ANNOTATED SAS OUTPUT (ASO)

Michael P. Meredith

BU-814-M in the Biometrics Unit series,
Cornell University, Ithaca, NY 14853

October 1983

Introduction

The present form of the Annotated SAS Output (ASO) has evolved from the original project aimed at illustrating common statistical methods using SAS (BU-664-M, BU-705-M). The primary goal of these annotated outputs has been to provide students with direction in using a statistical package to analyze data at the level of Statistics 602. This expanded version of the ASO should be of use to students as well as others outside the classroom.

Over the past five years there have been many people who have contributed to the ASO. Most of these people have been either students, lab instructors or undergraduate assistants involved with the conduct of Statistics 602. This list includes Anna Angelos, Suzanne Aref, Valerie Arneson, Calvin Berry, Foster Cady, Margaret Cecce, Patricia Firey, Laura Gnazzo, Jon Maatta, Charles McCulloch, Patricia Nolan, Norma Phalen, and Walter Piegorsch.

Description

A total of 15 data sets are employed to exemplify the application of SAS in the analysis of data. The data sets are derived from actual designed or observational experiments.

The name of the data set, which appears in the Table of Contents, and a description of the type of analysis that it serves to illustrate are provided below. The first 10 data sets are from _Analyzing Experimental Data by Regression_ (Allen and Cady, 1982), while references are given for the sources of the remaining five.

1) _Arsenic Data_ illustrate straight-line regression.
2) _Firefly Data_ illustrate multiple linear regression with two predictor variables.

3) Soymilk Data illustrate polynomial regression and lack-of-fit from a completely randomized design.

4) Electricity Load Data illustrate a model sequence when several straight lines may need to be fitted.

5) Potato Leafhopper Data illustrate an analysis of treatment means via a complete set of orthogonal contrasts. The experiment design is a one-way classification in a completely randomized design with equal replication.

6) Lymphocyte Data illustrate an analysis of a 2 x 2 factorial experiment laid out in a completely randomized design with equal replication.

7) Fat Digestibility Data illustrate an analysis of a 2 x 2 factorial experiment laid out in a randomized complete block design.

8) Protein Nutrition Data illustrate an analysis of a one-way completely randomized design with unequal replication. A set of nonorthogonal contrasts are presented as well as a complete orthogonal set.

9) Swamp pH Data illustrate the analysis of cell means in a 2 x 3 factorial with unequal replication. The first analysis follows that presented in Allen and Cady (1982). The second is an analysis of unweighted means which is discussed in Snedecor and Cochran (1980). The Least Square Means of SAS are illustrated.

10) Soybean Physiological Data illustrate a covariate analysis. Several ways are shown to estimate treatment means adjusted and unadjusted for the covariate. The "classical" ANCOVA table is given as is a test for homogeneity of slopes.

11) Potato Scab Data illustrate an analysis of a 2 x 3 factorial experiment where one treatment factor is qualitative with two levels and the other treatment factor is quantitative with three unequally spaced levels. Two equivalent analyses are presented: the first is a model sequence approach (see Allen and Cady, Unit 19) comparing quadratic regression curves; the second is an

analysis of cell means wherein appropriate single degree-of-freedom orthogonal polynomial contrasts and interaction contrasts are estimated. These data are taken from a larger set presented in Cochran and Cox (1957, p. 97).

12) Alcohol-Drug Data illustrate the analysis of a split-unit experiment. The cell means model is fitted as described in Allen and Cady (1982, p. 280). The approach presented allows the simple effects to be estimated more easily than in a default model specification. The whole unit analysis proceeds by analyzing the sums and the split-unit analysis is accomplished after removing the whole-unit variability. Such an approach can result in substantial savings of computing dollars and a strengthening of the understanding of such experiments. The usual approach is presented for completeness.

13) Urea Synthesis Data illustrate the analysis of a repeated measures experiment with two treatment groups and two repeated measurements upon each experimental unit. The pertinent hypothesis tests may be reduced to three independent two-sample t-tests based upon the within-subject sums or differences. Such an approach tends to render the analysis quite transparent. The usual ANOVA table is also presented for completeness. These data are taken from Brogan and Kutner (1980).

14) Hemoglobin Data illustrate the analysis of a two-period crossover experiment. The calculations for this design are exactly the same as those of the preceding Urea Synthesis Data. The distinction between the two designs lies in the method of treatment allocation. These data are taken from Grizzle (1965).

15) Milk Yield Data illustrate the analysis of a three-period three-treatment crossover design which is balanced for first order carry-over effects. The layout is in repeated latin squares. Treatment means adjusted and unadjusted for carry-over effects are given as are their standard errors. These data are taken from Cochran and Cox (1957, p. 135).

Table of Contents

---

* See the description for references.

ARSENIC DATA – Units 4, 5 and 6; Annotated Computer Output, p. 346 (BMDP)

```
    TITLE ARSENIC DATA; ← prints a heading on each page
    DATA ARSENIC;
    INPUT DIST ARSENIC;
    DEV= DIST-16.1;
    XO= 1;
    CARDS;
 2 3.19
 4 3.26
 8 1.82
10 1.02
12 1.85
15 2.05
21 1.34
23 0.79
30 0.66
36 0.30
```

These statements create a data set called ARSENIC, which has variables DIST, ARSENIC, DEV, and XO.
ARSENIC is the observed arsenic level,
DIST is the distance from the arsenic source,
DEV is the difference between DIST and the mean distance, 16.1,
XO is a column of ones.

```
  Ⓐ PROC PRINT;
        VAR XO DEV XO DIST; PROC PRINT prints the X matrices for parts Ⓑ and Ⓒ below.

    PROC REG SIMPLE USSCP;
     Ⓑ  MODEL ARSENIC= XO DEV/NOINT P SEQB SS1 SS2 CLM;
              OUTPUT OUT= NEW1 PREDICTED= YHAT1 RESIDUAL= RESID1
                 L95M= LOWER U95M= UPPER;
     Ⓒ  MODEL ARSENIC= XO DIST/NOINT P SEQB SS1 SS2;
     Ⓓ  MODEL ARSENIC= DIST/P SEQB SS1 SS2;
     Ⓔ  MODEL ARSENIC= DIST/NOINT P SEQB SS1 SS2;
     Ⓕ  MODEL ARSENIC= XO/NOINT;
              OUTPUT OUT= NEW2 PREDICTED= YBAR;
```

PROC REG is a regression procedure.
Various models and options are illustrated in parts Ⓑ through Ⓕ.
The OUTPUT statements save the predicted and residual values
in data sets called NEW1 and NEW2.

```
 Ⓖ PROC UNIVARIATE DATA= NEW1 PLOT FREQ NORMAL;
        VAR RESID1;
```

PROC UNIVARIATE gives descriptive statistics for the residuals obtained in Ⓑ. DATA = NEW1 specifies which data set the variable to be used is contained in. If the data set is not specified SAS uses the last data set created.

```
   PROC PLOT DATA= NEW1;
    Ⓗ  PLOT RESID1*YHAT1/VREF= 0; ← VREF = 0 draws a horizontal line  at RESID = 0.
    Ⓘ  PLOT ARSENIC*DIST= '*' YHAT*DIST= 'P'
             UPPER*DIST= 'U' LOWER*DIST= 'L'/OVERLAY;
```

Plot Ⓗ is a residual plot for models Ⓑ, Ⓒ and Ⓓ.
Plot Ⓘ prints observed, predicted, and confidence band values on the same set of axes using the symbols *, P, U, and L.

```
Ⓙ DATA DECOMP; MERGE NEW1 NEW2;
  DATA DECOMP; SET DECOMP;
  RES1= YHAT1-YBAR;
```

DATA DECOMP is a new data set which contains the mean of the ARSENIC values, YBAR (from part Ⓕ); the residuals left when the mean is subtracted from the predicted values, RES1; the ARSENIC values, and the residuals from Ⓑ. The MERGE statement combines data sets NEW1 and NEW2, and the SET statement allows the creation of the new variable RES1.

```
  PROC PRINT DATA= DECOMP; ← PROC PRINT prints the data decomposition.
       VAR ARSENIC YBAR RES1 RESID1;
  TITLE DATA DECOMPOSITION FOR ARSENIC DATA;
```

The X matrix
for the model
ARSENIC = X0 DEV
in part Ⓑ

The X matrix
for the model
ARSENIC = X0 DIST
in part Ⓒ

```
OBS     X0       DEV     X0      DIST

  1      1     -14.1      1        2
  2      1     -12.1      1        4
  3      1      -8.1      1        8
  4      1      -6.1      1       10
  5      1      -4.1      1       12
  6      1      -1.1      1       15
  7      1       4.9      1       21
  8      1       6.9      1       23
  9      1      13.9      1       30
 10      1      19.9      1       36
```

The options SIMPLE and USSCP (Uncorrected Sums of Squares and Cross Products) in the PROC REG statement produce this output:

```
                              ARSENIC DATA

                          DESCRIPTIVE STATISTICS

                     SUM          MEAN  UNCORRECTED        VARIAN STD DEVIATI

ARSENIC       16.28000000  1.62800000  35.7268000  1.02477333  1.01231089

X0            10.00000000  1.00000000  10.0000000  0.00000000  0.00000000

DEV            0.00000000  0.000000001126.9000000125.21111111  11.18977708

DIST         161.00000000 16.10000000003719.0000000125.21111111  11.18977708



                    SUMS OF SQUARES AND CROSSPRODUCTS

 SSCP           ARSENIC             X0             DEV          DIST        INTERCEP

 ARSENIC        35.7268          16.28         -88.068        174.04           16.28
 X0               16.28             10     2.13163E-14           161              10
 DEV            -88.068    2.13163E-14          1126.9        1126.9     2.13163E-14
 DIST            174.04            161          1126.9          3719             161
 INTERCEP         16.28             10     2.13163E-14           161              10
```

The circled numbers should be exactly zero.

SAS supplies an intercept which can produce the same sums of squares and cross products as X0 which was created as part of the data set.

MEAN AND SLOPE MODEL

Ⓑ MODEL ARSENIC = XO DEV/NOINT P SEQB SS1 SS2 CLM;

```
SEQUENTIAL PARAMETER ESTIMATES <- The option SEQB produces the sequential b's.

XO           1.628<- Coefficient for the mean model (see p.35).
DEV          1.628 -.078151<- Coefficients for the mean and slope model
                                    (see pp. 39 and 45).
DEP VARIABLE: ARSENIC
ANOVA:
                     SUM OF            MEAN
SOURCE     DF       SQUARES          SQUARE        F VALUE        PROB>F

MODEL       2     33.386414     16.693207         57.061         0.0001
ERROR       8      2.340386      0.292548= s^2
U TOTAL    10     35.726800   — U TOTAL stands for UNCORRECTED TOTAL.    ROOT MSE is the standard deviation.

    ROOT MSE      0.540877= s    R-SQUARE          0.9345
    DEP MEAN      1.628000= ȳ    ADJ R-SQ          0.9263  <- when NOINT is used, the reported R^2 is incorrect.
    C.V.          33.22342                                    See ASO, p.5 for the correct R^2.
   (Coefficient of variation)
NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.
```

The option NOINT instructs SAS not to supply an intercept since XO has been included in the model. When XO is in the model and the NOINT option is used, the model SS includes the SS for the mean.

```
                 PARTIAL
                PARAMETER       STANDARD    T FOR H0:
VARIABLE   DF    ESTIMATE          ERROR  PARAMETER=0    PROB > |T|     TYPE I SS

XO          1    1.628000= ȳ    0.171040          9.518       0.0001     26.503840= R(0)
DEV         1   -0.078151=      0.016112         -4.850       0.0013      6.882574= R(DEV|0)
                           slope
```

The options SS1 and SS2 in the MODEL statement produce the TYPE I SS (the sequential SS) and the TYPE II SS (the partial SS).

```
VARIABLE   DF    TYPE II SS

XO          1     26.503840= R(0|DEV)
DEV         1      6.882574= R(DEV|0)


                PREDICT   STD ERR LOWER95% UPPER95%
  OBS    ACTUAL    VALUE  PREDICT     MEAN     MEAN RESIDUAL

    1     3.190    2.730 0.284371    2.074    3.386 0.460075
    2     3.260    2.574 0.259352    1.976    3.172 0.686377
    3     1.820    2.261 0.215145    1.765    2.757 -.441020
    4     1.020    2.105 0.197269    1.650    2.560   -1.085
    5     1.850    1.948 0.183354    1.526    2.371 -.098418
    6     2.050    1.714 0.171956    1.317    2.111 0.336034
    7     1.340    1.245 0.188382 0.810647    1.679 0.094938
    8  0.790000    1.089 0.203997 0.618339    1.559 -.298760
    9  0.660000 0.541706 0.281803 -.108140    1.192 0.118294
   10  0.300000 0.072801 0.363402 -.765212 0.910815 0.227199

SUM OF RESIDUALS              7.07767E-15
SUM OF SQUARED RESIDUALS         2.340386
```

Should be exactly zero.

The CLM option prints the upper and lower confidence limits for a 95% confidence interval on μ at each observed X.

The P option in the MODEL statement prints each observed value, predicted value, and residual.

Note that using DEV instead of DIST as the X variable produces sequential b's that are the same as the partials. See Ⓒ.

INTERCEPT AND SLOPE MODEL

Ⓒ MODEL ARSENIC = XO DIST/NOINT P SEQB SS1 SS2;

```
SEQUENTIAL PARAMETER ESTIMATES ← The sequential coefficients are the diagonal elements of the SEQB output.

X0          1.628     ← Coefficents of the mean and slope model.
DIST      2.88623 -.078151 ← Coefficients of the intercept and slope model (see p.42).

DEP VARIABLE: ARSENIC

                   SUM OF          MEAN
SOURCE     DF     SQUARES        SQUARE        F VALUE       PROB>F

MODEL       2    33.386414     16.693207       57.061        0.0001 } The ANOVA is the same as in Ⓑ MODEL ARSENIC = XO DEV.
ERROR       8     2.340386      0.292548 = s²
U TOTAL    10    35.726800

    ROOT MSE   s=0.540877      R-SQUARE       0.9345
    DEP MEAN   ȳ=1.628000      ADJ R-SQ       0.9263
    C.V.         33.22342

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED. ← Same problem with R² as in Ⓑ.  See ASO, p.5.

                 PARTIAL
                PARAMETER      STANDARD    T FOR H0:                    (SEQUENTIAL)
VARIABLE   DF    ESTIMATE         ERROR  PARAMETER=0    PROB > |T|      TYPE I SS
                       intercept
X0          1    2.886226 ↗    0.310720        9.289        0.0001      26.503840 = R(0)
DIST        1   -0.078151 ↘    0.016112       -4.850        0.0013       6.882574 = R(DIST|0)
                       slope                                            printed by
                                                                        SS1 option
               (PARTIAL)
VARIABLE   DF   TYPE II SS

X0          1    25.241773 = R(0|DIST)
DIST        1     6.882574 = R(DIST|0)
                printed by SS2 option

                 PREDICT
 OBS   ACTUAL      VALUE RESIDUAL

   1    3.190      2.730 0.460075
   2    3.260      2.574 0.686377
   3    1.820      2.261 -.441020
   4    1.020      2.105   -1.085
   5    1.850      1.948 -.098418    Printed by P option,
   6    2.050      1.714 0.336034    same values as in Ⓑ.
   7    1.340      1.245 0.094938
   8 0.790000      1.089 -.298760
   9 0.660000 0.541706 0.118294
  10 0.300000 0.072801 0.227199

SUM OF RESIDUALS               6.40954E-15 ← should be exactly zero
SUM OF SQUARED RESIDUALS          2.340386 ← Error SS, see ANOVA above
```

The parameter estimates are different from those in Ⓑ, but the TYPE I SS are the same.

PROC REG does not compute F tests for the TYPE I or TYPE II SS. The t tests of the parameter estimates are equivalent to F tests of the TYPE II SS. To perform the independent F tests of the TYPE I SS, form the ratio

$$\frac{(\text{TYPE I SS})/\text{df}}{(\text{Residual SS})/(\text{Residual df})} = F^{\text{df}}_{\text{Resid df}}$$

INTERCEPT AND SLOPE MODEL

Ⓓ MODEL ARSENIC = DIST/P SEQB SS1 SS2;

[The NOINT option is not used, so SAS supplies an intercept. The model SS does not include the SS for the mean ($n\bar{y}^2$).]

```
SEQUENTIAL PARAMETER ESTIMATES

INTERCEP     1.628
DIST       2.88623 -.078151 } same sequentials as Ⓒ

DEP VARIABLE: ARSENIC

                     SUM OF          MEAN
SOURCE     DF       SQUARES        SQUARE       F VALUE       PROB>F

MODEL       1      6.882574      6.882574        23.526       0.0004
ERROR       8      2.340386      0.292548
C TOTAL     9      9.222960

    ROOT MSE   s = 0.540877      R-SQUARE        0.7462
    DEP MEAN   ȳ = 1.628000      ADJ R-SQ        0.7145
    C.V.           33.22342
```

ERROR mean square: 0.292548 $= s^2$, same as Ⓑ and Ⓒ

C TOTAL → Corrected total SS

R-SQUARE 0.7462 ← Correct R^2. Compare with calculated R^2 below.

ADJ R-SQ 0.7145 ← The adjusted R^2 is "adjusted" for the number of variables in the model.

```
                   PARTIAL
                  PARAMETER      STANDARD   T FOR H0:                   (SEQUENTIAL)
VARIABLE   DF      ESTIMATE         ERROR  PARAMETER=0    PROB > |T|    TYPE I SS

INTERCEP    1      2.886226      0.310720        9.289        0.0001    26.503840
DIST        1     -0.078151      0.016112       -4.850        0.0013     6.882574
                                                                        ↳same as Ⓒ

                  (PARTIAL)
VARIABLE   DF    TYPE II SS

INTERCEP    1     25.241773
DIST        1      6.682574
                   ↳same as Ⓒ

                  PREDICT
  OBS    ACTUAL     VALUE RESIDUAL

    1     3.190     2.730 0.460075
    2     3.260     2.574 0.686377
    3     1.820     2.261 -.441020
    4     1.020     2.105   -1.085
    5     1.850     1.948 -.098418      same as Ⓑ and Ⓒ
    6     2.050     1.714 0.336034
    7     1.340     1.245 0.094938
    8  0.790000     1.089 -.298760
    9  0.660000  0.541706 0.118294
   10  0.300000  0.072801 0.227199

SUM OF RESIDUALS                8.40994E-15
SUM OF SQUARED RESIDUALS           2.340386
```

When NOINT is used, the reported R^2 is incorrect. $R^2 = \dfrac{\text{Corrected Model SS}}{\text{Corrected Total SS}}$ where both model and total SS have been corrected for the mean. NOINT causes the SS for the mean to be included in both numerator and denominator. To calculate the correct R^2, subtract the SS for the mean ($n\bar{y}^2$) from the MODEL SS and TOTAL SS.

Ex.: Model Ⓒ $n\bar{y}^2 = 26.503840$

$$R^2 = \frac{33.386414 - n\bar{y}^2}{35.726800 - n\bar{y}^2} = 0.7462$$

(compare to above R^2)

$$R^2 = \frac{\text{Corrected Model SS}}{\text{Corrected Total SS}} = 1 - \frac{\text{RESIDUAL SS}}{\text{Corrected Total SS}}$$

$$R^2_{adj} = 1 - \left[\frac{(\text{Resid SS}) * \left(\frac{n-1}{n-p}\right)}{\text{Corrected Total SS}}\right]$$

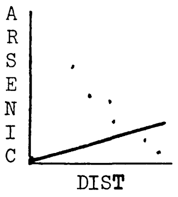
where n is the number of observations and p is the number of variables fitted, including the intercept or XO.

```
SEQUENTIAL PARAMETER ESTIMATES

DIST      .0467975 = slope of line through origin (see p. 53)

DEP VARIABLE: ARSENIC
                        ┌includes SS for the mean
                 SUM OF │          MEAN
SOURCE     DF   SQUARES │         SQUARE       F VALUE       PROB>F

MODEL       1   8.144641│       8.144641         2.658       0.1375
ERROR       9  27.582159┘       3.064684 = s^2, compare to previous models
U TOTAL    10  35.726800

    ROOT MSE    1.750624        R-SQUARE        0.2280
    DEP MEAN    1.628000        ADJ R-SQ        0.2280
    C.V.        107.5322

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

                PARAMETER      STANDARD   T FOR H0:                  (SEQUENTIAL)
VARIABLE   DF    ESTIMATE         ERROR  PARAMETER=0    PROB > |T|    TYPE I SS

DIST        1    0.046798      0.028706        1.630        0.1375     8.144641


              (PARTIAL)
VARIABLE   DF   TYPE II SS

DIST        1    8.144641


              PREDICT
  OBS    ACTUAL     VALUE RESIDUAL

   1     3.190 0.093595     3.096
   2     3.260 0.187190     3.073
   3     1.820 0.374380     1.446
   4     1.020 0.467975 0.552025
   5     1.850 0.561570     1.288
   6     2.050 0.701963     1.348
   7     1.340 0.982748 0.357252
   8 0.790000     1.076 -.286343
   9 0.660000     1.404 -.743926
  10 0.300000     1.685    -1.385

SUM OF RESIDUALS              8.745598
SUM OF SQUARED RESIDUALS      27.58216
```

LINE THROUGH ORIGIN

Ⓔ ARSENIC = DIST/NOINT;

NOINT is used, so SAS does not supply the intercept, and X0 is not specified in the model. Result: fitting a straight line through the origin.

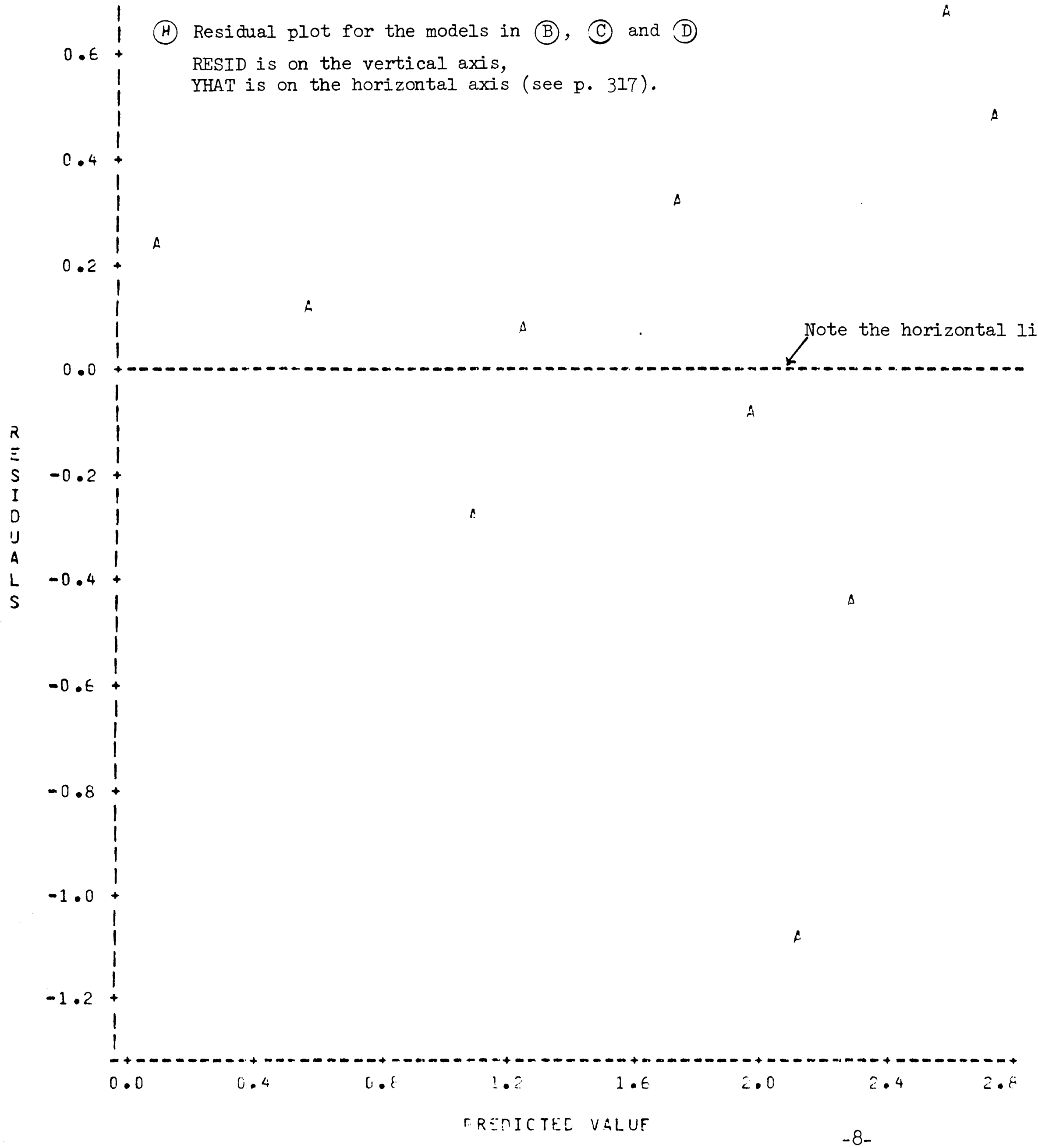


Such a model does not make much sense with the ARSENIC data set, although regression through the origin is useful in other circumstances.

RESIDUAL ANALYSIS

```
                                   UNIVARIATE

VARIABLE=RESID1         RESIDUALS

                MOMENTS                             QUANTILES(DEF=4)

N                 10  SUM WGTS         10  100% MAX  0.686377   99%      0.686377
MEAN      7.078E-16  SUM       7.078E-15   75% Q3    0.367045   95%      0.686377
STD DEV    0.509944  VARIANCE   0.260043   50% MED   0.106616   90%      0.663747
SKEWNESS  -0.960192  KURTOSIS    1.18871   25% Q1   -0.334325   10%     -1.02035
USS         2.34039  CSS         2.34039    0% MIN  -1.08472     5%     -1.08472
CV        7.205E+16  STD MEAN   0.161258                         1%     -1.08472
T:MEAN=0  4.389E-15  PROB>|T|          1   RANGE      1.7711
W:NORMAL   0.948196  PROB<W        0.619   Q3-Q1     0.70137
                                           MODE     -1.08472

                                   EXTREMES

                             LOWEST          HIGHEST
                          -1.08472          0.118294
                          -0.44102          0.227198
                          -0.29876          0.336034
                        -0.0984178          0.460075
                         0.0949383          0.686377

             STEM LEAF                     #             BOXPLOT
                0 57                       2                !  } 25% of residuals
                0 1123 (should approximate) 4            *--+--*
               -0 431  (a normal curve)    3             +-----+
               -0                                           |
               -1 1                        1                |  } 25% of residuals
                  ----+----+----+----+

 (All numbers are in tenths,
  i.e.,  -1 1=1.1)
                               NORMAL PROBABILITY PLOT
             0.75+                                   +    *
                 |                        *  *  *  *  *
           -0.25+               *  *  *  +                           <- should be ≈ a straight line if residuals are normally distributed.
                 |          +    +
           -1.25+          *
                  +----+----+----+----+----+----+----+----+----+----+
                      -2        -1        +0        +1        +2

                               FREQUENCY TABLE

                        PERCENTS                        PERCENTS
           VALUE COUNT  CELL   CUM          VALUE COUNT  CELL   CUM
        -1.08472     1 10.0  10.0      0.118294     1  10.0  60.0
        -0.44102     1 10.0  20.0      0.227198     1  10.0  70.0
        -0.29876     1 10.0  30.0      0.336034     1  10.0  80.0
        -.098418     1 10.0  40.0      0.460075     1  10.0  90.0
        .0949383     1 10.0  50.0      0.686377     1  10.0 100.0
```

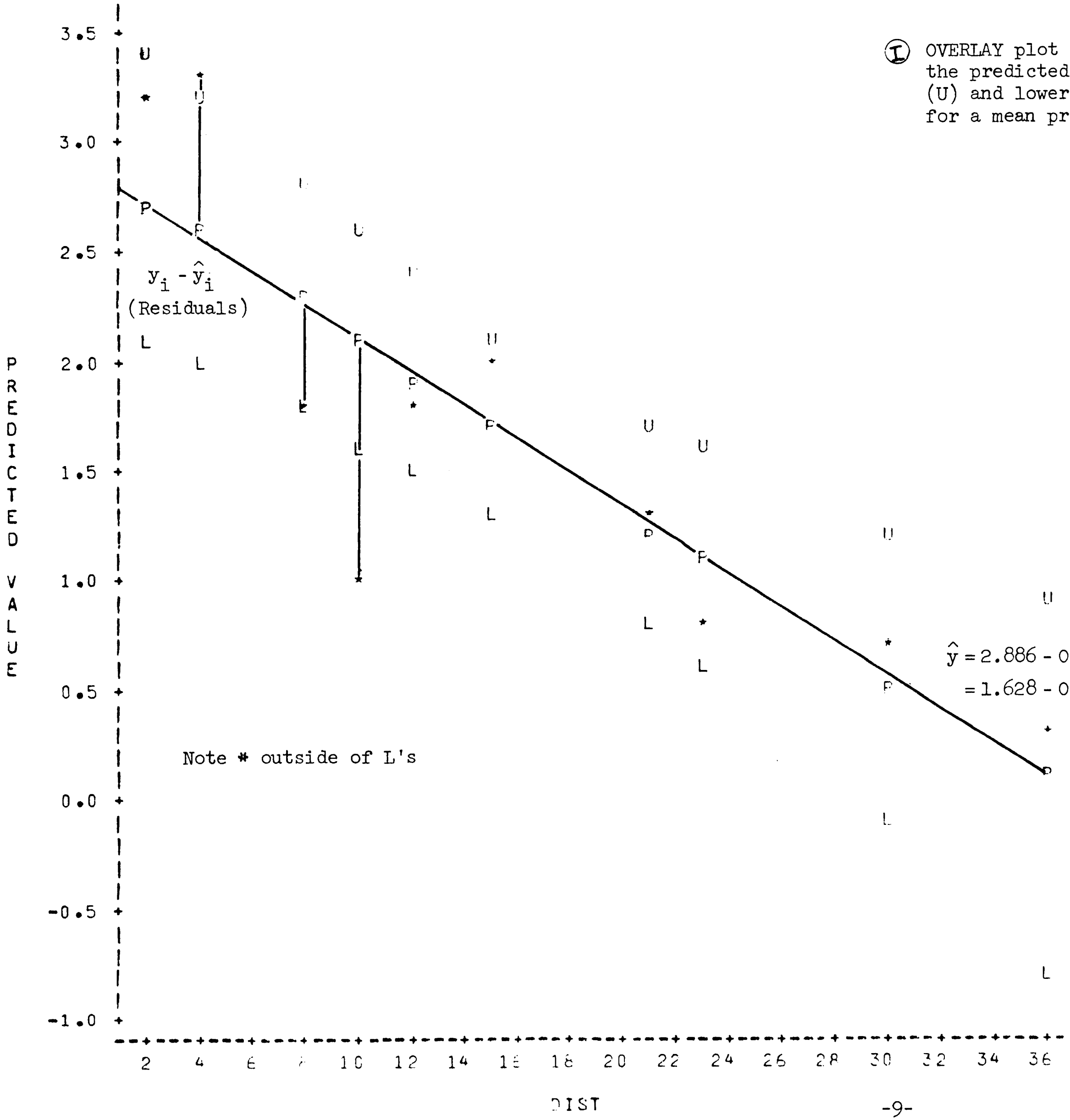
PLOT OF RESID1*YHAT1   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Ⓗ Residual plot for the models in Ⓑ, Ⓒ and Ⓓ
RESID is on the vertical axis,
YHAT is on the horizontal axis (see p. 317).

Note the horizontal line drawn by VREF = 0

```
PLOT OF ARSENIC*DIST     SYMBOL USED IS *
PLOT OF YHAT1*DIST       SYMBOL USED IS P
PLOT OF UPPER*DIST       SYMBOL USED IS U
PLOT OF LOWER*DIST       SYMBOL USED IS L
```



$y_i - \hat{y}_i$ (Residuals)

Note * outside of L's

$\hat{y} = 2.886 - 0.07815 * \text{DIST}$

$= 1.628 - 0.07815 * \text{DEV}$

Ⓘ OVERLAY plot of the observed values (*), the predicted values (P), and the upper (U) and lower (L) 95% confidence limits for a mean predicted value.

MEAN MODEL

Ⓕ ARSENIC = XO/NOINT

SS for the mean, equal to $n\bar{y}^2 = 10 * (1.628)^2$

```
DEP VARIABLE: ARSENIC

                      SUM OF            MEAN
SOURCE     DF        SQUARES          SQUARE        F VALUE        PROB>F

MODEL       1      26.503840       26.503840         25.863        0.0007
ERROR       9       9.222960        1.024773
U TOTAL    10      35.726800

    ROOT MSE        1.012311        R-SQUARE         0.7418
    DEP MEAN        1.628000        ADJ R-SQ         0.7418
    C.V.            62.18126

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.
```

Note that s^2 for this model is larger than for models Ⓑ, Ⓒ and Ⓓ. Fitting the slope reduces the estimate of σ^2.

```
                  PARAMETER       STANDARD    T FOR H0:
VARIABLE   DF      ESTIMATE          ERROR   PARAMETER=0    PROB > |T|    TYPE I SS

X0          1      1.628000       0.320121         5.086        0.0007            0
```

This model was run so that the predicted values, each equal to $\bar{y} = 1.628$, could be saved in the data set ONE, to be used in the data decomposition, part Ⓙ.

```
Ⓙ DATA DECOMPOSITION FOR ARSENIC DATA

OBS     ARSENIC      YBAR        RES1       RESID1

  1       3.19      1.628      1.1019       0.4601
  2       3.26      1.628      0.9456       0.6864
  3       1.82      1.628      0.6330      -0.4410
  4       1.02      1.628      0.4767      -1.0847
  5       1.85      1.628      0.3204      -0.0984
  6       2.05      1.628      0.0860       0.3360
  7       1.34      1.628     -0.3829       0.0949
  8       0.79      1.628     -0.5392      -0.2988
  9       0.66      1.628     -1.0863       0.1183
 10       0.30      1.628     -1.5552       0.2272
```

$$y_i = \bar{y} + (\hat{y}_i - \bar{y}) + (y_i - \hat{y}_i)$$

Each observation can be decomposed into the overall mean, the difference between the predicted values and the mean, and the difference between the observed and predicted values. See Unit 6, p. 50.

FIREFLY DATA – Units 10, 12, and 13; ACO, p. 351 (SAS)

```
TITLE FIREFLY DATA;
DATA FIREFLY;
INPUT FTIME LIGHT TEMP;
X0=1;
CARDS;
45 26  21.1
40 35  23.9
58 40  17.8
50 41  22.0
31 45  22.3
52 55  23.3
38 56  25.5
54 55  20.5
40 70  21.7
28 75  26.7
38 79  25.0
36 87  24.4
36 100 22.3
46 100 25.5
40 110 26.7
31 130 25.5
40 140 26.7
```

Creates the data set FIREFLY

FTIME is the response variable,
LIGHT and TEMP are two explanatory variables.

```
Ⓐ  PROC PRINT;
        VAR X0 LIGHT TEMP FTIME;
```

Prints the $\underset{\sim}{X} : \underset{\sim}{Y}$ matrix for Ⓒ

```
Ⓑ  PROC PLOT;
        PLOT FTIME*LIGHT FTIME*TEMP LIGHT*TEMP;
```

← 3 separate plots, FTIME against each explanatory variable and LIGHT vs. TEMP.

```
    PROC REG;
     Ⓒ MODEL FTIME=X0 LIGHT TEMP/NOINT P SS1 SS2 SECR    ;
           OUTPUT OUT=NEW1 PREDICTED=YHAT1 RESIDUAL=RESID1;
     Ⓓ MODEL FTIME= TEMP LIGHT/P SS1 SS2 SECR PARTIAL;
     Ⓔ MODEL FTIME=X0 TEMP/NOINT P SS1 SS2 SECR;
           OUTPUT OUT=NEW2 PREDICTED=YHAT2 RESIDUAL=RESID2;
```

PROC REG used to fit models with both LIGHT and TEMP, and then a reduced model

```
    PROC PLOT DATA=NEW1;
        PLOT RESID1*YHAT1/VREF=0;
```

Residual plot for Ⓒ and Ⓓ

```
    PROC PLOT DATA=NEW2;
        PLOT RESID2*TEMP/VREF=0;
```

Residual plot for Ⓔ

```
                        FIREFLY DATA

OBS     X0      LIGHT      TEMP      FTIME

  1      1        26       21.1        45
  2      1        35       23.9        40
  3      1        40       17.8        58
  4      1        41       22.0        50
  5      1        45       22.3        31
  6      1        55       23.3        52
  7      1        56       25.5        38
  8      1        55       20.5        54
  9      1        70       21.7        40
 10      1        75       26.7        28
 11      1        79       25.0        38
 12      1        87       24.4        36
 13      1       100       22.3        36
 14      1       100       25.5        46
 15      1       110       26.7        40
 16      1       130       25.5        31
 17      1       140       26.7        40
```

Ⓐ $\underset{\sim}{X} \mid \underset{\sim}{Y}$ **matrix for the general model:**
FTIME = XO LIGHT TEMP/NOINT in part Ⓒ**.**

FULL MODEL

Ⓒ FTIME = XO LIGHT TEMP/NOINT

```
SEQUENTIAL PARAMETER ESTIMATES ← The sequential coefficients are the diagonal elements of the SEQB output (see p. 99).

X0          41.3529 = ȳ ← See p. 94.
LIGHT       48.8962 -.103083 ← intercept and slope coefficients for FTIME = XO LIGHT (see p. 95).
TEMP        91.4745  6.9E-04 -2.12753 ← intercept and two slopes which define a tilted plane (see p. 99).

DEP VARIABLE: FTIME      Note that the slope in the LIGHT direction changes
                         sign when TEMP is added to the model.

                      SUM OF           MEAN
SOURCE       DF      SQUARES         SQUARE        F VALUE        PROB>F

MODEL         3    29527.857       9842.619        201.710        0.0001
ERROR        14      683.143     48.795924 = s²
U TOTAL      17    30211.000← Includes SS for the mean

     ROOT MSE      6.985408 = s   R-SQUARE        0.9774
     DEP MEAN     41.352941 = ȳ   ADJ R-SQ        0.9742
     C.V.          16.89217

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

                    PARAMETER      STANDARD   T FOR H0:
VARIABLE  DF         ESTIMATE         ERROR  PARAMETER=0    PROB > |T|    TYPE I SS
 Intercept
X0         1        91.474451     18.825103        4.859        0.0003    29071.118 = R(O)
LIGHT      1  0.0006919891       0.068339        0.010        0.9921      194.421 = R(L|O)
TEMP       1        -2.127529      0.917559       -2.319        0.0361      262.318 = R(T|L,O)
 Partial coefficients                 Slope of plane in LIGHT direction, see p. 100.
                                      Slope of plane in TEMP direction, see p. 100.

VARIABLE  DF     TYPE II SS

X0         1       1152.148 = R(O|T,L)
LIGHT      1    .005003143 = R(L|T,O)
TEMP       1        262.318 = R(T|L,O)
                when LIGHT is fitted last, it does not account for much of the remaining SS.

                  PREDICT
  OBS   ACTUAL      VALUE RESIDUAL

    1   45.000     46.602   -1.602
    2   40.000     40.651 -.650720
    3   58.000     53.632    4.368
    4   50.000     44.697    5.303
    5   31.000     44.062  -13.062
    6   52.000     41.941   10.059
    7   38.000     37.261 0.738795
    8   54.000     47.898    6.102
    9   40.000     45.356   -5.356
   10   28.000     34.721   -6.721
   11   38.000     38.341 -.340885
   12   36.000     39.623   -3.623
   13   36.000     44.100   -8.100
   14   46.000     37.292    8.708
   15   40.000     34.746    5.254
```

FULL MODEL

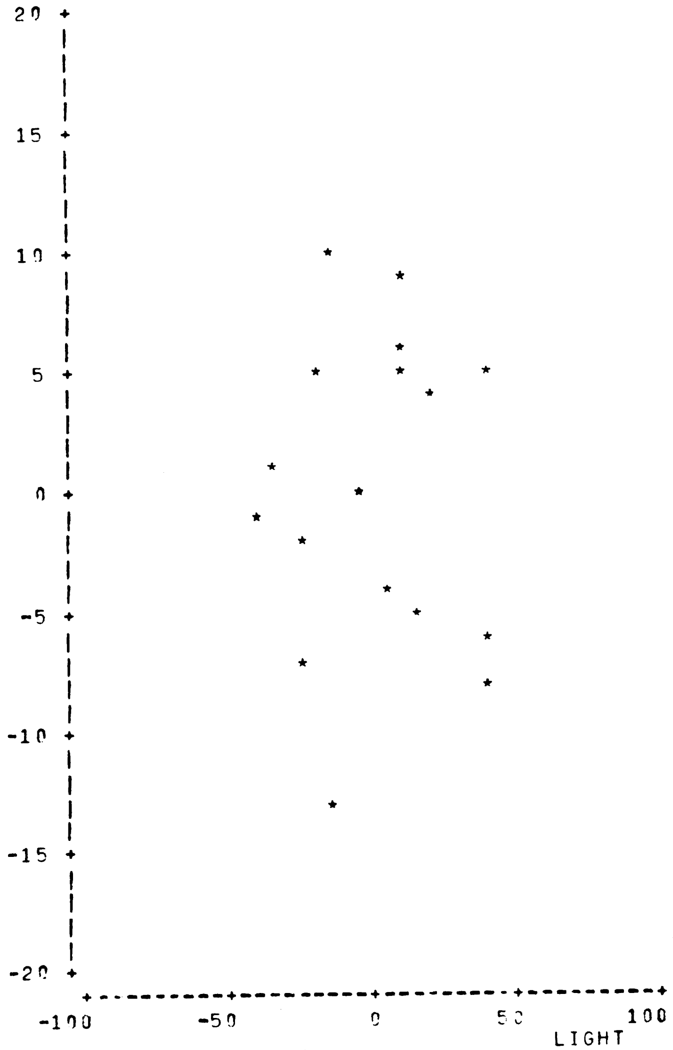
Ⓓ FTIME = TEMP LIGHT

```
                PREDICT
 OBS    ACTUAL    VALUE RESIDUAL

  16    31.000   37.312   -6.312
  17    40.000   34.766    5.234

SUM OF RESIDUALS            4.72511E-13 ≐ 0
SUM OF SQUARED RESIDUALS      683.1429

SEQUENTIAL PARAMETER ESTIMATES

INTERCEP   41.3529 = ȳ
TEMP       91.3816 -2.12144 ← intercept and slope coefficients for FTIME = TEMP (no NOINT)
LIGHT      91.4745 -2.12753  6.9E-04 ← same as Ⓒ
                                   Note that the sequential coefficients from ABDO are on
DEP VARIABLE: FTIME                  the diagonal of the SEQB output.  See p. 100.

                         SUM OF           MEAN
SOURCE      DF          SQUARES          SQUARE         F VALUE        PROB>F   same ANOVA as Ⓒ, except
                                                                                that the MODEL and TOTAL
MODEL        2          456.739         228.370           4.680        0.0182   SS are corrected for the
ERROR       14          683.143       48.795924 = s², same as Ⓒ                 mean.
C TOTAL     16         1139.882

    ROOT MSE           6.985408         R-SQUARE         0.4007
    DEP MEAN          41.352941         ADJ R-SQ         0.3151
    C.V.               16.89217

                      PARAMETER         STANDARD    T FOR H0:
VARIABLE    DF         ESTIMATE            ERROR  PARAMETER=0      PROB > |T|     TYPE I SS

INTERCEP     1        91.474451        18.825103        4.859          0.0003    29071.118 = R(O)
TEMP         1        -2.127529         0.917599       -2.319          0.0361      456.734 = R(T|O)
LIGHT        1     0.0006919891         0.068339        0.010          0.9921  0.005003143 = R(L|T,O)
                                                                               Again, LIGHT fitted
                                                                               after TEMP accounts
VARIABLE    DF      TYPE II SS                                                 for little of the
                                                                               remaining SS.
INTERCEP     1        1152.148                      = R(O|T,L)
TEMP         1         262.318                      = R(T|L,O)
LIGHT        1     0.005003143                      = R(L|T,O)
```

```
                      PREDICT
 OBS     ACTUAL         VALUE RESIDUAL

   1     45.000        46.602   -1.602
   2     40.000        40.651 -.650720
   3     58.000        53.632    4.368
   4     50.000        44.697    5.303
   5     31.000        44.062  -13.062
   6     52.000        41.941   10.059
   7     38.000        37.261 0.738795
   8     54.000        47.898    6.102
   9     40.000        45.356   -5.356
  10     28.000        34.721   -6.721
  11     38.000        38.341 -.340885
  12     36.000        39.623   -3.623
  13     36.000        44.100   -8.100
  14     46.000        37.292    8.708
  15     40.000        34.746    5.254
  16     31.000        37.312   -6.312
  17     40.000        34.766    5.234

SUM OF RESIDUALS                3.94351E-13
SUM OF SQUARED RESIDUALS           683.1429

                 Same as Ⓒ
```

PARTIAL REGRESSION RESIDUAL PLOTS

FTIME — $\widehat{\text{FTIME}}$ (INTERCEPT, TEMP)



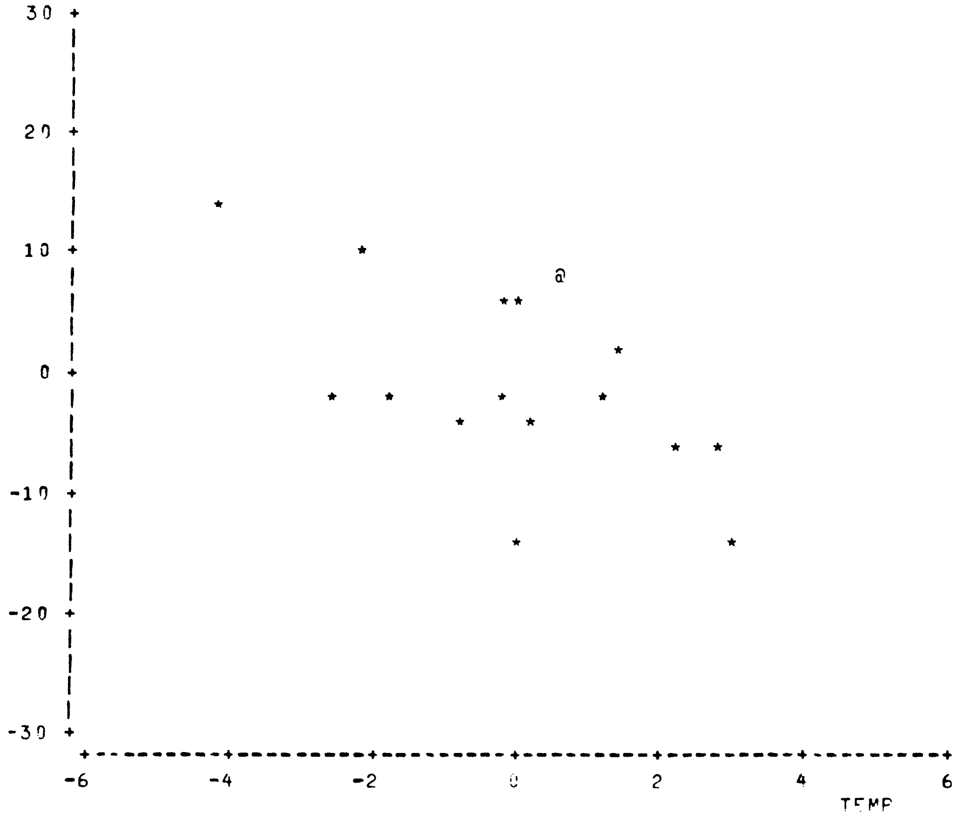
LIGHT — $\widehat{\text{LIGHT}}$ (INTERCEPT, TEMP)

There is little linear trend in this plot, so fitting LIGHT after an intercept and TEMP will account for a small portion of the remaining sum of squares, as in Ⓒ and Ⓓ.

The option PARTIAL in the model statement from Ⓓ produces these plots. These are similar to the plots on p. 141

PARTIAL REGRESSION RESIDUAL PLOTS

FTIME — $\widehat{\text{FTIME}}$ (INTERCEPT, LIGHT)



TEMP — $\widehat{\text{TEMP}}$ (INTERCEPT, LIGHT)

REDUCED MODEL

Ⓔ FTIME = XO TEMP/NOINT

```
SEQUENTIAL PARAMETER ESTIMATES

XO          41.3529 = ȳ
TEMP        91.3816 -2.12144 ← same estimates as second line, Ⓓ

DEP VARIABLE: FTIME

                      SUM OF          MEAN
SOURCE      DF       SQUARES        SQUARE        F VALUE       PROB>F

MODEL        2     29527.852     14763.926        324.174       0.0001
ERROR       15       683.148     45.543196 = s²
U TOTAL     17     30211.000

    ROOT MSE        6.748570      R-SQUARE        0.9774
    DEP MEAN       41.352941      ADJ R-SQ        0.9759
    C.V.            16.31944

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.
```

Note that s^2 has not increased very much from the s^2 in Ⓒ and Ⓓ where both TEMP and LIGHT were fitted.

```
                  PARAMETER      STANDARD   T FOR H0:
VARIABLE    DF     ESTIMATE         ERROR  PARAMETER=0    PROB > |T|    TYPE I SS

XO           1    91.381582     15.882436        5.754        0.0001    29071.118
TEMP         1    -2.121444      0.669902       -3.167        0.0064      456.734
```

Compare with standard errors of the estimates, part Ⓓ

SS for reduced model (FTIME = XO/NOINT)

= SS for the mean ($n\bar{y}^2$) ≐ 456.736

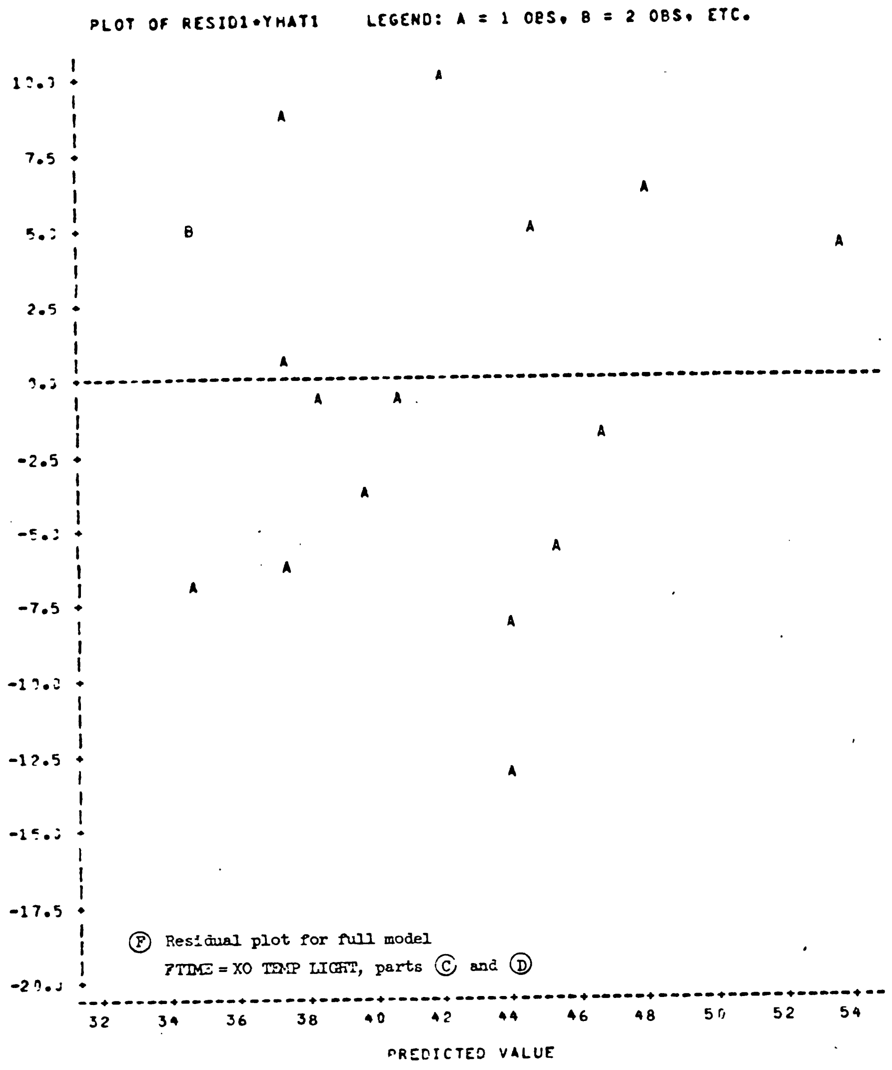
```
VARIABLE    DF    TYPE II SS     TOLERANCE

XO           1      1507.671   1805.465876
TEMP         1       456.734      1.000000


                 PREDICT
 OBS    ACTUAL     VALUE  RESIDUAL

   1    45.000    46.619    -1.619
   2    40.000    40.679  -.679071
   3    58.000    53.620     4.380
   4    50.000    44.710     5.290
   5    31.000    44.073   -13.073
   6    52.000    41.952    10.048
   7    38.000    37.285  0.715240
   8    54.000    47.892     6.108
   9    40.000    45.346    -5.346
  10    28.000    34.739    -6.739
  11    38.000    38.345  -.345482
  12    36.000    39.618    -3.618
  13    36.000    44.073    -8.073
  14    46.000    37.285     8.715
  15    40.000    34.739     5.261
  16    31.000    37.285    -6.285
  17    40.000    34.739     5.261

SUM OF RESIDUALS              3.65930E-13
SUM OF SQUARED RESIDUALS         683.1479
```

F test for the need of the general model with TEMP and LIGHT over the reduced model with TEMP (pp. 138-140):

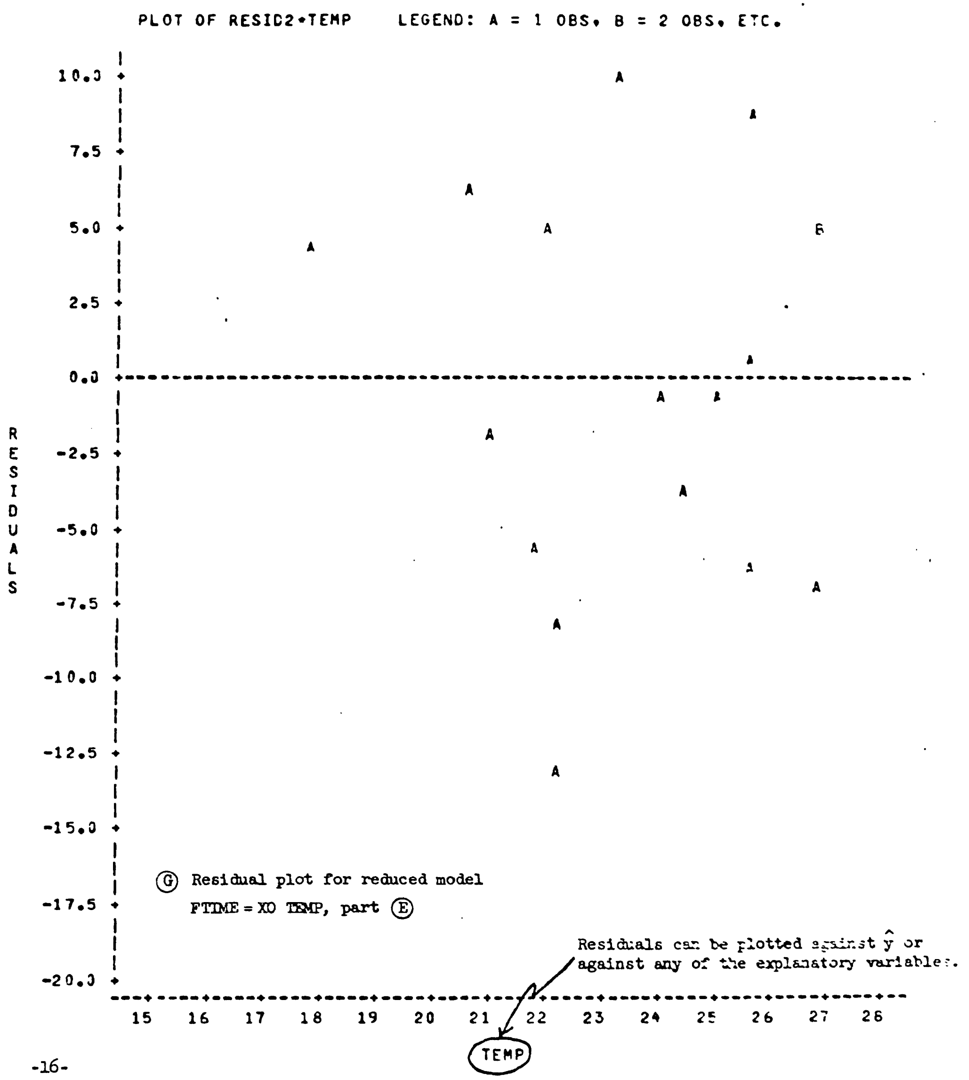
$$F'_{14} = \frac{[(\text{Model SS, general model}) - (\text{Model SS, reduced model})]/\text{difference in df between models}}{(\text{Residual SS, general model})/(\text{Residual df, general model})}$$

$$= \frac{(456.739 - 456.736)/1}{(683.143)/14} \doteq 0.0001$$ Reduced model is adequate .

This test is equivalent to the t-test of the LIGHT parameter estimate in Ⓒ or Ⓓ.



Ⓕ Residual plot for full model
FTIME = XO TEMP LIGHT, parts Ⓒ and Ⓓ



Ⓖ Residual plot for reduced model
FTIME = XO TEMP, part Ⓔ

Residuals can be plotted against $\hat{y}$ or against any of the explanatory variables.

SOYMILK DATA – Unit 11

```
TITLE'';
DATA SOYMILK;
INPUT TIME Y;
TIME=TIME/60.0;
TIME2=TIME*TIME;
TIME3=TIME2*TIME;
LOF=TIME;
CARDS;
0     2.74
0     2.25
0     2.34
12    3.14
12    2.68
12    2.83
30    3.44
30    3.53
30    3.63
60    3.68
60    3.75
60    3.51
```

The data set SOYMILK has variables TIME, TIME2 = $(\text{TIME})^2$, TIME3 = $(\text{TIME})^3$, LOF = TIME, and the response variable Y . The times are coded as TIME/60.0 to ensure numerical accuracy. Since there are four levels of the explanatory variable, the maximum degree polynomial that can be fitted is a cubic.

Ⓐ
```
PROC REG;
     MODEL Y=TIME TIME2 TIME3/SS1 SS2 SEQB;
```
— Fits the cubic polynomial model

```
PROC SORT; BY TIME;

PROC MEANS MEAN NOPRINT;
     BY TIME; VAR Y TIME;
     OUTPUT OUT=NEW MEAN=YBAR X;
```

In order to fit the cubic polynomial to the mean response, the mean response at each level of TIME must be found. The data must be sorted BY TIME before taking MEANS BY TIME. The data set NEW contains four observations (each TIME level). The variable names are YBAR (the mean response) and X (the mean response at TIME levels).

Ⓑ
```
PROC PRINT;
     VAR X YBAR;
```
Prints the four observations in NEW .

```
DATA NEW1;
SET NEW;
X2=X*X; X3=X2*X;
```

Data set NEW1 will have four observations: the response variable is YBAR and the explanatory variables X, X2 = X^2, and X3 = X^3 .

Ⓒ
```
PROC REG DATA=NEW1;
     MODEL YBAR=X X2 X3/SS1 SS2 SEQB;
```
← Fitting the cubic polynomial model to the mean response at each X level.

Ⓓ
```
PROC GLM DATA=SOYMILK;
     CLASS LOF;
     MODEL Y=TIME TIME2 LOF;
```

Fits quadratic model. PROC GLM (General Linear Models) used to fit the full model. Unlike PROC REG, only one MODEL statement may follow each PROC GLM. If no SS type is specified GLM gives Type I (sequentials) and Type IV (partials). LOF = lack of fit due to fitting a lower polynomial to the data. The CLASS statement constructs an indicator variable for each level of the LOF = TIME variable (see p. 166).

POLYNOMIAL MODEL ESTIMATION

Ⓐ Y = TIME TIME2 TIME3

```
SEQUENTIAL PARAMETER ESTIMATES

INTERCEP  3.12667
TIME      2.62141   1.18884
TIME2     2.41049   3.07182  -1.82743
TIME3     2.44333    1.9775   1.58417  -2.35833
```

The coefficients on the diagonal of the SEQB output are the sequential coefficients for the cubic polynomial model.

← The partial coefficients for the cubic polynomial model.

```
DEP VARIABLE: Y

                       SUM OF          MEAN
SOURCE        DF      SQUARES        SQUARE       F VALUE       PROB>F

MODEL          3     2.885800      0.961933        26.116       0.0001
ERROR          8     0.294667      0.036833 = s^2
C TOTAL       11     3.180467

    ROOT MSE         0.191920      R-SQUARE        0.9074
    DEP MEAN         3.126667      ADJ R-SQ        0.8726
    C.V.              6.13817
```

s^2 is an estimate of "pure error" here since the maximum degree polynomial has been fitted. The error has 8 df, 2 from each level of TIME.

```
                   PARAMETER       STANDARD    T FOR H0:
VARIABLE  DF        ESTIMATE          ERROR  PARAMETER=0    PROB > |T|    TYPE I SS

INTERCEP   1        2.443333       0.110805       22.051        0.0001      117.313
TIME       1        1.977500       1.485287        1.331        0.2197     2.406212
TIME2      1        1.584167       4.326547        0.366        0.7237     0.456351
TIME3      1       -2.358333       2.969196       -0.794        0.4500     0.023237


VARIABLE  DF     TYPE II SS

INTERCEP   1      17.909633
TIME       1       0.065291
TIME2      1    0.004938101
TIME3      1       0.023237
```

See p. 113

The pure error can be used to test for the "lack of fit" of the quadratic and straight line models.

Lack of fit of the quadratic model:

$$\frac{\underline{\text{SS full (cubic) model - SS reduced (quadratic) model/df}}}{\text{Pure Error SS/Pure Error df (= full model MSE)}}$$

$$= \frac{(2.88580-2.862563)/1}{.294667/8} = \frac{.023237}{.036833} = .6308$$

The quadratic model is adequate.

Ⓑ

DATA in NEW

SOYMILK BOILING TIME DATA

OBS	X	YBAR
1	0.0	2.44333
2	0.2	2.88333
3	0.5	3.53333
4	1.0	3.64667

mean Y for each level of X. See p. 113.

POLYNOMIAL MODEL ESTIMATION USING MEANS

Ⓒ Y = X X2 X3

Same parameter estimates as Ⓐ, so fitting the curve through the mean response is the same as using all the observations, except that with no replication, there are no df or SS for pure error.

```
SEQUENTIAL PARAMETER ESTIMATES

INTERCEP  3.12667
X         2.62141  1.18884
X2        2.41049  3.07182 -1.82743
X3        2.44333   1.9775  1.58417 -2.35833

DEP VARIABLE: YBAR

                     SUM OF          MEAN
SOURCE      DF      SQUARES        SQUARE       F VALUE        PROB>F

MODEL        3     0.961933      0.320644         0.000        0.0000
ERROR        0            0             .
C TOTAL      3     0.961933

    ROOT MSE              .      R-SQUARE        1.0000
    DEP MEAN       3.126667      ADJ R-SQ        0.0000
    C.V.                  0

                  PARAMETER      STANDARD    T FOR H0:
VARIABLE    DF     ESTIMATE         ERROR  PARAMETER=0    PROB > |T|      TYPE I SS

INTERCEP     1     2.443333             0            .        1.0000      39.104178
X            1     1.977500             0            .        1.0000       0.802071
X2           1     1.584167             0            .        1.0000       0.152117
X3           1    -2.358333             0            .        1.0000 0.007745551


VARIABLE    DF    TYPE II SS

INTERCEP     1             0
X            1             0
X2           1             0
X3           1             0
```

QUADRATIC MODEL EVALUATION
Ⓓ Y = TIME TIME2 LOF

```
                                  GENERAL LINEAR MODELS PROCEDURE

                                     CLASS LEVEL INFORMATION

                                  CLASS      LEVELS     VALUES

                                  LOF           4       0 1 0.2 0.5


                               NUMBER OF OBSERVATIONS IN DATA SET = 12

                  GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                  DF       SUM OF SQUARES         MEAN SQUARE        F VALUE

MODEL                    3           2.88580000          0.96193333          26.12

ERROR                    8           0.29466667          0.03683333         PR > F

CORRECTED TOTAL         11           3.18046667                             0.0002

~~R-SQUARE~~                C.V.              STD DEV              Y MEAN

~~0.907351~~              6.1382           0.19192012          3.12666667
                                    (Sequentials)
SOURCE                  DF             TYPE I SS    F VALUE    PR > F

TIME                     1            2.40621204      65.33    0.0001
TIME2                    1            0.45635131      12.39    0.0078
LOF                      1            0.02323665       0.63    0.4500
                                      (Partials)
SOURCE                  DF            TYPE IV SS    F VALUE    PR > F

TIME                     0            0.00000000          .         .
TIME2                    0            0.00000000          .         .
LOF                      1            0.02323665       0.63    0.4500
```

– Compare F = 0.63 with the annotated F statistic for Model Ⓐ, Y = TIME TIME2 TIME3 . In general, the lack of fit (LOF) partial sum of squares is the remaining sum of squares after the fitted model sum of squares. With t treatments (t levels of a quantitative variable for example), DF = t - p for LOF after fitting a p parameter model including one DF for the mean. In this output, DF = 4 - 3 = 1 but in general DF could be greater than 1 .

ELECTRICITY LOAD DATA – Unit 13; ACO, p. 363 (BMDP)

```
   OPTIONS LS=80 NODATE:
TITLE ELECTRICITY LOAD DATA;
DATA ELECLOAD;
INPUT DATE DAY $ TEMP Y;
X3=0: X4=0; X5=0; X6=0; X1=1: X2=TEMP;
IF DAY='SU' OR DATE=5 THEN LINK SUNDAY;
IF DAY='SA' THEN LINK SATURDAY;
RETURN;
SUNDAY: X5=1: X6=TEMP; X1=0: X2=0; RETURN;
SATURDAY: X3=1; X4=TEMP: Y1=0: X2=0; DAY='G'; RETURN:
CARDS:
```

$ used since DAY is a character variable

The link command causes the appropriately labeled series of statements to be executed. For example, if DAY = 'SA' then the five statements between SATURDAY: and RETURN are executed.

'SA' is changed to 'Q' for labeling purposes in Ⓖ

See Ⓐ and Ⓑ for the X variables created by the above statements.

Ⓐ
```
PROC PRINT:
     VAR X1 X2 X3 X4 X5 X6;
```
Prints the $\underset{\sim}{X}$ matrix for the full model: 3 intercepts and 3 slopes.

Ⓑ
```
PROC PRINT:
     VAR X1 X3 X5 TEMP;
```
Prints the $\underset{\sim}{X}$ matrix for the reduced model: 3 intercepts and 1 slope.

Ⓒ
```
PROC GLM;
     MODEL Y=X1 X2 X3 X4 X5 X6/NOINT P;
        OUTPUT OUT=NEW1 PREDICTED=YHAT1 RESIDUAL=RESID1;
```

Ⓓ
```
PROC PLOT DATA=NEW1;
     PLOT RESID1*YHAT1/VREF=0;
```
Residual plot for the full model.

Ⓔ
```
PROC GLM;
     MODEL Y=X1 X3 X5 TEMP/NOINT;
        OUTPUT OUT=NEW2 PREDICTED=YHAT2 RESIDUAL=RESID2;
```
Fits the reduced model.

```
PROC PLOT DATA=NEW2;
     PLOT RESID2*YHAT2/VREF=0;
     PLOT Y*TEMP=DAY;
```
Ⓕ PLOT RESID2*YHAT2/VREF=0; The residual plot for the reduced model.

Ⓖ PLOT Y*TEMP=DAY; Y vs. TEMP, and the character used to plot each observation (Y) is the data stored in DAY which corresponds to that observation (i.e., the first letter of the day in which the observation was taken).

Ⓐ The $\underset{\sim}{X}$ matrix for the full model in Ⓒ, p. 145

Separate Intercepts

Separate Slopes

ELECTRICITY LOAD DATA

OBS	X1	X2	X3	X4	X5	X6
1	1	77	0	0	0	0
2	1	75	0	0	0	0
3	0	0	1	74	0	0
4	0	0	0	0	1	78
5	0	0	0	0	1	77
6	1	72	0	0	0	0
7	1	79	0	0	0	0
8	1	81	0	0	0	0
9	1	85	0	0	0	0
10	0	0	1	83	0	0
11	0	0	0	0	1	82
12	1	75	0	0	0	0
13	1	77	0	0	0	0
14	1	75	0	0	0	0
15	1	77	0	0	0	0
16	1	77	0	0	0	0
17	0	0	1	82	0	0
18	0	0	0	0	1	74
19	1	72	0	0	0	0
20	1	71	0	0	0	0
21	1	73	0	0	0	0
22	1	78	0	0	0	0
23	1	78	0	0	0	0
24	0	0	1	74	0	0
25	0	0	0	0	1	77
26	1	79	0	0	0	0
27	1	73	0	0	0	0
28	1	75	0	0	0	0
29	1	69	0	0	0	0
30	1	71	0	0	0	0
31	0	0	1	68	0	0

X1 = 1 if weekday, 0 otherwise

X3 = 1 if Saturday, 0 otherwise

X5 = 1 if Sunday or holiday, 0 otherwise

X2 = temp if weekday, 0 otherwise

X4 = temp if Saturday, 0 otherwise

X6 = temp if Sunday or holiday, 0 otherwise

Ⓑ The $\underset{\sim}{X}$ matrix for the reduced model in Ⓔ

ELECTRICITY LOAD DATA

OBS	X1	X3	X5	TEMP
1	1	0	0	77
2	1	0	0	75
3	0	1	0	74
4	0	0	1	78
5	0	0	1	77
6	1	0	0	72
7	1	0	0	79
8	1	0	0	81
9	1	0	0	85
10	0	1	0	83
11	0	0	1	82
12	1	0	0	75
13	1	0	0	77
14	1	0	0	75
15	1	0	0	77
16	1	0	0	77
17	0	1	0	82
18	0	0	1	74
19	1	0	0	72
20	1	0	0	71
21	1	0	0	73
22	1	0	0	78
23	1	0	0	78
24	0	1	0	74
25	0	0	1	77
26	1	0	0	79
27	1	0	0	73
28	1	0	0	75
29	1	0	0	69
30	1	0	0	71
31	0	1	0	68

Separate Intercepts

Common Slope

Ⓒ FULL MODEL, 3 INTERCEPTS AND 3 SLOPES

Ⓒ Y = X1 X2 X3 X4 X5 X6/NOINT

```
DEPENDENT VARIABLE: Y

SOURCE                     DF       SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                       6    34740917.09721010     5790152.84953502        5432.09

ERROR                      25       26647.90278990        1065.91611160         PR > F

UNCORRECTED TOTAL          31    34767565.00000000                               0.0001
Like PROC REG, PROC GLM computes
an incorrect R^2 when NOINT is used.
R-SQUARE                 C.V.              STD DEV               Y MEAN

0.999234                3.1015          32.64837073        1052.67741935

                                    (sequentials)
SOURCE                     DF          TYPE I SS     F VALUE      PR > F

X1                          1   26243932.19047619   24621.01      0.0001
X2                          1      50779.42480620      47.64      0.0001
X3                          1    4723920.00000000    4431.79      0.0001
X4                          1      28880.00000000      27.09      0.0001
X5                          1    3692841.80000000    3464.48      0.0001
X6                          1        563.68192771       0.53      0.4739

                                     (partials)
SOURCE                     DF         TYPE IV SS     F VALUE      PR > F

X1                          1        613.59455571       0.58      0.4551
X2                          1      50779.42480620      47.64      0.0001
X3                          1        103.72400562       0.10      0.7577
X4                          1      28880.00000000      27.09      0.0001
X5                          1       1603.84116067       1.50      0.2314
X6                          1        563.68192771       0.53      0.4739

                        PARTIAL        T FOR H0:       PR > |T|        STD ERROR OF
PARAMETER              ESTIMATE      PARAMETER=0                         ESTIMATE

X1 weekday intercept=110.83421927          0.76         0.4551        146.08130199
X2   weekday slope = 13.30930233           6.90         0.0001          1.92829251
X3 Satur. intercept=-62.14285714          -0.31         0.7577        199.21039667
X4     Satur. slope = 13.57142857          5.21         0.0001          2.60728487
X5 Sunday intercept=539.65060241           1.23         0.2314        439.94005229
X6     Sunday slope =  4.12048193          0.73         0.4739          5.66620746
```

PROC GLM has no option which gives estimates of the sequential b's.

Ⓔ REDUCED MODEL, 3 INTERCEPTS AND A COMMON SLOPE
Ⓔ Y = X1 X3 X5 TEMP/NOINT

```
                         GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                    DF        SUM OF SQUARES         MEAN SQUARE         F VALUE

MODEL                      4     34738249.28002397     8684562.32000599        7998.55

ERROR                     27        29315.71997603        1085.76740652        PR > F

UNCORRECTED TOTAL         31     34767565.00000000                             0.0001

~~R-SQUARE~~                    C.V.             STD DEV             Y MEAN

~~0.999157~~                  3.1302         32.95098491       1052.67741935

                                                Sequentials
SOURCE                   DF                     TYPE I SS     F VALUE      PR > F

X1}                       1          R(1) = 26243932.19047619    24170.86      0.0001
X3} intercepts            1        R(3|1) =  4723920.00000000     4350.77      0.0001
X5}                       1      R(5|3,1) =  3692841.80000000     3401.14      0.0001
TEMP<-- slope             1    R(T|5,3,1) =    77555.28954779       71.43      0.0001

                                                Partials
SOURCE                                         TYPE IV SS     F VALUE      PR > F

X1                        1        R(1|3,5,T) =1934.52471220        1.78      0.1931
X3                        1           R(3|1,5,T) =0.00006809        0.00      0.9998
X5                        1      R(5|1,3,T) =  1325.61070453        1.22      0.2789
TEMP                      1      R(T|1,3,5) =77555.28954779       71.43      0.0001

                                           T FOR H0:        PR > |T|          STD ERROR OF
PARAMETER                   ESTIMATE      PARAMETER=0                            ESTIMATE

X1 weekday intercept=152.73674326                1.33          0.1931          114.42601814
X3  Satur. intercept=  0.02903497                0.00          0.9998          115.94506590
X5 Sunday intercept=-130.42869930               -1.10          0.2789          118.04118545
TEMP  common slope=  12.75552448                 8.45          0.0001            1.50924939
      \___________________________/
        Prediction equation given
               on p. 146
```

To evaluate the adequacy of the reduced model, form the F statistic

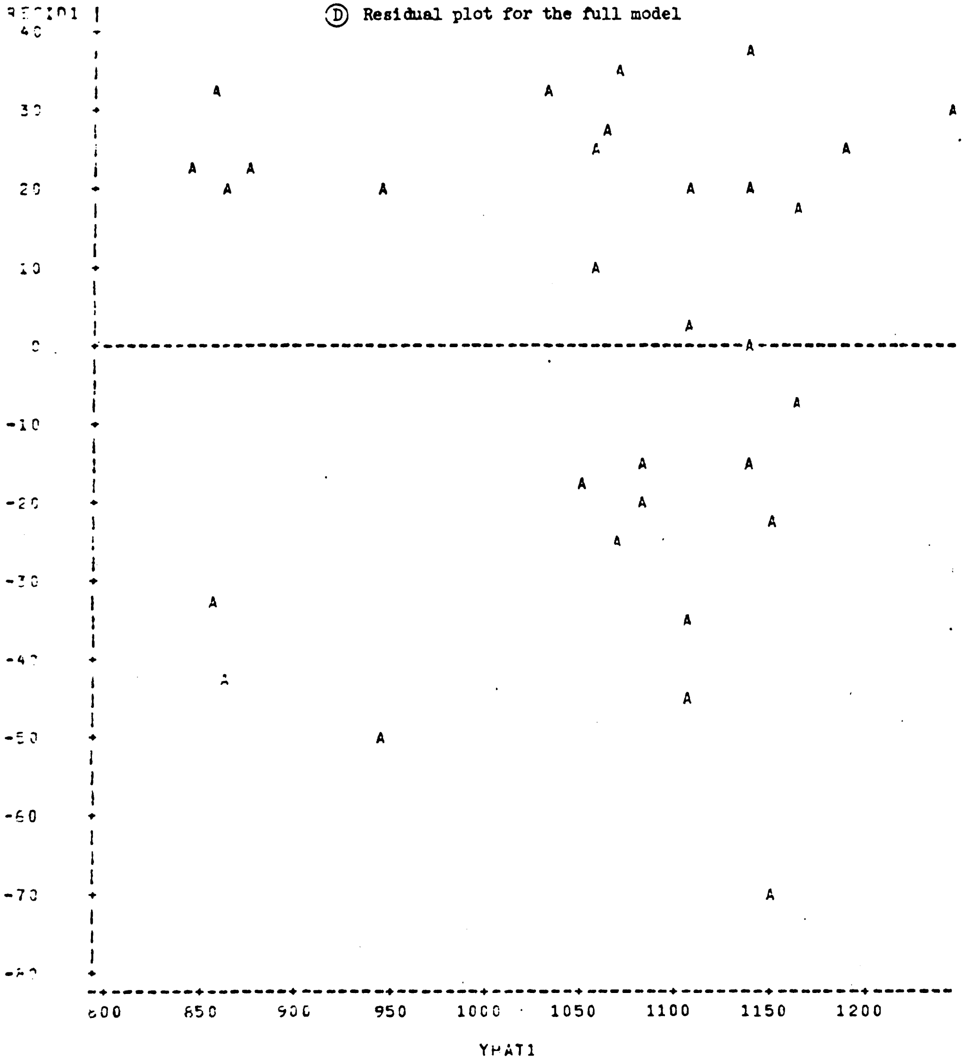
$$\frac{[(\text{Model SS, full model}) - (\text{Model SS, reduced model})]/\begin{matrix}\text{difference in df}\\ \text{between the 2 models}\end{matrix}}{(\text{Residual SS, full model})\,/\,(\text{Residual df, full model})}$$

$$= \frac{[34,740,917 - 34,738,249]/2}{1065.916} \doteq 1.25 = F^2_{25}$$

Based on the relatively small F value, the reduced model is judged to be sufficient and there is no need to fit separate slopes.

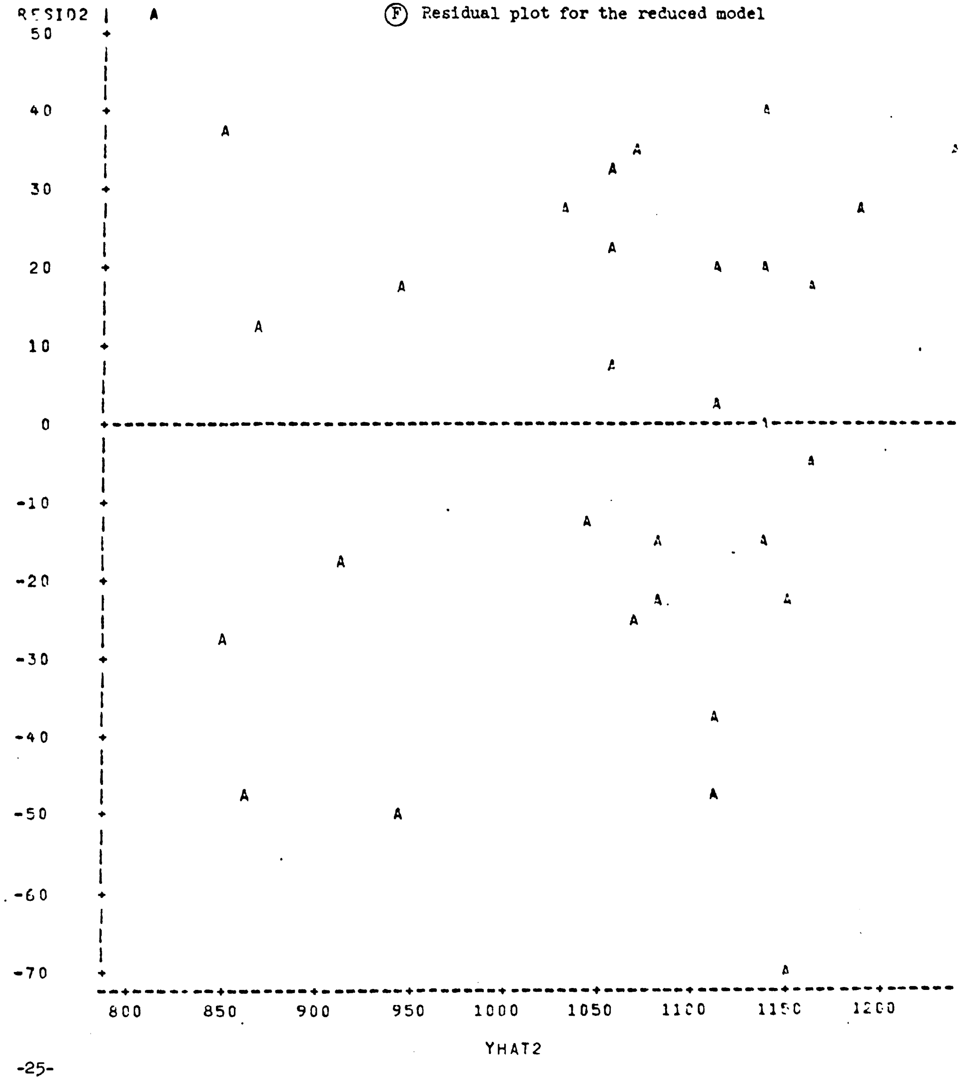
RESIDUAL ANALYSIS

PLOT OF RESID1*YHAT1   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



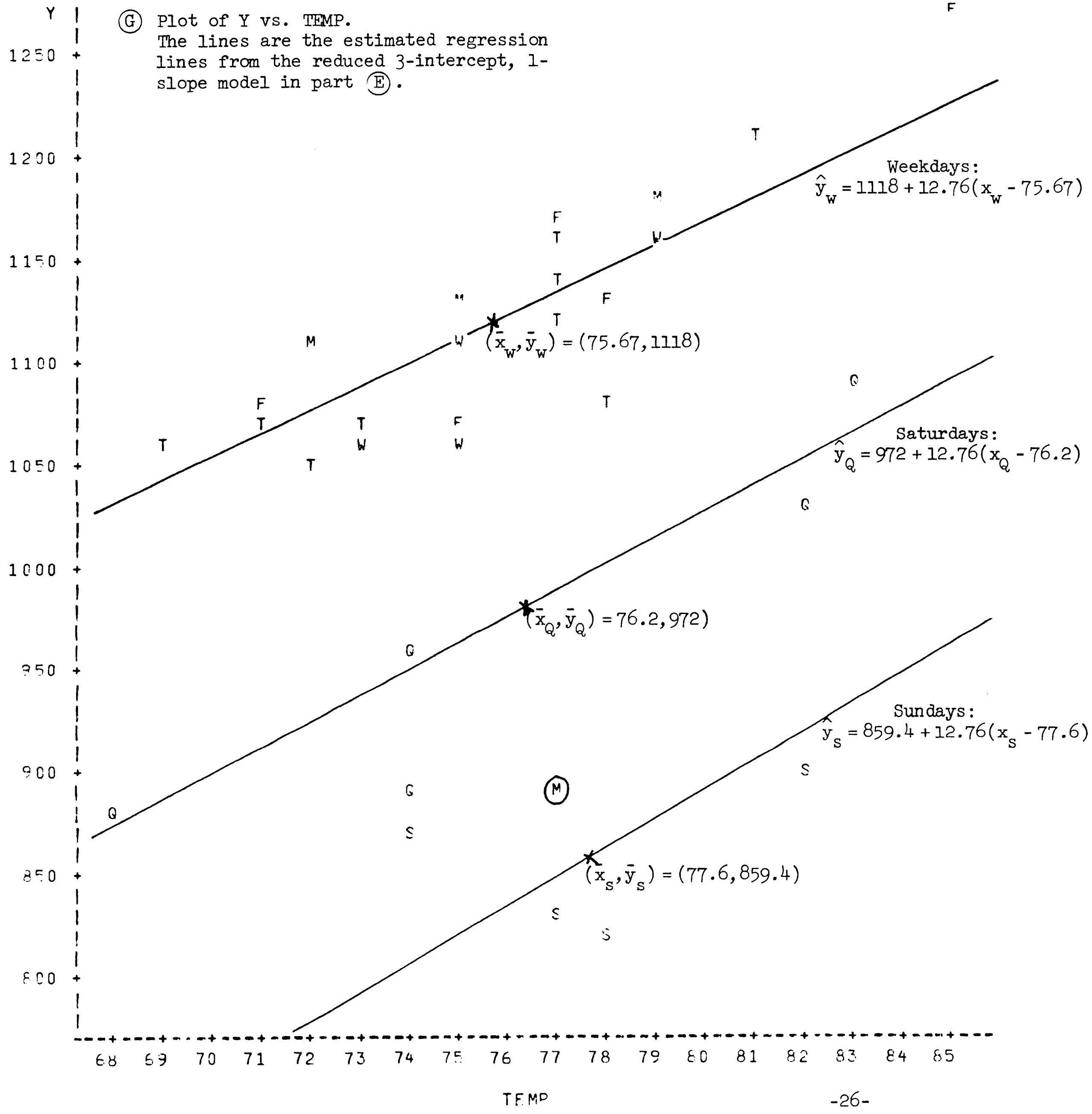
Ⓓ Residual plot for the full model

PLOT OF RESID2*YHAT2   LEGEND: A = 1 OBS, B = 2 OBS, ETC.



Ⓕ Residual plot for the reduced model

PLOT OF Y*TEMP   SYMBOL IS VALUE OF DAY



Ⓖ Plot of Y vs. TEMP.
The lines are the estimated regression lines from the reduced 3-intercept, 1-slope model in part Ⓔ.

LEAFHOPPER DATA – Unit 16

```
TITLE LEAFHOPPER DATA;
DATA LHOPPER;
INPUT TRTS $ DAYS;
```

DAYS is the response variable

$ indicates that TRTS is a non-numerical variable

```
CARDS;
CONTROL 2.3
CONTROL 1.7
SUCROSE 3.6
SUCROSE 4.0
GLUCOSE 3.0
GLUCOSE 2.8
FRUCTOSE 2.1
FRUCTOSE 2.3
```

The CLASS statement constructs the treatment indicator variables. CLASS is not available in PROC REG. Treatment indicators would have to be set up in the input statement as in the Electricity Load Data or the Soybean Physiological Data.

```
PROC GLM;
Ⓐ    CLASS TRTS;
     MODEL DAYS=TRTS/NOINT SOLUTION P XPX SSI;
          OUTPUT OUT=NEW PREDICTED=YHAT RESIDUAL=RESID;
     ESTIMATE 'CONTROL VS SUGARS'
          TRTS 3 -1 -1 -1 /DIVISOR=3 E;
     ESTIMATE '6-CARBONS VS SUCROSE'
          TRTS 0 -.5 -.5 1/E;
     ESTIMATE 'FRUCTOSE VS GLUCOSE'
          TRTS 0 -1 1 0/E;
     MEANS TRTS;
```

This model, following a CLASS statement, is the equivalent of a general means model. A SOLUTION option is needed after a CLASS statement so that the parameter estimates will be printed. XPX prints the $\underset{\sim}{X}'\underset{\sim}{X}$ matrix (see pp. 179-180).

ESTIMATE contrasts, p. 181

SAS orders levels of a classed variable alphabetically, or numerically, so the coefficients must be ordered: <u>C</u>ontrol, <u>F</u>ructose, <u>G</u>lucose, <u>S</u>ucrose.

← Calculates TRT means.

```
PROC PLOT;
Ⓑ    PLOT RESID*YHAT/VREF=0;
```

Residual plot for the general means model in Ⓐ

```
PROC GLM;
     CLASS TRTS;
     MODEL DAYS=TRTS/P XPX SSI;
     ESTIMATE 'CONTROL VS SUGARS'
          TRTS 3 -1 -1 -1 /DIVISOR=3 ;
     ESTIMATE '6-CARBONS VS SUCROSE'
          TRTS 0 -.5 -.5 1;
     ESTIMATE 'FRUCTOSE VS GLUCOSE'

           TRTS 0 -1 1 0;
```

```
                              LEAFHOPPER DATA

                       GENERAL LINEAR MODELS PROCEDURE

                          CLASS LEVEL INFORMATION

          CLASS      LEVELS     VALUES

          TRTS          4       CONTROL FRUCTOSE GLUCOSE SUCROSE

                  NUMBER OF OBSERVATIONS IN DATA SET = 8
```

The CLASS statement produces this output.

Note alphabetical ordering

With a CLASS statement, SAS creates indicator (dummy) variables and actually forms an X'X matrix as if indicator variables had been set up in the input statements.

```
                      LEAFHOPPER DATA

             GENERAL LINEAR MODELS PROCEDURE

              MATRIX ELEMENT REPRESENTATION

DEPENDENT VARIABLE: DAYS

               EFFECT                   REPRESENTATION


               TRTS      CONTROL        DUMMY001
                         FRUCTOSE       DUMMY002
                         GLUCOSE        DUMMY003
                         SUCROSE        DUMMY004
```

```
                           LEAFHOPPER DATA

                   GENERAL LINEAR MODELS PROCEDURE

                            THE X'X MATRIX

DEPENDENT VARIABLE: DAYS

                             DUMMY001      DUMMY002      DUMMY003      DUMMY004

   DUMMY001                 [ 2             0             0             0 ]
   DUMMY002                 [ 0             2             0             0 ]
   DUMMY003                 [ 0             0             2             0 ]
   DUMMY004                 [ 0             0             0             2 ]
```

```
ESTIMABLE FUNCTIONS FOR CONTROL VS SUGARS


EFFECT                       COEFFICIENTS


TRTS      CONTROL                    1
          FRUCTOSE            0.333333
          GLUCOSE             0.333333
          SUCROSE            -0.333333
```

Note the function of the DIVISOR = 3 option.

```
ESTIMABLE FUNCTIONS FOR 6-CARBONS VS SUCROSE


EFFECT                       COEFFICIENTS


TRTS      CONTROL                    0
          FRUCTOSE                -0.5
          GLUCOSE                 -0.5
          SUCROSE                    1
```

```
ESTIMABLE FUNCTIONS FOR FRUCTOSE VS GLUCOSE


EFFECT                       COEFFICIENTS


TRTS      CONTROL                    0
          FRUCTOSE                  -1
          GLUCOSE                    1
          SUCROSE                    0
```

The E option for each ESTIMATE statement prints the contrast vector (the c vector of cβ).

```
                  MEANS

TRTS                  N          DAYS

CONTROL               2    2.00000000
FRUCTOSE              2    2.20000000
GLUCOSE               2    2.90000000
SUCROSE               2    3.80000000
```

MEANS TRTS prints the treatment means. Compare to parameter estimates on next page.

GENERAL MEANS MODEL
Ⓐ Y = TRTS/NOINT

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: DAYS

SOURCE                  DF        SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                    4          63.38000000           15.84500000         211.27

ERROR                    4           0.30000000            0.07500000        PR > F

UNCORRECTED TOTAL        8          63.68000000                               0.0001

R-SQUARE              C.V.               STD DEV            DAYS MEAN

0.995289           10.0500            0.27386128           2.72500000

SOURCE                  DF              TYPE I SS    F VALUE     PR > F

TRTS                     4            63.38000000     211.27     0.0001 <- Ho:μc = μF = μc = μc = 0


                                                   T FOR H0:     PR > |T|      STD ERROR OF
PARAMETER     Sequentials from    ESTIMATE       PARAMETER=0                    ESTIMATE
              ABDO, p. 180

TRTS       CONTROL              2.00000000 = ȳc        10.33      0.0005        0.19364917
           FRUCTOSE             2.20000000 = ȳF        11.36      0.0003        0.19364917
Note order GLUCOSE              2.90000000 = ȳG        14.98      0.0001        0.19364917
           SUCROSE              3.80000000 = ȳS        19.62      0.0001        0.19364917

CONTROL VS SUGARS              -0.96666667             -4.32      0.0124        0.22360680
6-CARBONS VS SUCROSE            1.25000000              5.27      0.0062        0.23717082  Contrasts
FRUCTOSE VS GLUCOSE             0.70000000              2.56      0.0629        0.27386128

OBSERVATION           OBSERVED            PREDICTED              RESIDUAL
                        VALUE               VALUE

        1            2.30000000          2.00000000            0.30000000
        2            1.70000000          2.00000000           -0.30000000
        3            3.60000000          3.80000000           -0.20000000
        4            4.00000000          3.80000000            0.20000000
        5            3.00000000          2.90000000            0.10000000
        6            2.80000000          2.90000000           -0.10000000
        7            2.10000000          2.20000000           -0.10000000
        8            2.30000000          2.20000000            0.10000000
```

Control vs. Sugar contrast (see p. 181):

Estimate = -.9667 = 2.0 - .3333(2.2) - .3333(2.9) - .3333(3.8)

Standard Error= .2236 = SQRT[$\frac{\sigma^2}{2}(1+1/9+1/9+1/9)$]

GENERAL MEANS MODEL
Ⓒ Y = TRTS

```
                         LEAFHOPPER DATA

                GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: DAYS

SOURCE                   DF        SUM OF SQUARES          MEAN SQUARE       F VALUE

MODEL                     3            3.97500000           1.32500000         17.67

ERROR                     4            0.30000000           0.07500000       PR > F

CORRECTED TOTAL           7            4.27500000                            0.0090 ← H_0:μ_C = μ_F = μ_G = μ_S = μ

R-SQUARE               C.V.               STD DEV            DAYS MEAN

0.929825            10.0500            0.27386128           2.72500000


SOURCE                   DF               TYPE I SS   F VALUE     PR > F

TRTS                      3            3.97500000     17.67     0.0090

                                            T FOR H0:    PR > |T|    STD ERROR OF
PARAMETER                          ESTIMATE     PARAMETER=0                 ESTIMATE

CONTROL VS SUGARS               -0.96666667          -4.32      0.0124      0.22360680
6-CARBONS VS SUCROSE             1.25000000           5.27      0.0062      0.23717082
FRUCTOSE VS GLUCOSE              0.70000000           2.56      0.0629      0.27386128


OBSERVATION             OBSERVED              PREDICTED                RESIDUAL
                          VALUE                  VALUE

       1              2.30000000            2.00000000              0.30000000
       2              1.70000000            2.00000000             -0.30000000
       3              3.60000000            3.80000000             -0.20000000
       4              4.00000000            3.80000000              0.20000000
       5              3.00000000            2.90000000              0.10000000
       6              2.80000000            2.90000000             -0.10000000
       7              2.10000000            2.20000000             -0.10000000
       8              2.30000000            2.20000000              0.10000000

           SUM OF RESIDUALS                                          0.00000000
           SUM OF SQUARED RESIDUALS                                  0.30000000
           SUM OF SQUARED RESIDUALS - ERROR SS                      -0.00000000
           FIRST ORDER AUTOCORRELATION                              -0.20000000
           DURBIN-WATSON D                                           2.06666667
```

```
                    LEAFHOPPER DATA

           GENERAL LINEAR MODELS PROCEDURE

              CLASS LEVEL INFORMATION

CLASS      LEVELS     VALUES

TRTS            4     CONTROL FRUCTOSE GLUCOSE SUCROSE


     NUMBER OF OBSERVATIONS IN DATA SET = 8
```

LYMPHOCYTE DATA - Unit 17

	no stimulation	anti-IG stimulation
no ATP	treatment (0,0)	treatment (0,1)
ATP added	treatment (1,0)	treatment (1,1)

```
TITLE LYMPHOCYTE DATA;
DATA LYMPHCYT;
INPUT ATP IG Y;
CARDS
0 0 34
0 0 30
0 1 62.5
0 1 67.5
1 0 25.5
1 0 24.5
1 1 46.0
1 1 50.0
```

Presence of ATP or IG stimulation is denoted by a 1, otherwise a zero appears.

```
PROC GLM;
 Ⓐ  CLASSES ATP IG;
    MODEL Y= ATP IG ATP*IG/P;
         OUTPUT OUT= NEW RESIDUAL= RESID PREDICTED= YHAT;
```

Class on both ATP and IG

This model fits the main effects of ATP and IG, and their interaction, ATP * IG.

```
PROC GLM;
 Ⓑ  CLASSES ATP IG;
    MODEL Y= ATP*IG/NOINT SOLUTION SS1;
         ESTIMATE 'STIMULUS'
              ATP*IG 1 -1 1 -1/DIVISOR=2;
         ESTIMATE 'ATP PRESENCE'
              ATP*IG 1 1 -1 -1/DIVISOR=2;
         ESTIMATE 'INTERACTION'
              ATP*IG 1 -1 -1 1/DIVISOR=2;
PROC PLOT;
 Ⓒ  PLOT RESID*YHAT/VREF=0;
```

Refers to variables in CLASSES Statement.

Residual plot for Ⓐ and Ⓑ

Use the general means model to estimate contrasts. Contrasts among treatment means in part Ⓐ would be "non-estimable" because the main effects are fitted first.

The CLASSES statement orders the treatments numerically: (0,0); (0,1); (1,0); (1,1). The contrast coefficients must be in this order.

FACTORIAL ANALYSIS
Ⓐ Y = ATP IG ATP*IG

```
DEPENDENT VARIABLE: Y

SOURCE                    DF        SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                      3        1906.00000000         635.33333333          87.63

ERROR                      4          29.00000000           7.25000000        PR > F

CORRECTED TOTAL            7        1935.00000000                               0.0004


R-SQUARE                C.V.              STD DEV               Y MEAN

0.985013              6.3355           2.69258240          42.50000000


SOURCE                    DF             TYPE I SS    F VALUE     PR > F

ATP                        1         288.00000000       39.72     0.0032
IG                         1        1568.00000000      216.28     0.0001
ATP*IG                     1          50.00000000        6.90     0.0584


SOURCE                    DF            TYPE IV SS    F VALUE     PR > F

ATP                        1         288.00000000       39.72     0.0032
IG                         1        1568.00000000      216.28     0.0001
ATP*IG                     1          50.00000000        6.90     0.0584


OBSERVATION            OBSERVED             PREDICTED              RESIDUAL
                        VALUE                 VALUE

       1             34.00000000          32.00000000            2.00000000
       2             30.00000000          32.00000000           -2.00000000
       3             62.50000000          65.00000000           -2.50000000
       4             67.50000000          65.00000000            2.50000000
       5             25.50000000          25.00000000            0.50000000
       6             24.50000000          25.00000000           -0.50000000
       7             46.00000000          48.00000000           -2.00000000
       8             50.00000000          48.00000000            2.00000000

           SUM OF RESIDUALS                                      0.00000000
           SUM OF SQUARED RESIDUALS                             29.00000000
           SUM OF SQUARED RESIDUALS - ERROR SS                   0.00000000
           ~~FIRST ORDER AUTOCORRELATION~~                       ~~0.25000000~~
           ~~DURBIN-WATSON D~~                                   ~~2.22413793~~
```

Note that TYPE I and TYPE IV SS's
are the same due to orthogonality.
The SSI option could have been used.

GENERAL MEANS MODEL

(B) Y = ATP * IG/NOINT

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                  DF      SUM OF SQUARES         MEAN SQUARE        F VALUE

MODEL                    4      16356.00000000       4089.00000000         564.00

ERROR                    4         29.00000000          7.25000000 = s²    PR > F      s² same as for (A)

UNCORRECTED TOTAL        8      16385.00000000                             0.0001

R-SQUARE               C.V.            STD DEV              Y MEAN

0.998230             6.3355         2.69258240         42.50000000

SOURCE                  DF             TYPE I SS     F VALUE     PR > F

ATP*IG                   4        16356.00000000      564.00     0.0001

                                                T FOR H0:        PR > |T|      STD ERROR OF
PARAMETER                        ESTIMATE      PARAMETER=0                       ESTIMATE

ATP*IG      0 0             32.00000000 = ȳ00        16.81          0.0001        1.90394328
            0 1             65.00000000 = ȳ01        34.14          0.0001        1.90394328
            1 0             25.00000000 = ȳ10        13.13          0.0002        1.90394328
            1 1             48.00000000 = ȳ11        25.21          0.0001        1.90394328

STIMULUS                   -28.00000000 }           -14.71          0.0001        1.90394328 }
ATP PRESENCE                12.00000000 } p. 197      6.30          0.0032        1.90394328 }
INTERACTION                 -5.00000000 }            -2.63          0.0584        1.90394328 }
```

Compare significance levels to those of the TYPE I SS in part (A)

In (A) the contrasts are computed as part of the model. In (B) the contrasts are specified. More than three contrasts could have been run under the model. The contrasts do not need to be orthogonal for computation purposes.

FAT DIGESTIBILITY DATA - Unit 17; ACO, p. 365 (SAS)

```
TITLE FAT DIGESTIBILITY DATA;
DATA FAT_DIG;
INPUT PERIOD FAT $ LECITHIN Y;
X0=1; X1=0; X2=0; X4=0; X5=0; X6=0;
IF PERIOD=1 THEN X1=1;
IF PERIOD=2 THEN X2=1;
X3=1-X1-X2;
IF FAT='T' AND LECITHIN=0 THEN X4=1;
IF FAT='C' AND LECITHIN=0 THEN X5=1;
IF FAT='T' AND LECITHIN=1 THEN X6=1;
X7=1-X4-X5-X6;
CARDS;
```

Creating the indicator variables shown in Ⓐ

```
PROC PRINT;
Ⓐ  VAR X0-X7 Y;
```

Prints the $\underset{\sim}{X}|\underset{\sim}{Y}$ matrix for Ⓑ (see p. 201).

```
PROC GLM;
Ⓑ  MODEL Y=X0 X1-X7/NOINT;
      OUTPUT OUT=NEW RESIDUAL=RESID PREDICTED=YHAT;
   ESTIMATE 'W VS WO LECITHIN'
      X4 .5 X5 .5 X6 -.5 X7 -.5;
   ESTIMATE 'FAT DIFF WO LECITHIN'
      X4 1 X5 -1;
   ESTIMATE 'FAT DIFF W LECITHIN'
      X6 1 X7 -1;
```

Model: Equal means| Period Indicators| Treatment Indicators
and contrasts as discussed in Unit 17

```
PROC PLOT;
Ⓒ  PLOT RESID*YHAT/VREF=0;
```

Residual plot for Ⓑ

Ⓐ $\underset{\sim}{X}|\underset{\sim}{Y}$ matrix for model in Ⓑ

```
          Equal      Period Indicators      Treatment Indicators
          means
OBS        X0       X1      X2      X3      X4     X5     X6     X7        Y

  1         1        1       0       0       1      0      0      0      64.6
  2         1        0       1       0       1      0      0      0      52.4
  3         1        0       0       1       1      0      0      0      53.8
  4         1        1       0       0       0      1      0      0      66.0
  5         1        0       1       0       0      1      0      0      60.1
  6         1        0       0       1       0      1      0      0      64.4
  7         1        1       0       0       0      0      1      0      85.0
  8         1        0       1       0       0      0      1      0      68.9
  9         1        0       0       1       0      0      1      0      77.5
 10         1        1       0       0       0      0      0      1      96.0
 11         1        0       1       0       0      0      0      1      90.4
 12         1        0       0       1       0      0      0      1      98.2
```

MODEL SEQUENCE

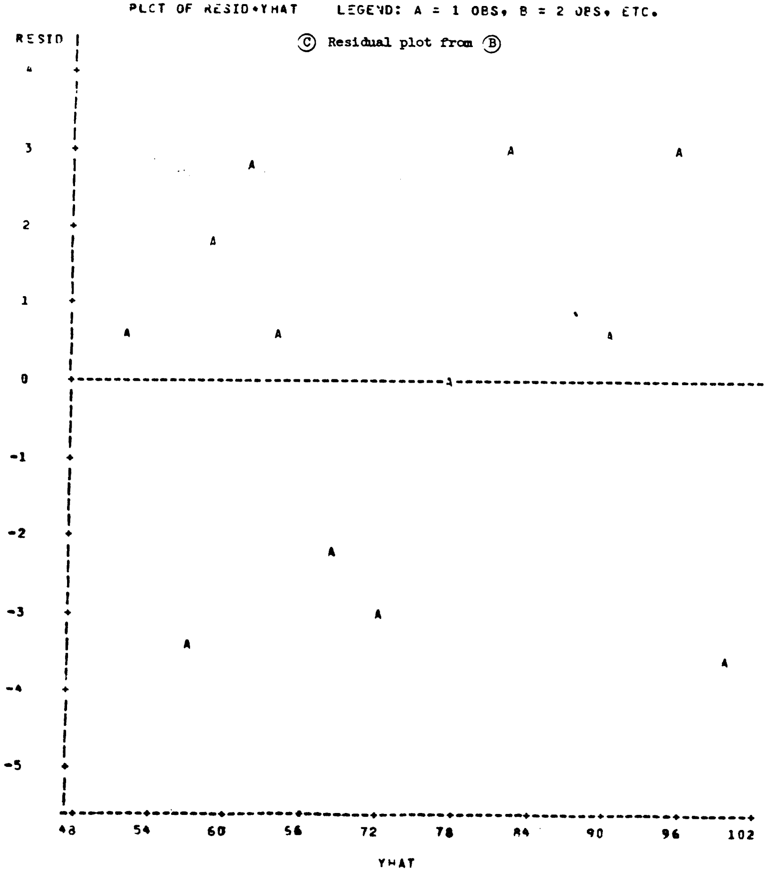
Ⓑ Y = Equal means | Periods | Treatments

```
                     GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                    DF        SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                      6        66867.48166667       11144.58027778         940.36

ERROR                      6           71.10833333          11.85138889        PR > F

UNCORRECTED TOTAL         12        66938.59000000                              0.0001


R-SQUARE              C.V.              STD DEV               Y MEAN

0.998938            4.7089           3.44258462          73.10833333


SOURCE                    DF               TYPE I SS     F VALUE      PR > F
                                                                               = 198.81 = R(1,2,3|0)
X0                         1    ny²=64137.94083333       5411.85      0.0001     Redundant
X1                         1   Period {137.76041667       11.62      0.0143
X2                         1     SS   {  61.05125000        5.15      0.0637
X3                         0          {   0.00000000          .           .
X4                         1   Treat-{1046.52250000       88.30      0.0001
X5                         1   ment  {1012.50000000       85.43      0.0001    = 2530.73
X6                         1    SS   { 471.70666667       39.80      0.0007    = R(4,5,6|0,1,2,3)
X7                         0              0.00000000          .           .
         X3 is redundant aftering fitting X0 X1 X2, therefore it does
         not account for any variability in the model.
SOURCE                    DF            TYPE IV SS     F VALUE      PR > F

X0                         0            0.00000000          .           .
X1                         0            0.00000000          .           .
X2                         0            0.00000000          .           .
X3                         0            0.00000000          .           .
X4                         0            0.00000000          .           .
X5                         0            0.00000000          .           .
X6                         0            0.00000000          .           .
X7                         0            0.00000000          .           .


                                                      T FOR H0:     PR > |T|    STD ERROR OF
PARAMETER                              ESTIMATE     PARAMETER=0                    ESTIMATE

X0      (ȳ4 - ȳ) + period 3 mean=95.23333333 B           39.12       0.0001      2.43427493
X1  period 1 mean - period 3 mean= 4.42500000 B            1.82       0.1190      2.43427493
X2  period 2 mean - period 3 mean=-5.52500000 B           -2.27       0.0637      2.43427493
X3                                 0.00000000 B              .           .               .
X4                   ȳ1 - ȳ4 = -37.93333333 B           -13.50       0.0001      2.81085857
X5                   ȳ2 - ȳ4 = -31.36666667 B           -11.16       0.0001      2.81085857
X6                   ȳ3 - ȳ4 = -17.73333333 B            -6.31       0.0007      2.81085857
X7                                 0.00000000 B              .           .               .
```

No matter which variable is fitted last, it will have 0 df and no sum of squares since it will be a linear combination of variables previously fitted. This redundancy causes the following message to appear. (See also note on following page.)

In general, the partial coefficients are not of interest but contrasts are appropriate (see next page).

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y


NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED
      INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS.
      THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE
      SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY
      THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER
      BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS
      (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS
      MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE
      FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT
      OF THE BIASED ESTIMATOR AND THE T VALUE TESTS
      HO: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE
      LETTER B ARE BLUE FOR THE PARAMETER.

                                              T FOR HO:    PR > |T|   STD ERROR OF
PARAMETER                       ESTIMATE     PARAMETER=0                 ESTIMATE
 Contrasts:
W VS WO LECITHIN            -25.78333333         -12.97      0.0001      1.98757716
FAT DIFF WO LECITHIN         -6.56666667          -2.34      0.0581      2.81085857
FAT DIFF W LECITHIN         -17.73333333          -6.31      0.0007      2.81085857
```



PROTEIN NUTRITION DATA – Unit 18

```
TITLE NUTRITION DATA;
DATA PROTEIN;
INPUT PROTEIN $ Y;
CARDS;
H 179
H 136
H 160
H 227
H 217
H 168      10 horsebean observations
H 108
H 124
H 143
H 140
L 309
L 229
L 181
L 141
L 260
L 203      12 linseed observations
L 148
L 169
L 213
L 257
L 244
L 271
S 243
S 230
S 248
S 327
S 329
S 250
S 193
S 271      14 soybean observations
S 316
S 267
S 199
S 177
S 158
S 248
```

Ⓐ

```
PROC GLM;
     CLASS PROTEIN;
     MODEL Y=PROTEIN/NOINT SOLUTION P SS1;
               OUTPUT OUT=NEW PREDICTED=YHAT RESIDUAL=RESID;
     ESTIMATE 'HORSEBEAN VS OILMEAL'  See p. 217
          PROTEIN 1 -.5 -.5;          Natural Contrasts
     ESTIMATE 'LINSEED VS SOYBEAN'
               PROTEIN 0 1 -1;        See p. 219
     ESTIMATE 'ORTHO H-B VS OILMEAL'  Ortho Contrasts
               PROTEIN 13 -6 -7/DIVISOR=13;
```

Using the **CLASS** statement to fit the general means model, followed by natural and orthogonal contrasts, as discussed in Unit 18.

Ⓑ

```
PROC PLOT;
     PLOT RESID*YHAT/VREF=0;
```

Residual plot for Ⓐ

GENERAL MEANS MODEL
Ⓐ Y = PROTEIN/NOINT

```
DEPENDENT VARIABLE: Y

SOURCE                  DF       SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                    3     1683997.43571429       561332.47857143        229.66

ERROR                   33       80659.56428571         2444.22922078        PR > F

UNCORRECTED TOTAL       36     1764657.00000000                              0.0001

R-SQUARE              C.V.             STD DEV               Y MEAN

0.954292           23.1656         49.43914664         213.41666667

SOURCE                  DF             TYPE I SS    F VALUE     PR > F ← Tests:  μH = μL = μS = 0  not  μH = μL = μS = μ

PROTEIN                  3      1683997.43571429     229.66     0.0001

                                                  T FOR H0:    PR > |T|     STD ERROR OF
PARAMETER                          ESTIMATE     PARAMETER=0                   ESTIMATE

PROTEIN             H         160.20000000 = ȳH      10.25     0.0001       15.63403090
                    L         218.75000000 = ȳL      15.33     0.0001       14.27185231
                    S         246.85714286 = ȳS      18.68     0.0001       13.21316773
     Contrasts:
HORSEBEAN VS OILMEAL          -72.60357143          -3.94     0.0004       18.41171677 ←In this case, the natural and ORTHO
LINSEED VS SOYBEAN            -28.10714286          -1.45     0.1578       19.44925628   contrasts give similar results.
ORTHO H-B VS OILMEAL          -73.68461538          -4.01     0.0003       18.39651430 ←

OBSERVATION          OBSERVED            PREDICTED              RESIDUAL          OBSERVED            PREDICTED              RESIDUAL
                       VALUE               VALUE                                    VALUE               VALUE

     1            179.00000000        160.20000000           18.80000000       244.00000000        218.75000000           25.25000000
     2            136.00000000        160.20000000          -24.20000000       271.00000000        218.75000000           52.25000000
     3            160.00000000        160.20000000           -0.20000000       243.00000000        246.85714286           -3.85714286
     4            227.00000000        160.20000000           66.80000000       230.00000000        246.85714286          -16.85714286
     5            217.00000000        160.20000000           56.80000000       248.00000000        246.85714286            1.14285714
     6            168.00000000        160.20000000            7.80000000       327.00000000        246.85714286           80.14285714
     7            108.00000000        160.20000000          -52.20000000       329.00000000        246.85714286           82.14285714
     8            124.00000000        160.20000000          -36.20000000       250.00000000        246.85714286            3.14285714
     9            143.00000000        160.20000000          -17.20000000       193.00000000        246.85714286          -53.85714286
    10            140.00000000        160.20000000          -20.20000000       271.00000000        246.85714286           24.14285714
    11            309.00000000        218.75000000           90.25000000       316.00000000        246.85714286           69.14285714
    12            229.00000000        218.75000000           10.25000000       267.00000000        246.85714286           20.14285714
    13            181.00000000        218.75000000          -37.75000000       199.00000000        246.85714286          -47.85714286
    14            141.00000000        218.75000000          -77.75000000       177.00000000        246.85714286          -69.85714286
    15            260.00000000        218.75000000           41.25000000       158.00000000        246.85714286          -88.85714286
    16            203.00000000        218.75000000          -15.75000000       248.00000000        246.85714286            1.14285714
    17            148.00000000        218.75000000          -70.75000000
    18            169.00000000        218.75000000          -49.75000000     SUM OF RESIDUALS                                  0.00000000
    19            213.00000000        218.75000000           -5.75000000     SUM OF SQUARED RESIDUALS                      80659.56428571
    20            257.00000000        218.75000000           38.25000000     SUM OF SQUARED RESIDUALS - ERROR SS              -0.00000000
                                                                             FIRST ORDER AUTOCORRELATION                       0.34539716
                                                                             DURBIN-WATSON D                                   1.30480762
```

SWAMP pH DATA

Ⓐ General Means Model Analysis Using Model Sequence

```
PROC GLM; CLASSES LOC TYPE;
   MODEL Y= LOC TYPE LOC*TYPE/P SS1 SS2;
   OUTPUT OUT=NEW1 PREDICTED=YHAT1 RESIDUAL=RESID1;
PROC PLOT;
   PLOT RESID1*YHAT1;
```

Objectives: (1) to fit the general means model with one model sequence so that the two "classical" ANOVA tables can be constructed;

(2) to examine residuals.

```
DEPENDENT VARIABLE: Y

SOURCE                DF      SUM OF SQUARES       MEAN SQUARE       F VALUE        PR > F        R-SQUARE          C

MODEL                  5          2.42538096        0.48507619          2.50        0.0536        0.300852        6.6

ERROR                 29          5.63633333        0.19435632                        STD DEV                     Y M

CORRECTED TOTAL       34          8.06171429                                       0.44085862      R(Mean) = (35)(6.66285)²
                                                                                                            = 1553.8

SOURCE                DF         TYPE I SS     F VALUE     PR > F        DF     TYPE II SS      F VALUE        PR

LOC                    1        {0.77487218        3.99     0.0553         1    1.33577182         6.87        0.0
TYPE|LOC               2  1.8176{1.04276977        2.68     0.0853         2    1.04276977         2.68        0.0
LOC*TYPE  (Interaction) 2        0.60773900        1.56     0.2266         2    0.60773900         1.56        0.2
                                                                  LOC|TYPE                      TYPE|LOC
```

ANOVA (1)

Source	df	SS
R(Mean)	1	1553.8
R(LOC)	1	0.7749
R(TYPE|LOC)	2	1.043
R(LOC*TYPE|LOC, TYPE|LOC)	2	0.6077
Residual	29	5.636

ANOVA (2)

Source	df	SS
R(Mean)	1	1553.8
R(TYPE)	2	1.8176 - 1.3358 = 0.4818
R(LOC| TYPE)	1	1.335
R(INTERACTION)	2	0.6077
Residual	29	5.636

If interaction is judged important ⇒ Ⓑ

If interaction is not judged to be important
⇒ both LOC and TYPE are needed (pg. 230) ⇒ Ⓒ or Ⓓ

Residual Analysis:

Not shown but the magnitude of the residuals is acceptable and the pattern is satisfactory (one or two values should be checked).

Ⓑ General Means Estimation

```
PROC GLM; CLASSES LOC TYPE;
   MODEL Y=LOC*TYPE/NOINT P CLM;

   ESTIMATE 'TEXT PG229 CONTRAST'
     LOC*TYPE 11 -6 -5 0 0 0/DIVISOR=11;
   ESTIMATE 'TEXT PG229 CONTRAST2'
     LOC*TYPE 0 1 -1 0 0 0;
   ESTIMATE 'TEXT PG229 CONTRAST3'
     LOC*TYPE 0 0 0 14 -8 -6/DIVISOR=14;
   ESTIMATE 'TEXT PG229 CONTRAST4'
     LOC*TYPE 0 0 0 0 1 -1;
```

Interaction is judged to be important

Objective: (1) to estimate the cell means from the general means model;

(2) to estimate column contrasts within each row (or row contrasts within each column).

```
DEPENDENT VARIABLE: Y

SOURCE                  DF      SUM OF SQUARES       MEAN SQUARE      F VALUE         PR > F        R-SQUARE          C

MODEL                    6      1556.20366667       259.36727778      1334.49         0.0001        0.996391        6.6

ERROR                   29         5.63633333         0.19435632                      STD DEV                       Y M

UNCORRECTED TOTAL       35      1561.84000000                                       0.44085862                  6.66285

SOURCE                  DF         TYPE I SS      F VALUE      PR > F        DF       TYPE IV SS       F VALUE        PR

LOC*TYPE                 6      1556.20366667     1334.49      0.0001         6     1556.20366667     1334.49        0.0

                                                 T FOR H0:      PR > |T|        STD ERROR OF
PARAMETER                          ESTIMATE     PARAMETER=0                       ESTIMATE

TEXT PG229 CONTRAST1             0.51590909          2.52        0.0176          0.20484945
TEXT PG229 CONTRAST2             0.34666667          1.30        0.2043          0.26695315
TEXT PG229 CONTRAST3             0.08571429          0.26        0.7988          0.33325779
TEXT PG229 CONTRAST4            -0.15000000         -0.63        0.5336          0.23809087
```

Interpretation: (1) Within the near location, community type North has a higher (0.5) pH than the average of the other two community types (p = .02).

(2) There is some evidence that the pH for the Mesic Community is higher than Shrub Community (p = .20).

(3) Within the away location, the pH for community types does not vary.

```
OBSERVATION          OBSERVED        PREDICTED        RESIDUAL       LOWER 95% CL     UPPER 95% CL
                      VALUE            VALUE                          FOR MEAN         FOR MEAN

     1              6.60000000      6.82500000      -0.22500000      6.50621858       7.14378142
     2              7.20000000      6.82500000       0.37500000      6.50621858       7.14378142
     3              7.20000000      6.82500000       0.37500000      6.50621858       7.14378142
     4              7.00000000      6.82500000       0.17500000      6.50621858       7.14378142
     5              6.80000000      6.82500000      -0.02500000      6.50621858       7.14378142
     6              6.40000000      6.82500000      -0.42500000      6.50621858       7.14378142
     7              7.00000000      6.82500000       0.17500000      6.50621858       7.14378142
     8              6.40000000      6.82500000      -0.42500000      6.50621858       7.14378142
     9              6.80000000      6.46666667       0.33333333      6.09856959       6.83476374
    10              7.00000000      6.46666667       0.53333333      6.09856959       6.83476374
    11              6.20000000      6.46666667      -0.26666667      6.09856959       6.83476374
    12              6.20000000      6.46666667      -0.26666667      6.09856959       6.83476374
    13              6.40000000      6.46666667      -0.06666667      6.09856959       6.83476374
    14              6.20000000      6.46666667      -0.26666667      6.09856959       6.83476374
    15              6.40000000      6.12000000       0.28000000      5.71676985       6.52323015
    16         →    5.20000000      6.12000000      -0.92000000      5.71676985       6.52323015
    17              6.20000000      6.12000000       0.08000000      5.71676985       6.52323015
    18              6.40000000      6.12000000       0.28000000      5.71676985       6.52323015
    19              6.40000000      6.12000000       0.28000000      5.71676985       6.52323015
    20              6.80000000      6.90000000      -0.10000000      6.26243716       7.53756284
    21              7.00000000      6.90000000       0.10000000      6.26243716       7.53756284
    22              6.20000000      6.75000000      -0.55000000      6.43121858       7.06878142
    23         →    5.60000000      6.75000000      -1.15000000      6.43121858       7.06878142
    24              7.20000000      6.75000000       0.45000000      6.43121858       7.06878142
    25              7.20000000      6.75000000       0.45000000      6.43121858       7.06878142
    26              7.20000000      6.75000000       0.45000000      6.43121858       7.06878142
    27              6.20000000      6.75000000      -0.55000000      6.43121858       7.06878142
    28              7.20000000      6.75000000       0.45000000      6.43121858       7.06878142
    29              7.20000000      6.75000000       0.45000000      6.43121858       7.06878142
    30              7.20000000      6.90000000       0.30000000      6.53190292       7.26809708
    31              6.80000000      6.90000000      -0.10000000      6.53190292       7.26809708
    32              7.00000000      6.90000000       0.10000000      6.53190292       7.26809708
    33              6.80000000      6.90000000      -0.10000000      6.53190292       7.26809708
    34              7.00000000      6.90000000       0.10000000      6.53190292       7.26809708
    35              6.60000000      6.90000000      -0.30000000      6.53190292       7.26809708
```

Predicted values are the cell means estimated from the general means model

	TYPE		
LOC	North	Mesic	Shrub
Near	6.825	6.467	6.12
Away	6.90	6.75	6.90

The last two columns give 95% confidence intervals on the population cell means

Ⓒ General Means Estimation (with No Interaction)

```
PROC REG DATA=SWAMP;
   MODEL Y=X1-X6/NOINT P CLM;
   RESTRICT X2-X3-X5+X6=0;
   RESTRICT X1-X3-X4+X6=0;
   NC: TEST X1-.5*X2-.5*X3+X4-.5*X5-.5*X6=0;
   NC2: TEST X2-X3+X5-X6=0;
  OC1: TEST .725773*X1-.395676*X2-.329897*X3+.274227*X4-.156701*X5
           -.117526*X6=0;
  OC2: TEST .00535834*X1+.440054*X2-.445412*X3-.00535834*X4
         +.559946*X5-.554588*X6=0;
   NR1: TEST X1+X2+X3-X4-X5-X6=0;
   OR1: TEST .206*X1+.442*X2+.352*X3-.206*X4-.442*X5-.352*X6=0;
```

```
DEP VARIABLE: Y
                        SUM OF          MEAN
SOURCE     DF          SQUARES        SQUARE       F VALUE       PROB>F
MODEL       4         1555.596       388.899      1930.770       0.0001
ERROR      31         6.244072      0.201422
U TOTAL    35         1561.840
     ROOT MSE         0.448800      R-SQUARE        0.9960
     DEP MEAN         6.662857      ADJ R-SQ        0.9956
     C.V.             6.735853
NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

                    PARAMETER      STANDARD    T FOR H0:
VARIABLE   DF        ESTIMATE         ERROR  PARAMETER=0   PROB > |T|

X1          1        6.756993      0.145537       46.428       0.0001
X2          1        6.391427      0.151220       42.266       0.0001
X3          1        6.319083      0.161362       39.161       0.0001
X4          1        7.172003      0.191737       37.405       0.0001
X5          1        6.806430      0.138410       49.176       0.0001
X6          1        6.734092      0.153873       43.764       0.0001
RESTRICT   -1        0.451440             .            .       1.0000
RESTRICT   -1        0.544005             .            .       1.0000
TEST: NC            NUMERATOR:     1.02702   DF:     1   F VALUE:    5.0989
                    DENOMINATOR: 0.201422    DF:    31   PROB >F :   0.0311
TEST: NC2           NUMERATOR:    .0322163   DF:     1   F VALUE:    0.1599
                    DENOMINATOR: 0.201422    DF:    31   PROB >F :   0.6919
TEST: OC1           NUMERATOR:     1.01055   DF:     1   F VALUE:    5.0171
                    DENOMINATOR: 0.201422    DF:    31   PROB >F :   0.0324
TEST: OC2           NUMERATOR:    .0322162   DF:     1   F VALUE:    0.1599
                    DENOMINATOR: 0.201422    DF:    31   PROB >F :   0.6920
TEST: NR1           NUMERATOR:     1.33577   DF:     1   F VALUE:    6.6317
                    DENOMINATOR: 0.201422    DF:    31   PROB >F :   0.0150
TEST: OR1           NUMERATOR:     1.33577   DF:     1   F VALUE:    6.6317
                    DENOMINATOR: 0.201422    DF:    31   PROB >F :   0.0150
```

Situation: the interaction is judged not to be important and both factors are needed in the model.

Objective: (1) to estimate the cell means for the restricted general means model (the two RESTRICT statements of no interaction forces any contrast among the three columns to be the same for each row or the difference between row means will be the same for all columns).

(2) to estimate contrasts (natural or ortho) among (i) the column cell means averaged over rows and (ii) row cell means averaged over columns.

Parameter estimates are the cell means estimated from the restricted general means model

	Community		
LOC	North	Mesic	Shrub
near	6.757	6.391	6.319
away	7.172	6.806	6.734

Note that the difference between rows is the same for each column (any contrast among columns is the same for each row).

Confidence intervals on the restricted population cell means are calculated with the CLM option.

Ⓓ General Means Estimation (no interaction) with Reduced Model

```
PROC GLM; CLASSES LOC TYPE;
   MODEL Y=LOC TYPE/P SS1 SS2;
   MEANS LOC TYPE;
   LSMEANS LOC TYPE;
   ESTIMATE 'LOC NAT MAIN EFFECT'
     LOC 1 -1;
   ESTIMATE 'TYPE NAT MAIN EFFECT1'
     TYPE 1 -.5 -.5;
   ESTIMATE 'TYPE NAT MAIN EFFECT2'
     TYPE 0 1 -1;
   ESTIMATE 'TYPE ORTHO MAIN EFFECT1'
TYPE 1 -.552577 -.447423;
```

```
DEPENDENT VARIABLE: Y

SOURCE                 DF      SUM OF SQUARES        MEAN SQUARE       F VALUE        PR > F      R-SQUARE           C

MODEL                   3          1.81764195         0.60588065          3.01        0.0451      0.225466         6.7

ERROR                  31          6.24407234         0.20142169                     STD DEV                       Y M

CORRECTED TOTAL        34          8.06171429                                     0.44880028                   6.66285


SOURCE                 DF          TYPE I SS     F VALUE     PR > F        DF            TYPE II SS      F VALUE      PR

LOC                     1         0.77487218        3.85     0.0589         1            1.33577182         6.63     0.0
TYPE                    2         1.04276977        2.59     0.0913         2            1.04276977         2.59     0.0

                                                  T FOR H0:      PR > |T|          STD ERROR OF
PARAMETER                          ESTIMATE     PARAMETER=0                          ESTIMATE

LOC NAT MAIN EFFECT             -0.41500335          -2.58        0.0150          0.16115307
TYPE NAT MAIN EFFECT             0.40174146           2.26        0.0311          0.17791409
TYPE NAT MAIN EFFECT             0.07233758           0.40        0.6919          0.18087522
TYPE ORTHO MAIN EFFE             0.39793817           2.24        0.0324          0.17765975

OBSERVATION           OBSERVED          PREDICTED           RESIDUAL
                        VALUE             VALUE

     1             6.60000000        6.75699933        -0.15699933        18      6.40000000      6.31908908       0.08091092
     2             7.20000000        6.75699933         0.44300067        19      6.40000000      6.31908908       0.08091092
     3             7.20000000        6.75699933         0.44300067        20      6.80000000      7.17200268      -0.37200268
     4             7.00000000        6.75699933         0.24300067        21      7.00000000      7.17200268      -0.17200268
     5             6.80000000        6.75699933         0.04300067        22      6.20000000      6.80643001      -0.60643001
     6             6.40000000        6.75699933        -0.35699933        23      5.60000000      6.80643001      -1.20643001
     7             7.00000000        6.75699933         0.24300067        24      7.20000000      6.80643001       0.39356999
     8             6.40000000        6.75699933        -0.35699933        25      7.20000000      6.80643001       0.39356999
     9             6.80000000        6.39142666         0.40857334        26      7.20000000      6.80643001       0.39356999
    10             7.00000000        6.39142666         0.60857334        27      6.20000000      6.80643001      -0.60643001
    11             6.20000000        6.39142666        -0.19142666        28      7.20000000      6.80643001       0.39356999
    12             6.20000000        6.39142666        -0.19142666        29      7.20000000      6.80643001       0.39356999
    13             6.40000000        6.39142666         0.00857334        30      7.20000000      6.73409243       0.46590757
    14             6.20000000        6.39142666        -0.19142666        31      6.80000000      6.73409243       0.06590757
    15             6.40000000        6.31908908         0.08091092        32      7.00000000      6.73409243       0.26590757
    16             5.20000000        6.31908908        -1.11908908        33      6.80000000      6.73409243       0.06590757
    17             6.20000000        6.31908908        -0.11908908        34      7.00000000      6.73409243       0.26590757
                                                                         35      6.60000000      6.73409243      -0.13409243
```

Note: Results from ESTIMATE statements are as in ⓒ

```
              MEANS

LOC           N              Y

 1            19      6.52631579 = [6.825(8) + 6.467(6) + 6.12(5)]/19     (weighted average of observations)
 2            16      6.82500000

TYPE          N              Y

 1            10      6.84000000 = [6.825(8) + 6.9(2)]/10
 2            14      6.62857143
 3            11      6.54545455


       LEAST SQUARES MEANS

        LOC                Y
                      LSMEAN

         1          6.48917169 = [6.757 + 6.391 + 6.319]/3     (unweighted average of estimated cell means)
         2          6.90417504

        TYPE               Y
                      LSMEAN

         1          6.96450100  =  [6.757 + 7.172]/2
         2          6.59892833
         3          6.52659076
```

Note: $6.904 - 6.489 = -0.415$

$6.965 - [6.599 + 6.527]/2 = 0.4017$

$6.599 - 6.527 = 0.072$

$6.965 - [(.552577)(6.599) + (.447423)(6.527)] = 0.398$

SWAMP DATA - Unweighted Analysis of Cell Means
see Snedecor & Cochran, 7 ed., p. 418

```
DATA SWAMP;
INPUT LOC TYPE TMT Y;
CARDS;
```

Ⓔ

```
PROC ANOVA; CLASS TMT;
     MODEL  Y = TMT;
     MEANS TMT / DEPONLY;
```

Ⓔ The six treatment combinations are indicated in the CLASS variable TMT.
The residual MS is an estimate of σ^2 .

```
PROC SORT; BY TYPE LOC;
```

Ⓕ

```
PROC MEANS MEAN NOPRINT; BY TYPE LOC;
     OUTPUT OUT=NEW MEAN=MY; VAR Y;
```

Ⓕ The six cell means are computed and placed
in the data set NEW.

Ⓖ

```
PROC ANOVA; CLASS TYPE LOC;
     MODEL  MY = LOC TYPE LOC*TYPE;
     MEANS LOC TYPE LOC*TYPE / DEPONLY;
```

Ⓖ The main effect SS and interaction SS are determined at this step.

Ⓔ One-way ANOVA on the six treatment combinations.

```
              ANALYSIS OF VARIANCE PROCEDURE

                 CLASS LEVEL INFORMATION

              CLASS      LEVELS    VALUES

              TMT           6      1 2 3 4 5 6

          NUMBER OF OBSERVATIONS IN DATA SET = 35


DEPENDENT VARIABLE: Y

SOURCE                  DF      SUM OF SQUARES       MEAN SQUARE     F VALUE

MODEL                    5          2.42538095        0.48507619        2.50

ERROR                   29          5.63633333        0.19435632      PR > F
                                                       = s^2
CORRECTED TOTAL         34          8.06171429                        0.0536

R-SQUARE              C.V.             STD DEV            Y MEAN

0.300852            6.6167          0.44085862        6.66285714

SOURCE                  DF           ANOVA SS    F VALUE    PR > F

TMT                      5         2.42538095       2.50    0.0536



                             MEANS are the cell means

                    TMT            N           Y

                     1             8      6.82500000
                     2             6      6.46666667
                     3             5      6.12000000
                     4             2      6.90000000
                     5             8      6.75000000
                     6             6      6.90000000
```

Note: s^2 is all that is to be used from this ANOVA table along with the associated error df = 29 .

Note: Since $\frac{\max(n_{ij})}{\min(n_{ij})} = \frac{8}{2} > 2$, the analysis of unweighted cell means is of dubious worth.

Calculate: $\frac{1}{n_h} = \frac{1}{2(3)}\left(\frac{1}{8}+\frac{1}{6}+\frac{1}{5}+\frac{1}{2}+\frac{1}{8}+\frac{1}{6}\right) = 0.21389$.   Thus, $n_h = 4.675$

Ⓖ There is no output from Ⓕ.

```
                    ANALYSIS OF VARIANCE PROCEDURE

DEPENDENT VARIABLE: MY

SOURCE                 DF        SUM OF SQUARES          MEAN SQUARE

MODEL                   5            0.47950231           0.09590046

ERROR                   0            0.00000000           0.00000000

CORRECTED TOTAL         5            0.47950231

R-SQUARE             C.V.               STD DEV              MY MEAN

1.000000           0.0000            0.00000000           6.66027778

SOURCE                 DF              ANOVA SS    F VALUE    PR > F

LOC                     1            0.21596713          .         .
TYPE                    2            0.13235093          .         .
TYPE*LOC                2            0.13118426          .         .
```

```
          ANALYSIS OF VARIANCE PROCEDURE

            CLASS LEVEL INFORMATION

          CLASS      LEVELS     VALUES

          TYPE            3     1 2 3

          LOC             2     1 2

      NUMBER OF OBSERVATIONS IN DATA SET = 6
```

= the number of cell means.

Type 1 = North
2 = Mesic
3 = Shrub

Loc 1 = Near
2 = Away

These are the desired SS which need to be multiplied by $n_h = 4.675$, the harmonic mean number of observations per cell.

ANOVA of Unweighted Cell Means

Source	df	SS	MS	F
LOC	1	1.00965	1.00965	5.19
TYPE	2	0.61874	0.30937	1.59
TYPE*LOC	2	0.61329	0.30664	1.58
ERROR	29	5.63633	0.19436	

$F^{.05}_{1,29} = 4.18$ $F^{.05}_{2,29} = 3.33$

$F^{.01}_{1,29} = 7.60$ $F^{.25}_{2,29} = 1.45$

where:

LOC SS = 4.675(0.21596713) = 1.00965
TYPE SS = 4.675(0.13235093) = 0.61874
TYPE*LOC SS = 4.675(0.13118426) = 0.61329
Error SS from part Ⓔ

Ⓖ continued

```
ANALYSIS OF VARIANCE PROCEDURE

               MEANS

LOC            N             MY

1              3     6.47055556
2              3     6.85000000


TYPE           N             MY

1              2     6.86250000
2              2     6.60833333
3              2     6.51000000
```

These are unweighted means, e.g. $6.471 = \frac{1}{3}(6.825 + 6.467 + 6.120)$.

Thus, they correspond to the LSMEANS from the model which includes LOC, TYPE and LOC*TYPE.

Note: The N are not appropriate on this page.

```
TYPE   LOC         N             MY

1      1           1     6.82500000
1      2           1     6.90000000
2      1           1     6.46666667
2      2           1     6.75000000
3      1           1     6.12000000
3      2           1     6.90000000
```

These are the six cell means -- compare with part Ⓔ

The standard error of any location mean is $\sqrt{\frac{s^2}{3(n_h)}} = \sqrt{\frac{0.19436}{3(4.675)}} = 0.1177$

The standard error of any type mean is $\sqrt{\frac{0.19436}{2(4.675)}} = 0.1442$

See p. 419 of Snedecor and Cochran (7 ed.) for a discussion of calculating the correct standard errors of comparisons among row means, column means and individual cell means.

SOYBEAN DATA — Unit 19; ACO, p. 373 (MINITAB)

```
/*LIMITS REGION=300K
/*FORMAT A,FORMS=024,DEST=WR2
//  EXEC SAS
OPTIONS LS=80 NODATE NONOTES;
DATA SOYBEAN;
TITLE SOYBEAN DATA;
INPUT LIGHT $ HEIGHT YIELD;
DEV = HEIGHT;
X1 = (LIGHT='C');
X2 = (LIGHT='L');
X3 = 1-X1-X2;
```

} X1 = (LIGHT = 'C') is equivalent to: IF LIGHT = 'C' THEN X1 = 1;
ELSE X1 = 0;

X1, X2, X3 are treatment indicators, HEIGHT is the covariate, YIELD is the response variable

```
CARDS;
C 48 12.2
C 52 12.4
C 42 11.9
C 35 11.3
C 40 11.8
C 48 12.1
C 60 13.1
C 61 12.7
C 50 12.4
C 33 11.4
C 48 12.3
C 51 12.2
C 56 12.6
C 65 13.2
C 51 12.3
L 63 16.6
L 50 15.8
L 63 16.5
L 33 15.0
L 38 15.4
L 45 15.6
L 50 15.8
L 46 15.8
L 50 16.0
L 49 15.8
L 35 15.0
L 50 16.2
L 62 16.7
L 49 15.8
L 52 15.9
S 52  9.5
S 54  9.5
S 53  9.6
S 45  8.8
S 57  9.5
S 62  9.8
S 52  9.1
S 67 10.3
S 55  9.5
S 40  8.5
S 41  8.6
S 67 10.4
S 55  9.4
```

Ⓐ
```
PROC STANDARD MEAN=0; VAR DEV;
```
← Sets mean of DEV = 0. It is equivalent to HEIGHT - $\overline{\text{HEIGHT}}$.

Ⓑ
```
PROC PRINT;
     VAR X1 X2 X3 DEV;
```
Prints the $\underset{\sim}{X}$ matrix for Ⓒ.

Ⓒ
```
PROC REG;
     MODEL YIELD=X1 X2 X3 DEV/NOINT SECF SS1 SS2;
     OUTPUT OUT=NEW PREDICTED=YHAT RESIDUAL=RESID;
     C_VS_TRT:  TEST X1-.5*X2-.5*X3=0;
     L_VS_S:  TEST X2-X3=0;
```
Model fitting the covariate deviations last.

Use TEST statements with PROC REG to test contrasts. Instead of specifying the coefficients as with ESTIMATE, use the equation that represents the contrasts under the null hypothesis.

Ⓓ
```
PROC PLOT DATA=NEW;
     PLOT RESID*YHAT=LIGHT/VREF=0;
     PLOT YIELD*HEIGHT=LIGHT;
```
Residual plot for Ⓒ.

Plot of the response variable vs. the covariate.

Ⓔ
```
PROC GLM;
     MODEL YIELD=HEIGHT X1 X2 X3;
```
Model fitting the covariate first.

Ⓕ
```
PROC GLM;
     CLASS LIGHT;
     MODEL YIELD=LIGHT HEIGHT LIGHT*HEIGHT;
```
Model fitting separate slopes after a common slope.

Ⓖ
```
PROC GLM; CLASS LIGHT;
     MODEL YIELD = LIGHT HEIGHT;
     LSMEANS LIGHT / STDERR;      ← Adjusted treatment means
     MEANS   LIGHT / DEPONLY;     ← Unadjusted treatment means
     ESTIMATE 'CONTROL VS TMT' HEIGHT 0 LIGHT 2 -1 -1 / DIVISOR=2;
     ESTIMATE 'LIGHT VS SHADE' HEIGHT 0 LIGHT 0  1 -1;
     ESTIMATE 'COMMON SLOPE  ' HEIGHT 1;
```

```
          SOYBEAN DATA

  OBS    X1    X2    X3     DEV

    1     1     0     0    -3.2
    2     1     0     0     0.8
    3     1     0     0    -9.2
    4     1     0     0   -16.2
    5     1     0     0   -11.2
    6     1     0     0    -3.2
    7     1     0     0     8.8
    8     1     0     0     9.8
    9     1     0     0    -1.2
   10     1     0     0   -18.2
   11     1     0     0    -3.2
   12     1     0     0    -0.2
   13     1     0     0     4.8
   14     1     0     0    13.8
   15     1     0     0    -0.2
   16     0     1     0    11.8
   17     0     1     0    -1.2
   18     0     1     0    11.8
   19     0     1     0   -18.2
   20     0     1     0   -13.2
   21     0     1     0    -6.2
   22     0     1     0    -1.2
   23     0     1     0    -3.2
   24     0     1     0    -1.2
   25     0     1     0    -2.2
   26     0     1     0   -16.2
   27     0     1     0    -1.2
   28     0     1     0    10.8
   29     0     1     0    -2.2
   30     0     1     0     0.8
   31     0     0     1     0.8
   32     0     0     1     2.8
   33     0     0     1     6.8
   34     0     0     1    -6.2
   35     0     0     1     5.8
   36     0     0     1    10.8
   37     0     0     1     0.8
   38     0     0     1    15.8
   39     0     0     1     3.8
   40     0     0     1   -11.2
   41     0     0     1   -10.2
   42     0     0     1    15.8
   43     0     0     1     3.8
   44     0     0     1    14.8
   45     0     0     1     4.8
```

Ⓑ The $\underset{\sim}{X}$ matrix for part Ⓒ

Control treatment indicator

Light treatment indicator

Shade treatment indicator

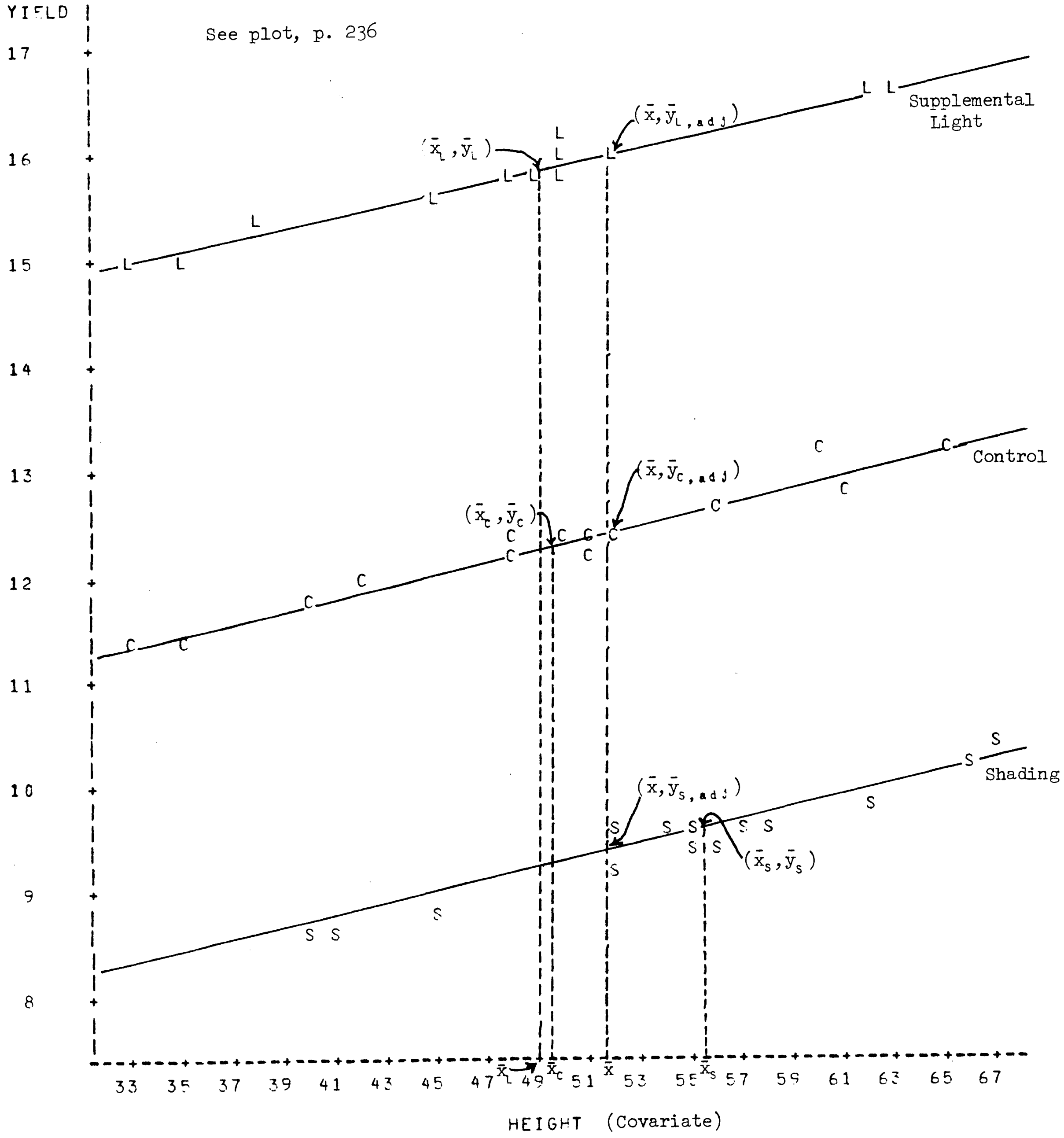
HEIGHT - $\overline{\text{HEIGHT}}$ (deviation of the covariate from its mean).

GENERAL MEANS MODEL
Ⓒ YIELD = Treatment indicators | covariate deviations

```
SEQUENTIAL PARAMETER ESTIMATES

X1          12.26 = ȳc
X2          12.26     15.86 = ȳL
X3          12.26     15.86    9.46667 = ȳs
DEV        12.369   15.9806    9.23709  .0583682 = common slope  ←  The first three partials are the adjusted treatment means (see p. 237).

DEP VARIABLE: YIELD

                         SUM OF           MEAN
SOURCE      DF          SQUARES          SQUARE         F VALUE        PROB>F

MODEL        4         7383.377        1845.844      110755.812        0.0001
ERROR       41         0.683301        0.016666
U TOTAL     45         7384.060

    ROOT MSE           0.129096        R-SQUARE          ~~0.9999~~
    DEP MEAN          12.528889        ADJ R-SQ          ~~0.9999~~
    C.V.                1.03039

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

                     PARAMETER        STANDARD     T FOR H0:
VARIABLE   DF         ESTIMATE           ERROR   PARAMETER=0     PROB > |T|      TYPE I SS

X1          1        12.368954       0.033592        368.213         0.0001       2254.614     ȳc (adj) = 12.37
X2          1        15.980628       0.033650        474.906         0.0001       3773.094     ȳL (adj) = 15.98
X3          1         9.237085       0.034469        267.984         0.0001       1344.267     ȳs (adj) =  9.24
DEV        1slope=0.058368     0.002231513         26.156         0.0001      11.402032

VARIABLE   DF     TYPE II SS

X1          1       2259.578
X2          1       3758.754
X3          1       1196.866
DEV         1      11.402032

TEST: C_VS_TRT     NUMERATOR:     0.562358   DF:     1     F VALUE:    33.7431
                   DENOMINATOR:  .0166659   DF:    41     PROB >F :    0.0001

TEST: L_VS_S       NUMERATOR:      315.604   DF:     1     F VALUE:  8937.1324
                   DENOMINATOR:  .0166659   DF:    41     PROB >F :    0.0001
```

TESTS of contrasts
These contrasts test differences between adjusted means.
The ESTIMATE option in GLM would provide the estimate and standard error (see Ⓖ).

Ⓓ PLOT OF YIELD*HEIGHT SYMBOL IS VALUE OF LIGHT



See plot, p. 236

NOTE: 5 OBS HIDDEN

Ⓔ Covariate/treatments model

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

SOURCE                  DF       SUM OF SQUARES          MEAN SQUARE       F VALUE

MODEL                    3         319.58914296         106.52971432       6392.08

ERROR                   41           0.68330149           0.01666589        PR > F

CORRECTED TOTAL         44         320.27244444                             0.0001


R-SQUARE              C.V.              STD DEV           YIELD MEAN

0.997866            1.0304           0.12909644          12.52888889


SOURCE                  DF          TYPE I SS    F VALUE     PR > F

HEIGHT                   1         1.77817807     106.70     0.0001
X1                       1         2.20679982     132.41     0.0001
X2                       1       315.60416506   18937.13     0.0001
X3                       0         0.00000000        .          .


SOURCE                  DF         TYPE IV SS    F VALUE     PR > F

HEIGHT                   1        11.40203184     684.15     0.0001
X1                       0         0.00000000        .          .
X2                       0         0.00000000        .          .
X3                       0         0.00000000        .          .


                                          T FOR H0:      PR > |T|       STD ERROR OF
PARAMETER                 ESTIMATE     PARAMETER=0                        ESTIMATE

INTERCEPT              6.24863365 B          49.02        0.0001        0.12746616
HEIGHT                 0.05836819            26.16        0.0001        0.00223151
X1                     3.13186885 B          64.07        0.0001        0.04888388
X2                     6.74354249 B         137.61        0.0001        0.04900394
X3                     0.00000000 B            .             .             .

NOTE: THE X'X MATRIX HAS BEEN DEEMED SINGULAR AND A GENERALIZED
      INVERSE HAS BEEN EMPLOYED TO SOLVE THE NORMAL EQUATIONS.
      THE ABOVE ESTIMATES REPRESENT ONLY ONE OF MANY POSSIBLE
      SOLUTIONS TO THE NORMAL EQUATIONS. ESTIMATES FOLLOWED BY
      THE LETTER B ARE BIASED AND DO NOT ESTIMATE THE PARAMETER
      BUT ARE BLUE FOR SOME LINEAR COMBINATION OF PARAMETERS
      (OR ARE ZERO). THE EXPECTED VALUE OF THE BIASED ESTIMATORS
      MAY BE OBTAINED FROM THE GENERAL FORM OF ESTIMABLE
      FUNCTIONS. FOR THE BIASED ESTIMATORS, THE STD ERR IS THAT
      OF THE BIASED ESTIMATOR AND THE T VALUE TESTS
      H0: E(BIASED ESTIMATOR) = 0. ESTIMATES NOT FOLLOWED BY THE
      LETTER B ARE BLUE FOR THE PARAMETER.
```

Produces the analysis of covariance table, p. 238.

Ⓕ Treatments/common slope/separate slopes model

```
                       SOYBEAN DATA

             GENERAL LINEAR MODELS PROCEDURE

                 CLASS LEVEL INFORMATION

                 CLASS     LEVELS    VALUES

                 LIGHT        3      C L S

        NUMBER OF OBSERVATIONS IN DATA SET = 45



                      SOYBEAN DATA

             GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

SOURCE                   DF      SUM OF SQUARES        MEAN SQUARE       F VALUE

MODEL                     5        319.66578021        63.93315604       4110.01

ERROR                    39          0.60666423         0.01555549        PR > F

CORRECTED TOTAL          44        320.27244444                           0.0001


R-SQUARE               C.V.             STD DEV         YIELD MEAN

0.998106             0.9955          0.12472166        12.52888889


SOURCE                   DF         TYPE I SS    F VALUE     PR > F

LIGHT                     2      308.18711111    9906.05     0.0001
HEIGHT                    1       11.40203184     732.99     0.0001
HEIGHT*LIGHT              2        0.07663726       2.46     0.0983
```

this line tests $H_0 : \beta_1 = \beta_2 = \beta_3$ $(=\beta)$
vs. H_a : different slopes needed.

```
SOURCE                   DF        TYPE IV SS    F VALUE     PR > F

LIGHT                     2       10.53027719     338.47     0.0001
HEIGHT                    1       11.47084949     737.41     0.0001
HEIGHT*LIGHT              2        0.07663726       2.46     0.0983
```

Ⓖ Treatments/common slope model

```
                         SOYBEAN DATA

                GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

SOURCE                  DF       SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                    3         319.58914296         106.52971432        6392.06

ERROR                   41           0.68330149           0.01666589        PR > F

CORRECTED TOTAL         44         320.27244444                              0.0001

R-SQUARE              C.V.              STD DEV           YIELD MEAN

0.997866            1.0304           0.12909644          12.52888889

SOURCE                  DF              TYPE I SS     F VALUE      PR > F

LIGHT  = R(L|μ)          2           308.18711111     9246.04      0.0001
HEIGHT = R(β|L,μ)        1            11.40203184      684.15      0.0001

SOURCE                  DF             TYPE IV SS     F VALUE      PR > F

LIGHT  = R(L|β,μ)        2           317.81096489     9534.77      0.0001
HEIGHT                   1            11.40203184      684.15      0.0001

                                           T FOR H0:        PR > |T|      STD ERROR OF
PARAMETER                   ESTIMATE     PARAMETER=0                        ESTIMATE

CONTROL VS TMT (1)       -0.23990239           -5.81          0.0001        0.04129927
LIGHT VS SHADE (2)        6.74354249          137.61          0.0001        0.04900394
COMMON SLOPE   (3)        0.05836819           26.16          0.0001        0.00223151

             MEANS = unadjusted treatment means

LIGHT          N          YIELD

C             15     12.2600000
L             15     15.8600000
S             15      9.4666667


               LEAST SQUARES MEANS = adjusted treatment means and their standard errors.

LIGHT          YIELD         STD ERR    PROB > |T|
              LSMEAN          LSMEAN    H0:LSMEAN=0

C         12.3689540       0.0335918         0.0001
L         15.9806276       0.0336501         0.0001
S          9.2370851       0.0344688         0.0001
```

$R(\beta|\mu) = R(L,\beta|\mu) - R(L|\beta,\mu)$

$= 319.58914 - 317.81096$

$= 1.77818$ = SS due to fitting a common slope before the treatments.

(1) and (2) are the contrasts among adjusted treatment means (same as the contrasts used in the TESTs of part Ⓒ).

(3) is the estimate of the common slope of the covariate.

Note that the F-value output for (2) in PROC REG is not correct.

Compare the MEANS and LSMEANS with the SEQB output of PROC REG in Ⓒ.

POTATO SCAB DATA - Comparison of regression lines in a 2 X 3 factorial experiment (two qualitative levels x three quantitative levels). P. 97, Cochran and Cox, 1957.

```
    DATA SCAB;
    INPUT YIELD TMT $ LEVEL X0 X01 X02 X1 X11 X12 X2 X21 X22;
    CARDS;
```

```
Ⓐ   PROC PRINT N;
```

```
Ⓑ   PROC REG;
        MODEL  YIELD = X0 X01 X02 X1 X11 X12 X2 X21 X22 /
                       NOINT P SEQB SS1 CLM;
```

Ⓑ and Ⓒ correspond to fitting a sequence of regression models

```
Ⓒ   PROC REG;
        MODEL  YIELD = X01 X02 X1 X2 / NOINT P SEQB SS1  CLM;
        OUTPUT OUT=NEW P=YHAT R=RES U95M=UPPER L95M=LOWER;
```

```
Ⓓ   PROC PLOT DATA=NEW;
        PLOT YIELD*LEVEL=TMT;
        PLOT RES*YHAT / VREF=0;
        PLOT YHAT*LEVEL='P' UPPER*LEVEL='U' LOWER*LEVEL='L' / OVERLAY;
```

A plot of the raw data as well as a residual plot and plot of predicted values for the model in part Ⓒ.

```
Ⓔ   PROC UNIVARIATE NORMAL PLOT DATA=NEW; VAR RES;
```

More analysis of residuals from the model in part Ⓒ.

```
Ⓕ   PROC GLM DATA=SCAB;  CLASS TMT LEVEL;
        MODEL  YIELD = TMT*LEVEL / NOINT P;
        ESTIMATE 'TMT'      TMT*LEVEL 1 1 1 -1 -1 -1  / DIVISOR=3;
        ESTIMATE 'B LIN'    TMT*LEVEL -4 -1 5 -4 -1 5 / DIVISOR=84;
        ESTIMATE 'T*B LIN'  TMT*LEVEL -4 -1 5 4 1 -5  / DIVISOR=42;
        ESTIMATE 'B QUAD'   TMT*LEVEL 2 -3 1 2 -3 1   / DIVISOR=108;
        ESTIMATE 'T*B QUAD' TMT*LEVEL 2 -3 1 -2 3 -1  / DIVISOR=54;
```

Ⓕ Fitting the cell means model and examining single degree-of-freedom contrasts which correspond to linear, quadratic, treatment and the respective interaction terms. The results are identical to those of part Ⓑ.

Recall that $b_{2.01} = \Sigma \ell_i y_i$ where the ℓ_i's may be computed from, say, the ORTHO algorithm. In this example $\ell_1 = \frac{2}{54}$, $\ell_2 = \frac{-3}{54}$ and $\ell_3 = \frac{1}{54}$. If the levels of the quantitative factor had been equally spaced, then the ℓ_i's could have been obtained from a table of orthogonal polynomial coefficients.

Ⓐ The data and indicator variables

```
OBS   YIELD   TMT   LEVEL   X0   X01   X02   X1   X11   X12    X2   X21   X22

  1      9     F      3     1     1     0     3     3     0     9     9     0
  2      9     F      3     1     1     0     3     3     0     9     9     0
  3     16     F      3     1     1     0     3     3     0     9     9     0
  4      4     F      3     1     1     0     3     3     0     9     9     0
  5     30     S      3     1     0     1     3     0     3     9     0     9
  6      7     S      3     1     0     1     3     0     3     9     0     9
  7     21     S      3     1     0     1     3     0     3     9     0     9
  8      9     S      3     1     0     1     3     0     3     9     0     9
  9     16     F      6     1     1     0     6     6     0    36    36     0
 10     10     F      6     1     1     0     6     6     0    36    36     0
 11     18     F      6     1     1     0     6     6     0    36    36     0
 12     18     F      6     1     1     0     6     6     0    36    36     0
 13     18     S      6     1     0     1     6     0     6    36     0    36
 14     24     S      6     1     0     1     6     0     6    36     0    36
 15     12     S      6     1     0     1     6     0     6    36     0    36
 16     19     S      6     1     0     1     6     0     6    36     0    36
 17     10     F     12     1     1     0    12    12     0   144   144     0
 18      4     F     12     1     1     0    12    12     0   144   144     0
 19      4     F     12     1     1     0    12    12     0   144   144     0
 20      5     F     12     1     1     0    12    12     0   144   144     0
 21     17     S     12     1     0     1    12     0    12   144     0   144
 22      7     S     12     1     0     1    12     0    12   144     0   144
 23     16     S     12     1     0     1    12     0    12   144     0   144
 24     17     S     12     1     0     1    12     0    12   144     0   144

N=24
```

Ⓑ Fitting the model sequence X0|X01 X02|X1|X11 X12|X2|X21 X22

which is: common mean|separate means|common linear|separate linear|common quadratic|separate quadratic

```
SEQUENTIAL PARAMETER ESTIMATES

X0         13.3333 = ȳ
X01        16.4167 -6.16667 = ȳF - ȳS
X02        16.4167 -6.16667  1.4E-14 = 0
X1         19.6458 -6.16667  1.7E-14 -0.46131  = b1.0 (common linear)
X11          18.75   -4.375  1.7E-14 -.333333 -.255952 = bF1.0 - bS1.0
X12          18.75   -4.375  1.7E-14 -.333333 -.255952  1.1E-13 = 0
X2         6.77083   -4.375  1.8E-14  3.77381 -.255952  1.6E-13 -.266204 = b2.01 (common quadratic)
X21        12.9167 -16.6667  3.3E-14  1.66667  3.95833 -2.2E-13 -0.12963 -.273148 = bF2.01 - bS2.01
X22        12.9167 -16.6667  3.3E-14  1.66667  3.95833 -2.2E-13 -0.12963 -.273148  1.2E-12 = 0

DEP VARIABLE: YIELD

                      SUM OF          MEAN
SOURCE       DF      SQUARES        SQUARE        F VALUE        PROB>F

MODEL         6     4721.000       786.833         22.374        0.0001
ERROR        18      633.000     35.166667
U TOTAL      24     5354.000
```

Standard errors and hypothesis tests are most easily determined from part Ⓕ.

```
    ROOT MSE       5.930149      ~~R-SQUARE         0.8818~~
    DEP MEAN      13.333333      ~~ADJ R-SQ         0.8489~~
    C.V.           44.47612

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.

NOTE: MODEL IS NOT FULL RANK. LEAST SQUARES SOLUTIONS FOR THE
      PARAMETERS ARE NOT UNIQUE. SOME STATISTICS WILL BE
      MISLEADING. A REPORTED DF OF 0 OR B MEANS THAT THE
      ESTIMATE IS BIASED. THE FOLLOWING PARAMETERS HAVE BEEN
      SET TO 0, SINCE THE VARIABLES ARE A LINEAR COMBINATION
      OF OTHER VARIABLES AS SHOWN.

X02       =+X0       -1*X01
X12       =+X1       -1*X11
X22       =+X2       -1*X21

                      PARAMETER      STANDARD   T FOR H0:
VARIABLE    DF         ESTIMATE         ERROR  PARAMETER=0    PROB > |T|    TYPE I SS

X0           B        12.916667      9.932877        1.300        0.2099     4266.667
X01          B       -16.666667     14.047209       -1.186        0.2509      228.167
X02          0                0             .            .             .            0
X1           B         1.666667      3.202646        0.520        0.6091    71.502976
X11          B         3.958333      4.529226        0.874        0.3937     5.502976
X12          0                0             .            .             .            0
X2           B        -0.129630      0.205450       -0.631        0.5360      118.080
X21          B        -0.273148      0.290550       -0.940        0.3596    31.080357
```

} not of use.

Ⓑ Results of the P and CIM option

```
                 PREDICT  STD ERR LOWER95% UPPER95%
OBS    ACTUAL      VALUE  PREDICT     MEAN     MEAN RESIDUAL
              (Cell Means)
  1     9.000      9.500    2.965    3.271   15.729 -.500000
  2     9.000      9.500    2.965    3.271   15.729 -.500000
  3    16.000      9.500    2.965    3.271   15.729    6.500
  4     4.000      9.500    2.965    3.271   15.729   -5.500
  5    30.000     16.750    2.965   10.521   22.979   13.250
  6     7.000     16.750    2.965   10.521   22.979   -9.750
  7    21.000     16.750    2.965   10.521   22.979    4.250
  8     9.000     16.750    2.965   10.521   22.979   -7.750
  9    16.000     15.500    2.965    9.271   21.729 0.500000
 10    10.000     15.500    2.965    9.271   21.729   -5.500
 11    18.000     15.500    2.965    9.271   21.729    2.500
 12    18.000     15.500    2.965    9.271   21.729    2.500
 13    18.000     18.250    2.965   12.021   24.479 -.250000
 14    24.000     18.250    2.965   12.021   24.479    5.750
 15    12.000     18.250    2.965   12.021   24.479   -6.250
 16    19.000     18.250    2.965   12.021   24.479 0.750000
 17    10.000      5.750    2.965 -.479346   11.979    4.250
 18     4.000      5.750    2.965 -.479346   11.979   -1.750
 19     4.000      5.750    2.965 -.479346   11.979   -1.750
 20     5.000      5.750    2.965 -.479346   11.979 -.750000
 21    17.000     14.250    2.965    8.021   20.479    2.750
 22     7.000     14.250    2.965    8.021   20.479   -7.250
 23    16.000     14.250    2.965    8.021   20.479    1.750
 24    17.000     14.250    2.965    8.021   20.479    2.750

SUM OF RESIDUALS                2.34479E-13
SUM OF SQUARED RESIDUALS                633
```

Ⓒ Fitting the model sequence XO1 XO2|X1|X2

which is: separate means|common linear|common quadratic

SEQUENTIAL PARAMETER ESTIMATES

XO1 10.25 = $\bar{y}_F$
XO2 10.25 16.4167 = $\bar{y}_S$
X1 13.4792 19.6458 -0.46131 = $b_{1.0}$
X2 1.5 7.66667 3.64583 -.266204 = $b_{2.01}$

```
DEP VARIABLE: YIELD

                     SUM OF          MEAN
SOURCE     DF       SQUARES        SQUARE        F VALUE        PROB>F

MODEL       4      4684.417      1171.104         34.980        0.0001
ERROR      20       669.583     33.479167
U TOTAL    24      5354.000

     ROOT MSE      5.786118      ~~R-SQUARE       0.8749~~
     DEP MEAN     13.333333      ~~ADJ R-SQ       0.8562~~
     C.V.          43.39589

NOTE: NO INTERCEPT TERM IS USED. R-SQUARE IS REDEFINED.


                PARAMETER      STANDARD    T FOR H0:
VARIABLE  DF     ESTIMATE         ERROR  PARAMETER=0    PROB > |T|     TYPE I SS

XO1        1     1.500000      6.954049        0.216        0.8314      1260.750
XO2        1     7.666667      6.954049        1.102        0.2833      3234.083   not useful
X1         1     3.645833      2.209610        1.650        0.1146     71.502976
X2         1    -0.266204      0.141747       -1.878        0.0750       118.080
```

```
                 PREDICT  STD ERR LOWER95% UPPER95%
 OBS   ACTUAL      VALUE  PREDICT     MEAN     MEAN RESIDUAL
             (=restricted cell means)
   1    9.000     10.042    2.362    5.114   14.969   -1.042
   2    9.000     10.042    2.362    5.114   14.969   -1.042
   3   16.000     10.042    2.362    5.114   14.969    5.958
   4    4.000     10.042    2.362    5.114   14.969   -6.042
   5   30.000     16.208    2.362   11.281   21.136   13.792
   6    7.000     16.208    2.362   11.281   21.136   -9.208
   7   21.000     16.208    2.362   11.281   21.136    4.792
   8    9.000     16.208    2.362   11.281   21.136   -7.208
   9   16.000     13.792    2.362    8.864   18.719    2.208
  10   10.000     13.792    2.362    8.864   18.719   -3.792
  11   18.000     13.792    2.362    8.864   18.719    4.208
  12   18.000     13.792    2.362    8.864   18.719    4.208
  13   18.000     19.958    2.362   15.031   24.886   -1.958
  14   24.000     19.958    2.362   15.031   24.886    4.042
  15   12.000     19.958    2.362   15.031   24.886   -7.958
```

```
                 PREDICT  STD ERR LOWER95% UPPER95%
 OBS   ACTUAL      VALUE  PREDICT     MEAN     MEAN RESIDUAL

  16   19.000     19.958    2.362   15.031   24.886 -.958333
  17   10.000      6.917    2.362    1.989   11.844    3.083
  18    4.000      6.917    2.362    1.989   11.844   -2.917
  19    4.000      6.917    2.362    1.989   11.844   -2.917
  20    5.000      6.917    2.362    1.989   11.844   -1.917
  21   17.000     13.083    2.362    8.156   18.011    3.917
  22    7.000     13.083    2.362    8.156   18.011   -6.083
  23   16.000     13.083    2.362    8.156   18.011    2.917
  24   17.000     13.083    2.362    8.156   18.011    3.917

SUM OF RESIDUALS             1.34115E-13
SUM OF SQUARED RESIDUALS        669.5833
```

```
                         GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

SOURCE                 DF      SUM OF SQUARES          MEAN SQUARE      F VALUE

MODEL                   6      4721.00000000          786.83333333        22.37

ERROR                  18       633.00000000           35.16666667      PR > F

UNCORRECTED TOTAL      24      5354.00000000                              0.0001

R-SQUARE             C.V.             STD DEV           YIELD MEAN

0.881771          44.4761          5.93014896          13.33333333

SOURCE                 DF          TYPE I SS    F VALUE     PR > F

TMT*LEVEL               6      4721.00000000      22.37     0.0001

SOURCE                 DF         TYPE IV SS    F VALUE     PR > F

TMT*LEVEL               6      4721.00000000      22.37     0.0001

                                       T FOR H0:        PR > |T|       STD ERROR OF
PARAMETER                 ESTIMATE    PARAMETER=0                        ESTIMATE

TMT                    -6.16666667          -2.55         0.0202         2.42097317
B LIN                  -0.46130952          -1.43         0.1710         0.32351615
T*B LIN                -0.25595238          -0.40         0.6971         0.64703230
B QUAD                 -0.26620370          -1.83         0.0835         0.14527499
T*B QUAD               -0.27314815          -0.94         0.3596         0.29054999

OBSERVATION               OBSERVED              PREDICTED             RESIDUAL
                            VALUE          VALUE (= cell means)

     1                  9.00000000           9.50000000          -0.50000000
     2                  9.00000000           9.50000000          -0.50000000
     3                 16.00000000           9.50000000           6.50000000
     4                  4.00000000           9.50000000          -5.50000000
     5                 30.00000000          16.75000000          13.25000000
     6                  7.00000000          16.75000000          -9.75000000
     7                 21.00000000          16.75000000           4.25000000
     8                  9.00000000          16.75000000          -7.75000000
     9                 16.00000000          15.50000000           0.50000000
    10                 10.00000000          15.50000000          -5.50000000
    11                 18.00000000          15.50000000           2.50000000
    12                 18.00000000          15.50000000           2.50000000
```

Ⓕ General Means Model

```
        GENERAL LINEAR MODELS PROCEDURE

            CLASS LEVEL INFORMATION

        CLASS       LEVELS      VALUES

        TMT              2      F S

        LEVEL            3      3 6 12

NUMBER OF OBSERVATIONS IN DATA SET = 24
```

Note that the parameter estimates are the same as those given in part Ⓑ under the SEQB output. Here we also have the standard errors and t-tests; however, we had to calculate the ℓ_i's for each contrast.

(Cell means)

TMT		LEVEL 3	6	12
	F	9.50	15.50	5.75
	S	16.75	18.25	14.25

The standard error of each cell mean is

$\sqrt{\frac{35.1667}{4}} = 2.965$. The cell means and

standard errors could have been printed by using the option:

LSMEANS TMT*LEVEL/STDERR;

Ⓕ continued

```
                         GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

OBSERVATION          OBSERVED            PREDICTED            RESIDUAL
                      VALUE                VALUE

     13           18.00000000         18.25000000         -0.25000000
     14           24.00000000         18.25000000          5.75000000
     15           12.00000000         18.25000000         -6.25000000
     16           19.00000000         18.25000000          0.75000000
     17           10.00000000          5.75000000          4.25000000
     18            4.00000000          5.75000000         -1.75000000
     19            4.00000000          5.75000000         -1.75000000
     20            5.00000000          5.75000000         -0.75000000
     21           17.00000000         14.25000000          2.75000000
     22            7.00000000         14.25000000         -7.25000000
     23           16.00000000         14.25000000          1.75000000
     24           17.00000000         14.25000000          2.75000000

          SUM OF RESIDUALS                                 0.00000000
          SUM OF SQUARED RESIDUALS                       633.00000000
          SUM OF SQUARED RESIDUALS - ERROR SS              0.00000000
          FIRST ORDER AUTOCORRELATION                     -0.63467615
          DURBIN-WATSON D                                  3.25701027
```

Note: In order to get the classical ANOVA table as well as the single degree-of-freedom contrasts and means with standard errors we could have used the statements:

```
PROC GLM; CLASS TMT LEVEL;
     MODEL YIELD = TMT LEVEL TMT*LEVEL/P CLM;
    [ same ESTIMATE statements]
     LSMEANS TMT LEVEL TMT*LEVEL/STDERR;
```

Note: To obtain the same results as in Ⓒ we would fit the model with interaction restricted to be zero That is,

```
PROC GLM; CLASS TMT LEVEL;
     MODEL YIELD = TMT LEVEL /P CLM;
```

ALCOHOL-DRUG DATA, p. 280.

Analysis of a Split-Unit Experiment

Y1 = Drug A response
Y2 = Drug B response
Y3 = Control response

$$YS = \frac{Y_1+Y_2+Y_3}{\sqrt{3}}$$

```
DATA WHOLE;
INPUT Y1 Y2 Y3;
SUBJECT = _N_;
ALCOHOL = 'YES';
IF _N_ > 6 THEN ALCOHOL='NO';
YS = (Y1+Y2+Y3)/SQRT(3);
XO = 1;
CARDS;


PROC SORT; BY ALCOHOL;
```

Ⓐ

```
PROC PRINT N; BY ALCOHOL;
```

Ⓑ

```
PROC GLM; CLASS ALCOHOL;
     MODEL  YS = XO ALCOHOL / NOINT;
     LSMEANS ALCOHOL / STDERR;
     ESTIMATE 'DIFFERENCE' ALCOHOL 1 -1;
```

Ⓒ

```
DATA SPLIT; SET WHOLE;
     Y=Y1; DRUG='A'; OUTPUT;
     Y=Y2; DRUG='B'; OUTPUT;
     Y=Y3; DRUG='C'; OUTPUT;
     DROP Y1-Y3 YS;
```

Rearranging the data to enable analysis of the subplot factor.

Ⓓ

```
PROC PRINT N;


PROC SORT; BY ALCOHOL SUBJECT;
```

PROC SORT must be used on the CLASS variables used in the ABSORB statement below. ABSORBing the whole plot factors reduces the storage requirements and hence the time and cost of the analysis.

Ⓔ

```
PROC GLM;
     ABSORB ALCOHOL SUBJECT;
     CLASS DRUG ALCOHOL;
     MODEL  Y = DRUG ALCOHOL*DRUG;
     ESTIMATE 'MAIN EFFECT: AB VS C' DRUG -1 -1 2 / DIVISOR=2;
     ESTIMATE 'MAIN EFFECT: A VS B ' DRUG  1 -1 0;
```

Effects model for the subplot factor and main effects contrasts.

Ⓕ

```
PROC GLM;
     ABSORB ALCOHOL SUBJECT;
```

```
      CLASSES ALCOHOL DRUG;
      MODEL  Y = ALCOHOL*DRUG / NOINT SS1 SS2;
      ESTIMATE 'AB VS CONTROL @ NO'
               ALCOHOL*DRUG -1 -1 2 0 0 0 / DIVISOR=2 E;
      ESTIMATE 'A VS B @ NO ALCOHOL'
               ALCOHOL*DRUG 1 -1 0 0 0 0;
      ESTIMATE 'AB VS CONTROL @ YES'
               ALCOHOL*DRUG 0 0 0 -1 -1 2 / DIVISOR=2;
      ESTIMATE 'A VS B @ YES ALCOHOL'
               ALCOHOL*DRUG 0 0 0 1 -1 0;
```

General means model and simple effects contrasts.

```
Ⓖ  PROC GLM; CLASSES SUBJECT ALCOHOL DRUG;
      MODEL  Y = XO ALCOHOL SUBJECT(ALCOHOL)
                 DRUG ALCOHOL*DRUG / SS1 SS2 NOINT;
      TEST H=ALCOHOL E=SUBJECT(ALCOHOL) / HTYPE=1 ETYPE=1;
      CONTRAST 'ALCOHOL DIFFERENCE' ALCOHOL 1 -1 /
               E=SUBJECT(ALCOHOL) ETYPE=1;
      ESTIMATE 'AB VS CONTROL @ NO' DRUG -1 -1 2
               ALCOHOL*DRUG -1 -1 2 0 0 0 / DIVISOR=2 E;
      ESTIMATE 'A VS B @ NO ALCOHOL' DRUG 1 -1 0
               ALCOHOL*DRUG  1 -1 0 0 0 0;
      ESTIMATE 'AB VS CONTROL @ YES' DRUG -1 -1 2
               ALCOHOL*DRUG 0 0 0 -1 -1 2 / DIVISOR=2;
      ESTIMATE 'A VS B @ YES ALCOHOL' DRUG 1 -1 0
               ALCOHOL*DRUG 0 0 0 1 -1 0;
      OUTPUT OUT=NEW2 P=P R=R;
```

This analysis performs both the whole-plot and sub-plot analyses all at one time. The simple effects contrasts are more difficult to specify and the computing costs are much steeper.

```
Ⓗ  PROC PLOT; PLOT R*P='*' / VREF=0;
```

Analysis of residuals

```
Ⓘ  PROC ANOVA; CLASSES SUBJECT ALCOHOL DRUG;
      MODEL  Y = ALCOHOL SUBJECT(ALCOHOL) DRUG ALCOHOL*DRUG;
      TEST H=ALCOHOL E=SUBJECT(ALCOHOL);
      MEANS ALCOHOL SUBJECT(ALCOHOL) DRUG ALCOHOL*DRUG / DEPONLY;
```

PROC ANOVA may be used since this experiment is balanced. The ESTIMATE option is not available, but the cell means and SS for the ANOVA table are computed and F-tests made.

```
--------------------------------- ALCOHOL=NO ---------------------------------

          OBS      Y1      Y2      Y3     SUBJECT       YS       XO

           1      2.83    2.55    2.63       7       4.62458     1
           2      2.93    2.42    2.73       8       4.66499     1
           3      3.58    3.99    3.38       9       6.32199     1
           4      2.98    3.07    2.78      10       5.09800     1
           5      2.32    2.15    2.12      11       3.80474     1
           6      2.73    3.23    2.53      12       4.90170     1

          N=6

--------------------------------- ALCOHOL=YES --------------------------------

          OBS      Y1      Y2      Y3     SUBJECT       YS       XO

            7     3.56    4.04    3.26       1       6.27002     1
            8     3.79    3.88    3.49       2       6.44323     1
            9     4.09    5.32    3.79       3       7.62102     1
           10     3.10    4.38    2.80       4       5.93516     1
           11     3.33    3.63    3.03       5       5.76773     1
           12     3.35    3.63    3.05       6       5.79082     1

          N=6
```

Ⓐ See p. 283 of Allen and Cady for a discussion of the assumed model.

Ⓑ Analysis of sums to test for whole-plot (alcohol) differences.

```
                    GENERAL LINEAR MODELS PROCEDURE

                       CLASS LEVEL INFORMATION

                      CLASS      LEVELS     VALUES

                      ALCOHOL       2       NO YES

               NUMBER OF OBSERVATIONS IN DATA SET = 12


DEPENDENT VARIABLE: YS

SOURCE                 DF      SUM OF SQUARES        MEAN SQUARE      F VALUE

MODEL                   2        382.70960556       191.35480278       328.00

ERROR                  10          5.83396111         0.58339611     PR > F
                                             = SS due to SUBJECT(ALCOHOL).
UNCORRECTED TOTAL      12        388.54356667                          0.0001

R-SQUARE             C.V.             STD DEV              YS MEAN

0.984985          13.6304          0.76380371           5.60366549

SOURCE                 DF           TYPE I SS   F VALUE     PR > F

XO                      1        376.81280278    645.90     0.0001
ALCOHOL                 1          5.89680278     10.11     0.0098

SOURCE                 DF          TYPE IV SS   F VALUE     PR > F

XO                      0          0.00000000      .          .
ALCOHOL                 1          5.89680278     10.11     0.0098

                                  T FOR H0:       PR > |T|      STD ERROR OF
PARAMETER             ESTIMATE  PARAMETER=0                      ESTIMATE

DIFFERENCE         -1.40199890        -3.18      0.0098        0.44098228


                          LEAST SQUARES MEANS

           ALCOHOL             YS        STD ERR    PROB > |T|
                           LSMEAN         LSMEAN   H0:LSMEAN=0

           NO          4.90266604     0.31182156        0.0001
           YES         6.30466494     0.31182156        0.0001
```

Compare SS found on this page with those in, say, analysis Ⓕ.

```
Ⓓ    OBS    SUBJECT    ALCOHOL    XO      Y     DRUG

       1       7          NO       1    2.83     A
       2       7          NO       1    2.55     B
       3       7          NO       1    2.63     C
       4       8          NO       1    2.93     A
       5       8          NO       1    2.42     B
       6       8          NO       1    2.73     C
       7       9          NO       1    3.58     A
       8       9          NO       1    3.99     B
       9       9          NO       1    3.38     C
      10      10          NO       1    2.98     A
      11      10          NO       1    3.07     B
      12      10          NO       1    2.78     C
      13      11          NO       1    2.32     A
      14      11          NO       1    2.15     B
      15      11          NO       1    2.12     C
      16      12          NO       1    2.73     A
      17      12          NO       1    3.23     B
      18      12          NO       1    2.53     C
      19       1          YES      1    3.56     A
      20       1          YES      1    4.04     B
      21       1          YES      1    3.26     C
      22       2          YES      1    3.79     A
      23       2          YES      1    3.88     B
      24       2          YES      1    3.49     C
      25       3          YES      1    4.09     A
      26       3          YES      1    5.32     B
      27       3          YES      1    3.79     C
      28       4          YES      1    3.10     A
      29       4          YES      1    4.38     B
      30       4          YES      1    2.80     C
      31       5          YES      1    3.33     A
      32       5          YES      1    3.63     B
      33       5          YES      1    3.03     C
      34       6          YES      1    3.35     A
      35       6          YES      1    3.63     B
      36       6          YES      1    3.05     C

     N=36
```

Ⓔ

```
                   GENERAL LINEAR MODELS PROCEDURE

                      CLASS LEVEL INFORMATION

                     CLASS       LEVELS     VALUES

                     DRUG           3       A B C

                     ALCOHOL        2       NO YES

              NUMBER OF OBSERVATIONS IN DATA SET = 36



                   GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                   DF       SUM OF SQUARES          MEAN SQUARE       F VALUE

MODEL                    15          14.47667500           0.96511167         13.66

ERROR                    20           1.41262222           0.07063111       PR > F

CORRECTED TOTAL          35          15.88929722                              0.0001


R-SQUARE               C.V.              STD DEV               Y MEAN

0.911096             8.2146           0.26576514           3.23527778


SOURCE                   DF              TYPE I SS    F VALUE     PR > F

ALCOHOL                   1             5.89680278    ~~83.49~~     ~~0.0001~~
SUBJECT(ALCOHOL)         10             5.83396111     ~~8.26~~     ~~0.0001~~
DRUG                      2             1.87722222      13.29     0.0002
DRUG*ALCOHOL              2             0.86868889       6.15     0.0083

SOURCE                   DF             TYPE IV SS    F VALUE     PR > F

DRUG                      2             1.87722222      13.29     0.0002
DRUG*ALCOHOL              2             0.86868889       6.15     0.0083

                                                     T FOR H0:    PR > |T|    STD ERROR OF
PARAMETER                            ESTIMATE      PARAMETER=0                  ESTIMATE

MAIN EFFECT: AB VS C              -0.40416667            -4.30      0.0003      0.09396217
MAIN EFFECT: A VS B               -0.30833333            -2.84      0.0101      0.10849817
```

Note that ALCOHOL and SUBJECT(ALCOHOL) are not included in the Type IV SS.

These contrasts are of little interest since the DRUG*ALCOHOL interaction is clearly important.

Ⓕ

General means model with whole-plot factor ABSORBed.

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                 DF      SUM OF SQUARES       MEAN SQUARE     F VALUE

MODEL                  15         14.47667500        0.96511167       13.66

ERROR                  20          1.41262222        0.07063111      PR > F

CORRECTED TOTAL        35         15.88929722                        0.0001


R-SQUARE             C.V.             STD DEV            Y MEAN

0.911096           8.2146          0.26576514        3.23527778


SOURCE                 DF         TYPE I SS   F VALUE    PR > F

ALCOHOL                 1        5.89680278     83.49    0.0001
SUBJECT(ALCOHOL)       10        5.83396111      8.26    0.0001
ALCOHOL*DRUG            4        2.74591111      9.72    0.0002


SOURCE                 DF        TYPE II SS   F VALUE    PR > F

ALCOHOL*DRUG            4        2.74591111      9.72    0.0002

                                           T FOR HO:    PR > |T|   STD ERROR OF
PARAMETER                  ESTIMATE      PARAMETER=0                  ESTIMATE

AB VS CONTROL @ NO      -0.20333333            -1.53      0.1416    0.13288257
A VS B @ NO ALCOHOL     -0.00666667            -0.04      0.9658    0.15343958
AB VS CONTROL @ YES     -0.60500000            -4.55      0.0002    0.13288257
A VS B @ YES ALCOHOL    -0.61000000            -3.98      0.0007    0.15343958
```

This is a set of simple effects contrasts
(see Table 22.3, p. 281 - Allen and Cady).

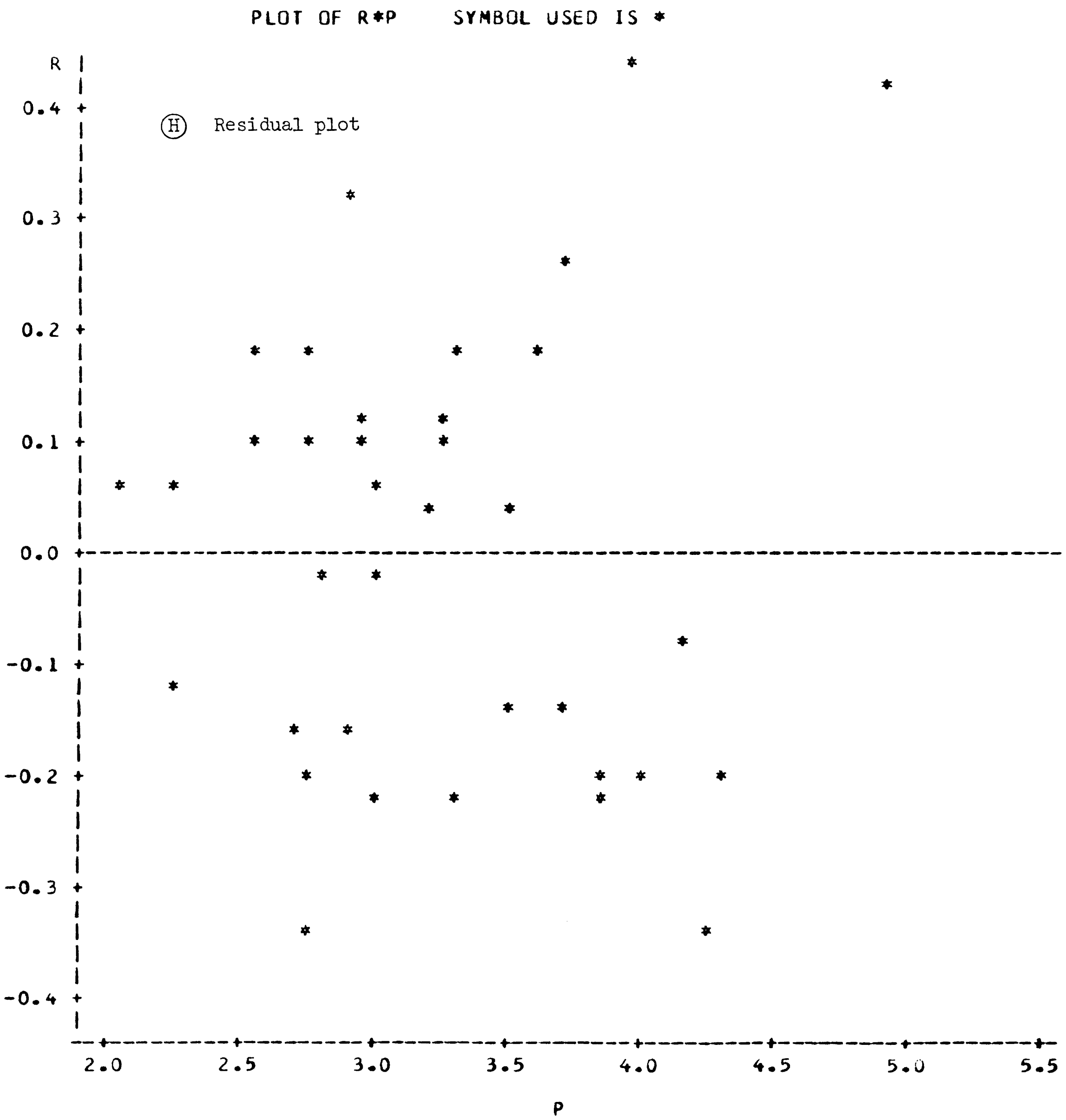
Ⓖ Combined whole-plot and split-plot analysis with simple effects contrasts.

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                 DF       SUM OF SQUARES          MEAN SQUARE      F VALUE

MODEL                  16         391.28947778          24.45559236       346.24

ERROR                  20           1.41262222           0.07063111       PR > F

UNCORRECTED TOTAL      36         392.70210000                            0.0001


R-SQUARE             C.V.              STD DEV               Y MEAN

0.996403           8.2146           0.26576514           3.23527778


SOURCE                 DF            TYPE I SS    F VALUE      PR > F

XO                      1         376.81280278    5334.94      ~~0.0001~~
ALCOHOL                 1           5.89680278      83.49      ~~0.0001~~
SUBJECT(ALCOHOL)       10           5.83396111       8.26      0.0001
DRUG                    2           1.87722222      13.29      0.0002
ALCOHOL*DRUG            2           0.86868889       6.15      0.0083


SOURCE                 DF           TYPE II SS    F VALUE      PR > F

XO                      0           0.00000000          .           .
ALCOHOL                 1           5.89680278      83.49      ~~0.0001~~
SUBJECT(ALCOHOL)       10           5.83396111       8.26      0.0001
DRUG                    2           1.87722222      13.29      0.0002
ALCOHOL*DRUG            2           0.86868889       6.15      0.0083


TESTS OF HYPOTHESES USING THE TYPE I MS FOR SUBJECT(ALCOHOL) AS AN ERROR TERM

SOURCE                 DF            TYPE I SS    F VALUE      PR > F

ALCOHOL                 1           5.89680278      10.11      0.0098

TESTS OF HYPOTHESES USING THE TYPE I MS FOR SUBJECT(ALCOHOL) AS AN ERROR TERM

CONTRAST               DF                   SS    F VALUE      PR > F

ALCOHOL DIFFERENCE      1           5.89680278      10.11      0.0098
```

Compare with the analysis in part Ⓑ.

```
                                              T FOR H0:     PR > |T|    STD ERROR OF
PARAMETER                       ESTIMATE    PARAMETER=0                   ESTIMATE

AB VS CONTROL @ NO           -0.20333333          -1.53      0.1416      0.13288257
A VS B @ NO ALCOHOL          -0.00666667          -0.04      0.9658      0.15343958
AB VS CONTROL @ YES          -0.60500000          -4.55      0.0002      0.13288257
A VS B @ YES ALCOHOL         -0.61000000          -3.98      0.0007      0.15343958
```

Simple effects as in Ⓕ.



Ⓗ Residual plot

Ⓘ

```
ANALYSIS OF VARIANCE PROCEDURE

            MEANS

ALCOHOL          N          Y

NO              18    2.83055556
YES             18    3.64000000

SUBJECT   ALCOHOL          N            Y

7         NO               3     2.67000000
8         NO               3     2.69333333
9         NO               3     3.65000000
10        NO               3     2.94333333
11        NO               3     2.19666667
12        NO               3     2.83000000
1         YES              3     3.62000000
2         YES              3     3.72000000
3         YES              3     4.40000000
4         YES              3     3.42666667
5         YES              3     3.33000000
6         YES              3     3.34333333

   DRUG          N            Y

   A            12    3.21583333
   B            12    3.52416667
   C            12    2.96583333

 ALCOHOL   DRUG          N            Y

 NO        A             6     2.89500000
 NO        B             6     2.90166667
 NO        C             6     2.69500000
 YES       A             6     3.53666667
 YES       B             6     4.14666667
 YES       C             6     3.23666667
```

Note: Only the MEANS output is included from the PROC ANOVA output. Although all the appropriate SS have been obtained via PROC GLM, PROC ANOVA would be less costly.

UREA SYNTHESIS DATA - Analysis of a repeated measures experiment

(Brogan & Kutner, 1980, The American Statistician 34:229-232)

```
DATA SHUNT;
INPUT PRE POST @@;
SUBJECT = _N_;
IF _N_ LE 8 THEN GROUP='NEW';
ELSE GROUP='OLD';
XO=1;
PREPOST = POST - PRE;
TMT     = POST + PRE;
TMT_X_PP= PREPOST;
IF GROUP='OLD' THEN PREPOST=-PREPOST;
CARDS;
```

These statements calculate the sums and differences required for the appropriate t-tests. See printout in part Ⓐ and Ⓑ.

```
PROC SORT; BY GROUP;
```

Ⓐ

```
PROC PRINT; BY GROUP;
```

Ⓑ

```
PROC TTEST; CLASS GROUP;
     VAR TMT PREPOST TMT_X_PP;
```

This gives the three appropriate t-tests for treatment effects, time effects and interaction.

Ⓒ

```
PROC GLM; CLASS GROUP;
     MODEL  TMT PREPOST TMT_X_PP = XO GROUP / NOINT;
     LSMEANS GROUP / STDERR;
     ESTIMATE 'NEW VS STD' GROUP 1 -1;
```

Same analysis as in Ⓑ except the analyses are performed using 1-way ANOVA. Compare with Ⓕ. Note that more than one dependent variable may appear to the left of the equal sign in the MODEL statement.

Ⓓ

```
DATA REPEAT; SET SHUNT;
Y= PRE; TIME=1; OUTPUT;
Y=POST; TIME=2; OUTPUT;
DROP PRE POST TMT PREPOST TMT_X_PP;
```

Rearrange data for a "split-plot" type analysis.

```
PROC PRINT; VAR SUBJECT XO GROUP TIME Y;
```

Ⓔ

```
PROC GLM; CLASS GROUP TIME SUBJECT;
     MODEL  Y = XO GROUP SUBJECT(GROUP)
                TIME TIME*GROUP / NOINT SS1 SS2 SS3 SS4 P;
     MEANS GROUP SUBJECT(GROUP) TIME TIME*GROUP / DEPONLY;
     TEST  H=GROUP E=SUBJECT(GROUP) / HTYPE=1 ETYPE=1;
     OUTPUT OUT=PLOT RESIDUAL=RES PREDICTED=P;
```

Combined ANOVA Table with two error terms.

Ⓕ

```
PROC PLOT; PLOT RES*P='*' / VREF=0;
```

Residual plot.

Ⓐ

```
---------------------------------- GROUP=NEW -----------------------------------

    OBS      PRE      POST     SUBJECT     XO     PREPOST     TMT     TMT_X_PP

     1       51        48         1         1        -3        99        -3
     2       35        55         2         1        20        90        20
     3       66        60         3         1        -6       126        -6
     4       40        35         4         1        -5        75        -5
     5       39        36         5         1        -3        75        -3
     6       46        43         6         1        -3        89        -3
     7       52        46         7         1        -6        98        -6
     8       42        54         8         1        12        96        12

---------------------------------- GROUP=OLD -----------------------------------

    OBS      PRE      POST     SUBJECT     XO     PREPOST     TMT     TMT_X_PP

     9       34        16         9         1        18        50       -18
    10       40        36        10         1         4        76        -4
    11       34        16        11         1        18        50       -18
    12       36        18        12         1        18        54       -18
    13       38        32        13         1         6        70        -6
    14       32        14        14         1        18        46       -18
    15       44        20        15         1        24        64       -24
    16       50        43        16         1         7        93        -7
    17       60        45        17         1        15       105       -15
    18       63        67        18         1        -4       130         4
    19       50        36        19         1        14        86       -14
    20       42        34        20         1         8        76        -8
    21       43        32        21         1        11        75       -11
```

PRE = preoperative response

POST = postoperative response

		Pre	Post
Treatment	New (n=8)	μ_{11}	μ_{12}
	Old (n=13)	μ_{21}	μ_{22}

Note that: TMT_X_PP = $y_{ij2} - y_{ij1} = d_{ij}$ (= within subject differences)

$$\text{PREPOST} = \begin{cases} d_{ij} & \text{if group} = \text{NEW} \\ -d_{ij} & \text{if group} = \text{OLD} \end{cases}$$

TMT = $y_{ij1} + y_{ij2} = z_{ij}$ (= within subject sums)

The assumed model may be written as:

$$y_{ijk} = \mu_{ik} + \delta_{j(i)} + \epsilon_{ijk} ,$$

$$\mu_{ik} = \mu + \alpha_i + \tau_k + (\alpha\tau)_{ik}$$

$$i = 1,2; \quad j = 1,\cdots,n_i; \quad k = 1,2$$

where:

μ_{ik} = cell mean of ith treatment, kth time combination

μ = overall mean

α_i = ith treatment effect

τ_k = kth time effect

$(\alpha\tau)_{ik}$ = treatment by time interaction

$\delta_{j(i)}$ = effect of subject j nested within ith treatment group

ϵ_{ijk} = random error

$\epsilon_{ijk} \sim N(0,\sigma_\epsilon^2)$

$\delta_{j(i)} \sim N(0,\sigma_\delta^2)$

$y_{ijk} \sim N(\mu_{ik},\sigma_\delta^2+\sigma_\epsilon^2)$.

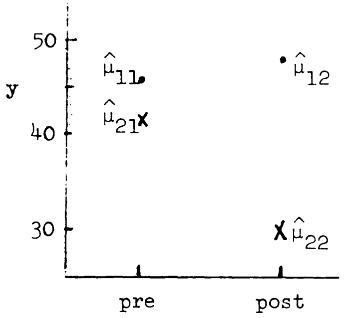
Ⓑ TTEST PROCEDURE

VARIABLE: TMT t-test uses $\bar{z}_1 - \bar{z}_2$ "difference of sums"

```
GROUP        N          MEAN       STD DEV     STD ERROR       MINIMUM        MAXIMUM

NEW          8   93.50000000   16.16875293    5.71651742   75.00000000   126.0000000
OLD         13   75.00000000   24.23839929    6.72252242   46.00000000   130.0000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL       2.0964       18.8       0.0498
EQUAL         1.9044       19.0       0.0721

FOR HO: VARIANCES ARE EQUAL, F'=     2.25 WITH 12 AND 7 DF      PROB > F'= 0.2891
---------------------------------------------------------------------------------
```

VARIABLE: PREPOST t-test uses $\bar{d}_1 + \bar{d}_2$ "sum of differences"

```
GROUP        N          MEAN       STD DEV     STD ERROR       MINIMUM        MAXIMUM

NEW          8    0.75000000    9.73579551    3.44212351   -6.00000000    20.00000000
OLD         13   12.07692308    7.63174880    2.11666628   -4.00000000    24.00000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL      -2.8031       12.3       0.0157
EQUAL        -2.9767       19.0       0.0078

FOR HO: VARIANCES ARE EQUAL, F'=     1.63 WITH 7 AND 12 DF      PROB > F'= 0.4375
---------------------------------------------------------------------------------
```



These are not helpful due to the significant interaction.

VARIABLE: TMT_X_PP t-test uses $\bar{d}_1 - \bar{d}_2$ "difference of differences"

```
GROUP        N          MEAN       STD DEV     STD ERROR       MINIMUM        MAXIMUM

NEW          8    0.75000000    9.73579551    3.44212351   -6.00000000    20.00000000
OLD         13  -12.07692308    7.63174880    2.11666628  -24.00000000     4.00000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL       3.1743       12.3       0.0078
EQUAL         3.3709       19.0       0.0032

FOR HO: VARIANCES ARE EQUAL, F'=     1.63 WITH 7 AND 12 DF      PROB > F'= 0.4375
```

Since the interaction is clearly important
we need to consider simple effects

Ⓒ

An alternative approach to finding the same information as in parts Ⓐ or Ⓕ.

```
                    GENERAL LINEAR MODELS PROCEDURE

                       CLASS LEVEL INFORMATION

                       CLASS      LEVELS     VALUES

                       GROUP         2       NEW OLD

                NUMBER OF OBSERVATIONS IN DATA SET = 21


DEPENDENT VARIABLE: TMT

SOURCE                      DF        SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                        2      143063.00000000        71531.50000000         153.05

ERROR                       19        8880.00000000          467.36842105         PR > F

UNCORRECTED TOTAL           21      151943.00000000                               0.0001

R-SQUARE                  C.V.              STD DEV              TMT MEAN

0.941557               26.3490          21.61870535           82.04761905

SOURCE                      DF          TYPE I SS    F VALUE    PR > F

XO                           1    141368.04761905     302.48    0.0001
GROUP                        1      1694.95238095       3.63    0.0721

SOURCE                      DF         TYPE IV SS    F VALUE    PR > F

XO                           0         0.00000000       .          .
GROUP                        1      1694.95238095       3.63    0.0721

                                          T FOR HO:       PR > |T|       STD ERROR OF
PARAMETER                 ESTIMATE      PARAMETER=0                        ESTIMATE

NEW VS STD             18.50000000             1.90         0.0721         9.71454938
```

```
         Ⓒ               GENERAL LINEAR MODELS PROCEDURE
DEPENDENT VARIABLE: PREPOST

SOURCE                    DF       SUM OF SQUARES         MEAN SQUARE      F VALUE

MODEL                      2       1900.57692308        950.28846154        13.25

ERROR                     19       1362.42307692         71.70647773      PR > F

UNCORRECTED TOTAL         21       3263.00000000                          0.0002


R-SQUARE                C.V.             STD DEV        PREPOST MEAN

0.582463            109.0965          8.46796775          7.76190476


SOURCE                    DF           TYPE I SS    F VALUE     PR > F

XO                         1       1265.19047619      17.64     0.0005
GROUP                      1        635.38644689       8.86     0.0078


SOURCE                    DF          TYPE IV SS    F VALUE     PR > F

XO                         0          0.00000000        .          .
GROUP                      1        635.38644689       8.86     0.0078

                                       T FOR HO:       PR > |T|       STD ERROR OF
PARAMETER                ESTIMATE    PARAMETER=0                       ESTIMATE

NEW VS STD           -11.32692308          -2.98       0.0078         3.80515343
```

```
DEPENDENT VARIABLE: TMT_X_PP

SOURCE                    DF       SUM OF SQUARES         MEAN SQUARE      F VALUE

MODEL                      2       1900.57692308        950.28846154        13.25

ERROR                     19       1362.42307692         71.70647773      PR > F

UNCORRECTED TOTAL         21       3263.00000000                          0.0002


R-SQUARE                C.V.             STD DEV       TMT_X_PP MEAN

0.582463            117.7664          8.46796775         -7.19047619


SOURCE                    DF           TYPE I SS    F VALUE     PR > F

XO                         1       1085.76190476      15.14     0.0010
GROUP                      1        814.81501832      11.36     0.0032


SOURCE                    DF          TYPE IV SS    F VALUE     PR > F

XO                         0          0.00000000        .          .
GROUP                      1        814.81501832      11.36     0.0032

                                       T FOR HO:       PR > |T|       STD ERROR OF
PARAMETER                ESTIMATE    PARAMETER=0                       ESTIMATE

NEW VS STD            12.82692308           3.37       0.0032         3.80515343
```

```
                     LEAST SQUARES MEANS

GROUP            TMT          STD ERR      PROB > |T|
              LSMEAN           LSMEAN     HO:LSMEAN=0

NEW       93.5000000        7.6433666          0.0001
OLD       75.0000000        5.9959501          0.0001

GROUP        PREPOST          STD ERR      PROB > |T|
              LSMEAN           LSMEAN     HO:LSMEAN=0

NEW        0.7500000        2.9938787          0.8049
OLD       12.0769231        2.3485917          0.0001

GROUP       TMT_X_PP          STD ERR      PROB > |T|
              LSMEAN           LSMEAN     HO:LSMEAN=0

NEW        0.7500000        2.9938787          0.8049
OLD      -12.0769231        2.3485917          0.0001
```

Ⓓ

```
 OBS   SUBJECT   XO   GROUP   TIME   Y

   1      1      1     NEW     1    51
   2      1      1     NEW     2    48
   3      2      1     NEW     1    35
   4      2      1     NEW     2    55
   5      3      1     NEW     1    66
   6      3      1     NEW     2    60
   7      4      1     NEW     1    40
   8      4      1     NEW     2    35
   9      5      1     NEW     1    39
  10      5      1     NEW     2    36
  11      6      1     NEW     1    46
  12      6      1     NEW     2    43
  13      7      1     NEW     1    52
  14      7      1     NEW     2    46
  15      8      1     NEW     1    42
  16      8      1     NEW     2    54
  17      9      1     OLD     1    34
  18      9      1     OLD     2    16
  19     10      1     OLD     1    40
  20     10      1     OLD     2    36
  21     11      1     OLD     1    34
  22     11      1     OLD     2    16
  23     12      1     OLD     1    36
  24     12      1     OLD     2    18
  25     13      1     OLD     1    38
  26     13      1     OLD     2    32
  27     14      1     OLD     1    32
  28     14      1     OLD     2    14
  29     15      1     OLD     1    44
  30     15      1     OLD     2    20
  31     16      1     OLD     1    50
  32     16      1     OLD     2    43
  33     17      1     OLD     1    60
  34     17      1     OLD     2    45
  35     18      1     OLD     1    63
  36     18      1     OLD     2    67
  37     19      1     OLD     1    50
  38     19      1     OLD     2    36
  39     20      1     OLD     1    42
  40     20      1     OLD     2    34
  41     21      1     OLD     1    43
  42     21      1     OLD     2    32
```

Ⓔ

```
                  GENERAL LINEAR MODELS PROCEDURE

                      CLASS LEVEL INFORMATION

CLASS      LEVELS     VALUES

GROUP         2       NEW OLD

TIME          2       1 2

SUBJECT      21       1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16 17 18 19 20 21


              NUMBER OF OBSERVATIONS IN DATA SET = 42
```

Ⓔ

```
                    GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: Y

SOURCE                  DF        SUM OF SQUARES          MEAN SQUARE       F VALUE

MODEL                   23        76921.78846154        3344.42558528         93.28

ERROR                   19          681.21153846          35.85323887       PR > F

UNCORRECTED TOTAL       42        77603.00000000                             0.0001


R-SQUARE              C.V.               STD DEV               Y MEAN

0.991222           14.5958            5.98775742          41.02380952


SOURCE                  DF           TYPE I SS    F VALUE     PR > F

XO                       1      70684.02380952    1971.48     0.0001
GROUP                    1        847.47619048      23.64     0.0001
SUBJECT(GROUP)          19       4440.00000000       6.52     0.0001
TIME                     1        542.88095238      15.14     0.0010
GROUP*TIME               1        407.40750916      11.36     0.0032

SOURCE                  DF          TYPE II SS    F VALUE     PR > F

XO                       0          0.00000000          .          .
GROUP                    1        847.47619048      23.64     0.0001
SUBJECT(GROUP)          19       4440.00000000       6.52     0.0001
TIME                     1        542.88095238      15.14     0.0010
GROUP*TIME               1        407.40750916      11.36     0.0032

SOURCE                  DF         TYPE III SS    F VALUE     PR > F

XO                       0          0.00000000          .          .
GROUP                    1        847.47619048      23.64     0.0001
SUBJECT(GROUP)          19       4440.00000000       6.52     0.0001
TIME                     1        317.69322344       8.86     0.0078
GROUP*TIME               1        407.40750916      11.36     0.0032

SOURCE                  DF          TYPE IV SS    F VALUE     PR > F

XO                       0          0.00000000          .          .
GROUP                    1        847.47619048      23.64     0.0001
SUBJECT(GROUP)          19       4440.00000000       6.52     0.0001
TIME                     1        317.69322344       8.86     0.0078
GROUP*TIME               1        407.40750916      11.36     0.0032

TESTS OF HYPOTHESES USING THE TYPE I MS FOR SUBJECT(GROUP) AS AN ERROR TERM

SOURCE                  DF           TYPE I SS    F VALUE     PR > F

GROUP                    1        847.47619048       3.63     0.0721
```

$$t = \frac{\left(\frac{8}{21}\right)(-0.75) + \left(\frac{13}{21}\right)(12.0769)}{\left\{2(35.85324)\left[\left(\frac{8}{21}\right)^2\left(\frac{1}{8}\right)+\left(\frac{13}{21}\right)^2\left(\frac{1}{13}\right)\right]\right\}^{\frac{1}{2}}}$$

$$= 3.89 \Rightarrow F = (3.89)^2 = 15.14 .$$

The Type I and II SS for TIME test equality of the weighted means -- weighted according to sample size. This is probably not desired and the Type III or IV should be used as they test equality of unweighted means.

Compare with the results of Ⓑ and Ⓒ.

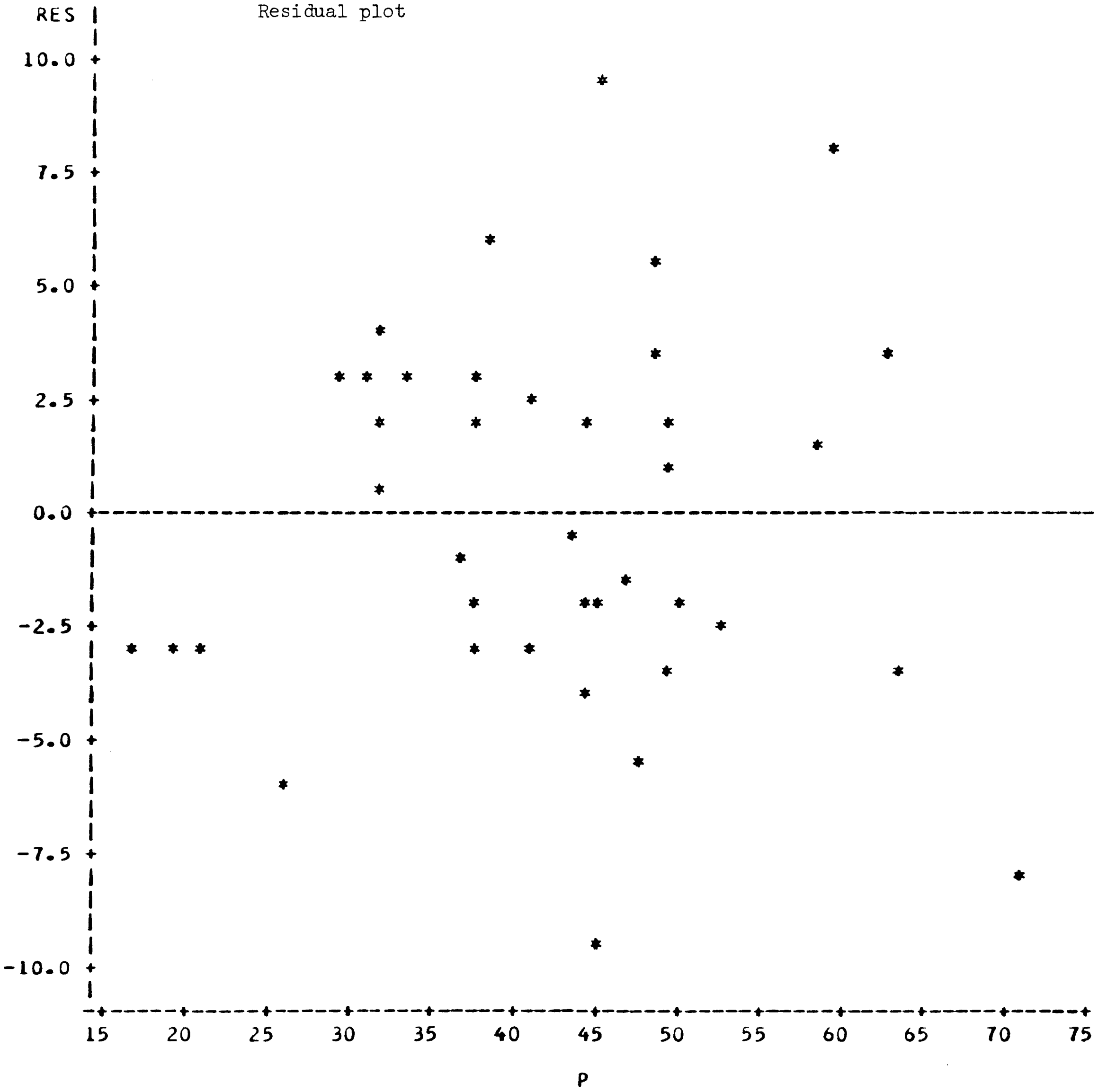
Ⓔ

```
          GENERAL LINEAR MODELS PROCEDURE

                         MEANS

            GROUP            N              Y

            NEW             16      46.7500000
            OLD             26      37.5000000

      SUBJECT   GROUP             N             Y

      1         NEW               2    49.5000000
      2         NEW               2    45.0000000
      3         NEW               2    63.0000000
      4         NEW               2    37.5000000
      5         NEW               2    37.5000000
      6         NEW               2    44.5000000
      7         NEW               2    49.0000000
      8         NEW               2    48.0000000
      9         OLD               2    25.0000000
      10        OLD               2    38.0000000
      11        OLD               2    25.0000000
      12        OLD               2    27.0000000
      13        OLD               2    35.0000000
      14        OLD               2    23.0000000
      15        OLD               2    32.0000000
      16        OLD               2    46.5000000
      17        OLD               2    52.5000000
      18        OLD               2    65.0000000
      19        OLD               2    43.0000000
      20        OLD               2    38.0000000
      21        OLD               2    37.5000000

           TIME             N              Y

           1               21      44.6190476
           2               21      37.4285714

         GROUP     TIME           N             Y

         NEW       1              8    46.3750000
         NEW       2              8    47.1250000
         OLD       1             13    43.5384615
         OLD       2             13    31.4615385
```

} These means are plotted in part Ⓐ.

Ⓕ PLOT OF RES*P   SYMBOL USED IS *

Residual plot



NOTE: 3 OBS HIDDEN

HEMOGLOBIN DATA – Analysis of a 2-Period Cross-over Design
(Grizzle, 1965, Biometrics 21: 467-480)

```
DATA HEMO;
INPUT Y1 Y2 @@;
SUBJECT = _N_;
IF _N_ LE 6 THEN GROUP='A1_B2';
ELSE GROUP='B1_A2';
TMT  = Y2-Y1;
RESID= Y2+Y1;
TREND= TMT;
IF GROUP='B1_A2' THEN TREND=-TMT:
XO = 1;
CARDS;
```

These statements calculate the sums and differences required for the appropriate t-tests. See printout for Ⓐ and Ⓑ.

```
PROC SORT; BY GROUP;
```

Ⓐ

```
PROC PRINT N; BY GROUP;
```

Ⓑ

```
PROC TTEST; CLASS GROUP;
     VAR TMT RESID TREND Y1;
```

PROC TTEST gives the appropriate four t-tests for treatment effects, carry-over effects, time trend and treatment differences within first period only.

Ⓒ

```
PROC GLM; CLASS GROUP;
     MODEL TMT RESID TREND Y1 = XO GROUP / NOINT;
     LSMEANS GROUP / STDERR;
     ESTIMATE 'CONTRAST' GROUP 1 -1;
```

This performs the same analyses as those found in part Ⓑ. Compare also with the results in part Ⓔ.

Ⓓ

```
DATA AOV; SET HEMO;
     Y=Y1; PERIOD=1; IF _N_ LE 6 THEN TRT='A'; ELSE TRT='B'; OUTPUT;
     Y=Y2; PERIOD=2; IF _N_ LE 6 THEN TRT='B'; ELSE TRT='A'; OUTPUT;
     DROP Y1 Y2 TMT RESID TREND;
```

Rearranging the data so that the "usual" ANOVA table may be constructed.

```
PROC PRINT; VAR SUBJECT XO GROUP PERIOD TRT Y;
```

Ⓔ

```
PROC GLM; CLASS TRT PERIOD SUBJECT GROUP;
     MODEL  Y = XO GROUP SUBJECT(GROUP)
                PERIOD TRT / NOINT P SS1 SS2 SS3 SS4;
     MEANS GROUP SUBJECT(GROUP) PERIOD TRT / DEPONLY;
     TEST H=GROUP E=SUBJECT(GROUP) / HTYPE=1 ETYPE=1;
     ESTIMATE 'TRT' TRT 1 -1;
     ESTIMATE 'PERIOD' PERIOD 1 -1;
```

Combined ANOVA table with two error terms.

Ⓕ

```
PROC PLOT; PLOT RES*P='*' / VREF=0;
```

Residual plot.

Ⓐ

```
-------------------------------- GROUP=A1_B2 ---------------------------------

 OBS      Y1       Y2     SUBJECT       TMT      RESID      TREND     XO

   1      0.2      1.0       1          0.8       1.2        0.8      1
   2      0.0     -0.7       2         -0.7      -0.7       -0.7      1
   3     -0.8      0.2       3          1.0      -0.6        1.0      1
   4      0.6      1.1       4          0.5       1.7        0.5      1
   5      0.3      0.4       5          0.1       0.7        0.1      1
   6      1.5      1.2       6         -0.3       2.7       -0.3      1

 N=6

-------------------------------- GROUP=B1_A2 ---------------------------------

 OBS      Y1       Y2     SUBJECT       TMT      RESID      TREND     XO

   7      1.3      0.9       7         -0.4       2.2        0.4      1
   8     -2.3      1.0       8          3.3      -1.3       -3.3      1
   9      0.0      0.6       9          0.6       0.6       -0.6      1
  10     -0.8     -0.3      10          0.5      -1.1       -0.5      1
  11     -0.4     -1.0      11         -0.6      -1.4        0.6      1
  12     -2.9      1.7      12          4.6      -1.2       -4.6      1
  13     -1.9     -0.3      13          1.6      -2.2       -1.6      1
  14     -2.9      0.9      14          3.8      -2.0       -3.8      1

 N=8
```

Y1 are the responses during period 1

Y2 are the responses during period 2

		Group (Sequence) 1(n=6)	2(n=8)
Period	1	A μ_{11}	B μ_{21}
	2	B μ_{12}	A μ_{22}

TMT $= y_{ij2} - y_{ij1} = d_{ij}$ (= within subject differences)

RESID $= y_{ij2} + y_{ij1} = z_{ij}$ (= within subject sums)

$$\text{TREND} = \begin{cases} d_{ij} & \text{if sequence is AB} \\ -d_{ij} & \text{if sequence is BA} \end{cases}$$

The assumed model appears as:

$$y_{ijk} = \mu_{ik} + \xi_{ij} + \epsilon_{ijk}$$

where

$$\mu_{ik} = \mu + \pi_k + \phi_\ell + \lambda_\ell$$

$$j = 1, \cdots, n_i; \quad i = 1,2; \quad k = 1,2; \quad \ell = 1,2$$

and

μ = **general mean,**

ξ_{ij} = **the effect of the *j*-th patient (subject) within the *i*-th sequence, which, for the sake of testing hypotheses, we must assume to be a normally distributed random variable with mean 0 and variance σ_s^2 ,**

π_k = **the effect of the *k*-th period,**

ϕ_l = **the direct effect of the *l*-th drug,**

λ_l = **the residual effect of the *l*-th drug, and**

ϵ_{ijk} = **the random fluctuation which is normally distributed with mean 0 and variance σ_e^2 , and is independent of the ξ_{ij} .**

The assumptions made about ξ_{ij} and ϵ_{ijk} imply that the variance of an observation is $\sigma_s^2 + \sigma_e^2$ and that two observations on an individual have covariance σ_s^2 . Observations made on different subjects are independent.

Note:

$$\text{(i)} \quad (\bar{y}_{1\cdot1} - \bar{y}_{1\cdot2}) - (\bar{y}_{2\cdot1} - \bar{y}_{2\cdot2}) = 2(\phi_1 - \phi_2) + (\lambda_2 - \lambda_1) + (\bar{\epsilon}_{1\cdot1} - \bar{\epsilon}_{1\cdot2} - \bar{\epsilon}_{2\cdot1} + \bar{\epsilon}_{2\cdot2})$$

$$\text{(ii)} \quad (\bar{y}_{1\cdot1} - \bar{y}_{1\cdot2}) + (\bar{y}_{2\cdot1} - \bar{y}_{2\cdot2}) = 2(\pi_1 - \pi_2) - (\lambda_1 + \lambda_2) + (\bar{\epsilon}_{1\cdot1} - \bar{\epsilon}_{1\cdot2} + \bar{\epsilon}_{2\cdot1} - \bar{\epsilon}_{2\cdot2})$$

$$\text{(iii)} \quad (\bar{y}_{1\cdot1} + \bar{y}_{1\cdot2}) - (\bar{y}_{2\cdot1} + \bar{y}_{2\cdot2}) = (\lambda_2 - \lambda_1) + 2(\bar{\xi}_{1.} - \bar{\xi}_{2.}) + (\bar{\epsilon}_{1\cdot1} + \bar{\epsilon}_{1\cdot2} - \bar{\epsilon}_{2\cdot1} - \bar{\epsilon}_{2\cdot2})$$

$$\text{(iv)} \quad \bar{y}_{1\cdot1} - \bar{y}_{2\cdot1} = (\phi_1 - \phi_2) + (\bar{\xi}_{1.} - \bar{\xi}_{2.}) + (\bar{\epsilon}_{1\cdot1} - \bar{\epsilon}_{2\cdot1})$$

Ⓑ TTEST PROCEDURE

VARIABLE: TMT   t-test uses $\bar{d}_1 - \bar{d}_2$ "difference of differences"

```
GROUP         N          MEAN        STD DEV      STD ERROR       MINIMUM       MAXIMUM

A1_B2         6    0.23333333     0.65625198     0.26791375   -0.70000000    1.00000000
B1_A2         8    1.67500000     1.99051321     0.70375270   -0.60000000    4.60000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL     -1.9145        8.9       0.0882
EQUAL       -1.6915       12.0       0.1165

FOR H0: VARIANCES ARE EQUAL, F'=     9.20 WITH 7 AND 5 DF      PROB > F'= 0.0265
---------------------------------------------------------------------------------
```

VARIABLE: RESID   t-test uses $\bar{z}_1 - \bar{z}_2$ "difference of sums"

```
GROUP         N          MEAN        STD DEV      STD ERROR       MINIMUM       MAXIMUM

A1_B2         6    0.83333333     1.32614730     0.54139737   -0.70000000    2.70000000
B1_A2         8   -0.80000000     1.47454593     0.52133071   -2.20000000    2.20000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL      2.1732       11.5       0.0515
EQUAL        2.1379       12.0       0.0538
```

There appear to be important carry-over effects. Thus, the above test for treatment effects is not free of residual effects and we must resort to the results of the first period only (variable Y1 below).

```
FOR H0: VARIANCES ARE EQUAL, F'=     1.24 WITH 7 AND 5 DF      PROB > F'= 0.8437
---------------------------------------------------------------------------------
```

VARIABLE: TREND   t-test uses $\bar{d}_1 + \bar{d}_2$ "sum of differences"

```
GROUP         N          MEAN        STD DEV      STD ERROR       MINIMUM       MAXIMUM

A1_B2         6    0.23333333     0.65625198     0.26791375   -0.70000000    1.00000000
B1_A2         8   -1.67500000     1.99051321     0.70375270   -4.60000000    0.60000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL      2.5342        8.9       0.0323
EQUAL        2.2390       12.0       0.0449

FOR H0: VARIANCES ARE EQUAL, F'=     9.20 WITH 7 AND 5 DF      PROB > F'= 0.0265
```

VARIABLE: Y1   t-test uses $\bar{y}_{1\cdot 1} - \bar{y}_{2\cdot 1}$ "independent two-sample t-test for the first period only"

```
GROUP         N          MEAN        STD DEV      STD ERROR       MINIMUM       MAXIMUM

A1_B2         6    0.30000000     0.75365775     0.30767949   -0.80000000    1.50000000
B1_A2         8   -1.23750000     1.50990775     0.53383301   -2.90000000    1.30000000

VARIANCES          T         DF    PROB > |T|

UNEQUAL      2.4953       10.8       0.0302
EQUAL        2.2746       12.0       0.0421
```

There appear to be treatment effects present. Note: Had we chosen to ignore residual effects the conclusion would likely have been different.

```
FOR H0: VARIANCES ARE EQUAL, F'=     4.01 WITH 7 AND 5 DF      PROB > F'= 0.1453
```

```
      Ⓒ          GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: TMT

SOURCE                   DF      SUM OF SQUARES        MEAN SQUARE      F VALUE

MODEL                     2         22.77166667        11.38583333         4.57

ERROR                    12         29.88833333         2.49069444       PR > F

UNCORRECTED TOTAL        14         52.66000000                          0.0334

R-SQUARE               C.V.             STD DEV           TMT MEAN

0.432428           149.2886          1.57819341         1.05714286

SOURCE                   DF          TYPE I SS    F VALUE    PR > F

XO                        1        15.64571429       6.28    0.0276
GROUP                     1         7.12595238       2.86    0.1165

                                     T FOR H0:      PR > |T|      STD ERROR OF
PARAMETER                ESTIMATE  PARAMETER=0                      ESTIMATE

CONTRAST              -1.44166667        -1.69        0.1165        0.85232186
```

```
DEPENDENT VARIABLE: RESID

SOURCE                   DF      SUM OF SQUARES        MEAN SQUARE      F VALUE

MODEL                     2          9.28666667         4.64333333         2.32

ERROR                    12         24.01333333         2.00111111       PR > F

UNCORRECTED TOTAL        14         33.30000000                          0.1406

R-SQUARE               C.V.             STD DEV         RESID MEAN

0.278879          1414.6063          1.41460634        -0.10000000

SOURCE                   DF          TYPE I SS    F VALUE    PR > F

XO                        1         0.14000000       0.07    0.7959
GROUP                     1         9.14666667       4.57    0.0538

                                     T FOR H0:      PR > |T|      STD ERROR OF
PARAMETER                ESTIMATE  PARAMETER=0                      ESTIMATE

CONTRAST               1.63333333         2.14        0.0538        0.76397474
```

```
                     LEAST SQUARES MEANS

GROUP            TMT         STD ERR       PROB > |T|
              LSMEAN          LSMEAN     HO:LSMEAN=0

A1_B2     0.23333333      0.64429476          0.7235
B1_A2     1.67500000      0.55797563          0.0110

GROUP          RESID         STD ERR       PROB > |T|
              LSMEAN          LSMEAN     HO:LSMEAN=0

A1_B2     0.83333333      0.57751062          0.1746
B1_A2    -0.80000000      0.50013887          0.1357

GROUP          TREND         STD ERR       PROB > |T|
              LSMEAN          LSMEAN     HO:LSMEAN=0

A1_B2     0.23333333      0.64429476          0.7235
B1_A2    -1.67500000      0.55797563          0.0110

GROUP             Y1         STD ERR       PROB > |T|
              LSMEAN          LSMEAN     HO:LSMEAN=0

A1_B2     0.30000000      0.51097334          0.5680
B1_A2    -1.23750000      0.44251589          0.0161
```

```
Ⓒ              GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: TREND

SOURCE                    DF        SUM OF SQUARES           MEAN SQUARE        F VALUE

MODEL                      2           22.77166667           11.38583333           4.57

ERROR                     12           29.88833333            2.49069444         PR > F

UNCORRECTED TOTAL         14           52.66000000                               0.0334

R-SQUARE                C.V.               STD DEV            TREND MEAN

0.432428            184.1226            1.57819341           -0.85714286

SOURCE                    DF            TYPE I SS    F VALUE     PR > F

XO                         1          10.28571429       4.13     0.0649
GROUP                      1          12.48595238       5.01     0.0449

                                           T FOR HO:       PR > |T|          STD ERROR OF
PARAMETER                  ESTIMATE      PARAMETER=0                           ESTIMATE

CONTRAST                 1.90833333             2.24         0.0449          0.85232186
--------------------------------------------------------------------------------------------

DEPENDENT VARIABLE: Y1

SOURCE                    DF        SUM OF SQUARES           MEAN SQUARE        F VALUE

MODEL                      2           12.79125000            6.39562500           4.08

ERROR                     12           18.79875000            1.56656250         PR > F

UNCORRECTED TOTAL         14           31.59000000                               0.0444

R-SQUARE                C.V.               STD DEV               Y1 MEAN

0.404915            216.3301            1.25162395           -0.57857143

SOURCE                    DF            TYPE I SS    F VALUE     PR > F

XO                         1           4.68642857       2.99     0.1093
GROUP                      1           8.10482143       5.17     0.0421

                                           T FOR HO:       PR > |T|          STD ERROR OF
PARAMETER                  ESTIMATE      PARAMETER=0                           ESTIMATE

CONTRAST                 1.53750000             2.27         0.0421          0.67595419
```

```
          OBS    SUBJECT    XO    GROUP    PERIOD    TRT      Y

Ⓓ          1        1       1     A1_B2       1       A      0.2
           2        1       1     A1_B2       2       B      1.0
           3        2       1     A1_B2       1       A      0.0
           4        2       1     A1_B2       2       B     -0.7
           5        3       1     A1_B2       1       A     -0.8
           6        3       1     A1_B2       2       B      0.2
           7        4       1     A1_B2       1       A      0.6
           8        4       1     A1_B2       2       B      1.1
           9        5       1     A1_B2       1       A      0.3
          10        5       1     A1_B2       2       B      0.4
          11        6       1     A1_B2       1       A      1.5
          12        6       1     A1_B2       2       B      1.2
          13        7       1     B1_A2       1       B      1.3
          14        7       1     B1_A2       2       A      0.9
          15        8       1     B1_A2       1       B     -2.3
          16        8       1     B1_A2       2       A      1.0
          17        9       1     B1_A2       1       B      0.0
          18        9       1     B1_A2       2       A      0.6
          19       10       1     B1_A2       1       B     -0.8
          20       10       1     B1_A2       2       A     -0.3
          21       11       1     B1_A2       1       B     -0.4
          22       11       1     B1_A2       2       A     -1.0
          23       12       1     B1_A2       1       B     -2.9
          24       12       1     B1_A2       2       A      1.7
          25       13       1     B1_A2       1       B     -1.9
          26       13       1     B1_A2       2       A     -0.3
          27       14       1     B1_A2       1       B     -2.9
          28       14       1     B1_A2       2       A      0.9
```

Ⓔ

```
                    GENERAL LINEAR MODELS PROCEDURE

                       CLASS LEVEL INFORMATION

             CLASS      LEVELS     VALUES

             TRT           2       A B

             PERIOD        2       1 2

             SUBJECT      14       1 2 3 4 5 6 7 8 9 10 11 12 13 14

             GROUP         2       A1_B2 B1_A2

                 NUMBER OF OBSERVATIONS IN DATA SET = 28


DEPENDENT VARIABLE: Y

SOURCE                      DF        SUM OF SQUARES          MEAN SQUARE        F VALUE

MODEL                       16           28.03583333           1.75223958           1.41

ERROR                       12           14.94416667           1.24534722         PR > F

UNCORRECTED TOTAL           28           42.98000000                              0.2780


R-SQUARE               C.V.                STD DEV               Y MEAN

0.652300          2231.9025             1.11595126          -0.05000000


SOURCE                      DF               TYPE I SS    F VALUE     PR > F

XO                           1              0.07000000       0.06     ~~0.8166~~
GROUP                        1              4.57333333       3.67     ~~0.0794~~
SUBJECT(GROUP)              12             12.00666667       0.80     ~~0.6446~~
PERIOD                       1              7.82285714       6.28     0.0276
TRT                          1              3.56297619       2.86     0.1165


SOURCE                      DF              TYPE II SS    F VALUE     PR > F

XO                           0              0.00000000        .         .
GROUP                        1              4.57333333       3.67     ~~0.0794~~
SUBJECT(GROUP)              12             12.00666667       0.80     ~~0.6446~~
PERIOD                       1              6.24297619       5.01     0.0449
TRT                          1              3.56297619       2.86     0.1165


SOURCE                      DF             TYPE III SS    F VALUE     PR > F

XO                           0              0.00000000        .         .
GROUP                        1              4.57333333       3.67     ~~0.0794~~
SUBJECT(GROUP)              12             12.00666667       0.80     ~~0.6446~~
PERIOD                       1              6.24297619       5.01     0.0449
TRT                          1              3.56297619       2.86     0.1165
```

The Type I SS test equality of weighted means for the PERIOD effect. This is probably not of interest.

The Type III SS give the appropriate tests of unweighted means for the TRT and PERIOD effects. (See Ⓐ, Ⓑ and ESTIMATE on next page.)

Ⓔ

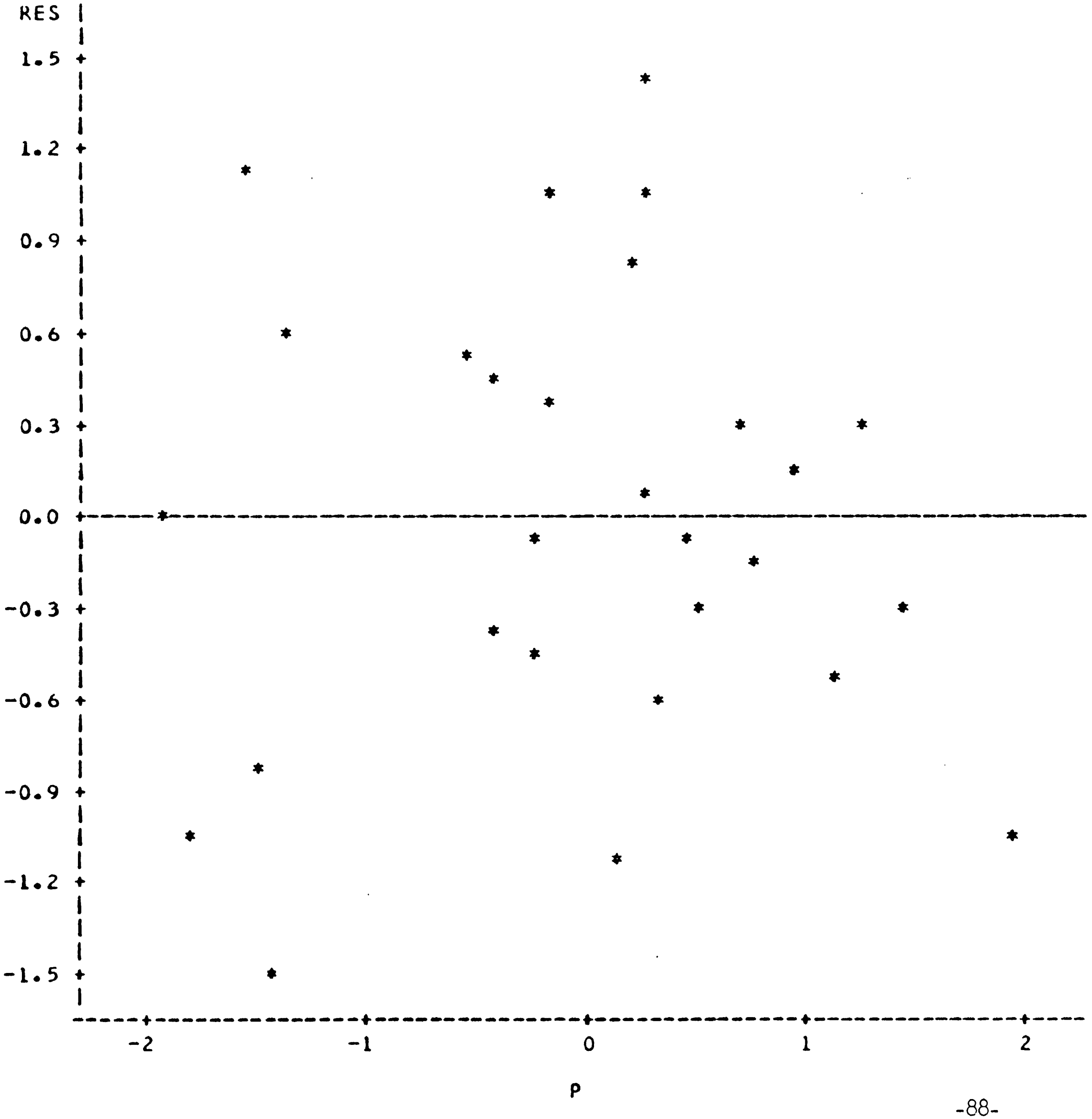
```
TESTS OF HYPOTHESES USING THE TYPE I MS FOR SUBJECT(GROUP) AS AN ERROR TERM

SOURCE                   DF          TYPE I SS    F VALUE    PR > F

GROUP  (Residual)         1         4.57333333       4.57    0.0538

                                     T FOR HO:      PR > |T|          STD ERROR OF
PARAMETER              ESTIMATE    PARAMETER=0                          ESTIMATE

TRT                  0.72083333           1.69        0.1165          0.42616093
PERIOD              -0.95416667          -2.24        0.0449          0.42616093
```

Compare with parts Ⓑ and Ⓒ.

```
                          MEANS

             GROUP          N           Y

             A1_B2         12     0.41666667
             B1_A2         16    -0.40000000


         SUBJECT    GROUP           N             Y

         1          A1_B2           2     0.60000000
         2          A1_B2           2    -0.35000000
         3          A1_B2           2    -0.30000000
         4          A1_B2           2     0.85000000
         5          A1_B2           2     0.35000000
         6          A1_B2           2     1.35000000
         7          B1_A2           2     1.10000000
         8          B1_A2           2    -0.65000000
         9          B1_A2           2     0.30000000
         10         B1_A2           2    -0.55000000
         11         B1_A2           2    -0.70000000
         12         B1_A2           2    -0.60000000
         13         B1_A2           2    -1.10000000
         14         B1_A2           2    -1.00000000


             PERIOD          N            Y

             1              14    -0.57857143
             2              14     0.47857143


              TRT          N            Y

              A           14     0.37857143
              B           14    -0.47857143
```

Ⓕ

PLOT OF RES*P   SYMBOL USED IS *



MILK YIELD DATA - Analysis of a 3 Period - 3 Treatment Cross-over Design
Balanced for Residual Effects. (p. 134ff in Cochran and Cox, 1957)

```
DATA COW;
INPUT SQUARE COW PERIOD TRT $ YIELD RA RB RC RES;
CARDS;
```

Ⓐ

```
PROC PRINT;
VAR SQUARE COW PERIOD TRT RA RB RC RES YIELD;
TITLE A CROSSOVER DESIGN WHEN THERE ARE POSSIBLE RESIDUAL EFFECTS;
```

Ⓑ

```
PROC GLM;
     CLASSES SQUARE COW PERIOD TRT;
     MODEL YIELD = COW PERIOD(SQUARE) RA RB RC TRT / SS1;
     MEANS  TRT / DEPONLY;
     LSMEANS TRT / STDERR;
```

Ⓒ

```
PROC GLM;
     CLASSES SQUARE COW PERIOD TRT;
     MODEL YIELD = COW PERIOD(SQUARE) TRT RA RB RC / SS1;
     ESTIMATE 'TRT A VS AVG(B+C)' TRT 1 -0.5 -0.5;
     ESTIMATE 'TRT B VS TRT C' TRT 0 1 -1;
     LSMEANS TRT / STDERR;
     MEANS TRT / DEPONLY;
     OUTPUT OUT=NEW PREDICTED=P RESIDUAL=R;
```

Ⓓ

```
PROC PLOT;
     PLOT R*P='*' / VREF=0;
```

A check on the residuals reveals no obvious violations of the assumed model.

SQUARE, COW, PERIOD, TRT and RES are CLASS variables. RA, RB and RC are 0,1 indicator variables which indicate the residual effects of each of the three treatments. YIELD is the response variable, coded milk yield.

Ⓐ CLASS, indicator and response variables

```
OBS    SQUARE    COW    PERIOD    TRT    RA    RB    RC    RES    YIELD

  1      1        1       1        A      0     0     0     0       38
  2      1        1       2        B      1     0     0     1       25
  3      1        1       3        C      0     1     0     2       15
  4      1        2       1        B      0     0     0     0      109
  5      1        2       2        C      0     1     0     2       86
  6      1        2       3        A      0     0     1     3       39
  7      1        3       1        C      0     0     0     0      124
  8      1        3       2        A      0     0     1     3       72
  9      1        3       3        B      1     0     0     1       27
 10      2        4       1        A      0     0     0     0       86
 11      2        4       2        C      1     0     0     1       76
 12      2        4       3        B      0     0     1     3       46
 13      2        5       1        B      0     0     0     0       75
 14      2        5       2        A      0     1     0     2       35
 15      2        5       3        C      1     0     0     1       34
 16      2        6       1        C      0     0     0     0      101
 17      2        6       2        B      0     0     1     3       63
 18      2        6       3        A      0     1     0     2        1
```

Ⓑ The model fitting residual effects before treatment effects.

```
A CROSSOVER DESIGN WHEN THERE ARE POSSIBLE RESIDUAL EFFECTS

                    GENERAL LINEAR MODELS PROCEDURE

                       CLASS LEVEL INFORMATION

                    CLASS      LEVELS     VALUES

                    SQUARE        2       1 2

                    COW           6       1 2 3 4 5 6          TRT   A = Roughage diet

                    PERIOD        3       1 2 3                      B = Limited grain diet

                    TRT           3       A B C                      C = Full grain diet

                 NUMBER OF OBSERVATIONS IN DATA SET = 18



DEPENDENT VARIABLE: YIELD

SOURCE                    DF        SUM OF SQUARES         MEAN SQUARE       F VALUE

MODEL                     13        20163.19444444       1551.01495726         31.14

ERROR                      4          199.25000000         49.81250000       PR > F

CORRECTED TOTAL           17        20362.44444444                           0.0023

R-SQUARE                C.V.               STD DEV          YIELD MEAN

0.990215             12.0761            7.05779711         58.44444444

SOURCE                    DF            TYPE I SS    F VALUE      PR > F

COW                        5        5781.11111111      23.21      0.0047
PERIOD(SQUARE)             4       11489.11111111      57.66      0.0009
RA }                       1     {    11.75555556       0.24      0.6525
RB }                       1     {    26.66666667       0.54      0.5049
RC }                       0     {     0.00000000        .           .
TRT                        2        2854.55000000      28.65      0.0043
```

The SS due to residual effects unadjusted for treatments is found as:

SS Residual effects (unadj.) = RA + RB + RC
= 11.7556 + 26.6667 + 0
= 38.4223 with 2 df.

Ⓒ The model-fitting treatments before residual effects.

```
        A CROSSOVER DESIGN WHEN THERE ARE POSSIBLE RESIDUAL EFFECTS

                         GENERAL LINEAR MODELS PROCEDURE

DEPENDENT VARIABLE: YIELD

SOURCE                  DF        SUM OF SQUARES        MEAN SQUARE       F VALUE

MODEL                   13        20163.19444444      1551.01495726         31.14

ERROR                    4          199.25000000        49.81250000       PR > F

CORRECTED TOTAL         17        20362.44444444                          0.0023


R-SQUARE              C.V.               STD DEV         YIELD MEAN

0.990215           12.0761            7.05779711        58.44444444


SOURCE                  DF           TYPE I SS     F VALUE     PR > F

COW                      5       5781.11111111       23.21     0.0047
PERIOD(SQUARE)           4      11489.11111111       57.66     0.0009
TRT                      2       2276.77777778       22.85     0.0065
RA                       1        258.67361111        5.19     0.0849
RB                       1        357.52083333        7.18     0.0553
RC                       0          0.00000000          .          .

                                                 T FOR H0:      PR > |T|     STD ERROR OF
PARAMETER                          ESTIMATE   PARAMETER=0                     ESTIMATE

TRT A VS AVG(B+C)              -23.93750000         -6.07        0.0037       3.94542853
TRT B VS TRT C                 -20.62500000         -4.53        0.0106       4.55578844
```

SS Residual effects (adj.) = RA + RB + RC
= 258.6736 + 357.5208 + 0
= 616.1944 with 2 df.

These are contrasts among the unadjusted treatment means.

Combining the results of Ⓑ and Ⓒ gives the ANOVA table found on page 135 of Cochran and Cox:

Source	df	SS	MS
Cows	5	5781.1111	
Periods w/i squares	4	11489.1111	
Treatments (unadj.)	2	2276.7778	1138.3889
Residual (adj.)	2	616.1944	308.0972
Residual (unadj.)	2	38.4223	19.1115
Treatments (adj.)	2	2854.5500	1427.2750
Error	4	199.2500	49.8125
Corrected Total	17	20362.4444	

```
A CROSSOVER DESIGN WHEN THERE ARE POSSIBLE RESIDUAL EFFECTS

           GENERAL LINEAR MODELS PROCEDURE

                           MEANS =  The unadjusted treatment means

             TRT           N         YIELD

             A             6     45.1666667
             B             6     57.5000000
             C             6     72.6666667
```

```
A CROSSOVER DESIGN WHEN THERE ARE POSSIBLE RESIDUAL EFFECTS

           GENERAL LINEAR MODELS PROCEDURE

                LEAST SQUARES MEANS =  The treatment means adjusted for residual effects

    TRT            YIELD          STD ERR     PROB > |T|
                  LSMEAN           LSMEAN     HO:LSMEAN=0

    A         42.4861111        3.1121960          0.0002
    B         56.1111111        3.1121960          0.0001
    C         76.7361111        3.1121960          0.0001
```

Notice that the adjustments are intuitively pleasing.

References

Allen, David M. and Cady, Foster B. (1982). *Analyzing Experimental Data by Regression*. Lifetime Learning Publications, Belmont, CA.

Aref, Suzanne and Piegorsch, Walter (1980). Analyzing experimental data by regression × SAS: Annotated outputs. Technical paper BU-705-M in the Biometrics Unit series, Cornell University, Ithaca, NY.

Arneson, Valerie, Firey, Patricia A. and McCulloch, Charles E. (1978). Statistical methods × SAS. Technical paper BU-664-M in the Biometrics Unit series, Cornell University, Ithaca, NY.

Brogan, Donna R. and Kutner, Michael H. (1980). Comparative analyses of pretest-posttest research designs. *The American Statistician* 34: 229-232.

Cochran, William G. and Cox, Gertrude M. (1957). *Experimental Designs*, 2*nd* ed. John Wiley, New York.

Grizzle, James E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics* 21: 467-480.

SAS Institute, Inc. (1982). *SAS Users' Guide*: *Statistics*. SAS Institute, Inc., Cary, NC, 584 pp.

Snedecor, George W. and Cochran, William G. (1980). Statistical Methods, 7*th* ed. Iowa State University Press, Ames, Iowa.