A Program for Outputting P Values and Confidence Intervals
In an Analysis of Variance Without the Use of the GLM or PRINTTO Procedures

Susan Foley Anton and Joseph J. Andary
Bristol-Myers Squibb Company

Introduction

Proc GLM is commonly used to repeatedly perform ANOVA (e.g., at each week of a study). The number of p values generated from Proc GLM in such applications is large, especially when pairwise comparisons are performed. The tabular display of these p values is desirable in statistical presentations. Because Proc GLM in versions 5 and 6.06 of the SAS® system does not allow the user to output these p values, the statistician is frequently required to hand-copy them from the Proc GLM output to a report. Such copying is tedious, inefficient and promotes errors. To avoid copying p values, we have previously used Proc PRINTTO to copy them from the Proc GLM output to a report. Such copying is tedious, inefficient and promotes errors. To avoid copying p values, we have previously used Proc PRINTTO to transcribe them from Proc GLM output to report tables. However, this is not an ideal way to automate the process, since SAS output formats change with different versions of the SAS® system, and programs incorporating Proc PRINTTO require revision with each new version. We have therefore developed a program which outputs p values (as well as standard errors, parameter estimates and other desired statistics) to a SAS® dataset while avoiding the use of Proc PRINTTO.

The purpose of this paper is to present the computer algorithm which is used to obtain those p values and statistics and store them in a SAS® dataset. The resulting dataset can subsequently be used to generate report-ready tables and graphs.

ANOVA

Method

The procedure which produces statistics from a 2-Way ANOVA without the use of PROC GLM is described below. The application used in demonstrating the procedure is a randomized clinical trial in which two drug treatments (variable TREAT) were compared at seven research sites (variable SITE) over a 6-week period. Efficacy assessments were performed at baseline (variable BASELINE) and at each week of the 6-week study (variable RESPONSE). The model for the treatment comparison examines the change from baseline in each treatment at each site, and the adjusted mean for each treatment over all sites. The mean of the dependent variable for each treatment and study site is obtained first and saved in an OUT= dataset. These means are then used in a second MEANS procedure run by treatment, to produce the mean of the dependent variable in each treatment group adjusted for unequal sample sizes at the sites.

1. First, a dummy-variable regression model is created so that a SAS® regression procedure can be used to perform a 2-Way ANOVA. The regression procedure of choice for this algorithm is Proc RSQUARE, since it can be used to output the Root Mean Square Error (RMSE) and the Degrees of Freedom for Error (EOF) to a SAS® output dataset (such as the OUTEST= dataset called MSE in Appendix I, step 1). Six dummy variables are used to represent the seven research sites, one dummy variable is used to represent the two treatments, and six dummy variables are used to represent the seven treatment by site interactions. The RMSE and the EDF are needed to generate p values (which are not output by version 5 or 6.06 of the SAS® system), and the standard errors of the adjusted means (which are needed for the calculation of confidence intervals and are not output by version 5 of the SAS® system).

2. Second, the MEANS procedure is used to obtain the mean change from baseline in each treatment at each site, and the adjusted mean for each treatment over all sites. The mean of the dependent variable for each treatment and study site is obtained first and saved in an OUT= dataset. These means are then used in a second MEANS procedure run by treatment, to produce the mean of the dependent variable in each treatment group adjusted for unequal sample sizes at the sites.

3. Third, the standard error of the mean change from baseline in the dependent variable in each treatment and study site is equal to the RMSE divided by the square root of the number of subjects:

\[ SE = \text{RMSE} \times \text{SORT (nSample)} \]

where nSample = # subjects in study i receiving treatment j

The standard error of the mean change from baseline in the dependent variable in each treatment over all study sites is equal to:

\[ SE = \text{SORT (average variance across all sites)} + \text{SORT (# sites)} \]

\[ \text{Variance} = (\text{Standard Error})^2 \]

Applying Steps 1-3 to the example data from a clinical trial of two drug treatments at seven sites, we obtain the following adjusted means and their standard errors:

<table>
<thead>
<tr>
<th>Treatment</th>
<th>Site</th>
<th>Adjusted Mean</th>
<th>Standard Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Treatment #1</td>
<td>1</td>
<td>-8.889</td>
<td>2.90710</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>2</td>
<td>-7.500</td>
<td>2.51762</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>3</td>
<td>-7.500</td>
<td>2.75792</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>4</td>
<td>-6.818</td>
<td>2.62967</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>5</td>
<td>-7.700</td>
<td>2.75792</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>6</td>
<td>-5.750</td>
<td>3.08345</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>7</td>
<td>-12.538</td>
<td>2.41885</td>
</tr>
<tr>
<td>Treatment #1</td>
<td>All Sites</td>
<td>-8.099</td>
<td>1.03288</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>1</td>
<td>-6.273</td>
<td>2.62967</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>2</td>
<td>-8.983</td>
<td>2.51762</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>3</td>
<td>-10.700</td>
<td>2.75792</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>4</td>
<td>-8.636</td>
<td>2.75792</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>5</td>
<td>-8.636</td>
<td>2.62967</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>6</td>
<td>-9.125</td>
<td>3.06345</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>7</td>
<td>-0.125</td>
<td>2.51762</td>
</tr>
<tr>
<td>Treatment #2</td>
<td>All Sites</td>
<td>-8.309</td>
<td>1.02246</td>
</tr>
</tbody>
</table>
4. Fourth, the adjusted mean of the dependent variable in one treatment group is then subtracted from the adjusted mean of the dependent variable in the other treatment group, yielding the adjusted mean difference in change from baseline between treatments.

The standard error of this adjusted mean difference in change from baseline between treatments is obtained by taking the square root of the sum of the variances of treatment #1 and treatment #2 which were computed in step 3:

\[
SE = \sqrt{\text{variance for treatment 1} + \text{variance for treatment 2}}
\]

The associated \( T \) statistic is equal to the adjusted mean difference in change from baseline between treatments divided by the standard error of that adjusted mean difference.

5. The probability of the \( T \) distribution is calculated next using the `TPROB` and `ABS` function. The \( p \) value for a two-sided test is equal to:

\[
P = (1 - \text{TPROB(ABS(T),EDf)}) \times 2
\]

with degrees of freedom equal to the total number of subjects in treatments 1 and 2 minus 2. \( 1\)-\( \text{TPROB} \) gives the probability that the value of the mean difference falls in the upper tail of the Student's \( T \) distribution.

6. The critical value for the 95% confidence limits around the mean difference between treatment #1 and treatment #2 is equal to:

\[
\text{TCRIT}._{95} = \text{TINV(0.025,DF)}
\]

where \( 0.025 = (1 - 0.95) \times 2 \), and \( \text{TINV} \) is the increase of \( \text{TPROB} \), the value \( z \) exceeded with probability \( = (1 - 0.025) \) by a student's \( T \) random variable.

7. The adjusted mean differences in change from baseline and their 95% confidence limits can then be transposed and plotted using the `GGRAPH` procedure to summarize the treatment effect in each study and over all studies:

8. Alternatively, calculation of the critical values can be eliminated, and the dataset containing the mean difference in change from baseline, \( T \) statistic and \( p \) values can be tabulated along with other variables of interest using the `PRINT` procedure:

**ANALYSIS OF EFFICACY SCORE OVER ALL SITES**

**TREATMENT 1**

<table>
<thead>
<tr>
<th>Week</th>
<th>Size</th>
<th>Baseline</th>
<th>Response Baseline</th>
<th>Standard Error</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>68</td>
<td>25.5</td>
<td>21.4</td>
<td>-4.0</td>
<td>0.0</td>
</tr>
<tr>
<td>2</td>
<td>73</td>
<td>25.4</td>
<td>19.8</td>
<td>-5.6</td>
<td>1.1</td>
</tr>
<tr>
<td>3</td>
<td>73</td>
<td>25.4</td>
<td>18.5</td>
<td>-6.9</td>
<td>1.2</td>
</tr>
<tr>
<td>4</td>
<td>73</td>
<td>25.4</td>
<td>17.3</td>
<td>-8.1</td>
<td>1.3</td>
</tr>
<tr>
<td>5</td>
<td>73</td>
<td>25.4</td>
<td>15.8</td>
<td>-8.6</td>
<td>1.3</td>
</tr>
<tr>
<td>6</td>
<td>73</td>
<td>25.4</td>
<td>17.3</td>
<td>-8.1</td>
<td>1.5</td>
</tr>
</tbody>
</table>

**TREATMENT 2**

<table>
<thead>
<tr>
<th>Week</th>
<th>Size</th>
<th>Baseline</th>
<th>Response Baseline</th>
<th>Standard Error</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>72</td>
<td>25.2</td>
<td>20.9</td>
<td>-4.3</td>
<td>0.9</td>
</tr>
<tr>
<td>2</td>
<td>74</td>
<td>25.4</td>
<td>20.2</td>
<td>-5.2</td>
<td>1.1</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
<td>25.4</td>
<td>18.6</td>
<td>-6.8</td>
<td>1.2</td>
</tr>
<tr>
<td>4</td>
<td>74</td>
<td>25.4</td>
<td>18.4</td>
<td>-6.9</td>
<td>1.3</td>
</tr>
<tr>
<td>5</td>
<td>74</td>
<td>25.4</td>
<td>18.2</td>
<td>-7.1</td>
<td>1.3</td>
</tr>
<tr>
<td>6</td>
<td>74</td>
<td>25.4</td>
<td>17.0</td>
<td>-8.3</td>
<td>1.5</td>
</tr>
</tbody>
</table>

The SAS® code for this ANOVA algorithm is presented in Appendix I.

**1-Way ANCOVA**

**Method**

The procedure which produces statistics from a 1-Way ANCOVA is analogous to the algorithm for the 2-Way ANOVA. The application used in demonstrating the procedure below is a study in which three treatments were compared at a single research site over a 6 week period. Efficacy assessments were performed at baseline (the covariate, \( X \)) and after 6 weeks of double-blind treatment (the response variable, \( Y \)). Since there were 3 treatment categories (the independent variable, \( T \)), pairwise comparisons between the treatments examined 2 categories at a time. The model for each pairwise comparison, \( Y = X T \), controls for baseline level of the dependent variable in assessing the effect of treatment on efficacy.

The steps of the algorithm are as follows:

1. First, a dummy-variable regression model is created so that the `RSQUARE` procedure can be used to output the Mean Square for Error (MSE) in an `OUTEST=` data set called MSE. The baseline value of the response variable is the covariate, the continuous variable, \( X \). The following two dummy variables are used to represent the three treatments:

<table>
<thead>
<tr>
<th>Dummy Variable</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>( T_1 )</td>
<td>1</td>
</tr>
<tr>
<td>( T_2 )</td>
<td>1</td>
</tr>
<tr>
<td>( T_3 )</td>
<td>0</td>
</tr>
</tbody>
</table>

- \( T_1 \) if observation is from treatment 1
- \( T_2 \) if observation is from treatment 2
- \( T_3 \) if otherwise

The SAS code for this ANOVA algorithm is presented in Appendix I.
This first regression model is:

\[ Y = X (T_1) (T_2) \]

2. Next, a second dummy-variable regression model is created for a 1-Way ANOVA so the RSQUARE procedure can be used to output the Sum of Squares for Error (SSE) of the baseline efficacy assessment, X, in an OUTEST= data set called SSE. This SSE is needed to compute the standard errors of the adjusted means. This second regression model is:

\[ X = (T_1) (T_2) \]

3. Next, the MEANS procedure is used to obtain the mean of the covariate, X, over all treatment groups. This quantity is needed to compute the value of the response variable, Y, adjusted for X, for each subject.

4. The following regression equation is then used to calculate Y adjusted for X for each subject:

\[ Y_{\text{adjusted}} = \text{Intercept} + b_1 (T_1) + b_2 (T_2) + b_3 (X_{\text{overall}}) \]

where \( b_1, b_2, \) and \( b_3 \) are regression coefficients obtained from Proc RSQUARE in step 1, and \( X_{\text{overall}} \) is the overall mean of the covariate computed in step 3.

5. The MEANS procedure is then used to obtain the mean of \( Y_{\text{adjusted}} \) within each treatment group.

6. The MEANS procedure is used a third time, this time to obtain the mean of the covariate, X, within each treatment group. This quantity is needed to compute the standard error of the adjusted mean difference between any two treatments.

7. The standard error of the adjusted mean of the response variable Y in each treatment group (from step 5) is equal to the following quantity:

\[ SE = \sqrt{\text{MSE} \times (1/X_N + (X_{MN} - \text{MNB&VAR})^2/SSE)} \]

where

- MSE = the Mean Square for Error of Y from step 1
- \( X_N \) = the # of subjects in treatment group 1
- \( X_{MN} \) = the mean of X in treatment group 1
- \( \text{MNB&VAR} \) = the mean of X in all treatment groups combined
- SSE = the Sum of Squares for Error of X from step 2

8. The adjusted mean of Y in one treatment group is then subtracted from the adjusted mean of Y in the other treatment group, yielding the difference in adjusted mean between the two treatments.

The standard error of the adjusted mean difference between the two treatments is equal to

\[ SE = \sqrt{\text{MSE} \times (1/\#1 + 1/\#2 + X_{DIF}^2/SSE)} \]

where

- MSE = the Mean Square for Error of Y from step 1
- \#1 = the # of subjects in treatment group 1
- \#2 = the # of subjects in treatment group 2
- \( X_{DIF} \) = the mean of X in treatment group 1 - (minus) the means of X in treatment group 2
- SSE = the Sum of Squares for Error of X from step 2

9. Calculation of the associated T statistic, the probability of the T distribution, and the critical value for the 95% confidence limits is the same as in steps 4, 5, 6 of the ANOVA procedure, where:

\[ T_{\text{difference}} = \text{mean difference} + SE_{\text{difference}} \]

\[ P = (1 - \text{TPROB}(|T_{\text{difference}}|))/2 \]

and \( TCRIT_{.05} = TINV(0.025,DF) \).

10. The adjusted mean difference in response between any two treatments and the associated statistics can be obtained and saved in a SAS dataset by applying steps 1 through 9 of the ANCOVA algorithm above to the data for the two treatments involved in the comparison.

11. Finally, the datasets containing the statistics from the pairwise treatment contrasts (contrast12, contrast13, and contrast23) can be merged and plotted using the GAXPLOT procedure to summarize differences in treatment outcomes.

12. Alternatively, calculation of the critical values can be eliminated, and the datasets containing the mean difference in change from baseline, T statistic and P values can be merged and tabulated along with other variables of interest using the PRINT procedure.

The SAS code for the ANCOVA algorithm is presented in Appendix II.

DESCRIPTION OF ENHANCEMENTS

The algorithms presented above can be adapted for N-Way ANOVA, N-Way ANOVA with multiple treatment groups, and ANCOVA and ANCOVA of ranks. Both the ANOVA and the ANCOVA algorithm are appropriate for balanced and unbalanced designs. The data used in the 2-Way ANOVA example was from a study with unequal sample sizes in each cell of the design (i.e., an unbalanced design).

For N-Way ANOVA, the algorithm presented for 2-Way ANOVA can be applied as follows, where all standard errors, T statistics and P values are computed as in the 2-Way ANOVA described above.

- Dummy variables must be created for all \( N \) independent variables, and for all interactions among them, resulting in the creation of a fully specified linear model. The RSQUARE procedure is then applied to this model as in step 1 of the ANOVA described above.

- A dataset is created for each treatment by \( N \) variable cell, which contains the mean change from baseline and the standard error of that mean, and has one observation for each value of site.

- Those datasets are merged using a MERGE statement, and the adjusted mean difference in change from baseline between each pair of treatments for each site and over all sites is obtained by subtraction.

For N-Way ANCOVA with multiple treatment groups, the algorithm for 2-Way ANOVA should be amended as described for N-Way ANOVA, while multiple treatment groups can be compared serially in a pairwise fashion as in the ANCOVA example.

1132
For the Analysis of Variance and Analysis of Covariance of ranked scores, use of the above algorithm is virtually identical to their use in the analysis of variance of raw scores. For example, consider the 2-Way ANOVA described in detail above, in which efficacy assessments were performed at baseline and at each week of the 6-week study. The RANK procedure can be used to rank the change in those efficacy scores from baseline and output the ranks to a SAS® dataset. The ranked change scores can then be substituted for the raw change scores for all the steps of the ANOVA procedure.

REFERENCES


Author Contact

Susan Foley Anton, Ph.D.
5 Research Parkway, Dept. 703
P.O. Box 5100
Wallingford, CT 06492-7660
(203) 294-7525

APPENDIX 1

********************************************************************
* ANOVA: SAS PROGRAM WHICH CALCULATES ADJUSTED *
* MEANS AND ASSOCIATED STATISTICS FROM A TWO-WAY *
* ANOVA.
********************************************************************

%MACRO ANALYZEVER (DSET);

**DWAR = EVAR DATA STATS; SET ADDDET:DWAR;
**EVAR IS THE RESPONSE VARIABLE CHOOSEN FOR THIS EXECUTION OF THE MACRO;

** 1) CREATE DUMMY VARS AND USE PROC RSQURARE TO *
** OBTAIN MEAN SQUARE ERROR - NEEDED TO COMPUTE *
** STANDARD ERROR OF THE ADJUSTED MEAN FOR EACH *
** TREAT X STUDY CELL. *

********************************************************************
** CREATE 1 DUMMY VARIABLE TO REPRESENT THE 2 TREATMENTS **;
**IF TREAT="TREAT1" THEN RX=1; ELSE RX=0;
**CREATE 6 DUMMY VARIABLES TO REPRESENT THE 7 STUDY CENTERS **;
ARRAY S SITE1-SITE6;
ARRAY I INTER1-INTER6;
N=1;
DO OVER S;
IF SITE=S TREAT="S1"; ELSE SITE=S; N=N+1; END;
********************************************************************
*** CREATE 6 DUMMY VARIABLES TO REPRESENT THE 7 INTERACTIONS ***;
DO OVER I;
I = SITE*RX;
END;
PROC RSQUARE DATA=SOUT=REST; VAR MND&VAR;
MODEL MND&VAR = SITE1 SITE2 SITE3 SITE4 SITE5 SITE6 RX
INTER1 INTER2 INTER3 INTER4 INTER5 INTER6 / INCLUDE=13 MSE RRMSE SSE;
DATA MSE (KEEP=RMSE_EFD); SET MSE;
PROC SORT DATA=S; BY TREAT SITE;
********************************************************************
** 2) USE PROC MEANS TO OBTAIN THE ADJUSTED MEAN **;
** OF THE CHANGE FROM BASELINE FOR EACH TREATMENT **;
** TREATMENT. *

********************************************************************
** OBTAIN THE MEAN FOR EACH TREATMENT AND SITE **;
** PROC MEANS DATA=STATS NOPRINT MEAN N STDERR **;
** BY TREATMENT SITE; VAR MND&VAR; OUTPUT OUT=MEAN **;
** OBTAIN THE ADJUSTED MEAN FOR EACH TREATMENT OVER *
** ALL SITES ****;
** PROC MEANS DATA=SITE MEAN NOPRINT MEAN SUM N;
** BY TREATMENT SITE; VAR MND&VAR NSAMPLE; OUTPUT OUT=TRTMEAN **;
** MEAN = NSAMPLE NSAMP **;
** SUM = SUM NSAMPLE **;
** DATA TRTMEAN (KEEP=TREAT SITE NSAMPLE MND&VAR); SET TRTMEAN **;
** SITE = 8;
** NSAMPLE = SUM NSAMPLE NSAMP **;
** PROC SORT DATA=SITE MEAN; BY TREAT SITE **;
** PROC SORT DATA=TRTMEAN; BY TREAT SITE **;
** DATA ALLMEAN; MERGE SITE MEAN TRTMEAN; BY TREAT SITE **;

** 3A) OBTAIN THE STANDARD ERRORS FOR THE MEAN *
** CHANGE FROM BASELINE FOR EACH TREATMENT X *
** STUDY CELL. **;

********************************************************************
** DATA ALLMEAN **;
** IF N=1 THEN SET MSE; SET ALLMEAN;
** SE = RMSE / (SORT(NSAMPLE)); **;
** IF SITE = 8 THEN SE=.; SE OVERALL MUST BE CALCULATED SEPARATELY; RUN; **;

********************************************************************
** 3B) OBTAIN THE STANDARD ERRORS FOR THE MEAN A_VAR *
** CHANGE FROM BASELINE FOR EACH TREATMENT OVER *
** ALL SITES. **;
** VARIANCE TREATMENT (OVERALL) = (SORT AVERAGE **;
** VARIANCE OVER ALL SITES) **;

********************************************************************
** DATA TEMP SE **;
** SET ALLMEAN; BY TREAT **;
** IF FIRST.TREAT THEN DO; *(SET ACCUMULATED **;
** TOTAL TO 0 EACH TIME A NEW VALUE OF TREAT IS **;
** ENCOUNTERED); **;
** SUM_VAR = 0;
** N SITES = 0;
** END; **;
** IF SITE NE 8; **;
** SUM_VAR + (SE**2); *(ADD UP VARIANCE **;
** OVER ALL SITES); **;
** N_SITES + 1; *(ADD UP # SITES FOR **;
** EACH TREATMENT); **;
** IF LAST. TREAT; **;
** *(IMPLIED: THEN OUTPUT); **;
** SE_TREAT = SQRT(SUM_VAR/N_SITES) / SQRT(N_SITES) **;
** KEEP SUM_VAR N_SITES SE_TREAT EDF **;
** DATA TEMP SE (KEEP=TREAT SITE SE_EFD); SET TEMP SE **;
** SITE=8; SE = SE_TREAT **;
** DATA ALLMEAN; MERGE ALLMEAN TEMP.SE; BY TREAT SITE **;

1133
**APPENDIX 2**

* ANCOVA: SAS PROGRAM WHICH RUNS A 1-WAY ANALYSIS
  + OF COVARIANCE AND COMPUTES THE GROUP MEANS
  + ADJUSTED FOR THE COVARIATE:
    * \( Y_{\text{adj}} = Y - b(X - \bar{X}) \)
    * \( i \)
  + where \( b \) is the pooled within-class regression coefficient
  + -Dependent Variable = \( Y = \text{BVAR} \)
  + -Covariate = \( X = \text{BVAR} \)
  + -Independent Variable = \( T = \text{Treatment} \)

**MACRO ANALAYVAR, DSET;**

1) CREATE DUMMY VARS AND USE PROC RSQUARE TO
  + OBTAIN MEAN SQUARE ERROR - NEEDED TO COMPUTE
  + STANDARD ERRORS OF THE ADJUSTED MEAN FOR EACH
  + TREATMENT.
  + **MODEL:**
    * \( Y = X (T1)(T2) \)
  + **END ANALYZE;**

**** THERE IS 1 STUDY SITE ****;

**** CREATE 2 DUMMY VARIABLES TO REPRESENT THE 3 TREATMENTS ****;

***** CREATE VARS EQUAL TO DUMMY VARS, SINCE PROC

***** OBTAIN THE OVERALL MEAN OF \( \text{BVAR} \) ****;

PROC MEANS DATA=STATS NOPRINT MEAN N; VAR \text{BVAR};
MODEL \text{BVAR} = T1_OTHER T2_OTHER /INCLUDE=5 STOP=4;
PROC SORT DATA=STATS; BY TREAT;

2) USE PROC REGRESS AGAIN, THIS TIME TO OBTAIN
  + SUM OF SQUARES FOR ERROR OF \( X (\text{BVAR}) \).
  + **MODEL:**
    * \( X = (T1)(T2) \)
  + **END ANALYZE;**

*** THERE IS 1 STUDY SITE ****;

*** CREATE 2 DUMMY VARIABLES TO REPRESENT THE 3 TREATMENTS ****;

**** CREATE VARS EQUAL TO DUMMY VARS, SINCE PROC

***** OBTAIN THE OVERALL MEAN OF \( \text{BVAR} \) ****;

PROC MEANS DATA=STATS NOPRINT MEAN N; VAR \text{BVAR};
MODEL \text{BVAR} = T1_OTHER T2_OTHER /INCLUDE=5 STOP=4;
PROC SORT DATA=STATS; BY TREAT;

3) USE PROC MEANS TO OBTAIN THE MEAN OF THE
  + COVARIATE OVER ALL TREATMENT GROUPS.
  + **END ANALYZE;**

DATA SSE_X (KEEP = _SSE_); SET SSE_X;
OUTPUT OUT=MEAN_ALL MEAN = MN8_BVAR N = N8_BVAR;

4) USE WITHIN-CLASS REGRESSION COEFFICIENT TO
  + COMPUTE \( y \) ADJUSTED FOR \( x \) FOR EACH SUBJECT.
A&VAR = INTERCEP + T1*OTHER + T10*OTHER + T2*OTHER + MNB&VAR;
RUN;

DATA MSE (KEEP = MSE EDF MNB&VAR); SET MSE;

***************************************************
* 5) USE PROC MEANS TO OBTAIN THE "ADJUSTED MEAN" *
* OF Y FOR EACH TREATMENT *
***************************************************
PROC SORT DATA=STATS; BY TREAT;
PROC MEANS DATA=STATS NOPRINT MEAN N STDERR; BY TREAT;
VAR A&VAR;
OUTPUT OUT= Y MEAN MEAN = MNA&VAR N = NSAMPlE;

***************************************************
* 6) USE PROC MEANS TO OBTAIN THE MEAN OF X *
* FOR EACH TREATMENT *
***************************************************
PROC SORT DATA=STATS; BY TREAT;
PROC MEANS DATA=STATS NOPRINT MEAN N STDERR; BY TREAT;
VAR B&VAR;
OUTPUT OUT= X MEAN MEAN = X MN N = X N;

***************************************************
* 7) OBTAIN THE STANDARD ERRORS FOR THE ADJUSTED *
* MEAN OF Y FOR EACH TREATMENT *
***************************************************
DATA MSE; MERGE MSE SSE X;
DATA Y STDERR (KEEP=TREAT Y SE X MN EDF NSE _SSE_); 
IF _N_=I THEN SET MSE; SET X_MEAN;
Y SE = SQRT( NSE * (I/X N + I/X NN * NNB&VAR)**2 /_SSE_);
RUN;

DATA Y MEAN (KEEP = TREAT Y MN Y N Y SE X MN EDF MSE_); 
MERGE Y_MEAN Y STDERR;

***************************************************
* 8) OBTAIN THE MEAN DIFFERENCE IN THE ADJUSTED *
* MEAN OF Y BETWEEN &TRY1 AND &TRTZ. *
***************************************************
DATA DIFMEAN (KEEP = CONTRAST MEAN DIF N DIF SE DIF T_DIF EDF_); 
MERGE GHIMEAN GlOHEAN PLAHEAN;
CONTRAST "&TRT1._&TRT2 11 ;
MEAN DIF = &TRT1. YMN - &TRTZ. YMN;
N-DIF = &TRT1.-N + &TRTZ._N;
X-DIF = &TRT1.-XMN - &TRTZ. XMN;
SE-DIF = SQRT (-MSE * (I/&TRT1. N + I/&TRTZ._N + _OF**2/ SSE »); 
T_OIF = MEAN DIF / SE DIF;

***************************************************
* 9A) CALCULATE THE PROBABILITY FOR THE T DISTRIBUTION. (2-SIDED TEST) *
***************************************************
DATA &DATA (KEEP = CONTRAST MEAN DIF N_DIF SE_DIF T_DIF P_DIF);
SET DIFMEAN;
P_DIF = (1-TPROB(ABS(T_DIF),_EDF_»*Z;

***************************************************
* 9B) CALCULATE THE CRITICAL VALUE FOR THE CONFIDENCE LIMITS AROUND THE MEAN DIFFERENCE BETWEEN &TRT1 AND &TRTZ. *
***************************************************
DATA &ClDATA (KEEP TCRIT1 TCRITZ);
SET &DATA;
DF = N_DIF-Z;
TCRIT1-= TINV(.025,DF);
Cl l = TCRIT1 * SE_DIF;
Cl:U = MEAN DIF + X1;

%MEND ANALA;
%MACRO ANALB(VAR, TRT1, TRTZ, DATA, CLDAlA);

***************************************************
* 10A) MERGE DATASETS CONTAINING ADJUSTED MEAN *
* DIFFERENCE, SE, T AND P *
***************************************************
DATA DATA123; MERGE &DATAl &DATAZ &DATAl; BY CONTRAST;

***************************************************
* 10B) MERGE DATASETS CONTAINING 2-WAY CONTRASTS *
***************************************************
DATA DATA123; MERGE &ClDATAl &ClDATAZ &clDATA3; BY CONTRAST;
%MEND ANALC;
%MACRO ANAlC(DATAI, DATAZ, DATA3, ClDATAI, ClDATAZ, CLDATA3, ANALNAME, WEEKNUM);

1135