TESTS OF ADDITIVITY IN TWO-WAY ANOVA MODELS WITH SINGLE SUBCLASS NUMBERS

MARIE ŠIMEČKOVÁ[1,2], PETR ŠIMEČEK[1] and DIETER RASCH[2]

simeckova.marie@vuzv.cz, simecek.petr@vuzv.cz, dieter.rasch@boku.ac.at

[1] Institute of Animal Science, Biometric Unit, Prague

[2] University of Natural Resources and Applied Life Sciences, Institute of Applied Statistics, Vienna

SUMMARY

The poster presents recent results on tests of interaction in two-way ANOVA mixed models without replication, and determination of the minimum size of experiment in this case.

INTRODUCTION

In many applications of statistical methods it is assumed that the response variable is a sum of several factor variables and a random noise. In a real world this may not be an appropriate model. For example, some patients may react differently to the same drug treatment. Or the influence of fertilizer may be influenced by the type of a soil. There might exist an interaction between factors. A testing for such interaction will be referred here as **testing of additivity hypothesis**.

If there is more than one observation per cell then standard ANOVA techniques may be applied. Unfortunately, in many cases it is infeasible to get more than one observation taken under the same conditions. For instance, it is impossible to ask the same student the same question twice.

We restrict ourselves to a case of two factors, i.e. two–array model, when the response in i^{th} row and j^{th} column is modeled as

$$\boldsymbol{y}_{ij} = \mu + \alpha_i + \boldsymbol{b}_j + \boldsymbol{\gamma}_{ij} + \boldsymbol{e}_{ij},\ i = 1, \ldots, a,\ j = 1, \ldots, b, \quad (1)$$

where α_i are levels of the fixed factor, $\sum_i \alpha_i = 0$, $\boldsymbol{b}_j$ and $\boldsymbol{\gamma}_{ij}$ are random factor and random interaction, both normally distributed with zero mean and variance σ_b^2 and σ_γ^2, resp., and the $\boldsymbol{e}_{ij}$ are normally distributed independent random variables with zero mean and variance σ^2.

The problem is to find a test for the hypothesis that there is no interaction between the fixed and the random factor ($\sigma_\gamma^2 = 0$).

There are several tests of additivity for the two-way ANOVA model without replication with fixed factors. We consider five of them: Tukey test, Mandel test, Johnson – Graybill test, locally best invariant (LBI) test and Tussel test.

First, the actual type-I-risk of all these tests is verified for the mixed ANOVA model by simulation. Second, their power is studied and a modification of Tukey test is proposed. Third, we found an approximate relation of the power of the Johnson – Graybill test and the parameters of model (1).

ADDITIVITY TESTS

We very shortly recall the five tests. Let $\bar{y}_{..}$ denotes the over all mean, $\bar{y}_{i.}$ and $\bar{y}_{.j}$ the i^{th} row's and the j^{th} column's means, respectively. The matrix R will stand for a residual matrix with respect to the main effects $r_{ij} = y_{ij} - \bar{y}_{i.} - \bar{y}_{.j} + \bar{y}_{..}$. The decreasingly ordered list of eigenvalues of RR^T will be denoted by $\kappa_1 \geq \kappa_2 \geq \ldots$, and its scaled version $\omega_i = \kappa_i / \sum_k \kappa_k$, $i = 1, 2, \ldots$.

If the interaction is present we may expect that some of the ω_i coefficients will be substantially higher than others.

Tukey test: Introduced in Tukey (1949). Tukey test first estimates row and column effects and then tests for the interaction of type $\gamma_{ij} = k\bar{y}_{i.}\bar{y}_{.j}$ ($k = 0$ implies no interaction). Tukey's test statistic S_T equals

$$S_T = MS_{int}/MS_{error},$$

where

$$MS_{int} = \frac{\left(\sum_i \sum_j y_{ij}(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})\right)^2}{\sum_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 \sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}$$

and

$$MS_{error} = \frac{\sum_i \sum_j (y_{ij} - \bar{y}_{\cdot\cdot})^2 - a\sum_j(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2 - b\sum_i(\bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot})^2 - MS_{int}}{(a-1)(b-1)-1}.$$

S_T is F-distributed with 1 and $(a-1)(b-1)-1$ degrees of freedom under the hypothesis of additivity.

Mandel test: Introduced in Mandel (1961). Mandel test statistic S_M equals

$$S_M = \frac{\sum_i (z_i - 1)^2 \sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}{a-1} \Big/ \frac{\sum_i \sum_j \left((y_{ij} - \bar{y}_{i\cdot}) - z_i(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})\right)^2}{(a-1)(b-2)},$$

where

$$z_i := \frac{\sum_j y_{ij}(\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})}{\sum_j (\bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot})^2}.$$

S_M is F-distributed with $a-1$ and $(a-1)\cdot(b-1)$ degrees of freedom under the additivity hypothesis.

Johnson – Graybill test: Introduced in Johnson and Graybill (1972). Johnson – Graybill test statistic is just $S_J = \omega_1$. The additivity hypothesis is rejected if S_J is high.

Locally best invariant (LBI) test: See Boik (1993). LBI test statistic equals

$$S_L = \sum_{k=1}^{\min(a,b)-1} \omega_k^2.$$

The additivity hypothesis is rejected if S_L is high.

Tussel test: See Tussel (1990). Tussel test statistic equals

$$S_U = \prod_{k=1}^{\min(a,b)-1} \omega_k.$$

The additivity hypothesis is rejected if S_U is low.

The definitions of Johnson – Graybill's, LBI and Tussel's test statistics presented above slightly differ from their original versions (in a multiplicative constant). For a, b fixed, a simulation may be used to get the critical values.

TYPE-I-RISK OF ADDITIVITY TESTS

The actual type-I-risk of all the five tests of additivity were verified for the mixed ANOVA model by simulation. Only the most common nominal type-I-risk 5 % was assumed.

For verifying the type-I-risk, the number of levels of the fixed factor was assumed $a = 3, 4, \ldots, 10$, number of levels of the random factor $\boldsymbol{b}$ was chosen between 4 and 50 (by 2 between 4 and 20, by 5 between 20 and 50), the variance of the random factor $\sigma_b^2 = 2, 5, 10$ and the variance of the random error $\sigma^2 = 1$.

In one step of the simulation a set of data was generated based on the model without interaction. Then the test of no interaction was performed. The percentage of significant test after several steps is assumed to be the actual level of the test.

The 10 000 simulations were repeated 10 times and the standard error of the estimation of the mean actual level was computed based on these 10 repetitions. Then the one-sided one sample t-test of the hypothesis "the actual p-level is equal to or lower than 0.05" was performed (on the 5 % level without correction to the multiple testing). The results of these tests are summarized in Table 1.

Test	$\hat{\alpha} \leq 0.05$	$\hat{\alpha} > 0.05$
Tukey test	349 (96.94)	11 (3.06)
Mandel test	348 (96.67)	12 (3.33)
Johnson Graybill test	339 (94.17)	21 (5.83)
LBI test	336 (93.33)	24 (6.67)
Tusell test	337 (93.61)	23 (6.39)

Table 1: Number and percentage of the simulated cases actual test level is equal to or lower than and greater than nominal level 5%.

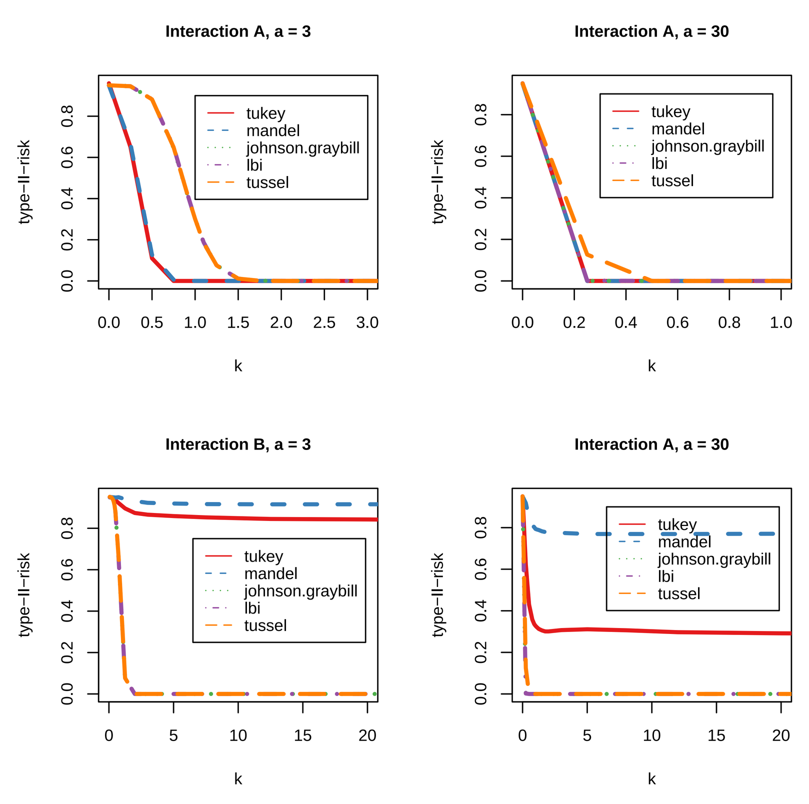
For the Tukey and Mandel tests in the vast majority (> 95 %) of cases the actual level is not significantly above the 0.05 level. In less than 4 % the type-I-risk is higher than the nominal 0.05. For the other tests the nominal level is higher than 0.05 in slightly more cases. However, this may also be false positives caused by multiple testing.

In the ANOVA models with both effects fixed there is an important assumption about summing of both effects to zero, i.e. $\sum_{i=1}^a a_i = \sum_{j=1}^b b_j = 0$. In the case of random model that is not valid. It is assumed that the expected value of random term $E(\boldsymbol{b}_j)$ equals zero, but in one particular case the sum is not zero (almost surely). It can cause inaccuracy of the tests. However, for high number of levels of the random factor $\boldsymbol{b}$, the sum converges to zero (law of large numbers) and this problem disappears.

For 5 % type-I-risk all these tests hold the level of type-I-risk and therefore the tests designed for fixed models can be used for the mixed models as well.

A MODIFICATION OF THE TUKEY TEST

As mentioned above the Tukey, Mandel, Johnson – Graybill, LBI and Tussel test hold the level of type-I-risk even when one factor is considered as random. In this part the power of these tests is compared by means of simulation. While Tukey test has relatively good power when the interaction is a product of the main effects, i.e., when $\boldsymbol{\gamma}_{ij} = k\alpha_i\boldsymbol{b}_j$ (interaction type A), its power for more general interaction (interaction type B) is very poor:

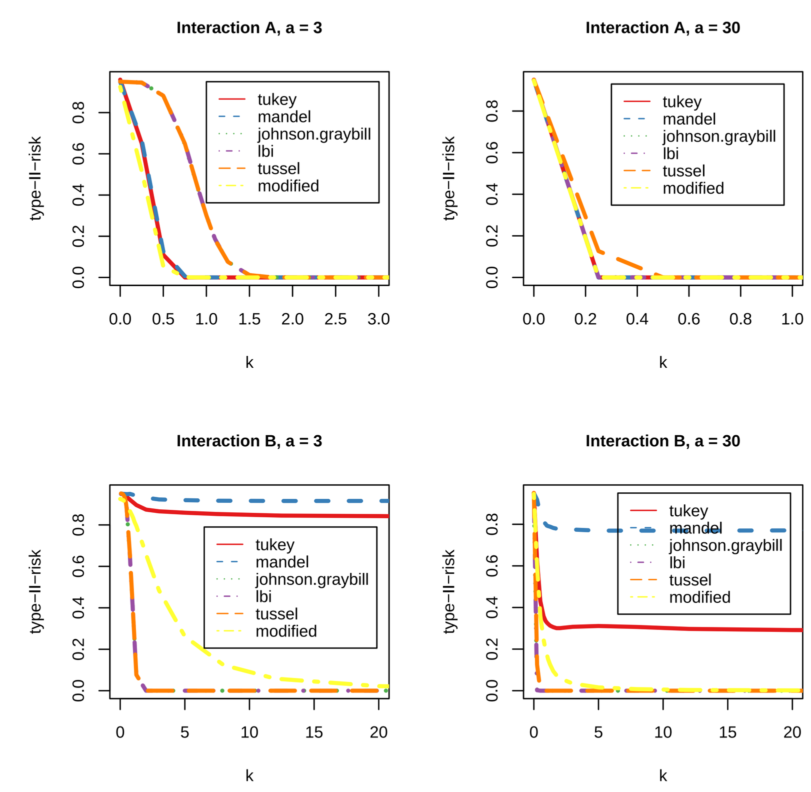


To increase the power of Tukey test we propose the following modification: In Tukey test a model $\boldsymbol{y}_{ij} = \mu + \alpha_i + \boldsymbol{b}_j + k\alpha_i\boldsymbol{b}_j + \boldsymbol{e}_{ij}$ is tested against a submodel $\boldsymbol{y}_{ij} = \mu + \alpha_i + \boldsymbol{b}_j + \boldsymbol{e}'_{ij}$. The estimators of row effects $\hat{\alpha}_i = \bar{y}_{i\cdot} - \bar{y}_{\cdot\cdot}$ and column effects $\hat{b}_j = \bar{y}_{\cdot j} - \bar{y}_{\cdot\cdot}$ are calculated in the same way in both models although the dependency of $\boldsymbol{y}_{ij}$ on these parameters is not linear for the full model.

The main idea behind the presented modification is that a full model

$$\boldsymbol{y}_{ij} = \mu + \alpha_i + \boldsymbol{b}_j + k\alpha_i\boldsymbol{b}_j + \boldsymbol{e}_{ij}$$

is fitted by a nonlinear regression and tested against a submodel $\boldsymbol{y}_{ij} = \mu + \phi_i + \boldsymbol{psi}_j + \boldsymbol{e}'_{ij}$ by a likelihood ratio test. The estimates of row and column effects therefore differ for each model. See the simulation results with modified test:



If the sample size is small bootstrapping without replacement or sampling from fitted model is used to control type-I-risk level.

THE POWER OF JOHNSON – GRAYBILL TEST

As was shown, the Tukey and Mandel tests are appropriate if the interaction is a product of the row and column effects (and a constant) $\boldsymbol{\gamma}_{ij} = k \cdot a_i \cdot \boldsymbol{b}_j$. The Johnson – Graybill, LBI and Tussel tests are a bit worse for this special case, but they are suitable in cases of more complex interactions. When planning experiments, it is hard to know the form of the interaction and therefore the three latter tests are more appropriate in the general case. In the following we will consider only the Johnson – Graybill test.

To state the formula for estimation of power, we assume an interaction term in the model (1) of the form

$$\boldsymbol{\gamma}_{ij} = k \cdot a_i \cdot \boldsymbol{c}_j, \quad (2)$$

where a_i are the row effects in (1), $\boldsymbol{c}_j$ are normally distributed random variables with zero expected value and variance σ_b^2, mutually independent to the random variable $\boldsymbol{b}_j$ and $\boldsymbol{e}_{ij}$. The k is a real constant.

The interaction in (2) is a random variable with zero mean. Its variance is equal to $var\ \boldsymbol{\gamma}_{ij} = k^2 \cdot a_i^2 \cdot \sigma_b^2$.

The power of a test gets higher if the distance of its alternative from the null hypothesis becomes greater. Based on simulations, it was found out that the power of the test of additivity depends on the a_i only through the sum of their squares $\sum_{i=1}^a a_i^2$. The power of Johnson – Graybill test for type-I-risk equal to 5 % can be approximately computed as

$$power = 1 - \frac{1}{a \cdot b \cdot k^4 \cdot \sigma_b^4 \sum_{i=1}^a a_i^2}. \quad (3)$$

The plan of an experiment means in our situation to set the number of levels of the random factor b (i.e. the number of blocks). From (3), for required power β and given a, k,σ_B^2 and $\sum a_i^2$ the number of levels should equals at least

$$b = \frac{1-\beta}{a \cdot k^4 \cdot \sigma_B^4 \sum_{i=1}^a a_i^2}.$$

Acknowledgment. The work was supported by the Minister of Agriculture of the Czech Republic MZE 0002701403.

References.

Alin, A. and Kurt, S. (2006). Testing non-additivity (interaction) in two-way ANOVA tables with no replication. *Stat. Meth. in Med. Res. 15*, 63–85.

Boik, R.J. (1993a). Testing additivity in two-way classifications with no replications: the locally best invariant test. *Journal of Applied Statistics 20*, 41–55.

Boik, R.J. (1993b). A comparison of three invariant tests of additivity in two-way classifications with no replications. *Computational Statistics and Data Analysis 15*, 411–424.

Johnson, D.E. and Graybill, 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. (1961). Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association 56*, 878–888.

Rasch, D. (1995). *Mathematische Statistik.* Johann Ambrosius Barth, Heidelberg dž" Leipzig und Wiley, Berlin.

Šimeček, P. and Šimečková (submitted). Modification of Tukey's Additivity Test. *Journal of Statistical Planning and Inference.*

Tukey, J.W. (1949). One degree of freedom for nonadditivity. *Biometrics 5*, 232–242.

Tusell, F. (1990). Testing for interaction in two-way ANOVA tables with no replication. *Computational Statistics and Data Analysis 10, 29–45.*