Analysis of Variance with GLM and Mixed Models in SAS: Some class notes

Monday, March 1, 2004

Robert A. Yaffee, Ph.D.
Statistics, Social Science, and Mapping Group
Academic Computing Services
75 Third Avenue
Level C-3
Phone: 212-998-3402
E-mail: yaffee@nyu.edu

Outline

1. Classification of Effects
2. General linear model
    1. Fixed Effects Models
        1. The one-way layout
3. Random Effects
    1. Two-Way Random Layout
    2. Solutions and estimates
4. Mixed Model theory
        1. Proper error terms
5. Two-way layout
6. Full-factorial model
    1. Contrasts with interaction terms
    2. Graphing Interactions

Definition of Mixed Models by their component effects

1. Mixed Models contain both fixed and random effects
2. Fixed Effects: factors for which the only levels under consideration are contained in the coding of those effects
3. Random Effects: Factors for which the levels contained in the coding of those factors are a random sample of the total number of levels in the population for that factor.

Examples of Fixed and Random Effects

1. Fixed effect:
2. Sex where both male and female genders are included in the factor, sex.
3. Agegroup: Minor and Adult are both included in the factor of agegroup
4. Random effect:
   1. Subject: the sample is a random sample of the target population

Classification of effects

1. There are main effects: Linear Explanatory Factors
2. There are interaction effects: Joint effects over and above the component main effects.
3. There are nested effects. Hierarchical designs contained nested effects: Patients are nested within doctors and doctors are nested within hospitals
4. Such effects may sometimes be fixed or random. Their classification depends on the experimental design
5. Between-subjects effects are those who are in one group or another but not in both. Experimental group is a fixed effect because the manager is considering only those groups in his experiment. One group is the experimental group and the other is the control group. Therefore, this grouping factor is a between- subject effect.
6. Within-subject effects are experienced by subjects repeatedly over time. Trial is a random effect when there are several trials in the repeated measures design; all subjects experience all of the trials. Trial is therefore a within-subject effect.
7. Operator may be a fixed or random effect, depending upon whether one is generalizing beyond the sample
8. If operator is a random effect, then the machine*operator interaction is a random effect.
9. There are contrasts: These contrast the values of one level with those of other levels of the same effect.

Classification of Effects-cont'd

Hierarchical designs have nested effects. Nested effects are those with subjects within groups.

An example would be patients nested within doctors and doctors nested within hospitals

SAS expresses nesting of effects by:

patients(doctors)

doctors(hospitals)

The General Linear Model

1. The main effects general linear model can be parameterized as

$$Y_{ij} = \mu + \alpha_i + b_j + \varepsilon_{ij}$$

where

Y_{ij} = *observation for ith* α

μ = *grand mean (an unknown fixed parm)*

α_i = *effect of ith value of* α $(a_i - \mu)$

b_j = *effect of jth value of b* $(b_j - \mu)$

ε_{ij} = experimental error $\sim N(\mathbf{0}, \sigma^2)$

A factorial model

If an interaction term were included, the formula would be

$$y_{ij} = \mu + \alpha_i + \beta_i + \alpha\beta_{ij} + e_{ij}$$

The interaction or crossed effect is the joint effect, over and above the individual main effects. Therefore, the main effects must be in the model for the interaction to be properly specified.

$$\alpha\beta_{i\,j} = (y_{ij} - \mu) - (\alpha - \mu) - (\beta - \mu)$$

$$= y_{ij} - \alpha - \beta + \mu$$

Higher-Order Interactions

If 3-way interactions are in the model, then the main effects and all lower order interactions must be in the model for the 3-way interaction to be properly specified.   For example, a 3-way interaction model would be:

$$y_{ijk} = \mu + a_i + b_j + c_k + ab_{ij} + ac_{ik} + bc_{jk} + abc_{ijk} + e_{ijk}$$

The General Linear Model

- In matrix terminology, the general linear model may be expressed as

$$Y = X\beta + \varepsilon$$

where

$Y = the\, observed\, data\, vector$

$X = the\, design\, matrix$

$\beta = the\, vector\, of\, unknown\, fixed\, effect\, parameters$

$\varepsilon = the\, vector\, of\, errors$

Programming the General Linear Model

- In the GLM procedure, one saves the data set plus the residuals, predicted values, and studentized residuals with an output statement in a data set called resdat.

```
PROC GLM;
   class machine operator;
Model yield=machine|operator;
   output out=resdat r=resid     p=pred
         student=stdres rstudent=rstud
         cookd=cksd h=lev;
Run;
```

Assumptions of the general linear model

$$E(\varepsilon) = \boldsymbol{0}$$

$$\text{var}(\varepsilon) = \sigma^2 I$$

$$\text{var}(Y) = \sigma^2 I$$

$$E(Y) = X\beta$$

GLM Assumptions-cont'd

1. Residual Normality.
2. Homogeneity of error variance
3. Functional form of Model: Linearity of Model
4. No Multicollinearity
5. Independence of observations
6. No autocorrelation of errors
7. No influential outliers

We have to test for these to be sure that the model is valid.
We will discuss the robustness of the model in face of violations of these assumptions.
We will discuss recourses when these assumptions are violated.

Residuals Diagnostics

- Residual diagnostics test the fulfillment of the assumptions of the general linear model

SAS test for residual normality.

```
Proc univariate data=resdat normal
plot;
   var resid;
Run;
```

Testing the assumption of Homogeneity of Error Variance

Levene's test is performed by computing

$$Levene's\ test\ \ for\ homogeneity\ of\ variance$$

$$\mathbf{1}.compute\ Z_{ij} = |Y_{ij} - \bar{Y}_{.j}|$$

$$\mathbf{2}.\ Perform\ an\ F\ test\ on\ the\ j\ groups$$

$$\mathbf{3}.\ A\ nonsignificant\ result\ indicates\ no\ heteroskedasticity$$

```
means  factor/  hovtest=levene(type=abs);
```

Graphically examining residuals for homogeneity

```
Proc gplot data=resdat;
  plot resid * pred;
Run;
```

- Analysis for lack of pattern;

Testing for outliers

```
Proc freq data=resdat;
    tables stdres cksd;
Run;
```

1. Look for standardized residuals greater than 3.5 or less than – 3.5
2. And look for high Cook's D (greater than 4*p/(n-p-1).

```
Proc gplot data=resdat;
   plot stdres*cksd;
Title 'Leverage of outliers';
Run;
```

Assessing the leverage of outliers

- Construct and analyze studentized residuals
- Construct and analyze the leverage of the high and low studentized residuals
- Use Cook's D to help determine how problematic the outliers are.

Studentized Residuals

$$e_i^{s} = \frac{e_i}{\sqrt{s^2_{(i)}(1 - h_i)}}$$

where

e_i^{s} = *studentized residual*

$s_{(i)}$ = *standard deviation where ith obs is deleted*

h_i = *leverage statistic*

Belsley et al (1980) recommend the use of studentized Residuals to determine whether there is an outlier.

Influence of Outliers

1. Leverage is measured by the diagonal components of the hat matrix.
2. The hat matrix comes from the formula for the regression of Y.

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

$$where \;\; X'(X'X)^{-1}X' = \; the\ hat\ matrix,\ H$$

Therefore,

$$\hat{Y} = HY$$

Leverage and the Hat matrix

1. The hat matrix transforms Y into the predicted scores.
2. The diagonals of the hat matrix indicate which values will be outliers or not.
3. The diagonals are therefore measures of leverage.
4. Leverage is bounded by two limits: 1/n and 1. The closer the leverage is to unity, the more leverage the value has.
5. The trace of the hat matrix = the number of variables in the model.
6. When the leverage > 2p/n then there is high leverage according to Belsley et al. (1980) cited in Long, J.F. <u>Modern Methods of Data Analysis</u> (p.262). For smaller samples, Vellman and Welsch (1981) suggested that 3p/n is the criterion.

Cook's D

1. Another measure of influence.
2. This is a popular one. The formula for it is:

$$Cook's\ D_i = \left(\frac{1}{p}\right)\left(\frac{h_i}{1-h_i}\right)\left(\frac{e_i^2}{s^2(1-h_i)}\right)$$

Cook and Weisberg(1982) suggested that values of D that exceeded 50% of the F distribution (df = p, n-p) are large.

Cook's D in SAS

Finding the influential outliers

Select those observations for which cksd > 4*p/n

Belsley suggests 4*p/(n-p-1) as a cutoff

```
If cksd > 4*p/(n-p-1);
Proc print; var cksd;
Run;
```

What to do with outliers

1. Check coding to spot typos
2. Correct typos
3. If observational outlier is correct, examine the dffits option to see the influence on the fitting statistics.
4. This will show the standardized influence of the observation on the fit. If the influence of the outlier is bad, then consider removal or replacement of it with imputation.

Decomposition of the Sums of Squares

1. Mean deviations are computed when means are subtracted from individual scores.
   1. This is done for the total, the group mean, and the error terms.
2. Mean deviations are squared and these are called sums of squares
3. Variances are computed by dividing the Sums of Squares by their degrees of freedom.
4. The total Variance = Model Variance + error variance

Formula for Decomposition of Sums of Squares

$$y_{ij} - \bar{y}_{..} = (y_{ij} - \bar{y}_{.j}) + (\bar{y}_{.j} - \bar{y}_{..})$$

total mean deviation = error within + group effect

and squaring terms $(y_{ij} - \bar{y}_{..})^2 = (y_{ij} - \bar{y}_{.j})^2 + (\bar{y}_{.j} - \bar{y}_{..})^2$

and summing the terms

$$\sum (y_{ij} - \bar{y}_{..})^2 = \sum (y_{ij} - \bar{y}_{.j})^2 + \sum (\bar{y}_{.j} - \bar{y}_{..})^2$$

SS total = SS error + SSmodel

Variance Decomposition

Dividing each of the sums of squares by their respective degrees of freedom yields the variances.

$$\frac{SStotal}{n-\mathbf{1}} = \frac{SSerror}{n-k} + \frac{SSmodel}{k-\mathbf{1}}$$

Total variance= error variance

+ model variance.

$$F_{\text{in fixed effects models}} = \frac{model\ variance}{error\ variance}$$

Proportion of Variance Explained

R^2 = proportion of variance explained.

SStotal = SSmodel + SSerrror

Divide all sides by SStotal

SSmodel/SStotal

=1 - SSError/SStotal

R^2=1 - SSError/SStotal

The Omnibus F test

The omnibus F test is a test that all of the means of the levels of the main effects and as well as any interactions specified are not significantly different from one another.

Suppose the model is a one way anova on breaking pressure of bonds of different metals.

Suppose there are three metals: nickel, iron, and Copper.

H_0: Mean(Nickel)= mean (Iron) = mean(Copper)
H_a: Mean(Nickel) ne Mean(Iron) or
Mean(Nickel) ne Mean(Copper)
or Mean(Iron) ne Mean(Copper)

Testing different Levels of a Factor against one another

- Contrast are tests of the mean of one level of a factor against other levels.

$$H_0 : \mu_1 = \mu_2 = \mu_3$$

$$H_a : \begin{cases} \mu_1 \neq \mu_2 \\ \mu_2 \neq \mu_3 \\ \mu_1 \neq \mu_3 \end{cases}$$

Contrasts-cont'd

- A contrast statement computes

$$F = \frac{\left[\begin{bmatrix} \beta \\ Z \end{bmatrix}' L'(L'\hat{V}^{-}L)^{-1} L \begin{bmatrix} \beta \\ Z \end{bmatrix}\right]}{rank(L)}$$

The estimated V^- is the generalized inverse of the coefficient matrix of the mixed model.
The L vector is the k'b vector.

The numerator df is the rank(L) and the denominator df is taken from the fixed effects table unless otherwise specified.

Estimate statements

- This is exactly like a contrast statement except only one-row L matrices are permitted. The actual estimate $L'\beta$ is printed along with its approximate standard error, t-test, and t-probability.
- They are useful in estimating random effects and their significance.

Testing Statistical Significance of Differences of levels of a factor or interaction

We may test whether different levels of a factor differ with respect to the mean level of the dependent variable in several ways. If there are three levels of a metal and we wish to test whether levels 2 and 3 are different, we would be asking whether Copper has a statistically significantly different in its breaking pressure from that of Iron.

We may use A Priori Contrasts or Estimates

An a priori contrast is a t-test between levels:

t= mean(Copper) – mean(Iron)/Se

the SAS syntax for testing it is:

Contrast 'Copper v Iron' metal 0 1 –1

An estimate is the mean response

= mean(copper) - mean(iron) + 0*mean(nickel)

The SAS syntax for estimating this is:

Estimate 'Copper v Iron' metal 0 1 – 1

We may use Post Hoc Tests. These are different sorts of tests that may or may not require equal variances to test whether one level of a factor or interaction is different from other levels.

means metal /scheffe alpha=.017;

Construction of the F tests in different models

The F test is a ratio of two variances (Mean Squares). It is constructed by dividing the MS of the effect to be tested by a MS of the denominator term. The division should leave only the effect to be tested left over as a remainder.

A Fixed Effects model F test for a = MSa/MSerror.
A Random Effects model F test for a = MSa/MSab
A Mixed Effects model F test for b = MSa/MSab
A Mixed Effects model F test for ab = MSab/MSerror

<u>**Expected Mean Squares for Different Designs**</u>
Source: Michaels, Brown, & Winer, 1993, *Statistical Principles of Experimental Design*, 304

Mean Squares	Case 1: Fixed a fixed, b fixed	Case 2: Mixed a fixed b random	Case 3: Random a random b random
MS_a	$\sigma_e^2 + nq\sigma_a^2$	$\sigma_e^2 + n\sigma_{ab}^2 + nq\sigma_a^2$	$\sigma_e^2 + n\sigma_{ab}^2 + nq\sigma_a^2$
Ms_b	$\sigma_e^2 + nq\sigma_b^2$	$\sigma_e^2 + np\sigma_b^2$	$\sigma_e^2 + n\sigma_{ab}^2 + np\sigma_b^2$
MS_{ab}	$\sigma_e^2 + n\sigma_{ab}^2$	$\sigma_e^2 + n\sigma_{ab}^2$	$\sigma_e^2 + n\sigma_{ab}^2$
MS_{error}	σ_e^2	σ_e^2	σ_e^2

The Mixed Model

The Mixed Model subsumes fixed and random effects.

It can be used to model merely fixed or random effects, by zeroing out the other parameter vector.

The F tests for the fixed, random, and mixed models differ.

Because the Mixed Model has the parameter vector for both of these and can estimate the error covariance matrix for each, it can provide the correct standard errors for either the fixed or random effects.

Mixed Model Theory

$Y = X\beta + Zu + \varepsilon$

where

Y = *the data vector*

X = *the design matrix*

β = *fixed effects parameter estimates*

Z = *design matrix for random effects*

u = *random effects*

ε = *error vector*

Mixed Model Theory-cont'd

Little et al.(p.139) note that u and e are uncorrelated random variables with 0 means and covariances, G and R, respectively.

$$\begin{aligned}&\textit{Because the}\\&\textit{covariance matrix}\\&V = ZGZ' \ + \ R,\\&\textit{the solution for}\\&\hat{\beta} = (X'\hat{V}^{-1}X)^{-}X'\hat{V}^{-1}y\\&u = \hat{G}Z'\hat{V}^{-}(y - X\hat{\beta})\end{aligned}$$

V^- is a generalized inverse. Because V is usually singular and noninvertible AVA = V^- is an augmented matrix that is invertible. It can later be transformed back to V.
The G and R matrices must be positive definite.
In Proc Mixed, A random statement defines G and a repeated statement defines R.

Linear Combinations

$$V(\beta) = (X'V^{-1}X)^{-}$$

$$\textit{If } L \textit{ is estimable, then } V(\beta) = L'(X'V^{-1}X)^{-}L$$

Mixed Model Assumptions

$$E\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = 0$$

$$Variance\begin{bmatrix} u \\ \varepsilon \end{bmatrix} = \begin{bmatrix} G & 0 \\ 0 & R \end{bmatrix}$$

Z and G matrices for the mean and variances of random effects

$$Z = \begin{pmatrix} 1 & & & & \\ 1 & & & & \\ 1 & & & & \\ & 1 & & & \\ & 1 & & & \\ & 1 & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \cdot \\ & & & & 1 \\ & & & & 1 \\ & & & & 1 \end{pmatrix} \qquad G = \begin{pmatrix} \sigma_1^2 & & & & & \\ & \sigma_1^2 & & & & \\ & & \cdot & & & \\ & & & \cdot & & \\ & & & & \cdot & \\ & & & & & \sigma_1^2 \end{pmatrix}$$

R matrix, the covariance structure of which is defined with a type option in the repeated command

$$R=\begin{pmatrix}
\sigma_1+\sigma^2 & \sigma_1 & \sigma_1 & & & & & \\
\sigma_1 & \sigma_1+\sigma^2 & \sigma_1 & & & & & \\
\sigma_1 & \sigma_1 & \sigma_1+\sigma^2 & & & & & \\
 & & & \cdot & & & & \\
 & & & & \cdot & & & \\
 & & & & & \sigma_1+\sigma^2 & \sigma_1 & \sigma_1 \\
 & & & & & \sigma_1 & \sigma_1+\sigma^2 & \sigma_1 \\
 & & & & & \sigma_1 & \sigma_1 & \sigma_1+\sigma^2
\end{pmatrix}$$

repeated / type=cs subject=id

GLM $\subset$ Mixed Model

The General Linear Model is a special case of the Mixed Model with Z = 0 (which means that Zu disappears from the model) and $R = \sigma^2 I$

SAS may require data set conversion for Mixed Model Processing

- The data structure for a repeated measures GLM is not the same as that for a Proc Mixed. One has to convert a horizontally repeated observation data set to a vertically repeated observation data set.
- The Horizontally repeated data set in sas appears as follows:

Horizontally Repeated (wide format) Data

```
data Miss;
input influent1  influent2 influent3 influent4
    influent5 influent6;
datalines;
21 21 20 14  7 41
27 11 19 24 15 42
29 18 20 30 18 35
17  9 11 21  4 34
19 13 14 31 28 30
12 23  . 27  .  .
29  .  .  .  .  .
20  .  .  .  .  .
20  .  .  .  .  .
proc print;
title5 'data miss‘; run;
```

Horizontally Repeated (Wide) Data

Command ===>

Mississippi River Influents 1
Data from Littell et al. (1996), p. 141
Nitrogen measurements in parts per million
Random Effects Model
data miss 07:47 Tuesday, November 5, 2002

Obs	influent1	influent2	influent3	influent4	influent5	influent6
1	21	21	20	14	7	41
2	27	11	19	24	15	42
3	29	18	20	30	18	35
4	17	9	11	21	4	34
5	19	13	14	31	28	30
6	12	23	.	27	.	.
7	29	.	.	.	.	.
8	20	.	.	.	.	.
9	20	.	.	.	.	.

Vertically Repeated (Long) Data Set

Mississippi River Influents 2
Data from Littell et al. (1996), p. 141
Nitrogen measurements in parts per million
Random Effects Model
new2 07:47 Tuesday, November 5, 2002

Obs	y	influent
1	21	1
2	21	2
3	20	3
4	14	4
5	7	5
6	41	6
7	27	1
8	11	2
9	19	3
10	24	4
11	15	5
12	42	6
13	29	1
14	18	2
15	20	3
16	30	4
17	18	5
18	35	6
19	17	1
20	9	2
21	11	3
22	21	4
23	4	5
24	34	6
25	19	1
26	13	2
27	14	3
28	31	4
29	28	5
30	30	6
31	12	1
32	23	2
33	.	3

Two Ways to Convert the data sets from Horizontally repeated (wide) to Vertically repeated (long) data sets

```
data new2;
  set miss;
/* Here we restructure the data set */
label type='Type of influent';
  y = influent1; influent=1; output;
  y = influent2; influent=2; output;
  y = influent3; influent=3; output;
  y = influent4; influent=4; output;
  y = influent5; influent=5; output;
  y = influent6; influent=6; output;
drop influent1-influent6;
proc print data=new2;
 title5 'new2';
 run;
```

More Elegant Conversion with Arrays

```
data new3;
  array influent{6} influent1-influent6;
   input influent1  influent2 influent3 influent4 influent5
      influent6;
do type = 1 to 6;
   y = influent(type);
   output;
   end;
drop influent1-influent6;
cards;
21 21 20 14  7 41
27 11 19 24 15 42
29 18 20 30 18 35
17  9 11 21  4 34
19 13 14 31 28 30
12 23  . 27  .  .
29  .  .  .  .  .
20  .  .  .  .  .
20  .  .  .  .  .
proc print data=new3;
title5 'New3';
run;
```

The Fixed-Effects Model

Suppose there are three groups: experimental control standard and three treatments: lowdose, medose, hidose. The dependent variable, Y, is the completeness of recovery. We assume that we have population data. We are not generalizing to larger levels than those we have in our sample. Therefore, all of our effects are deemed to be fixed. We analyze our model as a fixed effects model.

The Fixed Effects models may be programmed in SAS with PROC ANOVA, PROC GLM, or PROC MIXED. PROC ANOVA and PROC GLM have similar syntax.

```
PROC GLM;
   Class group treatment;
   Model Y = group treatment group*treatment;
   Means group treatment group*treatment/
              hovtest=Levene(type=abs)
              Scheffe alpha=.017;
   output out=resdat p=pred r=resid;
   run;
PROC UNIVARIATE normal plot data=resdat; var resid;
Title 'Test of Normality of Residuals';
Run;

PROC MIXED;
    Class group treatment;
    Model Y = group|treatment;
    Lsmeans group|treatment/pdiff;
  run;
```

Proc GLM vs. Proc Mixed

GLM has
- means
- lsmeans
- sstype 1,2,3,4
- estimates using OLS or WLS
- one has to program the correct F tests for random effects.
- cannot handle cases with missing values and drops them from the analysis.

Mixed has
- lsmeans
- sstypes 1 and 3
- estimates using maximum likelihood, general methods of moments, or restricted maximum likelihood
  - ML
  - MIVQUE0
  - REML
- gives correct std errors and confidence intervals for random effects
- Automatically provides correct standard errors for analysis.
- Can handle observation with missing data.
- Can handle a wide variety of covariance structures for random effects, repeated effects.
- More robust analysis.

Analysis of Fixed Effects model in the Mixed Model

SAS tests these effects by constructed a type III L matrix.

We analyze the breaking pressure of bonds made from three metals. We assume that we do not generalize beyond our sample and that our effects are all fixed.

Tests of Fixed Effects is performed with the help of the L matrix by constructing the following F test:

$$F = \frac{\hat{\beta}' L' [L(X' \hat{V}^{-1} X)^{-} L']^{-} L \hat{\beta}}{rank(L)}$$

Numerator df = rank(L)
Denominator df = RESID (n-rank(X)
df = Satherth

SAS Command Syntax: Comparison of SAS Syntax for Fixed Effects ANOVA with Glm and Mixed

```
title1 'General Linear and Mixed Models';
title2 'Ingot Data from SAS Course Notes';
title3 'p.11  SAS Institute, Cary, NC';
options ls=80 ps=55;
proc format;
  value met 1='Nickel'
                2='Iron'
               3='Copper';
data metalic;
 input ingot metals $ pressure;
if metals = 'nickel  ' then metal=1;
if metals = 'iron    ' then metal=2;
if metals = 'copper  ' then metal=3;
format metal met.;
```

Data and Preliminary Analysis

```
datalines;
1 nickel  67.0
1 iron   71.9
1 copper 72.2
2 nickel 67.5
2 iron  68.8
2 copper  66.4
3 nickel 76.0
3 iron    82.6
3 copper  74.5
4 nickel  72.7
4 iron    78.1
4 copper 67.3
5 nickel  73.1
5 iron  74.2
5 copper 73.2
6 nickel 65.8
6 iron  70.8
6 copper 68.7
7 nickel 75.6
7 iron  84.9
7 copper 69.0
proc print;
run;
proc sort; by metal;
run;
proc means; by metal; var pressure;
title 'Mean Pressure of ingots';
run;
proc univariate plot; var pressure; by metal;
title 'Comparative Box Plots';
run;
```

Proc Means Output

Command ===>

Mean Pressure of ingots 64

08:16 Wednesday, November 6, 2002

---------------------------------- metal=Nickel ----------------------------------

The MEANS Procedure

Analysis Variable : pressure

N	Mean	Std Dev	Minimum	Maximum
7	71.1000000	4.2559762	65.8000000	76.0000000

----------------------------------- metal=Iron -----------------------------------

Analysis Variable : pressure

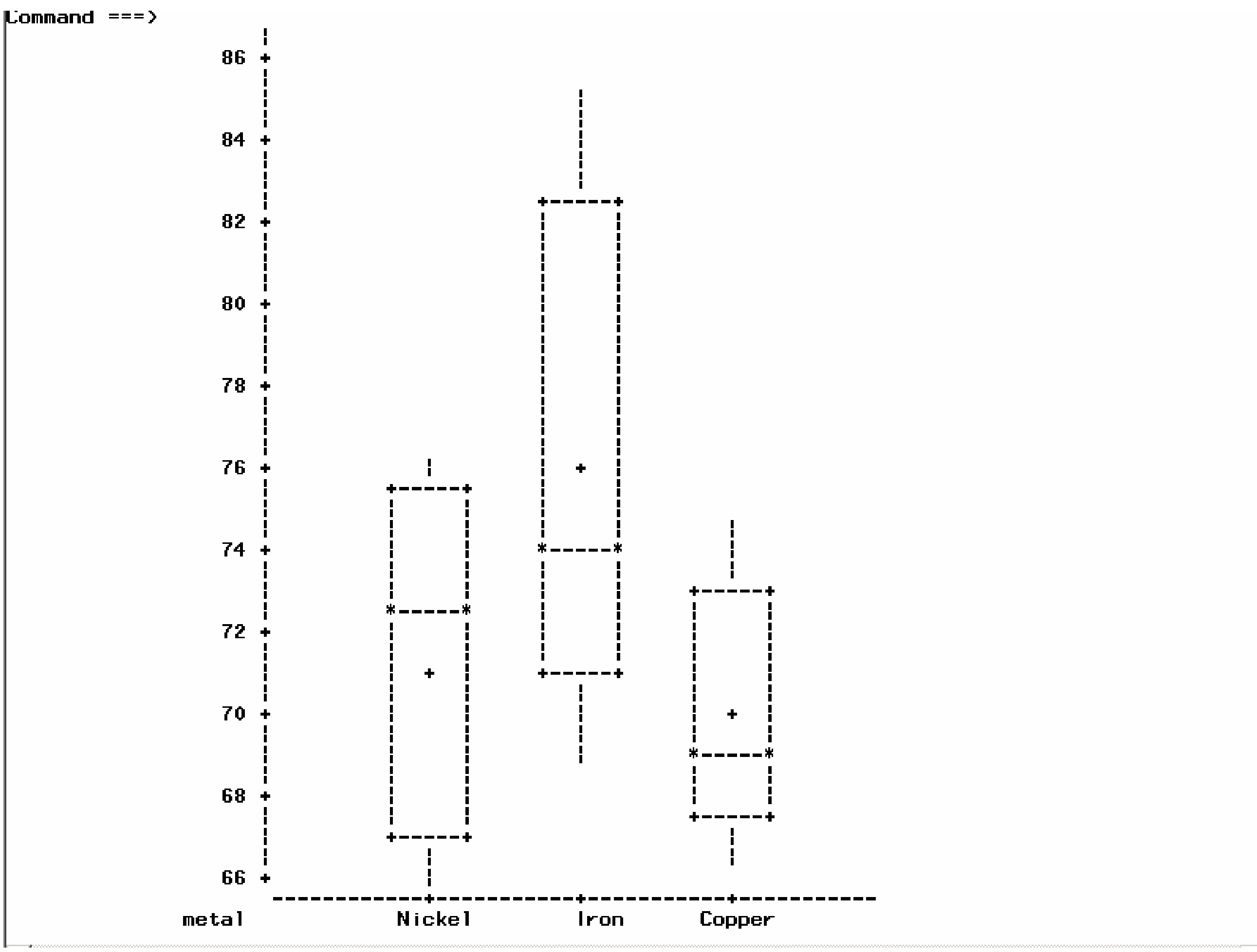
N	Mean	Std Dev	Minimum	Maximum
7	75.9000000	6.1378606	68.8000000	84.9000000

---------------------------------- metal=Copper ----------------------------------

Analysis Variable : pressure

N	Mean	Std Dev	Minimum	Maximum
7	70.1857143	3.1098921	66.4000000	74.5000000

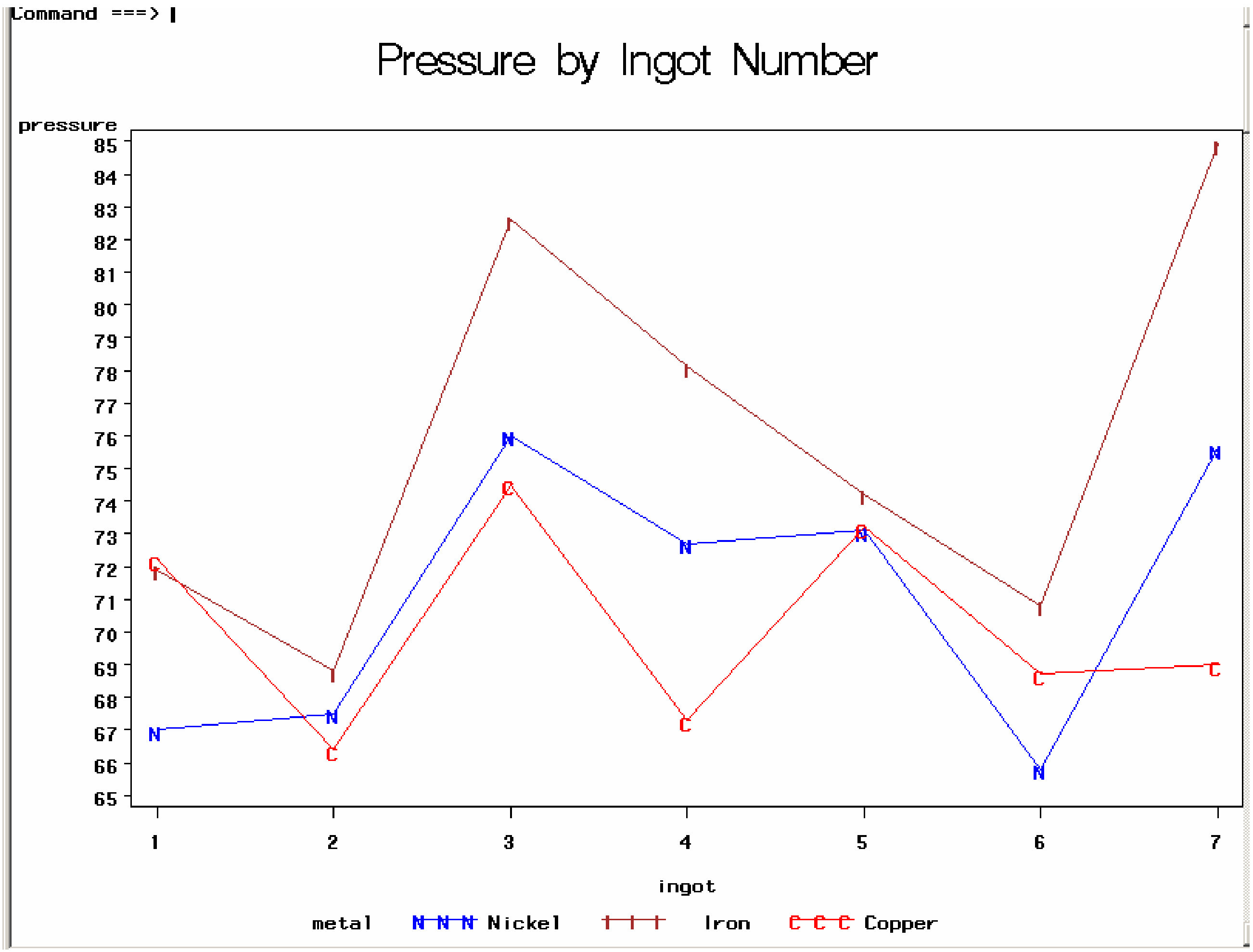
Box Plot Output from Proc Univariate



Graphing the Means

```
proc gplot data=metalic;
plot pressure*ingot=metal/haxis=axis1
   vminor=0;
   axis1 offset=(2,2) minor=none;
  symbol1 v='n' c=blue i=join;
  symbol2 v='i' c=brown i=join;
  symbol3 v='c' c=red i=join;
title 'Pressure by Ingot Number';
run;
```

Proc Gplot Output



SAS syntax for One-Way Fixed Effects model with Proc GLM

```
proc glm data=metalic;
  class metal;
  model pressure = metal;
  means metal/hovtest=levene (type=abs) scheffe tukey
     lsd;
  contrast 'nickel v iron' metal 1 -1 0;
  contrast 'nickel v copper' metal 1 0 -1;
  contrast 'iron v copper' metal 0 1 -1;
  estimate ‘nickel v iron’ metal 1 –1 0;
  estimate ‘nickel v copper’ metal 1 0 –1;
  estimate ‘iron v copper’ metal 0 1 –1;
output out=resdat1 r=resid1 p=pred1;
title 'One-Way Fixed Effects Model';
title2 'With Proc GLM';
run;
proc univariate data=resdat1 normal plot;
  var resid1;
title 'Test for normality of residuals';
run;
```

Contrasts

Constrasts are comparisons among means.
They are t tests between different levels of the same factor.

T = contrast/Se of contrast

$$t = \frac{c_1\bar{Y}_{.1} + c_2\bar{Y}_{.2} + \ldots + c_p\bar{Y}_{.p}}{\sqrt{MSerror \sum_{j=1}^{n} \frac{c_j^{\,2}}{n_j}}}$$

SAS contrast syntax

For simple contrasts of levels of a main effect, use 1 and –1 to identify the levels being contrasted.

contrast ‘copper v iron’ metal 1 –1 0;

Contrasts of interactions

Suppose you have 3 levels of a and 5 levels of b and you wish to compare the 2nd level of a with the 5th level of b in an interaction

In a 3 by 5 matrix, you are comparing
Cells 12 and 25

Variety						
Method	1	2	3	4	5	
A		X				
B					X	
C						

$$H_0 : \mu_{12} = \mu_{25}$$

In other words:

$$u + a_1 + b_2 + ab_{12} = u + a_2 + b_5 + ab_{25}$$

An Interaction contrast of ab12 and ab25

In other words:

$$(u + a_1 + b_2 + ab_{12}) - (u + a_2 + b_5 + ab_{25}) = 0$$

$$a_1 - a_2 + b_2 - b_5 + ab_{12} - ab_{25} = 0$$

$$a1 - a2 + 0a3 + 0b1 + b2 + 0b3 + 0b4 - b5$$

$$+ 0ab11 + ab12 + 0ab13 + 0ab14 + 0ab15$$

$$+ 0ab21 + 0ab22 + 0ab23 + 0ab24 - ab25$$

$$+ 0ab31 + 0ab32 + 0ab33 + 0ab34 + 0ab35 = 0$$

Contrast '*A*2 *v B*5' *method* 1 –1 0 *variety* 0 1 0 0 –1
method * var*iety* 0 1 0 0 0 0 0 0 0 –1 0 0 0 0 0;

Interpretation of the GLM output-1

```
        One-Way Fixed Effects Model                          145
               With Proc GLM
                             08:16 Wednesday, November 6, 2002

             The GLM Procedure

          Class Level Information

Class          Levels    Values

metal               3    Copper Iron Nickel

        Number of observations    21
```

GLM output-2

Command ===>

One-Way Fixed Effects Model 146
With Proc GLM
08:16 Wednesday, November 6, 2002

The GLM Procedure

Dependent Variable: pressure

Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	2	131.9009524	65.9504762	3.02	0.0738
Error	18	392.7485714	21.8193651		
Corrected Total	20	524.6495238			

R-Square	Coeff Var	Root MSE	pressure Mean
0.251408	6.452248	4.671120	72.39524

Source	DF	Type I SS	Mean Square	F Value	Pr > F
metal	2	131.9009524	65.9504762	3.02	0.0738

Source	DF	Type III SS	Mean Square	F Value	Pr > F
metal	2	131.9009524	65.9504762	3.02	0.0738

The R^2 = Model Variance/Total Variance
The Coeff Var = sd/mean

SAS syntax for One-Way Fixed Effects model with PROC Mixed

SAS Command Syntax

Class effects are all factors.
Only the **fixed effects are specified in the model statement**.

```
proc mixed data=metalic covtest mmeq;
   class metal;
   model pressure = metal;
  lsmeans metal/pdiff;
  lsmeans metal/adjust=bon;
Title 'One-Way Fixed Effects Model';
title2 'With PROC Mixed';
run;
```

Mixed statements

Proc Options:

mmeq : model equation reported

Class: specifies the discrete factors

Model statement:

model pressure=metal;

(model statement indicates the dependent variable=fixed effect;) Only fixed effects are included in the model statement.

LSmeans metal/pdiff;

There are only lsmeans statements in proc model. The pdiff statement requests that the differences between the levels of the fixed effects be estimated and reported.

LSmeans metal/adjust=bon;

Perform a bonferroni correction to the probability for multiple comparisons.

Mixed Model Syntax

For Models with random effects:

Random variety/type = cs subject=id;

The random statement contains only random effects of the model. This specifies the parameters of the random effects of the Zu part of the model– the variance matrix of which is the G matrix.

The Random statement

1. This specifies the random effects. Interactions with random effects are also random effects.
2. The solution option requests that the solution for the random parameters be printed. These estimates are best linear unbiased predictors:

The Repeated Statement

- The repeated statement specifies the covariance structure for repeated observations on the subjects, as in a repeated measures design.
- “In repeated measures situations, the Mixed model approach is both more flexible and applicable than the conventional univariate and multivariate appraches [SAS Stat Software: Changes and Enhancements through 6.11.”(1996). Cary, NC:SAS Institute, p.536].
- SAS Proc Mixed can handle a wide variety of covariance structures that none of the mentioned approaches can deal with; the conventional approaches depend on a Gaussian residual distribution.

Output Interpretation

```
Command ===>
                         One-Way Fixed Effects Model                          19
                               With PROC Mixed
                                            10:48 Thursday, November 7, 2002

                             The Mixed Procedure

                              Model Information

            Data Set                     WORK.METALIC
            Dependent Variable           pressure
            Covariance Structure         Diagonal
            Estimation Method            REML
            Residual Variance Method     Profile
            Fixed Effects SE Method      Model-Based
            Degrees of Freedom Method    Residual

                          Class Level Information

              Class     Levels    Values

              metal          3    Copper Iron Nickel

                                 Dimensions

                     Covariance Parameters             1
                     Columns in X                      4
                     Columns in Z                      0
                     Subjects                          1
                     Max Obs Per Subject              21
                     Observations Used                21
                     Observations Not Used             0
                     Total Observations               21
```

Output Interpretation- cont'd

- Mixed Model equations from option mmeq and covariance parameter estimate from option covtest

Command ===>

Mixed Model Equations

Row	Effect	metal	Col1	Col2	Col3	Col4	Col5
1	Intercept		0.9624	0.3208	0.3208	0.3208	69.6766
2	metal	Copper	0.3208	0.3208			22.5167
3	metal	Iron	0.3208		0.3208		24.3499
4	metal	Nickel	0.3208			0.3208	22.8100

Covariance Parameter Estimates

Cov Parm	Estimate	Standard Error	Z Value	Pr Z
Residual	21.8194	7.2731	3.00	0.0013

Information Criteria and Fixed Effect test

```
Command ===>
                One-Way Fixed Effects Model                        20
                      With PROC Mixed
                                 10:48 Thursday, November 7, 2002

                    The Mixed Procedure

                      Fit Statistics

            -2 Res Log Likelihood         112.4
            AIC (smaller is better)       114.4
            AICC (smaller is better)      114.7
            BIC (smaller is better)       115.3

               Type 3 Tests of Fixed Effects

                    Num      Den
          Effect     DF       DF     F Value    Pr > F

          metal       2       18        3.02    0.0738
```

$$Akaike\ Information\ Criterion$$

$$AIC = -2LL + 2d$$

$$where\ \ d = p + q \quad q = number\ of\ \text{cov}ariance\ \ parameters$$

$$p = rank\ \ of\ \ matrix$$

$$Schwarz\ \ Information\ criterion$$

$$BIC\ \ = -2LL + d\ \ln(n)$$

$$Hurvich\ \&Tsay$$

$$AICC\ = -2LL + 2q(N/N - q - 1)$$

The latter is best for smaller samples.

Fixed Effects Output Interpretation-cont'd

The Fixed effect is the independent factor in the one-way layout.

The Fixed effects are always specified only in the model statement.

The Type 3 test is one where the Sums of Squares can handle unbalanced designs. This is a simultaneous sums of squares, unlike that of a type I sum of squares, which is a sequential sum of squares.

Least Square Means and differences among them

Least Squares Means

Effect	metal	Estimate	Standard Error	DF	t Value	Pr > |t|
metal	Copper	70.1857	1.7655	18	39.75	<.0001
metal	Iron	75.9000	1.7655	18	42.99	<.0001
metal	Nickel	71.1000	1.7655	18	40.27	<.0001
metal	Copper	70.1857	1.7655	18	39.75	<.0001
metal	Iron	75.9000	1.7655	18	42.99	<.0001
metal	Nickel	71.1000	1.7655	18	40.27	<.0001

Differences of Least Squares Means

Effect	metal	_metal	Estimate	Standard Error	DF	t Value	Pr > |t|
metal	Copper	Iron	-5.7143	2.4968	18	-2.29	0.0344
metal	Copper	Nickel	-0.9143	2.4968	18	-0.37	0.7185
metal	Iron	Nickel	4.8000	2.4968	18	1.92	0.0705
metal	Copper	Iron	-5.7143	2.4968	18	-2.29	0.0344

Bonferroni adjustment for multiple comparisons

Differences of Least Squares Means

Effect	metal	_metal	Adjustment	Adj P
metal	Copper	Iron		.
metal	Copper	Nickel		.
metal	Iron	Nickel		.
metal	Copper	Iron	Bonferroni	0.1032

Two-Way Random Effects Models with Proc Mixed from Milliken and Johnson's Analysis of Messy Data Course.

```
proc mixed;               /* THE NEW PROC MIXED */
   class row col;
   model y = ;
   random row col row*col / solution;
   estimate 'grand mean' intercept 1;
   estimate 'row 1 col 1' intercept 1 | row 1 col 1 row*col 1;
   estimate 'row 1 col 2' intercept 1 | row 1 col 0 1 row*col 0 1;
   estimate 'row 1 col 3' intercept 1 | row 1 col 0 0 1 row*col 0 0 1;
   estimate 'row 2 col 1' intercept 1 | row 0 1 col 1 row*col 0 0 0 1;
   estimate 'row 2 col 2' intercept 1 | row 0 1 col 0 1 row*col 0 0 0 0 1;
   estimate 'row 2 col 3' intercept 1 | row 0 1 col 0 0 1 row*col 0 0 0 0 0
title2 'Two-Way Random Effects Model';
title3 ' With the Mixed Procedure';
   run;
```

There are no fixed effects specified on the right-hand side of the Model statement.
Both discrete factors specified in the class statement are mentioned in the random statement, along with their interaction.
The random specification is followed by the solution option.

The Random statement

1. This specifies the random effects. Interactions with random effects are also random effects.
2. The solution option requests that the solution for the random parameters be printed. These estimates are best linear unbiased predictors:

Estimate statements

- This is exactly like a contrast statement except only one-row L matrices are permitted. The actual estimate $L'\beta$ is printed along with its approximate standard error, t-test, and t-probability.
- They are useful in estimating random effects and their significance.

Output interpretation

```
Command ===>
Two-Way Random Effects Model
 With the Mixed Procedure

The Mixed Procedure

                  Model Information

Data Set                         WORK.TWOWAY
Dependent Variable               y
Covariance Structure             Variance Components
Estimation Method                REML
Residual Variance Method         Profile
Fixed Effects SE Method          Model-Based
Degrees of Freedom Method        Containment

             Class Level Information

Class     Levels    Values

row            2    1 2
col            3    1 2 3

            Dimensions

Covariance Parameters             4
Columns in X                      1
Columns in Z                     11
Subjects                          1
Max Obs Per Subject              14
Observations Used                14
Observations Not Used             0
Total Observations               14
```

More output

```
Command ===>

                    Iteration History

Iteration    Evaluations    -2 Res Log Like       Criterion

        0              1        71.76960400
        1              3        66.41375224      0.00177521
        2              2        66.38519638      0.00014277
        3              1        66.38203013      0.00000069
        4              1        66.38201547      0.00000000


                Convergence criteria met.


Covariance Parameter
      Estimates

Cov Parm       Estimate

row                   0
col                   0
row*col          9.2425
Residual         3.8398
```

Solutions and Estimates for Random Effects

Two-Way Random Effects Model
With the Mixed Procedure

The Mixed Procedure

Fit Statistics

-2 Res Log Likelihood	66.4
AIC (smaller is better)	70.4
AICC (smaller is better)	71.6
BIC (smaller is better)	67.8

Solution for Random Effects

Effect	row	col	Estimate	Std Err Pred	DF	t Value	Pr > |t|
row	1		0	.	.	.	.
row	2		0	.	.	.	.
col		1	0	.	.	.	.
col		2	0	.	.	.	.
col		3	0	.	.	.	.
row*col	1	1	-3.3858	1.5909	8	-2.13	0.0660
row*col	1	2	-0.7077	1.6852	8	-0.42	0.6856
row*col	1	3	4.2604	1.6852	8	2.53	0.0354
row*col	2	1	1.7763	1.6852	8	1.05	0.3226
row*col	2	2	0.4204	1.5909	8	0.26	0.7982
row*col	2	3	-2.3637	1.6852	8	-1.40	0.1983

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
grand mean	14.8547	1.3503	1	11.00	0.0577
row 1 col 1	11.4689	1.0729	1	10.69	0.0594
row 1 col 2	14.1470	1.2820	1	11.03	0.0575
row 1 col 3	19.1150	1.2820	1	14.91	0.0426
row 2 col 1	16.6310	1.2820	1	12.97	0.0490
row 2 col 2	15.2751	1.0729	1	14.24	0.0446
row 2 col 3	12.4910	1.2820	1	9.74	0.0651

The Mixed Model: A Two-Way Layout

- **data** grass;
- input method $ variety @;
- do i =**1** to **6**;
- input yield @;
- output;
- end;
- datalines;
- AA 1 22.1 24.1 19.1 22.1 25.1 18.1
- AA 2 27.1 15.1 20.6 28.6 15.1 24.6
- AA 3 22.3 25.8 22.8 28.3 21.3 18.3
- AA 4 19.8 28.3 26.8 27.3 26.8 26.8
- AA 5 20.0 17.0 24.0 22.5 28.0 22.5
- BB 1 13.5 14.5 11.5 6.0 27.0 18.0
- BB 2 16.9 17.4 10.4 19.4 11.9 15.4
- BB 3 15.7 10.2 16.7 19.7 18.2 12.2
- BB 4 15.1 6.5 17.1 7.6 13.6 21.1
- BB 5 21.8 22.8 18.8 21.3 16.3 14.3
- CC 1 19.0 22.0 20.0 14.5 19.0 16.0
- CC 2 20.0 22.0 25.5 16.5 18.0 17.5
- CC 3 16.4 14.4 21.4 19.9 10.4 21.4
- CC 4 24.5 16.0 11.0 7.5 14.5 15.5
- CC 5 11.8 14.3 21.3 6.3 7.8 13.8
- **proc print**;
- var method variety yield;
- title3 'Advanced Linear Modelswith Emphasis on Mixed Models';
- title4 'Two-Way Fixed Effects Model';
- title5 'Data from p. 45';
- **run**;

Mixed Model Syntax

method is a fixed effect
variety and its interaction with method are random effects

```
proc mixed data=grass covtest ratio ;
  class method variety;
  model yield= method;
  random variety method*variety;
  lsmeans method/pdiff;
  contrast 'method 1 v 2' method 1 -1 0;
  contrast 'method 1 v 3' method 1 0 -1;
  contrast 'method 2 v 3' method 0 1 -1;
  contrast 'method 1 v 2 and 3' method 2 -1 -1;
  estimate 'method 1 v 2' method 1 -1 0;
  estimate 'method 1 v 3' method 1 0 -1;
  estimate 'method 2 v 3' method 0 1 -1;
  estimate 'method 1 v 2 and 3' method 2 -1 -1;
title 'Two-Way Full-Factorial Mixed Model';
title2 'Mixed Effects Model';
run;
```

Mixed Model Output

```
Command ===>
                 Two-Way Full-Factorial Mixed Model                        45
                        Mixed Effects Model
                                 11:56 Thursday, November 7, 2002

                        The Mixed Procedure

                         Model Information

   Data Set                         WORK.GRASS
   Dependent Variable               yield
   Covariance Structure             Variance Components
   Estimation Method                REML
   Residual Variance Method         Profile
   Fixed Effects SE Method          Model-Based
   Degrees of Freedom Method        Containment


                      Class Level Information

    Class       Levels     Values

    method           3     AA BB CC
    variety          5     1 2 3 4 5


                            Dimensions

             Covariance Parameters               3
             Columns in X                        4
             Columns in Z                       20
             Subjects                            1
             Max Obs Per Subject                90
             Observations Used                  90
             Observations Not Used               0
             Total Observations                 90


                         Iteration History

 Iteration     Evaluations     -2 Res Log Like       Criterion

         0               1        523.51269564
```

Convergence attained

Class Level Information

Class	Levels	Values
method	3	AA BB CC
variety	5	1 2 3 4 5

Dimensions

Covariance Parameters	3
Columns in X	4
Columns in Z	20
Subjects	1
Max Obs Per Subject	90
Observations Used	90
Observations Not Used	0
Total Observations	90

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	523.51269564	
1	3	522.10817863	0.00003194
2	1	522.10265442	0.00000013
3	1	522.10263051	0.00000000

Convergence criteria met.

Mixed Output-cont'd

Command ===>

Two-Way Full-Factorial Mixed Model 46
Mixed Effects Model
11:56 Thursday, November 7, 2002

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Ratio	Estimate	Standard Error	Z Value	Pr Z
variety	0	0	.	.	.
method*variety	0.1061	2.0842	2.2523	0.93	0.1774
Residual	1.0000	19.6502	3.2089	6.12	<.0001

Fit Statistics

-2 Res Log Likelihood	522.1
AIC (smaller is better)	526.1
AICC (smaller is better)	526.2
BIC (smaller is better)	525.3

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
method	2	8	14.82	0.0020

Estimates

Label	Estimate	Standard Error	DF	t Value	Pr > |t|
method 1 v 2	7.3133	1.4641	8	4.99	0.0011
method 1 v 3	6.4033	1.4641	8	4.37	0.0024
method 2 v 3	-0.9100	1.4641	8	-0.62	0.5515
method 1 v 2 and 3	13.7167	2.5360	8	5.41	0.0006

Contrasts from Mixed output

Command ===>

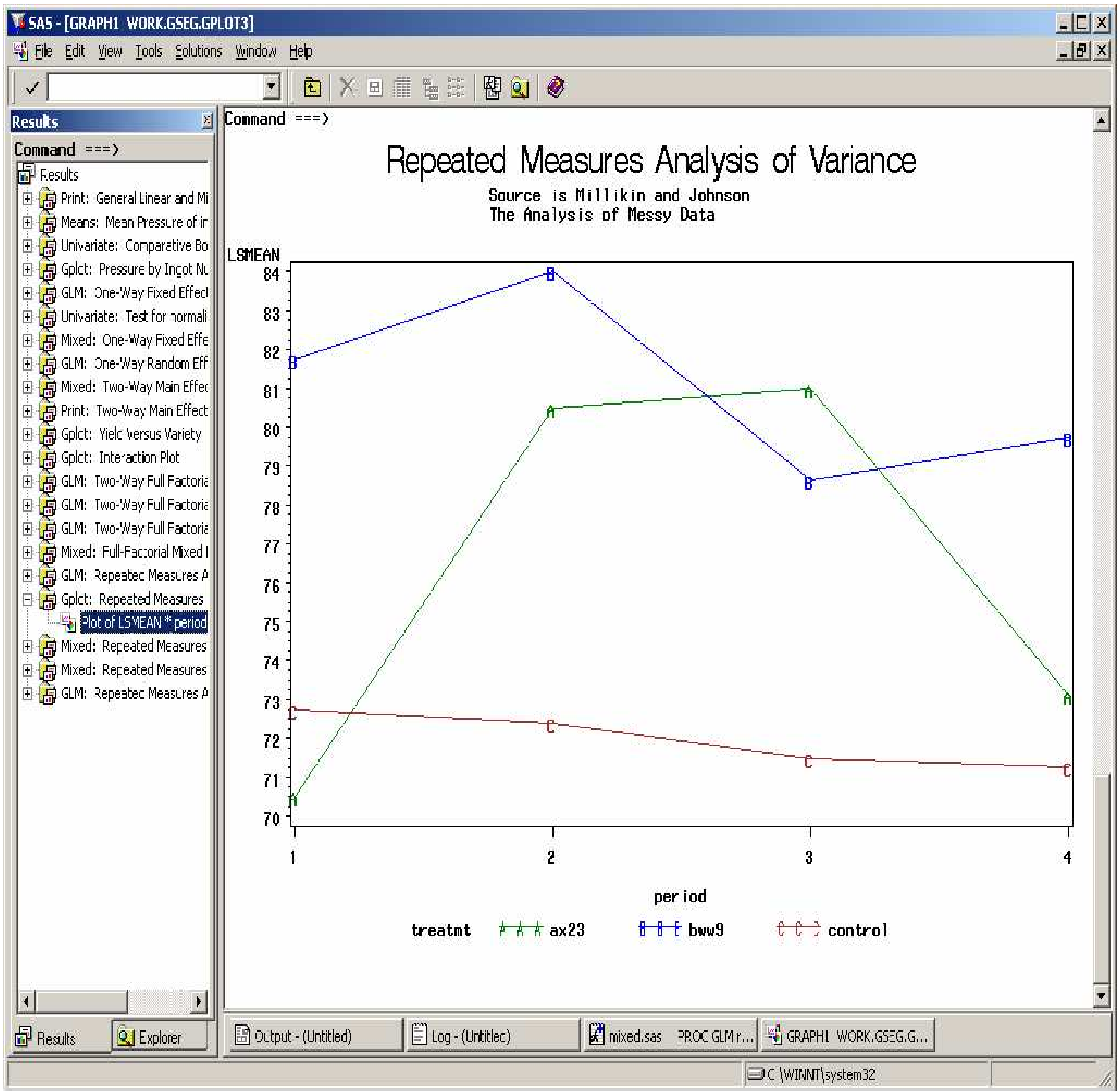
Contrasts

Label	Num DF	Den DF	F Value	Pr > F
method 1 v 2	1	8	24.95	0.0011
method 1 v 3	1	8	19.13	0.0024
method 2 v 3	1	8	0.39	0.5515
method 1 v 2 and 3	1	8	29.26	0.0006

LSmeans and Differences among them

```
Command ===>
                          Two-Way Full-Factorial Mixed Model                        47
                                 Mixed Effects Model
                                             11:56 Thursday, November 7, 2002

                                  The Mixed Procedure

                                  Least Squares Means

                                                Standard
    Effect      method      Estimate       Error        DF      t Value      Pr > |t|

    method      AA           23.0100      1.0353         8        22.23        <.0001
    method      BB           15.6967      1.0353         8        15.16        <.0001
    method      CC           16.6067      1.0353         8        16.04        <.0001

                          Differences of Least Squares Means

                                                     Standard
  Effect     method     _method     Estimate         Error       DF     t Value     Pr > |t|

  method     AA         BB            7.3133        1.4641        8        4.99      0.0011
  method     AA         CC            6.4033        1.4641        8        4.37      0.0024
  method     BB         CC           -0.9100        1.4641        8       -0.62      0.5515
```

Repeated Measures Anova with Proc Mixed



Data for Repeated Measures ANOVA from Millikin and Johnson Analysis of Messy Data

```
title 'Repeated Measures Analysis of Variance';
title2 'Source is Millikin and Johnson       ';
title3 'The Analysis of Messy Data           ';
data repeat;
input t1 t2 t3 t4 treatmt $;
label t1='Time1' t2='Time2' t3='Time3'
     t4='Time4';
datalines;
t1 t2 t3 t4 trt
72 86 81 77 ax23
78 83 88 81 ax23
71 82 81 75 ax23
72 83 83 69 ax23
66 79 77 66 ax23
74 83 84 77 ax23
62 73 78 70 ax23
69 75 76 70 ax23
85 86 83 80 bww9
82 86 80 84 bww9
71 78 70 75 bww9
83 88 79 81 bww9
86 85 76 76 bww9
85 82 83 80 bww9
79 83 80 81 bww9
83 84 78 81 bww9
69 73 72 74 control
66 62 67 73 control
84 90 88 87 control
80 81 77 72 control
72 72 69 70 control
65 62 65 61 control
75 69 69 68 control
71 70 65 65 control
;
/* Example of Repeated Measures Analysis Millikin and Johnson */
/* The Analysis of Messy Data 26.1
```

Repeated Measures Proc Mixed Syntax

```
data times; set lsm;
  period = substr(_NAME_,2,1);
symbol1 i=join c=green;
symbol2 i=join c=blue;
symbol3 i=join c=brown;
symbol4 i=join c=red;
proc gplot;
  plot lsmean*period=treatmt;
run;
data mixed; set repeat;
  person = _N_;
  y = t1; period = 1; output;
  y = t2; period = 2; output;
  y = t3; period = 3; output;
  y = t4; period = 4; output;
  drop t1--t4;
run;
proc mixed data=mixed;  /* repeated measures using proc mixed */
  class treatmt period person;
  model y = treatmt|period;
  repeated / type=ar(1) sub=person(treatmt);
  lsmeans treatmt*period;
run;
```

Repeated Measures Mixed Syntax

- Repeated Statement
  - Specifies the error matrix of Σ
  - This is called the R matrix in SAS
  - There is only 1 repeated statement
  - The within-subject (split plot) effect is specified.
- Repeated time/type=ar(1) subject=id;
- Subject = specifies the units of measurement which can be correlated within the id.
- Type= specifies the structure of the error covariance matrix R. Some types are un, ar(1), arh(1), cs, csh, hf, toep, etc.

Selected Covariance Structures

- Type=VC

$$VC(default) = \begin{bmatrix} \sigma_1^{\;2} & 0 & 0 & 0 \\ 0 & \sigma_2^{\;2} & 0 & 0 \\ 0 & 0 & \sigma_3^{\;2} & 0 \\ 0 & 0 & 0 & \sigma_4^{\;2} \end{bmatrix}$$

Type=Unstructured

$$Unstructured = \begin{bmatrix} \sigma_{11}^{\;2} & \sigma_{12}^{\;2} & \sigma_{13}^{\;2} & \sigma_{14}^{\;2} \\ \sigma_{21}^{\;2} & \sigma_{22}^{\;2} & \sigma_{23}^{\;2} & \sigma_{24}^{\;2} \\ \sigma_{31}^{\;2} & \sigma_{32}^{\;2} & \sigma_{33}^{\;2} & \sigma_{34}^{\;2} \\ \sigma_{41}^{\;2} & \sigma_{42}^{\;2} & \sigma_{43}^{\;2} & \sigma_{44}^{\;2} \end{bmatrix}$$

Selected Covariance Structures-Cont'd

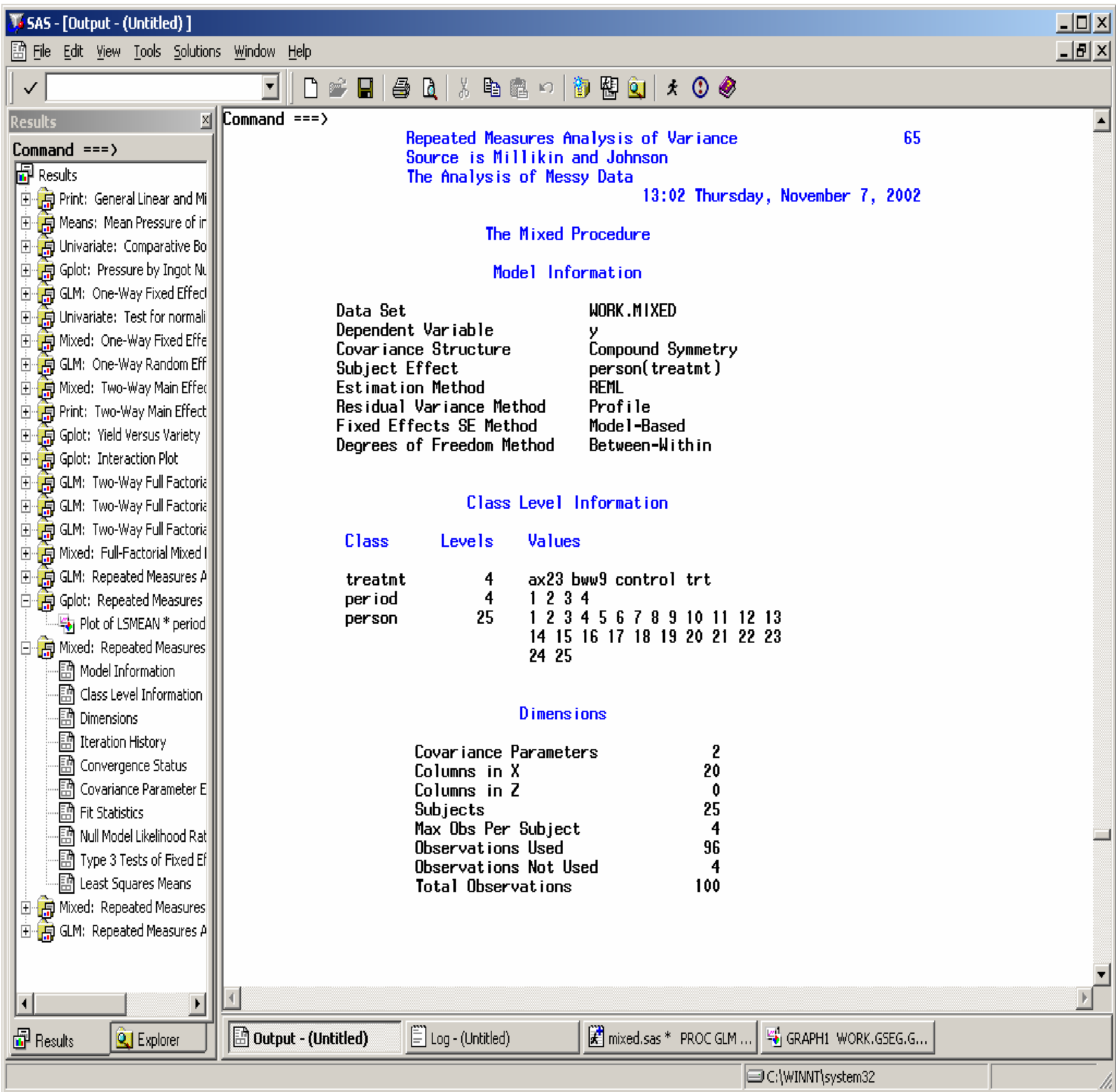
- Compound Symmetry type=CS

$$CS = \begin{bmatrix} \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 & \sigma_1^2 \\ \sigma_1^2 & \sigma_1^2 & \sigma_1^2 & \sigma^2 + \sigma_1^2 \end{bmatrix}$$

- First-Order Autoregressive type=AR(1)

$$AR(1) = \begin{bmatrix} 1 & \rho & \rho^2 & \rho^3 \\ \rho & 1 & \rho & \rho^2 \\ \rho^2 & \rho & 1 & \rho \\ \rho^3 & \rho^2 & \rho & 1 \end{bmatrix}$$

Basic Repeated Measures Output




```
Command ===>
                 Repeated Measures Analysis of Variance                65
                 Source is Millikin and Johnson
                 The Analysis of Messy Data
                                    13:02 Thursday, November 7, 2002

                          The Mixed Procedure

                           Model Information

             Data Set                     WORK.MIXED
             Dependent Variable           y
             Covariance Structure         Compound Symmetry
             Subject Effect               person(treatmt)
             Estimation Method            REML
             Residual Variance Method     Profile
             Fixed Effects SE Method      Model-Based
             Degrees of Freedom Method    Between-Within

                        Class Level Information

              Class      Levels    Values

              treatmt         4    ax23 bww9 control trt
              period          4    1 2 3 4
              person         25    1 2 3 4 5 6 7 8 9 10 11 12 13
                                   14 15 16 17 18 19 20 21 22 23
                                   24 25

                              Dimensions

                       Covariance Parameters            2
                       Columns in X                    20
                       Columns in Z                     0
                       Subjects                        25
                       Max Obs Per Subject              4
                       Observations Used               96
                       Observations Not Used            4
                       Total Observations             100
```

Convergence attained

Iteration History

Iteration	Evaluations	-2 Res Log Like	Criterion
0	1	557.24212530	
1	1	487.17690350	0.00000000

Convergence criteria met.

Fit statistics and Type 3 tests of Fixed Effects

Command ===>

The Analysis of Messy Data
13:02 Thursday, November 7, 2002

The Mixed Procedure

Covariance Parameter Estimates

Cov Parm	Subject	Estimate
CS	person(treatmt)	25.8016
Residual		7.2773

Fit Statistics

-2 Res Log Likelihood	487.2
AIC (smaller is better)	491.2
AICC (smaller is better)	491.3
BIC (smaller is better)	493.6

Null Model Likelihood Ratio Test

DF	Chi-Square	Pr > ChiSq
1	70.07	<.0001

Type 3 Tests of Fixed Effects

Effect	Num DF	Den DF	F Value	Pr > F
treatmt	2	21	5.95	0.0090
period	3	63	12.95	<.0001
treatmt*period	6	63	12.16	<.0001

LSMeans

Least Squares Means

Effect	treatmt	period	Estimate	Standard Error	DF	t Value	Pr > |t|
treatmt*period	ax23	1	70.5000	2.0334	63	34.67	<.0001
treatmt*period	ax23	2	80.5000	2.0334	63	39.59	<.0001
treatmt*period	ax23	3	81.0000	2.0334	63	39.83	<.0001
treatmt*period	ax23	4	73.1250	2.0334	63	35.96	<.0001
treatmt*period	bww9	1	81.7500	2.0334	63	40.20	<.0001
treatmt*period	bww9	2	84.0000	2.0334	63	41.31	<.0001
treatmt*period	bww9	3	78.6250	2.0334	63	38.67	<.0001
treatmt*period	bww9	4	79.7500	2.0334	63	39.22	<.0001
treatmt*period	control	1	72.7500	2.0334	63	35.78	<.0001
treatmt*period	control	2	72.3750	2.0334	63	35.59	<.0001

13:02 Thursday, November 7, 2002

The Mixed Procedure

Least Squares Means

Effect	treatmt	period	Estimate	Standard Error	DF	t Value	Pr > |t|
treatmt*period	control	3	71.5000	2.0334	63	35.16	<.0001
treatmt*period	control	4	71.2500	2.0334	63	35.04	<.0001

Fitting Mixed Models

There are a number of error covariance matrix patterns that can be used to estimate the model.

Several patterns are used.

Caveat: **As long as these models are nested,** the model with the lowest AICC is the one that is selected. Unless the models are nested, this criterion should not be used.

Among the types available:

Unstructured with : Type=un

Compound symmetry with: Type=cs

First-Order Autoregressive with

Type = ar(1)

Teopliz with : Type = toep

Compound symmetry with heterogeneity with: Type = csh

Etc.

Other Experimental Designs

- Multi-location designs
  - The tests are nested in the location
  - Such designs can include crossover designs where instead of doctors, there would be locations.

Latin Square Designs

Shrinkage may be a function of the material

Indicated by the letter and the setting of heat

Indicated by the position as well as the setting for the treatment.

This is an orthogonal design where the dv is a function of

- Position, setting, and type of material.

Other Experimental Designs-cont'd

- Lattice Square Designs (an orthogonal design)

	Position			
Sequence	1	2	3	4
1	A	C	D	B
2	C	A	B	D
3	D	B	A	C
4	B	D	C	A

Cross-over Design

- Crossover Designs: Used to compare 2 or more drugs or treatments. The treatment or drug is administered in a particular sequence to a patient over the time periods of observation.
- To counter the carryover effects of using the same sequence of treatment for each patient, the sequence is varied. One way to do this is to use a double latin square design.

Crossover Design-cont'd

- Suppose there are 6 patients, who all get a baseline physical. After that, they are given a sequence of three drugs—a placebo, a standard drug, and a test drug. Therefore, they have four visits to the 2 doctors running the tests.
- The structure of the experiment after the baseline visit might look like a double latin square:

Doctor 1					Doctor 2			
patient		1	2	3	4	5	6	
Visit	1	Baseline visit						
	2	A	B	C	A	B	C	
	3	B	C	A	B	C	A	
	4	C	A	B	C	A	B	

The Cross-over Design Model

$$y_{ijk} = \mu + seq + patient(sequence) + visit + treatmnt$$
$$+ resid\_pat + resid\_trtmt$$
$$patient(sequence) \text{ is random}$$

$$\left.\begin{array}{r} sequence \\ patient(sequence) \end{array}\right\} \text{is type I SS}$$

$$\left.\begin{array}{l} visit \\ treatment \end{array}\right\} \text{type III SS}$$

Crossover design SAS Syntax

```
Proc mixed order=internal;
  class sequence patient visit trt;
Model hr = sequence visit drug
     resid_trt  resid_pat/solution
     ddfm=satterth;
     random patient(sequence);
     lsmeans drug/pdiff cl e;
  run;
```

Hierarchical Linear Models

1. Hierarchical linear models multilevel random coefficient models.
2. SAS Proc Mixed can be used to model multi-level models.
3. Suppose we have a two level model
4. We need a random statement for coefficients of the first level of the model and meanses and sector are the second level predictors (Prof. Judy Singer,in “Using SAS Proc Mixed to Fit Multilevel models” gives the example):

```
Proc mixed covtest noitprint;
  class school;
  model mathach= meanses sector meanses*cses
     sector*cses/solution ddfm =   bw;
  random intercept cses/subject=school;
Run;
```

References

1. Latour, D. ,Littell, R.C., Brocklebank, J., Dockery, J., & Etchison, T. (1996) Advanced General Linear Models with an Emphasis on Mixed Models: Course Notes, Cary, NC: SAS Institute, pp.36-37.
2. Littell, R.C., Stroup, W., and Freund, R.J. (2002) SAS for Linear Models 4th ed., Cary,NC: SAS Institute, pp. 402-408.
3. Littell, R.C.,Milliken, G., Stroup, W., and Wolfinger, R. (1996): SAS for Mixed Models, Cary, NC: SAS Publishing, 139.
4. Milliken, G. and Johnson,D. (1984) Analysis of Messy Data, Vols 1 and 2, New York:Van Nostrand Reinhold. Course notes.
5. Singer, Judy. "Using SAS PROC Mixed to fit Multilevel Models (1998)." Journal of Educational and Behavioral Statistics, Winter, Vol.24, No.4, pp.324-355.