PhUSE 2010

Paper SP06

Useful Tips for Analysis of Variance (ANOVA) in Multicenter, Placebo Controlled Clinical Trials

Joanna Romaniuk, Quanticate, Warsaw, Poland

ABSTRACT

Analysis of Variance (ANOVA) is commonly performed on the data coming from multicenter, placebo controlled clinical trials in order to evaluate the size of the difference in efficacy between the study medication and placebo. PROC GLM and PROC MIXED procedures available in SAS® are useful tools that make it quite easy to conduct the analysis of variance. However, as far as ANOVA is concerned, it is crucial to understand the specificity of the data and to pay enough attention to the verification of model assumptions. The aim of the proposed paper is to provide some not obvious but very useful tips on how to handle the potential difficulties and address questions that may come up with the data coming from multicenter, placebo controlled clinical trials.

INTRODUCTION

Analysis of Variance is a statistical tool used to identify differences between experimental group means. ANOVA is commonly used in experimental designs with one dependent variable that is a continuous numerical parametric outcome measure and multiple experimental groups within one or more independent (categorical) variables. These independent variables are called factors and groups within each factor are referred to as levels. Depending on the number of factors included in the model we can distinguish *one-way* and *n-way* ANOVA. *One-way* ANOVA can be used when the researcher wants to examine the influence of only one independent variable (*factor*) on the dependent variable. This type of ANOVA simultaneously compares two or more group means that are based on independent samples from each group. In clinical trials it might be applied for comparing mean responses among number of drug dose groups or among patient's background information, such as age group or race (e.g. verify if the drug dose in randomized clinical trial significantly affects the patients' blood cholesterol level). The *Two-way* ANOVA is a statistical method for simultaneously analyzing two factors that affect a response. As in the *one-way* Analysis of Variance, there is a group effect (e.g. treatment group), but it also includes another source of variation – a blocking factor. Variation of the blocking factor can be separated from the error variation in order to give more precise group comparisons. In most types of ANOVA used in clinical trials, the main question the researcher wants to answer is whether there are any differences among the group population means based on the sample data. The null hypothesis says that 'there is no group effect' ('the mean responses are the same for all groups'). The alternative hypothesis is that 'the group effect is important' ('the group means differ for at least one pair of groups'). Since clinical studies frequently use factors such as study center, age group, gender as a blocking factor, the *two-way* Analysis of Variance is one of the most common ANOVA methods used in clinical data analysis.

THE FUNDAMENTALS OF THE ANOVA

In general, the ANOVA method seeks to detect sources of variation in the values of dependent variable and divide the total variability into components associated with each source. The total variability is the sum of squared deviations of each measurement from the overall mean and can be decomposed into a sum of squares (SS) due to suspected sources of variation (model sum of squares) and a sum of squares (SS) resulting from the error:

$$SS(Total) = SS(Model) + SS(Error)$$

The above sums of squares can be expressed as follows:

- ***Model Sum of Squares (Sum of Squares Between):***

$SS(Model) = \sum_{i=1}^{k} n_i(\overline{y}_i - \overline{y})^2$ with $k-1$ degrees of freedom,
n_i - number of observations in the *i*-th group,
$\overline{y}_i$ - mean value of the dependent variable in the *i*-th group,
$\overline{y}$ - overall mean,

k - number of categories (groups) of the factor (independent variable).

- ***Error Sum of Squares (Sum of Squares Within):***

$SS(Error) = \sum_{i=1}^{k} \sum_{j=1}^{n_i} (y_{ij} - \overline{y}_i)^2$ with $N - k$ degrees of freedom,
n_i - number of observations in the *i*-th group,
$\overline{y}_i$ - mean value of the dependent variable in the *i*-th group,
$\overline{y}$ - overall mean,
k - number of categories (groups) of the factor (independent variable),
N - number of observations from all groups.

The ANOVA is performed using *F* statistic based on the ratio of between-group to within-group variance estimates:

- ***F Statistic:***

$F = MS(Model)/MS(Error)$.

where:

- ***Model Mean Square:***

$MS(Model) = SS(Model)/(k-1)$.

- ***Error Mean Square:***

$MS(Error) = SS(Error)/(N-k)$.

ANOVA model can be presented in the table format:

Source of variation	*Sum of squares*	*DF*	*Mean Square*	*F Statistic*
Model	$SS(Model)$	$k-1$	$SS(Model)/(k-1)$	$F = MS(Model)/MS(Error)$
Error	$SS(Error)$	$N-k$	$SS(Error)/(N-k)$	
Total	$SS(Total)$	$N-1$		

When more than two group means are compared in ANOVA (e.g. there are three treatment groups in the study), the *F* statistic will only tell us whether there are significant differences in the group means as a whole. It will not tell us what are the differences between each groups and which group means differ from each other. Thus, a statistically significant Analysis of Variance is usually followed up with a multiple comparison procedures with the purpose of identifying which group means differ from each other. These tests are referred to as post hoc tests or multiple comparison procedures (MCPs). Post hoc tests involve multiple pair-wise comparisons. There are many post hoc tests available in SAS®. Most of them use standardized differences between the group means: t-Student's test, Dunnett's test, Bonferroni's test, Sidak's test or Scheffé's test. The standardized differences between the group means can be expressed in the following way: $t_{ij} = (\overline{y}_i - \overline{y}_j)/\hat{\sigma}_{ij}$, where i and j indicate the numbers of compared groups, $\overline{y}_i$ and $\overline{y}_j$ are the means or least square means estimated for i and j groups, $\hat{\sigma}_{ij}$ is the square root from the estimated variance for $\overline{y}_i - \overline{y}_j$ that can be expressed by the following formulas depending on the compared group means:

- $\hat{\sigma}_{ij}^2 = s^2(1/n_i + 1/n_j)$ for arithmetic means, where n_i and n_j are the numbers of subjects in the given groups and s^2 is the mean squared error with v degrees of freedom,
- $\hat{\sigma}_{ij}^2 = s^2(1/w_i + 1/w_j)$ for weighted arithmetic means where w_i and w_j are the sums of weights for groups i and j.
- $\hat{\sigma}_{ij}^2 = s^2 l_i'(\mathbf{X}'\mathbf{X})^{-l_j}$ for least square means defined as the linear combinations of parameters estimated in the model for each of the groups $li'b$ and $lj'b$.

When the statistic t_{ij} fulfils the following criteria: $P(|t_{ij}| \geq t_\alpha) = \alpha$, where t_α is the critical value from t-Student distribution for the given significance level α, the differences between compared group means can be assumed significant.

In order to perform analysis of variance for fixed effects PROC ANOVA procedure may be used only for balanced data and PROC GLM for both balanced and unbalanced data. PROC GLM is also useful in both

random effects models and repeated measures models (Multivariate ANOVA – MANOVA). PROC GLM procedure is also useful for random effects models, however in terms of random effects modelling, a more efficient procedure is available in SAS® - PROC MIXED, which can efficiently estimate various ANOVA models. When the researcher intends to perform the Analysis of Variance for categorical dependent variable, he can employ PROC CATMOD procedure. For generalized linear models estimation purposes PROC GENMOD procedure is available in SAS®.

BALANCED AND UNBALANCED DATA

In order to properly perform analysis of variance, the analyst has to understand how an unbalanced data set differs from a balanced one. ANOVA is straightforward when an experimental design is balanced, however the unequal cell sizes influence the computation of means, *F*-statistic and hypotheses tested. Then if incorrect computer procedures are applied, incorrect conclusions can be drawn, that is why it is crucial to be aware of the structure of the data.

The data design is balanced when all cell sizes are exactly equal. An unbalanced design is one in which the cell sizes are not exactly equal or/and some data are missing. Unequal cell sizes might be planned (e.g. sample sizes drawn in order to reflect unequal population sizes) or not (e.g. due to subjects withdrawals). Regardless of the reason of the unbalanced design, the use of simple ANOVA statistical procedures is not appropriate. The most often used solution to the problem of unbalanced data is to choose the appropriate Sum of Squares Test out of four tests available in SAS ®. These four tests give the same results when a design is balanced but when it is unbalanced, the sum of squares test different sets of null hypotheses.

The basic model for a simple two-factor ANOVA can be stated as follows: $y_{ijk} = \mu + \alpha_i + \beta_j + \alpha\beta_{ij} + \epsilon_{ijk}$, where the y_{ijk} are the measurements for subject k in the level i of factor A and in the level j of factor B,; μ is the overall mean; α_i's are the effects of factor A; β_j's are the effects of factor B; the $\alpha\beta_{ij}$'s are the interaction parameters and the ϵ_{ijk}'s are the subjects' error terms. *Two-way* ANOVA tests three hypotheses. One for the main effect for A ($H_0 : \alpha_1 = \alpha_1 = ... = \alpha_a = 0$); one regarding the main effect of B ($H_0 : \beta_1 = \beta_1 = ... = \beta_B = 0$) and one for interaction between A and B ($H_0 : all \quad \alpha\beta_{ij} = 0$).

SAS® Type I sum of squares is known as 'sequential sum of squares' and reflects the gradual decrease in error sums of squares as an effect is introduced to the model. Each term is adjusted for all terms previously fit in the model. Therefore, for a model in which the effects are in the following order: A, B, A*B, the sum of squares will reflect the following effects: $SS_A = SS_{re}(\alpha|\mu)$, $SS_B = SS_{re}(\beta|\mu, \alpha)$, $SS_{AB} = SS_{re}(\alpha\beta|\mu, \alpha, \beta)$. SS_A will reflect only an error for A, ignoring B and the interaction, SS_B will reflect the effect for B, but after the adjustment for the effect of A, nevertheless ignoring the interaction. S_{AB} will reflect the interaction effect, after the adjustment for both main effects. Therefore, the SS_{AB} will reflect the correct source of variability. But for unbalanced designs, the SS_A reflects A and also B and A*B; SS_B reflects B and also A*B. In type I test the total variability is the sum of particular sums of squares: $SS_{Total} = SS_A + SS_B + SS_{AB} + SS_{error}$. Thus, Type I Test is suitable only for balanced designs.

Type II sum of squares available in SAS® is referred to as 'EAD' – '**E**ach **AD**justed for the other'. Whether the model statement is ordered A, B, A*B or B, A, A*B the sums of squares will reflect the following: $SS_A = SS_{re}(\alpha|\mu, \beta)$, $SS_B = SS_{re}(\beta|\mu, \alpha)$ and $SS_{AB} = SS_{re}(\alpha\beta|\mu, \alpha, \beta)$. Both main effects are adjusted for the other, ignoring the interaction effects and the interaction sum of squares is corrected for the lower-order effects. Type II sums of squares is inappropriate if the interaction term cannot be assumed to be zero, which is rarely true.

Type III sums of squares are those recommended for general use in the ANOVA. They are referred to as 'partial sum of squares'. Every effect is adjusted for all other effects listed in the model statement: $SS_A = SS_{re}(\alpha|\mu, \beta, \alpha\beta)$, $SS_B = SS_{re}(\beta|\mu, \alpha, \alpha\beta)$ and $SS_{AB} = SS_{re}(\alpha\beta|\mu, \alpha, \beta)$. The Type III sum of squares will test the proper hypotheses $H_0 : \alpha_1 = \alpha_1 = ... = \alpha_a = 0$), $H_0 : \beta_1 = \beta_1 = ... = \beta_B = 0$, $H_0 : all \quad \alpha\beta_{ij} = 0$ and are considered the most appropriate for most research settings.

Type IV sums of squares are equal to the Type III sums of squares if there are no missing cells. However, the Type IV sums of squares are preferred if any cell size equals zero.

Assume analyzing data from multicenter placebo-controlled clinical trial with three treatment groups (A, B and Placebo) performed in 6 sites (1, 2, 3, 4, 5, 6). The primary endpoint is the worst possible pain score rated by patients in the 24 hour post surgery. Data extract can be seen below:

Obs	*Subject*	*Center*	*Race*	*Treatment*	*Pain*
1	1001	1	Black	A	1
2	1002	1	Black	B	2
3	1003	1	Black	Placebo	10
4	1004	1	Black	A	2
5	1005	1	Black	B	7
6	1006	1	Black	Placebo	9
7	1007	1	Black	A	3
8	1008	1	Black	B	8
9	1009	1	Black	Placebo	8
10	1010	1	Black	A	3
…	…	…	…	…	…

In order to investigate the design of the data the PROC FREQ procedure has to be performed:

```
ods rtf file="C:\mydata.rtf"  style=journal;
proc freq data=pain;
       tables treatment*center/ norow nocol nopercent;
run;
ods rtf close;
```

The procedure generates cross-table by treatment and center. Different numbers of observations in each table cell indicate that the data design is unbalanced.

Table of Treatment by Center							
Treatment(Treatment)	*Center(Center)*						
Frequency	*1*	*2*	*3*	*4*	*5*	*6*	*Total*
A	10	5	6	8	9	18	56
B	10	5	6	7	10	18	56
Placebo	9	5	7	7	9	18	55
Total	29	15	19	22	28	54	167

In order to compare different types of sums of squares computed for these unbalanced data the PROC GLM procedure has to be performed:

```
ods rtf file="C:\print.rtf" style=journal;
       proc glm data=pain;
              class treatment center;
              model pain=treatment center treatment*center;
              lsmeans treatment / pdiff;
              output out=predicted r=Residual predicted=Predicted
       rstudent=Studentiz;
       run;
ods rtf close;
```

The results are given below:

Source	*DF*	*Type I SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	481.56015	240.78007	61.17	<.0001
Center	5	139.91569	27.98313	7.11	<.0001
*Treatment*Center*	10	185.29061	18.52906	4.71	<.0001

Source	*DF*	*Type II SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	476.38654	238.19327	60.52	<.0001
Center	5	139.91569	27.98313	7.11	<.0001
*Treatment*Center*	10	185.29061	18.52906	4.71	<.0001

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	359.27493	179.63746	45.64	<.0001
Center	5	138.89181	27.77836	7.06	<.0001
*Treatment*Center*	10	185.29061	18.52906	4.71	<.0001

Source	*DF*	*Type IV SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	359.27493	179.63746	45.64	<.0001
Center	5	138.89181	27.77836	7.06	<.0001
*Treatment*Center*	10	185.29061	18.52906	4.71	<.0001

The sums of squares that are most appropriate for unbalanced data design with all cells non-missing (assuming that the interaction in the model is significant) are the Type III Sums of Squares. Although in this example all types of sums of squares indicate that all factors in the model are significant, we can notice that Type III SS calculated for treatment (359.27493) are sharply different from Type I SS computed for treatment (481.56015) and selection of inappropriate test may lead to false conclusions. For example, consider unbalanced design with empty cells:

Table of Treatment by Center							
Treatment(Treatment)	*Center(Center)*						
Frequency	*1*	*2*	*3*	*4*	*5*	*6*	*Total*
A	10	5	6	**0**	9	18	48
B	10	5	**0**	7	10	18	50
Placebo	**0**	5	7	7	9	18	46
Total	20	15	13	14	28	54	144

Different sums of squares computed for such a data design may bring different results. As we can see in the next four tables, with the assumed significance level $\alpha = 0.05$, Type I Sum of Squares indicates that treatment effect is not statistically significant (*Pr* > *F*=0.0502) in other words – it has no impact on the level of post-surgical pain. However, Types II, III and IV Sums of Squares indicate significant relationship between the treatment group and the level of pain reported by subjects within 24 hours post surgery ((*Pr* > *F)*<0.05). For such a data design (empty cells) the most appropriate sums of squares are the Type IV Sums of Squares.

Source	*DF*	*Type I SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	19.44379	9.72189	3.06	**0.0502**
Center	5	46.28160	9.25632	2.92	0.0158
*Treatment*Center*	7	195.46173	27.92310	8.80	<.0001

Source	*DF*	*Type II SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	28.74791	14.37395	4.53	**0.0126**
Center	5	46.28160	9.25632	2.92	0.0158
*Treatment*Center*	7	195.46173	27.92310	8.80	<.0001

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	35.60088	17.80044	5.61	**0.0046**
Center	5	45.97286	9.19457	2.90	0.0164
*Treatment*Center*	7	195.46173	27.92310	8.80	<.0001

Source	*DF*	*Type IV SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	27.21133	13.60566	4.29	**0.0158**
Center	5	40.85383	8.17076	2.57	0.0296
*Treatment*Center*	7	195.46173	27.92310	8.80	<.0001

An example of balanced data design can be seen in the table below (numbers of observations in each table cell are equal). It is vital to remember that Type I sums of squares in the ANOVA can be interpreted only for balanced data. Unbalanced data requires Type II, III or IV sums of squares.

Table of Treatment by Center							
Treatment(Treatment)	*Center(Center)*						
Frequency	*1*	*2*	*3*	*4*	*5*	*6*	*Total*
A	6	6	6	6	6	6	36
B	6	6	6	6	6	6	36
Placebo	6	6	6	6	6	6	36
Total	18	18	18	18	18	18	108

It is worth mentioning that sums of squares for unbalanced data are computed with the use of least squares means (the estimates for group means obtained from the ANOVA model). Assume that we want to build *two-way* ANOVA model including interaction. The estimated parameter for group *i* mean $\mu i. = \sum_{j=1}^{p} \mu_{ij}/p$ can be expressed as: $\hat{\mu_{i.}} = (\overline{y}_{i1} + \overline{y}_{ij} + ... + \overline{y}_{ip})/p = \sum_{j=1}^{p} \overline{y}_{ij}/p$, while the estimated parameter

for group *j* mean $\mu.j = \sum_{i=1}^{k} \mu_{ij}/k$ can be denoted as: $\hat{\mu_{.j}} = (\overline{y}_{1j} + \overline{y}_{2j} + ... + \overline{y}_{ij})/k = \sum_{j=1}^{k} \overline{y}_{ij}/k$.

The estimate for μ_{ij} is obtained as $\hat{\mu}_{ij} = \hat{y}_{ij}$, where $i = 1, ..., k$ - category of factor A; $j = 1, ..., p$ - category of factor B; $l = 1, ..., n$ - number of observation in the *i*-th category of factor A and *j*-th category of factor B; μ_{ij} - average value of dependent variable in *i*-th category of factor A and *j*-th category of factor B; $\mu_{i.}$ - average value of dependent variable in *i*-th category of factor A; $\mu_{.j}$ - average value of dependent variable in *j*-th category of factor B.

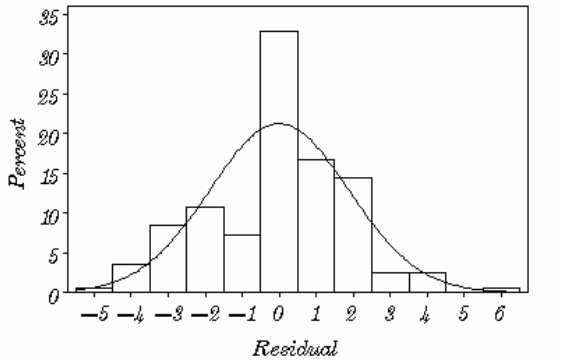
MODEL ASSUMPTIONS

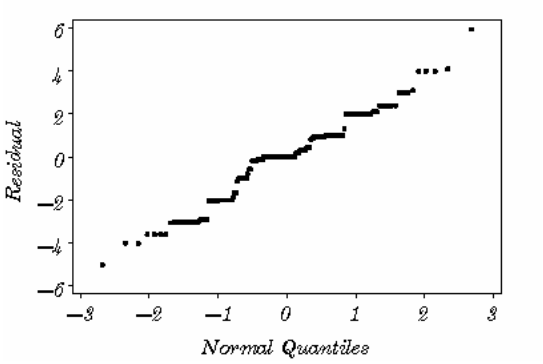
Before conducting ANOVA, researcher needs to consider three fundamental assumptions that underlie the analysis: error components associated with the scores of the dependent variable should be independent of each other, normally distributed with zero mean and an unknown but fixed variance. The normality assumption implies that every individual residual should be derived from a normal distribution. Equality of variance refers to the variance of the residuals, which have to be the same for all treatments. Although these assumptions are discussed separately, in practice they are interconnected – a violation of one of these assumption often affects the others. Failure to meet one or more of these assumptions has impact on the significance levels and the sensitivity of the *F* test. Therefore, deviations from these assumptions must be tested and corrected before the statistical analysis and interpretation of the results. Discrepancies between the data and the assumed model might be detected by studying the residuals (deviations from the observed and the predicted values according to the model). Residual plots can be useful to detect assumptions violations such as auto-correlated error (non-independence), variance heterogeneity (unequal variance) and the presence of outliers. If the ANOVA assumptions are valid, a residual plot (scatter plot between the predicted values and the residuals) will have a random distribution. When the residual plot has a systematic unexplained pattern, the ANOVA model will be inappropriate. A cyclic pattern in the residual plot is an indication for auto-correlation, non-independent error. If the residuals are not independent of each other, the validity of the *F* test can be impaired seriously. There is no transformation or adjustment that can mitigate the lack of independence of error. The lack of autocorrelation assumption should be secured by the proper randomization. If the residuals are auto-correlated the design of the experiment or the way in which the analysis is performed must be changed to deal with this problem. Therefore, it is very important to examine the residuals before performing ANOVA and interpreting the results. Residuals can be examined for homogeneity and normality by drawing box plots by treatments and normal probability plot, respectively, using the PROC UNIVARIATE in SAS®. The example of SAS code that might be useful in the verification of model assumptions is presented below:

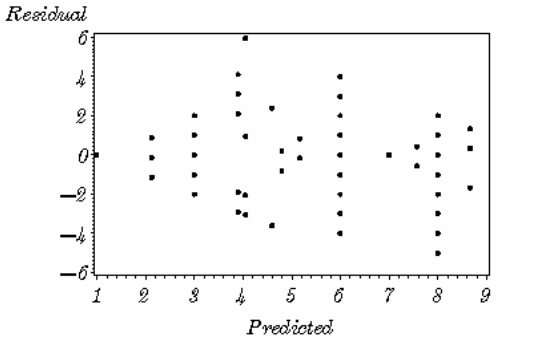
```
ods rtf file="C:\print.rtf" style=journal;
       proc glm data=pain;
              class treatment center;
              model pain=treatment center treatment*center  ;
              lsmeans treatment / pdiff;
              output out=predicted r=Residual predicted=Predicted
       rstudent=Studentiz;
       run;
ods rtf close;

data pred1;
       set predicted;
       Observation=_n_;
       label studentiz="Studentized residual";
run;

ods rtf file="C:\mydata.rtf" style=JOURNAL;
goptions reset = all;
libname gfont0 (sashelp);
goptions keymap=win1250;
goptions device=win target=winprtm FTEXT=centxi htext=3 HSIZE=4 VSIZE=3;
symbol1 color=black value=dot;
proc univariate data=pred1 normal;
       var residual;
       histogram residual / midpoints= 0 1 2 3 4 nrows=4 normal;
       qqplot residual;
run;

proc gplot data=pred1;
  plot Residual*Predicted Residual*Observation Studentiz*Predicted;
run;
quit;
ods rtf close;
```

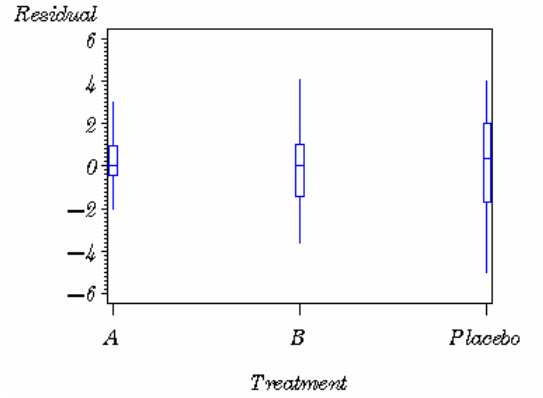
```
SYMBOL1 INTERPOL=BOX VALUE=NONE c=blue width = 1 bwidth = 1;
goptions hsize=4 vsize=3 htext=1;

proc gplot data=pred1 ;
       plot residual*treatment / hminor=0;
       run;
quit;
```

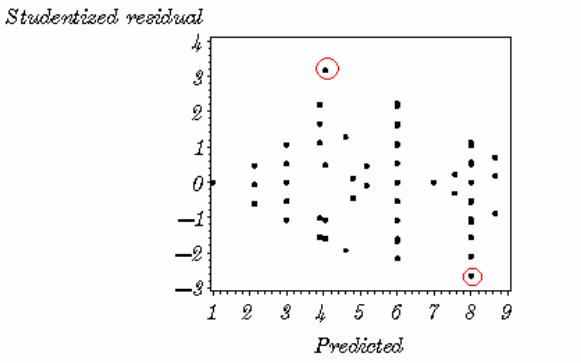
The results are as follows:











Both, histogram of residuals and quantile-quantile plot indicate non-normality, thus one of the ANOVA model assumptions is violated. Analysis of Residual vs Predicted values scatter plot does not show any systematic unexplained or cyclic pattern, therefore it can be assumed that residuals are not auto-correlated. Box-plots generated for residuals for each treatment group show unequal variances. What is more, it can be read from the Studentized residuals vs Predicted values scatter plot that there are two outliers (absolute values of studentized residuals greater than 2.5) that might distort the results of the analysis.

When the data seriously violates ANOVA assumptions, researchers have a few options:

(1) detect and eliminate outliers,
(2) apply a transformation to the response variable that will make it conform more closely to the assumed distribution,
(3) use a non-parametric (rank based) test (e.g. Kruskal-Wallis test),
(4) fit a different model, one that requires different distributional assumptions.

Most commonly used options include elimination of outliers and data transformations and these two methods will be discussed in this paper. The reason for failing ANOVA assumptions concerning homogeneity and normality may lie in outliers. Outliers are cases with unusual or extreme values on a particular variable, possibly indicating experimental error (e.g. participant failure to follow instructions, coding error, fatigue). Outliers might be detected by plotting the standardized residuals against predicted values. When the absolute value of the standardized residual is greater than 2.5, such an observation can be treated as an outlier and requires further investigation – the researcher has to verify whether the outliers result from the experimental error and if so, they have to be eliminated from the analyses or adequately adjusted to the distribution of the empirical data.

If elimination of outliers is not enough to fulfill model assumptions, an appropriate data transformation can be used in order to conform more closely to the assumptions. Data transformations are undertaken with two objectives: (1) to make error variances homogenous, (2) to produce normal error distribution. The data transformation implies the replacement of each observation by some function of its magnitude, followed by ANOVA. Therefore, the original data is changed into a new scale, resulting in observations that are expected to fulfill the assumptions of normality and homogeneity of variance. Since the transformation scale is used for all observations, the mean comparisons remain valid. Square-root and logarithmic transformations are the most commonly used for ANOVA of problematic data.

Square-root transformation is appropriate for data consisting of small numbers from rare events. For such data, the variance is proportional to the mean. If the treatment standard deviation is proportional to the treatment mean and the treatment effects are multiplicative, the log transformation is the best one. If the data contains zeros, 0.5 or 1 is added to the original data before square-root or log transformations. If the relationship between the treatment means and variances is unknown, it is recommended to use the data to estimate the appropriate transformation. Box and Cox proposed the power transformation where:

$$Y_t = \begin{cases} y^{\lambda} & \text{when } \lambda \neq 0 \\ \log y & \text{when } \lambda = 0 \end{cases},$$

Y_t is the transformed response and λ is the integer varying over the range of -3 to 3.
The following steps describe the power method:

(1) Estimate the treatment means $\overline{Y}$ (single factor or treatment combination means in case of two or more factors) and treatment standard deviations S.

(2) Calculates the logs of the $\overline{Y}$'s and the logs of the S's.

(3) Plot log(S) on log($\overline{Y}$) and examine for a linear relationship. A strong nonlinear relationship means that a power transformation is not suitable for such data and non-parametric, distribution-free methods such as Wilcoxon rank-sum test, Kruskal-Wallis test or Mann-Whitney U test are recommended.

(4) Regress log(S) on log($\overline{Y}$) and test for a linear relationship. If the relationship is significant the data should be transformed and the regression coefficient estimated ($\log(\overline{Y}) = \alpha + \beta \log S$).

(5) Estimate the power λ of the transformation by subtracting the regression coefficient β from 1. The value of λ indicates the suitable transformation. The most commonly used transformations and their power values are:

β	λ	Transformation
2.0	-1	Reciprocal
1.0	0	Log
0.66	0.33	Cubic root
0.5	0.5	Square root

To maintain the metric interpretation to the original scale, the transformed means and confidence intervals can be back transformed and reported within parenthesis along with the transformed means.

The most appropriate transformation can be easily determined by the SAS® system using the PROC TRANSREG procedure:

```
data pain1;
       set pred1;
       pain=pain+1;
       if (studentiz gt 2.5 or studentiz lt -2.5) then delete;
run;

ods rtf file="C:myfile.rtf"  style=JOURNAL;
       proc transreg data=pain1;
              model boxcox(pain)=class(treatment center treatment*center);
       run;
ods rtf close;
```

Results of the PROC TRANSREG procedure are presented below. It can be read from the table that the best transformation for the analyzed data is the power transformation with $\lambda = 0.75$. Thus, such a transformation should be performed on the data:

```
data pain2;
       set pain1;
       pain1=pain**.75;
run;
```

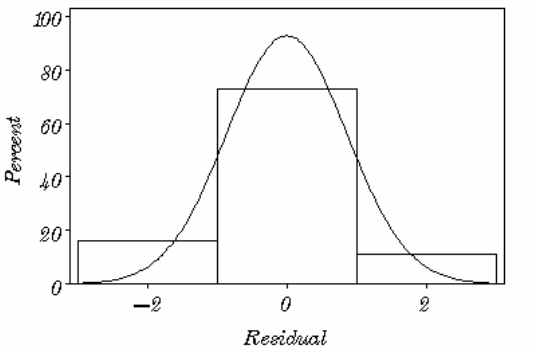
Transformation Information for BoxCox(Pain)

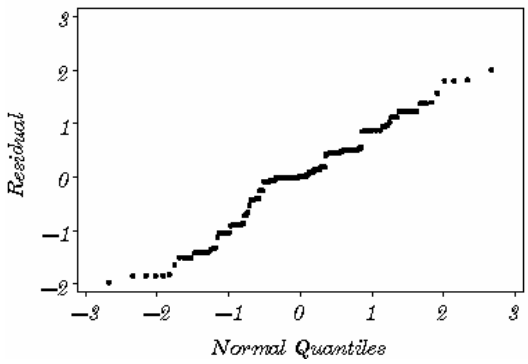
Lambda		*R-Square*	*Log Like*	
-3.00		0.59	-391.519	
-2.00		0.59	-273.629	
-1.00		0.60	-180.807	
0.50		0.59	-114.446	*
0.75		**0.59**	**-113.141**	<
1.00	+	0.58	-114.409	*
2.00		0.54	-140.730	
3.00		0.50	-190.481	

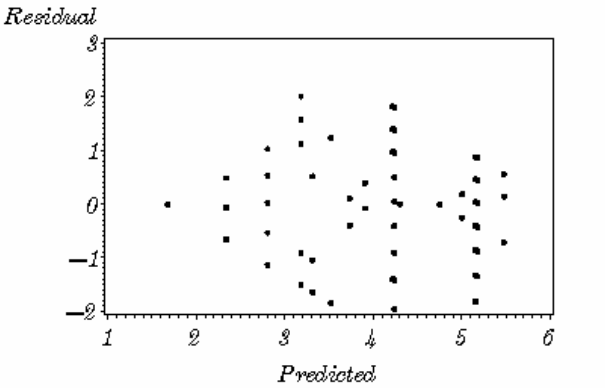
< - ***Best Lambda***
* - *Confidence Interval*
+ - *Convenient Lambda*

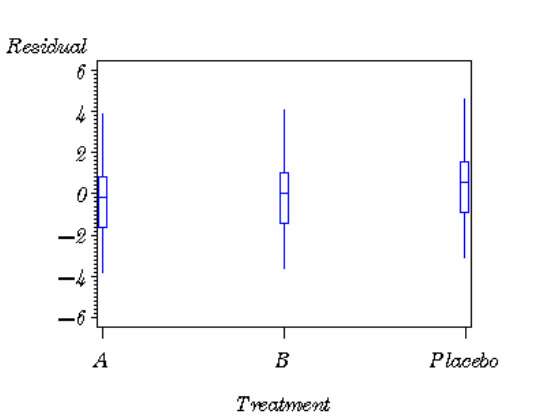
Algorithm converged.

ANOVA results for the transformed data are presented below. Residuals plots indicate that transformed data fulfils ANOVA assumptions (normality of residuals– histogram and quantile-quantile plot indicate that residuals distribution is similar to normal distribution, independence of error term – Residual vs Predicted values scatter plot shows no systematic pattern, homogeneity of variance of error term – box plots for residuals in each treatment group indicate similar variances), thus results might be interpreted.









Results obtained from *two-way* ANOVA model for transformed data show that treatment group center and the interaction between the treatment group and center have the significant influence on the level of pain suffered by subjects in the first 24 hours PS (*Pr*>*F*=<0.001).

Source	*DF*	*Sum of Squares*	*Mean Square*	*F Value*	*Pr > F*
Model	17	181.94724	10.70277	14.18	<.0001
Error	147	110.91501	0.75452		
Corrected Total	164	292.86226			

Source	*DF*	*Type III SS*	*Mean Square*	*F Value*	*Pr > F*
Treatment	2	80.621579	40.310789	53.43	<.0001
Center	5	30.052794	6.010558	7.97	<.0001
*Treatment*Center*	10	40.867448	4.086744	5.42	<.0001

Moreover, post-hoc test adequate for unbalanced data – (based on Least Squares Means) indicates that both drug treatment groups significantly differ from placebo with terms of efficiency in relieving post-operative pain (*Pr* > |*t*| =<0.0001).

Treatment	*Pain1 LSMEAN*	*LSMEAN Number*
A	3.510432	1
B	4.612522	2
Placebo	5.347940	3

Least Squares Means for effect Treatment *Pr > |t| for Ho: LSMean(i)=LSMean(j)* *Dependent Variable: pain1*			
i/j	*1*	*2*	*3*
1		<.0001	<.0001
2	<.0001		<.0001
3	<.0001	<.0001	

CONCLUSION

Although Analysis of Variance may seem quite straightforward, it is very easy to come to the false conclusions when it is performed inappropriately. In order to properly conduct ANOVA, the analyst should: (1) understand how an unbalanced data set differs from a balanced one; (2) know what sums of squares can be computed in SAS® and how to choose the best one for the given data design; (3) check for the existence of the outliers; (4) always verify model assumptions and, if they are not fulfilled, apply an adequate transformation to the response variable or use a non-parametric test or fit a different model, one that requires different distributional assumptions.

REFERENCES

[1] Doncaster C.P., Davey A. (2007), Analysis of Variance and Covariance. How to choose and construct models for the life sciences, Cambridge University Press, Cambridge.
[2] Fernandez G.C.J. (1992), Residual Analysis and Data Transformations: Important Tools in Statistical Analysis, HortScience, Vol 27(4).
[3] Gamst G., Meyers L.S., Guarino A.J. (2008), Analysis of Variance Design, Cambridge University Press, Cambridge.
[4] Gelman, A. (2005), Analysis of variance: why it is more important than ever (with discussion), Annals of Statistics 33.
[5] Hsu J.C. (1996), Multiple Comparisons Theory and Methods, Chapman & Hall, London.
[6] Iacobucci D. (1995), Analysis of Variance for Unbalanced Data, Northwestern University, American Marketing Association.
[7] Sawyer S.F. (2009), Analysis of Variance: The Fundamental Concepts, The Journal of Manual & Manipulative Therapy, Volume 17, Number 2.
[8] Walker G.A. (2002), Common statistical methods for clinical research with SAS examples, Second Edition, Cary, NC: SAS Institute Inc.
[9] SAS Online Documentation: http://support.sas.com/documentation/onlinedoc/sas9doc.html.

CONTACT INFORMATION

Joanna Romaniuk
Quanticate Polska Sp. z o.o.
Hankiewicza 2
02-103 Warsaw
Poland
Tel: +48(0) 22 576 21 40
Fax: +48(0) 22 576 21 59
E-mail: joanna.romaniuk@quanticate.com