ANALYSIS OF COVARIANCE: REPEATED MEASURES AND SPLIT-PLOT DESIGNS

George A. Milliken, Kansas State University

INTRODUCTION

Repeated measures or split-plot designs involve several sizes of experimental units (Milliken and Johnson 1984 chapter 5). The covariate could be measured on any of the sizes of experimental units. The size of the experimental unit on which the covariate is measured must be considered when constructing the covariance model in order to obtain an appropriate analysis. The models and analyses of the models differ from experiments where the covariate is measured on the large size of experimental unit than when the covariate is measured on the small size of experimental unit. This paper presents an introduction to the analysis of covariance by describing a case for split-plot designs with two sizes of experimental units. Specifically, the situation discussed is a model with the covariate is measured on the the subplot or smaller size of experimental unit. There are two levels of analysis, the between whole plot analysis and the within whole plot analysis. The second section explores the two types of analyses on the covariate part of the model detailing the cases for different assumptions on the form of the slopes. The last section discusses the analysis through an example.

The analysis of covariance models for split-plot and repeated measure designs take on the relevant characteristics pertaining to the analysis covariance models and the prohlems encountered when analyzing split-plot and repeated measures designs which are discussed in Chapters 5 and 24-28 of Milliken and Johnson (1984). In particular, the concept of size of experimental unit must be used to specify an appropriate model and is used when constructing the appropriate analysis. For example, if an experimental design involves three sizes of experimental units, then it is possible for a covariate to be measured on any of the three sizes of experimental units. Thus, in the terminology of a split-plot experiment, the covariate could he measured on a whole plot, or a covariate measured on a subplot, or a covariate measured on a sub-sub-plot. In terms of a repeated measures experiment the covariate could be measured on the largest size of experimental unit, or a covariate measured on the medium size of experimental unit or a covariate measured on the smallest experimental unit.

As described in Milliken and Johnson (1984), there is an analysis for each size of experimental unit as the model can be partitioned into groups of terms corresponding to each size of experimental unit.

Because of the possibility of heterogeneous slopes, the covariate can have an effect on the analysis of all sizes of experimental units. Even when the slopes are homogeneous, the covariate will have an effect on all sizes of experimental units larger than the one on which the covariate is measured. There is an estimate of the slope from all models for experimental units larger than the one on which the covariate was measured. The estimates could be combined into one estimate in the same manner as combining inter intra block estimates in an incomplete block design.

When constructing an appropriate covariance model, the covariate term is included with the segment of the model corresponding to the size of the experimental unit on which the covariate is measured. The following example is used to demonstrate the case when the covariate is measured on the smaller size of experimental unit for a model with two sizes of experimental units.

EXAMPLE:

The data set in Table 1 contains data from people who were given a blood pressure drug to determine the effect on their blood pressure. Their blood pressure was measured once a week for four weeks. The covariate is the amount of sodium in the persons diet on the day before the blood pressure measurement was obtained. The large size of experimental unit is the person, the entity which was randomly assigned to the DRUG, and the small size of experimental unit is a one week interval as each person's time in the experiment is split into four weeks. Since the amount of sodium was measured each week, the covariate is measured on the small size of experimental unit.

The repeated measures or split-plot model (ignoring the covariate)is:

$$y_{ijk} = \mu_{ik} + a_{ij} + \varepsilon_{ijk} \quad i = 1,2,\ldots,d, \quad j = 1,2,\ldots,n_i, \; k = 1,2,\ldots,t \qquad (1)$$

where μ_{ik} denotes the expected response of drug i at week k, a_{ij} denotes the random effect of subject j assigned to drug i which is assumed to be distributed iid $N(0, \sigma_a^2)$ and ε_{ijk} denotes the random effect of a blood pressure measurement taken at week k from subject j of drug i, which are assumed to be distributed iid $N(0, \sigma_\varepsilon^2)$.

By expressing μ_{ik} as

$$\mu_{ik} = \mu + \rho_i + \tau_k + (\rho\tau)_{ik}, \qquad (2)$$

the blood pressure model can be expressed in terms of the two sizes of experimental units as

$$y_{ijk} = \underbrace{\mu + \rho_i + a_{ij}}_{\text{subject part of the model}} + \underbrace{\tau_k + (\rho\tau)_{ik} + \varepsilon_{ijk}}_{\text{week part of model}} \qquad (3)$$

where $i = 1,2,\ldots,d$, $j = 1,2,\ldots,n_i$, and $k = 1,2,\ldots,t$. The measurement of the covariate, sodium content, is obtained from the

small size of experimental unit, or the week. An appropriate analysis of covariance model is

$$y_{ijk} = \underbrace{\mu + \rho_i + a_{ij}}_{\text{subject part of model}} + \underbrace{\tau_k + (\rho\tau)_{ik} + \beta_{ik}x_{ijk} + \varepsilon_{ijk}}_{\text{the week part}} \quad (4)$$

where $i = 1,2,\ldots,d$, $j = 1,2,\ldots,n_i$, and $k = 1,2,\ldots,t$, x_{ijk} is the measurement of sodium content on week k of subject j of drug i, and β_{ik} is the slope for the covariate sodium content corresponding to drug i and week k.

General Representation of the covariate part of the model. Before continuing with the analysis of the above data set, the general covariate model will be discussed. The covariance model, with a single covariate, can be expressed as

$$y_{ijk} = \mu_{ik} + \beta_{ik}x_{ijk} + e_{ij} + \varepsilon_{ijk}. \quad (5)$$

It is explicitly assumed that the covariate is linearly related to the dependent measure. One should plot the data or residuals from the model before continuing the analysis of the model. If the relationship is quadratic then a model with x_{ijk}^2 should be considered.

There are two types of comparisons which can he made;

(1) between large size experimental units or whole plots as

$$\bar{y}_{ij\cdot} = \sum_{k=1}^{t} y_{ijk}/t \quad (6)$$

and (2) within a large size experimental unit or a whole plot represented hy a contrast between the small size experimental units with each large size experimental unit as

$$\sum_{k=1}^{t} c_k y_{ijk} \text{ where } \sum_{k=1}^{t} c_k = 0. \quad (7)$$

The large size experimental unit model is

$$\bar{y}_{ij\cdot} = \bar{\mu}_{i\cdot} + \frac{1}{t}\sum_{k=1}^{t} \beta_{ik}x_{ijk} + e^{*}_{ij} \quad (8)$$

where $e^{*}_{ij} = e_{ij} + \bar{\varepsilon}_{ij\cdot}$ and

$\mathrm{Var}(e^{*}_{ij}) = (\sigma^2_{\varepsilon} + t\sigma^2_{e})/t.$

A contrast of small size experimental units within each large size experimental unit provides the model

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k\mu_{ik} + \sum_{k=1}^{t} c_k\beta_{ik}x_{ikj} + \sum_{k=1}^{t} c_k\varepsilon_{ijk} \quad (9)$$

which has variance $\sigma^2_{\varepsilon}\sum_{k=1}^{t} c_k^2$. or variance $c'\Sigma c$ for repeated measures where Σ is the covariance matrix of the repeated measures, assumed to be the same for all persons. This set of comparisons is free of the large size experimental unit errors.

The models 8 and 9 show that information about the covariate can be obtained from both models. There are conditions which provide simplifications of the covariate relationship which need to be explored in order to provide a better understanding of the appropriate approach to the analysis of the model.

There are two possible cases for the value of the covariate;

(1) the covariate can be measured on the large size experimental units, i.e., $x_{ijk} = x_{ij}$ or

(2) the covariate can he measured on the small size of experimental unit, represented by x_{ijk}.

There are five possible forms for which the slope can enter the model;

(1) common slope, $\beta_{ik} = \beta$,

(2) function of the levels of the large size experimental unit treatment as $\beta_{ik} = \beta_i$,

(3) function of the levels of the small size experimental unit treatment as $\beta_{ik} = \beta_k$

(4) an additive function of the levels of both treatments as $\beta_{ik} = \beta_{Bi} + \beta_{Wk}$, where β_{Bi} denotes that part of the slope for the between subject treatment and β_{Wk} denotes that part of the slope for the within subject treatment, or

(5) function of the levels of both treatments as β_{ik}.

To help determine the proper forms of the models, the covariate part of the model can be expressed as

$$\begin{aligned}\beta_{ik}x_{ijk} &= (\beta_{ik} - \bar{\beta}_{i\cdot})(x_{ijk} - \bar{x}_{ij\cdot}) \\ &\quad + (\beta_{ik} - \bar{\beta}_{i\cdot})\bar{x}_{ij\cdot} \\ &\quad + \bar{\beta}_{i\cdot}(x_{ijk} - \bar{x}_{ij\cdot}) + \bar{\beta}_{i\cdot}\bar{x}_{ij\cdot}. \quad (10)\end{aligned}$$

This paper is only considering the case covariate measured on the small size of experimental units. The covariate part of the model can be simplified by making assumptions on the form of the slope.

If the common slope model, $\beta_{ik} = \beta$, is appropriate, the large size experimental unit model of equation 8 is

$$\bar{y}_{ij\cdot} = \bar{\mu}_{i\cdot} + \beta\,\bar{x}_{ij\cdot} + e^{*}_{ij} \quad (11)$$

a common slope analysis of covariance model.

The comparison of small size experimental units within a large size of experimental unit is

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k\mu_{ik} + \beta\sum_{k=1}^{t} c_k(x_{ijk} - \bar{x}_{ij\cdot}) + \sum_{k=1}^{t} c_k\varepsilon_{ijk}. \quad (12)$$

For this case, two estimates of the slope can be obtained, one from model 11 and one from the small size of experimental unit analysis. If there are enough large experimental units,

the two estimators could be combined into a more efficient estimator.

If the slope is a function of the levels of the large size of experimental unit treatment, $\beta_{ik} = \beta_i$, the large size of experimental unit model is

$$\bar{y}_{ij\cdot} = \bar{\mu}_{i\cdot} + \beta_i \bar{x}_{ij\cdot} + e^*_{ij} \qquad (13)$$

a covariance model with unequal slopes

The comparison of the small size of experimental units within a large size of experimental unit is

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k \mu_{ik} + \beta_i \sum_{k=1}^{t} c_k (x_{ijk} - \bar{x}_{ij\cdot}) + \sum_{k=1}^{t} c_k \varepsilon_{ijk}. \qquad (14)$$

Again, estimates of β_i, $i = 1,2,\ldots,d$ can be obtained from model 13 and from the small experimental unit analysis which could be combined into a more efficient estimator. There also exists two tests of the equal slope hypothesis, $\beta_1 = \beta_2 = \ldots = \beta_d$.

If the slope is a function of the levels of the small size of experimental unit treatment, then the large size of experimental unit model is

$$\bar{y}_{ij\cdot} = \bar{\mu}_{i\cdot} + \frac{1}{k} \sum_{k=1}^{t} \beta_k x_{ijk} + e^*_{ij} \qquad (15)$$

a multiple covariate model (k covariates) and the within large size of experimental unit comparisons are

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k \mu_{ik} + \sum_{k=1}^{t} c_k (\beta_k - \bar{\beta}_{\cdot})(x_{ijk} - \bar{x}_{ij\cdot}) + \sum_{k=1}^{t} c_k \varepsilon_{ijk} \qquad (16)$$

The small size of experimental unit analysis provides estimates of contrasts of the β_k and a test of the equal slope hypothesis.

Estimates of the $(\beta_k - \bar{\beta}_{\cdot})$, $k = 1,2,\ldots,t$ can be obtained from model 16.

If the slope is an additive function of the levels of the two treatments as $\beta_{ik} = \beta_{Bi} + \beta_{Wk}$, the large size experimental unit model is

$$\bar{y}_{ij\cdot} = \bar{\mu}_{i\cdot} + \beta_{Bi} \bar{x}_{ij\cdot} + \sum_{k=1}^{t} \beta_{Wk} x_{ijk}/t + e^*_{ij}, \qquad (17)$$

a multiple covariate model with covariates

$$\bar{x}_{ij\cdot}, x_{ij1}, x_{ij2}, \ldots, x_{ijt}$$

when slopes correspond to $\bar{x}_{ij\cdot}$ are unequal for each whole plot treatment.

The within large size of experimental unit comparisons are

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k \mu_{ik} + \sum_{k=1}^{t} c_k \beta_{Wk} (x_{ijk} - \bar{x}_{ij\cdot}) + \beta_{Bi} \sum_{k=1}^{t} c_k (x_{ijk} - \bar{x}_{ij\cdot}) + \sum_{k=1}^{t} c_k \varepsilon_{ijk}. \qquad (18)$$

If there are no simplifying assumptions about the slopes, the large size of experimental unit model is in equation 8 aud the within large size of experimental unit comparisons are

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k \mu_{ik} + \sum_{k=1}^{t} c_k (\beta_{ik} - \bar{\beta}_{i\cdot}) x_{ijk} + \sum_{k=1}^{t} c_k \varepsilon_{ijk}. \qquad (19)$$

MAKING THE NECESSARY COMPARISONS

When conducting the analysis of the appropriate covariance model the treatments (DRUGS) need to be compared, depending on the whether there is interaction between the treatments. Due to lack of space, it is assumed that a common slope model is adequate to describe the data.

The large size of experimental unit model

$$\bar{y}_{ij\cdot} = \bar{\mu}_{i\cdot} + \beta \bar{x}_{ij\cdot} + e^*_{ij}, \qquad (20)$$

a common slope analysis of covariance model, is used to ohtain estimates of $\bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot}$ and of other interesting contrasts of the $\bar{\mu}_{i\cdot}$'s.

The small size experimental units comparisons

$$\sum_{k=1}^{t} c_k y_{ijk} = \sum_{k=1}^{t} c_k \mu_{ik} + \beta \sum_{k=1}^{t} c_k (x_{ijk} - \bar{x}_{ij\cdot}) + \sum_{k=1}^{t} c_k \varepsilon_{ijk}, \qquad (21)$$

are used to obtain estimates of

1) $\bar{\mu}_{\cdot 1} - \bar{\mu}_{\cdot 2}$, 2) $\mu_{11} - \mu_{12}$, 3) $\mu_{11} - \bar{\mu}_{1\cdot}$, and other interesting contrasts of these parameters.

If there is a DRUG*WEEK interaction, one possible comparison is to compare DRUGS at each WEEK. This information is obtained by combining information from two models. For example, to obtain estimates of $\mu_{11} - \mu_{21}$ use the relationship

$$\mu_{11} - \mu_{21} = [(\mu_{11} - \bar{\mu}_{1\cdot}) - (\mu_{21} - \bar{\mu}_{2\cdot})] + \{\bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot}\}$$

where $[(\mu_{11} - \bar{\mu}_{1\cdot}) - (\mu_{21} - \bar{\mu}_{2\cdot})]$ is a within person comparison and $\{\bar{\mu}_{1\cdot} - \bar{\mu}_{2\cdot}\}$ is a between person comparison. The estimates from the two models are independently distrihuted thus the

variance is the sum of the two variances. Let

SE1= STDERR of est$[(\mu_{11} - \bar{\mu}_{1.})-(\mu_{21} - \bar{\mu}_{2.})]$

and SE2= STDERR of est $(\bar{\mu}_{1.} - \bar{\mu}_{2.})$

then the

STDERR of est $(\mu_{11} - \mu_{21})$=SQRT$(SE1^2 + SE2^2)$

One can construct confidence intervals or carry out multiple comparisons between these means by using the approximate t value (Cochran and Cox 1957 pp 298-299). The approximate t value is

$t^* = (t_{df1}SE1^2 + t_{df2}SE2^2)/(SE1^2 + SE2^2)$

where df1 is the degrees of freedom associated with the sub-plot or WEEK error term and df2 is the degrees of freedom associated with the whole-plot or person error term.

The other approach is to approximate degrees of freedom using the Satterthwaite approximation which can be used to determine significance levels of t-statistics or construct multiple comparison procedures other than the LSD.

The approximate degrees of freedom are

$df = (SE1^2 + SE2^2)^2/((SE1^2)^2/df1 + (SE2^2)^2/df2)$.

If the data are from a repeated measures design, one would want to adjust the degrees of freedom for deviations of the repeated measures errors from the Huynh-Feldt condition. The REPEATED option will not work when the covariate is measured on the small size of experimental unit but the G-G epsilon H-F epsilon can be obtained by following the next steps:

a) Fitting the usual split-plot model with the covariate.
b) Compute the residuals for the model
c) Do a repeated measures analysis on the residuals without the covariate.
d) Use the G-G epsilon or H-F epsilon to correct the degrees of freedom of the small size experimental unit analysis from the split-plot analysis.

Example. The data in Table 1 are the blood pressure values and the sodium content values which were read in via the SAS System Code in Table 2. The person means were computed and are displayed in Table 3. The SAS System Code in Table 4 fits the person model to the person means. The variable NUM=4 is used as a weight variable in order for the sums of squares to be comparable across analyses. The results from fitting the person model are in Table 5. The sum of squares error(person) has 21 degrees of freedom. The estimate statements have been included to provide estimates of some of the comparisons, which would be used alone only if there is no DRUG*WEEK interaction. The LSMEANS provides adjusted means for comparing DRUGS averaged over WEEKS at the average NA value. The contrast statement is used to provide the sum of squares to test for equal intercepts (DRUG effects) as the Sum of squares due to DRUGS in the AOV table provides a test of the intercepts equal to zero (since the NOINT option is included). The SAS System Code in part b) of Table 4 does not have the NOINT option, thus the sum of squares due to DRUGS provides a test of equal intercepts, the results are in Table 6.

The next step is to carry out the within person part of the the analysis. Since this is a repeated measures design, one would want to check for deviations from the Huynh-Feldt condition. Unfortunately, when the covariate is measured on the small size of experimental unit and one is using a common slope model, the REPEATED option doses not enable one to carry out an appropriate within subjects repeated measures analysis. Table 7, parts a) and b), contain two possible models, but neither provides the correct within person analysis. Table 8 contains the analysis from part a), which provides a correct between person analysis, but has an incorrect within person analysis. Table 9 contains the results from part b), which are incorrect for both levels of analyses.

Table 10 contains the SAS System Code to construct a data set which can be used to provide the usual split-plot analysis (one observation for each combination of DRUG, SUBJECT and WEEK). Table 11 contains the SAS System Code to carry out the split-plot analysis. The results are in Table 12, which provide the correct within person analysis, but the incorrect between person analysis. The sum of squares error(person) denoted by SUBJ(DRUG) has 22 degrees of freedom rather than 21 as obtained from the between person analysis. The expected mean squares for SUBJ(DRUG) and DRUG show they are not appropriate sums of squares since the coefficients of VAR(SUBJ(DRUG)) are not 4. The value of the F-Statistic provided by the TEST option of the RANDOM statement is larger than the value of the F-Statistic in Table 5. The contrast statements are used to obtain the EMS of each contrast. The EMS's show that ERROR(week) is the correct error since these are within person comparisons. The LSMEANS are provided, but they are not correct since the model has not handled the covariate correctly.

Table 13 contains the SAS System Code for fitting a model with two covariates. The first covariate is the mean NA for each person, MNA, which is a person or large size of experimental unit covariate. The second covariate is the deviation of the individual NA values from the mean, NADEV, which is a week or small size of experimental unit covariate. Table 14 contains the results for this analysis which has the correct SS for error(person), Type III DRUG, and Type I MNA. All of the within person comparisons are correct. All estimates are correct except for MNA. Since the analysis is a within person analysis, by removing the person effects one also removes the MNA effect, hence the SSMNA is zero. The LSMEANS were not estimable from this analysis.

The SAS System Code in Table 15 constructs a data set of residuals from the split-plot model which can be used to carry out a repeated measures analysis. The SAS System Code in Table 16 provides the repeated measures analysis of the residuals and the results are in Table 17. All of the suns of squares in Table 17 are zero

except SSERROR. This analysis provides estimates of the G-G and H-F epsilons which can be used with the split-plot analysis to make necessary comparisons. The adjusted degrees of freedom associated with SSERROR(week) are .6809*65 = 56 (approximately).

All comparisons between DRUGs WEEKs and and combinations are straight forward except for comparing DRUGs at the same of different WEEKs. Table 18 contains the necessary computations for comparing DRUG 1 and DRUG 2 at WEEK 1.

Table 1. DATA SET where Covariate is measured on the smallest size of experimental unit

DRUG=Control

SUBJ	BP1	NA1	BP2	NA2	BP3	NA3	BP4	NA4
1	127	1089	118	654	122	745	127	821
2	137	1009	139	966	137	902	133	829
3	118	872	120	727	120	744	128	828
4	139	1000	137	997	133	868	136	903
5	129	767	136	913	130	726	127	672
6	119	606	123	798	123	779	120	640

DRUG=RZ23

SUBJ	BP1	NA1	BP2	NA2	BP3	NA3	BP4	NA4
1	117	946	110	1097	101	1009	109	808
2	137	826	120	670	110	729	115	752
3	131	1033	122	933	111	923	120	953
4	130	787	131	1081	109	654	135	995
5	141	1005	124	923	119	912	133	1084
6	125	775	112	606	101	642	109	644
7	133	637	135	1051	111	719	125	884
8	118	701	111	758	109	1063	115	1007

DRUG=XW31

SUBJ	BP1	NA1	BP2	NA2	BP3	NA3	BP4	NA4
1	105	941	91	704	109	887	118	779
2	110	642	116	1082	121	760	140	1045
3	109	961	93	651	97	614	107	687
4	114	786	97	608	112	883	124	904
5	116	739	104	796	114	813	136	1093

DRUG=YA01

SUBJ	BP1	NA1	BP2	NA2	BP3	NA3	BP4	NA4
1	130	743	133	900	119	831	114	915
2	145	1041	136	812	127	1020	120	850
3	138	884	135	1020	128	1049	109	713
4	135	644	140	860	119	601	122	1007
5	123	944	118	808	110	954	93	638
6	134	964	128	944	126	1087	107	793
7	141	1098	119	606	120	1013	110	837

Table 2. SAS System Code for reading the data set, computing the mean blood pressure and sodium content for each person, and print the data set.

```
DM 'LOG;CLEAR;OUTPUT;CLEAR;';
options ls=76 ps=56 pageno=1 nodate;
libname sastalk '\sas\mysas\amdiii\';
data drug; set sastalk.bp;
        * Read in the data set;
        * Compute mean BP and NA for each
                person;
mbp=mean(bp1,bp2,bp3,bp4);
mna=mean(na1,na2,na3,na4);
num=4;  * Number of observations in each mean;
run;
proc print; var drug subj bp1 na1 bp2 na2 bp3
na3 bp4 na4 mbp mna ;
Title 'Analysis of Covariance: Repeated Measures
and Split-plots';
Title2 'Covariate is measured on the smallest
size of experimental unit';
run;
```

Table 3. Data for the person or mean model

OBS	DRUG	MBP	MNA
1	Control	123.50	827.25
2	Control	136.50	926.50
3	Control	121.50	792.75
4	Control	136.25	942.00
5	Control	130.50	769.50
6	Control	121.25	705.75
7	RZ33	109.25	965.00
8	RZ33	120.50	744.25
9	RZ33	121.00	960.50
10	RZ33	126.25	879.25
11	RZ33	129.25	981.00
12	RZ33	111.75	666.75
13	RZ33	126.00	822.75
14	RZ33	113.25	882.25
15	XW31	105.75	827.75
16	XW31	121.75	882.25
17	XW31	101.50	728.25
18	XW31	111.75	795.25
19	XW31	117.50	860.25
20	YA01	124.00	847.25
21	YA01	132.00	930.75
22	YA01	127.50	916.50
23	YA01	129.00	778.00
24	YA01	111.00	836.00
25	YA01	123.75	947.00
26	YA01	122.50	888.50

Table 4. SAS System Code to fit the person model

a) model with NOINT option

```
proc glm; class drug;
model mbp=drug mna/solution noint;
weight num;
contrast 'DRUGS EQUAL' drug 1 -1 0 0 , drug 1 0
-1 0, drug 1 0 0 -1;
lsmeans drug/stderr pdiff;
estimate 'M1.-M2.' drug 1 -1 0 0;
estimate 'M1.-M3.' drug 1 0 -1 0;
estimate 'M2.-M4.' drug 0 1 0 -1;
title3 'Between subjects analysis for comparison
of Drugs';
Title4 'NUM is used as a weight to make sums of
squares comparable';
Title5 'NOINT used to obtain estimates and
standard errors of intercepts';
run;
```

b) Model without NOINT option

```
proc glm; class drug;
model mbp=drug mna/solution ;
weight num;
Title5 'OR use model w/o NOINT to obtain Drug
comparisons';
```

```
run;
```

Table 5. Between subjects analysis for comparison of Drugs. NUM is used as a weight to make sums of squares comparable. NOINT used to obtain estimates and standard errors of intercepts.

```
Dependent Variable: MBP
Weight:     NUM
                                Sum of          Mean
Source            DF          Squares        Square
Model              5       1535472.75     307094.55
Error(person)     21          3787.00        180.33
Uncor Total       26       1539259.75

Source        DF     Type III SS  F Value  Pr > F
DRUG           4      10692.3301    14.82  0.0001
MNA            1        975.2700     5.41  0.0301

Contrast      DF     Contrast SS  F Value  Pr > F

DRUGS EQUAL    3      3138.53948     5.80  0.0047


                                    Std Error of
Parameter              Estimate      Estimate

DRUG       Control   96.68025974   13.84919352
           RZ33      86.73459981   14.35419817
           XW31      80.40621274   13.76652192
           YA01      90.75611491   14.62448185
MNA                   0.03816035    0.01640922

                 Test of Ho: Pr>|T|  Std Error
Parameter   Est  Param=0              Estimate

M1.-M2.    9.9457   2.71    0.0132  3.67249866
M1.-M3.   16.2740   4.00    0.0006  4.06819312
M2.-M4.   -4.0215  -1.15    0.2613  3.48373890

        Least Squares Means

DRUG              MBP        Std Err  LSMEAN
                LSMEAN       LSMEAN   Number

Control    129.121328      2.766635      1
RZ33       119.175668      2.382881      2
XW31       112.847281      3.046592      3
YA01       123.197183      2.577873      4

        Pr > |T| HO: LSMEAN(i)=LSMEAN(j)

         i/j     1       2       3       4
         1    .      0.0132  0.0006  0.1365
         2  0.0132   .       0.1192  0.2613
         3  0.0006  0.1192   .       0.0184
         4  0.1365  0.2613  0.0184   .
```

Table 6. Person model where NUM is used as a weight to make sums of squares comparable and use model w/o NOINT to directly obtain Drug comparisons.

```
Dependent Variable: MBP
Weight:  NUM
                                Sum of          Mean
Source            DF          Squares        Square
Model              4       4326.97245     1081.74311
Error(person)     21       3786.99870      180.33327
Cor Tot           25       8113.97115

Source    DF     Type III SS  F Value     Pr > F
DRUG       3      3138.53948     5.80     0.0047
MNA        1       975.27005     5.41     0.0301


                                    Std Error of
Parameter              Estimate      Estimate

INTERCEPT          90.75611491 B  14.62448185
DRUG       Control  5.92414483 B   3.82608728
           RZ33    -4.02151510 B   3.48373890
           XW31   -10.34990217 B   4.04886472
           YA01     0.00000000 B   .
MNA                 0.03816035     0.01640922
```

Table 7. SAS System Code code to attempt to use the REPEATED option, but both analyses provide incorrect results.

a) model using MNA--the mean

```
Proc GLM; Class Drug;
Model bp1 bp2 bp3 bp4 = Drug Mna/solution;
Repeated Time 4;
Title3 'Analysis Using the REPEATED OPTION with
MNA as Covariate';
Title4 'Correct Between Subjects Analysis--
Incorrect Within Subjects analysis';
Run;
```

b) model using the individual NA values

```
Proc GLM; Class Drug;
Model bp1 bp2 bp3 bp4 = Drug na1 na2 na3 na4/
solution;
Repeated Time 4;
Title3 'Analysis Using the REPEATED OPTION with
NA1 NA2 NA3 NA4 as Covariates';
Title4 'Incorrect Between Subjects Analysis--
Incorrect Within Subjects analysis';
Run;
Analysis Using the REPEATED OPTION with MNA
as Covariate.  Correct Between Subjects
Analysis--Incorrect Within Subjects analysis.
```

Table 8. Repeated measures analysis using MNA as a covariate. Correct bewteen subject analysis and incorrect within subject analysis.

```
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects Effects

Source  DF     Type III SS  F Value     Pr > F

DRUG     3    3138.539477     5.80     0.0047
MNA      1     975.270049     5.41     0.0301
Error   21    3786.998701

Univariate Tests of Hypotheses for Within
 Subject Effects

Source: TIME
     DF     Type III SS F Value Pr > F  G - G
      3     96.82752958  1.33   0.2717 0.2742
Source: TIME*DRUG
     DF     Type III SS F Value Pr > F  G - G
      9   4221.01952069 19.36   0.0001 0.0001
```

```
Source: TIME*MNA
     DF      Type III SS F Value Pr > F  G - G
      3     63.59241154  0.88   0.4588 0.4424
Source: Error(TIME)
     DF      Type III SS
     63   1525.80193370
Greenhouse-Geisser Epsilon = 0.8226
       Huynh-Feldt Epsilon = 1.1181
```

Table 9. Repeated measures analysis using the individual NA values---neither analysis is correct.

```
Repeated Measures Analysis of Variance
Tests of Hypotheses for Between Subjects
Effects

Source    DF     Type III SS   F Value      Pr > F

DRUG       3    3758.103817     10.69      0.0003
NA1        1      33.754133      0.29      0.5981
NA2        1     511.771884      4.37      0.0511
NA3        1     105.378455      0.90      0.3556
NA4        1    1277.736556     10.90      0.0040
Error     18    2109.435387

Repeated Measures Analysis of Variance
Univariate Tests of Hypotheses for Within
Subject Effects
Source: TIME
     DF      Type III SS    Pr > F    G - G
      3     91.74138788    0.0058    0.0144
Source: TIME*DRUG
     DF      Type III SS    Pr > F    G - G
      9   3221.60557063    0.0001    0.0001
Source: TIME*NA1
     DF      Type III SS    Pr > F    G - G
      3    174.29943108    0.0001    0.0006
Source: TIME*NA2
     DF      Type III SS    Pr > F    G - G
      3    249.05523793    0.0001    0.0001
Source: TIME*NA3
     DF      Type III SS    Pr > F    G - G
      3    259.89362333    0.0001    0.0001
Source: TIME*NA4
     DF      Type III SS    Pr > F    G - G
      3    352.73089088    0.0001    0.0001
Source: Error(TIME)
     DF      Type III SS      Mean Square
     54    354.91637849       6.57252553

Greenhouse-Geisser Epsilon = 0.7002
       Huynh-Feldt Epsilon = 1.1031
```

Table 10. SAS System Code to construct a data set to be analyzed via the usual split-plot model.

```
data time;set drug;
*  Construct Data Set for Usual Split-plot
Analysis;
* Compute deviations of NA values from Subject
mean--NADEV;
bp=bp1; na=na1; nadev=na1-mna; time=1; output;
bp=bp2; na=na2; nadev=na2-mna; time=2; output;
bp=bp3; na=na3; nadev=na3-mna; time=3; output;
bp=bp4; na=na4; nadev=na4-mna; time=4; output;
run;
```

Table 11. SAS System Code to carry out the usual split-plot analysis with NA as a covariate. The between subject analysis is incorrect, but the within subject analysis is correct.

```
proc glm; class drug subj time;
model bp = drug subj(drug) time time*drug
na/solution;
Title3 'Model for Within Subject Comparisons';
Title4 'But incorrect Between Subject Analysis';
estimate 'M11-M12' time 1 -1 0 0 drug*time 1 -1
0 0  0 0 0 0  0 0 0 0
   0 0 0 0;
estimate 'M11-M13' time 1 0 -1 0 drug*time 1 0 -
1 0  0 0 0 0  0 0 0 0
   0 0 0 0;
estimate 'M11-M1.' time .75 -.25 -.25 -.25
  drug*time .75 -.25 -.25 -.25   0 0 0 0  0 0 0
0  0 0 0 0;
estimate 'M21-M2.' time .75 -.25 -.25 -.25
  drug*time 0 0 0 0 .75 -.25 -.25 -.25   0 0 0 0
0 0 0 0;
contrast 'M11-M12' time 1 -1 0 0 drug*time 1 -1
0 0  0 0 0 0  0 0 0 0
   0 0 0 0;
contrast 'M11-M13' time 1 0 -1 0 drug*time 1 0 -
1 0  0 0 0 0  0 0 0 0
   0 0 0 0;
contrast 'M11-M1.' time .75 -.25 -.25 -.25
  drug*time .75 -.25 -.25 -.25   0 0 0 0  0 0 0
0  0 0 0 0;
contrast 'M21-M2.' time .75 -.25 -.25 -.25
  drug*time 0 0 0 0 .75 -.25 -.25 -.25   0 0 0 0
0 0 0 0;
lsmeans drug time drug*time;
random subj(drug)/test;
output out=resid r=resbp;
* Compute Residuals for Use with REPEATED
OPTION;
run;
```

Table 12. Results of usual split-plot analysis using NA as a covariate. Correct Model for Within Subject Comparisons but incorrect for Between Subject Analysis.

```
Dependent Variable: BP
Source    DF        Squares F Value      Pr > F
Model     38    14845.5811   56.23      0.0001
Error     65      451.6400
Cor Tot 103    15297.2212

Source       DF      Type I SS   F Value Pr > F
DRUG          3     3351.70240    160.79 0.0001
SUBJ(DRUG)   22     4762.26875     31.15 0.0001
TIME          3     1368.56731     65.65 0.0001
DRUG*TIME     9     4225.28835     67.57 0.0001
NA            1     1137.75432    163.75 0.0001

Source       DF    Type III SS   F Value Pr > F
DRUG          3     3171.76836    152.16 0.0001
SUBJ(DRUG)   22     3820.89522     25.00 0.0001
TIME          3      832.06477     39.92 0.0001
DRUG*TIME     9     3628.30637     58.02 0.0001
NA            1     1137.75432    163.75 0.0001

Contrast     DF    Contrast SS   F Value Pr > F
```

```
M11-M12       1       12.89660       1.86 0.1778
M11-M13       1       13.86256       2.00 0.1626
M11-M1.       1       29.48223       4.24 0.0434
M21-M2.       1     1072.12715     154.30 0.0001

                      T for H0: Std Error of
Parameter Estimate    Param=0    Estimate

M11-M12    -2.0788     -1.36   1.52587060
M11-M13    -2.1723     -1.41   1.53796135
M11-M1.    -1.9429     -2.06   0.94321615
M21-M2.    10.0489     12.42   0.80897500

                                  Std Error
Parameter              Estimate   Estimate

NA                     0.029419   0.00229908

   Least Squares Means
Incorrect
 DRUG          BP        TIME        BP
               LSMEAN                LSMEAN

 Control  128.921751       1     125.445516
 RZ33     119.285745       2     120.246064
 XW31     112.573044       3     117.413041
 YA01     123.438330       4     121.114250


 DRUG     TIME    BP     DRUG     TIME    BP
                 LSMEAN                  LSMEAN
 Control  1     126.97   XW31     1     111.86
 Control  2     129.05   XW31     2     102.61
 Control  3     129.15   XW31     3     112.32
 Control  4     130.49   XW31     4     123.48
 RZ33     1     129.33   YA01     1     133.59
 RZ33     2     119.45   YA01     2     129.86
 RZ33     3     109.42   YA01     3     118.74
 RZ33     4     118.92   YA01     4     111.54

Source      Type III Expected Mean Square
DRUG        Var(Error) + 3.9431 Var(SUBJ(DRUG))
            + Q(DRUG,DRUG*TIME)
SUBJ(DRUG)  Var(Error) + 3.9386 Var(SUBJ(DRUG))
TIME        Var(Error) + Q(TIME,DRUG*TIME)
DRUG*TIME   Var(Error) + Q(DRUG*TIME)
NA          Var(Error) + Q(NA)
Contrast    Contrast Expected Mean Square
M11-M12     Var(Error) + Q(TIME,DRUG*TIME)
M11-M13     Var(Error) + Q(TIME,DRUG*TIME)
M11-M1.     Var(Error) + Q(TIME,DRUG*TIME)
M21-M2.     Var(Error) + Q(TIME,DRUG*TIME)

Tests of Hypotheses for Mixed Model
Analysis of Variance
Source: DRUG *
Error: 1.0011*MS(SUBJ(DRUG)) - 0.0011*MS(Error)

      DF     Type III MS
       3   1057.2561201
 Denominator    Denominator
       DF             MS      F Value    Pr > F
    22.00   173.86655865        6.081    0.0036
```

Table 13. SAS System Code to fit the usual split-plot model using two covariates, MNA and (NA-MNA)=nadev.

```
proc glm data=time; class drug subj time;
model bp = drug mna subj(drug) time time*drug
nadev/ solution;
Title3 'Model with Two Covariates MNA and
NADEV';
Title4 'Correct Sums of Squares for Between and
Within Subject Analyses';
Title5 'But LSMEANS are not estimable';
test b=drug e=subj(drug);
estimate 'M11-M12' time 1 -1 0 0 drug*time 1 -1
0 0  0 0 0 0  0 0 0 0
   0 0 0 0;
estimate 'M11-M13' time 1 0 -1 0 drug*time 1 0 -
1 0  0 0 0 0  0 0 0 0
   0 0 0 0;
estimate 'M11-M1.' time .75 -.25 -.25 -.25
  drug*time .75 -.25 -.25 -.25   0 0 0 0  0 0 0
0  0 0 0 0;
estimate 'M21-M2.' time .75 -.25 -.25 -.25
  drug*time 0 0 0 0 .75 -.25 -.25 -.25   0 0 0 0
0 0 0 0;
contrast 'M11-M12' time 1 -1 0 0 drug*time 1 -1
0 0  0 0 0 0  0 0 0 0
   0 0 0 0;
contrast 'M11-M13' time 1 0 -1 0 drug*time 1 0 -
1 0  0 0 0 0  0 0 0 0
   0 0 0 0;
contrast 'M11-M1.' time .75 -.25 -.25 -.25
  drug*time .75 -.25 -.25 -.25   0 0 0 0  0 0 0
0  0 0 0 0;
contrast 'M21-M2.' time .75 -.25 -.25 -.25
  drug*time 0 0 0 0 .75 -.25 -.25 -.25   0 0 0 0
0 0 0 0;
lsmeans drug /stderr pdiff e=subj(drug);
lsmeans time drug*time/stderr pdiff;
random subj(drug)/test;
```

Table 14. Results of the usual split-plot model with Two Covariates MNA and NADEV. The model provides Correct Sums of Squares for Between and Within Subject Analyses, But LSMEANS are not estimable.

```
Dependent Variable: BP
                 Sum of       Mean
Source    DF     Squares      Square
Model     38  14845.5811    390.6732
Error     65    451.6400      6.9483
Cor Tot  103  15297.2212

Source        DF      Type I SS F Value Pr > F
DRUG           3     3351.70240  160.79 0.0001
MNA            1      975.27005  140.36 0.0001
SUBJ(DRUG)    21     3786.99870   25.95 0.0001
TIME           3     1368.56731   65.65 0.0001
DRUG*TIME      9     4225.28835   67.57 0.0001
NADEV          1     1137.75432  163.75 0.0001

Source        DF    Type III SS F Value Pr > F
DRUG           3     3138.53948  150.57 0.0001
MNA            0        0.00000       .      .
SUBJ(DRUG)    21     3786.99870   25.95 0.0001
TIME           3      832.06477   39.92 0.0001
DRUG*TIME      9     3628.30637   58.02 0.0001
NADEV          1     1137.75432  163.75 0.0001
Contrast      DF   Contrast SS F Value Pr > F

M11-M12        1      12.89660    1.86 0.1778
M11-M13        1      13.86256    2.00 0.1626
M11-M1.        1      29.48223    4.24 0.0434
M21-M2.        1    1072.12715  154.30 0.0001
```

```
Tests of Hypotheses usng the Type III
MS Subj(Drug) as an error term
Source       DF     Type III SS F Value Pr > F
DRUG          3     3138.53948    5.80 0.0047

                    T for H0: Pr > |T|  Std Err
Parameter      Est   Param=0             Estimate

M11-M12    -2.0788     -1.36  0.1778     1.52587
M11-M13    -2.1723     -1.41  0.1626     1.53796
M11-M1.    -1.9429     -2.06  0.0434     0.94321
M21-M2.    10.0489     12.42  0.0001     0.80897

                          Std Error of
Parameter    Estimate     Estimate
MNA        0.02136752 B  0.03186168

NADEV      0.02941974    0.00229908

    But LSMEANS are not estimable

   Least Squares Means
  DRUG              BP       TIME        BP
                  LSMEAN               LSMEAN
  Control       Non-est       1       Non-est
  RZ33          Non-est       2       Non-est
  XW31          Non-est       3       Non-est
  YA01          Non-est       4       Non-est

  DRUG    TIME      BP  DRUG    TIME      BP
                 LSMEAN                LSMEAN
  Control 1    Non-est  XW31    1    Non-est
  Control 2    Non-est  XW31    2    Non-est
  Control 3    Non-est  XW31    3    Non-est
  Control 4    Non-est  XW31    4    Non-est
  RZ33    1    Non-est  YA01    1    Non-est
  RZ33    2    Non-est  YA01    2    Non-est
  RZ33    3    Non-est  YA01    3    Non-est
  RZ33    4    Non-est  YA01    4    Non-est

Source        Type III Expected Mean Square
DRUG          Var(Error) + 4 Var(SUBJ(DRUG))
              + Q(DRUG,DRUG*TIME)
MNA           0
SUBJ(DRUG)    Var(Error) + 4 Var(SUBJ(DRUG))
TIME          Var(Error) + Q(TIME,DRUG*TIME)
DRUG*TIME     Var(Error) + Q(DRUG*TIME)
NADEV         Var(Error) + Q(NADEV)

Contrast Contrast Expected Mean Square
M11-M12  Var(Error) + Q(TIME,DRUG*TIME)
M11-M13  Var(Error) + Q(TIME,DRUG*TIME)
M11-M1.  Var(Error) + Q(TIME,DRUG*TIME)
M21-M2.  Var(Error) + Q(TIME,DRUG*TIME)
```

Table 15. SAS System Code to manage data so residuals can be analyzed via the REPEATED option.

```
* Construct Data Sets with residuals to obtain
G-G epsilon;
data one; set resid; if time=1;rbp1=resbp;
run;

data two; set resid; if time=2; rbp2=resbp;
run;

data thr; set resid; if time=3; rbp3=resbp;
run;

data for; set resid; if time=4; rbp4=resbp;
run;

data residual; merge one two thr for; by drug
subj;
run;
```

Table 16. SAS System code to preform the repeated measures analysis on the residuals to obtain estimates of the epsilons

```
proc glm; class drug;
model rbp1 rbp2 rbp3 rbp4=drug;
repeated time 4;
Title3 'Repeated Measures Analysis on Residuals
to obtain G-G epsilon';
Title4 'To be used to adjust DFs in above
analyses';
run;
```

Table 17. Repeated Measures Analysis on Residuals to obtain G-G and H-F epsilons to be used to adjust DFs in split-plot analyses.

```
Tests of Hypotheses for Between Subjects
Effects
Source  DF     Type III SS
DRUG     3      0.00000000
Error   22     -0.00000000
Univariate Tests of Hypotheses for Within
Subject Effects

Source: TIME
     DF     Type III SS     Mean Square
      3      0.00000000      0.00000000

Source: TIME*DRUG
     DF     Type III SS     Mean Square
      9      0.00000000      0.00000000

Source: Error(TIME)
     DF     Type III SS     Mean Square
     66    451.64002201      6.84303064

Greenhouse-Geisser Epsilon = 0.6809
      Huynh-Feldt Epsilon = 0.8537
```

Table 18. Computations for Comparing the DRUG means at the same or different TIMES.

```
M1• - M2• = 9.9457  SE2=3.6725     df2=21
M11 - M1• = -1.9429 SE =.9432      df =.6809*65
M21 - M2• = 10.0489 SE =.8090      df =.6809*65

(M11-M1•)-(M21-M2) = -1.9429 - 10.0489 =-11.9918
```

with $SE1^2 = (.9432^2) + (.8090^2) = 1.5441$
based on 55.95 or 56 df = DF1
M11-M21=-11.9918 + 9.9457 = -2.0461

with variance $VAR=1.5441 + 3.6725^2 = 15.0314$
$t_{21}=2.0796$ $t_{56}=2.0032$

$t^* = 2.0718$

Approximate df = 25.96 or 26

t_c = -2.0461 / 3.8770 = -.5277

References

Huynh, H., and Feldt, L. S. 1970. Conditions Under Which Mean Square ratios in repeated measures designs have exact F-distributions. JASA 65: 1582-89.

Milliken, G. A. and Johnson, D. E. 1984. Analysis of Messy Data: Designed Experiments, Vol I. Van Nostrand Reinhold Co. Inc. NY.

SAS Institute Inc. SAS/STAT User's Guide, Release 6.03 Edition. Cary, NC: SAS Institute Inc., 1988.

George A. Milliken
Department Of Statistics
Dickens Hall
Kansas State University
Manhattan, KS 66506
913-532-6883