ANALYSIS OF A REPEATED MEASURES EXPERIMENT WITH MISSING DATA

Richard Allen, Computer Based Systems, Inc.

ABSTRACT

Repeated Measures problems often have missing values which complicate the analysis. Milliken and Johnson have presented a method for handling these type of problems in their book on the Analysis of Messy Data. This paper presents a SAS program that will do the analysis suggested in this book by using PROC GLM and PROC PRINTTO along with several MACRO variables. The output includes the proper F-tests for testing the effects, best estimates and estimated standard errors for functions of the parameters along with the approximate t-statistics, estimated degrees of freedom and p-values for testing the hypothesis that the function is zero. A numerical example is given along with the SAS code that will generate the output.

INTRODUCTION

Repeated measures designs have structures that involve more than one size of experimental unit, similar to split-plot type experiments. When a subject is measured over time at several different time points and time is a factor in the treatment structure, the subject is being split into parts, i.e. time intervals, and the response is measured on each of these parts. This structure leads to two experimental error variances corresponding to the two sizes of experimental units. The larger experimental unit being the subject or the collection of time units, while the smaller unit is the interval of time that the subject is exposed to the treatment.

Repeated measures designs differ from split-plot designs in that the levels of time cannot be randomly assigned to the time intervals. However, if the covariance matrices of the error terms have compound symmetry or, more generally, satisfy the Huynh-Feldt condition then the F-tests of the usual split-plot analysis are valid.

The model for the repeated measures or split-plot type experiment is

$$y_{ijk} = \mu+G_i+\delta_{k(i)}+T_j+GT_{ij}+\epsilon_{ijk}$$

where $\delta_{k(i)}$ represents the error contribution of the k^{th} subject in the i^{th} treatment group and the ϵ_{ijk} represents the error contribution of the j^{th} time period to the k^{th} subject in the i^{th} treatment group. These two error components are assumed to be distributed independently of each other, with the $\delta_{k(i)} \sim N(0,\sigma_\delta^2)$ and the $\epsilon_{ijk} \sim N(0,\sigma_\epsilon^2)$.

OUTLINE OF PROGRAM

The program developed here is based on a split-plot type analysis of repeated-measures data assuming that the covariance conditions are met. It is written to handle the case where some data is missing but will also work for balanced data. It uses two runs of PROC GLM to determine the appropriate F-tests for the effects and the approximate t-tests of the functions of the parameters of interest. For more detail on the formulas used to determine these tests in the missing data case see Milliken and Johnson (p. 403-407).

With a balanced data set the standard error estimates of linear functions of the parameters depend only on a few functions of the two experimental variances. Missing data, however, causes almost all of these linear functions to have standard error estimates that depend on different functions of the two experimental error variances. The first run of PROC GLM is used to determine the expected mean squares for these functions and the model effects using the RANDOM statement. The MODEL statement of the GLM will have terms in it for GROUP, TIME, GROUP*TIME and SUBJ(GROUP). The SUBJ(GROUP) effect is random.

The expected mean squares of interest that will be output by the RANDOM statement have the following form:

GROUP	$\sigma_\epsilon^2 + C*\sigma_\delta^2 + QF(\ )$
SUBJ(GROUP)	$\sigma_\epsilon^2 + Q*\sigma_\delta^2$
ERROR	σ_ϵ^2
CONTRASTS	$\sigma_\epsilon^2 + CC*\sigma_\delta^2 + QF'(\ )$

The coefficients C and Q along with the name and CC for each of the contrasts are chosen in a DATA step from the output into the PROC PRINTTO file for later use.

In the second run of PROC GLM all CONTRAST statements are changed to ESTIMATE statements so that the estimates and standard errors for the functions of the parameters may be obtained. These are again chosen in a DATA step from a PROC PRINTTO file along with information from the analysis of variance results.

All the information gathered above is then used to get the appropriate test for the main GROUP effect as well as corrected standard errors and approximate t-values and degrees of freedom for testing whether the functions of the parameters in the CONTRAST and ESTIMATE statements are zero.

The main GROUP effect is tested using the approximate F-statistic

$$F^* = \frac{MSgroup}{\hat{\sigma}_\varepsilon^2 + C*\hat{\sigma}_\delta^2}$$

where $\hat{\sigma}_\varepsilon^2$ = MSerror and

$\hat{\sigma}_\delta^2$ = (MSsubj(group) - MSerror) / Q.

The approximate degrees of freedom for the denominator of the above F-statistic uses Satterthwaite's approximation. They are

$$DDF = \frac{[C*MSsubj(group) + (Q-C)*MSerror]^2}{\frac{C^2*MSsubj(group)^2}{DFsubj(group)} + \frac{(Q-C)^2*MSerror^2}{DFerror}}$$

The corrected standard errors and approximate significance probabilities for testing the functions of the parameters are calculated from the output of the ESTIMATE statements gathered previously. To get the correct standard error we use the standard error printed to the PROC PRINTTO file as follows:

$$\hat{SE}(function) = SQRT[K*(MSerror+CC*\hat{\sigma}_\delta^2)]$$

where

$$K = SE(printed)^2 / MSerror$$

In order to test whether the function is zero we use a t-statistic where

$$T = \hat{EST}(function) / \hat{SE}(function)$$

An approximate significance probability is then found using this t-value and approximate degrees of freedom found as above using CC in the formula for DDF instead of C.

All the above calculations are then output into an organized and readable form using PUT statemnets in a DATA step.

EXAMPLE

The following example comes from Milliken and Johnson (p. 385). It is a small part of an experiment to determine the effects of a drug on the scores of patients on a test that measures depression. The data set is as follows:

		Score	
Patient	Treatment	Week 1	Week 2
1	Placebo	24	18
2	Placebo	22	--
3	Placebo	25	22
4	Drug	23	--
5	Drug	26	24

The data was entered into the program (see appendix) and the output is as follows:

REPEATED MEASURES ANALYSIS OF VARIANCE

SOURCE	DF	SS	MS	F-VALUE	PR>F
TRT	1	15.56	15.56	6.54	0.0804
WEEK	1	24.08	24.08	96.33	0.0646
TRT*WEEK	1	4.08	4.08	16.33	0.1544
PATIENT(TRT)	3	8.42	2.81	11.22	0.2152
ERROR	1	0.25	0.25		

ADJUSTED DENOMINATOR FOR TESTING TREATMENT EFFECT

3.11 2.38

MULTIPLE COMPARISONS

COMPARISON	EST	SE	T FOR H0:Δ=0	EST DF	PR>|T|
TW11	23.00	1.041	22.10	3.2	0.0001
TW12	17.00	1.155	14.72	4.0	0.0001
TW21	24.67	0.850	29.03	3.2	0.0001
TW22	22.17	0.898	24.70	3.7	0.0001
TRT1	20.00	1.041	19.22	3.2	0.0002
TRT2	23.42	0.837	27.96	3.0	0.0001
TRT DIFF	-3.42	1.336	-2.56	3.1	0.0806
WEEK1 M E	23.83	0.672	35.47	3.2	0.0001
WEEK2 M E	19.58	0.731	26.78	3.9	0.0001
WEEK DIFF	4.25	0.433	9.81	1.0	0.0646
INTERACTION	3.50	0.866	4.04	1.0	0.1544
WEEK1 DIFF	-1.67	1.344	-1.24	3.2	0.2987
WEEK2 DIFF	-5.17	1.462	-3.53	3.9	0.0252

REFERENCE:

Milliken, G.A. and Johnson, D.E. (1984) Analysis of Messy Data , Van Nostrand Reinhold Co., New York, NY

APPENDIX

```
*-------------------------------------*
*---Enter Total Number of Contrasts---*
*-------------------------------------*
%LET CT=13;
*-------------------------------------*
*---------Enter Model Effects---------*
*-------------------------------------*
%LET EFFECT1='TRTMNT';
%LET EFFECT2='WEEK';
%LET INTERACT='TRTMNT*WEEK';
%LET RANDOM='PATIENT(TRTMNT)';
*-------------------------------------*
*-------Enter First Contrast Name-------*
*-------------------------------------*
%LET CTR1='TW11';
*-------------------------------------*

DATA DRUG;
     INPUT PATIENT TRTMNT WEEK SCORE;
     CARDS;
  1  1  1  24
  1  1  2  18
  2  1  1  22
  1  2  1  25
  1  2  2  22
  2  2  1  23
  3  2  1  26
  3  2  2  24

PROC PRINTTO UNIT=20 NEW;
PROC GLM;
   CLASS PATIENT TRTMNT WEEK;
   MODEL SCORE=TRTMNT|WEEK
              PATIENT(TRTMNT) / SS3;
   RANDOM PATIENT(TRTMNT);

*---------------CONTRASTS-------------*;
   CONTRAST 'TW11'
             INTERCEPT 1
             TRTMNT 1 0
             WEEK 1 0
             TRTMNT*WEEK 1 0 0 0;
   CONTRAST 'TW12'
             INTERCEPT 1
             TRTMNT 1 0
             WEEK 0 1
             TRTMNT*WEEK 0 1 0 0;
   CONTRAST 'TW21'
             INTERCEPT 1
             TRTMNT 0 1
             WEEK 1 0
             TRTMNT*WEEK 0 0 1 0;
   CONTRAST 'TW22'
             INTERCEPT 1
             TRTMNT 0 1
             WEEK 0 1
             TRTMNT*WEEK 0 0 0 1;
   CONTRAST 'TRT1'
             INTERCEPT 1
             TRTMNT 1 0
             WEEK .5 .5
             TRTMNT*WEEK .5 .5 0 0
             PATIENT(TRTMNT) .5 .5 0 0 0;
```

```
   CONTRAST 'TRT2'
             INTERCEPT 1
             TRTMNT 0 1
             WEEK .5 .5
             TRTMNT*WEEK 0 0 .5 .5
             PATIENT(TRTMNT) 0 0
                  .33333 .33333 .33333;
   CONTRAST 'TRT DIFF'
             TRTMNT 1 -1;
   CONTRAST 'WEEK1 M E'
             INTERCEPT 1
             TRTMNT .5 .5
             WEEK 1 0
             TRTMNT*WEEK .5 0 .5 0
             PATIENT(TRTMNT) .25 .25
                 .166667 .166667 .166667;
   CONTRAST 'WEEK2 M E'
             INTERCEPT 1
             TRTMNT .5 .5
             WEEK 0 1
             TRTMNT*WEEK 0 .5 0 .5
             PATIENT(TRTMNT) .25 .25
                 .166667 .166667 .166667;
   CONTRAST 'WEEK DIFF'
             WEEK 1 -1
             TRTMNT*WEEK .5 -.5  .5 -.5;
   CONTRAST 'INTERACTION'
             TRTMNT*WEEK 1 -1 -1 1;
   CONTRAST 'WEEK1 DIFF'
             TRTMNT 1 -1
             TRTMNT*WEEK 1 0 -1 0;
   CONTRAST 'WEEK2 DIFF'
             TRTMNT 1 -1
             TRTMNT*WEEK 0 1 0 -1;
*-------------------------------------*;

PROC PRINTTO;

DATA TEST;
   INFILE FT20F001;
   RETAIN C Q;
   ARRAY CTRST (J) $ 20 CTRST1-CTRST&CT;
   ARRAY CC (J) CC1-CC&CT;
   ARRAY ACC (J) $ 20 ACC1-ACC&CT;
   INPUT @2 NAME $20. @;
   IF NAME=&EFFECT1 THEN DO;
     INPUT @2 NAME $20. DUM1 $ DUM2 $ C;
     OUTPUT;
   END;
   IF NAME=&RANDOM THEN DO;
     INPUT @2 NAME $20. DUM1 $ DUM2 $ Q;
     OUTPUT;
   END;
   IF NAME=&CTR1 THEN DO J=1 TO &CT;
      INPUT @2 CTRST $20. DUM1 $ DUM2 $
            ACC $ / ;
      CC=0;
      IF INDEX(ACC,'V')=1 THEN
         CC=INDEX(ACC,'V');
      ELSE CC=ACC;
      IF CC=. THEN CC=0;
      OUTPUT;
   END;
   KEEP C Q CTRST1-CTRST&CT CC1-CC&CT;
```

```
DATA PICK1; SET TEST;
   IF _N_=&CT+2;
PROC PRINTTO UNIT=20 NEW;
PROC GLM DATA=DRUG;
   CLASS PATIENT TRTMNT WEEK;
   MODEL SCORE=TRTMNT|WEEK
                PATIENT(TRTMNT) / SS3;
*----------------ESTIMATES--------------*;
   ESTIMATE 'TW11'
            INTERCEPT 1
            TRTMNT 1 0
            WEEK 1 0
            TRTMNT*WEEK 1 0 0 0;
   ESTIMATE 'TW12'
            INTERCEPT 1
            TRTMNT 1 0
            WEEK 0 1
            TRTMNT*WEEK 0 1 0 0;
   ESTIMATE 'TW21'
            INTERCEPT 1
            TRTMNT 0 1
            WEEK 1 0
            TRTMNT*WEEK 0 0 1 0;
   ESTIMATE 'TW22'
            INTERCEPT 1
            TRTMNT 0 1
            WEEK 0 1
            TRTMNT*WEEK 0 0 0 1;
   ESTIMATE 'TRT1'
            INTERCEPT 1
            TRTMNT 1 0
            WEEK .5 .5
            TRTMNT*WEEK .5 .5 0 0
            PATIENT(TRTMNT) .5 .5 0 0 0;
   ESTIMATE 'TRT2'
            INTERCEPT 1
            TRTMNT 0 1
            WEEK .5 .5
            TRTMNT*WEEK 0 0 .5 .5
            PATIENT(TRTMNT) 0 0
                 .33333 .33333 .33333;
   ESTIMATE 'TRT DIFF'
            TRTMNT 1 -1;
   ESTIMATE 'WEEK1 M E'
            INTERCEPT 1
            TRTMNT .5 .5
            WEEK 1 0
            TRTMNT*WEEK .5 0 .5 0
            PATIENT(TRTMNT) .25 .25
               .166667 .166667 .166667;
   ESTIMATE 'WEEK2 M E'
            INTERCEPT 1
            TRTMNT .5 .5
            WEEK 0 1
            TRTMNT*WEEK 0 .5 0 .5
            PATIENT(TRTMNT) .25 .25
               .166667 .166667 .166667;
   ESTIMATE 'WEEK DIFF'
            WEEK 1 -1
            TRTMNT*WEEK .5 -.5  .5 -.5;
   ESTIMATE 'INTERACTION'
            TRTMNT*WEEK 1 -1 -1 1;
   ESTIMATE 'WEEK1 DIFF'
            TRTMNT 1 -1
            TRTMNT*WEEK 1 0 -1 0;
   ESTIMATE 'WEEK2 DIFF'
            TRTMNT 1 -1
            TRTMNT*WEEK 0 1 0 -1;
*-------------------------------------*;
```

```
PROC PRINTTO;

DATA TEST;
   INFILE FT20F001;
   RETAIN DFE SSE MSE DFG SSG DFR SSR
          FR PR DFGR SSGR FGR PGR
          DFAG SSAG FAG PAG;
   ARRAY EST (J)  EST1-EST&CT;
   ARRAY SE (J) SE1-SE&CT;
   INPUT @2 NAME $20. @;
   IF NAME='ERROR' THEN DO;
      INPUT @2 NAME $20. DFE SSE MSE;
      OUTPUT;
   END;
   IF NAME=&EFFECT1 THEN DO;
      INPUT @2 NAME $20. DFG SSG;
      OUTPUT;
   END;
   IF NAME=&EFFECT2 THEN DO;
      INPUT @2 NAME $20. DFR SSR FR PR;
      OUTPUT;
   END;
   IF NAME=&INTERACT THEN DO;
      INPUT @2 NAME $20. DFGR SSGR FGR
            PGR;
      OUTPUT;
   END;
   IF NAME=&RANDOM THEN DO;
      INPUT @2 NAME $20. DFAG SSAG FAG
            PAG;
      OUTPUT;
   END;
   IF NAME=&CTR1 THEN DO J=1 TO &CT;
      INPUT @2 NAME $20. EST DUM1 DUM2
            SE;
      OUTPUT;
   END;
   KEEP DFE SSE MSE DFG SSG DFR SSR FR
        PR DFGR SSGR FGR PGR DFAG SSAG
        FAG PAG EST1-EST&CT SE1-SE&CT;
DATA PICK2; SET TEST;
   IF _N_=&CT+5;

DATA INFO; MERGE PICK1 PICK2;
   FILE PRINT;
   ARRAY CTRST (J) $ CTRST1-CTRST&CT;
   ARRAY CC (J) CC1-CC&CT;
   ARRAY EST (J)  EST1-EST&CT;
   ARRAY SE (J) SE1-SE&CT;
   ARRAY DF (J) DF1-DF&CT;
   ARRAY K (J) K1-K&CT;
   ARRAY T (J) T1-T&CT;
   ARRAY P (J) P1-P&CT;
   SIG1=SSAG/DFAG;
   SIGD=(SIG1-MSE)/Q;
   FG=(SSG/DFG)/(MSE+C*SIGD);
   DDFG=(C*SIG1+(Q-C)*MSE)**2/
        (((C*SIG1)**2/DFAG)
       +(((Q-C)*MSE)**2/DFE));
   PG=1-FPROB(FG,DFG,DDFG);
   IF PG<0.0001 THEN PG=0.0001;
   MSG=SSG/DFG;
   MSR=SSR/DFR;
   MSGR=SSGR/DFGR;
   MSAG=SSAG/DFAG;
   AMSAG=MSG/FG;
```

```
DO J=1 TO &CT;
   K=SE**2/MSE;
   SE=SQRT(K*(MSE+CC*SIGD));
   T=EST/SE;
   DF=(CC*SIG1+(Q-CC)*MSE)**2/
      (((CC*SIG1)**2/DFAG)
     +(((Q-CC)*MSE)**2/DFE));
   P=2*(1-TPROB(ABS(T),DF));
   IF P<0.0001 THEN P=0.0001;
END;

PUT @22 'REPEATED MEASURES ANALYSIS
OF VARIANCE' ///;
PUT @6 'SOURCE' @25 'DF'
    @31 'SUM OF SQUARES'
    @48 'MEAN SQUARE'
    @62 'F-VALUE' @ 72 'PR>F' /;
PUT @6 &EFFECT1 @23 DFG 5.
    @31 SSG 14.4 @48 MSG 11.4
    @62 FG 7.2 @72 PG 6.4;
PUT @6 &EFFECT2 @23 DFR 5.
    @31 SSR 14.4 @48 MSR 11.4
    @62 FR 7.2 @72 PR 6.4;
PUT @6 &INTERACT @23 DFGR 5.
    @31 SSGR 14.4 @48 MSGR 11.4
    @62 FGR 7.2 @72 PGR 6.4;
PUT @6 &RANDOM @23 DFAG 5.
    @31 SSAG 14.4 @48 MSAG 11.4
    @62 FAG 7.2 @72 PAG 6.4;
PUT @6 'ERROR' @23 DFE 5.
    @31 SSE 14.4 @48 MSE 11.4 / ;
PUT @6 'ADJUSTED DENOMINATOR FOR
TESTING TREATMENT EFFECT' / ;
PUT @22 DDFG 6.2 @48 AMSAG 11.4 /// ;
PUT @31 'MULTIPLE COMPARISONS' /;
PUT @54 'T FOR' @63 'ESTIMATED';
PUT @11 'COMPARISON' @29 'ESTIMATE'
    @42 'STDERR' @52 'H0:DIFF=0'
    @66 'DF' @75 'PR>|T|' / ;
DO J=1 TO &CT;
   PUT @7 CTRST $20. @29 EST 8.2
       @41 SE 7.3 @53 T 6.2
       @65 DF 4.1
       @75 P 6.4;
END;
```

For additional information contact:

Richard R. Allen
Computer Based Systems, Inc.
HFT-020
N.C.T.R. Circle
Jefferson, AR 72079-9502
501-541-4630