A SAS \textsuperscript{R} PROGRAM FOR ANOVA WHEN THE INPUT DATA ARE GROUP MEANS

Aloha E. Bryson, University of North Carolina at Chapel Hill

ABSTRACT
A four module computer program executed in SAS under the PROC MATRIX procedure was written to handle univariate between groups analysis of variance situations in which one might want to input group means and their variance-covariance matrix rather than input the raw data. Estimates of primary and secondary parameters and tests of hypotheses such as main effects and interactions may be obtained by either ordinary least squares or weighted least squares.

INTRODUCTION
Analysis of variance (ANOVA) computer programs generally require that the user input data for each experimental unit. The general linear univariate model, represented by the model equation

\[ E(Y) = XB \]

is then fit. The ordinary least squares estimates (OLS) of the primary parameters, \( \hat{\beta} \) and \( \hat{\sigma}^2 \) are estimated by

\[ \hat{\beta} = (X'X)^{-1}X'Y \]

and

\[ \hat{\sigma}^2 = (Y'Y - B'X'XB)/(n-k) \]

where
- \( Y \) = \( p \times 1 \) vector of observations on the response variable for \( n \) experimental units.
- \( X \) = \( p \times k \) design matrix. (For simplicity, \( X \) is assumed to be full rank).
- \( n \) = total sample size.
- \( k \) = rank of the design matrix.

Analyses based upon secondary parameters of the form

\[ \beta = CB - G \]

where \( C \) and \( G \) are appropriately defined matrices of contrasts, are generally the main interest in ANOVA. A test of these parameters is provided by

\[ F(m, n-m) = ((CB)'(C(X'X)^{-1}C')^{-1}(CB))/(m \sigma^2) \]

where \( m \) is the rank of the \( C \) matrix.

There are several situations in which it might be desirable to have a program in which the user only inputted the means of the groups, the number of observations per group, and the variance-covariance matrix of the means:

1. For data sets in which there are large numbers of observations and several models one wishes to fit, it could be more computationally efficient to input the means of the groups, the number of observations per group, and the variance-covariance matrix of the means, than to input the raw data to an ANOVA program.

2. In studies which involve national probability sampling, the variance-covariance matrix of means must often be estimated by special purpose programs because of the complex sampling schemes often employed. The estimate of error variance obtained from existing ANOVA programs assumes simple random sampling and thus is not appropriate for studies which have complex sampling schemes. An ANOVA program which would allow one to input the appropriate variance-covariance matrix or to include a design effect would be highly desirable.

3. A weighted least squares (WLS) ANOVA could easily be performed with a program that allowed input of a variance-covariance matrix of means. Presently, there is no easy way to perform a WLS with most ANOVA programs. A program that handles the three situations above was written in the SAS procedure PROC MATRIX. The program contains four modules which serve as "subroutines", each designed for a specific type of computation or analysis. Each subroutine may be accessed via LINKs in PROC MATRIX. The SAS code is presented in Appendix A.

The first module, OLSANOVA, computes the OLS estimates of \( \beta \), their standard errors and \( t \) statistics, and \( \hat{\sigma}^2 \) of group means on the response variable, a column vector of the variances of the group means, a column vector of sample sizes that correspond to the group means, an essence matrix, and a design effect. (If the sample is a simple random sample or if the vector of variances has taken into account the sampling scheme, the design effect is simply set equal to 1.00.) The equation for estimating \( \beta \) becomes

\[ \beta = (X'DX)^{-1}X'D\hat{y} \]

where \( X^* = qk \times (k < q) \) essence matrix.
After the model has been fit by OLSANOVA (assuming an OLS fit is desired), the second module, OLSTEST, may be used to perform additional analyses, such as tests of main effects and interactions. OLSTEST is the module one uses to obtain OLS estimates of the secondary parameters along with their tests of significance. The user must supply the constant matrix $C$. The equation for the $E$ statistic becomes

$$F(m, n-m) = (CB)'(C(X' DX')^{-1} C')^{-1}(CB)/(m\sigma^2)$$

where $B$, $C$, $X'$, $D$, and $m$ are as defined in the OLSANOVA module.

The third module, WLSANOVA, computes the WLS estimates of $B$ along with their standard errors and $t$ statistics. The equation for estimating $B$ becomes

$$B = (X'V^{-1}X')^{-1}X'V^{-1}Y'$$

where $X'$, $D$, and $Y'$ are as defined for OLS and $V$ is the variance-covariance matrix of the group means. It is the responsibility of the user to justify the use of the variance-covariance matrix (which often must be estimated) as the weighting matrix.

The fourth module, WLSTEST, may also be used to perform additional analyses, such as tests of main effects and interactions. WLSTEST is the module one uses to obtain WLS estimates of the secondary parameters along with their tests of significance. The equation for the $E$ statistic becomes

$$F(m, n-m) = (CB)'(C(X' V^{-1}X')^{-1} C')^{-1}(CB)/(mV),$$

where $C$, $B$, and $m$ are as defined with OLS and $V$ is the variance-covariance matrix of the group means.

The SAS code necessary to run the one-way ANOVA (4 levels of a between factor) for the sleep deprivation data from Kirk (1982, pg. 140) is presented in Appendix B. The SAS output is presented in Appendix C.

REFERENCES


Author’s Address
Aloha E. Bryson
3003 Darden Road
Greensboro, North Carolina 27407

SAS is the registered trademark of SAS Institute, Inc., Cary, N.C., USA.
**Appendix A**

```plaintext
OPTIONS NOSOURCE;
MACRO INITIAL
GO TO FINISH;

OLSANOVA:
  COLSX=NCOL(X);
  ROWSN=NROW(N);
  V=DIAG(VCV);
  DIAG_N=DIAG(N);
  SUMN=DIAG_N(+,+);
  N_MINUS1=DIAG_N-I(ROWSN);

---ESTIMATE BETA AND DETERMINE SS AND DF DEN---;
  BETA=INV(X'*DIAG_N*X)*X'*DIAG_N*YMEAN;
  SSMATRIX=DIAG_N*N_MINUS1*V;
  SSW=SSMATRIX(+,+);
  DFDEN=(SUMN#/DEFF)-COLSX;
  MSE=SSW#/DFDEN;

---DETERMINE STD_ERR, T, & P_VAL FOR BETA---;
  VCV_BETA=INV(X'*DIAG_N*X)*MSE;
  DIAG_VCV=DIAG(VCV_BETA);
  STD_ER1=SQRT(DIAG_VCV);
  STD_ERR=VECDIAG(STD_ER1);
  T_VALUE=BETA#/STD_ERR;
  P_VALUE=2*(1-PROBT(ABS(T_VALUE),DFDEN));

---PRINT RESULTS FROM OLSANOVA MODULE---;
  NOTE PAGE '-----------------------------';
  NOTE 'OLSANOVA MODULE BEGINS';
  NOTE '-----------------------------';
  NOTE VECTOR OF MEANS INPUTTED;
  PRINT YMEAN;
  NOTE VARIANCE COVARIANCE MATRIX INPUTTED;
  PRINT VCV;
  NOTE ESSENCE MATRIX;
  PRINT X;
  NOTE ESTIMATED PARAMETERS AND THEIR STANDARD ERRORS;
  TABLE=BETA!STD_ERR!T_VALUE!P_VALUE;
  TABNAME='BETA' 'STD_ERR' 'T_VALUE' '2 TAIL P_VAL';
  PRINT TABLE COLNAME=TABNAME;
  NOTE ESTIMATED SIGMA SQUARED;
  PRINT MSE;
  RETURN;

OLSTEST:

---DETERMINE DF NUM, F, & P_VAL FOR THETA---;
  DFNUN=NROW(C);
  C_BETA=C*BETA;
  V_CBETA=(C*VCV_BETA*C');
  INV_VCV=INV(V_CBETA);
  F=((C_BETA)'*INV_VCV+C_BETA)#/(DFNUM);
  P_VALUE=1-PROBF(F,DFNUM,DFDEN);

---DETERMINE T, AND P_VAL FOR EACH ELEMENT OF THETA---;
  DIAG_V=DIAG(V_CBETA);
  STD_ER=SQRT(DIAG_V);
  STD_ER=VECDIAG(STD_ER);
  T=C_BETA#/STD_ER;
  P=2*(1-PROBT(ABS(T),DFDEN));
```

636
**--PRINT RESULTS FROM ANALYSIS OF THETA--**;
  NOTE PAGE '----------------------------------------';
  NOTE 'IOLSTEST MODULE BEGINS I';
  NOTE '----------------------------------------';
  NOTE CONTRAST MATRIX;
  PRINT C;
  NOTE C BETA;
  PRINT C_BETA;
  NOTE F AND P VALUE FOR CONTRAST;
  NOTE CONTRAST MATRIX,
  TABLE2='F' 'DFNUM' 'DFDEN' 'P VALUE';
  PRINT TABLE2 COLNAME=TABNAME2;
  NOTE T AND P VALUE FOR EACH ELEMENT OF THETA;
  TABLE3=C_BETA 'STD_ERR' 'T' '2 TAIL P_VAL';
  PRINT TABLE3 COLNAME=TABNAME3;

RETURN;

WLSANOVA:
  COLSX=NCOL(X);
  ROWSN=NROW(N);
  DIAG_N=DIAG(N);
  SUMN=DIAG_N(+,+);
  BETA=CAPACIWSANOVA MODULBETA=
  DFDEN=ISUMN-DFEFF)-COLSX;
  VCV_BETA=CAPACIWSANOVA MODULVCV_BETA=
  STD_ERR1=CAPACIWSANOVA MODULSTD_ERR1=
  T_VALUE=BETAITSTD_ERR;
  P_VALUE=2*(1-PROBT(ABS(T_VALUE),DFDEN));

**--PRINT RESULTS FROM WLSANOVA MODULE--**;
  NOTE PAGE '----------------------------------------';
  NOTE 'I WLSANOVA MODULE BEGINS I';
  NOTE '-----------------------------------------';
  NOTE VECTOR OF MEANS INPUTTED;
  PRINT YMEAN;
  NOTE VARIANCE COVARIANCE MATRIX INPUTTED;
  PRINT VCV;
  NOTE ESSENCE MATRIX;
  PRINT Xi;
  NOTE ESTIMATED PARAMETERS AND THEIR STANDARD ERRORS;
  TABLE4=BETA 'STD_ERR' 'T_VALUE' '2 TAIL P_VAL';
  PRINT TABLE4 COLNAME=TABNAME4;

RETURN;

WLSTEST:
  **DETERMINE DF NUM, F, & P VALUE FOR CONTRAST**;
  DFPNUM=NROW(C);
  C_BETA=C*BETA;
  F=((C_BETA)**INV(CVCV_BETA+C)**C_BETA)/DFNUM;
  P_VAL=1-PROBF(F,DFNUM,DFDEN);

**--PRINT RESULTS FROM WLSTEST MODULE--**;
  NOTE PAGE '----------------------------------------';
  NOTE 'I WLSTEST MODULE BEGINS I';
  NOTE '----------------------------------------';
  NOTE CONTRAST MATRIX;
  PRINT C;
  NOTE C_BETA;

637
PRINT C_BETA;
NOTE F, DF NUM, DF DEN, & P_VALUE FOR CONTRAST;
TABNAMES='F' 'DF NUM' 'DF DEN' 'P VALUE';
TABLES=DFNUM DFDEN P VAL;
PRINT TABLES COLNAME=TABNAMES;
RETURN;
FINISH;
%
OPTIONS SOURCE;
/*

Appendix B

//job card
//**PW= // EXEC SAS
//SYSIN DD DSN=UNC.PSL.F795C.BRYSON.ANOVA,DISP=SHR
// DD *
PROC MATRIX
INITIAL;
N=(0 8 8 8)';
X=I(4);
VCV=(.2768 .1071 .1340 .2143)';
YMEAN=(2.75 3.5 6.25 9)';
DEFF=1;
LINK OLSANOVA;
C=1 -1 0 0/
1 0 -1 0/
1 0 0 -1;
LINK OLSTEST;

Appendix C

SLEEP DEPRIVATION DATA FROM KIPK PG 140
---------------------------------------------
| OLSANOVA MODULE BEGINS |-------------------------------|
---------------------------------------------
VECTOR OF MEANS INPUTTED

<table>
<thead>
<tr>
<th>YMEAN</th>
<th>COL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>2.75</td>
</tr>
<tr>
<td>ROW2</td>
<td>3.5</td>
</tr>
<tr>
<td>ROW3</td>
<td>6.25</td>
</tr>
<tr>
<td>ROW4</td>
<td>9</td>
</tr>
</tbody>
</table>

VARIANCE COVARIANCE MATRIX INPUTTED

<table>
<thead>
<tr>
<th>VCV</th>
<th>COL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>0.2768</td>
</tr>
<tr>
<td>ROW2</td>
<td>0.1071</td>
</tr>
<tr>
<td>ROW3</td>
<td>0.134</td>
</tr>
<tr>
<td>ROW4</td>
<td>0.2143</td>
</tr>
</tbody>
</table>

638
### ESSENCE MATRIX

<table>
<thead>
<tr>
<th></th>
<th>COL1</th>
<th>COL2</th>
<th>COL3</th>
<th>COL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROW2</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROW3</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ROW4</td>
<td>0</td>
<td>0</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

### ESTIMATED PARAMETERS AND THEIR STANDARD ERRORS

<table>
<thead>
<tr>
<th>TABLE</th>
<th>BETA</th>
<th>STD_ERR</th>
<th>T_VALUE</th>
<th>2 TAIL P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>2.75</td>
<td>0.42754</td>
<td>6.47754</td>
<td>5.0476E-07</td>
</tr>
<tr>
<td>ROW2</td>
<td>1.05</td>
<td>0.42754</td>
<td>2.4185</td>
<td>6.5475E-09</td>
</tr>
<tr>
<td>ROW3</td>
<td>6.25</td>
<td>0.42754</td>
<td>14.6091</td>
<td>1.2601E-14</td>
</tr>
<tr>
<td>ROW4</td>
<td>-9</td>
<td>0.42754</td>
<td>21.0334</td>
<td>0</td>
</tr>
</tbody>
</table>

### ESTIMATED SIGMA SQUARED

<table>
<thead>
<tr>
<th>MSF</th>
<th>COL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>1.4644</td>
</tr>
</tbody>
</table>

### SLEEP DEPRIVATION DATA FROM KIPK PG 140

### CONTRAST MATRIX

<table>
<thead>
<tr>
<th>C</th>
<th>COL1</th>
<th>COL2</th>
<th>COL3</th>
<th>COL4</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>1</td>
<td>-1</td>
<td>0</td>
<td>0</td>
</tr>
<tr>
<td>ROW2</td>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
</tr>
<tr>
<td>ROW3</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td>-1</td>
</tr>
</tbody>
</table>

### C BETA

<table>
<thead>
<tr>
<th>C_BETA</th>
<th>COL1</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>-0.75</td>
</tr>
<tr>
<td>ROW2</td>
<td>-1.5</td>
</tr>
<tr>
<td>ROW3</td>
<td>6.25</td>
</tr>
</tbody>
</table>

### F AND P VALUE FOR CONTRAST

<table>
<thead>
<tr>
<th>TABLE</th>
<th>F</th>
<th>DFNOM</th>
<th>DFDEN</th>
<th>P_VALUE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>44.273</td>
<td>3</td>
<td>28</td>
<td>9.3295E-11</td>
</tr>
</tbody>
</table>

### T AND P VALUE FOR EACH ELEMENT OF THEIA

<table>
<thead>
<tr>
<th>TABLE</th>
<th>C_BETA</th>
<th>STD_ERR</th>
<th>T_VALUE</th>
<th>2 TAIL P</th>
</tr>
</thead>
<tbody>
<tr>
<td>ROW1</td>
<td>-0.75</td>
<td>0.605062</td>
<td>-1.23954</td>
<td>0.225434</td>
</tr>
<tr>
<td>ROW2</td>
<td>-1.5</td>
<td>0.605062</td>
<td>-0.78452</td>
<td>0.44059279</td>
</tr>
<tr>
<td>ROW3</td>
<td>6.25</td>
<td>0.605062</td>
<td>10.3295</td>
<td>4.6874E-11</td>
</tr>
</tbody>
</table>

639