

A Simulation Study to Compare the Performance of Permutation Tests for Time by Group Interaction in an Unbalanced Repeated-Measures Design, Using Two Permutation Schemes

Mark S. Litaker and Bob Gutin, Medical College of Georgia, Augusta GA

ABSTRACT

The use of permutation tests for the analysis of data incorporating repeated measurements presents a choice of possible permutation schemes to generate the empirical null distribution of the test statistic. A commonly-recommended scheme keeps the vector of observations for each subject intact, with permutation based on group membership only. Another option is to permute the observations across the time points as well as relative to treatment group. These two permutation schemes are compared with respect to power and preservation of the nominal alpha level in testing for a time by group interaction using simulated data based on a randomized trial of effects of an exercise program. SAS® PROC IML is used to generate correlated observations under the null and alternative hypotheses which are then analyzed using PROC GLM. P-values are obtained for the usual F statistic and for each of the permutation tests. Proportions of significant interaction terms are calculated to obtain empirical power and alpha for each of the permutation schemes. The two schemes are then applied to the analysis of observed data from the study.

INTRODUCTION

In biological research it is often the case that observed data do not follow convenient distributions. Since many of the most-used methods of statistical analysis rely on the assumption that the data are sampled from a Normal distribution, use of these methods in many practical situations may give questionable results. While transformations may provide a remedy for some distributional problems, it is often the case that no suitable transformation exists. An alternative approach to analysis of such data is the use of permutation testing. Permutation tests are exact for all sample sizes regardless of the underlying distribution (Higgins, Noble; 1993). Assumptions regarding the form of the correlation structure of the parent distribution are not required. Two initial steps in the use of a permutation test are to identify a test statistic that differentiates between the null and alternative hypothesis situations, and to identify a permutation scheme that will generate the distribution of the test statistic under the null hypothesis (Good; 1995). Often there is a clear choice of permutation scheme based on the experimental design, but this is not always the case. In designs that incorporate repeated measurements on the same experimental units, one must decide whether to keep each subject's vector of observations intact or to permute across time points as well.

PERMUTATION TESTS

To perform a permutation test, the test statistic is calculated for the observed data. Then the data is permuted across groups, and the test statistic is calculated for each permutation. Typically a very large number of permutations is possible, so a random sample of permutations is selected instead of all possible arrangements. The test consists of calculating the area under the curve of this empirical null distribution of the test statistic that is at least as extreme as the test statistic that was calculated from the observed data. That is, the p-value is the fraction of the test statistics from the null distribution that are at least as extreme as that which resulted from the actual data. Typically, a

random sample of 100 to 500 repetitions is adequate to approximate the permutation distribution (O'Sullivan, et al; 1984).

THE REPEATED MEASURES SITUATION

Consider an experimental design with two groups and repeated measurements made at three time points. The effect that is of interest is the difference between groups in change across the times of observation. An appropriate test statistic is one whose magnitude reflects the amount of deviation between the observed data and the null situation. For the current experimental design an obvious choice is the usual F statistic for group by time interaction from a split-plot or repeated measure analysis of variance (ANOVA). To generate the permutation distribution of the F statistic, observations are permuted across group assignments. The order of observations for each subject may be maintained, or the observed values for each subject may be permuted across observation times as well. In the repeated measures design, subjects are randomized to one or the other of the experimental groups, but the repeated observations occur at fixed times. That is, subjects are not randomized to times. It has been suggested (Higgins) that the permutation scheme should correspond to the randomization scheme used in the experiment. However, this is not consistent with the situation with the usual Normal-based F test. The null hypothesis for the overall F test is that both of the main effects and the interaction effect are zero, that is, that neither time nor group is a meaningful categorization, whether experimental units are randomized to group only or to group and subplot. Thus, it is of interest to compare the performance of the two permutation schemes.

EXAMPLE: A STUDY OF THE EFFECTS OF PHYSICAL TRAINING IN CHILDREN

Study subjects were 80 obese children 7 – 11 years of age, who were randomized to participate in a physical training program during either the first or second four-month period of the study. The primary interest was in estimating changes in body composition due to the physical training program. For this evaluation, percent body fat as measured by dual energy x-ray absorptiometry was selected as the dependent variable. The statistical analysis was by mixed-model ANOVA, which was performed using SAS PROC GLM. Subject was included as a random effect, and group and time were fixed effects in the model. The statistical test which is of interest is the F test for the group by time interaction term.

THE SIMULATED DATA SETS

In order to compare the performance of the two permutation schemes, simulated data was generated having approximately the structure that was observed in the actual study. The simulated data sets each consist of 80 subjects and 3 observations. Correlated observations were produced by generating for each subject a "subject effect" consisting of a Uniform(0,20) random variable. Observations at each of the time points were generated by adding a Normal random number to the subject effect with means set to generate the desired group by time effects. The standard deviation of the Normal term was set at 3.5, giving a standard deviation of approximately 6.75 for the sum of the Uniform and Normal components. This is similar to the within-cell standard deviation of the observed data. Since the subject effect is Uniform rather than Normal, the resulting

distributions will be symmetric, but will be overdispersed relative to the Normal distribution.

EVALUATING THE PERMUTATION TESTS

Each of the simulated data sets was analyzed using both of the permutation schemes. In scheme 1, each subject's vector of observations remains intact, and these 3-element vectors are permuted across the group assignments. In scheme 2, the vectors of observations are permuted across groups, and then the three observations for each subject are randomly permuted across times. In order to calculate p-values to four decimal places, 1000 random permutations were performed for each test. For each of the permutation tests and the Normal-based ANOVA on each simulated data set, p-values for group by time interaction were written to text files for performance comparisons. Validity of the tests was evaluated by simulating data with no group by time interaction. In the simulations of the null situation, mean percent body fat was increasing over the times of observation in both groups. If the test rejects the null hypothesis in 5% of trials when $\alpha = .05$, when the null hypothesis is true, then it is a valid test. To evaluate the power of the tests, data was simulated with a group by time interaction. For group 1, the group which received the training program during months 0 – 4, mean percent fat values were 44, 42, and 43 for months 0, 4, and 8, respectively. For group 2, the corresponding means were 44, 45, and 43. The power of the test is the proportion of trials that result in rejection of the null hypothesis. At the 95% confidence level, this is the proportion of tests which yield a p-value of .05 or less.

RESULTS

Preliminary results of the simulation study show little evidence of a difference in performance between the two permutation tests. In 253 simulations under the null hypothesis, scheme 1 demonstrated an empirical alpha level of .0553, while scheme 2 showed .0593. The observed alpha for the F test was also .0593. Under the alternative situation, observed power in 285 simulations was 76.5% for the test using permutation scheme 1, 77.2% using scheme 2, and 76.8% for the F test.

Analysis of the observed data from the physical training study using the Normal-based ANOVA and each of the permutation tests gave similar results. The p-values for group by time interaction were .0007 for the ANOVA, and .002 for the permutation tests. In this case, the two permutation schemes gave identical results.

CONCLUSION

Based on these results, there is not a clear performance basis on which to choose between the permutation schemes. Permutation scheme 1 is a simpler procedure than scheme 2, and requires less computing time. This may suggest that scheme 1 would be preferable in practice. Performance of both permutation tests was similar to that of the Normal-theory test for the observed and simulated data in this study. The assumption of Normality is not seriously violated for these data, despite the use of a mixed distribution in the simulations. Thus, this study may not provide an adequate tool to distinguish differences in performance between the tests that might exist for data with a more extreme departure from Normality.

REFERENCES

1. SAS/IML® Software: Usage and Reference, Version 6, First Edition. SAS Institute, Inc., Cary, NC; 1980.
2. Higgins JJ, Noble W. A Permutation Test for a Repeated Measures Design, Proceedings of the 1993 Kansas State University Conference on Applied Statistics in Agriculture; 1993
3. Good P. Permutation Tests: A Practical Guide to Resampling Methods for Testing Hypotheses. Springer,

New York; 1995.

4. O'Sullivan F, Whitney P, Hinshelwood MM, Hauser ER. The Analysis of Repeated Measures Experiments in Endocrinology. J Animal Science 59 (4):1070-1079; 1984.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Author Name: Mark Litaker
Company: Medical College of Georgia
Address: Office of Biostatistics and Bioinformatics
City state ZIP: Augusta, GA 30912-4900
Work Phone: (706) 721-3785
Fax: (706) 721-6294
Email: mlitaker@mail.mcg.edu

APPENDIX.

options cleanup nosource nonotes;

```
*-----;
* provide a name for the dependent variable ;
*-----;
%let dep = fat;

*-----;
* identify 3 measurements of the dependent variable ;
*-----;
%let dep1 = fat1;
%let dep2 = fat2;
%let dep3 = fat3;

*-----;
* macro to repeatedly simulate data and analyze by ;
* permutation tests using different permutation schemes ;
*-----;

* permutation scheme 1 *;

%macro permute1(reps1);
  %do i = 1 %to &reps1;
  *-----;
  * reassign 3 observations per subject randomly to groups ;
  * keeping each subjects vector of observations intact ;
  *-----;

  proc iml;
  use &dep;
  read all var {subject &dep1 &dep2 &dep3} into &dep;
  rand1=j(nrow(&dep),1,1);
  do i = 1 to nrow(rand1);
    rand1(ii)=ranuni(-1);
  end;
  order=&dep||rand1;
  create order2 var {subject &dep1 &dep2 &dep3 rand1};
  append from order;
  run;

  proc sort data=order2;
  by rand1;
  run;
  proc iml;
  use order2;
  read all var {subject &dep1 &dep2 &dep3} into reorder;
  use fat;
  read all var {group} into group;
  permute1=group||reorder;
  create permute var {group subject &dep1 &dep2 &dep3};
  append from permute1;
  run;

  *-----;
  * analysis of variance for the ;
  * permuted data ;
  *-----;

  * rearrange data to have a single observation per record;
  data permute2;
  set permute;
  do i = 0 to 8 by 4;
    subject = subject;
    group = group;
    month = i;
    if month eq 0 then &dep = &dep1;
    else if month eq 4 then &dep = &dep2;
    else if month eq 8 then &dep = &dep3;
  output;
  end;
  keep subject group month &dep;
  run;
  proc glm data=permute2 noprint outstat=stats;
```

```
class subject group month;
model &dep = subject(group) group month group*month /ss3;
random subject(group);
run;
data stats;
set stats;
if _source_ = 'GROUP*month';
file "&dep 1.txt" mod;
put f;
run;
%end;
%mend permute1;

* permutation scheme 2;
* randomization by group and time;

%macro permute2(reps2);
  %do i = 1 %to &reps2;
  *-----;
  * reassign 3 observations per subject randomly to groups ;
  * and randomly to times ;
  *-----;

  proc iml;
  use &dep;
  read all var {&dep1 &dep2 &dep3} into &dep;
  rand1=j(ncol(&dep),1,1);
  do i = 1 to nrow(rand1);
    rand1(ii)=ranuni(-1);
  end;
  rand2=rand1';
  randcols=rand2//&dep;

  create order2 var {col1 col2 col3};
  append from randcols;
  run;

  proc transpose data=order2 out=order3;
  run;
  * permute columns;
  proc sort data=order3;
  by col1;
  run;
  data order3;
  set order3;
  drop _name_ col1;
  run;
  proc transpose data=order3 out=order2;
  run;
  data order2;
  set order2;
  drop _name_;
  &dep1=col1;
  &dep2=col2;
  &dep3=col3;
  drop col1 col2 col3;
  run;
  proc iml;
  use &dep;
  read all var {group} into group;
  rand2=j(nrow(group),1,1);
  do i = 1 to nrow(rand2);
    rand2(ii)=ranuni(-1);
  end;
  group=group||rand2;
  create group var {group rand};
  append from group;
  * permute rows;
  proc sort data=group;
  by rand;
  run;

  proc iml;
  use group;
```

```

read all var {group};
use order2;
read all var {&dep1 &dep2 &dep3} into &dep;
use &dep;
read all var {subject} into subject;
permute=subject||group||&dep;
create permute var {subject group &dep1 &dep2 &dep3};
append from permute;
run;

*-----;
* analysis of variance for the randomized data ;
*-----;

* rearrange data to have a single observation per record;
data permute2;
set permute;
do i = 0 to 8 by 4;
  subject = subject;
  group = group;
  month = i;
  if month eq 0 then &dep = &dep1;
  else if month eq 4 then &dep = &dep2;
  else if month eq 8 then &dep = &dep3;
  output;
end;
keep subject group month &dep;
run;
proc glm data=permute2 noprint outstat=stats;
class subject group month;
model &dep = subject(group) group month group*month /ss3;
random subject(group);
run;
data stats;
set stats;
if _source_ = 'GROUP*month';
file "&dep 2.txt" mod;
put f;
run;
%end;
%mend permute2;

*-----;
* macro to simulate data and perform 3 tests ;
*-----;
%macro simstudy(reps);
  %do j = 1 %to &reps;

*-----;
* SIMULATE data with one record per subject, ;
* three observations per subject in each record ;
*-----;

proc iml;
subject=j(80,1,1);
group1=j(40,1,1);
group2=j(40,1,2);
subeffect=j(80,1,1);
fixed1=j(80,1,1);
fixed2=j(80,1,1);
fixed3=j(80,1,1);
fat1=j(80,1,1);
fat2=j(80,1,1);
fat3=j(80,1,1);
group=group1//group2;

* generate the alternative hypothesis situation ;
* simulate data with group*time effect;

do i = 1 to 40;
  subject(ii) = i;
  subeffect(ii)= 20*ranuni(-1);
  fixed1(ii) = 34 + 3.5*rannor(-1);
  fixed2(ii) = 32 + 3.5*rannor(-1);

```

```

  fixed3(ii) = 33 + 3.5*rannor(-1);
end;
do i = 41 to 80;
  subject(ii) = i;
  subeffect(ii)= 20*ranuni(-1);
  fixed1(ii) = 34 + 3.5*rannor(-1);
  fixed2(ii) = 35 + 3.5*rannor(-1);
  fixed3(ii) = 33 + 3.5*rannor(-1);
end;

/* *OR generate the null situation;
* simulate data with time effect only;
do i = 1 to 80;
  subject(ii) = i;
  subeffect(ii)= 20*ranuni(-1);
  fixed1(ii) = 34 + 3.5*rannor(-1);
  fixed2(ii) = 35 + 3.5*rannor(-1);
  fixed3(ii) = 36 + 3.5*rannor(-1);
end;
*/
fat1=subeffect+fixed1;
fat2=subeffect+fixed2;
fat3=subeffect+fixed3;
fat=subject||group||fat1||fat2||fat3;
create fat var {subject group fat1 fat2 fat3};
append from fat;
run;

*-----;
* analysis of observed data (anova) ;
*-----;

* rearrange data to have a single observation per record;

data &dep2;
set &dep;
do i = 0 to 8 by 4;
  subject = subject;
  group = group;
  month = i;
  if month eq 0 then &dep = &dep1;
  else if month eq 4 then &dep = &dep2;
  else if month eq 8 then &dep = &dep3;
  output;
end;
keep subject group month &dep;
run;

* analysis of variance;

proc glm data=&dep2 noprint outstat=fstat;
class subject group month;
model &dep = subject(group) group month group*month /ss3;
random subject(group);
run;

```

```

* save the observed value of the F-statistic ;

data fstat;
set fstat;
if _source_ = 'GROUP*month';
file 'alt.txt' mod;
put _name_ _source_ f prob;
call symput('obs_f',f); *value of F from the observed data;
run;

* run permutation test using scheme 1;
%permute1(1000)
run;

data fstats;
infile "&dep 1.txt" missover;
input f;
if f eq . then delete;
p = 0;
if f ge &obs_f then p = 1;
label p='(p-value)';
run;
proc means data=fstats n mean noprint;
title1 'permutation 1: permute by group only';
title2 'group*time p-value';
title3 "dependent variable: &dep";
output out=pval1 n=n1 mean=pval1;
var p;
run;
data pval1;
set pval1;
file 'alt.txt' mod; * if simulating group*time effect;
* file 'null.txt' mod; * if simulating null ;
put n1 pval1;
run;

* run permutation test using scheme 2;
%permute2(1000)
run;

data fstats;
infile "&dep 2.txt" missover;
input f;
if f=. then delete;
p = 0;
if f ge &obs_f then p = 1;
label p='(p-value)';
run;
proc means data=fstats n mean noprint;
title1 'permutation 2: permuting by group and time';
title2 'group*time p-value';
title3 "dependent variable: &dep";
output out=pval2 n=n2 mean=pval2;
var p;
data pval2;
set pval2;
file 'alt.txt' mod;
put n2 pval2;
run;

* remove temporary files;
data clear;
space=.;
file "&dep 1.txt";
put space;
file "&dep 2.txt";
put space;
run;

%end;
proc datasets kill;
%mend simstudy;
run;

```

```

%let reps=200; * set number of simulations;
%simstudy(&reps)
run;

* calculate and print empirical power and alpha values;

data output;
infile 'alt.txt'; * for alternative H;
* infile 'null.txt'; * for null H;
input _name_ $ _source_ $ obsf obsp / n1 pval1 / n2 pval2;
if obsp le .05 then obs_power=1;
if obsp gt .05 then obs_power=0;
if pval1 le .05 then p1_power=1;
if pval1 gt .05 then p1_power=0;
if pval2 le .05 then p2_power=1;
if pval2 gt .05 then p2_power=0;
run;

proc means data=output n mean maxdec=4;
title1 'power for F and two permutation tests';
title2 'alternative hypothesis';
*title2 'null hypothesis';
var obs_power p1_power p2_power;
run;

```