{ij} - \mu_{ij})}{\partial \theta_k} = - \frac{\partial \mu_{ij}}{\partial \theta_k} .$$

Então:

$$\frac{\partial Q}{\partial \theta_k} = 2 \sum_{i,j} \left[- \frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right] \left(- \frac{\partial \mu_{ij}}{\partial \theta_k} \right) .$$

Fazendo-se $\frac{\partial Q}{\partial \theta_k} = 0$, tem-se:

$$\sum_{i,j} \left[\frac{f'(y_{ij} - \mu_{ij})}{f(y_{ij} - \mu_{ij})} \right] \frac{\partial \mu_{ij}}{\partial \theta_k} = 0 ,$$

que corresponde às equações de máxima verossimilhança.

O processo iterativo consiste em se tomar valores iniciais para os parâmetros a estimar, calcular-se os valores estimados (\hat{Y}_{ij}) e, então, obter os valores da variável transformada. Pelo método dos mínimos quadrados, obtêm-se novas estimativas dos parâmetros, e repete-se o processo, até convergir.

ANSCOMBE e TUKEY (1963) mostram como se comporta a função $-f'(z)/f(z)$ se a distribuição dos desvios for uma t de Student:

$$f(z) = \frac{\Gamma\left[\frac{k+1}{2}\right]}{\Gamma\left[\frac{k}{2}\right]} \cdot \frac{1}{\sqrt{k\pi}} \cdot \frac{1}{\left[1 + \frac{z^2}{k}\right]^{[(k+1)/2]}}$$

e

$$f'(z) = \frac{\Gamma\left[\frac{k+1}{2}\right]}{\Gamma\left[\frac{k}{2}\right]} \cdot \frac{1}{\sqrt{k\pi}} \cdot \frac{-\frac{k+1}{2} \left[1 + \frac{z^2}{k}\right]^{[(k+1)/2]-1} \cdot \frac{2z}{k}}{\left[1 + \frac{z^2}{k}\right]^{k+1}}$$

logo:

$$-\frac{f'(z)}{f(z)} = \frac{\frac{k+1}{k} \left[1 + \frac{z^2}{k} \right]^{[(k+1)/2]-1} \cdot z}{\left[1 + \frac{z^2}{k} \right]^{[(k+1)/2]}}$$

$$-\frac{f'(z)}{f(z)} = \frac{\frac{k+1}{k} \cdot z}{1 + \frac{z^2}{k}}$$

$$-\frac{f'(z)}{f(z)} = \frac{(k+1) \cdot z}{k+z^2}$$

Pode-se notar que, para pequenos valores de z , $-f'(z)/f(z)$ é quase proporcional a z , mas quando z cresce em magnitude, $-f'(z)/f(z)$ cresce até atingir um pico e, então, diminui até chegar a zero.

2.3.4 - Ausência de normalidade

NETER E WASSERMAN (1974) e SCHEFFÉ (1959), dentre outros, mostram que a análise de variância pode ser robusta em ausência de normalidade dos erros, mas para a aplicação do teste sobre os resíduos a existência de

normalidade faz-se necessária. A prática tem mostrado ser mais útil verificar a normalidade dos erros como um último procedimento da análise de resíduos, uma vez que a heterogeneidade de variâncias ou um dado discrepante pode conduzir à não normalidade dos mesmos.

Uma alternativa, quando se detecta a ausência de normalidade, pode ser a utilização de uma transformação de dados, como, por exemplo, as citadas em 2.3.1. Outra alternativa pode ser a análise dos dados através de testes não-paramétricos.

3. MATERIAL E MÉTODOS

3.1 - Material

Face aos objetivos deste estudo, julgou-se desejável apresentar a discussão de um exemplo numérico. Além disso, entendeu-se que a melhor caracterização seria obtida com dados realmente coletados experimentalmente. O material aqui utilizado é referente aos dados de 25 clones de seringueira (*Hevea* spp.), listados na Tabela 3, que fazem parte de ensaio de competição de clones de seringueira do Centro Nacional de Pesquisa da Seringueira e Dendê - EMBRAPA, Manaus, AM.

A coleta dos dados foi feita no ano agrícola de 1980, após um ano de plantio, e é referente aos se-

guintes caracteres:

- a) Produção de borracha seca, através da média de três mini-testes de produção, em mg;
- b) Caracteres relacionados com o vigor da planta:
 - b.1) altura da planta (AP), em m;
 - b.2) circunferência do caule (CIRC), em cm;
 - b.3) número de lançamentos (N.L.);
 - b.4) tamanho médio de lançamentos (T.M.L.), em m;
 - b.5) espessura de casca (E.C.), em mm;
- c) Caracteres relacionados com o sistema laticífero:
 - c.1) diâmetro dos vasos laticíferos (D.V.L.), em μ ;
 - c.2) densidade dos vasos laticíferos (DENS), em 5 mm de anéis;
 - c.3) número de anéis de vasos laticíferos (N.A.).

Dos dados apresentados, usou-se apenas a produção de borracha seca, através da média de três mini-testes de produção, em mg, cujos dados estão na Tabela 4.

Tabela 3 - Clones estudados e seus respectivos parentais e espécies.

CLONES	PARENTAIS	ESPÉCIES
IAN 6121	Fx 4073 (F4542 x PB86) x PB86	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
IAN 2925	Fx 516 (F4542 x AV363) x PB86	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
Fx 3864	PB86 x FB38	<i>H. brasiliensis</i> x <i>H. brasiliensis</i>
AC 68	Seleção primária	<i>H. brasiliensis</i> ?
IAN 3384	Fx 652 (F4542 x Tjiri) x PB186	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
PFB 26	Seleção primária	<i>H. brasiliensis</i> ?
AC 53	Seleção primária	<i>H. brasiliensis</i> ?
IAN 6717	Fx 43-655 (F4542 x Tjiri) x PB86	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
IAN 6158	Fx 43-655 (F4542 x Tjiri) x PB186	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
Fx 3925	F4542 x AV363	<i>H. benthamiana</i> x <i>H. brasiliensis</i>
PFB 4	Seleção primária	<i>H. brasiliensis</i> ?
AC 55	Seleção primária	<i>H. brasiliensis</i> ?
RO 54	Seleção primária	<i>H. brasiliensis</i> ?
PFB 1	Seleção primária	<i>H. brasiliensis</i> ?
IAN 2829	Fx 516 (F4512 x AV363) x PB86	<i>H. brasiliensis</i> ?
RO 60	Seleção primária	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
Fx 4037	F4542 x PB86	<i>H. brasiliensis</i> ?
IAN 717	PB86 x F4542	<i>H. brasiliensis</i> x <i>H. brasiliensis</i>
IAN 2945	Fx 516 (F4542 x AV363) x PB86	(<i>H. brasiliensis</i> x <i>H. benthamiana</i>)
Fx 25	F361 x AV49	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>
IAC 206	Poliplóide do IAN 873	<i>H. brasiliensis</i> x <i>H. brasiliensis</i>
IAC 207	Poliplóide do IAN 873	<i>H. brasiliensis</i> x <i>H. brasiliensis</i>
Alter do chão	Seleção primária	<i>H. brasiliensis</i> x <i>H. brasiliensis</i>
IAC 222	Poliplóide do IAN 873	<i>H. brasiliensis</i> x <i>H. brasiliensis</i>
IAN 4488	Fx 4421 (F4537 x PB86) x Tjiri	(<i>H. benthamiana</i> x <i>H. brasiliensis</i>) x <i>H. brasiliensis</i>

Tabela 4 - Dados de produção de borracha seca (P.B.S.).

CLONES	REPETIÇÕES	P.B.S.
IAN 6121	I	50,14
	II	39,76
	III	28,49
IAN 2925	I	56,58
	II	54,07
	III	56,62
Fx 3864	I	38,44
	II	34,01
	III	29,17
AC 68	I	20,99
	II	8,18
	III	7,47
IAN 3384	I	23,94
	II	20,32
	III	27,36
PFB 26	I	4,73
	II	10,74
	III	17,91
AC 53	I	30,64
	II	30,64
	III	24,11
IAN 6717	I	31,08
	II	26,14
	III	58,83
IAN 6158	I	32,89
	II	51,10
	III	30,68
Fx 3925	I	25,07
	II	12,86
	III	18,34

Tabela 4 - continuação.

CLONES	REPETIÇÕES	P.B.S.
PFB 4	I	36,45
	II	25,64
	III	34,99
AC 55	I	28,36
	II	19,90
	III	36,78
RO 54	I	10,04
	II	9,28
	III	21,60
PFB 1	I	33,42
	II	17,83
	III	16,11
IAN 2829	I	21,72
	II	18,39
	III	80,27
RO 60	I	17,40
	II	13,04
	III	11,71
Fx 4037	I	26,21
	II	31,28
	III	25,48
IAN 717	I	24,69
	II	37,88
	III	19,31
IAN 2945	I	42,62
	II	47,81
	III	43,95
Fx 25	I	55,53
	II	33,76
	III	32,30

Tabela 4 - continuação.

CLONES	REPETIÇÕES	P.B.S.
IAC 206	I	51,35
	II	65,96
	III	43,58
IAC 207	I	99,40
	II	90,55
	III	64,68
Alter do Chão	I	22,57
	II	20,57
	III	34,79
IAC 222	I	65,94
	II	48,31
	III	55,37
IAN 4488	I	30,35
	II	37,25
	III	43,78

O ensaio seguiu o esquema de lâ-tice 5x5, com 3 repetições, com seis plantas ú-teis por parcela de 126 m², no espaçamento 7m x 3m, e as análises da variância, para todos os caracteres, foram realizadas com médias de parcelas, segundo o esquema de lâ-tice, conforme o modelo matemático:

$$Y_{ijk} = m + t_i + b_{jk} + r_k + \varepsilon_{ijk} ,$$

onde:

Y_{ijk} é o valor da ijk -ésima observação;

m é a média geral;

t_i é o efeito do i -ésimo tratamento;

b_{jk} é o efeito do j -ésimo bloco da k -ésima repetição;

r_k é o efeito da k -ésima repetição;

$\varepsilon_{ijk} \sim N(0, \sigma^2)$ é o erro experimental associado à ijk -ésima observação.

3.2 - Métodos

3.2.1 - Obtenção dos resíduos

Para se estudar os resíduos, faz-se necessária a sua obtenção, como o primeiro passo para a realização das análises.

A obtenção dos resíduos será feita segundo NETER e WASSERMAN (1974) e BOX *et alii* (1978), através da decomposição de cada observação.

3.2.1.1 - Delineamento Inteiramente ao Acaso

De modo geral, o modelo linear para este delineamento é da forma:

$$Y_{ij} = m + t_i + \varepsilon_{ij} ,$$

onde: Y_{ij} = valor da ij -ésima observação;

m = média geral;

t_i = efeito do tratamento;

ε_{ij} = erro experimental associado à ij -ésima repetição, $\varepsilon_{ij} \sim N(0, \sigma^2)$;

então, cada observação pode ser escrita da seguinte forma:

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.}),$$

onde:

$\bar{Y}_{..}$ é a média amostral;

$(\bar{Y}_{i.} - \bar{Y}_{..})$ são os desvios entre tratamentos;

$(Y_{ij} - \bar{Y}_{i.})$ são os desvios dentro de tratamentos (resíduo).

Como o resíduo é a diferença entre o valor observado e o valor ajustado, e desde que:

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}),$$

$$Y_{ij} = \bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..}$$

$$Y_{ij} = \bar{Y}_{i.},$$

então,

$$e_{ij} = Y_{ij} - \hat{Y}_{ij}$$

fica:

$$e_{ij} = Y_{ij} - \bar{Y}_i .$$

Para a padronização dos resíduos:

1º) obtêm-se o quadrado médio do resíduo:

$$QM \text{ Res.} = \left[\sum_{i,j} (e_{ij})^2 \right] / g.l. \text{ Res.} ,$$

onde g.l. Res. = graus de liberdade do resíduo;

2º) divide-se o resíduo e_{ij} pela raiz quadrada do quadrado médio do resíduo, obtendo-se a nova variável d_{ij} :

$$d_{ij} = \frac{e_{ij}}{\sqrt{QM \text{ Res.}}} ,$$

que é chamado de "resíduo padronizado".

3.2.1.2 - Delineamento em Blocos Casualizados

Em geral, o modelo linear para este delineamento segue a forma:

$$Y_{ij} = m + t_i + b_j + \varepsilon_{ij} ,$$

onde:

Y_{ij} = valor da ij -ésima observação;

m = média geral;

t_i = efeito do i -ésimo tratamento;

b_j = efeito do j -ésimo bloco;

ε_{ij} = erro experimental, associado à ij -ésima observação, $\varepsilon_{ij} \sim N(0, \sigma^2)$;

e cada observação pode ser escrita como:

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..}) + (Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})$$

onde:

$\bar{Y}_{..}$ é a média geral;

$(\bar{Y}_{i.} - \bar{Y}_{..})$ são os desvios entre tratamentos;

$(\bar{Y}_{.j} - \bar{Y}_{..})$ são os desvios entre blocos;

$(Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..})$ são os resíduos.

Como o resíduo é a diferença entre o valor observado e o valor ajustado, e desde que:

$$Y_{ij} = \bar{Y}_{..} + (\bar{Y}_{i.} - \bar{Y}_{..}) + (\bar{Y}_{.j} - \bar{Y}_{..})$$

$$Y_{ij} = \bar{Y}_{..} + \bar{Y}_{i.} - \bar{Y}_{..} + \bar{Y}_{.j} - \bar{Y}_{..}$$

$$Y_{ij} = \bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}$$

então:

$$e_{ij} = Y_{ij} - \bar{Y}_{ij}$$

$$e_{ij} = Y_{ij} - [\bar{Y}_{i.} + \bar{Y}_{.j} - \bar{Y}_{..}]$$

$$e_{ij} = Y_{ij} - \bar{Y}_{i.} - \bar{Y}_{.j} + \bar{Y}_{..} \quad ,$$

e para obtenção do resíduo padronizado o processo é o mesmo descrito em 3.2.1.1.

3.2.1.3 - Delineamento em Quadrado Latino

O modelo linear para este delineamento é da forma:

$$Y_{ij(k)} = m + \lambda_i + c_j + t_k + \varepsilon_{ij(k)} \quad ,$$

onde:

$Y_{ij(k)}$ = ijk -ésima observação;

m = média geral;

λ_i = efeito da i -ésima linha;

c_j = efeito da j -ésima coluna;

t_k = efeito do k -ésimo tratamento;

$\varepsilon_{ij(k)}$ = erro experimental, associado à ijk -ésima observação, $\varepsilon_{ij(k)} \sim N(0, \sigma^2)$;

e cada observação pôde ser escrita como segue:

$$Y_{ij(k)} = \bar{Y}... + (\bar{Y}_{i..} - \bar{Y}...) + (\bar{Y}_{.j.} - \bar{Y}...) + (\bar{Y}_{..k} - \bar{Y}...) + \\ + (Y_{ij(k)} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}...)$$

onde:

$\bar{Y}...$ é a média amostral;

$(\bar{Y}_{i..} - \bar{Y}...)$ são os desvios entre linhas;

$(\bar{Y}_{.j.} - \bar{Y}...)$ são os desvios entre colunas;

$(\bar{Y}_{..k} - \bar{Y}...)$ são os desvios entre tratamentos;

$(Y_{ij(k)} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}...)$ são os resíduos.

Como o resíduo é a diferença entre o valor observado e o valor ajustado, e desde que:

$$\bar{Y}_{ij(k)} = \bar{Y}... + (\bar{Y}_{i..} - \bar{Y}...) + (\bar{Y}_{.j.} - \bar{Y}...) + (\bar{Y}_{..k} - \bar{Y}...)$$

$$\bar{Y}_{ij(k)} = \bar{Y}... + \bar{Y}_{i..} - \bar{Y}... + \bar{Y}_{.j.} - \bar{Y}... + \bar{Y}_{..k} - \bar{Y}...$$

$$\bar{Y}_{ij(k)} = \bar{Y}_{i..} + \bar{Y}_{.j.} + \bar{Y}_{..k} - 2\bar{Y}...$$

então,

$$e_{ij(k)} = Y_{ij(k)} - \bar{Y}_{ij(k)}$$

fica:

$$e_{ij(k)} = Y_{ij(k)} - \bar{Y}_{i..} - \bar{Y}_{.j.} - \bar{Y}_{..k} + 2\bar{Y}...$$

Para a obtenção do resíduo padronizado, o processo é o mesmo descrito em 3.2.1.1.

3.2.1.4 - Esquema em Látice

O modelo linear é:

$$Y_{ijk} = m + t_i + b_{jk} + r_k + \varepsilon_{ijk},$$

onde:

Y_{ijk} = valor da ijk -ésima observação;

m = média geral;

t_i = efeito do i -ésimo tratamento;

b_{jk} = efeito do j -ésimo bloco, na k -ésima repetição;

r_k = efeito da k -ésima repetição;

ε_{ijk} = erro experimental associado à ijk -ésima observação, $\varepsilon_{ijk} \sim N(0, \sigma^2)$;

e cada observação pode ser escrita como:

$$Y_{ijk} = \bar{Y}... + (\bar{Y}_{i..} - \bar{Y}...) + (\bar{Y}_{.jk} - \bar{Y}...) + (\bar{Y}_{..k} - \bar{Y}...) + \\ + (Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.jk} + \bar{Y}...)$$

onde:

$\bar{Y}...$ é a média amostral;

$(\bar{Y}_{i..} - \bar{Y}...)$ são os desvios entre tratamentos;

$(\bar{Y}_{.jk} - \bar{Y}_{..k})$ são os desvios entre blocos da k -ésima repetição;

$(\bar{Y}_{..k} - \bar{Y}_{...})$ são os desvios entre repetições;

$(Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.jk} + \bar{Y}_{...})$ são os resíduos.

Como o resíduo \bar{e} é a diferença entre o valor observado e o valor ajustado, e desde que:

$$\bar{Y}_{ijk} = \bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.jk} - \bar{Y}_{..k}) + (\bar{Y}_{..k} - \bar{Y}_{...})$$

Então, $e_{ijk} = Y_{ijk} - \bar{Y}_{ijk}$, fica:

$$e_{ijk} = Y_{ijk} - [\bar{Y}_{...} + (\bar{Y}_{i..} - \bar{Y}_{...}) + (\bar{Y}_{.jk} - \bar{Y}_{..k}) + (\bar{Y}_{..k} - \bar{Y}_{...})]$$

$$e_{ijk} = Y_{ijk} - \bar{Y}_{i..} - \bar{Y}_{.jk} + \bar{Y}_{...}$$

e para a obtenção do resíduo padronizado o processo é o mesmo descrito em 3.2.1.1.

3.2.2 - Análise gráfica dos resíduos

3.2.2.1 - Obtenção do gráfico dos resíduos contra a variável em estudo

Para a execução da análise gráfica dos resíduos, usar-se-á, neste trabalho, um programa em linguagem BASIC, para utilização em microcomputador PET/CBC 3032, desenvolvido por MATTIOLI (1983), chamado de PLOTRES.

Esse programa executa a análise gráfica dos resíduos, e para obtenção destes fazem-se necessários os valores de Y_{ij} estimados e dos seus respectivos resíduos padronizados, d_{ij} , não importando sua ordenação.

O programa executa a ordenação dos dados, fornece a maior e a menor ordenada, que corresponde ao valor de Y_{ij} estimado, e a maior e a menor abscissa, que corresponde aos resíduos padronizados para escolha da escala.

Para impressão do gráfico dos resíduos, basta, portanto, digitar o maior e o menor valor do resíduo, bem como o maior e o menor valor de Y_{ij} estimado.

3.2.2.2 - Obtenção do gráfico no "papel normal de probabilidade"

Para a construção deste gráfico, far-se-á primeiramente a ordenação dos resíduos normalizados, e em

seguida proceder-se-á à obtenção da probabilidade acumulada correspondente a cada resíduo d_{ij} , através da fórmula:

$$P_i = (i - 0,5)/n ,$$

onde i é a ordem do resíduo, e n é o numero total dos resíduos.

Esses valores de P_i são calculados na subrotina ORDE, depois de feita a ordenação dos resíduos.

No papel normal de probabilidade, o valor do resíduo ordenado corresponderá à abscissa dos eixos, e o valor de $P_i \cdot 100$ corresponderá à ordenada dos eixos, conforme a escala própria do papel.

3.2.3 - Inferências sobre os resíduos

Para a verificação de inferências sobre os resíduos, elaboraram-se dois programas em linguagem FORTRAN, para uso em um computador IBM 1130.

O programa ROSAN em linguagem FORTRAN, elaborado para a verificação de inferências sobre os resíduos, aceita para execução todos os delineamentos citados em 3.2.1, inclusive o esquema de Látice, fornecendo:

- a decomposição dos resíduos;
- o quadrado médio do resíduo;

- os resíduos padronizados;
- teste F para não-aditividade;
- teste de Cochran para verificação de heterocedasticidade de variâncias;
- teste de Lilliefors para verificação de normalidade;
- as probabilidades para a construção do papel normal de probabilidade; e
- transformação de dados.

O Programa ROSA1, também em linguagem FORTRAN, tem aplicação específica no delineamento em Blocos ao Acaso, para modificação de dados discrepantes.

3.2.3.1 - Teste F para não-aditividade

Para a verificação da aditividade do modelo linear adotado, usa-se o teste F para não-aditividade, na forma de BOX *et alii* (1978).

O procedimento para o teste é o seguinte.

A partir da obtenção de \hat{Y}_{ij} estimado, calcula-se uma nova variável, na forma seguinte:

$$q_{ij} = \hat{Y}_{ij}^2$$

onde \hat{Y}_{ij} é proveniente do delineamento em estudo.

Como um segundo passo do processo, obtem-se o valor estimado desta nova variável (q_{ij}), através do processo descrito em 2.2.1.

Uma vez obtidos q_{ij} e \hat{q}_{ij} , respectivamente, o passo seguinte é o cálculo da variável:

$$P = \sum_{i,j} (Y_{ij} - \bar{Y}_{ij})(q_{ij} - \hat{q}_{ij}) ,$$

onde $(Y_{ij} - \bar{Y}_{ij})$ é o resíduo e_{ij} da variável original, e $(q_{ij} - \hat{q}_{ij})$ é o resíduo da nova variável, obtido a partir de \bar{Y}_{ij} da variável original. Em seguida, procede-se ao cálculo da variável

$$Q = \sum_{i,j} (q_{ij} - \hat{q}_{ij})^2 ,$$

que nada mais é que uma soma de quadrados de resíduo.

Então a soma de quadrados para não-aditividade tem a seguinte forma:

$$SQ \bar{N}A = \frac{P^2}{Q}$$

ou

$$SQ \bar{N}A = \frac{\left[\sum_{i,j} (Y_{ij} - \bar{Y}_{ij})(q_{ij} - \hat{q}_{ij}) \right]^2}{\sum_{i,j} (q_{ij} - \hat{q}_{ij})^2}$$

e

$$SQ \text{ Novo Resíduo} = SQ \text{ Res.} - SQ \bar{N}A ,$$

onde SQ Res. é a soma de quadrados de resíduo da variável original (Y_{ij}).

Então:

$$QM \text{ Novo Resíduo} = \frac{SQ \text{ Novo Resíduo}}{g.l. \text{ Res.} - 1}$$

onde g.l. Res. = graus de liberdade do resíduo do delineamento em estudo.

Finalmente chega-se ao teste F para não-aditividade:

$$F_{\bar{N}A} = \frac{SQ \bar{N}A}{QM \text{ Novo Resíduo}}$$

O esquema do quadro de ANOVA para o teste de não-aditividade se o ensaio fosse em Blocos Casualizados é o seguinte:

C. de Variação	G.L.	S.Q.	Q.M.	F
Não-aditividade	1	SQ $\bar{N}A$	SQ $\bar{N}A$	$\frac{QM \bar{N}A}{QM \text{ Novo Res.}}$
Novo Resíduo	$(I-1)(J-1)-1$	$SQ \text{ Res.} - SQ \bar{N}A$	$\frac{SQ \text{ Res.} - SQ \bar{N}A}{(I-1)(J-1)-1}$	
RESÍDUO	$(I-1)(J-1)$	SQ Res.		

No programa ROSAN este teste é processado através da subrotina TESND. Para a execução desta subrotina se faz necessário o dimensionamento do delineamento que se está utilizando, a indicação de qual o delineamento usado, os valores dos \bar{Y} estimados, os valores dos resíduos, a soma de quadrados do resíduo da variável original, o número de graus de liberdade do resíduo e, ainda, o valor de F tabelado.

3.2.3.2 - Teste de Cochran

Para a verificação da existência de heterogeneidade de variância, quando o delineamento for Inteiramente ao Acaso, usar-se-á o teste de Cochran.

A estatística do teste é:

$$C_{\text{cal.}} = \frac{S_i^2 \text{ máx.}}{\sum_i S_i^2} .$$

A obtenção de S_i^2 é feita através dos resíduos e_{ij} , como segue:

$$S_i^2 = \sum_j \frac{e_{ij}^2}{J-1} ,$$

onde J é o número de repetições, e e_{ij} é obtido como em

3.2.1.1, ou seja:

$$e_{ij} = Y_{ij} - \bar{Y}_{i.}$$

A hipótese de nulidade, $H_0: \sigma_1^2 = \dots = \sigma_I^2$ será rejeitada se $C_{cal.} > C_{(I,J-1)}$, com $C_{(I,J-1)}$ retirado da tabela apropriada.

Este teste é feito através da subrotina TEVON, e para seu processamento faz-se necessário o dimensionamento do delineamento, a indicação de qual o delineamento utilizado, os resíduos, e o valor de tabela para o teste de Cochran.

Com respeito aos outros delineamentos, apenas será feita a análise gráfica dos resíduos, justificando-se este procedimento pelas restrições apresentadas em 2.2.2.

3.2.3.3 - Modificação de "outliers"

A modificação de "outliers" será utilizada somente depois que, feita a análise gráfica dos resíduos, for detectado algum resíduo padronizado maior que os valores compreendidos entre (-3; +3).

Para a aplicação da modificação de "outliers", deve-se, primeiramente, estudar qual a distribuição dos resíduos. Neste estudo somente será avaliado o caso em que os resíduos possam ser estudados através de uma distribuição t de Student.

Como primeiro passo, deve-se aplicar aos resíduos um teste de aderência.

Uma vez verificada a aderência dos resíduos a uma distribuição t de Student, com k graus de liberdade, o passo seguinte consiste em se obter os valores estimados das observações Y_{ij} , os resíduos, e os resíduos normalizados. Todos esses cálculos são obtidos no programa ROSAN, já citado.

A obtenção do resíduo que agora passara a ser chamado de resíduo transformado, será da seguinte forma:

$$d_{ij}(\text{transf.}) = - \frac{f'(d_{ij})}{f(d_{ij})}$$

sendo:

$$- \frac{f'(d_{ij})}{f(d_{ij})} = \frac{[(k+1) d_{ij}]}{[k + d_{ij}^2]},$$

onde k é o grau de liberdade da distribuição t de Student, e d_{ij} o resíduo normalizado.

Calculado o resíduo transformado, o valor de Y_{ij} será:

$$Y_{ij(\text{transf.})} = Y_{ij} + d_{ij(\text{transf.})} \cdot \sqrt{QM \text{ Res.}}$$

O "outlier" será aqui considerado modificação quando se obtiver o menor quadrado médio do resíduo.

O programa ROSA1 executa este procedimento apenas para o delineamento em Blocos ao Acaso. Para o funcionamento deste programa fazem-se necessários os valores das observações Y_{ij} , o grau de liberdade da distribuição t de Student e um valor de convergência. Este número que é chamado de valor de convergência serve como teste de diferença entre dois valores de Y_{ij} transformados.

3.2.3.4 - Teste de Lilliefors

O teste de Lilliefors (1967), conforme CAMPOS (1983), é uma modificação do teste de Kolmogorov-Smirnov, introduzido em 1933, para verificar se uma série de dados pertence a uma determinada distribuição com média e variância conhecida. Para a aplicação deste teste, inicialmente calculam-se a média e a variância dos dados. Neste caso, como os dados em questão são os resíduos das observações, a sua média será zero e a variância será o quadrado médio do resíduo.

Então a variável

$$Z_{ij} = \frac{e_{ij}}{\sqrt{QM \text{ Res.}}}$$

será, como pode ser observado, o resíduo normalizado d_{ij} obtido em 3.2.1.

Como passo seguinte para este teste, ordena-se esta variável Z_{ij} , que será agora chamada d_{ij} (uma vez que é o resíduo normalizado) e consideram-se:

$F(d_{ij})$ = proporção de valores esperados $\leq d_{ij}$, valor obtido a partir da distribuição normal;

$S(d_{ij})$ = proporção de valores obtidos $\leq d_{ij}$;

isto é, $S(d_{ij}) = k/n$, onde k é o número de valores obtidos a partir dos observados $\leq d_{ij}$, e n é o número de observações da amostra.

$$D = \text{Sup}_{d_{ij}} [|F(d_{ij}) - S(d_{ij})| , |F(d_{ij}) - S(d_{ij-1})|]$$

onde $\text{Sup}_{d_{ij}}$ = supremo em relação a d_{ij} , isto é, a máxima distância entre $F(d_{ij})$ e $S(d_{ij})$. Para sua determinação, consideram-se, em cada ponto d_{ij} , as diferenças

$$|F(d_{ij}) - S(d_{ij})| \text{ e } |F(d_{ij}) - S(d_{ij-1})|$$

e toma-se a maior delas.

Para a aplicação do teste, usa-se o maior valor de D , obtido através das diferenças entre $|F(d_{ij}) - S(d_{ij})|$.

Este teste é executado através das subrotinas ORDE e LILLI. A subrotina ORDE tem a única função de ordenar os resíduos normalizados dentro de um vetor. Processada a ordenação dos resíduos, é "chamada" a subrotina LILLI, que executa o teste de Lilliefors.

Para a sua execução faz-se necessário o dimensionamento do vetor dos resíduos, indicação de qual o delineamento que se está estudando, o vetor dos resíduos, e o valor de tabela para o teste de Lilliefors.

Para a obtenção dos valores de $F(d_{ij})$ que são valores obtidos a partir de uma distribuição normal, a subrotina LILLI gera-os através da função:

$$F = F2*(A1*T+A2*(T**2)+A3*(T**3)) , \quad 2/$$

se d_{ij} for menor que zero, e

$$F1 = 1-F ,$$

se d_{ij} for maior ou igual a zero, onde:

$$A1 = 0,4361836$$

$$A2 = -0,1201676$$

$$A3 = 0,9372980$$

$$T = 1/(1+0,3326*X)$$

$$F2 = 0,398942 \text{ EXP}(-0,5*X**2)$$

e X e o valor do resíduo ordenado.

2/ A função F foi retirada do programa em linguagem BASIC para distribuição normal, do livro de POOLE, M. e M. BORCHERS, 1979 - Some Common Basic Programs, 3ª ed., McGraw-Hill, Berkeley, Califórnia, 92 p.

3.2.3.5 - Teste de Kolmogorov-Smirnov

CAMPOS (1983) mostra um teste introduzido por Kolmogorov-Smirnov (1933) para verificação da adaptação de uma específica e bem conhecida distribuição $F(X)$, a dados provenientes de uma distribuição desconhecida $F_0(X)$.

A hipótese a ser verificada é:

$$H_0: F \equiv F_0 \quad ,$$

$$H_a: F \neq F_0$$

Para a aplicação do teste, inicialmente ordena-se o vetor dos resíduos d_{ij} . Como passo seguinte considera-se:

$F(d_{ij})$ = proporção de valores esperados $\leq d_{ij}$, valor obtido a partir da distribuição em estudo;

$S(d_{ij})$ = proporção de valores obtidos $\leq d_{ij}$;

isto é, $S(d_{ij}) = k/n$, onde k é o número de valores obtidos a partir dos observados $\leq d_{ij}$, e n é o número total de observações.

Seja:

$$D = \text{Sup}_{d_{ij}} [|F(d_{ij}) - S(d_{ij})| , |F(d_{ij}) - S(d_{ij-1})|]$$

onde $\text{Sup}_{d_{ij}}$ = supremo em relação a d_{ij} , isto é, a máxima distância entre $F(d_{ij})$ e $S(d_{ij})$. Para sua determinação consideram-se, em cada ponto d_{ij} , as diferenças $|F(d_{ij}) - S(d_{ij})|$

e $|F(d_{ij}) - S(d_{ij-1})|$ e toma-se a maior delas.

A hipótese de nulidade será rejeitada se $D \geq d$, onde d pode ser encontrado, por exemplo, na Tabela 5 de CAMPOS (1983).

Este teste será aplicado aqui, quando da observação de um dado discrepante for aplicada a "modificação de dados discrepantes", na possibilidade de se estudarem os resíduos através de uma distribuição t de Student.

3.2.4 - Transformação de dados

Como uma das propostas alternativas para amenizar violações às pressuposições no modelo adotado, como, por exemplo, não constância de variância, não normalidade dos erros, ou mesmo quando o modelo não é aditivo, uma tentativa de solução do problema é a transformação dos dados.

As transformações estudadas neste trabalho são aquelas em que a relação entre a média e a variância é dada pela expressão:

$$S_x^2 = a \bar{x}^b$$

ou, ainda,

$$\text{Log } S_x^2 = \text{Log } a + b \text{ Log } \bar{x} \quad .$$

Aplicando-se a fórmula geral da transformação de dados, descrita em 2.3.1.1, tem-se:

$$T(X) = \int \frac{c}{\sqrt{a} x^b} dx$$

$$T(X) = \frac{c}{\sqrt{a}} \int \frac{1}{x^{1/2}} dx$$

$$T(X) = \begin{cases} \frac{c}{\sqrt{a} [1-(b/2)]} x^{[1-(b/2)]} & , \text{ se } b \neq 2 \\ \frac{c}{\sqrt{a}} \text{ Log } x & , \text{ se } b=2 \end{cases}$$

Então, para valores de b entre zero e quatro ($0 \leq b \leq 4$), recorre-se à Tabela 2, apresentada em 2.3.1.1 (p. 46).

Estas transformações (Tabela 2), são executadas na subrotina TRANR, do programa ROSAN. Para o processamento desta subrotina faz-se necessário saber qual o delineamento em estudo, o indicativo de necessidade de se usar esta subrotina, e o valor de tabela para o teste t . Os esclarecimentos necessários para sua utilização estão

descritos no Apêndice 3 deste trabalho.

Para verificação da transformação de dados, obtêm-se primeiramente a média e a variância dos grupos de dados. Em seguida, o logaritmo de cada média e de cada variância para a construção da equação de regressão linear simples:

$$\text{Log } S_x^2 = a + b \text{ Log } \bar{x} .$$

A obtenção de \underline{a} e \underline{b} é feita através das fórmulas usuais para estimação de α e β de uma equação de regressão linear simples.

Uma vez obtidos \underline{a} e \underline{b} , é aplicado um teste de hipótese, para verificação da existência de regressão, isto é:

$$H_0: b = 0 ,$$

$$H_a: b \neq 0 .$$

Se este teste de hipótese for significativo a um nível α de confiança, é possível aplicar uma das transformações da Tabela 2, de acordo com o valor de b .

Se o teste de hipótese não for significativo, o programa para de ser executado, imprimindo a mensagem de necessidade de outro tipo de transformação que não as existentes no programa, devendo isto ser feito à parte.

Para a execução dos programas ROSAN e ROSA1,

os cartões de comando e de entrada de dados são explicados no Apêndice 1 deste trabalho.

4. APRESENTAÇÃO DE UM EXEMPLO

4.1 - Introdução

Conforme comentado anteriormente, apresenta-se aqui a discussão de um exemplo com base em dados reais, coletados experimentalmente.

É necessário que seja ressaltado inicialmente o fato de que a direção tomada neste capítulo não é a única possível. Assim, por exemplo, será aqui enfocado apenas o "plot" total de resíduos, ao estilo de ANSCOMBE e TUKEY (1963), e não o "plot" específico para cada tratamento, como pode ser observado em BOX *et alii* (1978) e NETER e WASSERMAN (1974). Além disso, a análise de resíduos aqui efetuada não será exaustiva, no sentido de verificar

todas as possíveis transgressões e nem todas as técnicas descritas para detectá-las.

Nesse contexto, apresenta-se "apenas um exemplo" dentre as possíveis direções alternativas a serem tomadas na análise exploratória de dados. Em resumo, entende-se que dificilmente dois pesquisadores iniciados na área da análise de resíduos processariam de modo semelhante a autópsia de um experimento, no entanto, embora seguindo rumos diferentes, os desenlaces podem ser igualmente eficientes, desde que prevaleça a honestidade científica e o bom senso dos pesquisadores.

4.2 - Análise de Resíduos Considerando o Modelo em Látice

Dos dados referentes ao experimento com clones de seringueira (*Hevea* spp), foi feita a decomposição dos resíduos, a obtenção dos valores de Y_{ij} estimados, o quadrado médio do resíduo e o resíduo normalizado, conforme a Tabela 5, que é resultado do programa ROSAN, para a construção do gráfico dos resíduos.

Com os valores de Y_{ij} estimados e os resíduos normalizados, foi feita a descrição gráfica dos resíduos através do programa PLOTRES, conforme mostra a Figura 10.

Tabela 5 - Valores de Y estimado e valores ordenados dos resíduos padronizados do experimento em látice, fornecidos pelo programa ROSAN.

Y estimado	RESÍDUO (padronizado)	Y estimado	RESÍDUO (padronizado)
75,0614473	-1,8147559	12,7974510	0,0164540
44,4981156	-1,6232278	53,0181199	0,0713576
45,4714432	-1,6112430	37,0347825	0,0953269
55,6481171	-1,2480855	33,5414505	0,0982666
80,2561036	-1,0566465	11,0461178	0,1230496
57,4947739	-0,9188460	34,5454484	0,1292007
103,3981020	-0,8751854	23,3901176	0,1417730
43,6947785	-0,8557572	31,1347771	0,1950484
68,4581148	-0,8030698	36,7247849	0,2059021
15,2581138	-0,7142030	21,9181175	0,2138167
41,7114870	-0,6384508	14,1361141	0,2851818
39,3121186	-0,6079691	43,5407792	0,2896144
17,7714462	-0,5244834	61,5514374	0,2977096
40,4087830	-0,5100564	16,3447800	0,3151211
32,6681137	-0,4381031	26,4041176	0,3307688
34,8427811	-0,4309577	17,5641212	0,3395875
72,1514436	-0,4200128	44,4747773	0,3843162
23,8281136	-0,4014709	12,7481212	0,4851670
35,0547791	-0,3992095	3,4161210	0,4968351
49,3634415	-0,3923352	19,6067810	0,5259606
29,6447792	-0,3754666	15,4607830	0,5752106
39,2774506	-0,3742910	16,7614517	0,6023005
24,8614502	-0,3711704	34,8341217	0,6068682
56,7414429	-0,3691346	20,7527847	0,6707261
47,9547806	-0,3618990	89,0081025	0,7049630
30,9727859	-0,3278449	20,0507851	0,7183482
29,4787788	-0,3248596	26,0541229	0,7276190
38,6781159	-0,2637608	44,6614457	0,7372982
22,6727829	-0,2281233	6,8547830	0,7445334
33,9567872	-0,2222894	-1,9585497	0,7623972
31,5081215	-0,2135612	5,6374540	0,7979440
9,9207859	-0,1662556	26,0407829	0,8031454
57,7214432	-0,1595163	9,3734493	0,8294211
17,8327827	-0,1168695	14,0407848	1,3146414
57,9421158	-0,0924026	30,5127869	1,3965895
9,4747829	-0,0878350	36,3907853	1,5222249
21,1347809	-0,0552726	53,1614457	1,8389818
11,7574539	-0,0032190		

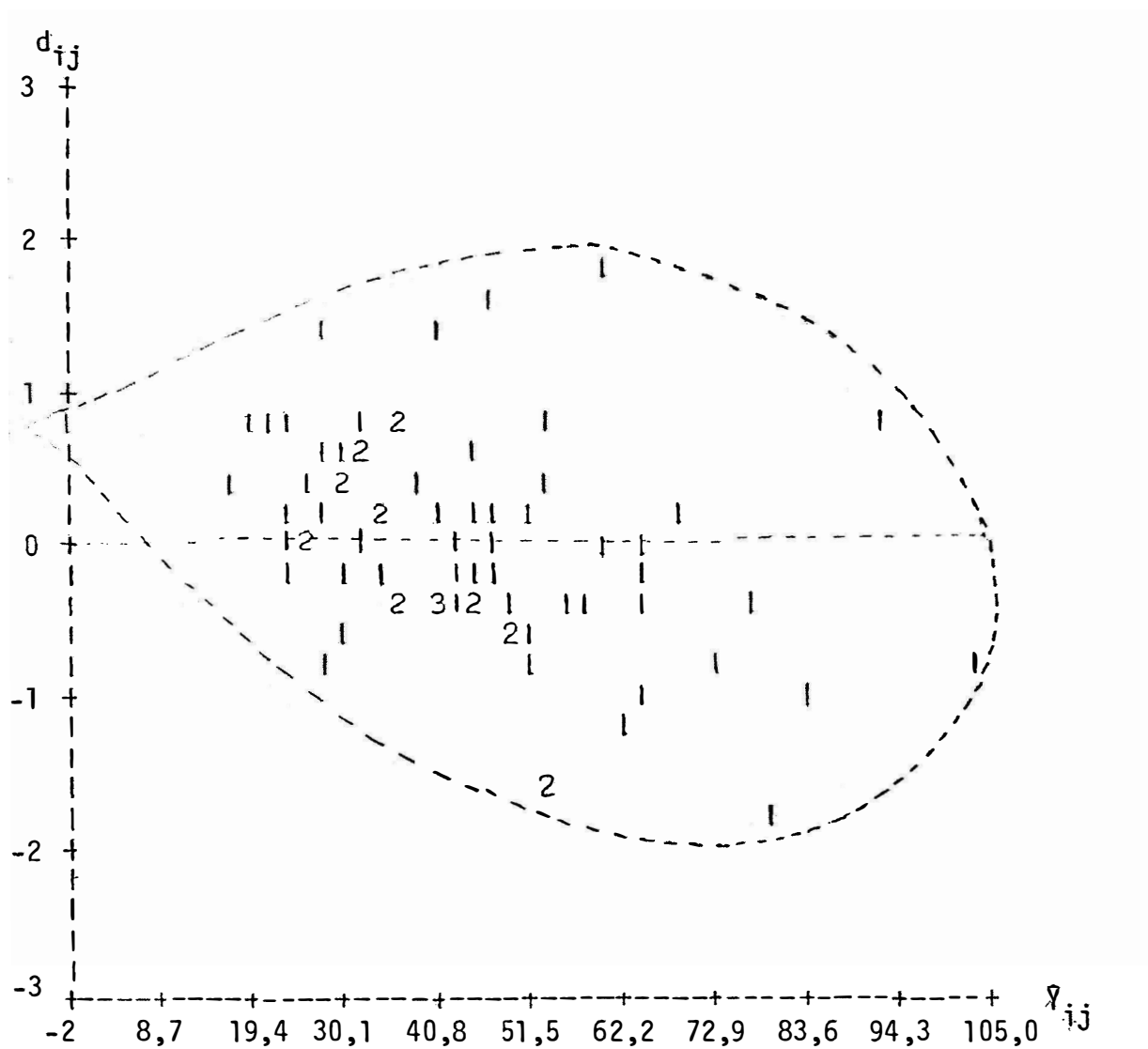


Figura 10 - Representação grãfica dos resĩduos padronizados versus \hat{Y}_{ij} estimados, de dados de um experimento em Lãtice.

- 1: "plot" de uma observação;
- 2: "plot" de duas observações;
- 3: "plot" de três observações.

Pode-se verificar que essa descrição gráfica não apresenta indícios de comportamento que evidenciem uma situação como a mostrada na Figura 1, que seria a condição ideal para se realizar a análise dos dados.

Nota-se, na Figura 10, uma tendência acen-tuadamente semelhante à Figura 4, sugerindo variância não constante dos resíduos. Por este gráfico, pode-se perceber a não existência de dados discrepantes que, conforme a literatura revista, seriam resíduos maiores que os compreendidos entre os valores de $(-3; 3)$. Exclui-se, assim, a possibilidade de que a heterogeneidade de variâncias detectada na análise gráfica tenha sido motivada por algum "outlier".

Muito embora a heterogeneidade de variâncias seja, por si só, extremamente importante e exija uma tomada de posição, efetuou-se ainda a análise gráfica da normalidade (Figura 11). Isto se deveu ao fato de que, sob normalidade, os dados com variâncias homogêneas podem, em geral, ser analisados sob o modelo generalizado de Gauss-Markov, através do método dos mínimos quadrados ponderados e, eventualmente, generalizados. Se, no entanto, os erros não são normais, então outras estratégias, como por exemplo, uma transformação de dados, podem ser mais adequadas.

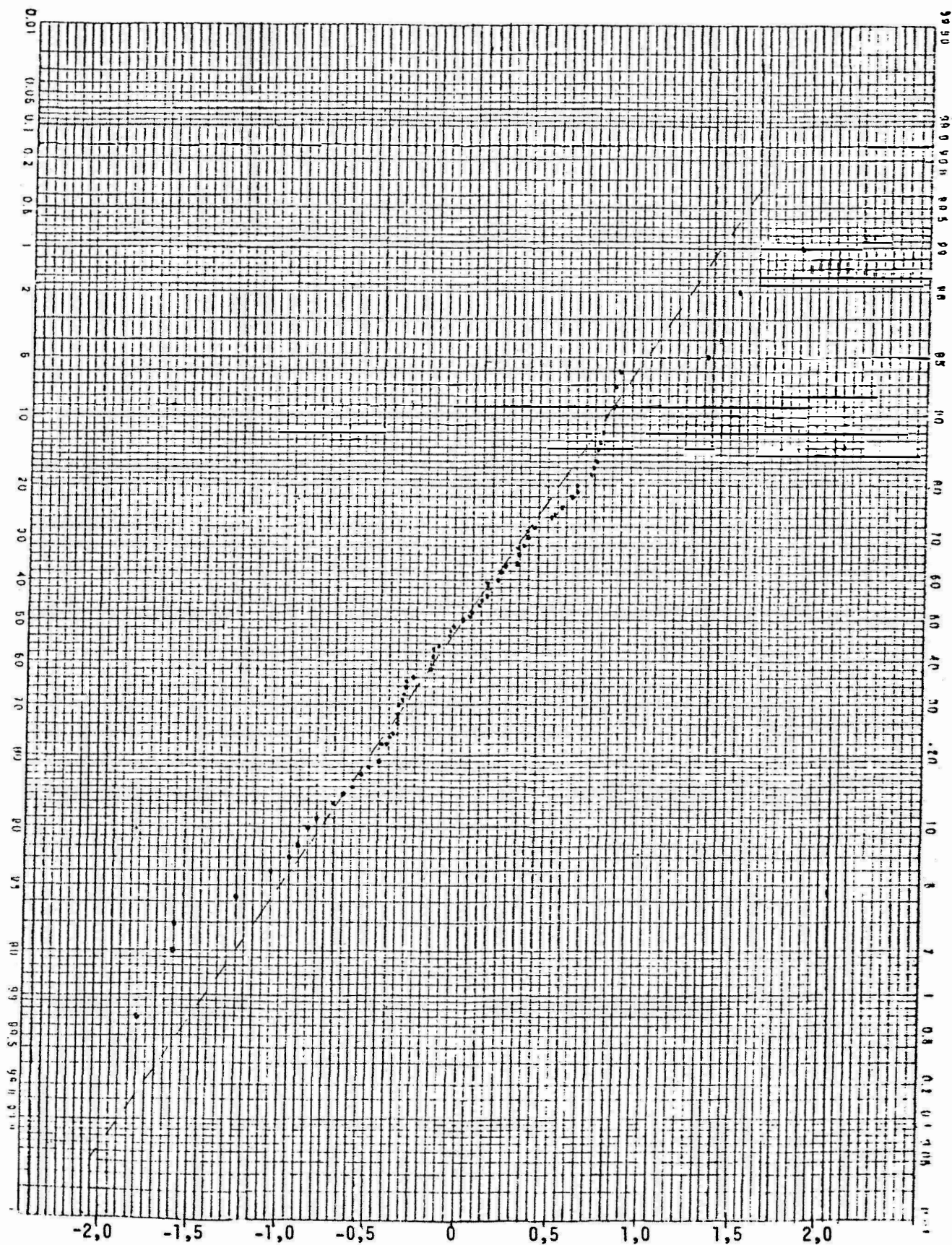


Figura 11 - Representação gráfica dos resíduos de um experimento em lâtilce, no "papel normal de probabilidade".

Observando-se a Figura 11, que mostra o comportamento dos resíduos num papel normal de probabilidade, nota-se uma razoável diferença entre esta e a Figura 9, que seria a situação na qual se aceitaria a normalidade dos resíduos.

A aplicação dos testes para aditividade e normalidade existentes no programa ROSAN apresentaram os resultados mostrados na Tabela 6. Observa-se que o teste de normalidade foi significativo, ao nível de 0,05 de probabilidade, concordando com a análise gráfica dos resíduos através do "papel normal de probabilidade".

Tabela 6 - Resultados do teste de não-aditividade e do teste de Lilliefors, fornecidos pelo programa ROSAN.

	Teste de Não-Aditividade	Teste de Lilliefors
Valor calculado	27,29019*	0,1501*
Valor tabelado	4,1186	0,1023

O teste de não-aditividade apresentado na Tabela 6 também foi significativo, ao nível de 0,05 de probabilidade, fato que não ficou suficientemente evidenciado graficamente.

O teste para verificação de heterogeneidade de variâncias deixou de ser efetuado tendo em vista as recomendações feitas por DEMÉTRIO (1984), descritas em 2.2.2, mas a análise gráfica mostra evidências ao fato de que a heterogeneidade pode existir.

Com as violações das pressuposições do modelo de Gauss-Markov, não se recomenda a aplicação direta da análise de variância aos dados. Faz-se necessário um estudo para se encontrar uma forma de amenizar as violações.

Como uma primeira tentativa, fez-se o estudo da viabilidade de se aplicar aos dados uma transformação que pudesse conduzir à aditividade do modelo, à homogeneidade da variância e à normalidade dos resíduos. Para tanto utilizou-se o programa ROSAN, que através da subrotina TRANR, a qual processa o método para transformação de dados (descrito em 3.2.4), apresentou os seguintes resultados:

Valor do coeficiente b:	$b = 0,62478$
Valor de \underline{t} calculado:	$t_{\text{calc.}} = 0,051802^{\text{NS}}$
Valor de \underline{t} tabelado:	$t_{\text{tab.}} = 2,160000 (\alpha=0,05)$

Verificou-se que não existe significância estatística que evidencie o uso de uma regressão linear

simples, a qual, pelo valor do coeficiente \underline{b} , indicaria uma transformação adequada aos dados.

Desse modo, não foram obtidas evidências que indicassem o uso de uma das transformações indicadas na Tabela 2.

Nesse momento, fez-se necessária uma tomada de posição. Sem dúvida, de acordo com a estratégia descrita anteriormente, alternativas como a exclusão de tratamentos ou a busca de resíduos específicos, dentre outras, poderiam ser adotadas.

No entanto, por se tratar de um esquema em Látice (Blocos Incompletos Balanceados), cuidou-se inicialmente de verificar sua eficiência, pois, segundo COCHRAN e COX (1967), em situações de baixa eficiência dos Látices, uma mudança para o modelo de Blocos Casualizados torna-se extremamente conveniente.

Tomou-se então, como base, o estudo de VASCONCELLOS (1982), que determinou a eficiência do Látice dos dados em questão, obtendo $E = 0,65$, valor considerado pequeno para COCHRAN e COX (1967).

Nesse contexto, optou-se pela adoção do modelo de Blocos Casualizados e os vinte e cinco tratamentos ficaram dispostos em três blocos, uma vez que as repetições do experimento em Látice passam a ser blocos.

4.3 - Análise de Resíduos Considerando o Modelo em Blocos Casualizados

Considerando-se o modelo em Blocos Casualizados, os dados ficam dispostos como na Tabela 7.

Tabela 7 - Dados de produção de borracha seca (PBS) dispostos em Blocos Casualizados.

B ₁	B ₂	B ₃
50,14	39,76	28,49
56,58	54,07	56,62
38,44	34,01	29,17
20,99	8,18	7,47
23,94	20,32	27,36
4,73	10,74	17,91
30,64	30,64	24,11
31,08	26,14	58,83
32,89	51,10	30,68
25,07	12,86	18,34
36,45	25,64	34,99
28,36	19,90	36,78
10,04	9,28	21,60
33,42	17,83	16,11
21,72	18,39	80,27
17,40	13,04	11,71
26,21	31,28	25,48
24,69	37,88	19,31
42,62	47,81	43,95
55,53	33,76	32,30
51,35	65,96	43,58
99,40	90,55	64,68
22,57	20,57	34,79
65,94	48,31	55,37
30,35	37,25	43,78

Com a matriz do novo modelo, processou-se novamente o programa, obtendo-se a decomposição dos resíduos, os Y_{ij} estimados, o quadrado médio do resíduo e os resíduos normalizados, conforme é mostrado na Tabela 8, que é o resultado do programa ROSAN para a construção do gráfico dos resíduos.

Com os valores de Y_{ij} estimados e dos resíduos normalizados, foi construído novamente o gráfico de resíduos, mostrado na Figura 12.

Verificou-se, através dessa representação, que ainda não se tem indícios de que os resíduos se comportem como na Figura 1, que seria a situação ideal para análise dos dados.

Esta descrição gráfica dos resíduos apresenta uma acentuada semelhança à Figura 2, indicativa de variância não constante dos resíduos. Verifica-se, ainda, a existência de um resíduo extremo, o qual, segundo a literatura revista, pode ser considerado um dado discrepante, uma vez que seu valor está fora do intervalo de (-3; 3).

Esse dado discrepante pode ser uma das causas da heterogeneidade de variância, e pode ainda levar os resíduos a não-normalidade.

Tabela 8 - Valores de Y_{ij} estimados e valores ordenados dos resíduos padronizados dos dados dispostos em Blocos ao Acaso, fornecidos pelo programa ROSAN.

Y estimado	RESÍDUO (padronizado)	Y estimado	RESÍDUO (padronizado)
85,4306184	-1,8377015	25,1020889	-0,1029159
38,3442231	-1,7671718	19,3106270	-0,0859600
41,3554230	-1,7389388	56,9854279	-0,0359050
40,0172959	-1,0208721	53,9742280	0,0084823
36,9008942	-0,9529985	56,3106308	0,0273983
54,1839524	-0,9390999	12,2675590	0,0684084
39,9120942	-0,7821815	29,6920891	0,0839485
41,0839615	-0,7779190	9,3442268	0,1236116
27,8472938	-0,7560739	35,3442307	0,1687781
38,7772980	-0,7171071	32,0908890	0,1699592
38,3554307	-0,7089710	32,9139634	0,1838570
12,3554268	-0,6753178	37,6808930	0,1841286
23,0072937	-0,6108335	15,2787590	0,1878598
26,5642281	-0,5901927	33,5887604	0,2533956
39,4520951	-0,5811473	24,4272919	0,2597246
54,7575532	-0,5710034	35,1020890	0,2956100
12,7672939	-0,4691352	26,6808891	0,3506244
34,4272919	-0,4655922	43,0108872	0,4250162
38,7475586	-0,4417046	19,9854240	0,4502971
30,5775566	-0,4372763	25,8742256	0,4787427
29,0172920	-0,4345962	37,6806336	0,5401686
14,8687572	-0,4276410	11,6806297	0,5516815
27,2054253	-0,4105191	14,1939602	0,6558887
16,9742241	-0,3643609	83,0942232	0,6602938
28,5220909	-0,3393748	13,4420910	0,6684528
24,1942253	-0,3209658	28,9006309	0,6978076
54,8587494	-0,3107389	57,7687531	0,7236560
46,0220872	-0,3012931	26,5306282	0,7314606
14,6039619	-0,2562930	40,6920929	0,8367198
20,6708908	-0,2515929	23,6820908	0,8624023
28,2106285	-0,2418277	25,5108910	1,0954248
28,8854256	-0,2369389	86,1054231	1,1773862
11,8575573	-0,2282718	41,7587586	1,2195999
10,4308910	-0,1993418	51,8475495	1,2498173
22,0908890	-0,1568319	36,4408951	1,2982304
57,0939560	-0,1526755	39,2372971	1,7351560
45,3472901	-0,1237454	40,6806260	3,5060873
29,5754280	-0,1076398		

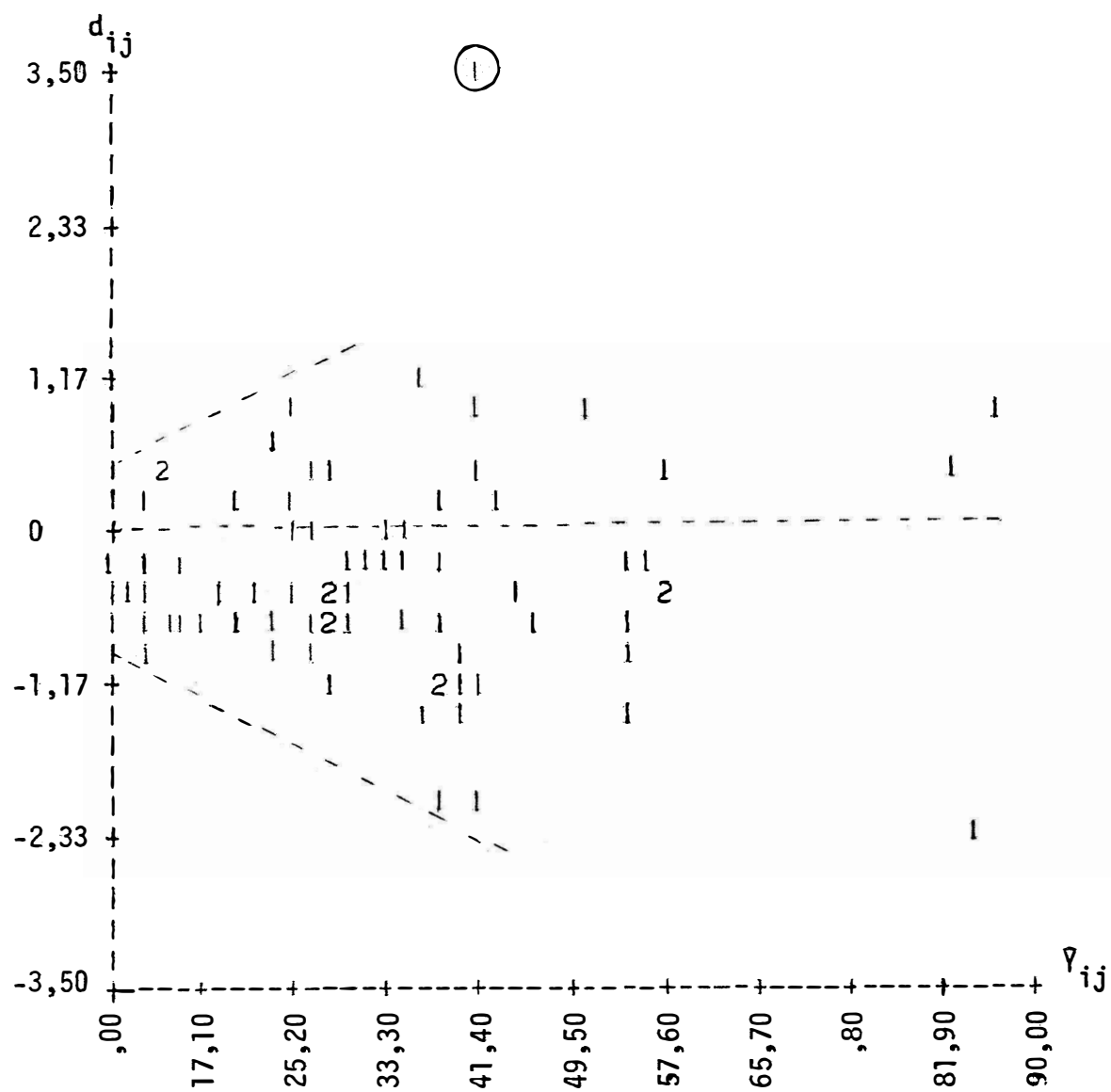


Figura 12 - Representação gráfica dos resíduos padronizados versus Y_{ij} estimados, dos dados dispostos em Blocos ao Acaso.
 1: "plot" de uma observação;
 2: "plot" de duas observações.

Para a verificação da normalidade dos resíduos, fez-se o gráfico deles, num "papel normal de probabilidade".

O comportamento dos resíduos, mostrado na Figura 13, não apresenta semelhança com a Figura 9, que seria a condição na qual se aceitaria a normalidade dos resíduos.

Observa-se, nesse gráfico, a localização de um ponto extremo, como verificado anteriormente. Este ponto provavelmente deve conduzir os resíduos à não-normalidade.

Embora as constatações visuais sejam suficientes para indicar transgressões das pressuposições adotadas, aplicaram-se os testes encontrados no programa ROSAN.

A Tabela 9 mostra o resultado do teste de Lilliefors para a normalidade dos resíduos. Verifica-se que o teste foi significativo ao nível de 0,05 de probabilidade, concordando, deste modo, com a análise gráfica dos resíduos feita através do "papel normal de probabilidade". Pode-se observar ainda, nesta tabela, que o teste para não-aditividade não foi significativo, isto é, não se rejeita a hipótese de que os dados podem ser estudados por um modelo aditivo.

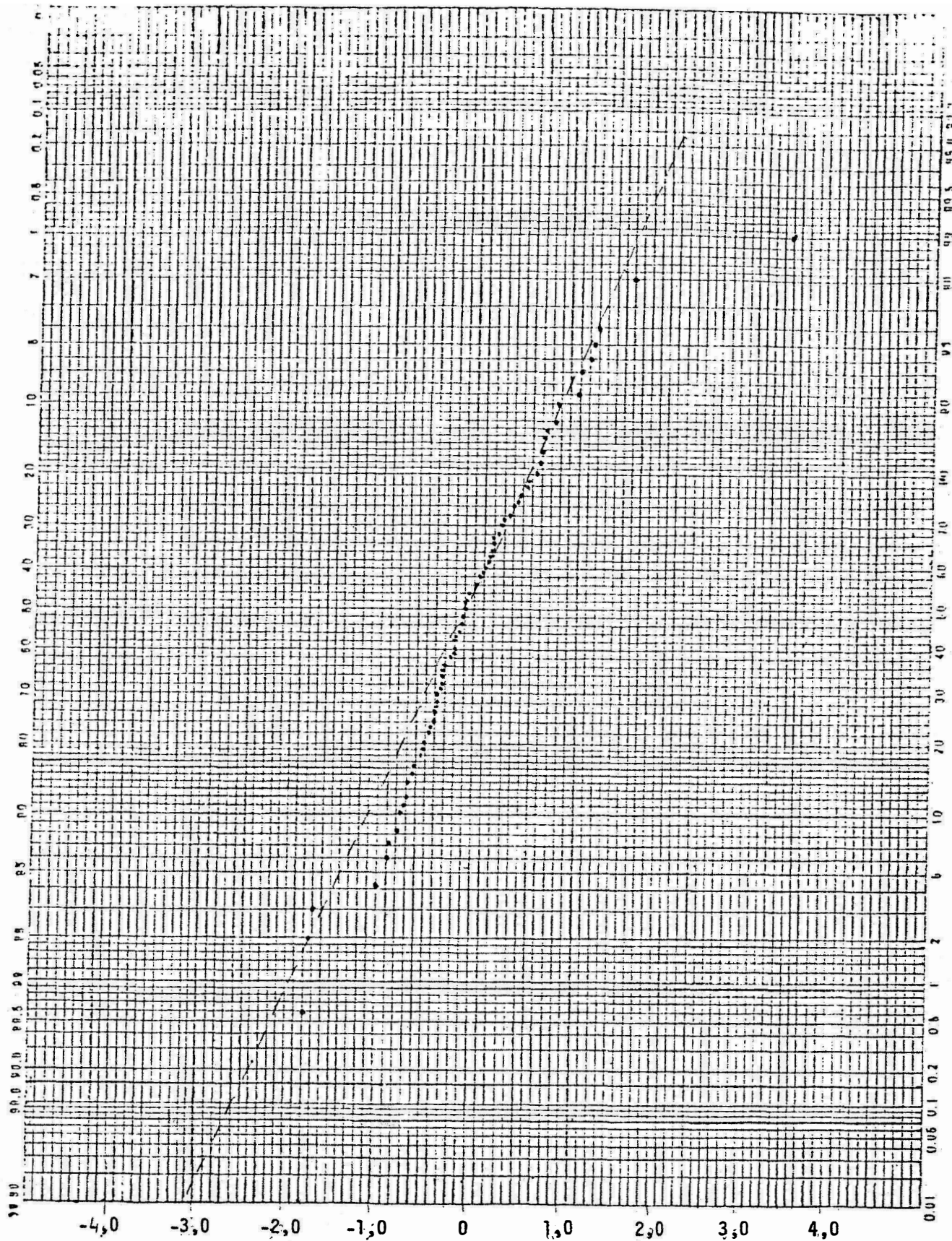


Figura 13 - Representação gráfica dos resíduos dos dados dispostos em Blocos ao Acaso, no "papel normal de probabilidade".

Tabela 9 - Resultados do teste de não-aditividade e teste de Lilliefors, fornecidos pelo programa ROSAN.

	Teste de Não-Aditividade	Teste de Lilliefors
Valor calculado	0,14043 ^{NS}	0,1370*
Valor tabelado	4,0400	0,1023

O teste para verificação de heterogeneidade de variâncias deixou de ser efetuado, tendo em vista as recomendações feitas por DEMÉTRIO (1984), descritas em 2.2.2, mas a análise gráfica mostra fortes evidências de existência de heterogeneidade de variâncias.

Como os dados não atendem às pressuposições feitas, não é recomendado o uso da análise de variância aos dados da maneira como eles se apresentam.

Verificou-se a existência de um resíduo superior a três (3), sendo o mesmo considerado como um dado discrepante ("outlier").

Com respeito a esta observação discrepante, como a literatura recomenda um estudo cauteloso quando se detecta um "outlier", optou-se por fazer a modificação de "outliers" sugerida por ANSCOMBE e TUKEY (1963), e descri

ta em 2.3.3.1, ao invés de simplesmente descartá-la. Essa escolha se deveu ao fato de que, descartada a observação, para que a análise fosse feita de modo usual (modelos balanceados), ela deveria ser estimada como se fosse uma parcela perdida e isto nos pareceu ser um contrasenso. Na verdade essa situação sempre fica caracterizada, desde que o modelo adotado não seja o de um fator inteiramente casualizado, e aí reside a grande vantagem do método proposto por ANSCOMBE e TUKEY (1963).

A grande dificuldade em se aplicar a metodologia da modificação de "outliers", reside no fato de se desconhecer qual a verdadeira distribuição dos resíduos.

Com vistas a este aspecto, construiu-se um histograma com todos os resíduos, e através da comparação visual destes, com gráficos da distribuição t de Student, observou-se que o histograma apresentava semelhança a esta distribuição. A partir deste ponto, foram aplicados vários testes de aderência de Kolmogorov-Smirnov aos resíduos, para se verificar qual seria o grau de liberdade adequado para a distribuição acima citada, através da qual se estudariam os resíduos.

O teste de Kolmogorov-Smirnov foi aplicado para 8, 9 e 10 graus de liberdade, supostamente conheci-

dos, respectivamente. Para nove graus de liberdade, verificou-se que a diferença máxima foi $D = 0,136$. Este resultado foi comparado com o valor $d = 0,154$, retirado da tabela de CAMPOS (1983) para o teste ao nível de 0,05 de probabilidade, e verificou-se a não rejeição da hipótese de nulidade do teste, isto é, não se rejeitou a hipótese de que os dados podem ser estudados através de distribuição t de Student, com nove graus de liberdade.

Uma vez verificada essa possibilidade, obtiveram-se os resíduos transformados, na forma descrita em 2.3.3.1. Esses cálculos foram executados através do programa ROSA 1, e os valores de Y_{ij} transformados constam da Tabela 10.

Com esses novos valores de Y_{ij} processou-se novamente o programa, obtendo-se a decomposição dos resíduos, os valores estimados de Y_{ij} e os resíduos normalizados, conforme mostra a Tabela 11.

Com os valores de Y_{ij} estimados e os resíduos normalizados, construiu-se novamente o gráfico dos resíduos, conforme mostra a Figura 14.

Tabela 10 - Valores de Y_{ij} transformados, obtidos através do programa ROSA1.

B_1	B_2	B_3
50,8679	41,9231	32,4729
54,6394	53,5601	58,9416
42,3672	36,8503	26,5765
20,4374	4,7008	3,4600
21,0337	16,8699	30,2811
5,7612	12,3237	19,5466
31,6277	34,2960	20,9214
27,4101	23,6181	39,4933
28,9094	40,2015	26,7924
28,0049	9,7589	17,9213
40,0992	24,1865	38,7529
26,0661	17,7489	34,6573
6,3547	5,6633	19,2515
30,0392	14,1574	12,5556
23,1679	20,0240	35,6474
20,6492	14,1145	8,0669
22,3570	33,0682	22,2341
20,6701	30,5733	16,2642
39,1351	50,5879	42,7115
46,4113	29,5587	29,0550
47,6131	56,8539	42,1093
100,1015	94,7095	82,3624
18,5824	16,4905	31,0499
65,6581	46,4512	53,9467
32,1709	39,4822	45,8650

Tabela 11 - Valores de Y_{ij} estimados e valores ordenados dos resíduos padronizados, obtidos a partir dos Y_{ij} transformados, fornecidos pelo programa ROSAN.

Y estimado	RESÍDUO (padronizado)	Y estimado	RESÍDUO (padronizado)
91,9107515	-1,2087004	43,6644516	-0,1206314
41,0602341	-1,1242777	54,8716202	-0,1170831
41,2742539	-1,1141412	54,3067398	-0,0945147
32,9392396	-1,1079866	18,0813217	-0,0202563
14,4313679	-1,0975346	30,8358955	0,1002326
34,7842865	-1,0390021	11,1368732	0,1502376
28,4679832	-0,9553027	12,8699054	0,1575500
53,9450379	-0,9486258	40,3476715	0,1994299
17,1547394	-0,9362202	37,7657395	0,2172825
24,7504730	-0,8863114	33,8577042	0,3788256
46,0323639	-0,8731071	55,2333222	0,4694221
48,3783951	-0,7935881	90,9841615	0,4715816
12,3107013	-0,7539551	36,2337342	0,4893202
18,4370232	-0,7445142	16,1644020	0,5677193
22,0221901	-0,7288767	33,8658219	0,6186438
13,7964878	-0,7252941	37,1521988	0,6601542
9,0523548	-0,7079218	94,2786562	0,7370994
34,5279542	-0,6928069	27,5414009	0,8550479
27,7739677	-0,6857200	38,6923218	0,9079719
23,9284668	-0,6767450	43,6421662	0,9146881
28,7668724	-0,6517697	12,0634556	0,9472723
28,1673012	-0,6328616	20,4492340	0,9564529
33,8553010	-0,6260886	42,7378693	0,9937163
24,8728066	-0,6137980	22,2478561	1,0169084
31,4873848	-0,5943262	57,2395325	1,0656859
32,0613709	-0,5887918	24,4794731	1,0872256
21,3212738	-0,5634877	25,6770554	1,1367864
20,6339722	-0,5245115	11,4202671	1,1414561
33,6013718	-0,5117511	20,8049355	1,1689429
24,3901024	-0,4709056	9,9427891	1,1783664
24,6157684	-0,4534446	47,4518128	1,1901872
8,1257724	-0,4335584	21,0956078	1,1997578
17,5104408	-0,4244527	21,5605545	1,2012324
9,0162067	-0,4244358	36,8958664	1,2045350
25,4060555	-0,4015298	30,5608025	1,2203922
50,7463074	-0,3966241	29,6934547	1,2405376
57,6012345	-0,3749306	25,7993889	1,2466354
28,0449676	-0,2504997		

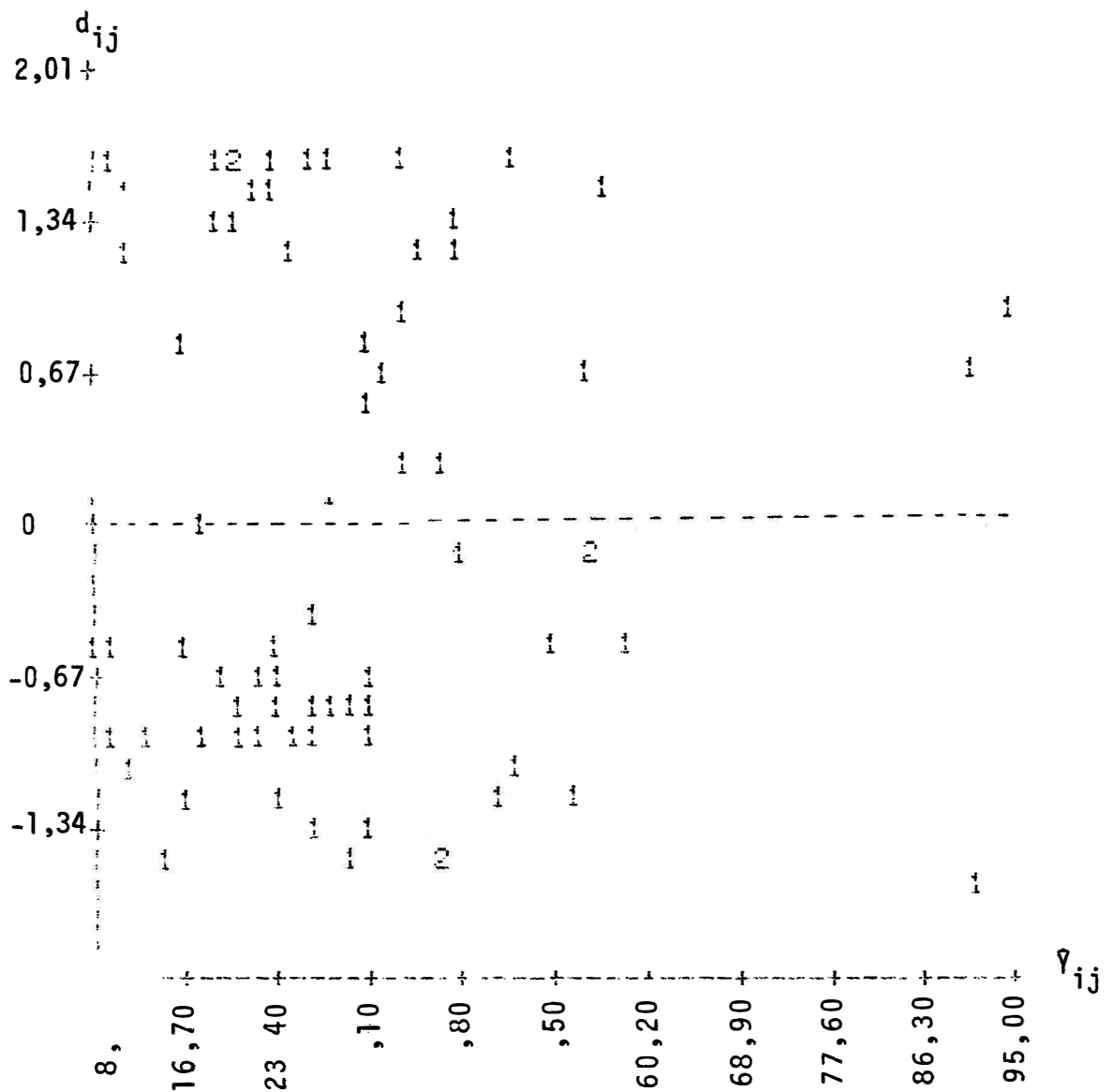


Figura 14 - Representação gráfica dos resíduos padronizados versus Y_{ij} estimado, depois da "modificação de dados discrepantes".
 1: "plot" de uma parcela;
 2: "plot" de duas parcelas.

Verifica-se, através dessa representação gráfica, uma acentuada semelhança com a Figura 1, que indica a situação ideal para se realizar a análise dos dados.

Dos testes existentes no programa ROSAN, a Tabela 12 apresenta o resultado do teste de não-aditividade, que se mostrou não significativo, e do teste de Lilliefors, significativo ao nível de 0,05 de probabilidade, fato que se poderia esperar, uma vez que os resíduos foram estudados através de uma distribuição t de Student.

O teste de heterogeneidade não foi executado pela mesma razão anteriormente descrita, mas pode-se observar, através da representação gráfica dos resíduos, que não existem evidências para que se suspeite da existência de heterocedasticidade de variâncias.

Tabela 12 - Resultados do teste de não-aditividade e teste de Lilliefors, fornecidos pelo programa ROSAN.

	Teste de não aditividade	Teste de Lilliefors
Valor calculado	0,06904 ^{NS}	0,1394*
Valor tabelado	4,0400	0,1023

Observa-se que apenas a normalidade não foi atendida, mas autores como SCHEFFÉ (1959), NETER e WASSERMAN (1974), dentre outros, mostram que o teste F é robusto contra a ausência de normalidade.

Com a obtenção dos resultados dos testes, e principalmente pela análise gráfica dos resíduos, verifica-se a adequacidade dos "novos" dados para a realização da análise através do modelo de Gauss-Markov.

5. CONCLUSÕES

Dados os objetivos e as diretrizes estabelecidas, as conclusões relacionadas a seguir não são privilégio deste estudo, mas, sim, de domínio da bibliografia consultada. Parece, no entanto, prudente relacioná-las como subsídios para os usuários do modelo de Gauss-Markov.

- 1 - A análise gráfica de resíduos apresenta, sobre a inferência com resíduos, a vantagem de não depender da estrutura dos dados.
- 2 - A existência de valores discrepantes leva, em geral, à ausência de normalidade e à heterogeneidade de variâncias.

- 3 - O processo da modificação de "outliers" pode ser uma alternativa extremamente conveniente quando é adotado um modelo com dois ou mais fatores, pois evita a estimação de parcelas perdidas.
- 4 - A forma geral para obtenção de transformações adequadas, sugerida por RAO (1965), sempre que aplicável, constitui-se num processo eficiente, no sentido de que direciona o processo e evita a procura desordenada de boas transformações.
- 5 - Se os erros apresentam variâncias heterogêneas, então, dentre outras, duas alternativas podem ser estudadas: sob normalidade, o método dos mínimos quadrados ponderados, e na ausência de normalidade, uma transformação de dados.
- 6 - Se os erros apresentam correlação, de modo análogo, duas alternativas, dentre outras, podem ser consideradas: sob normalidade, a adoção do modelo generalizado de Gauss-Markov ou a inclusão no modelo de algum parâmetro como, por exemplo, o tempo, que poderá estar motivando tal correlação. Na ausência de normalidade, pode ser mais prudente a busca de uma transformação adequada.

6. REFERÊNCIAS BIBLIOGRÁFICAS

- ANSCOMBE, F.J., 1961. Examination of residual. IN: Berkeley Symposium of Mathematical Statistics and Probability, 4, Berkeley, ed. by J. Reyzon Berkeley. University of Califórnia, 1: 1-36.
- ANSCOMBE, F.J., 1973. Graphs in statistical analysis. American Statistician, 27(2): 17-21.
- ANSCOMBE, F.J. e J.W. TUKEY, 1963. The examination and analysis of residuals. Technometrics, 5: 141-160.
- BARBOSA, J.C., 1983. O problema da heterocedasticidade na análise de variância. Seminários de Estatística e Experimentação Agronômica, Piracicaba, ESALQ/USP. 14 p. [Mi meografado].

- BOX, G.E.P.; W.G. HUNTER Jr. e J.S. HUNTER, 1978. Statistics for Experimenters: An Introduction to Design, Data Analysis and Model Building. John Wiley, New York, 653 p.
- CAMPOS, H. de, 1983. Estatística Experimental Não-Paramétrica. 4ª ed., Deptº de Matemática e Estatística da ESALQ/USP., Piracicaba, 343 p.
- COCHRAN, W.G., 1947. Some consequences when the assumptions for the analysis of variance are not satisfied. Biometrics, 3: 22-38.
- COCHRAN, W.G. e G.M. COX, 1967. Experimental Design. 2ª ed. John Wiley, New York. 661 p.
- COX, D.R. e E.J. SNELL, 1968. A general definition of residuals. J.R. Stat. Soc., B, 30: 248-275.
- DE FINNETTI, B., 1961. The Bayesian Approach to the Rejection of Outliers. 1 Proceedings of the Fourth Berkeley Symposium on Mathematical Statistics and Probability. University of California Press. 199-210.
- DEMÉTRIO, C.G.B., 1978. Transformação de Dados. Efeitos sobre a análise de variância. Piracicaba, ESALQ/USP, 113 p. [Dissertação de Mestrado].
- DEMÉTRIO, C.G.B., 1984. Testes para homogeneidade de variâncias para ensaios em blocos casualizados. Seminários de Estatística e Experimentação Agronômica, Piracicaba, ESALQ/USP, 10 p. [Mimeografado].
- DIXON, W.J. e F.J. MASSEY Jr., 1969. Introduction to Statistical Analysis. 3ª ed. McGraw-Hill Book Company, New York, 638 p.

- DRAPER, N.R. e H. SMITH, 1966. Applied Regression Analysis. John Wiley, New York, 407 p.
- DURBIN, J. e G.S. WATSON, 1971. Testing for serial correlation in least squares regression. III. Biometrika, 56: 1-19.
- FIGUEIREDO, G.C., 1977. Teste baseado no número total de chorrilhos. Seminários de Estatística e Experimentação Agronômica, Piracicaba, ESALQ/USP. 10 p. [Mimeografado].
- HAN, C.P., 1969. Testing the homogeneity of variances in a two-way classification. Biometrics, 25(1): 153-158.
- JEFFREYS, H., 1939, 1948. Theory of Probability. University Press, Oxford.
- KEMPTHORNE, O., 1952. The design and analysis of experiments. John Wiley & Sons, Inc., New York. 631 p.
- LINDGREN, B.W., 1976. Statistical Theory. 3ª ed. McMillan Pub., New York, 614 p.
- MAGE, D.T., 1982. An objective graphical method for testing normal distributional assumptions using probability plots. The American Statistician, 36(2): 116-120.
- MATTIOLLI, C.H., 1983. Métodos de seleção de equações de regressão linear múltipla. Piracicaba, ESALQ/USP, 104 p. [Dissertação de Mestrado].
- NETER, J. e W. WASSERMAN, 1974. Applied Linear Statistical Models - Regression Analysis of Variance, and Experimental Designs. Richard D. Irwin, Inc. Illinois, 842 p..

- PEARSON, E.S. e H.O. HARTLEY, 1956. Biometrika Tables for Statisticians. 1ª ed. Cambridge, The University Press, 238 p.
- POOLE, M. e M. BORCHERS, 1979. Some Common Basic Program. 3ª ed. McGraw-Hill. Berkeley, California, 92 p.
- RAO, C.R., 1965. Linear Statistical Inference and Its Applications. J. Wiley & Sons, Inc. New York, 522 p.
- SCHEFFÉ, H., 1959. The Analysis of Variance. John Wiley & Sons, Inc. New York, 477 p.
- SEBER, G.A.F., 1977. Linear Regression Analysis. John Wiley, New York, 465 p.
- SNEDECOR, G.W., 1956. Statistical Methods. 5ª ed. Iowa State University Press, Ames, Iowa.
- STEEL, R.G.D. e T.H. TORRIE, 1960. Principles and Procedures of Statistics. McGraw-Hill Book Co., Inc., New York, 481 p.
- TUKEY, J.W., 1949. One degree of freedom of non-additivity. Biometrics, 5: 232-242.
- VASCONCELLOS, M.E. da C., 1983. Análise do coeficiente de caminhamento ("Path Coefficient") e estimativas de parâmetros genéticos em clones de seringueira (*Hevea* spp). Piracicaba, ESALQ/USP, 77 p. [Dissertação de Mestrado].

7. A P È N D I C E S

APÊNDICE 1 - INSTRUÇÕES PARA USO DO PROGRAMA ROSAN PARA COMPUTADORES IBM 1130

1 - Cartões de leitura necessários para a execução do programa:

a) 1º Cartão: NDEL

A variável NDEL indica quantas análises se deseja fazer. Seu formato é inteiro e seu tamanho é I2. Isto significa que o programa aceita até 99 análises.

b) 2º Cartão: ITRAN

Esta variável ITRAN indica se o usuário deseja ou não o uso das transformações de dados existentes no programa, da seguinte forma:

0 se não deseja transformação,

1 se deseja transformação.

Seu formato é inteiro e seu tamanho é I2.

c) 3º Cartão: JDEL, CTAB, FCAL, TAB, DTAB

Descrevendo cada variável:

JDEL: esta variável indica qual o delineamento dos dados que serão estudados, sendo:

- 1: se for Inteiramente ao Acaso;
- 2: se for Blocos ao Acaso;
- 3: se for Quadrado Latino;
- 4: se for o esquema em Látice Quadrado.

Seu formato é inteiro e seu tamanho é 12.

CTAB: valor retirado da tabela para o teste de Cochran. Esta variável é real e seu formato é F8.5.

FCAL: valor retirado de uma tabela de distribuição F, para o teste de não-aditividade. Esta variável é real e seu formato é F8.5.

TAB: valor retirado da tabela de uma distribuição t de Student, para verificação da existência ou não de regressão linear simples, para o uso das transformações existentes no programa. Esta variável é real e seu formato é F8.5.

DTAB: valor retirado de uma tabela de Lilliefors, para o teste de normalidade dos resíduos. Esta variável é real e seu formato é F8.5.

d) 40 Cartão: tamanho do delineamento.

Se for Inteiramente ao Acaso: M,N, onde,

M = número de repetições; e

N = número de tratamentos.

Se for Blocos ao Acaso: M,N, onde,

M = número de blocos; e

N = número de tratamentos.

Se for Quadrado Latino: M,N, onde,

M = número de colunas; e

N = número de linhas.

Se for Látice Quadrado: N,M,L, onde:

N = número de linhas por repetição;

M = número de colunas por repetição;

L = número de repetições.

Todas as variáveis são inteiras, e para os três primeiros casos, seu formato é 2I2, e para o último 3I2.

e) Leitura dos dados para execução do programa:

Inteiramente ao Acaso: em cada cartão cabem até 8 dados com o formato F10.4, como repetições de um tratamento.

Blocos ao Acaso: em cada cartão cabe até 8 dados com o formato de F10.4, como blocos de um tratamento.

Quadrado Latino: em cada cartão entre o número de tratamento e o seu valor respectivo, sendo que cada cartão comporta até 5 tratamentos e seu valor respectivamente. O seu formato é: 5(I2,F6.3).

Látice Quadrado: em cada cartão o processo é o mesmo que para o quadrado latino, observando-se apenas a ordem das repetições, cabendo, no total, 10 observações por cartão. Seu formato é 5(I2,F6.3).

APÊNDICE 2 - PROGRAMA PARA GRÁFICOS DE RESÍDUOS, ELABORADO POR
MATIOOLI (1983) PARA MICROCOMPUTADOR PET/CBC 3032

```

100 REM ***** PROGRAMA PARA PLOTAR RESÍDUOS *****
110 POKE59468,14: DIM X(100,15),X*(100)
135 INPUT"Q# DE LINHAS, # DE COLUNAS":NL,NC
150 FORI=1TONL
160 FORJ=1TONC
165 PRINT"LINHA ":I;"-COLUNA ":J;
170 INPUTX(I,J)
180 NEXTJ,I
190 CLOSE2
  INPUT"QUAL COLUNA É A ABSCISSA":AB
  INPUT"QUAL COLUNA É A ORDENADA":OD
  PRINT"AGUARDE UM INSTANTE !"
  M=NL-1:AI=1E20:AS=-1E20
  K=0:FORI=1TOM
    IFX(I,AB)<=AI THENAI=X(I,AB)
232 IFX(I,AB)>=ASTHENAS=X(I,AB)
240 IFX(I,OD)>=X(I+1,OD) THEN310
250 FORJ=1TONC
  TR=X(I,J)
  X(I,J)=X(I+1,J)
280 X(I+1,J)=TR
290 NEXT
300 K=1
310 NEXT
320 M=M-1:IFK=1 THEN230
  PRINT"ORDENADA \MINIMA"=X(NL,OD)
340 PRINT"ORDENADA \MAXIMA"=X(1,OD)
350 PRINT"ABSCISSA \MINIMA"=AI
  PRINT"ABSCISSA \MAXIMA"=AS
370 PRINT"IXE A ESCALA HORIZONTAL"
380 INPUT"LIMITE INFERIOR":LI
390 INPUT"LIMITE SUPERIOR":LS
  PRINT"IXE A ESCALA VERTICAL"
410 INPUT"LIMITE INFERIOR":VI
420 INPUT"LIMITE SUPERIOR":VS

  X*(0)="+":FORI=1TO10:X*(0)=X*(0)+"----+":NEXT
466 OPEN3,4:OPEN1,4,1:OPEN2,4,2:PRINT#2,"S9,99":PRINT#3
470 FORI=1TOLH:X*(I)=" ":FORJ=1TOLV:X*(I)=X*(I)+" ":NEXT:NEXT
490 FORI=5TOLHSTEP5:X*(I)=" "+MID$(X*(I),2):NEXT
495 FORI=1TONL:K=1
500 ILX=INT((X(I,OD)-VI)*LH/(VS-VI)+.5)
505 IFILX>LV THENILX=LV
510 COM=INT((X(I,AB)-LI)*LV/(LS-LI)+.5)+1
520 FX=VAL(MID$(X*(ILX),COM,1))
525 IFFX=K THENK=K+1:GOTO525
530 X*(ILX)=MID$(X*(ILX),1,COM-1)+MID$(STR$(K),2)+MID$(X*(ILX),COM+1)
535 NEXTI:PRINT#2,C$
540 II=5:FORI=LHTO0STEP-1
570 PRINT#3,TAB(18)X*(I)
590 NEXT
580 PRINT"VAVANCE=0 PAPEL PARA COMENTARIOS"
610 GETY$:IFY$="" THEN610
620 IFY$<"A" THEN610
640 INPUT"NOME DA VARIÁVEL":NV$
650 INPUT"COMENTARIO":C$
655 PRINT#3,"#C$
660 PRINT#3,"ABSCISSA"=NV$+"LI"R"LS"!"
680 CLOSE1:CLOSE2:CLOSE3
READY.

```

APENDICE 3 - PROGRAMA ROSAN ELABORADO PARA COMPUTADOR IBM 1130.

```

// FOR
*ONE WORD INTEGERS
*LIST SOURCE PROGRAM
  SUBROUTINE ORDE(JDFL,N,M,L,RES,RES1,RNORM,RNOR1,YEST,YEST1,DTAR,YORD,
1RD,RNORD,K,P,DSUP,S)
  DIMENSION RES(25,10),RS1(5,5,5),RNORM(25,10),RNOR1(5,5,5),RNORD(25
10),YORD(20),RORD(20),YST(25,10),YEST1(5,5,5),P(20),S(20)
  GO TO (660,660,660,665),JDFL
  COLOCANDO OS RESIDUOS UM VETOR
660 K=0
  DO 680 I=1,N
  DO 685 J=1,M
  K=K+1
  RORD(K)=RES(I,J)
  RNORD(K)=RNORM(I,J)
  YORD(K)=YEST(I,J)
  P(K)=(K-0.5)/(N*M)
  A=FLOAT(K)

685 S(K)=A/3
680 CONTINUE
C   ORDENANDO OS RESIDUOS NORMALIZADOS
  L=M*N
  L1=L-1
  DO 690 I=1,L1
  L2=I+1
  DO 695 J=L2,L
694 A=RNORD(I)
  B=YORD(I)
  C=RORD(I)
  RNORD(I)=RNORD(J)
  YORD(I)=YORD(J)
  RORD(I)=RORD(J)
  RNORD(J)=A
  YORD(J)=B
  RORD(J)=C
695 CONTINUE
690 CONTINUE
  CALL LILLI(JDFL,N,M,L,RES,RES1,RNORM,RNOR1,YEST,YEST1,DTAR,YORD,KN
1ORD,K,P,DSUP,S)
  RETURN
C   LATICE QUADRADO
C   COLOCANDO NUM VETOR OS RESIDUOS
665 K1=0
  DO 700 K=1,L
  DO 705 I=1,N
  DO 710 J=1,M
  K1=K1+1
  RORD(K1)=RES1(K,I,J)
  RNORD(K1)=RNOR1(K,I,J)
  YORD(K1)=YEST1(K,I,J)
  P(K1)=(K1-0.5)/(L*N*M)
  A=FLOAT(K1)
  B=FLOAT(L*N*M)
710 S(K1)=A/B

700 CONTINUE
C   ORDENANDO OS RESIDUOS NORMALIZADOS
  K2=M*N*L

```


PAGE 2

```

      K3=K2-1
      DO 715 I=1,K3
      K4=I+1
      DO 720 J=K4,K2
      IF(RNORD(I)-RNORD(J))720,720,721
721  A=RNORD(I)
      B=YORD(I)
      C=RORD(I)
      RNORD(I)=RNORD(J)
      RORD(I)=RORD(J)
      YORD(I)=YORD(J)
      RNORD(J)=A
      YORD(J)=B
      RORD(J)=C
720  CONTINUE
715  CONTINUE
      K=K1
      CALL LILLI(JDEL,N,M,L,RES,RES1,RNORM,RNOR1,YEST,YEST1,DTAP,YORD,R
1ORD,K,P,DSUP,S)
      RETURN
      END

```

FEATURES SUPPORTED
ONE WORD INTEGERS

COFF REQUIREMENTS FOR ORDE
COMMON C VARIABLES 178 PROGRAM 698

RELATIVE ENTRY POINT ADDRESS IS 0086 (HEX)

END OF COMPILATION

// DUP

*DELETE ORDE
CART ID 0002 DB ADDR 54FA DB CNT 002E

*STORE WS UA ORDE
CART ID 0002 DB ADDR 5D41 DB CNT 002E

PAGE 1

// JOB

LOG DRIVE	CART SPEC	CART AVAIL	PHY DRIVE
0000	0002	0002	0000

ACTUAL 16K CONFIG 16K

// FOR

*ONE WORD INTEGERS

*LIST SOURCE PROGRAM

```

SUBROUTINE TEVON(JDEL,RES,N,M,L,RES1,N1,N2,COCAL,CTAB)
DIMENSIONRES(25,10),RES1(5,5,5),N2(5,5,5),N1(5,5),SIGM2(25)
GO TO (350,360,360,360),JDEL

```

350 DO 380 I=1,N

SIGM2(I)=0

DO 385 J=1,M

385 SIGM2(I)=SIGM2(I)+RES(I,J)**2

SIGM2(I)=SIGM2(I)/(M-1)

380 CONTINUE

C ORDERANDO AS VARIANCIAS

I1=N-1

DO 390 I=1,I1

I2=I+1

DO 395 J=I2,N

IF(SIGM2(I)-SIGM2(J))395,396,396

396 A=SIGM2(I)

SIGM2(I)=SIGM2(J)

SIGM2(J)=A

395 CONTINUE

390 CONTINUE

C TESTE DE COCHRAN

SOMAT=0

DO 400 I=1,N

400 SOMAT=SOMAT+SIGM2(I)

COCAL=SIGM2(N)/SOMAT

WRITE(3,6002)COCAL,CTAB

6002 FORMAT(30X,'TESTE DE COCHRAN',2X,F10.4,/,30X,'VALOR TABELADO',2X,F10.4,/,)

360 RETURN

END

FEATURES SUPPORTED

ONE WORD INTEGERS

CORE REQUIREMENTS FOR TEVON

COMMON C VARIABLES

272

RELATIVE ENTRY POINT ADDRESS IS 005C (HEX)

END OF COMPILATION

// DUP

*DELETE	TEVON			
CART ID 0002	DB ADDR 54FA	DB CNT	0013	

*STORE	WS UA TEVON			
CART ID 0002	DB ADDR 505F	DB CNT	0013	

```

// JOB
LOG DRIVE   CART SPEC   CART AVAIL  PHY DRIVE
  0000       0002       0002         0000

V2 M11   ACTUAL 16K   CONFIG 16K

// FOR
*ONE WORD INTEGERS
*LIST SOURCE PROGRAM
  SUBROUTINE TESND(JDEL ,N,M,L,YEST,RES,SCORES,GLRES,N1,N2,RES1,YEST1,
  1ESTF)
  DIMENSIONYEST(25,10),RES(25,10),N1(5,5),N2(5,5,5),RES1(5,5,5),YEST
  11(5,5,5),Q(25,25),Z(25,25),XMTG(25),XMBG(25),XMLQ(25),XMC(25),G1(
  15,5,5),XMRG(5,5),QEST1(5,5,5),Z1(5,5,5),QEST(25,25)
  GO TO (480,491,482,483),JDEL
C     CALCULO DE R(I,J)

C     BLOCOS CASUALIZADOS
481  X2=0
      DO 495 I=1,N
      DO 501 J=1,M
      Q(I,J)=YEST(I,J)**2
501  X2=X2+Q(I,J)
      XMTG(I)=X2/M
      X2=0

      XMGO=0
      X3=0
      DO 505 J=1,M
      DO 510 I=1,N
      X3=X3+Q(I,J)
510  XMGO=XMGO+Q(I,J)
      XMRG(J)=X3/M
      X3=0
505  CONTINUE
      XMGO=XMGO/(M*N)
      PCALC=0
      QCALC=0
      DO 515 I=1,N
      DO 520 J=1,M
520  QEST(I,J)=XMTG(I)+XMRG(J)-XMGO
515  CONTINUE
      DO 525 I=1,N
      DO 530 J=1,M
      Z(I,J)=Q(I,J)-QEST(I,J)
      QCALC=QCALC+Z(I,J)**2
530  PCALC=PCALC+RES(I,J)*Z(I,J)
525  CONTINUE
      GO TO 526
C     CUADRADO LATINO
482  DO 540 I=1,N
      DO 545 J=1,M
545  Q(I,J)=YEST(I,J)**2
540  CONTINUE
      X4=0
      X5=0
      XMGO=0
      DO 550 I=1,N

```

PAGE 2

```

DO 555 J=1,M
X4=X4+Q(I,J)
X5=X5+Q(J,I)
555 XMGQ=XMGQ+Q(I,J)
XMLQ(I)=X4/N
XMG(I)=X5/N
X4=0
550 X5=0
XMGQ=XMGQ/(N*M)
C CALCULO DA MEDIA DE TRATAMENTO
X6=0
DO 560 K=1,N
DO 565 I=1,N
DO 570 J=1,M
IF(N1(I,J)-K)570,571,570
570 CONTINUE
571 X6=X6+Q(I,J)
565 CONTINUE
XMTQ(K)=X6/N
560 X6=0
C CALCULO DE QEST
DO 580 K=1,N
DO 585 I=1,N
DO 590 J=1,M
IF(N1(I,J)-K)590,591,590
590 CONTINUE
591 QEST(I,J)=XMLQ(I)+XMG(I)+XMTQ(K)-XMGQ*2
585 CONTINUE
580 CONTINUE
PCALC=0
QCALC=0
DO 600 K=1,N
DO 605 I=1,N
DO 610 J=1,M
IF(N1(I,J)-K)610,611,610
610 CONTINUE
611 Z(I,J)=Q(I,J)-QEST(I,J)
QCALC=QCALC+Z(I,J)**2
605 PCALC=PCALC+RES(I,J)*Z(I,J)
600 CONTINUE
GO TO 526
C LATICE QUADRADO
483 X7=0
I3=M*N
DO 615 IK=1,I3
DO 620 K=1,L
DO 625 I=1,M
DO 625 J=1,N
IF(N2(K,I,J)-IK)625,626,625
625 CONTINUE
626 Q1(K,I,J)=YEST1(K,I,J)**2
C CALCULO DA MEDIA DE TRATAMENTO
620 X7=X7+Q1(K,I,J)
XMTQ(IK)=X7/L
X7=0
615 CONTINUE
C CALCULO DA MEDIA DO BLOCO NA REP K ,MEDIA GERAL
X8=0
XMGQ=0

```

PAGE 3

```

DO 630 K=1,L
DO 635 I=1,N
DO 640 J=1,M
XR=XB+Q1(K,I,J)
640 XMGO=XMGO+Q1(K,I,J)
XMRQ(K,I)=XR/M
635 XR=0
630 CONTINUE
XMGO=XMGO/(L*N*M)
QCALC=0
PCALC=0
I6=M*N
DO 645 IK=1,I6
DO 650 K=1,L
DO 655 I=1,N
DO 655 J=1,M
IF(R2(K,I,J)-IK)655,656,655
655 CONTINUE
656 QEST1(K,I,J)=XHTQ(IK)+XMRQ(K,I)-XMGO
Z1(K,I,J)=Q1(K,I,J)-QEST1(K,I,J)
QCALC=QCALC+Z1(K,I,J)**2
650 PCALC=PCALC+RFS1(K,I,J)*Z1(K,I,J)
645 CONTINUE
526 SQAD=(PCALC**2)/QCALC
SQRF=SQRF5-SQAD
QMR=SQRF/(QLF5-1)
ESTF=SQAD/QMR
RETURN
END

```

FEATURES SUPPORTED
ONE WORD INTEGERS

CORE REQUIREMENTS FOR TESND
COMMON C VARIABLES 4798 PROGRAM 1

RELATIVE ENTRY POINT ADDRESS IS 12B7 (HEX)

END OF COMPILATION

// DUP

```

*DELETE          TESND
CART ID 0002    DB ADDR 54FA    DB CNT  0052

*STORE          WS UA TESND
CART ID 0002    DB ADDR 5D22    DB CNT  0052

```

PAGE 1

// JOB

LOG DRIVE	CART SPEC	CART AVAIL	PHY DRIVE
0000	0002	0002	0000

ACTUAL 16K CONFIG 16K

// FOR

*ONE WORD INTEGERS

*LIST SOURCE PROGRAM

SUBROUTINE TRANR(JDEL,N,M,L,Y,Y1,N1,N2,TAB,IFLAG)

DIMENSIONY(25,10),Y1(5,5,5),XMED(50),XM(50),SIGM2(50),N1(5,5),N2(5

1,5,5),T1(25,25),TY1(5,5,5)

IFLAG=1

GO TO (920,920,930,935),JDEL

920 AMED=0

AMG=0

DO 940 I=1,N

DO 945 J=1,M

945 AMED=AMED+Y(I,J)

XMED(I)=AMED/M

A=XMED(I)

XM(I)=ALOG(A)*0.4343

AMG=AMG/M

ASM=0

SMG=0

DO 950 I=1,N

DO 955 J=1,M

955 ASM=ASM+Y(I,J)**2

SIGM2(I)=(ASM-M*(XMED(I)**2))/(M-1)

R=SIGM2(I)

SIGM2(I)=ALOG(R)*0.4343

ASM=0

950 SMG=SMG+SIGM2(I)

SMG=SMG/M

1150 R1=0

R2=0

DO 960 I=1,N

R1=R1+(XM(I)-AMG)*(SIGM2(I)-SMG)

960 R2=R2+(XM(I)-AMG)**2

WRITE(3,966) R3

966 FORMAT(40X,'VALOR DO COEFICIENTE',F10.5/)

R0=SMG-R3*AMG

SQRE=0

DO 965 I=1,N

965 SQRE=SQRE+(SIGM2(I)-R0-R3*XM(I))**2

OMRE=SQRE/(N-2)---

R4=OMRE/R2

TCAL=R3/SQRT(R4)

WRITE(3,988)TCAL,TAB

988 FORMAT(40X,'VALOR DE T CALCULADO',F10.6/40X,'VALOR TABELADO',F10.6
1/)

IF(TCAL-TAB)999,999,975

999 WRITE(3,9000)

9000 FORMAT(20X,'NAO EXISTE A TRANSFORMACAO NO PROGRAMA',/20X,'ESTUDE A
1TRANSFORMACAO APROPRIADA',/20X,'FAÇA A TRANSFORMACAO A PARTE',/20X,

PAGE 2

```

1)ENTRE COM O NUMERO ZERO NO CARTAO - ITRAN'/)
  IFLAG=2
  GO TO 992
975 IR=IFIX(R3+0.5)
  IF(IR-4)8001,8001,999
8001 IF(IR-0)999,8002,979
8002 WRITE(3,8003)
8003 FORMAT(20X,'PELO VALOR DE B NAO E NECESSARIO A TRANSFORMACAO'/)
  GO TO 992
979 IF(JDEL-4)2222,2223,2222
2222 GO TO (980,981,982,983),IR
980 DO 991 I=1,N
  DO 991 J=1,M
  TY(I,J)=SQRT(Y(I,J))
991 Y(I,J)=TY(I,J)
  GO TO 992
981 DO 993 I=1,N
  DO 993 J=1,M
  T1=Y(I,J)
  TY(I,J)=ALOG(T1)*0.4343
993 Y(I,J)=TY(I,J)
  GO TO 992
982 DO 994 I=1,N
  DO 994 J=1,M
  TY(I,J)=1/SQRT(Y(I,J))
994 Y(I,J)=TY(I,J)
  GO TO 992
983 DO 995 I=1,N
  DO 995 J=1,M
  TY(I,J)=1/Y(I,J)
995 Y(I,J)=TY(I,J)

930 AMED=0
  ANG=0
  DO 1100 K=1,N
  DO 1101 I=1,M
  DO 1105 J=1,M
  IF(N1(I,J)-K)1105,1115,1105
1105 CONTINUE
1115 AMED=AMED+Y(I,J)
1101 CONTINUE
  XMED(K)=AMED/N
  A=XMED(K)
  XM(K)=ALOG(A)*0.4343
  AMED=0
1100 ANG=ANG+XM(K)
  ANG=ANG/N
  ASM=0
  DO 1120 K=1,N
  DO 1125 I=1,M
  DO 1130 J=1,M
  IF(N1(I,J)-K)1130,1135,1130
1130 CONTINUE
1135 ASM=ASM+Y(I,J)**2
1125 CONTINUE
  SIGM2(K)=(ASM-N*(XMED(I)**2))/(N-1)
  R=SIGM2(K)
  SIGY2(K)=ALOG(R)*0.4343
  ASM=0

```

PAGE 3

```

1120 SMG=SMG+SIGM2(K)
      SMG=SMG/N
      GO TO 1150
935  AMED=0
      L1=N*M
      DO 1155 IK=1,L1
      DO 1160 K=1,L
      DO 1165 I=1,N
      DO 1165 J=1,M
      IF(N2(K,I,J)-IK)1165,1170,1165
1165 CONTINUE
1170 AMED=AMED+Y1(K,I,J)
1160 CONTINUE
      XMED(IK)=AMED/L
      A=XMED(IK)
      XM(IK)=ALOG(A)*0.4343
      AMED=0
1155 AMG=ANG+XM(IK)
      AMG=ANG/L1
      SMG=0
      ASM=0
      DO 1175 IK=1,L1
      DO 1180 K=1,L
      DO 1185 I=1,N
      DO 1185 J=1,M
      IF(N2(K,I,J)-IK)1185,1190,1185
1185 CONTINUE
1190 ASM=ASM+Y1(K,I,J)**2
1180 CONTINUE
      SIGM2(IK)=(ASM-L*(XMED(IK)**2))/(L-1)
      B=SIGM2(IK)
      SIGM2(IK)=ALOG(B)*0.4343
      ASM=0
1175 SMG=SMG+SIGM2(IK)
      SMG=SMG/L
      N=L1
      GO TO 1150
2223 GO TO(2224,2225,2226,2227),IB
2224 DO 2228 K=1,L
      DO 2228 I=1,N
      DO 2228 J=1,M
      TY1(K,I,J)=SORT(Y1(K,I,J))
2228 Y1(K,I,J)=TY1(K,I,J)
      GO TO 992
2225 DO 2229 K=1,L
      DO 2229 I=1,N
      DO 2229 J=1,M
      T1=Y1(K,I,J)
      TY1(K,I,J)=ALOG(T1)*0.4343
2229 Y1(K,I,J)=TY1(K,I,J)

2226 DO 3000 K=1,L
      DO 3000 I=1,N
      DO 3000 J=1,M
      TY1(K,I,J)=1/SORT(Y1(K,I,J))
3000 Y1(K,I,J)=TY1(K,I,J)
      GO TO 992
2227 DO 3005 K=1,L
      DO 3005 I=1,N

```


PAGE 4

```
DO 3005 J=1,M
  TY1(K,I,J)=1/Y1(K,I,J)
3005 Y1(K,I,J)=TY1(K,I,J)
992 RETURN
END
```

FEATURES SUPPORTED
ONE WORD INTEGERS

CODE REQUIREMENTS FOR TRANR
COMMON 0 VARIABLES 1844 PROGRAM 1802

RELATIVE ENTRY POINT ADDRESS IS 07D6 (HEX)

END OF COMPILATION

// DUP

```
*DELETE          TRANR
CART ID 0002    DB ADDR 54FA    DB CNT 0078
```

```
*STORE          WS UA TRANR
CART ID 0002    DB ADDR 5D04    DB CNT 0078
```

```

PAGE 1
// JOB
LOG DRIVE    CAPT SPEC    CART AVAIL    PHY DRIVE
   0000         0002         0002         0000

V2 M11  ACTUAL 16K  CONFIG 16K

// FOR
*ONE WORD INTEGERS
*LIST SOURCE PROGRAM
  SUBROUTINE LILLI(JDEL,N,Y,L,RES,RES1,RNORM,RNOR1,YEST,YEST1,DTAB,Y
  1ORD,RNORD,K,P,DSUP,S)
DIMENSION YORD(80),RNORD(80),F1(80),S(80),D(80),P(80)
A1=C.4361836
A2=-0.1201676
A3=C.9372980
DO 50 I=1,K
X=RNORD(I)
IF(X-C)60,65,65
  60 X=ABS(X)
  T=1/(1+0.3326*X)
  F2=C.393942*EXP(-0.5*X**2)
  F1(I)=F2*(A1*T+A2*(T**2)+A3*(T**3))
  GO TO 50

  F2=C.393942*EXP(-0.5*X**2)
  F1(I)=1-F2*(A1*T+A2*(T**2)+A3*(T**3))
  50 CONTINUE
  D(I)=ABS(F1(I)-S(I))
  DO 80 I=2,K
  J=I-1
  D1=ABS(F1(I)-S(J))
  D2=ABS(F1(I)-S(I))
  IF(D1-D2)92,100,100
  100 D(I)=D1

  92 D(I)=D2
  80 CONTINUE
  WRITE(3,355)
  WRITE(3,357)
  357 FORMAT(15X,'Y ESTIMADO',5X,'ERRO PADRAO',5X,'VALOR DE F',5X,'VALOR
  1R DE S',5X,'F(X)-S(X)',/)
  WRITE(3,355)
  DO 76 I=1,K
  WRITE(3,77)YORD(I),RNORD(I),F1(I),S(I),D(I)
  77 FORMAT(15X,F12.7,5X,F10.7,5X,F10.7,5X,F10.7,5X,F10.7)
  78 CONTINUE
  WRITE(3,355)
  WRITE(3,1000)
  1000 FORMAT('1')
  K1=K-1
  DO 150 I=1,K1
  K2=I+1
  DO 200 J=K2,K
  IF(D(I)-D(J))200,200,300
  200 A=D(I)
  D(I)=D(J)
  D(J)=A
  200 CONTINUE

```

PAGE 2

```

150 CONTINUE
    DSUP=D(K)
    WRITE(3,355)
355 FORMAT(20X,30('*'))/
    WRITE(3,350)DSUP,DTAP
350 FORMAT(10X,'TESTE DE LILLIFORS',2X,F10.4,///40X,'VALOR TABELADO',2X
1,F10.4)
    WRITE(3,355)
    WRITE(3,1000)
    WRITE(3,400)
400 FORMAT(20X,'ERRO PADRAO',10X,'VALOR DA FUN. ACUM.//')
    WRITE(3,355)
    DO 450 I=1,K
    WRITE(3,500)PNORD(I),P(I)
500 FORMAT(20X,F12.6,15X,F10.5)
450 CONTINUE
    WRITE(3,355)
    RETURN
    END

```

FEATURES SUPPORTED
ONE WORD INTEGERS

CORE REQUIREMENTS FOR LILLI
COMMON C VARIABLES 352 PROGRAM 692

RELATIVE ENTRY POINT ADDRESS IS 01E2 (HEX)

END OF COMPILATION

// DUP

```

*DELETE          LILLI
CART ID 0002    DB ADDR 54FA    DB CNT  0032

```

```

*STORE          WS UA LILLI
CART ID 0002    DB ADDR 503C    DB CNT  0032

```

PAGE 1

// JOP

LOG DRIVE CART SPEC CART AVAIL PHY DRIVE
 0000 0002 0002 0000

V2 V11 ACTUAL 16K CONFIG 16K

// EQP

*IOCS(CARD,1132PRINTER)

*ONE WORD INTEGERS

*LIST SOURCE PROGRAM

```

C    PROGRAMA DE ANALISE DE RESIDUO
C    O PROGRAMA PRINCIPAL DECOMPOE O RESIDUO DETERMINA O TIPO E O
C    GRAU DE LIBERDADE DO RESIDUO. CHAMA AS SUBROTINAS DE TESTE DE
C    MAG ADITIVIDADE TESTE DE HOMOGENEIDADE DE VARIANCIA - TESTE DE
C    COCHRAN, TESTE DE LILLIFFORS E LIBERACAO DOS RESIDUOS
C    PARA ENTRADA DOS DADOS
C    O PRIMEIRO CARTAO - NUMERO DE ANALISES A FAZER
C    O SEGUNDO CARTAO SE DESEJA TRANSFORMACAO
C    SE DESEJA TRANSFORMACAO, ENTRE COM O NUMERO 1(UM)
C    SE NAO DESEJA, ENTRE COM O NUMERO 0(ZERO)
C    O TERCEIRO CARTAO, PRIMEIRA POSICAO NUMERO DE DELINEAMENTO
C    SEGUNDA POSICAO VALOR TARELADO PARA TESTE DE COCHRAN
C    TERCEIRA POSICAO VALOR TAR. PARA TESTE DE MAG ADITIVIDADE
C    QUARTA POSICAO VALOR TAR. DE T PARA RES.
C    QUINTA POSICAO VALOR TAR. PARA TESTE DE LILLIFFORS
C    OS NUMEROS DOS DELINEAMENTOS 11 - INTENSIDADE AD ACASO
C    12 - BLOCOS CASUALIZADOS, 13 - QUADRADO LATINO,
C    14 - LATICE.
C    O QUARTO CARTAO O TAMAHO DO DELINEAMENTO
C    INTEIRAMENTE AD ACASO - PRIMEIRO TRAT. DEPOIS TRATAMEN
C    BLOCOS CASUALIZADOS - PRIMEIRO BLOCOS, DEPOIS TRAT.
C    QUADRADO LATINO - QUALQUER TIPO DE ENTRADA
C    LATICE - TRATAMENTOS POR LINHA, BLOCOS DENTRO DE CADA SER.
C    E O NUMERO DE REPETICOES.
C    DIMENSIONY(25,10),XMT(25),RES(25,10),YEST(25,10),RNOV(25,10),YMM
C    1(25),XVL(5),XVC(5),X1(5,5),Y1(5,5),X2(5,5),Y2(5,5),X3(5,5),
C    Y3(5,5),RES1(5,5,5),RNOV1(5,5,5),XORD(50),YORD(50),E(50),XAC
C    00AD(2,5)MDEL
C    5    FORMAT(12)
C        IDEL=1
C    3008    READ(2,22)ITRAN
C        22    FORMAT(12)
C            READ(2,4)JDEL,CTAR,FCAL,TAR,RTAR
C            4    FORMAT(12,4F8,5)
C            GO TO(2),25,30,40),JDEL
C    1    LEITURA DOS DADOS
C        20    READ(2,45)M,M
C        25    FORMAT(2I2)
C        M=NUMERO DE TRATAMENTOS,M=NUMERO DE REPETICOES
C        WRITE(3,6)
C            6    FORMAT(50X,1A MATRIZ DO DELINEAMENTO//)
C            DO 30 I=1,M
C        30    READ(2,55)(Y(I,J),J=1,Y)
C        35    FORMAT(8F10,4)
C            DO 51 I=1,Y
C        51    WRITE(3,52)(Y(I,J),J=1,Y)
C        52    FORMAT(20X,8(F10,4,2X))
C    2    DECOMPOSICAO DOS RESIDUOS

```

PAGE 2

```

7 X1=0.
  DO 100 I=1,M
  DO 105 J=1,N
105 X1=X1+Y(I,J)
  XMT(I)=X1/M
  X1=0
100 CONTINUE
  SCRES=0
  DO 110 I=1,N
  DO 115 J=1,M
  YEST(I,J)=XMT(I)
  RES(I,J)=Y(I,J)-YEST(I,J)
115 SCRES=SCRES+RES(I,J)**2
110 CONTINUE
  GLRES=N*(M-1)
  QMRES=SCRES/GLRES
  DO 116 I=1,N
  DO 116 J=1,M
  RNOPI(I,J)=RES(I,J)/SQRT(QMRES)
116 CONTINUE
  GO TO 5555
25 READ(2,26)M,N
26 FORMAT(2I2)
  WRITE(3,1000)
1000 FORMAT('11')
  WRITE(3,21)
21 FORMAT(20X,50('1*1'))
5003 FORMAT(1RX,'A MATRIZ DE DEFLINEAMENTO')
  DO 27 I=1,M
27 READ(2,28)((Y(I,J),J=1,N)
28 FORMAT(2F10.4)
  DO 14 I=1,M
14 WRITE(3,11)((Y(I,J),J=1,N)
11 FORMAT(8X,9(F10.4,2X))
C DECOMPOSICAO DOS RESIDUOS
8 X2=0.
  X3=0
  XMG=0
  DO 120 I=1,M
  DO 125 J=1,N
125 X2=X2+Y(I,J)
  XMT(I)=X2/M
  X2=0
120 CONTINUE
  DO 130 J=1,N
  DO 135 I=1,M
  X3=X3+Y(I,J)
135 XMG=XMG+Y(I,J)
  XMR(J)=X3/M
  X3=0.
130 CONTINUE
  XMG=XMG/(M*N)
C OBTENCAO DE Y EST, RES E RES NORMALIZADO
  DO 136 I=1,M
  DO 140 J=1,N
  YEST(I,J)=XMT(I)+XMR(J)-XMG
  RES(I,J)=Y(I,J)-YEST(I,J)

```

```

PAGE      3

140 SQRES=SQRES+RES(I,J)**2
136 CONTINUE
      GLRES=(I-1)*(M-1)
      QMPRES=SQRES/GLRES
      WRITE(3,5005)QMPRES
5005 FORMAT(40X,'QMPRESIDUO',5X,F16.9)
      DO 137 I=1,M
      DO 138 J=1,N
138 RNOPI(I,J)=RES(I,J)/SQRT(QMPRES)
137 CONTINUE
      GO TO 5555
C      CUADRADO LATINO
30 READ(2,60)M,N
60 FORMAT(2I2)
      DO 65 I=1,N
      READ(2,75)(M1(I,J),Y(I,J),J=1,M)
75 FORMAT(5(I2,F6.3))
65 CONTINUE
C      DECOMPOSICAO DE RESIDUO
9 X4=0.
  X5=0
  X6=0
  XMG=0
      DO 150 I=1,N
      DO 155 J=1,M
155 X4=X4+Y(I,J)
      X4L(I)=X4/M
      X4=0
150 CONTINUE
      DO 200 J=1,M
      DO 205 I=1,N
205 X5=X5+Y(I,J)
      X5C(J)=X5/M
      X5=0
200 CONTINUE
      DO 225 I=1,M
      DO 230 J=1,N
230 XMG=XMG+Y(I,J)
225 CONTINUE
      XMG=XMG/(M*N)
      X4=0
      DO 210 K=1,M
      DO 215 I=1,N
      DO 220 J=1,N
      IF(M1(I,J)-K)220,221,220
220 CONTINUE
221 X6=X6+Y(I,J)
215 CONTINUE
      X4T(K)=X6/M
210 X6=0
C      CALCULO DE YEST, RES E RESNORM
SQRES=0
      DO 235 K=1,M
      DO 240 I=1,M
      DO 245 J=1,N
      IF(M1(I,J)-K)245,244,245
245 CONTINUE
246 YEST(I,J)=X4L(I)+X5C(J)+X4T(K)-(XMG*2)
      RES(I,J)=Y(I,J)-YEST(I,J)

```

PAGE 4

```

240 SORRES=SORRES+RES(I,J)**2
245 CONTINUE
      GLRES=(K-1)*(N-2)
      OVRRES=SORRES/GLRES
      DO 247 K=1,N
      DO 248 I=1,M
      DO 249 J=1,Y
      IF(I 1(I,J)-K)249,244,249
249 CONTINUE
244 ENORM(I,J)=RES(I,J)/SQRT(OVRRES)
249 CONTINUE
247 CONTINUE
      GO TO 5555
C     LATICIE QUADRADO
40 READ(2,175)M,N,L
775 FORMAT(3I2)
C     M=NUMERO DE LINHAS -TRATAMENTOS
C     N=NUMERO DE BLOCOS
C     L=NUMERO DE REPETICOES
C     LEITURA DOS SAPOS
      WRITE(3,21)
      WRITE(3,5006)
      DO 80 K=1,L
      DO 80 I=1,M
      READ(2,95)(H2(K,I,J),Y1(K,I,J),J=1,Y)
95 FORMAT(5(12,F6.3))
      WRITE(3,96)(H2(K,I,J),Y1(K,I,J),J=1,M)
96 FORMAT(20X,5(12,2X,F6.4,2X))
      DO CONTINUE
      WRITE(3,21)
10 X0=0.
C     DECOMPOSICAO DOS RESIDUOS

      DO 241 IK=1,K3
      DO 242 K=1,L
      DO 250 I=1,M
      DO 250 J=1,Y
      IF(I 2(K,I,J)-IK)250,251,250
250 CONTINUE
251 X0=X0+Y1(K,I,J)
242 CONTINUE
      XMT(IK)=X0/L
      X0=0
241 CONTINUE
C     CALCULO DA MEDIA DO BLOCO J NA REP K
      X10=0
      DO 255 K=1,L
      DO 260 I=1,M
      DO 265 J=1,Y
265 X10=X10+Y1(K,I,J)
      XMR(K,I)=X10/M
      X10=0
260 CONTINUE
255 CONTINUE
C     CALCULO DA MEDIA GERAL
      XUG=0
      DO 270 K=1,L
      DO 275 I=1,M
      DO 280 J=1,Y

```

PAGE 5

```

260 XMG=XMG+Y1(K,I,J)
275 CONTINUE
270 CONTINUE
XMG=XMG/(L*M*N)
SQRES=0
C CALCULO DE YEST,RESIDUO,RNORM,OMRES
K9=M*N
DO 285 IK=1,K9
DO 290 K=1,L
DO 295 I=1,M
DO 295 J=1,N
IF(N2(K,I,J)-PK)295,296,295
295 CONTINUE
296 YEST1(K,I,J)=XMT(IK)+XMR9(K,I)-XMG
RES1(K,I,J)=Y1(K,I,J)-YEST1(K,I,J)
298 SQRES=SQRES+RES1(K,I,J)**2
285 CONTINUE
GLRES=L*N*(M-1)-((M**2)-1)
OMRES=SQRES/GLRES
K9=M*N
DO 299 IK=1,K9
DO 297 K=1,L
DO 299 I=1,M
DO 299 J=1,N
IF(N2(K,I,J)-IK)299,301,299
299 CONTINUE
301 RNOR1(K,I,J)=RES1(K,I,J)/SQRT(OMRES)
297 CONTINUE
299 CONTINUE
5555 IF(ITRAN-1)5556,5557,5556
5556 CALL TEVON(JDEL,M,N,L,RES,Y1,M1,N2,COCAL,CTAR)
WRITE(3,6001)
6001 FORMAT(10X,60(1-1))
IF(JDEL-1)490,485,490
490 CALL TESNO(JDEL,M,N,L,YEST,RES,SQRES,GLRES,M1,N2,KES1,YEST1,FSTE),
WRITE(3,6003)FSTE,FCAL
6003 FORMAT(30X,1FSTE DE MAC ANITIVIDADE),2X,F10.5,/,30X,'VALOR TABELA
1001,2X,F10.4)
WRITE(3,6001)
WRITE(3,1000)
485 CALL ORDE(JDEL,M,N,L,RES,YEST1,RNORM,RNOR1,YEST,YEST1,DTAR,YORD,
*DO,K,P,NSUP,S)
IDEL=IDEL+1
IF(IDEL=NDL)4001,4001,4004
4001 GO TO 3008
5557 CALL TRANR(JDEL,M,N,L,Y,Y1,M1,N2,TAR,IFLAG)
IF(IFLAG-2)3044,4005,3044
3044 ITRAN=0
GO TO(7,8,9,10),JDEL
4005 IDEL=IDEL+1
IF(IDEL=NDL)3008,3008,4004
4004 STOP
END

```

FEATURES SUPPORTED
ONE WORD INTEGERS
IOCS

CORE REQUIREMENTS FOR
COMMON 0 VARIABLES 4022 PROGRAM 21RC

APÊNDICE 4 - PROGRAMA ROSAT PARA MODIFICAÇÃO DE "OUTLIERS" ELABORADO PARA COMPUTADOR IBM 1130

```

*LIST SOURCE PROGRAM
SUBROUTINE RESID(N,M,Y,YEST,RNORM,QVRES,RES)
  DIMENSION Y(25,10),YEST(25,10),RNORM(25,10),XMT(25),XMB(25),RES(25
  1,10)
C   DECOMPOSICAO DOS RESIDUOS
C   MEDIA TRAT,MEDIA BLOCO,MEDIA GERAL
  X2=0.
  X3=0
  XMG=0
  DO 120 I=1,N
  DO 125 J=1,M
125  X2=X2+Y(I,J)
  XMT(I)=X2/M
  X2=0
120  CONTINUE
  DO 130 J=1,M
  DO 135 I=1,N
  X3=X3+Y(I,J)
135  XMG=XMG+Y(I,J)
  XMB(J)=X3/M
  X3=0
130  CONTINUE

  SQRES=0
C   OBTENCAO DE Y EST,RES E RES NORMALIZADO

  DO 140 J=1,M
  YEST(I,J)=XMT(I)+XMB(J)-XMG
  RES(I,J)=Y(I,J)-YEST(I,J)
140  SQRES=SQRES+RES(I,J)**2
126  CONTINUE
  GLRES=(N-1)*(M-1)
  QVRES=SQRES/GLRES
  WRITE(3,5005)QVRES
5005  FORMAT(40X,'QVRESIDUO',F16.9)

  DO 138 J=1,M
138  RNORM(I,J)=RES(I,J)/SQRT(QVRES)
137  CONTINUE
  RETURN
  END

```

FEATURES SUPPORTED
ONE WORD INTEGERS

CORE REQUIREMENTS FOR RESID
COMMON C VARIABLES 118 PROGRAM 350

// DUP

*DELETE RESID
CARD ID 0002 DB ADDR 5A7F DB CNT 0018

*STORE WS UA RESID
CARD ID 0002 DB ADDR 5D55 DB CNT 0018

PAGE 1

// JOB

LOG DRIVE	CART SPEC	CART AVAIL	PHY DRIVE
0000	0002	0002	0000

V2 M11 ACTUAL 16K CONFIG 16K

// FOR

*IOCS(CARD,1132PRINTER)

*ONE WORD INTEGERS

*LIST SOURCE PROGRAM

```

C     ESTE PROGRAMA FAZ MODIFICACAO DE OUTLIERS
C     O PROGRAMA TEM UMA SUBROUTINE PARA DECOMP.DOS RESIDUOS
C     O DELINEAMENTO TEM DE SER BLOCOS AO ACASO
C     PRIMEIRO CARTAO E NUMERO DE BLOCOS E NUM.DE TRATAMENTOS
C     SEGUNDO CARTAO E VALOR DE CONVERGENCIA,GPAUS DE LIBERDADE
C     PARA ENTRADA DE DADOS
C     TERCEIRO CARTAO OS DADOS DO DELINEAMENTO
C     O PROGRAMA TRANSFORMA OS RESIDUOS, PARA EM SEGUIDA CALCULAR
C     UM NOVO VALOR DE Y, TRANSFORMANDO O OUTLIERS.
C     DIMENSION Y(25,10),YEST(25,10),RNORM(25,10),WEST(25,10),RES(25,10)
C     LEITURA DOS DADOS
      READ(2,45)M,N
45  FORMAT(2I2)
      READ(2,5)VC,VK
      5  FORMAT(F8.6,F5.3)
      WRITE(3,21)
21  FORMAT(20X,60('*'))/
      WRITE(2,5003)
5003 FORMAT(18X,'A MATRIZ DO DELINEAMENTO')/
      DO 27 I=1,N
27  READ(2,29)(Y(I,J),J=1,M)
29  FORMAT(8F10.4)
      DO 14 I=1,N
14  WRITE(3,11)(Y(I,J),J=1,M)
11  FORMAT(8X,8(F10.4,2X))
      DO 22 I=1,N
      DO 22 J=1,M
      WEST(I,J)=0.
22  CONTINUE
      CALL RESID(N,M,Y,YEST,RNORM,QYRES,RES)
      DO 20 I=1,N
      DO 20 J=1,M
      DIF=ABS(YEST(I,J)-WEST(I,J))
      IF(DIF-VC)20,30,30
20  CONTINUE
      GO TO 210
30  DO 40 I=1,N
      DO 40 J=1,M
40  WEST(I,J)=YEST(I,J)
      DO 80 I=1,N
      DO 80 J=1,M
80  RNORM(I,J)=((VK+1)*RNORM(I,J)/(VK+RNORM(I,J)**2))
      DO 90 I=1,N
      DO 90 J=1,M
90  Y(I,J)=YEST(I,J)+SQRT(QYRES)*RNORM(I,J)
      A=QYRES
      CALL RESID(N,M,Y,YEST,RNORM,QYRES,RES)
      WRITE(3,17)

```

PAGE 2

```

17 FORMAT(20X, 'LA MATRIZ DE Y TRANSFORMADOS')
DO 15 I=1,N
WRITE(3,16)(Y(I,J),J=1,M)
16 FORMAT(20X,6(F10.5,2X))
15 CONTINUE
IF(OMPRES-A)8,8,210
210 STOP
END

```

FEATURES SUPPORTED
ONE WORD INTEGERS
IOCS

CORE REQUIREMENTS FOR
COMMON 0 VARIABLES 2518 PROGRAM 496

END OF COMPILATION

// DUP

```

*DELETE          ROSA1
CART ID 0002    DB ADDR 5A7F    DB CNT  0020

```

```

*STORE          WS  UA  ROSA1
CART ID 0002    DB ADDR 5D4D    DB CNT  0021

```

// XEQ ROSA1 1

*LOCAL ROSA1,RESID