



Data transformation in biological assays

Sylvio Péllico Netto¹, Alexandre Behling¹

¹Universidade Federal do Paraná, Departamento de Ciências Florestais, Avenida Pref. Lothário Meissner, 900, Jardim Botânico, CEP 80210-170, Curitiba, PR, Brazil

***Corresponding author:**

alexandre.behling@yahoo.com.br

Index terms:

Analysis of variance
Acacia mearnsii
Radiation

Termos para indexação:

Análise de variância
Acacia mearnsii
Radiação

Received in 24/03/2019
Accepted in 23/11/2020
Published in 31/03/2021

Abstract - The analysis of variance is the statistical test most used for comparison of three or more means simultaneously. Its application requires, however, the compliance to some assumptions, with main emphasis on normality of the data and homoscedasticity of variances. When such requirements are not met, one of the alternatives is the data transformation to enable the continuity of the experimental evaluation. With the proposition of the Tukey's data transformation system, understood as a power transformation system, i.e. the application of n^{th} root on a data set ($X^{1/n}$) this statistical procedure has methodologically evolved to ensure such solutions. In the present research we proposed a complement to this system, denominated here as transformation in four steps, with inclusion of two hypothesis tests to evaluate normality and homoscedasticity. This was applied on experimental data to evaluate the amount of radiation available at soil level within stands of *Acacia mearnsii* De Wild. We have proposed a model for data transformation to simultaneously obtain homoscedasticity and normality. The methodology was appropriate to ensure these two statistical aspects on the experimental data, allowing comparison of eight treatments by conventional analysis of variance.

Transformação de dados em ensaios biológicos

Resumo - A análise de variância é o teste estatístico mais utilizado para a comparação de três ou mais médias simultaneamente. Sua aplicação exige, no entanto, o cumprimento de algumas condicionantes, com ênfase principal na normalidade dos dados e homoscedasticidade das variâncias. Quando tais requisitos não são atendidos, uma das alternativas é a transformação de dados para permitir a continuidade da avaliação experimental. Com a proposição do sistema de transformação de dados de Tukey, entendido como um sistema de transformação de potência, ou seja, a aplicação de $n^{\text{ésima}}$ raiz em um conjunto de dados ($X^{1/n}$), este procedimento estatístico evoluiu metodologicamente para garantir tais soluções. No presente trabalho foi proposto um complemento a esse sistema, denominado aqui de transformação em quatro passos, com a inclusão de dois testes de hipóteses para avaliar a normalidade e homoscedasticidade. Isto foi aplicado em dados experimentais para avaliar a quantidade de radiação disponível ao nível do solo dentro de povoamentos de *Acacia mearnsii* De Wild. Um modelo para transformação de dados foi proposto para obter simultaneamente homoscedasticidade e normalidade. A metodologia foi apropriada para garantir esses dois aspectos estatísticos nos dados experimentais, permitindo a comparação de oito tratamentos pela análise de variância convencional.



Introduction

Parametric and non-parametric statistical methods applied to biological experiments are widely used. The most applied are the Student's t-test, when comparing only two means and the analysis of variance (ANOVA), when three or more means are compared. It must be emphasized that the execution of these requires attendance to certain premises. It is assumed that the experimental errors associated with the treatments are independent, normally distributed and with homoscedasticity of variances (Cochran, 1947).

Still, the best choice to measure the selected variables must be appropriately chosen to ensure the functionality of the linear additive model (Steel & Torrie, 1960). The randomization of treatments that integrate the experiment should be applied to ensure the independence of experimental errors.

Homoscedasticity of variances is essential, because when ANOVA is applied to test the null hypothesis of equality between means of the experimental samples taken from a population, it is calculated the overall arithmetic mean, which means they are estimators of a single parameter, i.e., there is a single population from which the different treatments are only repeated samples taken from a unique structural set of data.

In some biological populations, asymmetry occurs on the data distribution with relative frequency and, in this case, the lack of normality can also generate heteroscedasticity. This statistical condition is referenced by Steel & Torrie (1960) as a regular type of heteroscedasticity, which appears owing to the variability of the data showing pronounced asymmetries in relation to their averages. To detect the statistical significance for unequal variances at 95% or 99% of probability, it is possible to satisfy the assumption of homoscedasticity with data transformation, which will propitiate a new scale with approximation of them to normality. This transformation is a way to make the means and variances independents and, therefore, their homogeneity.

Even though data transformation has made it possible in most cases, adverse situations of heteroscedasticity of variances in addition to no achievement of normality for experimental treatments may remain, which require special considerations. For these cases, Steel & Torrie (1960) recommend the use of other methods of statistical analysis, such as the application of weighted ANOVA as presented in Rahman & Sen (2009).

The concept of data transformation gained expressiveness with the proposal formalized by Tukey (1957), when he focused on a single system as alternative to transform data in order to reach normality and homoscedasticity, based on what he called power transformation. Later, Box & Cox (1964) improved the Tukey's system and diversified it with special cases like indeterminacies and other situations as: negative logarithmic values and occurrences of zeros in the dataset. The authors' experience was an important step for realizing that in adverse circumstances in datasets was not possible achieving normality and homoscedasticity simultaneously. These situations motivated the author's proposal for a new data transformation to solve the above referenced adverse conditions.

The main aim of the present research was to propose a complement to the Tukey's data transformation system, denominated by the authors as transformation in four steps, with addition of two hypothesis tests to evaluate normality and homoscedasticity simultaneously.

Material and methods

Concepts on data transformation

Consider a random variable X with normal distribution $N(\mu_x, \sigma_x)$. Under these circumstances, if an ANOVA test is applied to the data set, the result of the F test is expected to test H_0 for equality of means of T experimental treatments sampled in the population. However, if the random variable X deviates from normality, i.e., the distribution presents asymmetry, then this condition will be transferred to the distribution of experimental errors and will produce more significant results in the application of the F test than expected, besides the loss of efficiency in the ANOVA. Under these circumstances, the mean of each treatment may not estimate the mean of the entire population (Snedecor & Cochran 1967). When taking samples in populations, two conditions must be observed: 1) The occurrence of non-normality, which can be easily evaluated by the application of statistical tests, such as Kolmogorov-Smirnov (Kolmogorov, 1933; Smirnov, 1948), when the number of replications is sufficient; 2) The heteroscedasticity of variances, which can be easily detected by the application of the statistical tests of Bartlett (1937) or Hartley (1950).

With respect to non-normality, data transformation has been an alternative technique to solve the problem, because the probability distribution of the errors usually

is not known, which turns the ANOVA a complex problem to be solved. Alternatively, the application of a generalized linear model is also a recommended technique to be used. In the context of this manuscript the transformative procedure, as proposed by Snedecor & Cochran (1967), i.e., the variable Y represents the set of transformed data obtained as a function of the variable X , whose asymmetry has been statistically pre-detected. Then:

$$Y = f(X) \quad (1)$$

Aiming at achieving normality for the Y variable and considering that the data are measurable in a continuous scale, then $Y > 0$. By the application of a mathematical function to each of the data comprising the replications of the treatments in test, then (Equation 1) will be equated as:

$$Y_i = f(X_i) \quad (2)$$

According to McDonald (2009), many biological variables do not meet the assumptions of a parametric statistical test: non-normality of treatment data and no homogeneity of variances, or both. Using a parametric statistical method, such as ANOVA, data in these circumstances may give a misleading result. In some cases, transforming data will be an alternative to attend the assumptions of the ANOVA. According to Osborne (2002), data transformations are commonly used tools that can serve many functions in quantitative analysis of data. The three data transformations most applied in statistical analysis for improving the normality of variables are: square root, log, and inverse of the treatment values. Ramsey & Shafer (2012) also pointed out that there are useful transformations for positive measurements with skewed distributions, where the means and standard deviations differ between groups, for example, the following transformations: square root, reciprocal, arcsine of square root, logit and logarithmic of treatment values.

Under a variety of circumstances when conducting field experiments, we have detected the need to make data transformations. According to Parsad (2005), Baran & Warry (2008), McDonald (2009) and Fowler et al. (2009), the most common transformations used in biological data are: logarithmic, square root and arcsine. The application of logarithmic transformation is suitable for data where the variance is proportional to square of the mean or the coefficient of variation (ratio between

standard deviation and mean) is constant or where effects are multiplicative (Parsad, 2005). The application of logarithmic transformation is performed by applying the base 10 or the neperian base.

Square root transformation is appropriate for data sets where the variance is proportional to the mean (Parsad, 2005). The application of square root to the data set is a special case of what can be nominated as power transformation, i.e., in which the square root (n^{th} root on a set of data i.e., $X^{1/n}$) is the special case when $n = 2$.

Trigonometric transformations (arcsin) are also applied to special situations in which the dataset occurs only in the range $0 \leq X \leq 1.0$, where $Y_i = f(\arcsin X_i)$. According to Parsad (2005), this transformation is appropriate for proportions, i.e., data obtained from count or data expressed as decimal fractions and percentages. The distribution of percentages is binomial, and its transformation turn to be a normal distribution.

Another possibility is the inverse transformation, or reciprocal of data, which is proper to situations where part of them is in the range $0 < X \leq 1.0$, and this transformation works favorably to approximate the data to the unit $Y_i = f(X_i^{-1})$. Another part, i.e., for the condition when $X \geq 1.0$, it is desirable application of the reverse of the inverse transformation which is very effective, or $Y_i = f(1 - X_i^{-1})$.

Tukey (1957) agglutinated the various data transformations in a proposal called by him power transformation, in which the transformed data result from the application of a monotonic function to observed data, expressed as:

$$\begin{aligned} Y_i &= X_i^\lambda & \text{if } \lambda &\neq 0 \\ &= \log X_i & \text{if } \lambda &= 0 \end{aligned} \quad (3)$$

This transformation, understood as a system for transforming data, was later modified by Box & Cox (1964) to eliminate the condition of discontinuity occurring when $\lambda = 0$. They have proposed the modified Tukey's system, i.e.:

$$\begin{aligned} Y_i &= (X_i^\lambda - 1) / \lambda & \text{if } \lambda &\neq 0 \\ &= \log X_i & \text{if } \lambda &= 0 \end{aligned} \quad (4)$$

This transformation is valid only for $X_i > 0$.

Evolutionarily, the main contribution to improve the Tukey's data transformation system was performed by Box & Cox (1964). These authors suggested changes applicable in the case of occurrence of negative values and the system, as shown in (Equation 4), came to be expressed by:

$$\begin{aligned}
 Y_i &= (X_i + \lambda_2)^{\lambda_1} / \lambda_1 & \text{if } \lambda_1 \neq 0 \\
 &= \log(X_i + \lambda_2) & \text{if } \lambda_1 = 0
 \end{aligned}
 \tag{5}$$

Where λ_1 is the transformation parameter and λ_2 is a parameter chosen to satisfy the condition $X_i > 0$.

Other situations proposed to modify the Tukey's system were presented in Sakia (1992), involving the following suggestions: a) when negative values occur in the data (Manly, 1976); b) when the data set presents little asymmetry, named modular transformation (John & Draper, 1980); and c) when the probability distribution of the data is known and is asymptotic in the abscissa axis, as in the normal distribution (Bickel & Doksum, 1981).

The logarithmic transformation was used as the first step, as suggested by Parsad (2005), and, consequently, $\lambda_1 = 0$ in the Box and Cox transformation. After the logarithmic transformation, not always it is achieved the conditions of normality and homoscedasticity of variances. Consequently, a new transformation should be performed to ensure both conditions simultaneously.

In some conditions, this requirement may not be reached and, therefore, one of the assumptions for the application of ANOVA is not attended. Consequently, the calculation of the residuals will be affected, which will increase the chance for the occurrence of errors of type I and II, i.e., to accept the null hypothesis when it is false or vice-versa.

This problem was analyzed by Box & Cox (1964), who have proposed the method of maximum likelihood and the Bayesian solution to estimate the transformation's λ parameter, however later Draper & Cox (1969) concluded that such procedures are robust to achieve normality only when one can get a reasonable symmetry in their discrepancies, however not always such a transformation can also result in homoscedasticity (Zarembka, 1974). The same author evaluated the statistical implications for the estimation of the λ parameter and said there is bias in its obtaining. From there, he went on to analyze the relationship between the variance and the mean and picked up what was defined in (Equation 1), for the population of the variable X , for what he considered the approximation of first order of Taylor series to better understand this relationship.

Proposed transformations

In the process of data transformation, several attempts were applied following the suggestions made by the authors previously mentioned, in which the transformed data of the distributions should, as far as

possible, approximate to the unit. Such a condition is more obviously achieved with the application of n^{th} root on a set of data i.e. $(X^{1/n})$, since both data ≥ 1.0 as well as ≤ 1.0 will tend to one, but the results did not always have proved effective, because the uniformity effect occurred proportionally, and the variances remained with heteroscedasticity.

Consider, after we have performed the logarithmic transformation to reduce the scale of the data, that we could additionally apply a second transformation to achieve normality and homoscedasticity. This proposal consists in the formulation of three new hypotheses as consequence:

Hypothesis 1: It is possible to seek rapprochement between the ranges of variation of data in t experimental treatments, which could be achieved when their extreme values converge to equality.

If this convergence is achieved, the ranges of variation will change from a condition of maximum dispersion to a condition of lower dispersion, approaching symmetry, i.e. it will fluctuate along a rectangular uniform trend, whose extremes will be parametrically equal. In these circumstances, the other ranges of variation will oscillate around their arithmetic mean. Such a situation, not considering the simultaneous condition of normality and homoscedasticity, can be forced with the formulation of a second hypothesis:

Hypothesis 2: It is possible to find the convergence of other ranges of variation by applying a third transformation on the original data, maintaining the double transformation previously performed, aiming to bring them as close as possible of their mean.

If these three transformations can force homoscedasticity, there still may remain non-normality in the population of data triply transformed. In these circumstances a fourth transformation is proposed until normality is achieved, maintaining though the previous transformations already performed.

Thus, we propose the formulation of the third hypothesis:

Hypothesis 3: Normality will be reached when the effect of a fourth transformation results in a new population with asymmetry close to zero.

To test the first hypothesis (HGV), consider the ranges of variation of " t " treatments, i.e. $(X_{jmax} - X_{jmin}) = GV_j$, i.e. GV is the amplitude of each treatment in which $j = 1, 2, 3 \dots t$. Proceed to the ordination of these series in ascending order.

Consider additionally the parameter δ , generator of homoscedasticity, which will be achieved for each two treatments when:

$$GV_{j+1} - GV_j = 0$$

This fact is equivalent to the formulation of the following null hypothesis:

$$H_0: GV_{j+1} = GV_j$$

The alternative hypothesis is therefore:

$$H_1 = GV_{j+1} \neq GV_j$$

In the circumstances in which we shall compare “t” treatments, we propose that such hypotheses be formulated for the extreme values of the transformed series:

$$H_0: GV_{max} = GV_{min}$$

$$H_1: GV_{max} \neq GV_{min}$$

This hypothesis test will be conducted by considering the variable Y_i as a result of the first two transformations, i.e. first the logarithmic to achieve normality or quasi-normality [$\log(GV)$] and the second to achieve homoscedasticity, applying to it the inverse of the ranges of variation [$(GV)^{-1}$].

$$Y = \log(GV) [(GV)^{-1}]^\delta \quad (6)$$

Assuming their equality in the extreme ranges of variation, as is presented in the null hypothesis, we can obtain the estimator for the transformer parameter $\hat{\delta}$:

$$\log GV_{max} [(GV_{max})^{-1}]^\delta = \log GV_{min} [(GV_{min})^{-1}]^\delta \quad (7)$$

By isolating $\hat{\delta}$ we have:

$$\hat{\delta} = \log \left[\log GV_{min} (\log GV_{max})^{-1} \right] \left\{ \log \left[GV_{min} (GV_{max})^{-1} \right] \right\}^{-1} \quad (8)$$

We should apply the double transformation to all sampling data of the variable X , for acceptance of H_0 :

$$Y_{li} = [(X_i^{\lambda_1} - 1) \lambda_1^{-1}] [(X_i)]^\delta = \log X_i (X_i)^\delta \quad (9)$$

Where $\lambda_1 = 0$ and $\hat{\delta}$ equal to the value obtained in Equation 8.

When analyzing the results of this double transformation we realized that homoscedasticity was achieved, but normality remained only approximate

(evaluated with the Kolmogorov-Smirnov’s test). This resulted in the formulation of the second hypothesis to achieve greater approximation of ranges of variation in Equation 10.

$$Y_{2i} = [(X_i^{\lambda_2} - 1) \lambda_2^{-1}] [(X_i^{\lambda_2})]^\delta = \log X_i (X_i^{\lambda_2})^\delta \quad (10)$$

Where $\lambda_1 = 0$, λ_2 to be determined by applying the adjustment of a polynomial of n^{th} degree resulting from successive sampling points generated by the assessment of the chi-square test, equated as a function of λ_2 in a range that contains the most appropriate possible transformations and $\hat{\delta}$ = value obtained in Equation 8.

Consider, in addition, the parameter λ_3 , generator of normality, obtaining it by forcing the triple transformation result in a normal distribution with symmetry (S) equal or close to zero.

This additional condition is equivalent to the formulation of the following null hypothesis (HS):

$$H_0: S = 0$$

The alternative hypothesis is therefore:

$$H_1: S \neq 0$$

To test such hypothesis, we finally propose a fourth transformation, incorporating the parameter λ_3 to generate normality, and the transformed variable obtained in Equation (11).

$$Z_i = (\log X_i)^{\lambda_3} [(X_i)^{\lambda_2}]^\delta \quad (11)$$

The hypothesis test was conducted by taking successive values for the parameter λ_3 , in an interval that contains the most appropriate possible transformations, such that the population of transformed data Z_i reaching the optimal condition of normality and homoscedasticity, i.e. when both Kolmogorov-Smirnov’s and Bartlett’s tests applied to this population result in no statistical significance, the first at 95% and the second at 99% probability levels. For both, the successive values of symmetry (S) are prospected in function of λ_3 , i.e. $S = f(\lambda_3)$. As homoscedasticity and normality behave in a reverse manner, the final solution will be found for the value of S that meets the acceptance of the null hypotheses.

In the context of acceptance of H_0 for the second hypothesis of this research, we chose a λ_3 value, which generates asymmetry as close as possible to zero.

Application on experimental data

The variable to be used for describing the problem of non-normality and heteroscedasticity of variances is the incident radiation on ground level (IRGL).

The quantity of radiation available at ground level within a population is a function of a number of factors, such as: morphometric and physiological characteristics, growth habit, leaf incidence angle and arrangement of leaves of the species that compose the forest plantation. In addition, photosynthetically active radiation incidence (PARI) is also determined by canopy characteristics, as well as the location and size of existing spaces on the canopy, commonly called sun flecks. The light which passes through the canopy of the trees provokes points of sunshine at the soil level with different intensity of radiation, producing a heterogeneous environment of light incidence. In addition, this variable does not present normal distribution, because the points with lower intensities of radiation at ground level tend to be more frequent than points with higher intensities of occurrence, causing a positive asymmetric distribution.

The analysis was carried out to verify the effect of the environment and of the age of the black wattle (*Acacia mearnsii* De Wild.) stands on the incidence of photosynthetically active radiation on ground level. Several commercial plantations of black wattle were selected from regions of high concentration of the species in the municipalities of Cristal (Hillside of the Southeast - HS) and Piratini (Mountain of the Southeast - MS), Rio Grande do Sul State, Brazil.

In each municipality the stands were selected in a range of ages: one, three, five and seven years old. In Cristal, the evaluated areas are located between coordinates 30°54' S and 50°40' W and in Piratini, between 31°25' S and 52°58' W, all in the altitude of 320 to 370 m above sea level.

In each stand, a north-facing hillside was chosen, and 45 sample plots were used as replications of the experiment for the assessment of photosynthetically active radiation (PAR), for which five measurements were taking in each point and calculated the average of them.

The readings of PARI were carried out at 30 cm height from the surface using a pyranometer (LI-COR 1600) in $\mu_{\text{mol}} \text{ cm}^{-2} \text{ s}^{-1}$. The experimental design was a factorial completely randomized design, in which the treatments were assigned as: the environmental condition

(HS and MS) and age of the stands (one, three, five and seven years old), organized in 45 replications.

The errors in the ANOVA model, because of no controlled factors, should be normally distributed with a unique parametric variance, i.e. all treatments present homogeneous variances. This implies that the experimental data should also present a normal distribution. After the evaluation at 95% probability level for the Kolmogorov-Smirnov's test, and 99% probability level for the Bartlett's test, it was found statistical significance for both conditions. Then it was necessary to proceed the application of the transformations envisaged in Equation 11. We also used an analysis of variance to verify the occurrence of errors of type I and II. All statistical tests were conducted using SAS software for academics and Julia language.

Results

Implementation of the proposed transformations

The ratios were calculated between the standard deviation and the average of each treatment, and was observed approximating condition of proportionality of these results (Figure 1).

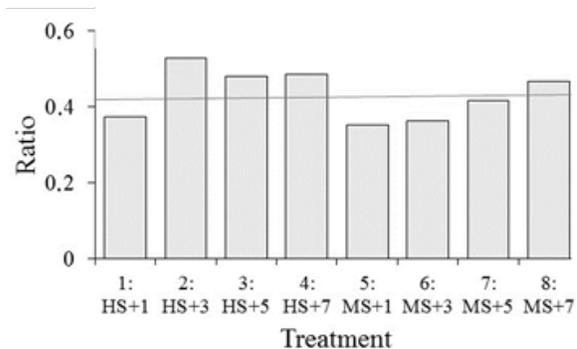


Figure 1. Ratio between the standard deviation and the mean of the treatment (for the variable radiation photosynthetically active incident on ground level) from an experiment conducted in black wattle stands in the state of Rio Grande do Sul, Brazil (HS = Hillside of the Southeast, MS = Mountain of the Southeast. Four stand ages: 1, 3, 5 and 7 years).

Applying the transformation to the data, the p value that was $-0.48 \approx -0.5$ have approximated to zero, while the logarithmic transformation resulted in values near the unit, which is the initial premise to be performed in all transformations to achieve normality and homoscedasticity of variances.

The maximum and minimum values, the ranges of variation, the means, and the variance of each treatment are shown in Table 1. The largest variance was 35,885.71 and the smallest 8,617.77, i.e. representing a ratio of 4.16. The lower range of variation was present in treatment 6, while the largest in treatment 1 (Figure 2). The Bartlett's

test resulted in a value of $X^2=48.69$, therefore significant at 99% probability. The asymmetry value was 0.53, tending toward a moderate condition (Figure 2); the Kolmogorov-Smirnov's test was significant, confirming the non-normality of the data.

Table 1. Statistics after the application of the transformation to the variable (radiation photosynthetically active incident on ground level) from an experiment conducted in black wattle stands in the state of Rio Grande do Sul, Brazil.

Transformation	Treatment	Maximum Value	Minimum Value	Range of variation	Mean	Variance
0	1: HS+1	931.40000	112.20000	819.20000	494.30222	34,258.96113
0	2: HS+3	700.20000	64.80000	635.40000	322.25778	29,108.29931
0	3: HS+5	870.40000	88.20000	782.20000	393.31111	35,885.70646
0	4: HS+7	557.40000	68.20000	489.20000	256.06800	15,447.01277
0	5: MS+1	748.40000	137.35600	611.04400	432.62996	23,482.14578
0	6: MS+3	434.80000	91.20000	343.60000	254.70222	8,617.77386
0	7: MS+5	705.60000	98.80000	606.80000	374.50658	24,344.55462
0	8: MS+7	549.60000	56.20000	493.40000	246.53333	13,346.14545
1	1: HS+1	2.96914	2.04999	0.91914	2.65667	0.03865
1	2: HS+3	2.84522	1.81158	1.03365	2.43153	0.08030
1	3: HS+5	2.93972	1.94547	0.99425	2.53625	0.05942
1	4: HS+7	2.74617	1.83378	0.91238	2.34928	0.05955
1	5: MS+1	2.87413	2.13785	0.73629	2.60535	0.03003
1	6: MS+3	2.63829	1.95999	0.67829	2.37348	0.03142
1	7: MS+5	2.84856	1.99476	0.85380	2.52689	0.04796
1	8: MS+7	2.74005	1.74974	0.99031	2.33664	0.05616
2	1: HS+1	1.02023	0.97887	0.04136	1.01643	0.00006
2	2: HS+3	1.02023	0.94268	0.07755	1.00487	0.00047
2	3: HS+5	1.02023	0.96464	0.05559	1.01152	0.00019
2	4: HS+7	1.02018	0.94662	0.07356	1.00163	0.00044
2	5: MS+1	1.02023	0.98899	0.03124	1.01602	0.00004
2	6: MS+3	1.01898	0.96676	0.05222	1.00580	0.00017
2	7: MS+5	1.02023	0.97165	0.04858	1.01208	0.00015
2	8: MS+7	1.02016	0.93103	0.08913	1.00103	0.00047
3	1: HS+1	12.04319	5.39030	6.65289	9.39690	2.13600
3	2: HS+3	10.88551	4.25683	6.62867	7.83109	3.32116
3	3: HS+5	11.75959	4.86942	6.89017	8.52817	2.87664
3	4: HS+7	10.02702	4.35439	5.67263	7.23935	2.24475
3	5: MS+1	11.14707	5.85910	5.28797	8.97635	1.63070
3	6: MS+3	9.15531	4.93950	4.21580	7.33744	1.25556
3	7: MS+5	10.91543	5.11020	5.80523	8.43222	2.24903
3	8: MS+7	9.97584	3.99335	5.98248	7.15048	2.05472
4	1: HS+1	1.77274	1.52443	0.24830	1.66938	0.00311
4	2: HS+3	1.72710	1.49796	0.22913	1.61182	0.00419
4	3: HS+5	1.76157	1.51055	0.25102	1.63672	0.00397
4	4: HS+7	1.69326	1.49962	0.19364	1.58943	0.00268
4	5: MS+1	1.73742	1.53864	0.19878	1.65285	0.00238
4	6: MS+3	1.65910	1.51228	0.14683	1.59134	0.00160
4	7: MS+5	1.72828	1.51667	0.21160	1.63271	0.00309
4	8: MS+7	1.69125	1.49430	0.19695	1.58615	0.00239

Where: Two environments: HS = Hillside of the Southeast, MS = Mountain of the Southeast. Four stand ages: 1, 3, 5 and 7 years. Transformations: 0: $Y_i = X_i$, 1: $Y_i = \log X_i$, 2: $Y_i = \log X_i(X_i)^{-0.1566}$, 3: $Y_i = \log X_i [(X_i)^{-1.30786}]^{-0.1566}$ and 4: $Y_i = (\log X_i)^{-0.785} (X_i)^{0.2087}$.

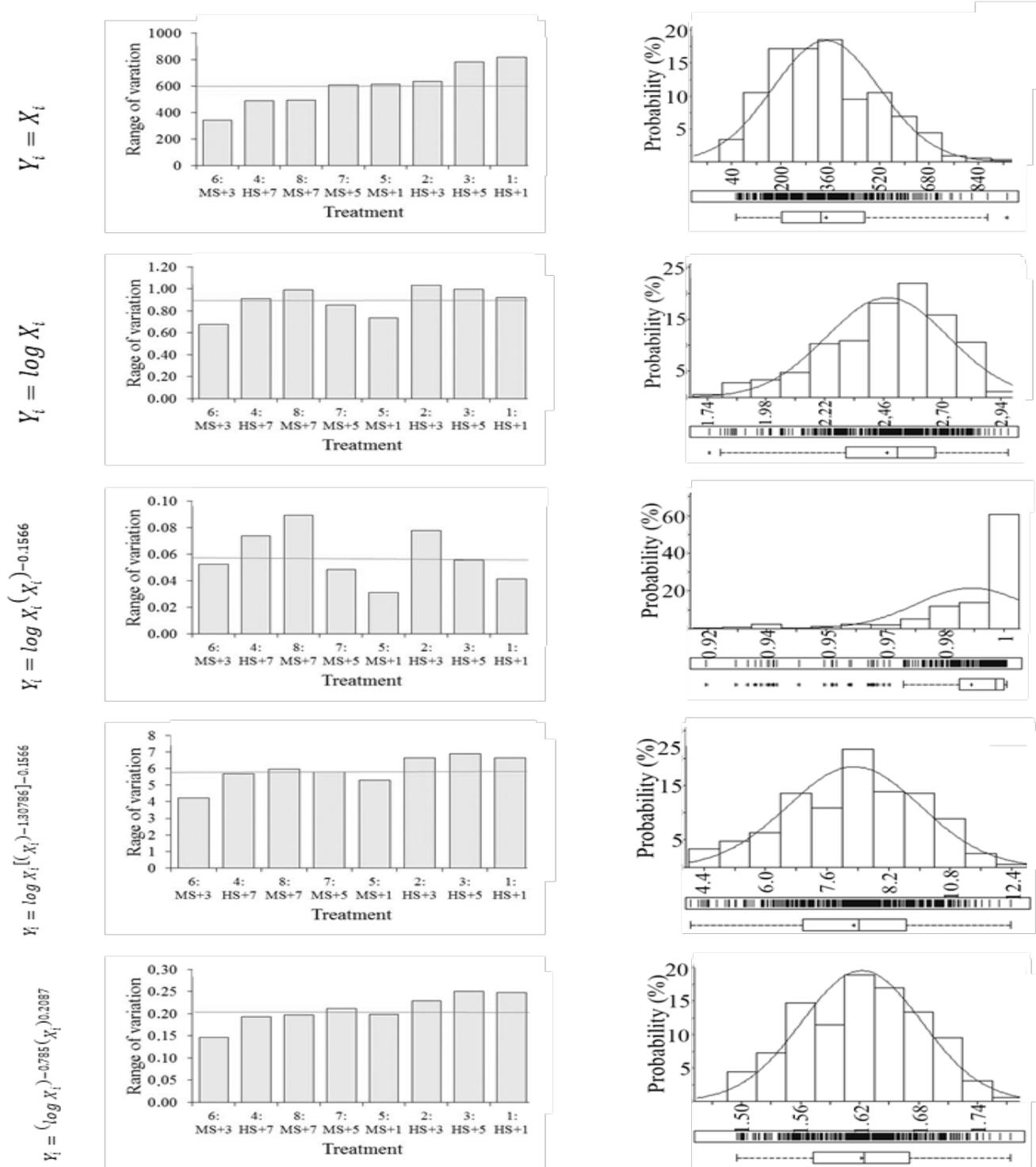


Figure 2. Ranges of variation and probability distributions after applying the transformations. The variable is incident radiation at ground level in stands of black wattle under the effect of two environments (HS = Hillside of the Southeast, MS = Mountain of the Southeast. Four stand ages: 1, 3, 5 and 7 years).

At first, the data were transformed to $Y_i = \log X_i$. The statistical summary is presented in Table 1. The application of the Bartlett's test resulted in a chi-square value of $\chi^2 = 18.21$, significant at 95% probability. As far as normality is concerned, the application of the Kolmogorov-Smirnov's test resulted in a value of 0.0702, therefore normality was rejected. The value of asymmetry was -0.65, tending to a moderate condition (Figure 2). Considering such circumstances, we decided to apply the transformations proposed by the authors (Equation 11).

Assuming the equality of the extreme ranges of variation, as specified in the null hypothesis presented in (7), we have :

$$\log 819.20 [(819.20)^{-1}]^{\delta} = \log 343.60 [(343.60)^{-1}]^{\delta} \quad (12)$$

By isolating δ we have:

$$\delta = \log \left[\log 343.60 (\log 819.20)^{-1} \right] \left\{ \log \left[343.60 (819.20)^{-1} \right] \right\}^{-1} \quad (13)$$

Therefore,

$$\hat{\delta} = 0.1566 \quad (14)$$

For the acceptance of H_0 (HGV), double transformation was applied to all sampling data of the variable X , as presented in (Equation 9).

$$Y_{1i} = \log X_i (X_i)^{-0.1566} \quad (15)$$

After processing the data, the results of the Bartlett's and Kolmogorov-Smirnov's tests remained significant, revealing that the data still presented heteroscedasticity and non-normality. Although transformation (Equation 15) resulted in lower ranges of variation, the probability distribution of the data worsened, emphasizing an asymmetric distribution. This detracted in reaching the goal of homoscedasticity.

Thus, it was proposed the inclusion of the coefficient λ_2 , in order to force the values which are distanced from the mean, consequence of the performed transformation, by the introduction of the λ_2 parameter, to go back to an approximation of the mean. Therefore, the new adjusted transformation was:

$$Y_{2i} = (\log X_i) [(X_i)^{\lambda_2}]^{\delta} \quad (16)$$

The third step was the application of the second hypothesis test (HS), in order to take successive values for the parameter λ_2 in the Equation 11, such that the population of transformed data would reach the optimal condition of homoscedasticity, i.e. when the Bartlett's

test applied to this population result in no statistical significance at 99% probability level.

The successive values of Chi-square (χ^2) in the Bartlett's test were related in function of λ_2 , i.e. $\chi^2 = f(\lambda_2)$, for a range between $-2.5 \leq \lambda_2 \leq -0.5$. After pairs of values had been collected, a second-degree polynomial was adjusted to the sequence of results obtained between the specified interval for the variable λ_2 (Figure 3).

To obtain the minimum value of the Chi-square in the Bartlett's test, the first derivative of the polynomial was equated to zero and solved, i.e.:

$$\frac{\partial y}{\partial x} = 3.7144x + 4.8579 = 0 \quad (17)$$

This resulted in a value of $x = -1.30786 = \lambda_2$, and therefore the transformation obtained was:

$$Y_{2i} = (\log X_i) [(X_i)^{-1.30786}]^{-0.1566} \quad (18)$$

The Bartlett's test resulted in a value of $\chi^2 = 13.62$, therefore not significant at 95% and 99% of probability levels, i.e. homoscedasticity was finally reached. The application of the Kolmogorov-Smirnov's test resulted in a significant value, still presenting non-normality. The value of asymmetry was -0.25838, tending to a moderate condition.

In view of these circumstances, it was proposed the inclusion of the λ_3 parameter, such that to force the normality, resulting in the final set of transformations proposed by the authors:

$$Z_i = (\log X_i)^{\lambda_3} [(X_i)^{\lambda_2}]^{\delta} \quad (19)$$

The successive asymmetry values were solved in function of λ_3 , i.e. $S = f(\lambda_3)$, for a range between $-1.0 \leq \lambda_3 \leq 2.5$ (Figure 3).

The minimization of asymmetry and homogeneity of variances were obtained for the value of $\lambda_3 = -0.785$, therefore the final transformation, after rationalization (ϖ) in (Equation 20), is presented in (Equation 21).

$$\varpi = \hat{\lambda}_2 \hat{\delta} = [-1.30786(-0.1566)] = 0.2087 \quad (20)$$

$$Z_i = (\log X_i)^{-0.785} (X_i)^{0.2087} \quad (21)$$

Once applied the procedures as suggested in (Equation 21), the data were transformed to the variable Z_i . The largest variance resulted in 0.000419 and the lowest in 0.00016, i.e. a ratio of 2.62. The new transformed variable (Z_i) was submitted to Bartlett's test and resulted in a value of $\chi^2 = 14.00$, therefore not significant at 95% and 99% probability levels. The Kolmogorov-Smirnov's test presented a value of 0.04614, which identifies a normal distribution and a value of asymmetry equal to - 0.06395, i.e. close to zero. These results demonstrate the effectiveness of the proposed methodology.

The entire procedure proposed in this research is based on the application of the hypothesis test for the equality ($H_0: GV_{max} = GV_{min}$) of the ranges of variation of the treatments, which implicitly was detected after the rejection of the null hypothesis, both for normality as well as to homoscedasticity of variances.

It was observed that λ_3 is the generator of normality since successive Bartlett's tests have resulted in minor variations. Therefore, the proposed transformation for the parameter $\hat{\delta}$ strength the ratios between the variances of the treatments to decrease, so to approximate closer as much as possible to one; the parameter λ_2 corrects the changes which were not included in the calculation of $\hat{\delta}$, while the parameter λ_3 strength asymmetry get closer to zero. This is expected when using data transformation to create a suitable condition for the application of the analysis of variance.

The original and transformed data were submitted to the analysis of variance and the results are presented in Table 2.

Table 2. Analysis of variance of the variable radiation incident at ground level in stands of black wattle under the effect of two environments (Hillside of the Southeast – HS and Mountain of the Southeast - MS), in four stand ages: 1, 3, 5 and 7 years.

Source of Variation	Degrees of Freedom	Mean Square	
		X_i	Z_i
Environment (E)	1	178,040.0967 **	0.01109 ^{ns}
Stand Age (S)	3	854,218.0970 **	0.09872**
E * S	3	71,155.9901 ^{ns}	0.00172 ^{ns}

Where X_i is the original variable and Z_i is the transformed variable. ns - not significant and ** - highly significant.

Discussion

Properties of the proposed transformations

Observing Table 2, it was evident that results of F test for environmental effect would be leading to a type II error, i.e. to reject the null hypothesis when it is true, as previously was stated by Snedecor & Cochran (1967).

The application of logarithms to the data was consider appropriate in this research, reason why we opted for this transformation as the first step of the four others proposed by the authors.

The transformation proposed by the authors was:

$$Z_i = (\log X_i)^{\lambda_3} [(X_i)^{\lambda_2}]^{\hat{\delta}} \tag{22}$$

Thus, we have:

When $\lambda_3 = 1$, the transformation is equivalent to

$$Z_i = (\log X_i) [(X_i)^{\lambda_2}]^{\hat{\delta}} \tag{23}$$

When $\lambda_3 = 0$, the transformation is equivalent to

$$Z_i = [(X_i)^{\lambda_2}]^{\hat{\delta}} = (X_i)^{\varpi} \tag{24}$$

Where ϖ is equal to the product of λ_2 by $\hat{\delta}$.

From these applications we can verify that λ_2 and $\hat{\delta}$ ($Z_i = [(X_i)^{\lambda_2}]^{\hat{\delta}}$) are generators of homoscedasticity, while λ_3 of normality [$Z_i = (\log X_i)^{\lambda_3}$]. Complementary to the expression given in (Equation 22) we have a general case of other restrictive conditions presented in (Equation 23) and (Equation 24) or in other circumstances.

Taking the case, since $\lambda_2 = -1.30786$ and $\hat{\delta} = 0.1566$, the application of the Chi-square test on the new obtained variable (Z_i) resulted in a value of $\chi^2 = 13.99$, therefore not significant at 95% and 99% probability levels, however, with the condition of non-normality, verified by means of the Kolmogorov-Smirnov's test, and a value of asymmetry of -0.37941, i.e., tending to moderate value. With the modification of the coefficient ϖ on the variable X , it is possible to force the asymmetry closer to zero, however when reaching this point it becomes difficult to achieve homogeneity of variances, since both procedures are inversely proportional (Figure 3).

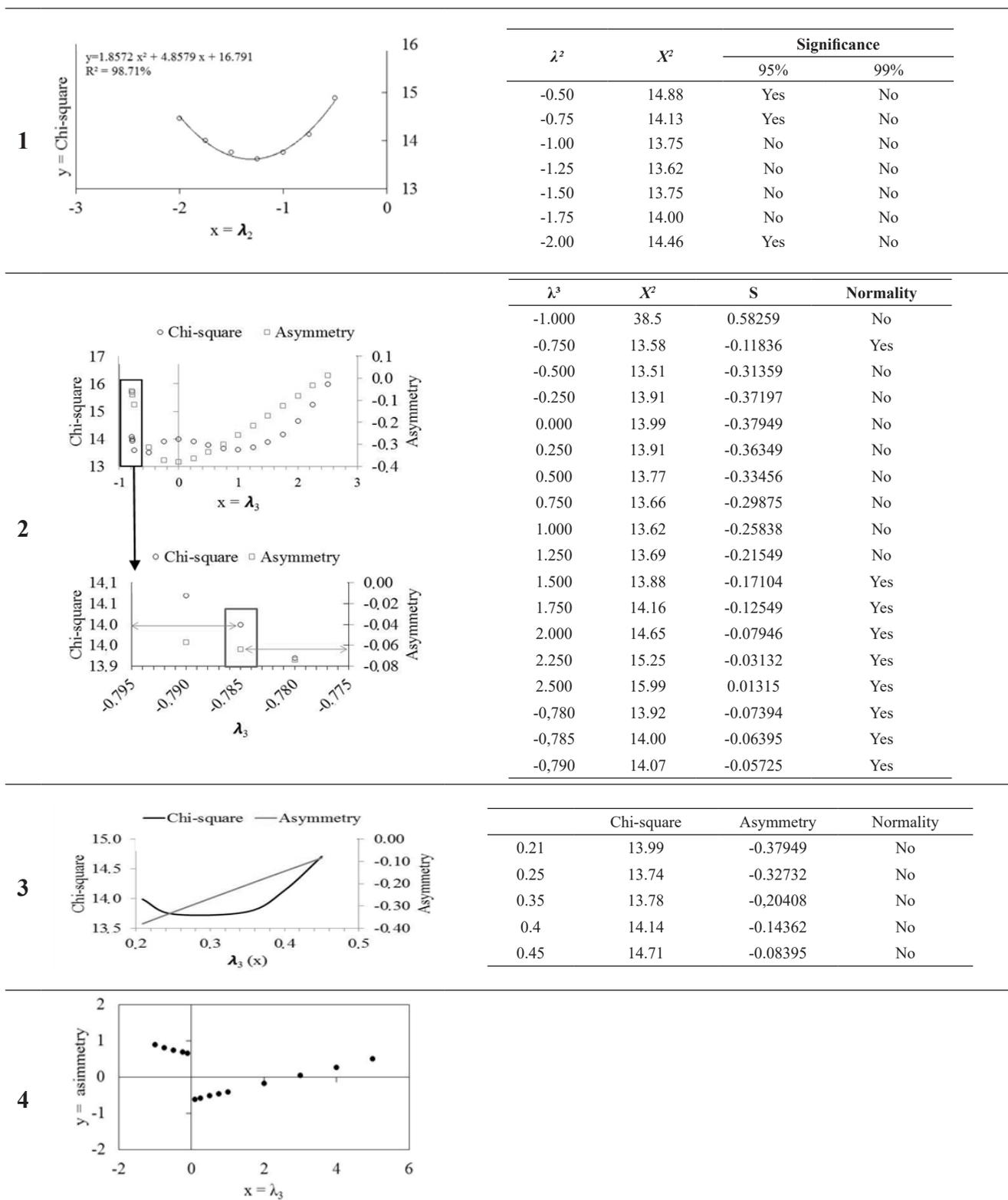


Figure 3. Test of normality and homogeneity of variances. 1 - Modeling the value of Chi-square (X^2) of the Bartlett's test as a function of λ_2 . 2 - Minimization of asymmetry and the homoscedasticity of variances. 3 - Relationship between the normality of the data and homoscedasticity of variances. 4 - Variation of asymmetry in function of λ_3 .

Taking $\lambda_2 = 0$ and $\hat{\delta} = 0$, we get the transformation $Z_i = (\log X_i)^{\lambda_3}$. The successive values of symmetry (S) were equated in function of λ_3 , i.e. $S = f(\lambda_3)$, for a range of $-1.0 \leq \lambda_3 \leq 5$ (Figure 2). It was observed that it is possible to achieve symmetry equal to zero for $\lambda_3 = 2.8076$, however with heteroscedasticity of variances at 95% probability level ($\chi^2=16.50$).

In this research it was demonstrated that the inclusion of parameters that generate normality and homoscedasticity in the same transformation procedure results a balance to achieve these two assumptions, which is an ideal situation for the application of analysis of variance. It was also observed that it is preferable to work first with heteroscedasticity of variances and after with normality once the values of Bartlett's test have not changed much when the asymmetry was forced to move closer to zero.

Some relevant points about the proposed transformations

Data transformation have been used frequently to approximate the residuals to the normal distribution and to achieve homoscedasticity (Sokal & Rohlf, 2012). For example, in Brazilian Journal of Forestry Research, analysis of variance was one of the most applied statistical test, as can be evidenced in several papers (Amaral et al., 2013; Faria et al., 2013; Freitas et al., 2013; Guareschi & Pereira, 2013; Guimarães Junior et al., 2013; Haliski et al., 2013; Lucena et al., 2013; Maeda & Bognola, 2013; Oliveira & Alixandre, 2013; Pias et al., 2013; Rocha et al., 2013; Rosado et al., 2013; Rossa et al., 2013; Santos et al., 2013; Ataíde et al., 2014; Bastos et al., 2014; Carpanezzi & Gualtieri, 2014; Dias Júnior et al., 2014; Flores et al., 2014; Leite et al., 2014; Medeiros et al., 2014; Mossanek et al., 2014; Oliveira et al., 2014; Prevedello et al., 2014; Protásio et al., 2014; Silva et al., 2014; Souza et al., 2014; Téó et al., 2014; Turchetto & Fortes, 2014; Alves et al., 2017; Benedetti et al., 2017; Boscardin et al., 2017; Cerqueira et al., 2017; Dutra et al., 2017; Freire et al., 2017; Lima & Stape, 2017; Lustosa et al., 2017; Mascarenhas et al., 2017; Navroski et al., 2017; Nobile et al., 2017; Oliveira et al., 2017; Ribeiro et al., 2017; Salles et al., 2017; Talgatti et al., 2017; Tavares et al., 2017; Venial et al., 2017).

According to Sokal & Rohlf (2012), in many cases it is possible to assume that analysis of variance can be rectified by transformation of the original data into a new scale. Ribeiro et al. (2017) and Oliveira et al. (2017) presented an interesting discussion about data

transformation, evidencing the most frequent situations related to data transformation and its impact on ANOVA assumptions and experimental accuracy.

The proposed transformations have started from the knowledge of other proposals hitherto existing, but they could not meet simultaneously normality and homoscedasticity of variances. The authors have conceived the development of this methodology in two tests of hypothesis and have complemented them with the application of the transformation from Tukey's and Box-Cox's ideas, in which they have presented other algebraic developments.

The proposed transformations were conceived by introducing mathematical effects to get simultaneously the two conditions pointed out before. Certainly, this problem is complex, and the possible users of these transformations should follow step by step the conditions presented in the manuscript, i.e. applying two tests of hypotheses, and finding solutions to achieve both goals.

The proposed transformations (Transformation in Four Steps - TFS) are particularly interesting in cases where the conditions of normality and homoscedasticity are not simultaneously met, which does not apply in the case of Box-Cox's transformation. TFS is not a direct competitor with the Box-Cox's transformation, but is an alternative that can assist in more specific cases.

Referring to the condition of wider ranges of variation observed in the data, the proposed transformations are robust because the modeling of these ranges is performed to find a balance between them. Homogeneity of variances is consequence of equality of ranges of variations (GV). Consequently, the changes proposed by the authors aimed to, primarily, homogenize GV . In addition, as GV is obtained by the difference between the extreme values of each treatment, the possible influence of outliers, in our view, is smaller than any other more complex solution applied to the variances.

There are several statistical tests to verify that the data come from a normal distribution, standing among them more frequently used the Shapiro-Wilk's test (for cases where the sample is composed of less than 50 observations), and the Kolmogorov-Smirnov's test (for cases where the sample is of moderate to large size). Under the conditions of this research, the Kolmogorov-Smirnov's test was applied, because there are in the experiment 360 observations from 8 treatments conducted in 45 replications. For this case the Shapiro-Wilk's test is less appropriated when compared to the Kolmogorov-Smirnov's test.

The statistical model applied in this study was completely randomized in factorial design. The proposed transformations may be used to any other type of experimental design (completely randomized design, randomized block, Latin Squares, and others), because the scheme used to perform local control does not interfere in the proposed methodology. The methodology can also be used in regression analysis, providing that the independent variable is grouped into classes.

Conclusions

In some experimental evaluations, in which analysis of variance is applied, we can face asymmetric distributions of data and heteroscedasticity of variances. This was the case for the evaluation of radiation at ground level in stands of *Acacia mearnsii* De Wild., in Rio Grande do Sul, Brazil, but without attending the assumptions for ANOVA, in spite of the application of all alternatives of data transformation for biometric variables so far available in the literature.

The proposal presented the so-called transformation in four steps, which includes two additional hypothesis testing, one for heteroscedasticity and another for asymmetry evaluations.

The data transformation proposed is robust and is unrestricted to be applied to any set of experimental biological data, after detecting a priori no attendance to the assumptions required for the application of the conventional analysis of variance.

References

- Alves, A. R. et al. Nutrientes na biomassa aérea e na serapilheira em áreas de caatinga em Floresta, PE. **Pesquisa Florestal Brasileira**, v. 37, n. 92, p. 413-420, 2017. <https://doi.org/10.4336/2017.pfb.37.92.1060>.
- Amaral, C. S. et al. Crescimento de *Dalbergia miscolobium* em substrato de rejeito da mineração de quartzo fertilizado. **Pesquisa Florestal Brasileira**, v. 33, n. 74, p. 179-187, 2013. <https://doi.org/10.4336/2013.pfb.33.74.423>.
- Ataide, G. M. et al. Avaliação preliminar da embebição de sementes de jacarandá-da-Bahia. **Pesquisa Florestal Brasileira**, v. 34, n. 78, p. 133-139, 2014. <https://doi.org/10.4336/2014.pfb.34.78.520>.
- Bastos, M. C. et al. Mineral content of young leaves of yerba mate. **Pesquisa Florestal Brasileira**, v. 34, n. 77, p. 63-71, 2014. <https://doi.org/10.4336/2014.pfb.34.77.594>.
- Baran, E. & Warry, F. **Simple data analysis for biologists**. Cambodia: WorldFish and the Fisheries Administration, 2008. 67 p.
- Bartlett M. S. Add link under of statistical methods of research in agriculture and applied biology. **Journal of the Royal Statistical Society**, v. 4, p. 137-183, 1937.
- Benedetti, E. L. et al. Alumínio estimula o crescimento radicular de erva-mate? **Pesquisa Florestal Brasileira**, v. 37, n. 90, p. 139-147, 2017. <https://doi.org/10.4336/2017.pfb.37.90.983>.
- Bickel, P. J. & Doksum, K. A. An analysis of transformations revisited. **Journal of the American Statistical Association**, v. 76, p. 296-311, 1981.
- Boscardin, J. et al. Crescimento de nogueira-pecã sob diferentes preparos do solo e coveamentos: coleópteros como bioindicadores. **Pesquisa Florestal Brasileira**, v. 37, n. 92, p. 587-596, 2017. <https://doi.org/10.4336/2017.pfb.37.92.1496>.
- Box, G. E. P. & Cox, D. R. An analysis of transformations. **Journal of the Royal Statistical Society: Series B**, v. 26, p. 211-252, 1964.
- Carpanezzi, F. B. & Gualtieri, S. C. J. Alelopatia de extratos aquosos foliares da exótica invasora *Pittosporum undulatum* na germinação e crescimento do capim-arroz. **Pesquisa Florestal Brasileira**, v. 34, n. 79, p. 173-179, 2014. <https://doi.org/10.4336/2014.pfb.34.79.599>.
- Cerqueira, P. H. A. et al. Adubação residual na produção de mudas clonais de eucalipto. **Pesquisa Florestal Brasileira**, v. 37, n. 90, p. 119-129, 2017. <https://doi.org/10.4336/2017.pfb.37.90.934>.
- Cochran, W. G. Add consequences when the assumptions for the analysis of variance are not satisfied. **Biometrics**, v. 3, p. 22-38, 1947.
- Dias Júnior, A. F. et al. Caracterização de briquetes produzidos com resíduos agroflorestais. **Pesquisa Florestal Brasileira**, v. 34, n. 79, p. 225-234, 2014. <https://doi.org/10.4336/2014.pfb.34.79.613>.
- Draper, N. R. & Cox, D. R. On distributions and their transformations to variable normality test. **Journal of the Royal Statistical Society, Series B**, v. 31, p. 472-476, 1969.
- Dutra, T. R. et al. Efeito da salinidade na germinação e crescimento inicial de plântulas de três espécies arbóreas florestais. **Pesquisa Florestal Brasileira**, v. 37, n. 91, p. 323-330, 2017. <https://doi.org/10.4336/2017.pfb.37.91.1447>.
- Faria, J. C. et al. Uso de resíduos orgânicos no crescimento de mudas de *Mimosa setosa*. **Pesquisa Florestal Brasileira**, v. 33, n. 76, p. 409-418, 2013. <https://doi.org/10.4336/2013.pfb.33.76.501>.
- Flores, A. V. et al. Efeito do substrato, cor e tamanho de sementes na germinação e vigor de *Melanoxylon brauna*. **Pesquisa Florestal Brasileira**, v. 34, n. 78, p. 141-147, 2014. <https://doi.org/10.4336/2014.pfb.34.78.558>.
- Fowler, J. et al. **Practical statistics for field biology**. 2nd. ed. New Jersey: John Wiley & Sons, 2009. 272 p.
- Freire, J. M. et al. Efeito do substrato sobre o crescimento de mudas de *Mimosa bimucronata* inoculadas com estirpes de rizóbio. **Pesquisa Florestal Brasileira**, v. 37, n. 90, p. 131-138, 2017. <https://doi.org/10.4336/2017.pfb.37.90.1132>.
- Freitas, T. A. S. et al. Crescimento e ciclo de produção de mudas de *Eucalyptus* em recipientes. **Pesquisa Florestal Brasileira**, v. 33, n. 76, p. 419-428, 2013. <https://doi.org/10.4336/2013.pfb.33.76.575>.

- Guareschi, R. F. & Pereira, M. G. Carbono, matéria orgânica leve e frações oxidáveis do carbono orgânico sob sistemas de aléias. **Pesquisa Florestal Brasileira**, v. 33, n. 74, p. 109-114, 2013. <https://doi.org/10.4336/2013.pfb.33.74.450>.
- Guimarães Junior, J. B. et al. Produção de painéis aglomerados da madeira de desrama de *Acacia mangium*. **Pesquisa Florestal Brasileira**, v. 33, n. 76, p. 387-391, 2013. <https://doi.org/10.4336/2013.pfb.33.76.434>.
- Haliski, S. et al. Caracterização morfológica de frutos, sementes, plântulas e germinação de sementes de *Casearia decandra*. **Pesquisa Florestal Brasileira**, v. 33, n. 75, p. 253-259, 2013. <https://doi.org/10.4336/2013.pfb.33.75.451>.
- Hartley, H. O. The use of range in analysis of variance. **Biometrika**, v. 37, p. 271-280, 1950. <https://doi.org/10.2307/2332380>.
- John, J. A. & Draper, N. R. An alternative family of transformations. **Applied Statistics**, v. 29, p. 190-197, 1980.
- Kolmogorov, A. Sulla determinazione empirica di una legge di distribuzione. **Giornale dell'Istituto Italiano degli Attuari**, v. 4, p. 83-91, 1933.
- Leite, T. S. et al. Crescimento e partição da biomassa de mudas de mulungu sob adubação fosfatada e inoculação micorrízica. **Pesquisa Florestal Brasileira**, v. 34, n. 80, p. 407-415, 2014. <https://doi.org/10.4336/2014.pfb.34.80.642>.
- Lima, I. L. & Stape, J. L. Caracterização da madeira serrada em clones de *Eucalyptus*. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 55-62, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1113>.
- Lucena, V. B. et al. Influência de fungos micorrízicos-arbusculares em paricá (*Schizolobium amazonicum*) cultivado no estado do Pará. **Pesquisa Florestal Brasileira**, v. 33, n. 75, p. 235-241, 2013. <https://doi.org/10.4336/2013.pfb.33.75.38>.
- Lustosa, B. M. et al. Strategies of two tropical woody species to tolerate salt stress. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 63-72, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1332>.
- Maeda, S. & Bognola, I. A. Propriedades químicas de solo tratado com resíduos da indústria de celulose e papel. **Pesquisa Florestal Brasileira**, v. 33, n. 74, p. 169-177, 2013. <https://doi.org/10.4336/2013.pfb.33.74.417>.
- Manly, B. F. Exponential data transformation. **The Statistician**, v. 25, p. 37-42, 1976.
- Mascarenhas, A. R. P. et al. Atributos físicos e estoques de carbono do solo sob diferentes usos da terra em Rondônia, Amazônia Sul-Occidental. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 19-27, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1295>.
- McDonald, J. H. **Handbook of biological statistics**. Maryland: Sparky House Publishing, 2009. 313 p.
- Medeiros, J. G. F. et al. Sanidade e germinação de sementes de *Clitoria fairchildiana* tratadas com extratos de plantas. **Pesquisa Florestal Brasileira**, v. 33, n. 76, p. 403-408, 2014. <https://doi.org/10.4336/2013.pfb.33.76.541>.
- Mossanek, E. A. O. et al. Indução de perfilhos em mudas de pupunheira. **Pesquisa Florestal Brasileira**, v. 34, n. 77, p. 39-48, 2014. <https://doi.org/10.4336/2014.pfb.34.77.549>.
- Navroski, M. C. et al. Tolerância ao frio e características silviculturais do cedro-australiano no sul do Brasil. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 47-54, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1257>.
- Nobile, F. O. et al. Soluções nutritivas para produção de porta-enxertos de seringueira. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 73-80, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1347>.
- Oliveira, L. dos R. et al. Crescimento de mudas de cedro-rosa em diferentes substratos. **Pesquisa Florestal Brasileira**, v. 34, n. 79, p. 187-195, 2014. <https://doi.org/10.4336/2014.pfb.34.79.605>.
- Oliveira, J. J. & Alixandre, T. F. Parâmetros biométricos de mudas de sabiá micorrizadas sob níveis de fósforo em Latossolo Amarelo. **Pesquisa Florestal Brasileira**, v. 33, n. 74, p. 159-167, 2013. <https://doi.org/10.4336/2013.pfb.33.74.467>.
- Oliveira, M. K. T. et al. Uso de substratos orgânico-minerais na produção de mudas de *Erythrina velutina*. **Pesquisa Florestal Brasileira**, v. 37, n. 91, p. 235-242, 2017. <https://doi.org/10.4336/2017.pfb.37.91.1029>.
- Osborne, J. W. Notes on the use of data transformations. **Practical Assessment, Research & Evaluation**, n. 8, v. 6, 2002.
- Parsad, R. Transformation of data. In: Parsad, R. et al. (ed.). **A diagnostic study of design and analysis of field experiments**. New Delhi: Indian Agricultural Statistics Research Institute, 2005. p. 637-647.
- Pias, O. H. C. et al. Doses de fertilizante de liberação controlada no índice de clorofila e na produção de mudas de grápiá. **Pesquisa Florestal Brasileira**, v. 33, n. 73, p. 19-26, 2013. <https://doi.org/10.4336/2013.pfb.33.73.419>.
- Prevedello, J. et al. Agregação e matéria orgânica de um argissolo sob diferentes preparos do solo para plantio de Eucalipto. **Pesquisa Florestal Brasileira**, v. 34, n. 78, p. 149-158, 2014. <https://doi.org/10.4336/2014.pfb.34.78.456>.
- Protásio, T. P. et al. Qualidade da madeira e do carvão vegetal oriundos de floresta plantada em Minas Gerais. **Pesquisa Florestal Brasileira**, v. 34, n. 78, p. 111-123, 2014. <https://doi.org/10.4336/2014.pfb.34.78.657>.
- Rahman, L. & Sen, K. Problems of usual weighted analysis of variance (WANOVA) in randomized block design (RBD) with more than one observation per cell when error variances vary from cell to cell. **Bangladesh Journal Agricultural Research**, v. 34, p. 395-401, 2009. <https://doi.org/10.3329/bjar.v34i3.3964>.
- Ramsey, F. L. & Schafer, D. W. **The statistical sleuth: a course in methods of data analysis**. 3rd ed. Boston: Cengage Learning, 2012. 784 p.
- Ribeiro, F. P. et al. Conteúdo de nutrientes na serapilheira em três fisionomias do Cerrado do Distrito Federal. **Pesquisa Florestal Brasileira**, v. 37, n. 92, p. 465-473, 2017. <https://doi.org/10.4336/2017.pfb.37.92.1312>.
- Rocha, J. H. et al. Composto de lodo de esgoto como substrato para mudas de eucalipto. **Pesquisa Florestal Brasileira**, v. 33, n. 73, p. 27-36, 2013. <https://doi.org/10.4336/2013.pfb.33.73.33>.

- Rosado, A. M. et al. Avaliação da tolerância à quebra por vento em árvores de eucalipto via teste de resistência. **Pesquisa Florestal Brasileira**, v. 33, n. 75, p. 309-315, 2013. <https://doi.org/10.4336/2013.pfb.33.75.50>.
- Rossa, U. B. et al. Fertilização de liberação lenta no crescimento de mudas de paricá em viveiro. **Pesquisa Florestal Brasileira**, v. 33, n. 75, p. 227-234, 2013. <https://doi.org/10.4336/2013.pfb.33.75.42>.
- Sakia, R. M. The Box-Cox transformation technique: a review. **The Statistician**, v. 41, p. 169-178, 1992.
- Salles, E. A. P. B. et al. Desinfestação e introdução in vitro de segmentos nodais de *Acacia mearnsii*. **Pesquisa Florestal Brasileira**, v. 37, n. 92, p. 485-491, 2017. <https://doi.org/10.4336/2017.pfb.37.92.1392>.
- Smirnov, N. Table for estimating the goodness of fit of empirical distributions. **Annals of Mathematical Statistics**, v. 19, p. 279-281, 1948.
- Snedecor, G. W. & Cochran, G. W. **Statistical methods**. The Iowa State University Press, 1967. 593 p.
- Santos, L. W. et al. Qualidade de mudas de pau-ferro produzidas em diferentes substratos e condições de luz. **Pesquisa Florestal Brasileira**, v. 33, n. 74, p. 151-158, 2013. <https://doi.org/10.4336/2013.pfb.33.74.344>.
- Silva, N. M. et al. Efeito de homeopatia na germinação de sementes de ipê amarelo. **Pesquisa Florestal Brasileira**, v. 34, n. 79, p. 181-186, 2014. <https://doi.org/10.4336/2014.pfb.34.79.540>.
- Sokal, R. R. & Rohlf, F. J. **Biometry**: the principles and practice of statistics in biological research. New York: W. H. Freeman and Co, 2012. p. 889.
- Souza, C. C. et al. Densidade de minicepas em minijardim clonal na produção de mudas de eucalipto. **Pesquisa Florestal Brasileira**, v. 34, n. 77, p. 49-56, 2014. <https://doi.org/10.4336/2014.pfb.34.77.512>.
- Steel, R. D. & Torrie, J. H. **Principles and procedures of statistics with special reference to the biological sciences**. McGraw-Hill Book Company, 1960. 481 p.
- Talgatti, M. et al. Compósito madeira-plástico a partir de resíduos de três espécies florestais. **Pesquisa Florestal Brasileira**, v. 37, n. 91, p. 277-283, 2017. <https://doi.org/10.4336/2017.pfb.37.91.1385>.
- Tavares, C. J. et al. Crescimento inicial de plantas de pequi após aplicação de 2,4-D. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 81-87, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1280>.
- Téo, S. J. et al. Poda da parte aérea, visando melhor qualidade de mudas de *Pinus taeda*, em Caçador, SC. **Pesquisa Florestal Brasileira**, v. 34, n. 77, p. 57-62, 2014. <https://doi.org/10.4336/2014.pfb.34.77.46>.
- Tukey, J. W. The Comparative anatomy of transformations. **Annals of Mathematical Intelligence Statistics**, v. 28, p. 602-632, 1957. <https://doi.org/10.1214/aoms/1177706875>.
- Turchetto, F. & Fortes, F. O. Aporte e decomposição de serapilheira em Floresta Estacional Decidual na região do Alto Uruguai, RS. **Pesquisa Florestal Brasileira**, v. 34, n. 80, p. 391-397, 2014. <https://doi.org/10.4336/2014.pfb.34.80.735>.
- Venial, L. R. et al. Biometria e armazenamento de sementes de genótipos de cacaueiro. **Pesquisa Florestal Brasileira**, v. 37, n. 89, p. 39-46, 2017. <https://doi.org/10.4336/2017.pfb.37.89.1239>.
- Zarembka, P. Transformation of variables in econometrics. In: Zarembka, P. **Frontiers of econometrics**. London: Academic Press, 1974. p. 81-104.