

Testing the additivity assumption in repeated measures ANOVA

Robert J. Zambarano

University of Texas Computation Center

In social and biological sciences, a useful and frequently used statistical technique is the two-way mixed effects ANOVA design without replication. This is known as a split-plot design, or more generally as a repeated measures ANOVA. There are some unique assumptions associated with using this approach. Sphericity, for example, is an assumption that must be met for mixed effects designs; this assumption is documented elsewhere (e.g., Huyhn & Mandeville, 1979). A less well recognized assumption is additivity. For repeated measures ANOVA, the lack of attention to this condition may be due to the fact that the presence or absence of additivity in the data has no effect on the type I error rate. However, non-additivity has a drastic effect on the type II error rate; as the violation of this assumption becomes more severe, the power of this analysis decreases. Furthermore, any interpretation and generalization of significant effects becomes questionable when additivity does not hold.

Consider the following two-way factorial ANOVA design: factor A has a levels, factor B has b levels, and their factorial combination A by B produces $a*b$ cells. If only a single observation is made for each cell, this is known as a two-way design without replication. An appropriate model for this design is

$$Y_{ij} = C + A_i + B_j + INT_{ij} . \quad (1.1)$$

The Y_{ij} is the observed value of the dependent variable where i represents a particular corresponding level of A ranging from 1 to a , and j represents a particular corresponding level of B ranging from 1 to b . C represents the model constant, or intercept. A_i , B_j , and INT_{ij} represent the model parameters. A_i is the parameter corresponding to observations made at the i th level of factor A. B_j is the parameter corresponding to observations made at the j th level of factor B. INT_{ij} is the parameter corresponding to observations made at the i th level of factor A within the j th level of factor B. In order to obtain a unique least squares solution for such a model, we will impose the condition that $A_a = B_b = INT_{aj} = INT_{ib} = 0$. For the remaining parameters, the hypotheses that the A_i s or the B_j s do not equal zero will be referred to as main effects, while the hypothesis that the INT_{ij} s do not equal zero will be referred to as an interaction effect. No error term is required for this model because the number of Y_{ij} s is equal to the sum of the

number of parameters in the right side of the model; the model is saturated. In this situation, mean squares can be calculated for A, B, and INT, but calculation of a mean square for error is not possible. In the two-way mixed effects design such as the repeated measures ANOVA, the appropriate F ratio denominator for the test of the main effects of interest is the INT mean square. In the two-way fixed effects design however, only if the experimenter assumes that there is no interaction effect can the mean square for the INT term then be used as the denominator in F ratios testing the A and B effects. This assumption is known as additivity. When assuming additivity, the model becomes

$$y_{ij} = C + A_i + B_j + \text{Error}_{ij}. \quad (1.2)$$

Again, to obtain a unique least squares solution, $A_a = B_b = 0$.

The assumption of additivity is generally described in the context of the two-way fixed effects design without replication (e.g., Tukey, 1949; Mandel, 1961; Mandel, 1971; Johnson & Graybill, 1972). Clearly, the assumption is crucial in this case since the validity of the F ratios testing the main effects depend upon an accurate assessment of additivity or an accurate partitioning of the non-additive components. If the INT mean square or some partition thereof contains any variance due to an interaction effect, then the F ratio using this term as a denominator will not have the appropriate expected value under the null hypothesis. Furthermore, if a non-additive component of the INT mean square can be identified, it represents a significant interaction effect and must be interpreted. Of course, the presence of an interaction supersedes the interpretation of main effects for factor A and/or factor B.

Use of the fixed effects design without replication is rare in behavioral or biological sciences. However, use of the mixed effects design without replication in the form of the repeated measures ANOVA is common. In one basic form familiar to social scientists, the random factor is Subjects and the fixed factor is Measurement Occasions. Measurement Occasions typically represent fixed levels of some experimental treatment or condition of interest. Subjects are selected at random. Each subject is measured once at each measurement occasion so the design is completely crossed, but there is no replication. (Replication could be achieved by measuring subjects more than once at each measurement occasion.) Investigators who use this methodology are generally only interested in testing differences between measurement occasions. Actually though, anyone concerned about the power of this methodology has no choice but to be interested in the presence or absence of a Subject by Measurement Occasion interaction. The presence of an interaction means the data are non-additive, and the presence of non-additivity has a great effect on the type II error rate.

Furthermore, if an interaction is present, its interpretation will subsume any interpretation of the main effect for Measurement Occasion.

To understand why power is lost when non-additivity is present, recall that in a two-way mixed effects design, the fixed factor is confounded with the interaction of the fixed and random factors. Thus the components of variance for each mean square are as follows:

$$MS_{\text{fixed}} = \text{Effect of fixed factor} + \text{Effect of interaction} + \text{Error}$$

$$MS_{\text{interaction}} = \text{Effect of interaction} + \text{Error}$$

The F ratio testing the fixed factor would thus be

$$F = \frac{\text{Effect of fixed factor} + \text{Effect of interaction} + \text{Error}}{\text{Effect of interaction} + \text{Error}}$$

which returns an expected F value of approximately one if there is no effect of treatment in the MS_{fixed} . Since variance due to the interaction effect is contained in both the F numerator and denominator, the type I error rate is not affected by additivity or the lack. Of course, as the fixed factor effect become larger, so does the expect value of F. However, notice also that as the interaction effect becomes larger it increases the size of both numerator and denominator. The presence of any interaction here has the same effect as an increase in experimental error -- power is lost.

Perhaps even more important than the reduction of power is the issue of interpretation. Consider the following example. A researcher is interested in testing the effectiveness of a new therapy for treating the occurrence of hallucinatory episodes in schizophrenics. Pre-treatment counts of the number of episodes during two days of constant observation are obtained from a group of five subjects. Then the therapy is administered and subjects' conditions are re-evaluated at one week and one month from the start of treatment. Here is the data: (see table 1)

Subject	A	B	C	D	E	Mean
Pre-treatment	5	9	11	14	18	11.4
One week	11	9	8	7	5	8.0
One month	10	8	7	5	3	6.6

Look at the column of means and notice that between pre-treatment and the one month

follow-up, subjects displayed an average of about five less hallucinatory episodes. If a significant F is obtained from a repeated measures ANOVA, should the researcher conclude that the treatment was effective? Perhaps, but look again at the data. The treatment seems to be most effective for subjects who have the most hallucinatory episodes pre-treatment. The subject who initially had the fewest episodes experienced an increase across treatment. This is a Subject by Treatment interaction. Any interpretation of the main effect for treatment without reference to the interaction would in this case result in a serious inaccuracy. When describing the effects of the treatment, an accurate description of this data must take individual differences into account. The differential response of the subjects to the treatment becomes the issue of interest: to better understand the treatment, we must be able to identify those individuals who were benefited, and those who were not.

Clearly, the presence of non-additivity in repeated measures data presents a problem. However, if non-additive components can be identified, then the interaction implied by these components can be described. By identifying and describing such an interaction, the researcher has acquired a vastly improved understanding of the data at hand.

F ratio tests for additivity

The definition of the parameters for model 1.1 imply a design matrix composed of a set of "dummy code" binary contrasts. Of course, given any equivalent full-rank design matrix, the variance estimates for each effect will remain constant. However, individual parameter estimates will vary depending on the specific contrast each degree of freedom is based upon. For example, model 1.1 could be equivalently parameterized as follows.

$$y_{ij} = C + \sum_{k=1}^a A_k + \sum_{l=1}^b B_l + \sum_{k=1}^a \sum_{l=1}^b INT_{kl} . \quad (1.3)$$

All definitions for model 1.1 hold, except the parameters are redefined. A_k is one of a group of a parameters for orthogonal polynomial contrasts. Orthogonal polynomials are created from linear contrast weights where the weight for the i th group is the mean of that group. B_l is one of a group of b parameters for orthogonal polynomial contrasts. Orthogonal polynomial are created from linear contrast weights where the weight for the j th group is the mean of that group. INT_{kl} is one of a group of $a*b$ parameters for the interaction of A by B. Given the set of design matrix vectors for A and for B, the design matrix vector for INT_{kl} will correspond to the direct products

of the design vectors $A_k * B_l$. As before, in order to obtain a unique least squares solution, the restriction is imposed that the parameters $A_a = B_b = INT_{ak} = INT_{lb} = 0$. INT_{11} , which is the direct product of the two linear-contrast design vectors, gives the contrast for Tukey's test for additivity (1949): the mean square for this single degree of freedom is divided by the mean square for the remaining INT_{kl} to obtain an F test for non-additivity. In practice, equivalent results can be obtained by producing a new variable. This variable will have values for each observation being the product of the mean for the corresponding level of A and the mean for the corresponding level of B. This variable can now be used as a covariate in PROC GLM, and the test of its significance is Tukey's test for additivity. The following SAS code could be used. Putting the data into the context of the current discussion, row_num could be identifying unique subjects, and col_num could be identifying unique measurement occasions.

```
*This program demonstrates Tukey's test for non-additivity;
*Data is from Tukey, J.W. (1949). One degree of freedom for non-additivity,
Biometrics, 5, pp. 232-242;
```

```
Data tukey; input cell_val col_num row_num;
cards;
14 1 1
 2 1 2
 2 1 3
 2 2 1
 0 2 2
 1 2 3
 1 3 1
 2 3 2
 5 3 3
 2 4 1
 2 4 2
 0 4 3
;
```

```
*The first PROC SUMMARY calculates the column means, and outputs them to a
data set work.col;
```

```
proc summary; by col_num; var cell_val; output out=col mean=cmean;
proc sort data=tukey; by row_num;
```

```
*The second PROC SUMMARY calculates the row means, and outputs them to a data
set work.row;
```

```
proc summary; by row_num; var cell_val; output out=row mean=rmean;
```

```
*The next data step crosses the row means and column means so that each row
mean is paired with each column mean;
```

```
data means;
```

```

do j=1 to nobsl;
  set row nobs=nobs1 point=j;
  do k=1 to nobs2;
    set col nobs=nobs2 point=k;
    output;
  end;
end;
stop;

*The next data step merges work.means with work.tukey, and the covariate
representing Tukey's additivity test is produced;

data new; merge tukey means; by row_num col_num;
tuk_vec=cmean*rmean;

*PROC GLM is used to perform Tukey's additivity test;

proc glm;
class row_num col_num;
model cell_val=row_num col_num tuk_vec;
run;
endsas;

```

The output would be as follows:

The SAS System 11:51 Sunday, September 13, 1992 1

General Linear Models Procedure
Class Level Information

Class	Levels	Values
ROW_NUM	3	1 2 3
COL_NUM	4	1 2 3 4

Number of observations in data set = 12

The SAS System 11:51 Sunday, September 13, 1992 2

General Linear Models Procedure

Dependent Variable: CELL_VAL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	6	123.36593625	20.56098938	3.13	0.1158
Error	5	32.88406375	6.57681275		
Corrected Total	11	156.25000000			
R-Square		C.V.	Root MSE	CELL_VAL Mean	
0.789542		93.25563	2.56452973	2.75000000	

Source	DF	Type I SS	Mean Square	F Value	Pr > F
--------	----	-----------	-------------	---------	--------

ROW_NUM	2	24.50000000	12.25000000	1.86	0.2486
COL_NUM	3	46.91666667	15.63888889	2.38	0.1862
TUK_VEC	1	51.94926958	51.94926958	7.90	0.0375
Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW_NUM	2	24.50000000	12.25000000	1.86	0.2486
COL_NUM	3	46.91666667	15.63888889	2.38	0.1862
TUK_VEC	1	51.94926958	51.94926958	7.90	0.0375

This output corresponds to the results published by Tukey (1949). From the variance accounted for by the interaction effect, Tukey's test parses the variance accounted for by a single parameter. The significance of the variance associated with this parameter is tested against the variance accounted for by the remaining interaction parameters. It cannot be over emphasized, however, that Tukey's test examines only one of many possible hypothesis concerning non-additivity. The condition described and tested by this method is neither a necessary nor a sufficient condition for non-additivity. Tukey's test looks at only one possible non-additive data configuration, but others are quite possible.

The general method used of identifying one or more INT parameters as representing a hypothesized pattern of non-additivity is not limited to Tukey's single degree of freedom. Other's have proposed tests of non-additivity along similar lines (Mandel, 1961). Instead of relying on "cook-book" hypotheses, however, it would be more useful for investigators to consider their own data, determining the patterns of non-additivity that may be likely to occur. From this knowledge, contrast parameters can be designed for each main effect which will produce a set of interaction contrasts, one or more of which will represent the hypothesized pattern of non-additivity. These interaction contrasts can be used to create covariates that can be tested using PROC GLM as above.

The multiplicative interaction model, and a characteristic root test for non-additivity

As an alternative to developing an hypothesis concerning non-additivity, the investigator can use the following method to determine the exact non-additive components of his or her sample. To partition the INT component from model 1.1, the following equivalency is given:

$$INT_{ij} = \sum_{k=1}^e \lambda_k \alpha_{ki} \gamma_{kj}.$$

Then we can rewrite model 1.1 as

$$y_{ij} = C + A_i + B_j + \lambda_1 \alpha_{1i} \gamma_{1j} + \lambda_2 \alpha_{2i} \gamma_{2j} + \dots + \lambda_e \alpha_{ei} \gamma_{ej} \quad (1.3)$$

As with model 1.1, i corresponds to each of the a levels of A , and j corresponds to each of the b levels of B . The value for e will be the rank of the matrix $Z'Z$, where Z is the a by b matrix of INT parameters from model 1.1; row addresses correspond to each value of i and column addresses correspond to each value of j (Mandel, 1971). In this case, the values of λ_k^2 will be the e characteristic roots of $Z'Z$ or Z^*Z' . The values of α_{ki} will be the e characteristic vectors of Z^*Z' . The values of γ_{kj} will be the e characteristic vectors of $Z'Z$. Now we can test for non-additivity by $H_0: \lambda_1=0$ vs. $H_1: \lambda_1 \neq 0$. This test is given as

$$U = \lambda_1^2 / (\lambda_1^2 + \lambda_2^2 + \dots + \lambda_k^2).$$

A table of critical values of U is can be found in Miliken & Johnson, (1989).

If a significant non-additive parameter is detected via this method, the pattern of additive can be discerned by examining the matrix product of $\alpha_{li} * \gamma_{lj}$. These values represent the non-additive component of the expected value for each observation.

Of course, the variance contained in the row by column matrix of INT parameters will be equivalent to the variance contained in the matrix of row by column residuals from model 1.2. These residuals are easily obtained from PROC GLM. These residuals are then used as input to PROC IML, where eigen analysis is performed. The code to run such analysis is as follows, again using Tukey's data.

```
proc glm data=tukey;
class row_num col_num;
model cell_val=row_num col_num;
output out=tukmat r=resid;
proc sort data=tukmat; by col_num row_num;
proc iml;
use tukmat;
read all var { resid col_num row_num };
close tukmat;
z=j(3,4,0);
do i=1 to 3;
  do j=1 to 4;
    z[i,j]=resid[(i+(3*(j-1))),1];
  end;
end;
call eigen(L2,ec,(z`*z));
call eigen(L2,er,(z*z`));
u=L2[1,]/(L2[1,]+L2[2,]+L2[3,]);
L1=L2[1,]##.5;
print u;
nonadd=L1#((er[,1]#ec[1,1])/(er[,1]#ec[2,1])/(er[,1]#ec[3,1])/(er[,1]#ec[4,1]));
```

```
print row_num col_num resid nonadd Ll;
;
```

The output would be as follows:

The SAS System 08:24 Sunday, September 27, 1992 1

General Linear Models Procedure
Class Level Information

Class	Levels	Values
ROW_NUM	3	1 2 3
COL_NUM	4	1 2 3 4

Number of observations in data set = 12

The SAS System 08:24 Sunday, September 27, 1992 2

General Linear Models Procedure

Dependent Variable: CELL_VAL

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	5	71.41666667	14.28333333	1.01	0.4849
Error	6	84.83333333	14.13888889		
Corrected Total	11	156.25000000			

R-Square	C.V.	Root MSE	CELL_VAL Mean
0.457067	136.7335	3.76017139	2.75000000

Source	DF	Type I SS	Mean Square	F Value	Pr > F
ROW_NUM	2	24.50000000	12.25000000	0.87	0.4671
COL_NUM	3	46.91666667	15.63888889	1.11	0.4171

Source	DF	Type III SS	Mean Square	F Value	Pr > F
ROW_NUM	2	24.50000000	12.25000000	0.87	0.4671
COL_NUM	3	46.91666667	15.63888889	1.11	0.4171

U

0.9339679

ROW_NUM	COL_NUM	RESID	NONADD	L1
1	1	6	5.9584904	8.9012141
2	1	-2.75	-2.405775	
3	1	-3.25	-3.552716	
1	2	-1	-1.01949	
2	2	0.25	0.411625	
3	2	0.75	0.6078651	
1	3	-3.666667	-3.780384	
2	3	0.5833333	1.5263516	
3	3	3.0833333	2.2540321	
1	4	-1.333333	-1.158617	
2	4	1.9166667	0.4677981	
3	4	-0.583333	0.6908185	

Using an alpha level of $p < 0.05$, the null hypothesis is rejected. It is interesting to compare the values of the residual with the values of the non-additive component..

The main problem of this approach is that unique values of U exist for each unique experimental design with an unreplicated factor. Complete tables do not exist, and the investigator wishing to use this method is generally limited to two-factor unreplicated experiments. Less problematic, but perhaps more important, this is an exploratory method. Unreplicated results must be generalized with caution. Also, interpretation of the non-additive component is difficult. However, as a means of identifying certain subjects who do not behave similarly across measurement occasion, this is quite useful.

This is a brief demonstration of these approaches, additional analyses are possible. It is hoped that the reader is now at least convinced that non-additivity is something to consider. Analysis of hypotheses concerning non-additivity in repeated measures ANOVA designs is within the ability of anyone with a basic understanding of the concepts and a good working knowledge of SAS PROC GLM and PROC IML.

Citations

- Huynh, H. & Mandeville, G.K. (1979). Validity conditions in repeated measures designs. Psychological Bulletin, 86, 964-973.
- Johnson, D.E. & Greybill, F.A. (1972). An analysis of a two-way model with interaction and no replication. Journal of the American Statistical Association, 67, 862-868.
- Mandel, J. (1971). A new analysis of variance model for non-additive data. Technometrics, 13, 1-18.
- Milliken, G.E. & Johnson, D.E. (1989). Analysis of Messy Data (Vol. 2). New York: Van Nostrand Reinhold Company.
- Tukey, J.W. (1949). One degree of freedom for non-additivity. Biometrics, 5, 232-242.