INFLUÊNCIA DE TRANSFORMAÇÕES DE DADOS SOBRE A ESTIMAÇÃO DOS COMPONENTES DE VARIÂNCIA

Ivani Pozar Otsuk
(Engenheira-Agrônoma)

Orientador: Prof. Dr. Décio Barbin

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.

P I R A C I C A B A
Estado de São Paulo - Brasil
Fevereiro - 1991

Ficha catalográfica preparada pela Seção de Livros da
Divisão de Biblioteca e Documentação - PCAP/USP

088i Otsuk, Ivani Pozar
Influência de transformações de dados sobre a estim
mação dos componentes de variância. Piracicaba, 1991.
66p.

Diss.(Mestre) - ESALQ
Bibliografia.

1. Análise de variância 2. Dado estatístico - Trans
formação 3. Estatística aplicada 4. Genética - Estatís
tica 5. Método da máxima verossimilhança restrita 6. Mo
delo matemático I. Escola Superior de Agricultura Luiz
de Queiroz, Piracicaba

CDD 519.5

INFLUÊNCIA DE TRANSFORMAÇÕES DE DADOS SOBRE A
ESTIMAÇÃO DOS COMPONENTES DE VARIÂNCIA

Ivani Pozar Otsuk

Aprovada em: 05.04.1991

Comissão julgadora:

Prof. Dr. Décio Barbin

ESALQ/USP

Prof.^a Dr.^a Clarice G. B. Demétrio

ESALQ/USP

Prof. Dr. Sérgio do Nascimento Kronka

FCAV/UNESP

Prof. Dr. DÉCIO BARBIN
Orientador

Aos meus pais

André e Cecília

OFEREÇO

Ao meu esposo *Massami*

e aos meus filhos *Wesley e Wilson*

DEDICO

AGRADECIMENTOS

AO professor Dr. Décio Barbin pela orientação na realização do presente trabalho, pelo estímulo e amizade.

Ao professor José Eduardo Corrente pela orientação na área de simulação, pela sua dedicação e incentivo.

Aos professores do curso de pós-graduação em Estatística e Experimentação Agronômica da ESALQ/USP, pelos ensinamentos transmitidos.

Aos colegas do curso de pós-graduação, pela amizade e companheirismo.

Ao CNPq e à CAPES pelas bolsas de estudo concedidas.

Ao Instituto de Zootecnia pela oportunidade concedida para a realização deste trabalho.

Ao Dr. Benedicto do Espírito Santo de Campos chefe da Seção de Estatística e Técnica Experimental do Instituto de Zootecnia, pelo apoio concedido para a execução do trabalho.

Aos colegas da Seção de Estatística e Técnica Experimental, do Instituto de Zootecnia pelo apoio e amizade.

A todos que, direta ou indiretamente, colaboraram para a realização deste trabalho.

SUMÁRIO

	Página
RESUMO	iv
SUMMARY	vi
1. INTRODUÇÃO	1
2. REVISÃO DE LITERATURA	3
2.1. Componentes de variância	3
2.2. Transformação de dados	10
3. METODOLOGIA	24
3.1. Modelo aleatório	24
3.2. Estimação dos componentes de variância	27
4. RESULTADOS E DISCUSSÃO	33
5. CONCLUSÕES	60
6. REFERENCIAIS BIBLIOGRÁFICAS	61

INFLUÊNCIA DE TRANSFORMAÇÕES DE DADOS SOBRE A ESTIMAÇÃO DOS COMPONENTES DE VARIÂNCIA

Autora: Ivani Pozar Otsuk

Orientador: Prof. Dr. Décio Barbin

RESUMO

Os modelos de efeitos aleatórios têm sido usados com muita frequência na experimentação, principalmente no melhoramento genético (coeficiente de herdabilidade, índice de seleção, etc.). Em geral o uso de tais modelos leva a estimar os componentes de variância, cujos métodos são bastante discutidos na literatura.

O problema surge quando não se pode supor uma distribuição normal dos dados. Neste caso, é recomendado o uso de transformações que normalizem esses dados (ou pelo menos que se obtenha uma distribuição simétrica dos mesmos) e se tenha homogeneidade de variâncias. Essas transformações têm sido indicadas por vários autores dependendo da distribuição dos dados.

Assim, utilizando-se a proporção de variação explicada e feita uma simulação de um modelo inteiramente ao acaso com dados balanceados, analisou-se o efeito da transformação sobre as estimativas de

componentes de variância para os efeitos de tratamentos e resíduo supondo distribuição Binomial e Poisson.

O método utilizado para a estimação dos componentes de variância foi o de Máxima Verossimilhança Restrita e para gerar os dados usaram-se as subrotinas do Sistema SAS.

Verificou-se que, as proporções de variações explicadas não diferiram muito na presença de transformações, confirmando as indicações da literatura de que para se estimarem os componentes de variância, não é necessário que os dados sigam uma distribuição normal.

INFLUENCE OF DATA TRANSFORMATIONS ON ESTIMATION OF VARIANCE COMPONENTS

Author: Ivani Pozar Otsuk

Adviser: Prof. Dr. Décio Barbin

SUMMARY

Models with random effects are being used intensively in researches for genetic breeding (genetic parameters). Usually, application of these models is done to get the estimation of variance components, although the methods used in these estimations are get largely discussed in the literature.

The problem arises when data are not normally distributed. In this case, it is suggested the use of the data transformation that normalizes these data (or at least one which gives symmetric distribution of these data) and get homogeneous variance. These transformations are being indicated by several authors considering the data distributions.

Thus, the use of a proportion accounted for the explained variance and by simulation in a model for a completely randomized design, it was analysed the effect of the data transformation on the estimation of variance

components for treatment effects and it was supposed that data follow the Binomial and Poisson distributions.

The method used for estimation of variance components was the Restricted Maximum Likelihood and to get these data it was utilized the subroutines of SAS system.

It was verified that the proportion of explained variances demonstrated no great differences when the data were transformed, and it was confirmed the literature indication that for variance components estimation it is not necessary the normal distribution of the data.

1 - INTRODUÇÃO

Os componentes de variância, que são variâncias associadas aos efeitos aleatórios de um modelo matemático, têm grande importância e aplicação tanto no melhoramento vegetal quanto animal, (obtenção de índices de repetibilidade, coeficiente de herdabilidade, etc), bem como em outras áreas, tais como médica e industrial.

Existem hoje vários métodos de estimação de componentes de variância cuja eficiência ainda é bastante discutida, devido ao fato de nem sempre se obterem estimadores com propriedades ótimas (não viciado, variância mínima), em geral, quando se trata de modelos não balanceados.

Um outro problema que ainda surge é a distribuição amostral dos dados experimentais. Para se fazer a estimação de componentes de variância, não precisamos ter normalidade dos dados, mas para os testes de efeitos do modelo, essa suposição se faz necessária. Em tais casos, em geral, é preciso recorrer a uma transformação dos dados, tendo em vista o comportamento da média e da variância de cada tratamento.

BARBOSA (1983) fornece uma tabela na qual dependendo da relação existente entre média e variância, uma transformação é sugerida. Obtidos então, os dados transformados, que se supõe terem distribuição normal ou pelo menos simétrica, faz-se novamente a estimação dos componentes de variância.

O objetivo deste trabalho é verificar as possíveis alterações que as transformações de dados podem trazer às estimativas dos componentes de variância. Isso é feito analisando a proporção de variação explicada com os dados sem transformação e com a transformação usada na literatura.

2 - REVISÃO DE LITERATURA

2.1. Componentes de variância

A análise de variância tem sido largamente usada para testar a significância de efeitos de tratamentos, ou melhor, para testar a existência de diferenças entre médias de tratamentos e também estimar os componentes de variância.

Apenas a título de esclarecimento do termo componentes de variância, consideremos o modelo aleatório com dois fatores, dado por:

$$Y_{ij} = u + \alpha_i + \varepsilon_{ij}$$

com $i=1, \dots, I$ e $j=1, \dots, J$, onde:

Y_{ij} é um vetor de observações

u é um vetor de médias fixado

α_i um efeito aleatório de um fator α com

$$E(\alpha_i) = 0 \text{ e } \text{Var}(\alpha_i) = \sigma_\alpha^2 \text{ e } \text{Cov}(\alpha_i, \alpha_{i'}) = 0 \text{ (} i \neq i' \text{)},$$

ε_{ij} é um erro experimental com $E(\varepsilon_{ij}) = 0$ e $\text{Var}(\varepsilon_{ij}) = \sigma_\varepsilon^2$

e sendo os ε_{ij} independentes e normalmente distribuídos.

Deve-se ressaltar também que os efeitos aleatórios não são correlacionados.

Nota-se que a variância da observação

será dada por: $\text{Var}(Y_{ij}) = \sigma_{\alpha}^2 + \sigma_{\epsilon}^2$.

Então, σ_{α}^2 e σ_{ϵ}^2 são chamados componentes de variância de Y_{ij} . Portanto, componentes de variância são variâncias associadas aos efeitos aleatórios de um modelo matemático. Esses componentes são de grande importância e aplicação no campo genético, (índices de seleção, coeficientes de herdabilidade, etc...) e também na Estatística Experimental (para a determinação do denominador para o teste F na análise de variância).

Considerando esses componentes como parâmetros, existe a preocupação de se buscarem estimadores eficientes para eles. Em geral os métodos de estimação para componentes de variância diferem quando os dados são balanceados e não balanceados. Os vários métodos de estimação de componentes de variância, permitem ao pesquisador uma opção de escolha, embora a eficiência desses métodos seja ainda bastante discutida.

Um modo usual de estimar os componentes de variância consiste em igualar os quadrados médios observados (QM) da análise da variância aos seus valores esperados [E(QM)], substituindo-se nestas expressões, as variâncias por suas estimativas.

Este método é chamado de Método da Análise da Variância (Método da ANOVA). Estes estimadores são as soluções das equações lineares nos parâmetros.

Os estimadores dos componentes de variância obtidos por este método apresentam as propriedades de

serem não tendenciosos e de variância mínima.

Estas duas propriedades não necessitam da suposição de normalidade para os elementos aleatórios do modelo, apenas que eles tenham média e covariância iguais a zero e variâncias finitas. Isto porque os valores esperados dos quadrados médios da ANOVA não usam suposições de normalidade, e assim os estimadores dos componentes de variância obtidos por este método não vão depender de exigências de normalidade (SEARLE, 1971).

Entretanto, sob a suposição de normalidade, os estimadores dos componentes de variância obtidos pelo método da ANOVA em dados balanceados, são baseados em estatísticas suficientes minimais, resultado encontrado em GRAYBILL & HULTGUIST (1961).

SINGHAL (1986) estudando o efeito de normalidade sobre o valor médio do erro de certos estimadores truncados, na comparação com estimadores da Análise da Variância, conclui que o valor médio de estimadores truncados e o problema de obter estimativas negativas não são afetados por parâmetros de não normalidade, somente a variância dos estimadores é deslocada notavelmente na direção e magnitude da curtose.

Apesar das facilidades de cálculo, e de algumas propriedades ótimas, os estimadores de componentes de variância pelo método da ANOVA em dados balanceados, apresentam um inconveniente que é o de produzir estimativas negativas e quando aplicado a dados não

balanceados é equivalente ao Método 1 de Henderson. Se os dados não são balanceados e o modelo é misto, os estimadores por ele obtidos são tendenciosos (SEARLE, 1971).

Outros métodos podem ser citados: Métodos 1, 2 e 3 de Henderson, Máxima Verossimilhança (MV), Máxima Verossimilhança Restrita (MVR), Estimadores Quadráticos Não Tendenciosos de Norma Mínima (Minimum Norm Quadratic Unbiased Estimator - MINQUE), Estimadores Quadráticos Não Tendenciosos de Variância Mínima (Minimum Variance Quadratic Unbiased Estimator - MIVQUE), além de outros.

Segundo HENDERSON (1953), o Método 1 baseia-se no cálculo das somas de quadrados não corrigidas para em seguida equacioná-las a seus valores esperados, como na análise da variância para dados balanceados. Por este motivo é também chamado de Método da Análise da Variância para estimar componentes de variância. Este método quando usado para modelos aleatórios produz estimadores não viciados para todos os componentes de variância o que não é válido quando usado em modelos mistos. (SEARLE, 1968, 1971).

O Método 2 de Henderson (HENDERSON, 1953) surgiu com o objetivo de contornar o problema de tendenciosidade dos estimadores dos componentes de variância em modelos mistos. Este método consiste inicialmente em estimar os efeitos fixos do modelo, para em seguida ajustar os dados, e finalmente aplicar o Método

1 de Henderson aos dados ajustados. Os estimadores obtidos por este método não são viciados, mas, não é definido de maneira única, pois envolve uma inversa generalizada e também não pode ser usado quando há interação entre os efeitos fixos e aleatórios, (SEARLE, 1968, 1971).

O Método 3 de Henderson, também chamado de ajuste de constante, desenvolvido por HENDERSON (1953), apropriado para modelos mistos, produz estimadores não tendenciosos dos componentes de variância e é o método mais indicado para modelos desse tipo, pois não faz restrições quanto às interações entre os efeitos fixos e aleatórios.

Como característica comum aos três métodos de Henderson tem-se que, quando usados em dados balanceados, reduzem-se ao método da ANOVA, e produzem estimativas negativas dos componentes de variância.

Na tentativa de contornar o problema de estimativas negativas, surgiu, talvez, como primeiro trabalho, o de HERBACH (1959), que usa os estimadores de Máxima Verossimilhança para modelos aleatórios simples em dados balanceados. Esses estimadores apresentam como vantagem o fato de levarem a estimativas não negativas e como desvantagem o fato de serem truncados e viciados.

THOMPSON (1962) também na tentativa de resolver o problema de estimativas negativas, desenvolveu um algoritmo simples, chamado "pool-the-minimum violator", que consiste na maximização da função de verossimilhança

dos quadrados médios, sujeita a uma série de restrições, (princípio da máxima verossimilhança restrita).

Como alternativa para estimação pelo Método da Máxima Verossimilhança, CORBEIL & SEARLE (1976a) usaram a estimação de Máxima Verossimilhança Restrita, baseada na transformação sugerida por PATTERSON & THOMPSON (1971) onde dividiram a função verossimilhança, sob condições de normalidade em duas partes, de modo que a maximização da parte livre dos efeitos fixos forneça os estimadores de Máxima Verossimilhança Restrita para os componentes de variância e a outra forneça os estimadores para os efeitos fixos.

CORBEIL & SEARLE (1976b) mostram que os estimadores de Máxima Verossimilhança Restrita para dados balanceados são idênticos aos estimadores de análise de variância, mas se os dados não são balanceados, isso não ocorre.

HICKS (1964) desenvolveu um outro método conhecido como Método Prático de Hicks para obtenção de esperança de quadrado médio ECQM. Este método é usado para modelos fixos e aleatórios e para todos os delineamentos, mas, também apresenta como desvantagem o de produzir estimativas negativas.

RAO (1970, 1971a, 1971b, 1972) desenvolveu dois métodos de estimação denominados Estimadores Quadráticos Não Tendenciosos de Norma Mínima (MINQUE - Minimum Norm Quadratic Unbiased Estimator) e Estimadores

Quadráticos Não Tendenciosos de Variância Mínima (MIVQUE - Minimum Variance Quadratic Unbiased Estimator) que podem produzir estimativas negativas e, são equivalentes, quando aplicados a dados normalmente distribuídos, uma vez que estes têm coeficientes de curtose zero.

SEARLE (1971) enumera vários procedimentos alternativos, dentre outros, quando encontramos uma estimativa negativa para um componente de variância:

I) Aceitar a estimativa negativa, admitindo que o verdadeiro valor do componente de variância é zero. Isto pode parecer lógico, mas como parte da estimação foi truncada, isto perturba as propriedades dos estimadores.

II) A estimativa negativa pode ser tomada como evidência de que o verdadeiro valor do componente é zero. Isto não é satisfatório quando queremos utilizar uma estimativa da soma de componentes de variância, um dos quais tendo estimativa negativa. A estimativa da soma incluindo este componente é menor que a soma sem ele.

III) Interpretação da estimativa negativa como indicação de um modelo errôneo

ANDERSON (1965) sugere que quando ocorrerem estimativas negativas, é porque há possibilidade de haver problemas nos cálculos ou na especificação do modelo, portanto, primeiramente o experimentador deve conferir os cálculos realizados, e se esses cálculos estiverem corretos, ele recomenda a reconsideração do modelo, mas, se mesmo assim permanecer o problema da estimativa

negativa, sugere considerar o valor do componente como sendo zero.

Em geral, os dados obtidos nem sempre possuem uma distribuição normal. Isso, como já citado, não interfere na estimação de componentes de variância. Ocorre que, para continuar a análise, isto é, utilizar testes de hipóteses, é preciso que a normalidade ou pelo menos que a simetria da distribuição dos dados esteja presente e também que haja homogeneidade de variâncias.

Diante desse fato, surgiram muitos trabalhos aplicados em várias áreas, onde os autores sugerem o uso de uma transformação de dados.

2.2 Transformação de dados

BARTLETT (1936) afirmou que a análise da variância não pode ser usada sem consideração cuidadosa sobre a variação que ocorreu realmente com os dados. Portanto, para estabilizar a variância, quando se têm dados em que o desvio padrão é proporcional à média, pode-se utilizar a transformação logarítmica, enquanto que, quando se tem a variância proporcional à média, a transformação raiz quadrada.

Segundo BARTLETT (1947) os requisitos para uma transformação ideal são:

a) a variância da variável transformada não deve depender da média dessa variável;

b) a média aritmética da variável transformada deve ser normalmente distribuída;

c) a média aritmética da variável transformada deve ser uma estimativa da média verdadeira;

d) os efeitos reais devem ser lineares e aditivos.

ANSCOMBE (1948) considerou a transformação $\sqrt{X + K}$ onde X tem distribuição de Poisson com média $\underline{m}$. Se $\underline{m}$ for grande e $K=3/8$, o valor transformado terá variância aproximadamente igual a $1/4$, resultado devido a A. H. L. Johnson. Considerou ainda que quando X tem distribuição Binomial deve-se usar a transformação

$$\text{arc sen } \sqrt{\frac{X + K}{n + 2K}}$$

onde $\underline{n}$ é igual ao número total de dados. Para $K=3/8$, $\underline{m}$ e $(\underline{n} - \underline{m})$ grandes, a variância é aproximadamente constante e igual a

$$\frac{1}{4} \left(n + \frac{1}{2} \right)^{-1}$$

Para o caso de X ter distribuição Binomial Negativa com média $\underline{m}$ e expoente $\underline{r}$ ($\underline{m}, \underline{r} > 0$), a transformação adequada

$$\text{arc senh } \sqrt{(X+K)/(r-2K)}$$

(para $\underline{r} > 2$ e $\underline{m}$ grande) ou

$$\ln \left(X + \frac{1}{2} \underline{r} \right)$$

para $\underline{r} > 1$ e $\underline{m}$ grande. Considerou também os casos em que $\underline{m}$ é um valor pequeno.

Segundo BENNETT & FRANKLIN (1954), às vezes, em uma grande classe de dados os efeitos não são aditivos sob qualquer transformação, de tal modo que o uso

de uma transformação feita para estabilizar a variância pode não fazer efeito neste caso. Uma transformação poderá ser usada em casos onde a variância está relacionada como valor da média $\bar{m}$ dos dados. A natureza dessa relação pode ser conhecida por métodos gráficos, plotando-se s_m^2 contra $\log \bar{X}_{ij}$.

Se: $\sigma_m^2 = f(m)$, então, para qualquer transformação $\phi(m)$ temos aproximadamente

$$\sigma_{\phi}^2 = (d\phi / dm)^2 f(m)$$

ou, se σ_{ϕ}^2 é para ter valor constante K^2 ,

$$\phi = \int \frac{K \, d_m}{\sqrt{f(m)}}$$

Assim, dependendo da expressão $f(m)$ seguem-se algumas transformações mais comuns:

a) $f(m) = m$. Uma relação desse tipo ocorre com dados que têm uma distribuição de Poisson. A transformação apropriada é $\phi = \sqrt{X}$ se os dados apresentarem valores altos, caso contrário a transformação $\phi = \sqrt{X + 1/2}$ tem sido proposta. Qualquer das duas transformações fornece variância estável;

b) $f(m) = m(1-m)$. Esta é uma relação apropriada para dados de contagem distribuídos na forma Binomial e conduz à transformação $\phi(m) = \arcsin \sqrt{m}$. Neste caso a correção de continuidade pode ser feita substituindo $\bar{m}$ por $(m+1/2)$.

c) $f(m) = m^2$. Esta corresponde ao caso de um coeficiente de variação constante, e a transformação

apropriada é :

$$\phi(m) = \log(m).$$

Às vezes, a relação entre a média e a variância pode ser melhor descrita por:

$$f(m) = a + bm + cm^2$$

de tal modo que quando a extensão da variação de m em um experimento particular é limitada a valores grandes, a transformação $\log(m)$ pode ser apropriada. Alternativas como $\log(m+1)$ ou $\log(m+c)$, $0 < c < 1$ têm sido propostas como melhores aproximações para tais casos.

Segundo COCHRAN (1947), os fatores que causam mais distúrbios na análise da variância são: assimetria extrema, presença de erros grosseiros, anomalia de certos tratamentos, mudanças na variância residual e falta de aditividade dos efeitos reais sendo que os principais métodos para contornar essas dificuldades seriam: omissão de certos tratamentos, observações ou repetições; desdobramento de variância residual em seus componentes e transformação dos dados antes da análise.

Segundo LE CLERG *et alii.* (1962), uma regra geral útil para resultados colhidos em números inteiros é usar transformação raiz quadrada especialmente para dados onde os números por parcela variam de 10 a 100. Quando os números são todos maiores que 10, a raiz quadrada de cada número é tomada diretamente. Em experimentos onde os totais das parcelas são menores que 10, particularmente nos casos em que há zeros, é mais apropriado adicionar

algum valor, tais como 0.5 ou 1.0, para cada contagem antes que a raiz quadrada seja tomada. Entretanto, se a amplitude entre os valores não for tão grande, nenhuma transformação será necessária.

BOX & COX (1964) sugeriram uma família de transformações baseada em uma função de verossimilhança e sua distribuição, para a escolha de transformações adequadas que tornam o modelo linear, homocedástico e normal. A família de transformações é a seguinte:

$$Y^{(\lambda)} = \begin{cases} Y^{(\lambda)} & \text{se } \lambda \neq 0 \\ \log Y & \text{se } \lambda = 0 \end{cases} \quad (1)$$

Citam que essa particular família que foi estudada por Tukey (1957) para $|\lambda| \leq 1$ foi alterada por estes autores para evitar a descontinuidade em $\lambda = 0$ e consideraram a modificação:

$$Y^{(\lambda)} = \begin{cases} \frac{Y^{(\lambda)} - 1}{\lambda} & \text{se } \lambda \neq 0 \\ \log Y & \text{se } \lambda = 0 \end{cases} \quad (2)$$

Ambas as famílias são válidas para $Y > 0$. Eles assumem que, para algum λ , as observações

TUKEY, J.W. On the comparative anatomy of transformations. *The Annals of Mathematical Statistics*, Baltimore, 28: 602-632, 1957.

transformadas $Y^{(\lambda)} = [Y_i^{(\lambda)}]$ satisfazem à suposição da teoria normal, isto é, $Y_i^{(\lambda)} \sim N_n(X\beta, \sigma^2 I_n)$. Sob essa suposição a função de verossimilhança para as observações originais Y é:

$$(2\pi\sigma^2)^{-(1/2)n} \exp \left[-1/2\sigma^2 (Y^{(\lambda)} - X\beta)' (Y^{(\lambda)} - X\beta) \right] J \quad (3)$$

onde

$$J = \prod_{i=1}^n \left| dY_i^{(\lambda)} / dY_i \right| = \prod_{i=1}^n Y_i^{\lambda-1}$$

é o valor absoluto do Jacobiano da transformação. Para λ fixo (3) é a função de verossimilhança correspondente a um problema de quadrados mínimos padrão, exceto para o fator J . Então, o valor máximo da função de verossimilhança é:

$$(2\pi\sigma^2)^{-(1/2)n} e^{-(1/2)nJ}$$

onde:

$$\hat{\sigma}^2 = Y^{(\lambda)'} (I_n - X(X'X)^{-1}X') Y^{(\lambda)} = \hat{\sigma}^2(\lambda, Y)$$

O logaritmo da função de máxima verossimilhança a menos de uma constante

$$\begin{aligned} L_{\text{máx}}(\lambda) &= -1/(2n) \log \left\{ \hat{\sigma}^2(\lambda, Y) \right\} + \log J \\ L_{\text{máx}}(\lambda) &= -1/(2n) \log \left\{ \hat{\sigma}^2(\lambda, Z) \right\} \end{aligned}$$

onde:

$$Z_i^{(\lambda)} = Y_i^{(\lambda)} / J^{1/n} = Y_i^{(\lambda)} / \bar{Y}^{\lambda-1}, \lambda \neq 0 = \bar{Y} \log Y_i$$

$\lambda = 0$ e $\bar{Y}$ é a média geométrica de Y_i , isto é:

$$\bar{Y} = (\prod_{i=1}^n Y_i)^{1/n}.$$

onde:

$$Z_i^{(\lambda)} = Y_i^{(\lambda)} / J^{1/n} = Y_i^{(\lambda)} / \bar{Y}^{\lambda-1}, \lambda \neq 0 = \bar{Y} \log Y_i$$

$\lambda = 0$ e $\bar{Y}$ é a média geométrica de Y_i , isto é:

$$\bar{Y} = \left(\sum_{i=1}^n Y_i \right)^{1/n}.$$

Os autores sugerem plotar $L_{\text{máx}}(\lambda)$ versus λ para um conjunto de valores e registrar o valor maximizador $\hat{\lambda}$. Um valor mais preciso de $\hat{\lambda}$ pode ser obtido desenvolvendo as equações $dL_{\text{máx}}(\lambda)/d\lambda = 0$. Eles discutem também a estimação de λ de um ponto de vista amostral e Bayesiano. Para testar a adequacidade de uma transformação específica $\lambda = \lambda_0$, os autores sugerem o teste estatístico da razão da verossimilhança

$$D = -2 \left\{ L_{\text{máx}}(\lambda_0) - L_{\text{máx}}(\hat{\lambda}) \right\},$$

que é assintoticamente distribuído com um χ^2_1 quando a hipótese nula $H : \lambda = \lambda_0$ é verdadeira.

SCHLESSELMAN (1971) fez uma nota sobre a transformação de Box & Cox, mostrando que esses autores apontaram que, desde que uma análise da variância não seja mudada por uma transformação linear a família de transformações (1) é equivalente a (2). Mas, falharam em mencionar, entretanto, que em modelos que não atentam para a remoção de uma constante aditiva, a definição (2) conduz a procedimentos que não são invariantes em escala para $\lambda \neq 0$ levando a uma alteração da soma de quadrados. Portanto, modificações da família de transformações são sugeridas tais que preservam ambas continuidade na transformação e invariância na escala dos procedimentos de verossimilhança.

DEMÉTRIO (1978) cita visando à estabilização

de variâncias as transformações:

- a) $\text{arc sen } \sqrt{X}$: variáveis Binomiais (porcentagem);
- b) raiz quadrada: variável Poisson - variância proporcional à média;
- c) logarítmica : variância proporcional ao quadrado da média;
- d) arco seno hiperbólico : variáveis Binomiais negativas.

Este mesmo autor , concluiu, entre outras coisas, que as indicações da literatura quanto à adição de uma constante aos dados em transformação devem ser feitas somente quando eles forem mais próximos de zero; quando todas as condições da análise da variância estiverem satisfeitas, deve-se evitar a transformação dos dados, não só por questões práticas , como também por inversões na ordenação das médias que ela pode trazer.

Segundo COCHRAN (1947) a heterogeneidade das variâncias pode ser classificada como regular e irregular. 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, assim os dados podem ser transformados de modo que passem a ter distribuição aproximadamente normal e as médias e variâncias independentes, resultando em variâncias

homogêneas. A heterogeneidade é do tipo irregular quando aparentemente não existe 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 também a homocedasticidade.

Segundo STEEL & TORRIE (1980) quando as variâncias são proporcionais aos quadrados das médias dos tratamentos ou os desvios padrões são proporcionais às médias, a transformação logarítmica homogeneiza as variâncias. A base 10 é usada por conveniência, sendo qualquer base satisfatória. Os efeitos que são multiplicativos na escala original de medida se tornam aditivos na escala logarítmica. A transformação logarítmica é usada para números inteiros positivos quando cobrem uma larga amplitude de valores. Não pode ser usada diretamente para valores zero; ou quando alguns dos valores são menores que 10, é desejável se ter uma transformação que atue como a raiz quadrada para valores pequenos e como a logarítmica para valores grandes. A adição de 1 a cada número antes de se tomarem os logaritmos tem o efeito desejado. Aplicam a transformação angular ou $\arcsen \sqrt{X}$ a dados Binomiais expressos como frações decimais ou porcentagens e é especialmente recomendado quando a porcentagem cobre uma ampla gama de valores. Os valores arco seno são expressos tanto em graus como em radianos. A variância das observações resultantes

é aproximadamente constante, sendo $821/n$ quando os dados transformados são expressos em graus e $1/(4n) = 0.25/n$, quando em radianos, sendo n o denominador comum para todas as frações. Essa variância torna claro que todas as porcentagens deveriam ser baseadas no mesmo número de observações.

HOAGLIN et alii (1982) citam a tabela de transformações potências mais usadas, que são as principais transformações da "escala de transformações potências de Tukey", através da qual, partindo da relação linear entre o logaritmo da dispersão dos quartis dos dados com a mediana destes, consegue-se facilmente uma transformação adequada para eliminar essa dependência, com o uso apenas de uma regressão linear simples entre esses parâmetros.

Segundo PARENTE (1984) quando um teste para heterogeneidade de variâncias detecta a existência de heterocedasticidade, e ainda, a relação entre a média e a 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, citado por PARENTE (1984):

TEOREMA Rao (1965): Seja $[T_n]$, $n = 1, 2, \dots$, uma sequência de estatísticas, tais que:

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

Seja g uma função de uma única variável e admita-se que existe a derivada de primeira ordem g' , $g'(\theta) \neq 0$. Então,

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

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

$$g(T_n) - g(\theta) = (T_n - \theta) \left[g'(\theta) + E_n \right], \text{ onde } E_n \xrightarrow{P} 0$$

quando $T_n \rightarrow \theta$.

Isto implica que, para qualquer ε arbitrário, $|E_n| < \varepsilon$ e sempre que $|T_n - \theta| < \delta$.

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

$$\begin{aligned} \sqrt{n} \left[g(T_n) - g(\theta) \right] &= \sqrt{n} (T_n - \theta) g'(\theta) = \\ &= \sqrt{n} (T_n - \theta) E_n \xrightarrow{P} 0 \end{aligned}$$

desde que $\sqrt{n} (T_n - \theta)$ tem uma distribuição assintótica e $E_n \xrightarrow{P} 0$. Então, a distribuição

RAO, C.R. Linear statistical inference and its applications. J. Wiley & Sons, Inc. Nova Iorque, 1965, 522p.

assintótica de:

$$\sqrt{n} \left[g(T_n) - g(\theta) \right]$$

é a mesma de:

$$\sqrt{n} (T_n - \theta) g'(\theta)$$

que é:

$$\text{Normal} \left(0, \left[g'(\theta) \sigma(\theta) \right]^2 \right)$$

completando a prova.

Toma-se agora: $g'(\theta)\sigma(\theta)=C$ (constante independente de θ), a variância assintótica da estatística transformada $g(T_n)$ será independente de θ .

Então, $g(T_n)$ pode ser obtido 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.

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,

$$\log S_x^2 = \log a + b \log \bar{X}$$

Aplicando-se a fórmula geral da transformação, tem-se,

$$T(X) = \int \frac{C}{\sqrt{a X^b}} dX$$

$$T(X) = \frac{C}{\sqrt{a}} \int \frac{dX}{X^{b/2}}$$

$$T(X) = \frac{C}{\sqrt{a} \left[1 - \frac{b}{2} \right]} X^{[1 - (b/2)]} \quad , \text{se } b \neq 2$$

$$\frac{C}{\sqrt{a}} \log X \quad , \text{se } b = 2$$

onde b é obtido fazendo-se a regressão de $\log S_x^2$ contra $\log \bar{X}$. As transformações apropriadas, tomando-se o valor de C adequado são:

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

$$Y = X \left[1 - \frac{b}{2} \right] \quad , \text{se } b \neq 2$$

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

TABELA 1: Valores do coeficiente de regressão linear simples, b , com a transformação correspondente.

RELAÇÃO ENTRE $S_x^2 \propto \bar{X}$	VALORES DE b	$1 - b/2$	TRANSFORMAÇÃO $Y = X^{[1 - (b/2)]}$
$S_x^2 \propto \bar{X}^0$	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	$Y = \log X$
$S_x^2 \propto \bar{X}^3$	3	$-1/2$	$Y = 1/\sqrt{X}$
$S_x^2 \propto \bar{X}^4$	4	-1	$Y = 1/X$

FONTE: BARBOSA (1983)

3- METODOLOGIA

3.1. Modelo aleatório

Seja um experimento inteiramente ao acaso, cujo modelo matemático é:

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

onde:

Y_{ij} é um vetor de observações com $i=1,2,\dots,I$ e $j=1,2,\dots,J$;

m é a média geral;

t_i é o efeito de tratamentos;

ε_{ij} é o erro experimental.

Considerando o modelo aleatório tem-se:

m é uma constante sempre fixa, portanto,

$$E(m) = m \text{ e } E(m^2) = m^2$$

t_i é aleatório com $E(t_i) = 0$ e $E(t_i^2) = \sigma_t^2$;

ε_{ij} sempre aleatório com $E(\varepsilon_{ij}) = 0$ e $E(\varepsilon_{ij}^2) = \sigma_\varepsilon^2$ e independentes, onde pode-se supor que sejam normalmente distribuídos.

Deve-se ressaltar também que os efeitos aleatórios não são correlacionados.

Para a verificação da proporção de variação explicada, com dados originais e transformados, foi escolhido o modelo inteiramente ao acaso com efeitos aleatórios e feita uma simulação de dados que seguem uma distribuição Binomial e Poisson.

Foram simulados 100 experimentos para cada distribuição, com 4 tratamentos e 100 repetições, num total de 400 dados para cada simulação.

Para a geração dos dados das distribuições de Poisson e Binomial, foram utilizadas as funções aleatórias do Sistema SAS (1987a, 1987b) (Statistical Analysis System).

Para a distribuição de Poisson, utilizou-se a função aleatória RANPOI, que exige o parâmetro $\lambda > 0$ da distribuição, mais um valor inicial, semente, chamado "SEED" para iniciar o processo, cuja forma é expressa por,

RANPOI(SEED, λ)

Foram escolhidos os valores :

amostra 1: $\lambda_1 = 1$

amostra 2: $\lambda_2 = 2$

amostra 3: $\lambda_3 = 3$

amostra 4: $\lambda_4 = 4$

que produzem médias de tratamentos diferentes para cada experimento. Os dados foram obtidos atribuindo-se valores diferentes para "SEED" para cada simulação, uma vez que para o mesmo valor inicial, a subrotina RANPOI gera o mesmo valor.

A seguir foram utilizadas as transformações $\sqrt{x+0.5}$, $\log(x + 0.5)$ e fez-se uma análise da variância univariada dos dados através do PROC UNIVARIATE do SAS para se verificar qual a melhor transformação para a referida distribuição no sentido de se obter uma distribuição normal ou pelo menos simétrica. Foram feitos também testes para homogeneidade de variâncias.

Para a distribuição Binomial utilizou-se a função aleatória RANBIN, que exige os parâmetros $\underline{n}$ e $\underline{p}$ da distribuição Binomial onde $\underline{n} > 0$ inteiro, $0 < p < 1$, e também a semente "SEED" para iniciar o processo, cuja forma é expressa por:

RANBIN (seed, n,p)

Essa função gera observações de uma variável Binomial com média np e variância $np(1-p)$.

Foram escolhidos os valores:

amostra 1: $n_1 = 10$, $p_1 = 0.01$

amostra 2: $n_2 = 9$, $p_2 = 0.05$

amostra 3: $n_3 = 8$, $p_3 = 0.1$

amostra 4: $n_4 = 7$, $p_4 = 0.15$

produzindo médias de tratamentos diferentes para cada experimento, mudando-se sempre a semente "SEED".

As transformações utilizadas neste caso foram : $\sqrt{x + 0.5}$, $\log(x + 0.5)$ e $\arcsin \sqrt{\frac{x}{100}}$. De maneira análoga a anterior foi feita também uma análise da variância univariada dos dados para verificar qual a

transformação que os normaliza, ou pelo menos obter-se uma simetria para a distribuição dos mesmos. Testes para homogeneidade de variâncias também foram feitos.

3.2 Estimação dos componentes de variância

Para a estimação dos componentes de variância usou-se o procedimento PROC VARCOMP do SAS e o método utilizado foi o de Máxima Verossimilhança Restrita (MVR), pois, para dados balanceados são equivalentes aos estimadores da análise de variância e também para evitar o aparecimento de variâncias negativas.

O método consiste no seguinte :

Consideremos um modelo aleatório escrito na forma (CORBEIL & SEARLE (1976a))

$$y = Xu + U_1 b_1 + \dots + U_e b_e + e \quad (1)$$

onde:

y é um vetor de N observações,

u é um vetor de K constantes desconhecidas, relativas aos efeitos fixos do modelo,

X é uma matriz $N \times K$, de posto coluna completo, correspondente para u e com $K < N$,

U_i é uma matriz delineamento $N \times m_i$ associada ao i -ésimo fator de efeito aleatório, com $\sum_c m_i + K < N$,

b_i é um vetor de m_i variáveis aleatórias independentes e normalmente distribuídas com elementos de média 0 e variância σ_i^2 , com os b_i 's mutuamente independentes.

e é um vetor de N variáveis aleatórias normalmente e independentemente distribuídas com elementos de média 0 e variância σ^2 e independente dos b_i 's.

Portanto, y tem uma distribuição normal multivariada com média e variância:

$$E(y) = Xu \text{ e } \text{var}(y) = V = H \sigma^2 \quad (2)$$

onde:

$$H = \sum_{i=1}^c \gamma_i U_i U_i' + I_N \text{ para } \gamma_i = \frac{\sigma_i^2}{\sigma^2} \quad (3)$$

$$\text{e } X = \begin{bmatrix} 1_{n_1} & 0 & \dots & 0 \\ 0 & 1_{n_2} & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & \dots & 0 & 1_{n_t} \end{bmatrix} = \sum_{i=1}^t 1_{n_i} \quad (4)$$

onde 1_{n_i} é um vetor de n_i 1's e $\sum^+$ representa a soma direta de matrizes.

O logaritmo da função verossimilhança para $y \sim N(Xu, H\sigma^2)$ é:

$$\begin{aligned} \lambda = & -\frac{1}{2} N \log 2\pi - \frac{1}{2} N \log \sigma^2 - \frac{1}{2} \log |H| - \\ & - \frac{1}{2 \sigma^2} (y - Xu)' (y - Xu) \end{aligned} \quad (5)$$

Para separar a expressão acima em duas partes, uma das quais é livre de u , Patterson e Thompson (1971) citado por CORBEIL & SEARLE (1972) sugeriram a transformação singular:

PATTERSON, H.D. & R. THOMPSON. Recovery of interblock information when block sizes are unequal. *Biometrika*, Cambridge, 58: 545-554, 1971.

$$y' [S_1 H^{-1} X]$$

onde:

$$S = I - X(X'X)^{-1}X' = \sum_{i=1}^{K_t} \left(I_{n_i} - n_i^{-1} J_{n_i} \right) \quad (6)$$

é simétrica e idempotente, com J_{n_i} começando na matriz $n_i \times n_i$ com todos elementos unitários.

Uma vez que SX é nulo, Sy tem distribuição $N(0, SHS\sigma^2)$ independente de $X'H^{-1}y$.

Vê-se que a distribuição de Sy é livre de efeitos fixos u . Essa função verossimilhança, portanto, forma a base da derivação de estimadores de componentes de variância envolvidos no $H\sigma^2$. Porém, para evitar a singularidade de SHS surgindo na forma de S dado em (6), usa-se uma S alternativa obtida a partir da eliminação da n_1 -ésima, (n_1+n_2) -ésima, $(n_1+n_2+n_3)$ -ésima ..., e $(n_1+n_2+\dots+n_K)$ -ésima linha. Cada matriz tem ordem $(N-K) \times N$, e denotando por T tem-se:

$$T = \sum_{i=1}^{K_t} \left[(I_{n_{t-1}} \mid 0 \mid 1_{n_{t-1}}) - n_K^{-1} J_{(n_{t-1})} \times n_t \right] \quad (7)$$

$$= \sum_{i=1}^{K_t} \left(I_{n_{t-1}} - n_t^{-1} J_{n_{t-1}} \quad \vdots \quad -n_t^{-1} 1_{n_{t-1}} \right) \quad (8)$$

onde:

01_a é um vetor de zeros de ordem a e $J_{(n_{t-1})} \times n_t$ é uma matriz de ordem $(n_{t-1}) \times n_t$ onde todos os elementos são unitários. A partir do X de (4) vê-se que:

$$TX = 0 \quad (9)$$

pela própria natureza de T e:

$$T'(TT')^{-1}T = S \quad (10)$$

A transformação agora usada é :

$$z = \begin{bmatrix} T \\ X'H^{-1} \end{bmatrix} \quad y = \begin{bmatrix} Ty \\ X'H^{-1}y \end{bmatrix} \quad (11)$$

cuja distribuição a partir de (9) será:

$$z \sim N \left\{ \begin{bmatrix} 0 \\ X'H^{-1}Xu \end{bmatrix}, \begin{bmatrix} THT'\sigma^2 & 0 \\ 0 & X'H^{-1}X\sigma^2 \end{bmatrix} \right\} \quad (12)$$

Essa transformação é não singular porque X' e T de (4) e (7) têm posto linha completo e a partir de (9) as linhas de T são linearmente independentes das de X' .

Considera-se agora o logaritmo da função verossimilhança. De (11) e (12), os logaritmos das funções de verossimilhança de Ty e de $X'H^{-1}y$ os quais são denotados por λ_1 e λ_2 , respectivamente são :

$$\begin{aligned} \lambda_1 = & -\frac{1}{2} (N-K) \log 2\pi - \frac{1}{2} (N-K) \log \sigma^2 - \frac{1}{2} \log |TH T'| \\ & - \frac{1}{2\sigma^2} y'T'(THT')^{-1} Ty \end{aligned} \quad (13)$$

e

$$\begin{aligned} \lambda_2 = & -\frac{1}{2} K \log 2\pi - \frac{1}{2} K \log \sigma^2 - \frac{1}{2} \log |X'H^{-1}X| - \\ & - \frac{1}{2\sigma^2} (y - Xu)' H^{-1} X (X'H X)^{-1} X'H^{-1} (y^{-1} - Xu) \end{aligned} \quad (14)$$

Como λ_1 não envolve μ os estimadores de σ^2 e os γ_i 's são chamados estimadores de Máxima Verossimilhança

Restrita (MVR), cujos valores de $\hat{\sigma}^2$ e $\hat{\gamma}$'s são que maximizam λ_1 e a derivação de (13) tem-se:

$$\frac{\partial \lambda_1}{\partial \sigma^2} = - \frac{1}{2\sigma^2} (N-K) + \frac{1}{2\sigma^4} y' T' (T H H')^{-1} T y \quad (15)$$

e

$$\begin{aligned} \frac{\partial \lambda_1}{\partial \gamma_i} = & - \frac{1}{2} \text{tr} [U_i' T' (T H T')^{-1} T U_i] + \\ & + \frac{1}{2\sigma^2} y' T' (T H T')^{-1} T U_i U_i' T' (T H T')^{-1} y, i=1,2,\dots,c \end{aligned} \quad (16)$$

onde $\text{tr}(Q)$ é o traço da matriz Q .

Igualando (15) e (16) a zero têm-se os estimadores de Máxima Verossimilhança Restrita.

As equações resultantes não têm solução analítica e precisam ser resolvidas numericamente.

Um procedimento iterativo é atribuir primeiro um valor inicial para

$$\gamma' = \{\gamma_1, \dots, \gamma_c\}$$

e o procedimento da solução é:

$$(i) \text{ calcula-se } \hat{\sigma}^2 = y' T' (T H T')^{-1} T y / (N-K) \quad (17)$$

(ii) usam-se os γ -valores, e $\hat{\sigma}^2$ de (17), para calcular novos valores de γ que torna os elementos de (16) mais próximos de zero. A repetição de (i) e (ii) terminando em (i), é continuado até que um grau de precisão é alcançado.

Como os valores das estimativas dos componentes de variância obtidas com dados transformados e não transformados não são comparáveis, pois apresentam

unidades de medidas distintas de acordo com a transformação usada, comparamos as porcentagens em relação à estimativa da variância total, ou seja, a proporção de variação explicada, que é dada por:

$$\phi_t = \frac{\hat{\sigma}_t^2}{\hat{\sigma}_{total}^2} \times 100$$

onde:

ϕ_t é a proporção da variação explicada devida a tratamentos.

$\hat{\sigma}_t^2$ é a estimativa do componente de variância devida a tratamentos.

$\hat{\sigma}_{total}^2$ é a estimativa do componente de variância total.

Portanto:

$$\hat{\sigma}_{total}^2 = \hat{\sigma}_t^2 + \hat{\sigma}_\varepsilon^2$$

$\hat{\sigma}_\varepsilon^2$ é a estimativa do componente de variância residual.

4 - RESULTADOS E DISCUSSÃO

De acordo com os dados simulados segundo o modelo inteiramente ao acaso, distribuição de Poisson, as estimativas dos componentes de variância para efeitos de tratamentos dos dados originais e transformados, utilizando-se as transformações Raiz Quadrada e Logarítmica, são apresentados na Tabela 1. O método de estimação usado foi o da Máxima Verossimilhança Restrita.

Na tabela 2 têm-se as estimativas dos componentes de variância para o efeito residual.

A porcentagem de variação explicada em relação ao total, para efeito de tratamentos, dos dados sem transformação e com transformações Raiz Quadrada e Logarítmica, apresentados na Tabela 3, revelam uma diferença bastante pequena entre as proporções de variação explicada em relação ao total para dados sem transformação e transformados. Verificou-se que essa diferença, em geral, é de no máximo, 4% . Observou-se também que a transformação que fornece a maior proporção de variação explicada, considerando a distribuição de Poisson, na maioria dos casos foi a transformação Raiz Quadrada,

confirmando os referencias de literatura.

Como as diferenças entre dados transformados e não transformados são pequenas em relação à proporção de variação explicada em relação ao total, poder-se-ia adotar um procedimento de trabalho com os dados sem transformação. Ocorre que os testes F para efeitos de tratamentos perderiam o poder de decisão, uma vez que os dados não são normais e as variâncias são heterogêneas, o que não ocorre em termos de estimação dos componentes de variância, já que não se faz nenhuma exigência para este procedimento. Assim, seria mais interessante o uso da transformação, que neste caso, foi a transformação Raiz Quadrada.

Na Tabela 4 tem-se a proporção de variação explicada em relação ao total para o efeito residual.

Tabela 1- Valores de $\hat{\sigma}_t^2$ para dados simulados com distribuição de Poisson no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
001	1,682736	0,158603	0,056124
002	1,611740	0,148990	0,005115
003	1,697881	0,159224	0,055524
004	2,113456	0,197377	0,055524
005	1,397301	0,134996	0,048777
006	1,717586	0,142814	0,043379
007	1,848749	0,166410	0,054822
008	1,458323	0,135849	0,048298
009	1,779607	0,170821	0,061008
010	1,863434	0,136554	0,043181
011	1,571288	0,153519	0,055992
012	1,559681	0,148084	0,055893
013	1,841126	0,160584	0,052866
014	1,680820	0,140885	0,044345
015	1,764221	0,153764	0,050638
016	1,525587	0,147954	0,052945
017	1,652522	0,163201	0,059559
018	1,602986	0,149712	0,051539
019	1,627604	0,149612	0,051476
020	1,949466	0,168307	0,053760
021	1,919469	0,167094	0,053409
022	1,652269	0,151464	0,051763
023	0,152353	0,196355	0,066722
024	1,853072	0,173904	0,061898
025	1,350828	0,123126	0,039964
026	1,512776	0,140361	0,048793
027	1,461244	0,132213	0,045352
028	1,689360	0,147333	0,049382
029	1,495755	0,130360	0,041306
030	1,513994	0,137105	0,047603
031	1,536313	0,148680	0,053448
032	1,591622	0,142658	0,047603
033	1,495298	0,143568	0,051446
034	1,871089	0,163439	0,053373
035	1,963622	0,171679	0,056827
036	1,219583	0,117548	0,041896
037	1,692226	0,160429	0,056716
038	1,703098	0,159199	0,055890
039	1,825457	0,179388	0,066816
040	1,975782	0,180434	0,061867

Tabela 1- Valores de $\hat{\sigma}_t^2$ para dados simulados com distribuição de Poisson no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
041	1,599606	0,143103	0,046858
042	2,054323	0,187023	0,063085
043	1,720271	0,164889	0,059420
044	2,048123	0,184052	0,064065
045	1,248024	0,115243	0,039606
046	1,841761	0,168047	0,057632
047	1,533541	0,149386	0,054249
048	2,070149	0,181506	0,059473
049	2,130082	0,185586	0,060900
050	1,696328	0,152653	0,052092
051	1,244669	0,111837	0,036431
052	1,466975	0,140238	0,050040
053	1,489911	0,131986	0,042819
054	1,536718	0,144426	0,050253
055	1,329602	0,117905	0,040363
056	1,535670	0,138595	0,046232
057	1,723987	0,162313	0,057132
058	1,554672	0,150611	0,053684
059	1,831115	0,165862	0,056847
060	1,447952	0,136629	0,047706
061	1,941850	0,187249	0,068904
062	1,388691	0,127171	0,042976
063	1,889659	0,172614	0,058804
064	1,410791	0,121958	0,038381
065	1,571365	0,143772	0,049004
066	1,826818	0,171581	0,059812
067	1,344924	0,117247	0,037631
068	1,684581	0,160980	0,056745
069	1,702995	0,155390	0,052506
070	1,322632	0,119764	0,040565
071	1,745218	0,160411	0,056148
072	1,833240	0,171605	0,060715
073	1,647426	0,143898	0,047432
074	1,309106	0,116226	0,037749
075	1,843188	0,164177	0,063818
076	1,767421	0,163746	0,056978
077	2,308586	0,198333	0,063670
078	1,545830	0,143890	0,051020
079	1,824688	0,164908	0,056686
080	1,626497	0,155038	0,054582

Tabela 1- Valores de $\hat{\sigma}_i^2$ para dados simulados com distribuição de Poisson no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
081	1,611477	0,136554	0,043181
082	1,716619	0,159035	0,054464
083	1,490614	0,138016	0,047715
084	1,345516	0,124345	0,047715
085	1,530159	0,153105	0,057817
086	1,676122	0,148895	0,049754
087	1,632930	0,149555	0,051267
088	1,708022	0,163819	0,059163
089	1,650704	0,152123	0,050833
090	1,950912	0,179168	0,067930
091	1,045392	0,103028	0,037443
092	2,082250	0,020548	0,076457
093	1,544819	0,139476	0,046986
094	1,389454	0,133213	0,047354
095	1,412528	0,120942	0,038374
096	1,718754	0,158622	0,055534
097	1,842422	0,154594	0,048465
098	1,519605	0,140713	0,038836
099	1,502153	0,139794	0,048522
100	1,866832	0,184020	0,070085

Tabela 2- Valores de $\hat{\sigma}_e^2$ para dados simulados com distribuição de Poisson no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0.5}$ e $\log(X + 0.5)$.

N	X	$\sqrt{X + 0.5}$	$\log(X + 0.5)$
001	2,577651	0,215255	0,077786
002	2,341792	0,199027	0,072903
003	2,280393	0,202409	0,077644
004	2,592348	0,218410	0,073168
005	2,609949	0,225328	0,082786
006	2,728030	0,225926	0,080387
007	2,347500	0,213352	0,082942
008	2,695908	0,219405	0,076581
009	2,116792	0,190791	0,073797
010	2,698326	0,211968	0,072089
011	2,171665	0,192334	0,072084
012	2,744620	0,228375	0,082004
013	2,600656	0,216880	0,077042
014	2,520378	0,212115	0,078767
015	2,956186	0,225865	0,077404
016	2,144948	0,184048	0,066772
017	2,141868	0,182530	0,065703
018	2,486084	0,207810	0,073405
019	2,417145	0,207489	0,075666
020	2,366717	0,202559	0,074747
021	2,235302	0,189511	0,069180
022	3,086414	0,249086	0,086953
023	2,707196	0,209916	0,072401
024	2,569393	0,204937	0,071809
025	2,367171	0,200410	0,071164
026	2,751540	0,243225	0,090854
027	2,615505	0,266337	0,077910
028	2,390958	0,196508	0,069101
029	2,534495	0,204485	0,070659
030	2,430555	0,205141	0,090030
031	2,236034	0,195053	0,072293
032	2,854292	0,242448	0,090030
033	2,435479	0,206066	0,074990
034	2,626943	0,200253	0,069596
035	3,005680	0,245029	0,088149
036	2,090782	0,197243	0,077487
037	2,314267	0,200536	0,076013
038	2,549797	0,221792	0,084422
039	2,745277	0,229184	0,081135
040	2,499550	0,207846	0,076815

Tabela 2- Valores de $\hat{\sigma}_e^2$ para dados simulados com distribuição de Poisson no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0.5}$ e $\log(X + 0.5)$.

N	X	$\sqrt{X + 0.5}$	$\log(X + 0.5)$
041	2,368560	0,196279	0,067926
042	2,347626	0,192411	0,069169
043	2,539545	0,221268	0,082473
044	2,688282	0,229892	0,087040
045	2,504165	0,218296	0,080345
046	2,666615	0,200671	0,066320
047	2,079141	0,181665	0,066968
048	2,274267	0,192272	0,071120
049	2,884595	0,227408	0,080840
050	2,323055	0,193917	0,068562
051	2,189747	0,199672	0,075977
052	2,384974	0,211999	0,079291
053	2,615327	0,225512	0,083827
054	2,108131	0,188892	0,072943
055	2,488964	0,206660	0,070820
056	2,282146	0,209619	0,080176
057	2,562145	0,210579	0,075642
058	2,291767	0,192447	0,070430
059	2,397601	0,204488	0,074485
060	2,251756	0,204137	0,077325
061	2,595000	0,224502	0,083392
062	2,200075	0,209041	0,083064
063	2,354696	0,198547	0,073354
064	2,393786	0,198674	0,070106
065	2,585479	0,219901	0,077990
066	2,604797	0,222005	0,082568
067	2,447525	0,215324	0,078896
068	2,371868	0,210613	0,080764
069	2,282954	0,196142	0,072987
070	3,033383	0,253241	0,090790
071	2,661515	0,219396	0,078967
072	2,715959	0,225794	0,081628
073	2,529317	0,209400	0,074058
074	2,576539	0,218197	0,078730
075	2,484595	0,202446	0,073294
076	2,800378	0,229920	0,083451
077	2,457146	0,197086	0,070285
078	2,441362	0,209084	0,075442
079	2,517777	0,199911	0,071188
080	2,312978	0,205776	0,078775

Tabela 2- Valores de $\hat{\sigma}_e^2$ para dados simulados com distribuição de Poisson no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0.5}$ e $\log(X + 0.5)$.

N	X	$\sqrt{X + 0.5}$	$\log(X + 0.5)$
081	2,721388	0,211968	0,072089
082	2,388712	0,198962	0,073315
083	2,597118	0,227018	0,077103
084	2,401767	0,202208	0,077103
085	2,620707	0,220777	0,080550
086	2,603661	0,224313	0,084932
087	2,415428	0,206762	0,093356
088	2,226893	0,190445	0,068737
089	2,228358	0,202721	0,077900
090	2,430832	0,207523	0,077848
091	2,289873	0,202481	0,076283
092	2,277601	0,196546	0,072698
093	2,450403	0,210812	0,079639
094	2,635403	0,223729	0,082047
095	2,668231	0,225792	0,081140
096	2,487095	0,207752	0,075353
097	2,497928	0,203450	0,072798
098	2,741994	0,237453	0,084114
099	2,684696	0,228271	0,082616
100	2,653484	0,217682	0,076955

Tabela 3- Porcentagem de variação explicada de $\hat{\sigma}_1^2$ para dados simulados com distribuição de Poisson, para 100 experimentos, no modelo inteiramente ao acaso usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
001	39	42	42
002	41	43	41
003	43	44	42
004	45	48	42
005	35	37	37
006	39	39	35
007	44	44	40
008	35	38	39
009	46	47	45
010	41	39	37
011	42	44	44
012	36	39	39
013	41	43	41
014	40	40	36
015	37	40	40
016	42	45	44
017	44	47	48
018	39	42	41
019	40	42	40
020	45	45	42
021	46	47	44
022	35	38	37
023	44	48	48
024	42	46	46
025	36	38	36
026	35	37	35
027	36	38	37
028	41	43	42
029	37	39	37
030	38	40	35
031	41	43	43
032	36	37	35
033	38	41	41
034	42	45	43
035	40	41	39
036	37	37	35
037	42	44	43
038	40	42	40
039	40	44	45
040	44	46	45

Tabela 3- Porcentagem de variação explicada de $\hat{\sigma}_t^2$ para dados simulados com distribuição de Poisson, para 100 experimentos, no modelo inteiramente ao acaso usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
041	40	42	41
042	47	49	48
043	40	43	42
044	43	45	42
045	33	34	33
046	41	46	47
047	42	45	45
048	48	49	45
049	42	45	43
050	42	44	43
051	36	36	32
052	38	40	39
053	36	37	34
054	42	43	41
055	35	37	36
056	40	40	37
057	40	44	43
058	40	44	43
059	43	45	43
060	39	40	38
061	43	45	45
062	39	38	34
063	45	46	44
064	37	38	35
065	38	39	39
066	41	44	42
067	35	35	32
068	42	43	42
069	43	44	42
070	30	32	31
071	40	42	42
072	40	43	43
073	39	41	39
074	34	35	32
075	43	45	43
076	39	42	41
077	48	50	48
078	39	41	40
079	42	45	44
080	41	43	41

Tabela 3- Porcentagem de variação explicada de $\hat{\sigma}_t^2$ para dados simulados com distribuição de Poisson, para 100 experimentos, no modelo inteiramente ao acaso usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
081	37	39	37
082	42	44	43
083	36	39	38
084	36	38	38
085	37	41	42
086	39	40	37
087	40	42	41
088	43	46	46
089	42	43	39
090	45	47	47
091	31	34	33
092	48	51	51
093	39	40	37
094	35	37	37
095	35	35	32
096	41	43	42
097	42	43	40
098	36	37	32
099	36	38	37
100	41	46	48

Tabela 4 - Porcentagem de variação explicada de $\hat{\sigma}_e^2$ para dados simulados com distribuição de Poisson, para 100 experimentos, no modelo inteiramente ao acaso usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
001	61	58	58
002	59	57	59
003	57	56	58
004	55	52	58
005	65	63	63
006	61	61	65
007	56	56	60
008	65	62	61
009	54	53	55
010	59	61	63
011	58	56	56
012	64	61	61
013	59	57	59
014	60	60	64
015	63	60	60
016	58	55	56
017	56	53	52
018	61	58	59
019	60	58	60
020	55	55	58
021	54	53	56
022	65	62	63
023	56	52	52
024	58	54	54
025	64	62	64
026	65	63	65
027	64	62	63
028	59	57	58
029	63	61	63
030	62	60	65
031	59	57	57
032	64	63	65
033	62	59	59
034	58	55	57
035	60	59	61
036	63	63	65
037	58	56	57
038	60	58	60
039	60	54	55
040	56	54	55

Tabela 4 - Porcentagem de variação explicada de $\hat{\sigma}_e^2$ para dados simulados com distribuição de Poisson, para 100 experimentos, no modelo inteiramente ao acaso, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X + 0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
041	60	58	59
042	53	51	52
043	60	57	58
044	57	55	58
045	67	66	67
046	59	54	53
047	58	55	55
048	52	51	55
049	58	55	57
050	58	56	57
051	64	64	68
052	62	60	61
053	64	63	66
054	58	57	59
055	65	63	64
056	60	60	63
057	60	56	57
058	60	56	57
059	57	55	57
060	61	60	62
061	57	55	55
062	61	62	66
063	55	54	56
064	63	62	65
065	62	61	61
066	59	56	58
067	65	65	68
068	58	57	58
069	57	56	58
070	70	68	69
071	60	58	58
072	60	57	57
073	61	59	61
074	66	65	68
075	57	55	57
076	61	58	59
077	52	50	52
078	61	59	60
079	58	55	56
080	59	57	59

Tabela 4 -Porcentagem de variação explicada de $\hat{\sigma}_e^2$ para dados simulados com distribuição de Poisson, para 100 experimentos, no modelo inteiramente ao acaso usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$ e $\log(X+0,5)$.

N	X	$\sqrt{X + 0,5}$	$\log(X + 0,5)$
081	63	61	63
082	58	56	57
083	64	61	62
084	64	62	62
085	63	59	58
086	61	60	63
087	60	58	59
088	57	54	54
089	58	57	61
090	55	53	53
091	69	66	67
092	52	49	49
093	61	60	63
094	65	63	63
095	65	65	68
096	59	57	58
097	58	57	60
098	64	63	68
099	64	62	63
100	59	54	52

A Tabela 5 apresenta as estimativas dos componentes de variância para dados simulados a partir da distribuição Binomial para o efeito de tratamentos no modelo inteiramente ao acaso dos dados originais e transformados, utilizando-se as transformações Raiz Quadrada, Logarítmica e Arco Seno. O método de estimação usado também foi o de Máxima Verossimilhança Restrita.

Na Tabela 6 têm-se as estimativas dos componentes de variância para o efeito residual.

A porcentagem de variação explicada em relação ao total para efeito de tratamentos dos dados sem transformação e com as transformações Raiz Quadrada, Logarítmica e Arco Seno apresentados na Tabela 7, também revelam uma diferença bastante pequena entre as proporções de variação explicada em relação ao total para dados sem transformação e transformados.

A proporção de variação explicada em relação ao total para o efeito residual, é apresentada na Tabela 8.

Tabela 5 - Valores de $\hat{\sigma}_t^2$ para dados simulados com distribuição Binomial no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{\frac{X}{100}}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
001	0,173670	0,035405	0,024969	0,001045
002	0,250095	0,048136	0,031855	0,001359
003	0,106719	0,021831	0,015131	0,000646
004	0,189387	0,036306	0,024122	0,010347
005	0,131590	0,027760	0,019664	0,000839
006	0,231897	0,045675	0,049004	0,001296
007	0,224126	0,044434	0,301721	0,001293
008	0,159199	0,032309	0,022354	0,000958
009	0,171166	0,033853	0,022778	0,000966
010	0,123116	0,025751	0,018078	0,000769
011	0,204253	0,040782	0,027713	0,001180
012	0,156983	0,031525	0,021594	0,000922
013	0,164594	0,033914	0,023643	0,001009
014	0,137922	0,030015	0,021760	0,000931
015	0,164716	0,029476	0,023252	0,000996
016	0,140388	0,030843	0,022516	0,000964
017	0,221517	0,029476	0,031212	0,001333
018	0,172191	0,033738	0,022612	0,000964
019	0,165542	0,034039	0,023728	0,001015
020	0,131975	0,029475	0,019622	0,000844
021	0,141451	0,030339	0,021747	0,000928
022	0,150741	0,029475	0,023481	0,001009
023	0,241775	0,049254	0,034163	0,001465
024	0,193273	0,039655	0,027563	0,001179
025	0,150266	0,031040	0,021587	0,000918
026	0,137656	0,028165	0,019453	0,000832
027	0,143153	0,031485	0,023050	0,000989
028	0,187523	0,035124	0,022912	0,000980
029	0,178762	0,033356	0,021640	0,000921
030	0,149744	0,028379	0,018512	0,000788
031	0,158230	0,033841	0,024237	0,001035
032	0,126533	0,024950	0,016812	0,0007148
033	0,178970	0,035760	0,024420	0,001041
034	0,156320	0,031727	0,021828	0,000929
035	0,137068	0,028226	0,019697	0,000844
036	0,162281	0,032574	0,022178	0,000943
037	0,180315	0,034738	0,023035	0,000978
038	0,174310	0,033223	0,021767	0,000923
039	0,175893	0,035479	0,024311	0,001035
040	0,165432	0,033447	0,022958	0,000977

Tabela 5- Valores de $\hat{\sigma}_t^2$ para dados simulados com distribuição Binomial no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{\frac{X}{100}}$.

N	X	$\sqrt{X + 0,5}$	Log(X+0,5)	Arcsen $\sqrt{\frac{X}{100}}$
041	0,162087	0,032305	0,021930	0,000935
042	0,178090	0,035721	0,024431	0,001045
043	0,139245	0,027428	0,018471	0,000787
044	0,179372	0,037471	0,026322	0,001121
045	0,182139	0,036267	0,024602	0,001050
046	0,210334	0,042217	0,028710	0,001219
047	0,192391	0,039647	0,027553	0,001173
048	0,163321	0,031970	0,021284	0,000902
049	0,197326	0,039414	0,026644	0,001289
050	0,144834	0,029657	0,020435	0,000864
051	0,237547	0,044936	0,029356	0,001231
052	0,214654	0,042569	0,028819	0,001232
053	0,243943	0,047981	0,032358	0,001380
054	0,214435	0,040919	0,026721	0,001132
055	0,160424	0,031911	0,031627	0,000920
056	0,175250	0,036515	0,025605	0,001093
057	0,170500	0,034410	0,023769	0,001024
058	0,154267	0,031159	0,021391	0,000913
059	0,155752	0,031679	0,021897	0,000938
060	0,153863	0,030675	0,020896	0,000894
061	0,163856	0,034242	0,024113	0,001030
062	0,150546	0,030036	0,020391	0,000867
063	0,155538	0,031244	0,021546	0,000926
064	0,197438	0,029476	0,026909	0,001149
065	0,138637	0,027730	0,019021	0,000820
066	0,152706	0,032525	0,023180	0,000987
067	0,139619	0,029565	0,021099	0,000907
068	0,213412	0,029516	0,031705	0,001351
069	0,151464	0,029516	0,022876	0,000980
070	0,191046	0,038640	0,026509	0,001127
071	0,142968	0,030754	0,022058	0,000950
072	0,175940	0,034151	0,022670	0,000964
073	0,171374	0,037325	0,027282	0,001177
074	0,116868	0,023306	0,015816	0,000672
075	0,147252	0,029926	0,020579	0,000873
076	0,160024	0,033581	0,023674	0,001086
077	0,217115	0,041974	0,027774	0,001182
078	0,183921	0,037915	0,026527	0,001137
079	0,176226	0,034331	0,022895	0,000973
080	0,127388	0,026565	0,018681	0,000800

Tabela 5- Valores de $\hat{\sigma}_1^2$ para dados simulados com distribuição Binomial no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{\frac{X}{100}}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
081	0,167617	0,033476	0,022879	0,000980
082	0,177592	0,034980	0,023559	0,001003
083	0,129800	0,026566	0,018429	0,000789
084	0,178882	0,036311	0,025154	0,000789
085	0,161809	0,032699	0,022491	0,000959
086	0,152922	0,032542	0,023260	0,000997
087	0,159866	0,031214	0,020861	0,000888
088	0,092859	0,020200	0,014784	0,000637
089	0,217814	0,041079	0,024604	0,001147
090	0,193808	0,040043	0,027892	0,001188
091	0,173481	0,034578	0,023558	0,001007
092	0,177880	0,035978	0,024543	0,001040
093	0,137344	0,029552	0,021114	0,000900
094	0,178165	0,033392	0,021535	0,000911
095	0,154089	0,030682	0,020850	0,000891
096	0,190638	0,038537	0,026394	0,001123
097	0,173242	0,034948	0,023981	0,001024
098	0,198523	0,039399	0,026734	0,001140
099	0,154080	0,031592	0,021831	0,000928
100	0,132369	0,027675	0,019477	0,000831

Tabela 6- Valores de $\hat{\sigma}_e^2$ para dados simulados com distribuição Binomial no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{\frac{X}{100}}$.

N	X	$\sqrt{X + 0,5}$	Log(X+0,5)	Arcsen $\sqrt{\frac{X}{100}}$
001	0,495479	0,093929	0,063492	0,002710
002	0,606843	0,104326	0,066150	0,002826
003	0,541414	0,103587	0,070073	0,002987
004	0,646211	0,109323	0,069172	0,002970
005	0,423459	0,086339	0,061317	0,002625
006	0,525706	0,096132	0,063274	0,002697
007	0,590984	0,101995	0,065031	0,002780
008	0,545756	0,100912	0,066663	0,002839
009	0,550454	0,103189	0,065016	0,002920
010	0,481257	0,095259	0,066043	0,002817
011	0,562317	0,094747	0,062707	0,002666
012	0,546387	0,099501	0,065726	0,002810
013	0,467953	0,089514	0,061477	0,002631
014	0,436893	0,087061	0,060463	0,002576
015	0,533333	0,093287	0,066070	0,002818
016	0,456954	0,089408	0,061496	0,002622
017	0,556944	0,093286	0,063907	0,002719
018	0,609368	0,108570	0,070286	0,003002
019	0,503484	0,095214	0,064128	0,002734
020	0,582474	0,093286	0,069031	0,002940
021	0,437348	0,088398	0,062055	0,002645
022	0,460479	0,093286	0,059956	0,002567
023	0,471590	0,086042	0,057072	0,002439
024	0,535126	0,097259	0,063779	0,002722
025	0,490336	0,094877	0,064516	0,002753
026	0,545134	0,104107	0,068427	0,002922
027	0,457701	0,089619	0,061334	0,002607
028	0,640959	0,110391	0,070317	0,003010
029	0,591792	0,107859	0,070183	0,002999
030	0,540504	0,100104	0,067176	0,000468
031	0,480030	0,093061	0,063329	0,002690
032	0,483282	0,095192	0,066038	0,002824
033	0,549646	0,101083	0,066933	0,002855
034	0,514646	0,098856	0,066952	0,002850
035	0,545545	0,104708	0,069127	0,002947
036	0,471388	0,090447	0,062027	0,003609
037	0,545101	0,100354	0,066845	0,002858
038	0,535606	0,100152	0,067117	0,002871
039	0,537323	0,099341	0,065925	0,002813
040	0,566717	0,104299	0,068797	0,002928

Tabela 6- Valores de $\hat{\sigma}_e^2$ para dados simulados com distribuição Binomial no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{\frac{X}{100}}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
041	0,463383	0,088922	0,061137	0,002621
042	0,570959	0,102271	0,066831	0,002856
043	0,572904	0,105640	0,070295	0,003004
044	0,506111	0,096927	0,069172	0,002760
045	0,656010	0,115825	0,073400	0,003114
046	0,509015	0,094710	0,063025	0,002684
047	0,470858	0,090146	0,061181	0,002609
048	0,554494	0,103764	0,069325	0,002954
049	0,553989	0,102797	0,067702	0,002872
050	0,445681	0,089731	0,062821	0,002680
051	0,603004	0,104744	0,066808	0,002856
052	0,614014	0,104895	0,065784	0,002809
053	0,522323	0,094703	0,061973	0,002643
054	0,625259	0,109886	0,070252	0,002996
055	0,571918	0,105439	0,069706	0,002970
056	0,568813	0,100874	0,065342	0,002796
057	0,552500	0,097830	0,063887	0,002742
058	0,526716	0,098045	0,065795	0,002816
059	0,577827	0,104692	0,068516	0,002922
060	0,569469	0,105557	0,069711	0,002965
061	0,497727	0,094555	0,063922	0,002726
062	0,549217	0,102194	0,068157	0,002911
063	0,600176	0,107596	0,069871	0,002979
064	0,566111	0,093286	0,066024	0,002820
065	0,636212	0,111559	0,071258	0,003044
066	0,445075	0,088370	0,061124	0,002601
067	0,503685	0,094863	0,064111	0,002742
068	0,457272	0,093286	0,058872	0,002501
069	0,469443	0,093286	0,061340	0,002625
070	0,511969	0,096100	0,064281	0,002736
071	0,492878	0,093093	0,059102	0,002676
072	0,596110	0,108009	0,070619	0,003008
073	0,440277	0,084111	0,057483	0,002457
074	0,556515	0,106381	0,071588	0,003045
075	0,483964	0,095372	0,065690	0,002797
076	0,530277	0,099872	0,066703	0,002834
077	0,570934	0,103060	0,066115	0,002847
078	0,558358	0,100026	0,064769	0,002780
079	0,546540	0,101007	0,066981	0,002859
080	0,557828	0,101912	0,067171	0,002875

Tabela 6- Valores de $\hat{\sigma}_e^2$ para dados simulados com distribuição Binomial no modelo inteiramente ao acaso, para 100 experimentos, usando-se dados sem transformação (X) e com as transformações $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{\frac{X}{100}}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
081	0,597449	0,106051	0,068484	0,002923
082	0,563207	0,103364	0,068043	0,002902
083	0,529519	0,099628	0,067243	0,002877
084	0,575504	0,104947	0,065900	0,002819
085	0,528358	0,099997	0,067202	0,002861
086	0,461060	0,088682	0,060222	0,002567
087	0,525959	0,098494	0,066194	0,002831
088	0,463207	0,091340	0,064113	0,002753
089	0,556262	0,098345	0,064003	0,002746
090	0,585681	0,106409	0,068851	0,002916
091	0,527651	0,096293	0,064127	0,002750
092	0,574469	0,103849	0,067645	0,002883
093	0,484594	0,093286	0,063699	0,002719
094	0,566767	0,105003	0,070054	0,002991
095	0,605353	0,107113	0,069576	0,002981
096	0,506104	0,094137	0,062915	0,002687
097	0,484873	0,091283	0,061451	0,002627
098	0,564292	0,101320	0,066005	0,002816
099	0,468585	0,089491	0,061357	0,002635
100	0,499747	0,096281	0,065719	0,002806

Tabela 7 - Porcentagem de variação explicada de $\hat{\sigma}_t^2$ para dados simulados com distribuição Binomial, para 100 experimentos, no modelo inteiramente ao acaso, usando-se dados sem transformação (X) e com as transformações: $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{X / 100}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
001	26	27	28	29
002	29	32	32	32
003	16	17	18	18
004	23	25	26	26
005	24	24	24	24
006	31	32	32	32
007	28	30	32	32
008	23	24	25	25
009	24	25	25	25
010	20	21	22	21
011	29	30	31	31
012	22	24	25	25
013	26	27	28	28
014	24	26	26	27
015	24	24	26	26
016	24	26	27	27
017	28	24	33	33
018	22	24	24	24
019	25	26	27	27
020	18	24	22	22
021	24	26	26	26
022	24	24	28	28
023	34	36	37	38
024	27	29	30	30
025	23	25	25	25
026	20	21	22	22
027	24	26	27	27
028	23	24	25	25
029	23	24	24	23
030	22	22	21	21
031	25	27	28	28
032	21	21	20	20
033	25	26	27	27
034	23	24	25	25
035	10	21	22	22
036	26	26	26	26
037	25	26	26	25
038	25	25	24	24
039	25	26	27	27
040	23	24	25	25

Tabela 7 - Porcentagem de variação explicada de $\hat{\sigma}_t^2$ para dados simulados com distribuição Binomial, para 100 experimentos, no modelo inteiramente ao acaso, usando-se dados sem transformação (X) e com as transformações: $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{X / 100}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
041	26	27	26	26
042	24	26	27	27
043	20	21	21	21
044	26	28	29	29
045	22	24	25	25
046	29	31	31	31
047	29	31	31	31
048	23	24	23	23
049	26	28	28	28
050	25	25	25	24
051	28	30	30	30
052	26	29	30	30
053	32	34	34	34
054	25	27	28	27
055	22	23	24	24
056	23	26	28	28
057	24	26	27	27
058	23	24	24	24
059	21	23	24	24
060	21	23	23	23
061	25	27	27	27
062	22	23	23	23
063	21	22	24	24
064	26	24	29	29
065	18	20	21	21
066	25	27	27	28
067	22	24	25	25
068	22	24	35	35
069	24	24	27	27
070	27	29	29	29
071	22	25	27	26
072	23	24	24	24
073	28	31	32	32
074	17	18	18	18
075	23	24	24	24
076	23	25	26	26
077	28	29	29	29
078	25	27	29	29
079	24	25	25	25
080	19	21	22	22

Tabela 7 - Porcentagem de variação explicada de $\hat{\sigma}_1^2$ para dados simulados com distribuição Binomial, para 100 experimentos, no modelo inteiramente ao acaso, usando-se dados sem transformação (X) e com as transformações: $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{X / 100}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
081	22	24	25	25
082	24	25	26	28
083	20	21	22	22
084	24	26	28	28
085	23	24	25	25
086	25	27	28	28
087	23	24	24	24
088	17	18	19	19
089	28	29	30	30
090	25	27	29	29
091	25	26	27	27
092	24	26	27	27
093	22	24	25	25
094	24	24	23	23
095	20	22	23	23
096	27	29	29	29
097	26	28	28	28
098	26	28	29	29
099	25	26	26	26
100	21	22	23	23

Tabela 8 - Porcentagem de variação explicada de $\hat{\sigma}_e^2$ para dados simulados com distribuição Binomial, para 100 experimentos, no modelo inteiramente ao acaso, usando-se dados sem transformação (X) e com as transformações: $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{X / 100}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
001	74	73	72	71
002	71	68	63	68
003	84	83	82	82
004	87	75	69	74
005	76	76	76	76
006	69	68	64	68
007	72	70	67	68
008	77	76	73	75
009	76	75	74	75
010	80	79	78	79
011	71	70	69	69
012	78	76	74	75
013	74	73	71	72
014	76	74	74	73
015	76	76	71	74
016	76	74	73	73
017	72	76	62	67
018	78	76	72	76
019	75	74	71	73
020	82	76	78	78
021	76	74	74	74
022	76	76	69	72
023	66	64	63	62
024	73	71	70	70
025	77	75	75	75
026	80	79	76	78
027	76	74	68	73
028	77	76	75	75
029	77	76	73	77
030	78	78	73	79
031	75	73	71	72
032	79	79	80	80
033	75	74	73	73
034	77	76	75	75
035	81	79	74	78
036	74	74	74	74
037	75	74	74	75
038	75	75	76	76
039	75	74	73	73
040	77	76	75	75

Tabela 8 - Porcentagem de variação explicada de $\hat{\sigma}_e^2$ para dados simulados com distribuição Binomial, para 100 experimentos, no modelo inteiramente ao acaso usando-se dados sem transformação (X) e com as transformações: $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{X / 100}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
041	74	73	74	74
042	76	74	73	73
043	80	79	75	79
044	74	72	71	71
045	78	76	75	75
046	71	69	69	69
047	71	69	69	69
048	77	76	77	77
049	74	72	72	72
050	75	75	75	76
051	72	70	67	70
052	74	71	68	70
053	68	66	66	66
054	75	73	72	73
055	78	77	74	76
056	77	74	71	72
057	76	74	73	73
058	77	76	75	76
059	79	77	75	76
060	79	77	77	77
061	75	73	73	73
062	78	77	76	77
063	79	78	76	76
064	74	76	71	71
065	82	80	79	79
066	75	73	73	72
067	78	76	74	75
068	78	76	65	65
069	76	76	68	73
070	73	71	71	71
071	78	75	73	74
072	77	76	71	76
073	72	69	67	68
074	83	82	82	82
075	77	76	76	76
076	77	75	72	74
077	72	71	71	71
078	75	73	71	71
079	76	75	75	75
080	81	79	78	78

Tabela 8 - Porcentagem de variação explicada de $\hat{\sigma}_e^2$ para dados simulados com distribuição Binomial, para 100 experimentos, no modelo inteiramente ao acaso, usando-se dados sem transformação (X) e com as transformações: $\sqrt{X + 0,5}$, $\log(X + 0,5)$ e $\arcsen \sqrt{X / 100}$.

N	X	$\sqrt{X + 0,5}$	$\log(X+0,5)$	$\text{Arcsen} \sqrt{\frac{X}{100}}$
081	78	76	75	75
082	76	75	74	72
083	80	79	78	78
084	76	74	72	72
085	77	76	71	75
086	75	73	71	72
087	77	76	74	76
088	83	82	81	81
089	72	71	68	70
090	75	73	66	71
091	75	74	73	73
092	76	74	73	73
093	78	76	72	75
094	76	76	76	77
095	80	78	76	77
096	73	71	67	71
097	74	72	72	72
098	74	72	71	71
099	75	74	74	74
100	79	78	77	77

5- CONCLUSÕES

Verifica-se pelos resultados da simulação feita aqui que:

1- Quando se é possível comprovar na prática que os dados experimentais têm distribuição de Poisson, as proporções de variação explicada não diferem muito na presença de transformações. Porém essas se mostraram necessárias uma vez que as variâncias não são homogêneas acarretando problemas nos testes da análise de variância. Concluimos também que a transformação recomendada neste caso é a Raiz Quadrada que mostrou ter um melhor desempenho na proporção de variação explicada em relação à transformação Logarítmica.

2- Quando os dados experimentais têm distribuição Binomial, as proporções de variação explicada também não diferem muito na presença de transformações. Neste caso, a transformação Logarítmica não mostrou nenhuma diferença quanto à proporção de variação explicada em relação à transformação arco seno, portanto não sendo necessária.

3- Em face dos resultados obtidos podem-se usar em melhoramento genético as estimativas dos componentes de variância oriundas dos dados não transformados.

5-REFERÊNCIAS BIBLIOGRÁFICAS

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

ANSCOMBE, F.J. The transformation of poisson, binomial and negative binomial data. *Biometrika*, Cambridge, 35: 246-254, 1948.

BARBOSA, J.C. O problema da heterocedasticidade na análise da variância. *Seminários de Estatística e Experimentação Agronômica*, Piracicaba, ESALQ/USP. 14p. (Mimeografado), 1983.

BARTLETT, M.S. The square root-transformation in the analysis of variance. *Journal of the Royal Statistical Society Supplement*, Londres, 3(1): 68-78, 1936

BARTLETT, M.S. The use of transformations. *Biometrics*. Raleigh, 3: 39-52, 1947.

- BENNETT, C.A. & FRANKLIN, N.L. Statistical analysis in chemistry and the chemical industry, Nova Iorque, John Wiley & Sons, 1954. 724p.
- BOX, G.E.P. & D.R. COX. An analysis of transformations. *Journal of the Royal Statistical Society*. Londres, 26: 221-243, 1964.
- COCHRAN, W.G. Some consequences when the assumptions for the analysis of variance are not satisfied. *Biometrics*, Raleigh, 3: 22-38, 1947.
- CORBEIL, R.R. & S.R. SEARLE. Restricted Maximum Likelihood (REML) estimation of variance components in the mixed model. *Technometrics*, Richmond, 18: 31-38, 1976a.
- CORBEIL, R.R. & S.R. SEARLE. A comparison of variance component estimators. *Biometrics*, Nova Iorque, 32: 779-791, 1976b.
- DEMETRIO, C.G.B. Transformação de dados. Efeitos sobre a análise da variância. Piracicaba, 1978. 113p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).

GRAYBILL, F. A. & HULTGUIST, R.A. Theorems concerning eisenhart's Modell II. *Annals of Mathematical Statistics*, Colorado. 32: 261-269, 1961.

HENDERSON, C.R. Estimation of variance and covariance components. *Biometrics*, Richmond, 9: 226-252, 1953.

HERBACH, L.H. Properties of Modell II - Type analysis of variance tests, a: optimum nature of the F-test for Modell II in the balance case. *Annals of Mathematical Statistics*, Colorado, 30: 939-959, 1959.

HICKS, C.R. Fundamental concepts in the design of experiments. Nova Iorque, Holt, Rinehart and Winston, 1964, 349p.

HOAGLIN, D.C., F. MOSTELLER & J.W. TUKEY. Understanding robust and exploratory data analysis. Nova Iorque Wiley, 1982, 447p.

LE CLERG, E.L., W.H. LEONARD & A.G. CLARK. Transformation of experimental data. Field plot technique. Burgess Publishing Co., Minneapolis, 1962, 270p.

- PARENTE, R.C.P. Aspectos da análise de resíduos. Piracicaba, 1984. 139p. (Mestrado - Escola Superior de Agricultura "Luiz de Queiroz"/USP).
- PATTERSON, H.D. & R. THOMPSON. Recovery of interblock information when block sizes are unequal. *Biometrika*, Cambridge, 58: 545-554, 1971.
- RAO, C.R. Estimation of heterocedastic variances in linear models. *Journal of the American Statistical Association*, Boston, 65: 161-172, 1970.
- RAO, C.R. Estimation of variance and covariance components - MINQUE theory. *Journal of the Multivariate Analysis*, Nova Iorque, 1: 257-275, 1971a.
- RAO, C.R. Minimum variance quadratic unbiased estimation of variance components. *Journal of the Multivariate Analysis*, Nova Iorque, 1: 445-456, 1971b.
- RAO, C.R. Estimation of variance and covariance components in linear models. *Journal of the American Statistical Association*, Boston, 67: 112-115, 1972.
- RUSSELL, T.S. & R.A. BRADLEY. One-way variances in a two-way classification. *Biometrika*, Cambridge, 45: 111-129, 1958.

SAS Institute Inc. SAS Language Guide for Personal Computers, Version 6 Edition, Cary. NC, SAS Institute Inc, 1987, 429p.

SAS Institute Inc. SAS/STAT Guide for Personal Computers, Version 6 Edition. Cary NC, SAS Institute Inc., 1987, 1028p.

SCHLESSELMAN, J.J. Data transformation in two-way analysis of variance. *Journal of the American Statistical Association* Washington 68(342): 369-377, 1971.

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

SEARLE, S.R. Topics in variance component estimation. *Biometrics* Richmond 27: 1-76, 1971.

SINGHAL, R.A. Comparison of variance components for non-normal samples. *Sankhyā B*, 48: 167-175, 1986.

STEEL, R.G.D. & J.H. TORRIE. principles and procedures of statistics a biometrical approach. Segunda edição MacGraw-Hill International Book Company, Singapore, 1980, 481p.

THOMPSON, W.A. The problem of negative estimates of variance components. *Annals of Mathematical Statistics*. Florida, 33: 273-289, 1962.