11-1-2003

JMASM8: Using SAS To Perform Two-Way Analysis Of Variance Under Variance Heterogeneity

Scott J. Richter
University of North Carolina at Greensboro, sjricht2@uncg.edu

Mark E. Payton
Oklahoma State University, mpayton@okstate.edu

Follow this and additional works at: http://digitalcommons.wayne.edu/jmasm

🌀 Part of the Applied Statistics Commons, Social and Behavioral Sciences Commons, and the Statistical Theory Commons

Recommended Citation

Richter, Scott J. and Payton, Mark E. (2003) "JMASM8: Using SAS To Perform Two-Way Analysis Of Variance Under Variance Heterogeneity," *Journal of Modern Applied Statistical Methods*: Vol. 2 : Iss. 2 , Article 25.
DOI: 10.22237/jmasm/1067646300
Available at: http://digitalcommons.wayne.edu/jmasm/vol2/iss2/25

JMASM Algorithms and Code
JMASM8: Using SAS To Perform Two-Way Analysis Of Variance Under Variance Heterogeneity

Scott J. Richter
Department of Mathematical Sciences
University of North Carolina at Greensboro

Mark E. Payton
Department of Statistics
Oklahoma State University

We present SAS code to implement the method proposed by Brunner et al. (1997) for performing two-way analysis of variance under variance heterogeneity.

Key words: ANOVA, heteroscedasticity, SAS

Introduction

Brunner et al. (1997) suggested a method to perform tests on main effects and interaction in two-way analysis of variance that allows variance heterogeneity. Their approach is to use a generalization of chi-square approximations dating back to Patnaik (1949) and Box (1954). Their statistic is identical to the classical ANOVA F-statistic, and thus their method can be regarded as a robust extension of the classical ANOVA to heteroscedastic designs. They recommend that their method should always be preferred (even in the homoscedastic case) to the classical ANOVA. Richter and Payton (2003) found that the performance of their statistic compares favorably to that of the usual ANOVA F-statistic for sample sizes of at least n = 7 per factor combination.

In this article, we present a SAS program (SAS Institute, Cary, N.C.) for implementing the Brunner et al. (1997) method.

Scott Richter is an Assistant Professor in the Mathematical Sciences Department at the University of North Carolina at Greensboro. His email address is sjricht2@uncg.edu. Mark Payton is a Professor in the Department of Statistics at Oklahoma State University. His email address is mpayton@okstate.edu.

Brunner Method

The method of Brunner et al. (1997) is a small sample adjustment to the well-known Wald statistic, which permits heterogeneous variance but is known to have inflated Type I error rates for small sample sizes. Consider a two-way layout with a levels of factor A and b levels of factor B. Assume a set of independent random variables

$$X_{ij} \sim N(\mu_i, \sigma_i^2), i = 1, ..., ab.$$

Let $\mathbf{\mu} = (\mu_1, \mu_2, ..., \mu_{ab})'$ denote the vector containing the $a \bullet b$ population means. Then the hypotheses of no main effects and interaction can be written as

$$H_0(A) : \mathbf{M}_A \mathbf{\mu} = 0$$
$$H_0(B) : \mathbf{M}_B \mathbf{\mu} = 0$$
$$H_0(AB) : \mathbf{M}_{AB} \mathbf{\mu} = 0,$$

where

$$\mathbf{M}_A = \mathbf{P}_a \otimes \frac{1}{b} \mathbf{J}_b$$

$$\mathbf{M}_B = \frac{1}{a} \mathbf{J}_a \otimes \mathbf{P}_b$$

$$\mathbf{M}_{AB} = \mathbf{P}_a \otimes \frac{1}{b} \mathbf{J}_b.$$

Here $\mathbf{P}_c = \mathbf{I}_c - \dfrac{1}{c}\mathbf{J}_c$, where $\mathbf{I}_c$ is a $c \times c$ identity matrix, $\mathbf{J}_c$ a $c \times c$ matrix of 1's, and the symbol $\otimes$ represents the Kronecker product of the matrices. The vector of observed cell means is denoted by $\overline{\mathbf{X}} = \left(\overline{X}_1, ..., \overline{X}_{ab} \right)'$ and the estimated covariance matrix is given by $\hat{\mathbf{S}}_N = N \bullet diag\left\{ \dfrac{S_1^2}{n_1}, ..., \dfrac{S_{ab}^2}{n_{ab}} \right\}$, where S_i^2 is the ith sample variance and $N = \sum\limits_{i=1}^{ab} n_i$.

For a complete cross-classification, the test statistic is

$$ FB = \frac{N \bullet \overline{\mathbf{X}}'\mathbf{M}\overline{\mathbf{X}}}{\dfrac{1}{(n-1)} tr\left(\hat{\mathbf{S}}_N\right)}, $$

which has an approximate F distribution with

$$ f_{num} = \frac{\dfrac{1}{(n-1)^2} \bullet \left[tr\left(\hat{\mathbf{S}}_N\right) \right]^2}{tr\left(\mathbf{M}\hat{\mathbf{S}}_N\mathbf{M}\hat{\mathbf{S}}_N \right)} $$

numerator and

$$ f_{den} = \frac{\left[tr\left(\hat{\mathbf{S}}_N\right) \right]^2}{tr\left(\hat{\mathbf{S}}_N^2 \boldsymbol{\Lambda}\right)} $$

denominator degrees of freedom, where

$$ \boldsymbol{\Lambda} = diag\left\{ \frac{1}{n_1 - 1}, ..., \frac{1}{n_{ab} - 1} \right\} $$

(Brunner, 1997).

References

Box, G. E. P. (1954). Some theorems on quadratic forms applied in the study of analysis of variance problems, I. Effect of inequality of variance in the one-way classification. *The Annals of Mathematical Statistics*, *25*, 290-302.

Brunner, E., Dette, H., & Munk, A. (1997). Box-type approximations in nonparametric factorial designs. *Journal of the American Statistical Association*, *92*, 1494-1502.

Patnaik, P. B. (1949). The noncentral χ^2 and F-distributions and their applications. *Biometrika*, *36*, 202-232.

Richter, S. J., & Payton, M. E. (2003). Performing two-way analysis of variance under variance heterogeneity. *Journal of Modern Applied Statistical Methods*, *2*(1), 152-160.

Appendix

The following program can be used to perform the FB test described above.

```
/* Program to compute unadjusted and Box-adjusted F-ratios
and p-values for a two-way layout. */

data rcht;
input a b RESP @@;
datalines;
<data>
;
proc sort;
by a b;
```

```
/* Run Proc Mixed to calculate variances */
proc mixed data=rcht; class a b;
model RESP=a|b;
repeated/type=un(1) group=a*b;

/* Create data sets for covariances, means and class levels for input
   to Proc IML */
ods listing exclude covparms; ods output covparms=tempcov;
ods listing exclude classlevels; ods output classlevels=levels;
ods listing exclude dimensions; ods output dimensions=sizes;

/* Suppress printing of Proc Mixed tables */
ods listing exclude fitstatistics;
ods listing exclude reml;
ods listing exclude ConvergenceStatus;
ods listing exclude IterHistory;
ods listing exclude lrt;
ods listing exclude modelinfo;
ods listing exclude tests3;

/* Use Proc GLM to calculate and output Type III F-statistics */
proc glm data = rcht noprint outstat=fstats;
class a b; model RESP=a|b/ss3; run;

/* Use Proc Means to calculate and output cell means and sample sizes
*/
proc means data=rcht noprint;
var RESP; by a b;
output mean=means n=n out=tempss;

/* Begin Proc IML to calculate adjusted df and p-values */
proc iml;

/* Create matrices from data sets created above */
use levels;
read point 1 var {levels} into nla;
read point 2 var {levels} into nlb; use sizes;
read point 1 var {value} into parms;
read point 8 var {value} into nobs; use tempss;
read all var {n} into ni; use tempcov;
read all var {estimate} into tsighat;
sighat=diag(tsighat);
do i=1 to parms;
  sighat[i,i]=sighat[i,i]/ni[i];
end;
shat=nobs*sighat; use tempss;
read all var {means} into Xbar;
results=j(3,4,0); use fstats;
read point 2 var {F} into FHA;
read point 3 var {F} into FHB;
read point 4 var {F} into FHAB;
```

```
read point 2 var {df} into dfa;
read point 3 var {df} into dfb;
read point 4 var {df} into dfAB;
read point 1 var {df} into dfe;
RESULTS[1,1]=FHA;
RESULTS[2,1]=FHB;
RESULTS[3,1]=FHAB;

/* Calculations for Box-type adjustment */
MA=(i(nla)-(1/nla)*j(nla))@(1/nlb*j(nlb));
DMA=diag(MA);
denA=trace(DMA*Shat);
dmas=dma*shat;
QA=nobs*Xbar`*MA*Xbar;
FNA=QA/denA;
RESULTS[1,3]=FNA;

MB=((1/nla)*j(nla))@(i(nlb)-(1/nlb)*j(nlb));
DMB=diag(MB);
denb=trace(DMb*Shat);
Qb=nobs*Xbar`*Mb*Xbar;
FNB=Qb/denb;
RESULTS[2,3]=FNB;

MAB=(i(nla)-(1/nla)*j(nla))@(i(nlb)-(1/nlb)*j(nlb));
DMAB=diag(MAB);
denAb=trace(DMAb*Shat);
QAb=nobs*Xbar`*MAb*Xbar;
FNAB=QAb/denAb;
RESULTS[3,3]=FNAB;
Lambda=DIAG(1/NI);

/* Calculate adjusted df */
fA=((trace(DMA*Shat))**2)/(trace((MA*Shat)*(MA*Shat)));
foA=(trace(DMA*Shat))**2/(trace(DMA**2*Shat**2*Lambda));
fB=((trace(DMB*Shat))**2)/(trace((MB*Shat)*(MB*Shat)));
foB=(trace(DMB*Shat))**2/(trace(DMB**2*Shat**2*Lambda));
fAB=((trace(DMAB*Shat))**2)/(trace((MAB*Shat)*(MAB*Shat)));
foAB=(trace(DMAB*Shat))**2/(trace(DMAB**2*Shat**2*Lambda));

/* Calculate p-values */
adjpvalA=1-probf(FnA,fA,foA);
RESULTS[1,4]=ADJPVALA;
adjpvalB=1-probf(FnB,fB,foB);
RESULTS[2,4]=ADJPVALB;
adjpvalAB=1-probf(FnAB,fAB,foAB);
RESULTS[3,4]=ADJPVALAB;
pvala=1-probf(fHa,dfa,dfe);
RESULTS[1,2]=PVALA;
pvalb=1-probf(fHb,dfb,dfe);
RESULTS[2,2]=PVALB;
```

```
pvalab=1-probf(fHab,dfab,dfe);
RESULTS[3,2]=PVALAB;

/* Print results */
headings={'      ANOVA F' '     P-value' '     Adjusted F' '     P-
value'};
EFFECT={A,B,AB}; mattrib RESULTS colname=HEADINGS
FORMAT=7.3 label='Results';
print effect RESULTS;

quit;
```