ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY -
A GENERALIZED SAS MACRO

Shiferaw G. Mariam and Daniel W. Griffin
Research Laboratories
Ortho Pharmaceutical Corporation
Raritan, New Jersey 08869-0602

INTRODUCTION

The analysis of two-factor studies with n=1 depends on the assumption that the two factors do not interact. A plot of the residuals versus the predicted values sometimes reveals a relationship suggesting nonadditivity between the study factors. A SAS macro for a general approach to the test devised by Tukey is given. It may be used with any analysis of variance from any design (two-factor ANOVA, randomized block, latin square) or from any regression analysis. If the Tukey test indicates the presence of interaction effects, the macro performs a series of simple transformations such as a square root or logarithmic transformation to see if the interactions can be removed or made unimportant. When a suitable transformation of the response is selected for data covering a wide range of values, a considerable gain in precision is possible.

METHODOLOGY

When there is one observation in each cell of a factorial experiment, there can be no within-cell variation and hence no direct estimate of the experimental error. Assume the following two models for an observed data set:

\[ Y_{ij} = \mu + \alpha_i + \beta_j + \epsilon_{ij} \]  

(1)

and

\[ Y_{ij} = \mu + \alpha_i + \beta_j + (ab)_{ij} + \epsilon_{ij} \]  

(2)

where \( i = 1, \ldots, a \) and \( j = 1, \ldots, b \).

In (1), no interaction effect is assumed; thus all sources of variation other than the main effects are considered to be part of the experimental error. This model is considered the additive model. In (2), an interaction term is postulated. This model is considered the nonadditive model. When each cell has only one observation, one cannot include the interaction term \((ab)_{ij}\) in (2), because the error sum of squares (SSE) would equal zero.

Tukey devised a test which may be used for examining whether or not model (1) or (2) is appropriate for a given observed data set. The test is based on a partitioning of the experimental error sum of squares into a term attributed to nonadditivity (interaction) and a remainder sum of squares. Tukey proposed the following model for a two-factor ANOVA when \( n=1 \):

\[ Y_{ij} = \mu + \alpha_i + \beta_j + R_{ij} + \epsilon_{ij} \]  

(3)

where \( R \) is a regression coefficient. Note the \( R_{ij} \) is a more restricted interaction effect when compared to \((ab)_{ij}\) in model (2).

Using model (3), a test for \( R=0 \) is equivalent to a test of the hypothesis that the product term in (3) does not contribute to the prediction of \( Y_{ij} \). In order to test for \( R=0 \), least squares procedures are used. The least squares estimator of \( R \), assuming the other parameters are known, turns out to be:

\[ R = \frac{\sum \sum (\bar{Y}_{i.} - \bar{Y}) (\bar{Y}_{.j} - \bar{Y}) Y_{ij}}{\sum \sum (\bar{Y}_{i.} - \bar{Y})^2 \sum (\bar{Y}_{.j} - \bar{Y})^2} \]

The sample counterpart of the nonadditivity (interaction) sum of squares

\[ \Sigma \Sigma R^2 \alpha_i^2 \beta_j^2 \]

is:

\[ \text{SSA} = \frac{\left[ \sum \sum (\bar{Y}_{i.} - \bar{Y})(\bar{Y}_{.j} - \bar{Y}) Y_{ij}\right]^2}{\sum \sum (\bar{Y}_{i.} - \bar{Y})^2 \sum (\bar{Y}_{.j} - \bar{Y})^2} \]

The analysis of variance table for the case \( n=1 \) is shown in Table 1 and displays the decomposition of the sum of squares terms. If interaction effects exist, SSE reflects not only the random error variation but also the nonadditive effects. Thus the pure error sum of
TABLE 1
Two-Way Analysis of Variance With One Observation Per Cell

<table>
<thead>
<tr>
<th>Source of Variation</th>
<th>SS</th>
<th>df</th>
<th>MS</th>
<th>F</th>
</tr>
</thead>
<tbody>
<tr>
<td>Factor A</td>
<td>(SSA = \sum (Y_i - \bar{Y})^2)</td>
<td>(a-1)</td>
<td>(MSA = \frac{SSA}{a-1})</td>
<td>(MEA)</td>
</tr>
<tr>
<td>Factor B</td>
<td>(SSB = \sum (Y_j - \bar{Y})^2)</td>
<td>(b-1)</td>
<td>(MSB = \frac{SSB}{b-1})</td>
<td>(MSE)</td>
</tr>
<tr>
<td>Error</td>
<td>(SSE = \sum (Y_{ij} - \bar{Y})^2)</td>
<td>((a-1)(b-1))</td>
<td>(MSE = \frac{SSE}{(a-1)(b-1)})</td>
<td>(MSE)</td>
</tr>
<tr>
<td>Nonadditivity</td>
<td>(SSAB = \frac{\left[ \sum (Y_i - \bar{Y})^2 \sum (Y_j - \bar{Y})^2 \right]}{2} )</td>
<td>1</td>
<td>(MSE = \frac{SSAB}{1})</td>
<td>(MSE)</td>
</tr>
<tr>
<td>Remainder</td>
<td>(SSPE = SSE - SSAB)</td>
<td>((a-1)(b-1)-1)</td>
<td>(MSE = \frac{SSPE}{(a-1)(b-1)-1})</td>
<td>(MSE)</td>
</tr>
<tr>
<td>Total</td>
<td>(SSD = \sum (Y_{ij} - \bar{Y})^2)</td>
<td>(ab-1)</td>
<td>(MSE)</td>
<td></td>
</tr>
</tbody>
</table>

Squares (SSPE) can be obtained as follows:

\[
SSPE = SSE - SSAB
\]

It can be shown\(^3\) that if \(R=0\), SSPE and SSAB are independently distributed, and the test statistic:

\[
F^* = \frac{SSAB}{SSPE} \cdot \frac{1}{(a-1)(b-1)-1}
\]

is distributed as \(F(1, (a-1)(b-1)-1)\). For testing

- \(H_0: R=0\) (no interactions)
- \(H_1: R\neq 0\) (interactions \(a_i b_j\))

with Type I error at \(\alpha\), one concludes:

\[
H_0 \text{ if } F^* \leq F(1-\alpha; 1, (a-1)(b-1)-1)
\]

\[
H_1 \text{ if } F^* > F(1-\alpha; 1, (a-1)(b-1)-1)
\]

A significant result may indicate the need for transformation of the data. One way to choose a suitable transformation is to make the analyses for various transformations and then select a suitable one that shows no evidence of the existence of transformable interaction.\(^4\) Once an appropriate transformation is selected and the data analyzed on the new scale, all inferences regarding study factor effects must be made with respect to the new scale. In most behavioral research situations, inferences based on \(\log Y_i\)'s or \(\sqrt{Y_i}\)'s, for example, are just as meaningful as inferences based on untransformed responses\(^5\).

Note that it is not possible to find transformations which will eliminate nonadditivity in all cases. In some cases there is an intrinsic interaction between the factors which cannot be considered a function of the choice of the scale of measurement.

---

**MACRO DESCRIPTION**

The theory underlying Tukey's test for a two-factor experiment has been extended to higher-order factorial experiments. There are special formulas available to perform this test for particular designs. A general approach\(^5\) is outlined below which may be used with two-factor ANOVA, randomized block, or latin square designs. For each model selected:

1. Obtain the fitted values, \(\hat{Y}\), for each cell.
2. Calculate the residuals, \(Y - \hat{Y}\), for each cell.
3. Treat the values \(Z = Y^2\) as data and fit the model to the \(Z\) as in Step 2, and then calculate the residuals \(Z - \hat{Z}\) for each cell.
4. Compute \(S_P\), the sum of products of \((Y-Y)\) and \((Z-Z)\).
5. Compute \(SSZ\), the sum of squares of the residuals \((Z-Z)\).
6. Calculate the lack of fit sum of squares for nonadditivity as:

\[
SSAB = \frac{(SP)^2}{SSZ}
\]

7. Compute the pure error sum of squares as:

\[
SSPE = SSE - SSAB
\]

8. Compute the test statistic as:

\[
F^* = \frac{SSAB}{SSPE} \cdot \frac{1}{(df\ of\ SSE)-1}
\]

676
9. Compare $F^*$ with $F(1 - \alpha; 1, (\text{df of SSE}) - 1)$.

The TUKEY1DF macro renames the untransformed data entered via cards as macro variables for use in computing residuals and predicted values of the response variable. Using PROC PRINTTO, the GLM procedure output is routed to a dummy file for later recall. Three residual plots are generated as follows:

1) $Y - \hat{Y}$ versus $Y$
2) $Z - \hat{Z}$ versus $Z$, where $Z = Y^2$ and $Z - \hat{Z}$ versus $Y - \hat{Y}$

in the hope that these relationships will suggest transformable nonadditivity.

The error sum of squares is partitioned into a component due to pure error and one due to nonadditivity. Intermediate calculations are performed, the results of which are routed to a second dummy file. The two dummy files are recalled and concatenated, forming an enhanced GLM procedure output.

Through a conditional IF statement, the statistical significance of the nonadditivity term is evaluated. If there is no evidence of transformable interaction, the program execution terminates. If, however, the result indicates the need for data transformation, the analysis proceeds and performs a series of transformations from which the data analyst can choose. The data analyst can change the type and number of transformations performed thereby controlling the appropriateness of certain transformations.

An example of TUKEY1DF output for a two-factor study design is displayed in the appendix in addition to the SAS code.

ACKNOWLEDGMENTS

The authors wish to thank Raymond D. Buck for his helpful comments and suggestions with respect to the development of the transformation loop.

REFERENCES


AUTHORS' ADDRESS

Ortho Pharmaceutical Corporation
Route 202
P.O. Box 300
Raritan, New Jersey 08869-0602
APPENDIX

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY -
A GENERALIZED SAS RACCO

NO TRANSFORMATION OF THE RESPONSE

<table>
<thead>
<tr>
<th>OBS</th>
<th>BLOCK</th>
<th>TREATMT</th>
<th>RESPONSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>2</td>
<td>A</td>
<td>80</td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>B</td>
<td>85</td>
</tr>
<tr>
<td>3</td>
<td>4</td>
<td>C</td>
<td>88</td>
</tr>
<tr>
<td>4</td>
<td>5</td>
<td>D</td>
<td>90</td>
</tr>
<tr>
<td>5</td>
<td>6</td>
<td>E</td>
<td>95</td>
</tr>
<tr>
<td>6</td>
<td>7</td>
<td>F</td>
<td>100</td>
</tr>
<tr>
<td>7</td>
<td>8</td>
<td>G</td>
<td>105</td>
</tr>
<tr>
<td>8</td>
<td>9</td>
<td>H</td>
<td>110</td>
</tr>
<tr>
<td>9</td>
<td>10</td>
<td>I</td>
<td>115</td>
</tr>
</tbody>
</table>

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY -
A GENERALIZED SAS RACCO

NO TRANSFORMATION OF THE RESPONSE

PLOT OF R VS Y  
LEGEND: A = 1 OBS, B = 2 OBS, ETC.

RESPONSE PREDICTED VALUES

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY -
A GENERALIZED SAS RACCO

NO TRANSFORMATION OF THE RESPONSE

PLOT OF R VS P  
LEGEND: A = 1 OBS, B = 2 OBS, ETC.

RESPONSE**2 PREDICTED VALUES

678
ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIONALITY
A GENERALIZED BAYESIAN
NO TRANSFORMATION OF THE RESPONSE
PLOT OF R ONE Y  LEGEND: A = 1 OBS, B = 2 OBS, ETC.

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIONALITY
A GENERALIZED BAYESIAN
NO TRANSFORMATION OF THE RESPONSE
GENERAL LINEAR MODELS PROCEDURE
CLASS LEVEL INFORMATION
CLASS LEVELS VALUES
BLOCK 4 1 2 3 4
TREATMENT 5 1 2 3 4 5
NUMBER OF OBSERVATIONS IN DATA SET = 20

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIONALITY
A GENERALIZED BAYESIAN
NO TRANSFORMATION OF THE RESPONSE
GENERAL LINEAR MODELS PROCEDURE
DEPENDENT VARIABLE: RESPONSE
SOURCE DF SUM OF SQUARES MEAN SQUARE F VALUE PR > F
MODEL 7 1104.00000000 157.71428571 30.14 0.0001
ERROR 12 42.80000000 3.53333333
CORRECTED TOTAL 19 1146.80000000
R-SQUARE .946178 ROOT MSE 3.240328764799 79.60000000
SOURCE DF TYPE I SS F VALUE PR > F
TREATMENT 4 331.50000000 80.91 0.0001
SOURCE DF TYPE III SS F VALUE PR > F
TREATMENT 4 331.50000000 80.91 0.0001
SOURCE DF SUM OF SQUARES F VALUE PR > F
NONADDITIVITY 11 7.72 0.0180

679
ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY
A GENERALIZED SAS MACRO
THE COMMON LOGARITHM OF THE RESPONSE

<table>
<thead>
<tr>
<th>OBS</th>
<th>BLOCK</th>
<th>TREATY</th>
<th>LOG RESP</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>1</td>
<td>1.7443</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>2</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>3</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>1</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>2</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td>3</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>1</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>2</td>
<td>1.7483</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td>3</td>
<td>1.7483</td>
<td></td>
</tr>
</tbody>
</table>

LOG RESP PREDICTED VALUES

A GENERALIZED SAS MACRO
THE COMMON LOGARITHM OF THE RESPONSE
PLOT OF X Y*Y
LEGEND: A = 1 OBS. B = 2 OBS. ETC.

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY
A GENERALIZED SAS MACRO
THE COMMON LOGARITHM OF THE RESPONSE
PLOT OF X Y*Y
LEGEND: A = 1 OBS. B = 2 OBS. ETC.

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY
A GENERALIZED SAS MACRO
THE COMMON LOGARITHM OF THE RESPONSE
PLOT OF X Y*Y
LEGEND: A = 1 OBS. B = 2 OBS. ETC.

LOG RESP PREDICTED VALUES

680
ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY -
A GENERALIZED SAS MACRO
THE COMMON LOGARITHM OF THE RESPONSE
PLOT OF R vs Y. LEGEND: A = 1 OBS; B = 2 OBS, ETC.

ONE DEGREE OF FREEDOM TEST FOR TRANSFORMABLE NONADDITIVITY -
A GENERALIZED SAS MACRO
THE COMMON LOGARITHM OF THE RESPONSE
GENERAL LINEAR MODELS PROCEDURE
CLASS LEVEL INFORMATION
CLASS LEVELS VALUES
BLOCK 4 1 3 4
TREATMENT 5 1 2 3 4 5
NUMBER OF OBSERVATIONS IN DATA SET = 20

GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: LOG RESP
SOURCE DF SUM OF SQUARES MEAN SQUARE F VALUE PR > F
MODEL 7 0.04064314 0.00606314 40.30 0.0001
ERROR 15 0.00012994
CORRECTED TOTAL 19 0.00012994
R-SQUARE 0.959196 C.V. 0.000172980 0.00014415
SOURCE DF TYPE I SS F VALUE PR > F
BLOCK 2 0.01791297 34.05 0.0001
TREATMENT 4 0.01282517 19.59 0.0001
SOURCE DF TYPE III SS F VALUE PR > F
BLOCK 2 0.01282517 19.59 0.0001
TREATMENT 4 0.01282517 19.59 0.0001
SOURCE DF SUM OF SQUARES F VALUE PR > F
NONADDITIVITY 15 0.00012994 0.01 0.0001
DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

4. User defines the RESPONSE, transformation, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.

PROC ANOVA: 

DATA MONTAGE: 

1. User modifies 'RENAME' statement to accommodate the 'X#' variables defined in DATA
   NO TRANS.

2. User defines a RESOURCE of INGREDIENT. "X" DATES.

3. User defines the RESPONSE, TRANSFORMATION, and units.
   All INGREDIENTS.