University of São Paulo

Ribeirão Preto Medical School

Models based on skewed distributions for paired data, with applications to health data

Modelos baseados em distribuições assimétricas para dados pareados, com aplicações à dados de saúde

Guilherme Zubatch da Cunha

Ribeirão Preto

2020

Models based on skewed distributions for paired data, with applications to health data

Modelos baseados em distribuições assimétricas para dados pareados, com aplicações à dados de saúde

Thesis submitted to the Ribeirão Preto Medical School, University of São Paulo, as part of the requirements to obtain the Degree of Doctor in Science (DSc).

Area of concentration: Public Health

Supervisor: Prof. Dr. Edson Zangiacomi Martinez

Ribeirão Preto

2020

Ficha Catalográfica

Cunha, Guilherme Zubatch da

Models based on skewed distributions for paired data, with applications to health data, 2020. 63p. : il. ; 30 cm

Modelos baseados em distribuições assimétricas para dados pareados, com aplicações à dados de saúde

Tese de Doutorado, apresentada à Faculdade de Medicina de Ribeirão Preto/USP

Área de concentração: Saúde Púbica
Orientador: Martinez, Edson Zangiacomi

1. Biostatistics. 2. Paired data. 3. Regression models. 4. Analysis of Variance. 5. Health data. 6. Longitudinal data.

1. Bioestatística 2. Dados pareados 3. Modelos de regressão 4. Análise de Variância 5. Dados em saúde 6. Dados longitudinais.

Declaration

I hereby declare that the work presented in this thesis is my own effort and that the thesis is my own composition. I have not presented previously any part of the thesis for any other degree.

I authorize the reproduction of this work; by any means conventional or electronic, for study and research purposes, provided that the source is cited properly.

Guilherme Zubatch da Cunha

Financial support and resources

The financial and academic support from the following resources made this work possible:

- Coordination for the Improvement of Higher Education Personnel – CAPES.
- Foundation for Teaching Support, and Research Assistance (FAEPA), from Hospital of Clinics, belonging to Ribeirão Preto Medical School (HC-FMRP, USP).
- Ribeirão Preto Medical School (FMRP, USP).

Name: Guilherme Zubatch da Cunha

Title: Models based on skewed distributions for paired data, with applications to health data

Título: Modelos baseados em distribuições assimétricas para dados pareados, com aplicações à dados de saúde

Thesis submitted to the Ribeirão Preto Medical School, University of São Paulo, as part of the requirements to obtain the Degree of Doctor in Science (DSc).

Examination committee

President: Prof. Dr. Edson Zangiacomi Martinez

Institution: FMRP/USP Signature _____

Members

Prof. Dr. _____

Institution: _____ Signature _____

Prof. Dr. _____

Institution: _____ Signature _____

Prof. Dr. _____

Institution: _____ Signature _____

ABSTRACT

Cunha G. Z. **Modelos baseados em distribuições assimétricas para dados pareados, com aplicações à dados de saúde**. 2020. Tese (Doutorado em Ciências) – Faculdade de Medicina de Ribeirão Preto, Programa de Pós-Graduação em Saúde Pública. Universidade de São Paulo, Ribeirão Preto. 2020. 63p.

Esta tese tem por objetivo introduzir e explorar o interesse na análise pré-teste / pós-teste em estudos epidemiológicos. Em um primeiro artigo, aspectos conceituais são descritos e diretrizes metodológicas para esse tipo de análise são sistematizadas. Definimos os conceitos-chave de pré-teste e pós-teste e como eles podem ser aplicados à epidemiologia. Os métodos para avaliar o pré-teste e o pós-teste são muito comuns e existem várias maneiras de resolver o mesmo problema. O objetivo aqui é apresentar alguns deles: teste t de Student, análise de variância (ANOVA) nos escores de ganho, análise de covariância (ANCOVA), ANOVA em escores residuais e ANOVA com medidas repetidas. O segundo artigo apresenta uma família geral de distribuições assimétricas aplicadas no pré-teste / pós-teste em estudos epidemiológicos, considerando uma abordagem longitudinal. Situações em que a suposição de normalidade não é válida são muito comuns, especialmente quando os dados apresentam uma distribuição assimétrica. Para considerar possíveis efeitos de assimetria, propomos o uso de análises longitudinais em uma perspectiva bayesiana. Nos dois artigos, são relatados os resultados de uma aplicação em um conjunto de dados reais relacionados à esclerose múltipla.

Palavras chave: Bioestatística; Dados pareados; Modelos de regressão; Análise de Variância; Dados em saúde; Dados longitudinais.

RESUMO

Cunha G. Z. **Models based on skewed distributions for paired data, with applications to health data**. 2020. Thesis (Doctorate in Sciences) – Ribeirão Preto Medical School, Public Health Postgraduate Program. University of São Paulo, Ribeirão Preto. 2020. 63p.

This thesis aims to introduce and raise interest on pretest/posttest analysis in epidemiological studies. In a first article, conceptual aspects are described and methodological guidelines for this type of analysis are systematized. We defined the key concepts of pretest and posttest and how they can be applied to epidemiology. Methods to assess pretest and posttest are very common and there are many ways to resolve the same problem. The objective here is to present some of them: t-test, analysis of variance (ANOVA) on the gain scores, analysis of covariance (ANCOVA), ANOVA on residual scores, and repeated measures ANOVA. The second article presents a general family of skew-symmetric distributions applied on pretest/posttest in epidemiological studies by considering a longitudinal approach. Situation where the normality assumption is not valid are very common, especially when data present a skewed distribution. For taking into account possible skewness effects, we propose the use of longitudinal analysis from a Bayesian perspective. In both articles, results of an application to a real data set related to multiple sclerosis are reported.

Key words: Biostatistics; Paired data; Regression models; Analysis of Variance; Health data; Longitudinal data.

Acknowledgements

I would like to express my sincere gratitude to my research supervisor, Dr. **Edson Zangiacomi Martinez**, for his support and friendship.

My special thanks to **Sergio Carlos Nascimento** and **Paula Maria Pereira Merichelo** without them I would have missed all the registration deadlines.

The conclusion of this project could not have been achieved without the support of my friends, **Fernanda Lang** and **Paulo Pereira**.

I thank the professors and students of the Post-Graduate Program in Public Health at the Ribeirão Preto Medical School-USP for their support during my research work.

Contents

ARTICLE 1:
Approaches for pretest/posttest analysis

ARTICLE 1: Approaches for pretest/posttest analysis

Guilherme Zubatch da Cunha [1], and Edson Zangiacomi Martinez [1]

[1] Ribeirão Preto Medical School, University of São Paulo (USP), Ribeirão Preto, SP, Brazil

---

Address for correspondence: Guilherme Zubatch da Cunha, Ribeirão Preto Medical School, University of São Paulo (USP), Av. Bandeirantes 3900, Monte Alegre, CEP 14049-900, Ribeirão Preto, SP, Brazil.

E-mail: zubatch@hotmail.com.

Phone: (+55) 16 3602 2569.

Fax: (+1) 999 888 666.

Abstract: This paper aims to introduce and raise interest on pretest/posttest analysis in epidemiological studies. Conceptual aspects are described and methodological guidelines for this type of analysis are systematized. We defined the key concepts of pretest and posttest and how they can be applied to epidemiology. Methods to assess pretest and posttest are very common and there are many ways to resolve the same problem. The objective here is to present some of them: t-test, Analysis of variance (ANOVA) on the gain scores, Analysis of covariance (ANCOVA), ANOVA on residual

scores, and Repeated measures ANOVA. Results of an application to a real data set related to multiple sclerosis are reported.

Key words: Pretest/ posttest; Analysis of covariance; Analysis of variance.

1 Introduction

A common research design is the two-group pretest/posttest design with one dependent variable (Bonate, 2000; Brogan and Kutner, 2010; Dugard and Todman, 1995). The objective may be, for example, to establish the superiority or equivalence between two treatments. Inferring about the target population based on sample results requires special care in the study planning, since factors affecting the response should be evaluated and controlled.

The basic premise behind the pretest/posttest design involves obtaining a pretest measure of the outcome of interest prior to administering some treatment, followed by a posttest on the same measure after treatment occurs. In general, a common analysis is to set a difference score for each subject (posttest minus pretest or vice versa). The null hypothesis is that the means or medians of the (relative) differences are the same for each group.

In many cases the Student t-test or analysis of variance is used, although nonparametric tests could also be used, for example, the Mann- Whitney U test, or the median test, the Wilcoxon rank test is the most common, or their analogs for more than two groups. If there are three or more samples, it is common to use the Kruskal Wallis

and Friedman tests in the non-parametric case (Brown and Hettmansperger, 2002; Zimmerman and Zumbo, 1993).

Covariance analysis, where the pretest score is used as the covariate, is another method used for analyzing this design. An analysis of covariance (ANCOVA) will be superior to its analysis of covariance (ANOVA) counterpart in two distinct respects (i.e., increased statistical power and control), so long as a good covariate is used (Vogt and Johnson, 2011; Goldberg and Scheiner, 2001).

Another option for analyzing this kind of design is to consider the pretest and posttest as a longitudinal data, where the term longitudinal data means that each subject is measured repeatedly over time. The main interest is in characterizing as the measures changes over time and making previous versions of the results. Repeated-Measures and longitudinal designs are discussed in details by a number of authors, such as Cnaan et al. (1997), Crowder (2017), Laird and Ware (1982) and Zeger and Liang (1986).

This article illustrates some of the equivalences and differences between the difference score analysis and the repeated-measures/split-plot or profile analysis. Application to real data and major discussion are for a two-group pretest/posttest design where subjects are not matched. Concluding remarks indicate how the results can be extended easily to more than two groups.

2 Methods

In pretest/posttest designs, there are three ways of using a baseline measure:

(a) using only the posttest responses, the baseline measures are ignored.

(b) using change measures, or say, the difference or gain scores.

(c) using an ANCOVA where the baseline measures are considered as covariate. In this case, one can use either the post measure or the gain score as an outcome. The estimated treatment effects will be identical.

The following statistical methods are traditionally used in comparing groups with pretest and posttest data (Dimitrov and Rumrill, 2003):

(a) Student t-test,

(b) Analysis of variance (ANOVA) on the gain scores,

(c) Analysis of covariance (ANCOVA),

(d) ANOVA on residual scores, and

(e) Repeated measures ANOVA.

In all these methods, the use of pretest scores helps to reduce the error variance, thus producing more powerful tests than designs with no pretest data (Stevens, 2012). Generally speaking, the power of the test represents the probability of detecting differences between the groups when such differences exist.

2.1 General assumptions

When you choose to analyze your data using a methodology, part of the process involves checking for some assumptions to ensure that the data you want to analyze

can actually be analyzed. If the data violates one or more assumptions, the results of the analysis may be incorrect or misleading. In particular, a small sample size may increase vulnerability to breaches of assumptions. If the normality assumption is violated, or outlier values are present, the test may not be the most powerful test available, and this may mean the difference between detecting a true difference or not.

1) Implicit factors: are factors that imply some connection between items in a sample. A common implicit factor is time, for example, values collected over time can be serially correlated (time here is the implicit factor). The word "factor" is extremely broad, and means that virtually anything can be an implicit factor. Implicit factors can be difficult to detect. An index chart of the data (value of plotted data in relation to row number) may reveal patterns in the chart that could suggest possible effects over time.

2) Outliers: in the analysis, an outlier is an observation that presents a high degree of inconsistency in the series. Outliers are simply single data points within your data that do not follow the usual pattern (Hawkins, 1980). Outliers tend to increase the estimate of sample variance, thus decreasing the calculated t statistic and lowering the chance of rejecting the null hypothesis. The boxplot and normal probability plot (normal Q-Q plot) may suggest the presence of outliers in the data.

3) Nonnormality: A normal distribution, sometimes called the Gaussian curve, is a distribution that occurs naturally in many situations. Several tests, including the one sample Z-test, t-test and ANOVA assume normality (Bonett and Woodward, 1990). Signs of nonnormality are lack of symmetry or heavy-tailedness or light-tailedness. The histogram, boxplot, and normal probability plot, along with the normality test,

can provide information on the normality of the population distribution.

4) Skewness: is a measure of symmetry. A distribution, or data set, is symmetric if it looks the same to the left and right of the center point. If the population from which the data were sampled is skewed, then the one-sample t test may incorrectly reject the null hypothesis that the population mean is the hypothesized value even when it is true. The effect of skewness on t-test have been investigated by Johnson (1978), Hall (1992), Abramovitch and Singh (1985), Lim and Lim (2016) and many others. Consider the one sample t-test. Based on simulation studies, skewness of the distribution does not affect the t-test as much as the confidence interval. However, for small samples, or highly skewed distributions, the above asymptotic result may not give a very good approximation, and so the type 1 error rate may deviate from the nominal 5% level. Skewness can be assessed either informally or formally. Histogram is an effective graphical technique for showing both the skewness and kurtosis of data set. Formally examining the sample skewness statistic or conducting a test for skewness. Usually tests for skewness are not resistant to outliers, so one could be considering the possibility that apparent skewness is in fact due to one or more outliers.

5) Small sample sizes: sample size calculation is part of the initial stages of conducting a clinical, epidemiological or laboratory study (Nayak, 2010; Chow et al., 2017). With a small sample, estimates and models may not be adequate. A smaller sample than the ideal may be difficult to detect breaches of assumptions as well as cause a false and inflated rate of discovery, inflated effect size estimation, low statistical power and also low reproducibility. To avoid such problems, additional care is required in the experimental design.

2.2 Student t-test

The simplest test to evaluate two-group pretest/posttest data is the Student t-test, a statistical procedure used to determine whether the mean difference between two sets of observations is zero (Lehmann, 2012; Pagano and Gauvreau, 2018). Student t-test is only used when it is a single group and it is not possible to used co-variables. The outcome of these tests is the acceptance or rejection of the null hypothesis (H_0). The null hypothesis generally states that: "Any differences, discrepancies, or suspiciously outlying results are purely due to random and not systematic errors". The alternative hypothesis (H_A) states exactly the opposite.

This test assumes: A normal (Gaussian) distribution for the populations of the random errors, there is no significant difference between the standard deviations of both population samples. Whether or not the observations are independent is an important factor to consider, it is also important to take into account the variability associated with population and sample values. Note that in the mean tests, we use the population variance value or an appropriate estimator.

Thus, in order to compare the means, we need to analyze what happens with the variances in the two populations. In some situations, they are known by previous studies, censuses or even assumptions. If population variations are unknown, there is still the question of whether they are the same or different. Some authors argue that in the case of unequal population variability, the means test should not be performed because the populations are already different. Next, a summary of possible situations will be presented in the comparison of two populations.

Case 1: Dependent samples (paired t-test)

In the case of dependent samples, we wish to compare two population averages and, for each sample unit, we performed two measures of the characteristic of interest. In general, these observations correspond to n measures taken before and after a given intervention.

The measures taken before and after the intervention will be represented by the random variables X_i and Y_i, respectively. In this way, the produced effect can be represented, by the i-th individual, by the variable $D_i = X_i - Y_i$. Assuming, for $i = 1, ..., n$, that D_i follows a normal distribution with mean μ_D and variance σ_D^2, or say,

$$D_i \sim N(\mu_D, \sigma_D^2),$$

the null and alternative hypotheses are respectively:

$$H_0 : \mu_D = 0 \text{ (the intervention has an effect)}$$

and

$$H_a : \mu_D \neq 0 \text{ (the intervention has no effect)}$$

The parameter μ_D is estimated by the sample mean and, $barD$ since we usually do not have information about σ_D^2, we estimate its value by S_D^2 a, given by

$$S_D^2 = \frac{1}{n-1} \sum_{n=1}^{n} (D_i - \bar{D})^2$$

The hypothesis test is performed using the

$$T = \frac{\bar{D} - \mu_D}{S_D/\sqrt{n}},$$

which, under H_0, follows a t-Student distribution with $n - 1$ degrees of freedom (Casella and Berger, 2002; Cox, 2006; Rohde, 2014).

Case 2. Independent samples with known variances

Consider now a test where the means of two populations are independent when the corresponding variances are known. Obtaining information about the value of the population variance can be obtained from previous studies or similar experiments. The interest is to determine if there is a statistically significant difference between the population means, the hypotheses are written in terms of $M_D = \mu_1 - \mu_2$. Thus, the null and alternative hypotheses are respectively:

$$H_0 : M_D = 0$$

and

$$H_a : M_D \neq 0.$$

The M_D estimator is given by

$$\bar{D} = \bar{X} - \bar{Y},$$

given the following assumptions:

$$X_i \sim N(\mu_1, \sigma_X^2), i = 1, 2, ..., n_1,$$

and

$$Y_i \sim N(\mu_2, \sigma_Y^2), i = 1, 2, ..., n_2.$$

By the property of independence between these variables, we have that $\bar{D}$ follows a normal distribution with mean $E(\bar{D}) = \mu_D$ and variance given by:

$$Var(\bar{D}) = Var(\bar{X} - \bar{Y}) = Var(\bar{X}) + Var(\bar{Y}) = \frac{\sigma_X^2}{n_1} + \frac{\sigma_Y^2}{n_2}$$

Note that the independence between the samples is required to obtain this result, since in this case the covariance between the samples is zero (Casella and Berger,

2002; Cox, 2006). With this information, we process the hypothesis test in the usual way.

Case 3A: Independent samples with unknown and equal variances

Let us supose that for both populations we have the same σ^2 (unknown) variance. Suppose that the interest is to test the following null and alternative hypothesis:

$$H_0 : \mu_X = \mu_Y$$

and

$$H_A : \mu_X \neq \mu_Y.$$

Again, let us consider the $\bar{D}$ estimator defined by the $\bar{X} - \bar{Y}$ difference. Given the assumption of independence between the samples, we have that the mean and the variance of $\bar{D}$ are given respectively by

$$E(\bar{D}) = \mu_X - \mu_Y$$

and

$$Var(\bar{D}) = \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right).$$

In addition, also considering the normality of the data, it follows that

$$\bar{D} \sim N \left(\mu_X - \mu_Y, \sigma^2 \left(\frac{1}{n_1} + \frac{1}{n_2} \right) \right).$$

Consequently,

$$\frac{\bar{D} - (\mu_X - \mu_Y)}{\sigma \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}} \sim N(0, 1).$$

As the variance of the population σ^2 is unknown, it will need to be estimated. Considering that S_X^2 and S_Y^2 are not addicted, we will use as an estimate for σ^2 a combination of them, given by:

$$S_P^2 = \frac{(n_1 - 1)S_X^2 + (n_2 - 1)S_Y^2}{n_1 + n_2 - 2}$$

(Casella and Berger, 2002; Cox, 2006). Note that S_P^2 is a weighted average between S_X^2 and S_Y^2, with weighting given by the degrees of freedom $n_1 - 1$ and $n_2 - 1$. In this way, all available information is used to estimate σ^2. In addition, it can be shown that S_c^2 is not addict to σ^2. In this way, we can work with the Student t-distribution, that is, let T be a statistics given by

$$T = \frac{\bar{D} - (\mu_X - \mu_Y)}{S_P\sqrt{\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}},$$

following, under H_0 a t-Student distribution with $n_1 + n_2 - 2$ degrees of freedom.

Case 3B: Independent samples with unknown and different variances

The hypothesis test for the case in which the variances are known and unequal is theoretically more complicated (Gronow, 1951). Thus, without going into more detail, we consider the same hypotheses presented in Case 3A, but the quantity to be used for the test will be

$$T = \frac{\bar{D} - (\mu_X - \mu_Y)}{\sqrt{S_X^2/n_1 + S_Y^2/n_2}}$$

As in the previous case, T also follows a t-Student distribution, but the degrees of freedom v are corrected by expression

$$v = \frac{(S_X^2/n_1 + S_Y^2/n_2)^2}{(S_X^2/n_1)^2 + (S_Y^2/n_2)^2}$$

(Satterthwaite, 1946). The sequence of the test is similar to all cases. If $T > t_{th}$ then H_0 is rejected, where t_{th} is a percentile of the t-Student distribution for a given significance level.

2.3 ANOVA on gain scores

ANOVA can be used to compare two or more groups (Scheffe, 1999). When the averages of two population are compared using ANOVA, it is equivalent to using a t-test to compare the means of independent groups. For single-group tests that compare pretest to posttest, analysts have been left with three options: a Student t-test of gain scores, paired t-tests, and repeated measures ANOVA. A more straightforward approach is to compute gain scores by subtracting the pretest scores from the posttest scores and running a one-way ANOVA (or t-test) on them.

The assumptions of ANOVA are the following:

(i) All populations involved follow a normal distribution.

(ii) All populations have the same variance (or standard deviation).

(iii) The samples are randomly selected and independent of one another.

If one defines $d_i = x_{ia} - x_{ib}$, then $M(d_i) = M(X_{ia}) - M(X_{ib})$ for each individual i, where $M(.)$ denotes the mean of a variable. A Student t-test of whether $M(x_{ai})$ is equal to $M(x_{bi})$ actually tests whether $M(X_{ai}) - M(X_{bi})$ is equal to zero, and this is identical to a test of whether $M(d)$ is equal zero.

The analysis of gains focuses on the improvements from pretest to posttest for whole groups. The use of gain scores in measurement of change has been criticized by a number of authors (Gupta et al., 1988; Linn and Slinde, 1977), but it can provide a reliable and unbiased estimate of true change. If all individuals grow at nearly the same rate, gain scores show that you cannot detect individual differences that do not

exist.

However, gain scores, for purposes such as the analysis of outcomes in randomized control trials can offer a better interpretation. Gains scores tells precisely how scores changed from pretest to posttest. It tells whether each group remained constant, increased or decreased, and by precisely how much.

The general approach to a gain score analysis is: (1) to compute the gain score, and then (2) analyze those gain scores in an analysis of variance with treatment as the between-subjects factor. The improvement (gain) can be calculated as:

$$Gain = posttest - pretest$$

When you compute a gain score in this manner, a positive gain score indicates that the posttest score was greater than the pretest score, and a negative gain score indicates that the posttest score was less than the pretest score. A linear additive statistical model for the ANOVA, called the effects model, is given by:

$$Y_{ij} = \mu + \tau_i + \epsilon_{ij},$$

where μ is the overall mean, τ_i are the deviations from the overall mean due to the treatment levels and ϵ_{ij} are the error terms. Under the null hypothesis, where the treatment effects are all zero, the reduced model can be written $Y_{ij} = \mu + \epsilon_{ij}$ and the SS_{Error} for the reduced model can be denoted by $SS_{Error_{(Reduced)}}$.

Under the alternative hypothesis, where the treatment effects are not all zero, the full model can be written $Y_{ij} = \mu_i + \epsilon_{ij}$ and the SS_{Error} for the full model can be denoted $SS_{Error_{(Full)}}$.

The treatment sum of squares (TSS) is

$$SST = SS_{Error(Reduced)} - SS_{Error(Full)}$$

and the General Linear Test can be employed to test the null hypothesis. Let be the null hypothesis that the group effect is equal to zero. If we reject this null hypothesis, then we conclude that the group effect is significant.

2.4 ANCOVA with pretest-posttest data

In most cases a more appropriate analysis would be an analysis of covariance (ANCOVA) (Cochran, 1957; Elashoff, 1969). In ANCOVA, the post-test measurement can be considered the dependent variable, the treatment can be considered the design factor, and the pretest scores the independent variables, along with other possible explanatory variables.

The purpose of using the pretest scores as a covariate in ANCOVA with a pretest-posttest design is to reduce the error variance and eliminate systematic bias. Moreover, the ANCOVA will almost always provide a more powerful test of the hypothesis of interest in comparison to the repeated measures ANOVA approach (Gourlay, 1955; Huck and McLean, 1975; Singer and Andrade, 1997).

Using ANCOVA implies making several assumptions, besides those already associated with ANOVA (such as population distribution of error in the subjacent population from which the sample is derived is normal, homogeneity of variance) the ANCOVA method suposses that the relationship between the covariate and the dependent variable is linear, that the covariate is measured without error, that is the within group regression slopes are homogeneous, and that the treatments do not affect the covari-

ate.

While it is not uncommon that gain or difference scores are used as the outcome for ANCOVA (Pascarella et al., 2003; Pike, 2004), the dependent variable of the ANCOVA in this study is primary post-test results for all the projects considered. Although the two approaches are mathematically equivalent, as shown in the equation below:

$$post_i = \beta_0 + \beta_1 T_i + e_i,$$

where $post_i$ represents gains from pretest to posttest, β_0 is the intercept, β_1 is average treatment effect, and T_i is treatment indicator. In addition, e_i refers to the residual component of the model. If we add $pret_i$ to both sides of the above equation, which is a t-test in a regression framework (Tu et al., 2008), we get:

$$post_i = \beta_0 + \beta_1 T_i + \beta_2 pret_i + e_i,$$

which is equivalent to the first one (for a similar argument where gain-ANCOVA also controls for baseline measure, see Eriksson and Häggström (2014)), although the implicit assumptions underpinning the two equations are different (Tu et al., 2008). That said, we still prefer post-ANCOVA to gain-ANCOVA for the following reasons. When baseline and final measures are on different scales, transforming them into comparable scales so that the difference can be calculated may, in our view, impose a rather different distribution to the data, let alone that there are usually multiple ways to transform the data (McElreath, 2018).

2.5 ANOVA Repeated-measures analysis

Another method of analyzing data from this study design is to consider the pretest and posttest as a repeated-measures/split-plot design or as a profile of two measure-

ments for each subject. Repeated-measures/split-plot designs are discussed in detail by Winer (1962), Jones and Nachtsheim (2009) and Stell et al. (1980). Repeated measures data has to do with experiments where you are taking observations repeatedly over time. Under a repeated measures experiment, experimental units are observed at multiple points in time. With this type of data, we are looking at only a single response variable but measured over time.

Statistical properties of the repeated-measures analysis of variance for this example and compare it with the statistical properties of the ANCOVA/ANOVA analysis. Using the model proposed by Winer (1962), we have

$$X_{ijk} = \mu + \alpha_i + \pi_{(k(i))} + \beta_j + \alpha\beta_{ij} + \beta\pi_{(jk(i))} + e_{(m(ijk))},$$

where $j = 1, 2$ (pretest $= 1$, posttest $= 2$), $i = 1, 2$ (group $1 = 1$, group $2 = 2$), and $k = 1, 2, n_i$. In addition, X_{ijk} is the observed value of subject k within group i at time j, μ is the overall mean, α_i is the effect of group i, $\pi_{(k(i))}$ is the effect of subject k nested within group i, $beta_j$ is the effect of the repeated-measures variable (i.e., pretest and posttest), $\alpha\beta_{ij}$ is the interaction of group i with level j of the repeated measures factor, $\beta\pi_{(jk(i))}$ is the interaction of subject k within group i with level j of the repeated-measures factor. The following constraints are imposed on the parameters:

$$\alpha_{\cdot} = \beta_{\cdot} = \alpha\beta_{\cdot j} = 0, j = 1, 2,$$

where

$$\alpha\beta_{\cdot j} = \sum_i \alpha\beta_{ij},$$

and so on. In the design under discussion, the repeated-measures factor and the group factor have each two levels.

3 Application to real data

Multiple sclerosis (MS) is an unpredictable, often disabling disease of the central nervous system that disrupts the flow of information within the brain, and between the brain and body (Goldenberg, 2012). The main disorders are intense fatigue, balance deficit and motor incoordination. In this study, the effects of physical therapy on the balance and quality of life of MS patients were evaluated. Data was obtained from analytical studies, where 20 subjects were randomized into two groups: Group 1 - submitted to fifteen sessions of physical therapy program for MS, three times per week, and Group 2 - underwent conventional physiotherapy once a week (Rodrigues et al., 2008). The objective of the study is to evaluate whether specific physiotherapy has a better effect in relation to traditional physiotherapy, longitudinally evaluating the alteration of individuals.

Tables 1 and 2 present the values of DEFU index (Functional Determination of Quality of Life Scale in individuals with MS) for each patient of the Groups 1 and 2, respectively, considering the periods before and after the intervention and their respective means. Figure 1 suggests that the effect in Group 1 improves from pre to post.

3.1 Statistical methods for analysis of pretest-posttest data

Comparing groups with pretest and posttest data: the paired t-test, analysis of variance (ANOVA) on the gain scores, analysis of covariance (ANCOVA), ANOVA on residual scores, and (4) repeated measures ANOVA. The main discussion here is to show several ways to compare experimental and control groups in posttest scores,

Table 1: Results from balance tests, through the BERG Balance Scale, and quality of life test, through the DEFU, before and after the fifteen sessions of physical therapy program for MS, three times per weeks (Group 1) (Rodrigues et al., 2008).

	BERG index		DEFU index	
Subject	Pre	Post	Pre	Post
1	56	58	51	58
2	42	46	42	50
3	56	60	66	68
4	52	58	61	66
5	52	59	91	99
6	54	62	109	109
7	53	57	142	144
8	53	60	149	152
9	39	41	62	68
10	53	62	86	86
Mean	51	56.3	85.9	90
SD	37.1	7.0	35.2	33.7

Table 2: Results from balance tests, through the BERG Balance Scale, and quality of life test, through the DEFU, before and after the conventional physiotherapy once a week (Group 2) (Rodrigues et al., 2008).

	BERG index		DEFU index	
Subject	Pre	Post	Pre	Post
11	53	53	63	66
12	48	50	78	76
13	36	38	97	97
14	56	56	88	88
15	18	18	86	88
16	13	13	87	91
17	49	51	79	79
18	54	56	77	78
19	16	19	46	49
20	28	31	153	156
Mean	37.1	38.5	85.4	86.8
SD	17.1	17.1	26.3	26.5

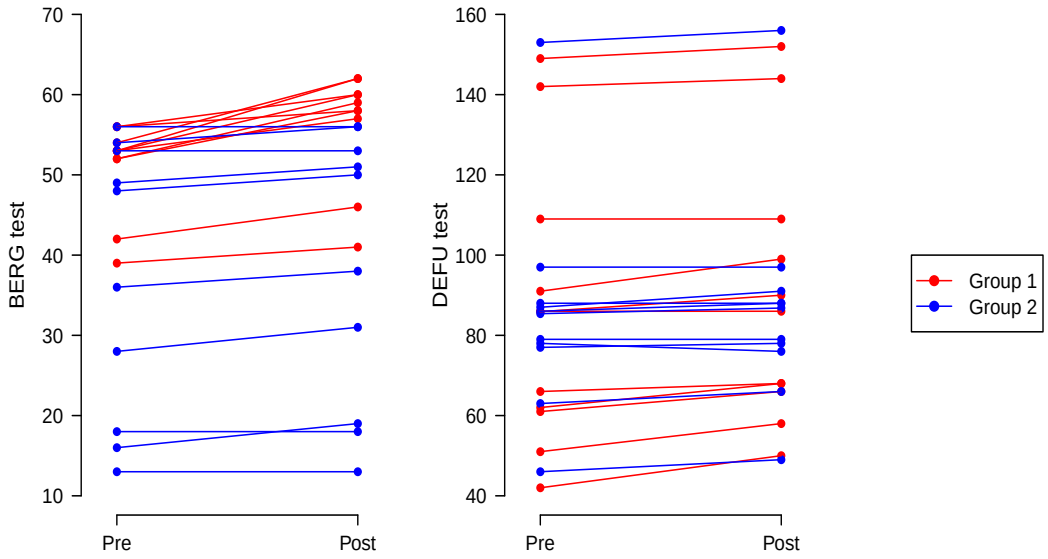

Figure 1: Pretest and posttest charts (ladder plots) for DEFU and BERG indexes.

while controlling for pretest differences or mean gain scores, that is, the difference between the posttest mean and the pretest mean.

Paired Student t-test

Results from paired Student t-test can be easily obtained by using the function t.test(x1,x2,paired=TRUE) of the R package, where x1 and x2 denote the pre- and post-measures. Table 3 shows results for t-tests applied to data present in Tables 1 and 2.

ANOVA on posttest scores

Results from ANOVA can be easily obtained by using the function aov() of the R package. Using only posttest measures and the groups (1 and 2) as a independent variable, results from ANOVA are showed in Table 4. In this table, DF denotes the

Table 3: Results from paired t tests.

Test	Group	mean of the differences	95% CI	t value	df	p-value
BERG	Group 1	-5.3	(-7.05 , -3.55)	-6.84	9	7E-05
BERG	Group 2	-1.4	(-2.30 , -0.49)	-3.50	9	0.007
DEFU	Group 1	-4.1	(-6.32 , -1.88)	-4.17	9	0.002
DEFU	Group 2	-1.4	(-2.76 , -0.04)	-2.33	9	0.04

degrees of freedom, SS denotes the sum of squares, MS denotes the mean squares, and F denotes the F statistics obtained from the quotient of the mean squares for treatment and residuals.

Table 4: ANOVA applied for the BERG and DEFU indexes.

Test	Source	DF	SS	MS	F	p-value
BERG	Treatment	1	1584	1584.2	9.305	0.00689
	Residuals	18	3065	170.3		
DEFU	Treatment	1	51	51.1	0.05	0.826
	Residuals	18	18416	1023.1		

When we apply a Student t-test for independent samples to this data set, comparing posttest means scores between the two groups, we obtain for the BERG indexes a test statistics t equal to 3.0504 that follows, under the null hypothesis, a Student t distribution with 18 degrees of freedom. We can observe that $t^2 = 3.0504^2$ is approximately equal to 9.305 (the F statistics showed in Table 4). For the DEFU indexes, we obtain a test statistics t equal to 0.22371. Observe that, in this case, t^2 is approximately 0.05, equal to correspondent F statistics in Table 4, with identical p-values. If we compare an ANOVA with a standard regression model we will have the same coefficients. The one-factor ANOVA as well the t-test on posttest scores, is

the not satisfactory because it ignores useful information contained in the pretest. To guarantee the useful information we can incorporate the pre-post nature of the data. The variable now is the change score (sometimes called the gain or difference score).

Standard regression model

In this approach, the intercept of the linear model correspond to the mean for the reference group (Group 1), while the coefficient for group (treat) is the difference between their respective means. This could save you the time of having to conduct post-hoc tests to compare the means between groups after running ANOVA.

Table 5: Standard Regression applied for the BERG and DEFU indexes.

Test	Coefficients	Estimate	Std.Error	t-value	p-value
BERG	(Intercept)	56.30	4.126	13.64	6.22E-11
	Treat	-17.80	5.835	-3.050	0.00689
DEFU	(Intercept)	90.00	10.12	8.898	5.22E-08
	Treat	-3.20	14.30	-0.224	0.826

The great advantage of linear regression is that it facilitates the inclusion of several covariates (though this can also be easily accomplished through ANCOVA when you are interested in including just one covariate). The point is simply that we have run the same model three different ways, and produced identical results in each case.

ANOVA on gain scores

Table 6 shows the results of the ANOVA on gain scores.

ANCOVA with pretest-posttest data

An alternative approach is ANCOVA in which the pretest is used as the covariate and

Table 6: ANOVA on gain scores applied for the BERG and DEFU indexes.

Test	Source	DF	SS	MS	F	p-value
BERG	Treatment	1	76.05	76.05	19.98	0.000296
	Residuals	18	68.5	3.81		
DEFU	Treatment	1	36.45	36.45	5.5	0.0307
	Residuals	18	119.30	6.63		

the posttest is used as the dependent variable. The target variable is different from the previous model (crude values versus change). However, note that an ANCOVA is a sequential regression model that examines the effect of the treatment while controlling the pretest results.

Table 7: ANCOVA with pretest-posttest data applied for the BERG and DEFU indexes.

Test	Source	DF	SS	MS	F	p-value
BERG	Pretest	1	1584.2	1584	395.4	3.29E-13
	Treatment	1	2996.5	2996	747.9	1.71E-15
	Residuals	17	68.1	4.00		
DEFU	Pretest	1	51	51	8.304	0.0104
	Treatment	1	18311	18311	2969.8	2.00E-16
	Residuals	17	105	6.00		

The similarity between the t test on the change can also be shown between the t-test and ANOVA on gain scores. Table 7 shows that the null hypothesis is rejected and is assumed the alternative hypothesis. The true difference in means is not equal to zero with 95% confidence interval.

Since Lords Paradox was introduced (Pike, 2004; Eriksson and Häggström, 2014; Xiao et al., 2019), there has been some back and forth among various papers ever since debating whether one should use ANCOVA vs. t-test on change scores. As such, conducting separate power analyses for the t-test vs. ANCOVA makes no sense, as they belong to the same model, and can also be (essentially) equivalent if pretest and posttest are strongly correlated. Lords Paradox (Lord, 1967, 1969) occurs when a continuous covariate is controlled for and the relationship between a continuous outcome and group status indicator changes in both magnitude and direction.

ANOVA with repeated measures

Now that weve gone through that, lets do a repeated measures ANOVA with treatment as the between subjects effect and score as the repeated measure (pre-post).

Table 8 presents the analysis of variance of repeated measures. In this case the total number of subjects is considered as 20. The hypotheses of interest were from the interaction and simple effects test on the equality of the pre-intervention population means between the groups. The interaction test is significant (F = 19.98 with 1 and 18 df, $p < (0.005)$; Therefore, it is concluded that pre / post-media change in group 1 is significantly different from pret/ posttest mean change in group 2 (Figure 1).

Assuming the group factor and the pre/post factor to be fixed effects, Winer (1971) shows that the appropriate F tests are as indicated in the F ratio column of Table 2. All the main effects were significant, in this way:

(1) The ratio MSG/MSE tests the null hypothesis that there is group main effect. This is equivalent to testing whether the sum of the pretest and posttest observations on each subject does not have the same population mean in the two

Table 8: ANOVA with repeated measures applied to groups and treatments, considering the BERG and DEFU indexes.

Test	Source	DF	SS	MS	F	p-value
BERG	**Between Subjects**					
	Groups	1	2512	2512.2 (MSg)	7.588	0.013
	Subjects Within Groups	18	5959	331.1 (Mse)		
	Within Groups					
	Groups x Pre/Post	1	112.22	112.22 (MSp)	58.98	4.36e-07
	(Pre/Post) x Subjects	1	38.03	38.03 (Mgp)	19.98	0.000296
	Within Groups	18	34.25	1.90 (Mspe)		
DEFU	**Between Subjects**					
	Groups	1	34	34.2 (MSg)	0.016	0.900
	Subjects Within Groups	18	37711	2095.1 (Mse)		
	Within Groups					
	Groups x Pre/Post	1	75.63	75.63 (MSp)	22.82	0.000151
	(Pre/Post) x Subjects	1	18.23	18.23 (Mgp)	5.50	0.030684
	Within Groups	18	59.65	3.31 (Mspe)		

groups.

(2) The ratio MSP/MSpE tests the null hypothesis that there is pretest/posttest main effect and is equivalent to testing whether the population mean of the pretest observations is not the same as the population mean of the posttest observations.

(3) The ratio MSGP/MSPE tests the null hypothesis that there is interaction between the group main effect and the pre/post main effect. This ratio also tests

whether the difference between pretest and posttest observations does not have the same population mean in both groups.

4 Conclusion

The first one, a one-factor ANOVA on posttest scores, is the least satisfactory because it ignores potentially useful information contained in the pretest.

There are several ways to treat the data, so it is not necessary to force data to suit the technique (for example, transformations of target variables, categorizing numerical covariates), it is important to find adequate techniques to represent the data.

ANCOVA is no different than the standard linear regression. A t-test on change scores is the simplest part of a repeated measures ANOVA result that is a special case of mixed models.

These results demonstrate that the various F tests in the repeated-measures analysis of variance can be obtained by using simple t tests on linear combinations of the pretest and posttest scores.

Since the difference-score analysis is embedded in the repeated-measures analysis, the repeated-measures analysis provides more information about the data at hand. Fewer assumptions, however, are required in the difference-score analysis. The difference-score analysis assumes only homogeneous variances for the difference scores and a normally distributed error term with mean zero. It is easy to show that if the assumptions of the repeated-measures analysis are satisfied, then the assumptions of the difference-score analysis are also met. However, the converse is not true.

Acknowledgements

We want to thank the financial and academic support from Brazilian federal government agency CAPES (Coordination for the Improvement of Higher Education Personnel)....

References

Abramovitch, L., and Singh, K. (1985). Edgeworth corrected pivotal statistics and the bootstrap. *The Annals of Statistics*, **13**, 116-132.

Bonate, P. L. (2000). *Analysis of pretest-posttest designs*. Chapman and Hall/CRC.

Bonett, D. G., and Woodward, J. A. (1990). Testing residual normality in the ANOVA model. *Journal of Applied Statistics*, **17**, 383–387.

Brogan, D. R., and Kutner, M. H. (1980). Comparative analyses of pretest-posttest research designs. *The American Statistician*, **34**, 229–232.

Brown, B. M., and Hettmansperger, T. P. (2002). Kruskal–Wallis, multiple comparisons and Efron dice. *Australian & New Zealand Journal of Statistics*, **44**, 427–438.

Casella, G., and Berger, R. L. (2002). *Statistical inference*. 2nd Edition. Thomson Press.

Chow, S. C., Shao, J., Wang, H., and Lokhnygina, Y. (2017). *Sample size calculations in clinical research*. Chapman and Hall/CRC.

Cnaan, A., Laird, N. M., and Slasor, P. (1997). Using the general linear mixed model to analyse unbalanced repeated measures and longitudinal data. *Statistics in Medicine*, **16**, 2349–2380.

Cochran, W. G. (1957). Analysis of covariance: its nature and uses. *Biometrics*, **13**, 261–281.

Cox, D. R. (2006) *Principles of Statistical Inference* Cambridge University Press

Crowder, M. (2017). *Analysis of repeated measures*. Routledge.

Dimitrov, D. M., and Rumrill Jr, P. D. (2003). Pretest-posttest designs and measurement of change. *Work*, **20**, 159–165.

Dugard, P., and Todman, J. (1995). Analysis of pretestposttest control group designs in educational research. *Educational Psychology*, **15**, 181–198.

Elashoff, J. D. (1969). Analysis of covariance: A delicate instrument. *American Educational Research Journal*, **6**, 383–401.

Eriksson, K., and Häggström, O. (2014). Lords paradox in a continuous setting and a regression artifact in numerical cognition research. *PLoS One*, **9**, e95949.

Goldenberg, M. M. (2012). Multiple sclerosis review. *Pharmacy and Therapeutics*, **37**, 175–184.

Goldberg, D. E., and Scheiner, S. M. (2001). *ANOVA and ANCOVA: field competition experiments*. In: Scheiner, S. M., and Gurevitch J. Design and analysis of ecological experiments, 2nd ed., p. 77–98. Oxford University Press.

Gourlay, N. (1955). F-test bias for experimental designs in educational research. *Psychometrika*, **20**, 227–248.

Gronow, D. G. C. (1951). Test for the significance of the difference between means in two normal populations having unequal variances. *Biometrika*, **38**, 252–256.

Gupta, J. K., Srivastava, A. B. L., and Sharma, K. K. (1988). On the optimum predictive potential of change measure. *The Journal of Experimental Education*, **56**, 124–128.

Hall, P. (1992). On the removal of skewness by transformation. *Journal of Royal Statistical Society, Series B*, **54**, 221–228.

Hawkins, D. M. (1980). *Identification of outliers*. Chapman and Hall. London.

Huck, S. W., and McLean, R. A. (1975). Using a repeated measures ANOVA to analyze the data from a pretest-posttest design: a potentially confusing task. *Psychological Bulletin*, **82**, 511–518.

Johnson, N. J. (1978). Modified t-test and confidence intervals for asymmetrical population. *Journal of the American Statistical Association*, **73**, 536–544.

Jones, B., and Nachtsheim, C. J. (2009). Split-plot designs: What, why, and how. *Journal of Quality Technology*, **41**, 340–361.

Laird, N. M., and Ware, J. H. (1982). Random-effects models for longitudinal data. *Biometrics*, **38**, 963-974.

Lehmann, E. L. (2012). *"Student" and small-sample theory*. In: Rojo J. (eds). Selected works of E. L. Lehmann, p. 1001–1008. Springer, Boston.

Lim, W. K., and Lim, A. W. (2016). A comparison of usual t-test statistic and modified t-test statistics on skewed distribution functions. *Journal of Modern Applied Statistical Methods*, **15**, 67–89.

Linn, R. L., and Slinde, J. A. (1977). The determination of the significance of change between pre-and posttesting periods. *Review of Educational Research*, **47**, 121–150.

Lord, F. M. (1967). A paradox in the interpretation of group comparisons. *Psychological Bulletin*, **68**, 304–305.

Lord, F. M. (1969). Statistical adjustments when comparing preexisting groups. *Psychological Bulletin*, **72**, 336–337.

McElreath, R. (2018). *Statistical rethinking: A Bayesian course with examples in R and Stan*. Chapman and Hall/CRC.

Nayak, B. K. (2010). Understanding the relevance of sample size calculation. *Indian Journal of Ophthalmology*, **58**, 469–470.

Pagano, M., and Gauvreau, K. (2018). *Principles of biostatistics*. 2nd Edition. Chapman and Hall/CRC.

Pascarella, E. T., Wolniak, G. C., and Pierson, C. T. (2003). Explaining student growth in college when you don't think you are. *Journal of College Student Development*, **44**, 122–126.

Pike, G. R. (2004). Lords paradox and the assessment of change during college. *Journal of College Student Development*, **45**, 348–353.

Rodrigues, I. F., Nielson, M. B. P., and Marinho, A. R. (2008). Evaluation of the physical therapy on the balance and on the quality of life in multiple sclerosis patients. *Revista Neurocincias*, **16**, 269–274.

Rohde, C. A. (2014). *Introductory statistical inference with the likelihood function*. Springer International Publishing.

Satterthwaite, F. E. (1946). An approximate distribution of estimates of variance components. *Biometrics Bulletin*, **2**, 110–114.

Scheffe, H. (1999). *The analysis of variance*. John Wiley & Sons.

Singer, J. M., and Andrade, D. F. (1997). Regression models for the analysis of pretest/posttest data. *Biometrics*, **53**, 729–735.

Stell, R. G. D., Torrie, J. H., and Dickey, D. (1980). *Principles and procedures of statistics: a biometrical approach*. MacGraw-Hill.

Stevens, J. P. (2012). *Applied multivariate statistics for the social sciences*. 3rd Edition. Routledge.

Tu, Y. K., Baelum, V., and Gilthorpe, M. S. (2008). A structural equation modelling approach to the analysis of change. *European Journal of Oral Sciences*, **116**, 291–296.

Vogt, W. P., and Johnson, B. (2011). *Dictionary of statistics & methodology: A nontechnical guide for the social sciences*. Sage.

Xiao, Z., Higgins, S., and Kasim, A. (2019). An Empirical Unraveling of Lord's Paradox. *The Journal of Experimental Education*, **87**, 17–32.

Winer, B. J. (1962). *Statistical principles in experimental design*. McGraw-Hill Book Company.

Zeger, S. L., and Liang, K. Y. (1986). Longitudinal data analysis for discrete and continuous outcomes. *Biometrics*, **42**, 121–130.

Zimmerman, D. W., and Zumbo, B. D. (1993). Relative power of the Wilcoxon test, the Friedman test, and repeated-measures ANOVA on ranks. *The Journal of Experimental Education*, **62**, 75–86.

ARTICLE 2:
Pretest/posttest analysis based on skew-symmetric distributions

ARTICLE 2: Pretest/posttest analysis based on skew-symmetric distributions

Guilherme Zubatch da Cunha [1], and Edson Zangiacomi Martinez [1]

[1] Ribeirão Preto Medical School, University of São Paulo (USP), Ribeirão Preto, SP, Brazil

---

Address for correspondence: Guilherme Zubatch da Cunha, Ribeirão Preto Medical School, University of São Paulo (USP), Av. Bandeirantes 3900, Monte Alegre, CEP 14049-900, Ribeirão Preto, SP, Brazil.

E-mail: zubatch@hotmail.com.

Phone: (+55) 16 3602 2569.

Fax: (+1) 999 888 666.

Abstract: This paper presents a general family of skew-symmetric distributions applied on pretest/posttest in epidemiological studies by considering a longitudinal approach. Situation where the normality assumption is not valid are very common, especially when data present a skewed distribution. For taking into account possible skewness effects, we propose the use of longitudinal analysis from a Bayesian perspective. Results of an application to a real data set related to multiple sclerosis are reported.

Key words: Pretest/posttest; Bayesian; skew-symmetric distributions.

1 Introduction

A common research design is the two-group pretest/ posttest design with one dependent variable. The objective may be to establish the superiority or the equivalence between treatments. Inferring about the target population based on sample results requires special care in planning, factors affecting the response should be evaluated and controlled.

The statistical analysis for these designs can be approached from several viewpoints, usually involving a selection of two samples and a comparison of the results obtained. In general, the response variable is measured on an interval scale or ratio, a metric commonly used in analyses is the difference score for each subject (posttest minus pretest or vice versa). The null hypothesis is that the means or medians of the (relative) differences are the same for each group.

Another option for analyzing this kind of design is to consider the pretest and posttest as a longitudinal data set, the term longitudinal data implies that each subject is measured repeatedly over time. The main goal is to find the pattern (if it exists) in which the variable of interest changes over time. Repeated-measures and longitudinal designs are discussed in details by Winer (1962) and Stell et al. (1980).

The restrictive assumption of Gaussian errors is common in linear mixed model. An alternative that relax this assumption was proposed by (Strandén and Gianola, 1999),

who modelled the residual term using a Students t density. This kind of heavy-tailed distribution allows for more extreme residual values and, as a consequence, deviations from the Gaussian distribution such as referential treatment (Kühn et al., 1994) or other causes of outliers or abnormal phenotypic records (Jamrozik et al., 1994).

Nevertheless, both Gaussian and Student's t distributions are symmetric, and little investigation into alternative approaches which enable skewness on the error term of skewness for the residual term has been done.

The aims of this study are: to include nonsymmetric residual distributions in the linear mixed models; to implement and compare these procedures with the standard linear mixed models approach. Bayesian implementation of skewed distributions using the procedure suggested by Fernández and Steel (1998) and von Rohr and Hoeschele (2002), which involves Markov chain Monte Carlo (MCMC) techniques, requires a Metropolis–Hastings step (Hastings, 1970; Chib and Greenberg, 1995; Gilks et al., 1993) to sample the asymmetry parameter.

2 Methods

In situations where the use of the normal distribution is not adequate, the interest in describing the characteristics of a data set is common, either regarding its skewness or its nature. One way to deal with asymmetry is to use transformations in the data, however Gómez et al. (1993) indicate that an available alternative is to fit an asymmetric model, such models are objects of intense development in the recent statistical literature. The main idea is to replace the normal distribution with other distributions that preserve some of its properties.

2.1 Skew-symmetric distributions

A family of asymmetric distributions is described by Azzalini (1985), assuming the density function

$$f\left(x|\lambda\right) = 2g(x)G(\lambda x),$$

where $x \in \mathbb{R}$, $\lambda \in \mathbb{R}$ is the skewness parameter, $g\left(\cdot\right)$ is a probability density function symmetric about zero and $G\left(\cdot\right)$ is a distribution function such that the derivarite G' exists. A proof of the fact that

$$s\left(\lambda\right) = \int_{-\infty}^{\infty} f\left(x|\lambda\right) dx = 1$$

for any λ is presented by Gómez et al. (1993), considering that $G\left(\cdot\right)$ is a bounded function and $xg\left(x\right)G'(\lambda x)$ is an odd function. In this way,

$$\frac{d}{d\lambda} s\left(\lambda\right) = \int_{-\infty}^{\infty} xg\left(x\right)G'(\lambda x) = 0$$

so that $s\left(\lambda\right)$ is constant. Thus, $s\left(\lambda\right) = 1$ for any λ, since $s\left(0\right) = 1$. A more general family of asymmetric distributions (Azzalini, 1985) considers

$$f\left(x\right) = 2g(x)G(h\left(x\right)), \tag{2.1}$$

where $h\left(x\right)$ is an odd function. A proof that (2.1) is a probability density function is also provided by Gómez et al. (1993), assuming that G' exists.

The skew-normal (SN) distribution introduced by Azzalini (1985) is denoted by $X \sim SN(\mu, \sigma^2, \lambda)$, and its probability density function (pdf) is given by

$$f_I\left(x|\mu, \sigma, \lambda\right) = 2\phi\left(z\right)\Phi\left(\lambda z\right),$$

where $\mu \in \mathbb{R}$ is the location parameter, $\lambda \in \mathbb{R}$ is the skewness parameter, $\phi\left(\cdot\right)$ and $\Phi\left(\cdot\right)$ are the density and the cumulative density functions of the normal distribution,

respectively, and

$$z = \frac{x - \mu}{\sigma}. \tag{2.2}$$

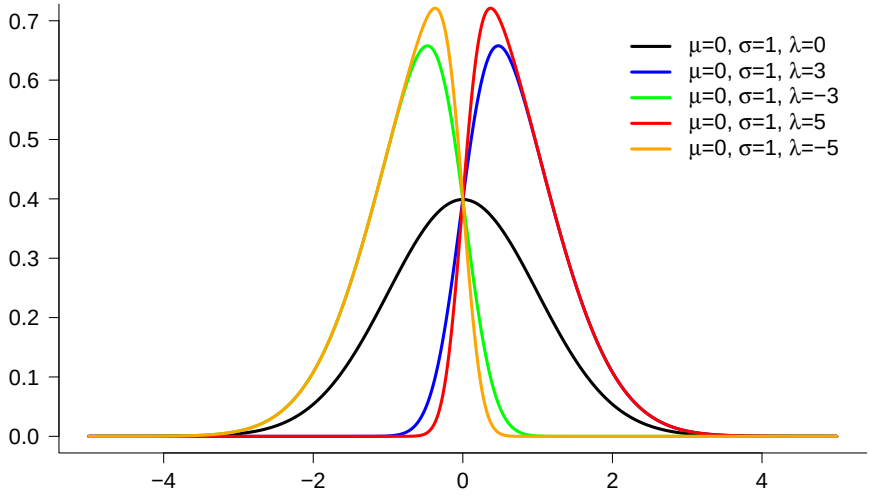

Figure 1: Graphics of the probability density function of the skew-normal (SN) distribution.

It may be noted that when $\lambda = 0$, the normal distribution is recovered, the skewness increases as the λ gets larger. Figure 1 shows that the distribution is skewed on the right when the value of $\lambda > 0$ and it is skewed on the left when $\lambda <0$.

If $\mu = 0$ and $\sigma = 1$, the mean of X is given by

$$E(X) = \delta\sqrt{\frac{2}{\pi}}$$

and the variance of X is given by

$$Var(X) = 1 - \frac{2\delta^2}{\pi},$$

where

$$\delta = \frac{\lambda}{\sqrt{1 + \lambda^2}}.$$

Thus, in the general case, we have

$$E(X) = \mu + \sigma\delta\sqrt{\frac{2}{\pi}}$$

and

$$Var(X) = \sigma^2\left(1 - \frac{2\delta^2}{\pi}\right).$$

In order to get more flexibility, Arellano-Valle et al. (2004) proposed the skew-generalized normal (SGN) distribution, denoted by $X \sim SGN(\mu, \sigma^2, \lambda_1, \lambda_2)$ and with *pdf* given by

$$f_{II}\left(x|\mu, \sigma, \lambda_1, \lambda_2\right) = 2\phi\left(z\right)\Phi\left(\frac{\lambda_1 z}{\sqrt{1+\lambda_2 z^2}}\right),$$

where $\lambda_1 \in \mathbb{R}$, $\lambda_2 \geq 0$ and z is given by (2.2).

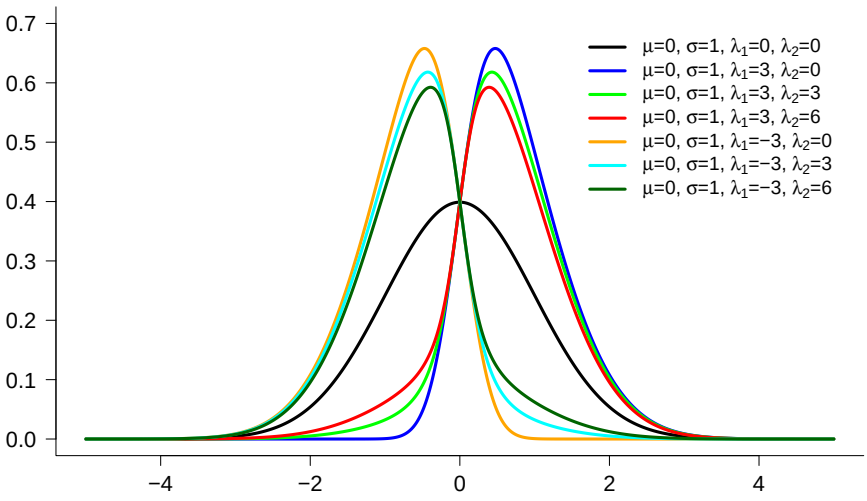

Figure 2: Graphics of the probability density function of the skew-generalized normal (SGN) distribution.

Figure 2 shows that the distribution is asymmetric on the right when the value of $\lambda_1 > 0$ and that asymmetries on the left when $\lambda_1 < 0$.

In addition, the skew-curved normal (SCN) distribution, also introduced by Arellano-Valle et al. (2004), denoted by $X \sim SCN(\mu, \sigma^2, \lambda)$, is a special case of the SGN distribution where $\lambda = \lambda_1$ and $\lambda_2 = \lambda_1^2$. The *pdf* of the SCN distribution is defined by

$$f_{III}(x|\mu, \sigma, \lambda) = 2\phi(z)\,\Phi\left(\frac{\lambda z}{\sqrt{1+\lambda^2 z^2}}\right),$$

where z is given by (2.2). Graphics of the *pdf* of the skew-curved normal (SCN) distribution are given in the Figure 3.

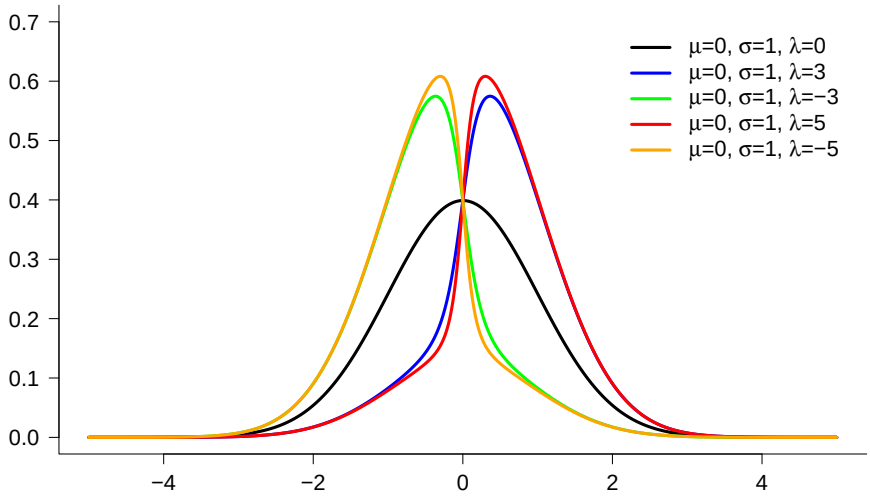

Figure 3: Graphics of the probability density function of the skew-curved normal (SCN) distribution.

The skew-Student-t-normal (StN) distribution was introduced by Gómez et al. (1993), denoted by $X \sim StN(\mu, \sigma^2, \lambda, \nu)$

$$f_{IV}(x|\mu, \sigma, \lambda, \nu) = \frac{2}{\sigma}\phi_t(z, \nu)\,\Phi(\lambda z),$$

where λ is the skewness parameter, z is given by (2.2), $\phi_t(\cdot, \nu)$ denotes the probability density function of the standard Student's t distribution with ν degrees of freedom and $\Phi(\cdot)$ is the standard normal distribution function. If $\nu = 1$, we obtain the skew

Cauchy-normal model. The StN distribution is regarded by Cabral et al. (2008) as a good alternative to model heavy tailed data with strong asymmetrical nature, specially because it has a larger range of skewness than other skewed distributions.

If $v = 1$, there is a Cauchy-Normal model. A special case is where $\lambda = 0$, where we have the density function of the normal Student-t distribution with v degrees of freedom. Figure 4 shows that the distribution is asymmetrical on the right when the value of $\lambda > 0$ and that asymmetries the left when $\lambda < 0$.

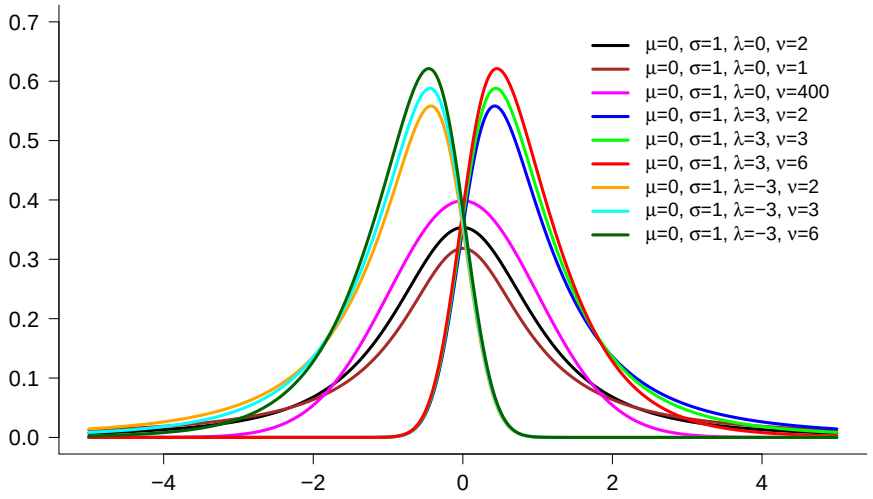

Figure 4: Graphics of the probability density function of the skew-Student-t-normal (StN) distribution.

The moments of order n of the skew Student-t-normal distribution with $\nu > n$, $\mu = 0$ and $\sigma = 1$ is given by

$$E(X^n) = \frac{v^{n/2}\Gamma\left(\frac{n+1}{2}\right)\Gamma\left(\frac{\nu-n}{2}\right)}{\sqrt{\pi}\Gamma\left(\frac{\nu}{2}\right)},$$

for n odd. If n is even, a more complex expression is given by xxx. In particular, the second moment is given by

$$E(X^2) = \frac{v}{v-2}.$$

Thus, if $\mu = 0$ and $\sigma = 1$,we have, for $\nu > 1$,

$$E(X) = \sqrt{\frac{v}{\pi}} \frac{\Gamma\left(\frac{\nu-1}{2}\right)}{\Gamma\left(\frac{\nu}{2}\right)},$$

and an expression for the variance can be obtained from the relation $Var(X) = E(X^2) - E^2(X)$.

3 Real data

As an exemple, let us consider a real data set from a study in physical therapy. Multiple sclerosis (MS) is an unpredictable, often disabling disease of the central nervous system that disrupts the flow of information within the brain, and between the brain and body (Goldenberg, 2012). The main disorders are intense fatigue, balance deficit and motor incoordination. In this study, the effects of physical therapy on the balance and quality of life of MS patients were evaluated. Data was obtained from analytical studies, where 20 subjects were randomized into two groups: Group 1 - submitted to fifteen sessions of physical therapy program for MS, three times per week, and Group 2 - underwent conventional physiotherapy once a week (Rodrigues et al., 2008). The objective of the study is to evaluate whether specific physiotherapy has a better effect in relation to traditional physiotherapy, longitudinally evaluating the alteration of individuals.

Tables 1 and 2 present the values of DEFU index (Functional Determination of Quality of Life Scale in individuals with MS) for each patient of the Groups 1 and 2, respectively, considering the periods before and after the intervention and their respective means. Figure 1 suggests that the effect in Group 1 improves from pre to post.

Table 1: Results from balance tests, through the BERG Balance Scale, and quality of life test, through the DEFU, before and after the fifteen sessions of physical therapy program for MS, three times per weeks (Group 1) (Rodrigues et al., 2008).

	BERG index		DEFU index	
Subject	Pre	Post	Pre	Post
1	56	58	51	58
2	42	46	42	50
3	56	60	66	68
4	52	58	61	66
5	52	59	91	99
6	54	62	109	109
7	53	57	142	144
8	53	60	149	152
9	39	41	62	68
10	53	62	86	86
Mean	51	56.3	85.9	90
SD	37.1	7.0	35.2	33.7

Table 2: Results from balance tests, through the BERG Balance Scale, and quality of life test, through the DEFU, before and after the conventional physiotherapy once a week (Group 2) (Rodrigues et al., 2008).

	BERG index		DEFU index	
Subject	Pre	Post	Pre	Post
11	53	53	63	66
12	48	50	78	76
13	36	38	97	97
14	56	56	88	88
15	18	18	86	88
16	13	13	87	91
17	49	51	79	79
18	54	56	77	78
19	16	19	46	49
20	28	31	153	156
Mean	37.1	38.5	85.4	86.8
SD	17.1	17.1	26.3	26.5

Let us consider an analysis on gain scores, where x_{ia} denotes the pretest values and x_{ib} denotes the posttest values. If one defines $d_i = x_{ia} - x_{ib}$, we have $M(d_i) = M(X_{ia}) - M(X_{ib})$ for each individual i, where $M(.)$ denotes the mean of a variable. The general approach to a gain score analysis is: (1) to compute the gain score, and then (2) analyze those gain scores in a regression model with treatment as the between-subjects factor. The improvement (gain) can be calculated as:

$$Gain = posttest - pretest$$

When you compute a gain score in this manner, a positive gain score indicates that the posttest score was greater than the pretest score, and a negative gain score indicates that the posttest score was less than the pretest score. A linear additive statistical model for the ANOVA, called the effects model, is given by:

$$Y_i = \mu + \tau_i + \epsilon_i,$$

where μ is the overall mean, τ_i are the deviations from the overall mean due to the treatment levels and ϵ_i are the error terms.

Considering that obtaining means for the skew generalized normal (SGN) and skew-curved normal (SCN) distribution is a difficult task, let us consider in our analysis only the skew-normal (SN) and the skew-Student-t-normal (StN) distributions. In the regression model, in both models, let us replace the location parameter μ with

$$\mu_i = \mu_0 + \mu_1 g_i,$$

where g_i is a dummy variable corresponding to the groups ($g_i = 0$ if group 1 and $g_i = 1$ otherwise).

3.1 Bayesian implementation

For fitting the model described in previous section, we consider a Bayesian approach, using a Gibbs sampler (Gelfand and Smith, 1990). In a Bayesian structure, the joint posteriori distribution of the model parameters is obtained by combining the joint prior distribution of the parameters and the likelihood function by equation. For both models based on the skew-normal (SN) and the skew-Student-t-normal (StN) distributions, let us consider the following prior distributions:

$$\mu_0 \sim N(0, 10),$$

$$\mu_1 \sim N(0, 10),$$

$$\sigma \sim gamma(1, 1),$$

and

$$\lambda \sim N(0, 10),$$

where $N(0, 10)$ denotes a normal distribution with mean 0 and variance 10, and $gamma(1, 1)$ denotes a gamma distribution with both parameters equal to 1. To simulate samples from the joint posterior distribution, we could consider MCMC (Markov Chain Monte Carlo) algorithms implemented in the OpenBUGS software, where we only need to specify the distribution for the data and the prior distribution for the parameters. Highest Posterior Density (HPD) intervals were obtained from all parameters of interest. HPD are the shortest interval among all of the possible Bayesian credible intervals. We used the package MCMCpack (Martin et al., 2011) of the R software.

Convergence analysis of the chains for the different models was performed using the Geweke criterion (Geweke, 1992). This approach divides each simulated chain into two

windows, and it thus compares the mean of the sampled values in the first window to the mean of the sampled values in the last window. It is produced a Z statistic based on the difference between the two means divided by the asymptotic standard error of their difference, where the variance is determined by spectral density estimation. This Z statistic approaches a standard normal distribution if the chain has converged.

The conditional predictive ordinate (CPO) proposed by Gelfand and Smith (1990) was used for comparing model fit and complexity of the considered models. For each individual, larger values of CPO imply better fits of the model. As a summary statistic of CPO over all individuals we can utilize the logarithm of the pseudomarginal likelihood (LPML), as proposed by Ibrahim et al. (2001). Larger value of LPML indicates better fit of the regression model.

4 Results

Table 3 shows the results from the models based on the SN and StN distributions. Standard deviations (SD) for each simulated chain, naive standard errors (SE), HPD intervals, Z Geweke statistics and LPML values are also presented. In the Bayesian calculations, a burn-in period of 5000 samples was discarded and the chain was then subsampled at every 200th iterate to get a final sample size of 2500.

All Z Geweke values are into the interval (-2.5,2.5), suggesting a satisfactory convergence for all simulated chains. The LPML values are close to each other, but the LPML for the model based on the SN distribution suggest that this is the model more adequate to these data.

Table 3: Results from Bayesian modeling.

Model		Mean	SD	Naive SE	HPD interval	Z Geweke	LPML
SN	μ_0	0.7141	1.275	0.02550	(-1.9905 2.8369)	-1.86	-9.59
	μ_1	-2.7437	1.702	0.03403	(-6.3014 0.5008)	1.64	
	σ	7.8044	1.970	0.03939	(4.2145 11.6055)	-0.30	
	λ	11.8503	5.830	0.11660	(2.6965 23.4752)	1.57	
StN	μ_0	0.805	1.226	0.02452	(-1.8139 2.7827)	-2.02	-12.84
($\nu = 2$)	μ_1	-2.627	1.636	0.03272	(-6.1233 0.5410)	1.86	
	σ	7.924	2.045	0.04090	(4.3392 12.0264)	-0.47	
	λ	12.074	5.896	0.11791	(2.7830 24.0402)	2.15	
StN	μ_0	0.8577	1.213	0.02426	(-1.6295 2.9704)	-0.51	-15.52
($\nu = 1$)	μ_1	-2.5589	1.564	0.03128	(-5.9456 0.4260)	1.08	
	σ	8.1222	2.130	0.04260	(4.2476 12.2861)	-1.51	
	λ	12.2034	5.955	0.11910	(2.6212 23.9794)	1.43	

5 Conclusion

This study brought out the advantages of using models developed for the skew-normal distribution in addition to the conventional techniques used in data modeling.

In this paper we consider a Bayesian skew-symmetric approach to models where the dependent variable is prestest/posttest. The Bayesian method presented here is quite flexible and easily applicable.

The the data analysis, we considered the following distributions: skew normal, skew Generalized normal, skew cumulative normal , and normal distribution. This approach provides flexibility in capturing a broad range of non-normal and asymmetric behavior that may occur at the level of either within unit error or random effects. In addition, our approach represents an alternative to the use of other methodologies such as normal and thick-tailed symmetric distributions.

Furthermore, the methods presented here can be easily generalized to the multivariate longitudinal data.This kind of skew modelling approach is important in many biostatistical application areas, allowing accurate inference of the parameters while adjusting for the skewness in the data.

Acknowledgements

We want to thank the financial and academic support from Brazilian federal government agency CAPES (Coordination for the Improvement of Higher Education Personnel).

References

Arellano-Valle, R. B., Gmez, H. W., and Quintana, F. A. (2004). A new class of skew-normal distributions. *Communications in Statistics – Theory and Methods*, **33**, 1465–1480.

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics*, **12**, 171–178.

Cabral, C. R. B., Bolfarine, H., and Pereira, J. R. G. (2008). Bayesian density estimation using skew student-t-normal mixtures. *Computational Statistics & Data Analysis*, **52**, 5075–5090.

Chib, S., and Greenberg, E. (1995). Understanding the Metropolis-Hastings algorithm. *The American Statistician*, **49**, 327–335.

Fernández, C., and Steel, M. F. (1998). On Bayesian modeling of fat tails and skewness. *Journal of the American Statistical Association*, **93**, 359–371.

Gelfand, A. E., and Smith, A. F. (1990). Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association*, **85**, 398–409.

Geweke, J. (1992). *Evaluating the accuracy of sampling-based approaches to the calculation of posterior moments (with discussion).* In J. M. Bernardo et al. (Eds.), Bayesian Statistics 4 (pp. 156163). Oxford: Oxford University Press.

Gilks, W. R., Clayton, D. G., Spiegelhalter, D. J., Best, N. G., McNeil, A. J., Sharples, L. D., and Kirby, A. J. (1993). Modelling complexity: applications of Gibbs sampling in medicine. *Journal of the Royal Statistical Society: Series B*, **55**, 39–52.

Gómez, H. W., Venegas, O., and Bolfarine, H. (2007). Skewsymmetric distributions generated by the distribution function of the normal distribution. *Environmetrics*, **18**, 395–407.

Goldenberg, M. M. (2012). Multiple sclerosis review. *Pharmacy and Therapeutics*, **37**, 175–184.

Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, **57**, 97–109.

Ibrahim, J., Chen, M., and Sinha, D. (2001). *Bayesian Survival Analysis*. SpringerVerlag, New York.

Jamrozik, J., Strandén, I., and Schaeffer, L. R. (2004). Random regression test-day models with residuals following a Student's-t distribution. *Journal of Dairy Science*, **87**, 699–705.

Kühn, O., May, V., and Schreiber, M. (1994). Dissipative vibrational dynamics in a curvecrossing system. *The Journal of Chemical Physics*, **101**, 10404–10415.

Martin, A. D., Quinn, K. M., and Park, J. H. (2011). MCMCpack: Markov chain Monte Carlo in R. *Journal of Statistical Software*, **42**, 1–21.

Rodrigues, I. F., Nielson, M. B. P., and Marinho, A. R. (2008). Evaluation of the physical therapy on the balance and on the quality of life in multiple sclerosis patients. *Revista Neurocincias*, **16**, 269–274.

Stell, R. G. D., Torrie, J. H., and Dickey, D. (1980). *Principles and procedures of statistics: a biometrical approach*. MacGraw-Hill.

Strandén, I., and Gianola, D. (1999). Mixed effects linear models with t-distributions for quantitative genetic analysis: a Bayesian approach. *Genetics Selection Evolution*, **31**, 25–42.

von Rohr, P., and Hoeschele, I. (2002). Bayesian QTL mapping using skewed Student-t distributions. *Genetics Selection Evolution*, **34**, 1.

Winer, B. J. (1962). *Statistical principles in experimental design*. McGraw-Hill Book Company.