How To: Analyze a Repeated Measures Experiment Using STATGRAPHICS Centurion

by

Dr. Neil W. Polhemus

July 26, 2006

Introduction

In a typical repeated measures experiment, multiple measurements are made on the same experiment unit or subject at several different times. For example, an individual may be given a drug and the effect of the drug monitored for some length of time after it is administered. As with a split-plot design, repeated measures designs typically have two different size experimental units: a smaller unit consisting of each time interval at which measurements are taken, and a larger unit consisting of the subject who is given the treatment. To analyze the data properly, it is necessary to account for the different unit sizes.

This “How To” guide shows how STATGRAPHICS Centurion can be used to analyze typical repeated measures experiments. Two examples are considered.

Example #1

The first example comes from Analysis of Messy Data - Volume 1: Designed Experiments (Van Nostrand Reinhold, 1992) by Milliken and Johnson. In this study, 2 experimental drugs and a control were each administered to 8 subjects (for a total of 24 subjects). The heart rate of each subject was measured at 4 different times after the drug was administered. The data are saved in the file *howto12a.sf6*, a portion of which is shown below:

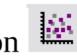
Subject	*Drug*	*Time*	*Heart Rate*
1	AX23	T1	72
1	AX23	T2	86
1	AX23	T3	81
1	AX23	T4	77
2	BWW9	T1	85
2	BWW9	T2	86
2	BWW9	T3	83
2	BWW9	T4	80
3	CONTROL	T1	69
3	CONTROL	T2	73
3	CONTROL	T3	72
3	CONTROL	T4	74
4	AX23	T1	78
4	AX23	T2	83
4	AX23	T3	88
4	AX23	T4	81
…	…	…	…

The goal of the experiment is to compare the effects of the two drugs to that of the control.

Step 1: Plot the Data

The first step when analyzing any new data set is to plot it. In this case, a coded scatterplot is very useful.

Procedure: X-Y Scatterplot

To plot the experimental data, let's begin by pushing the *X-Y Scatterplot* button  on the main toolbar. On the data input dialog box, indicate the variables to be plotted on each axis as shown below:

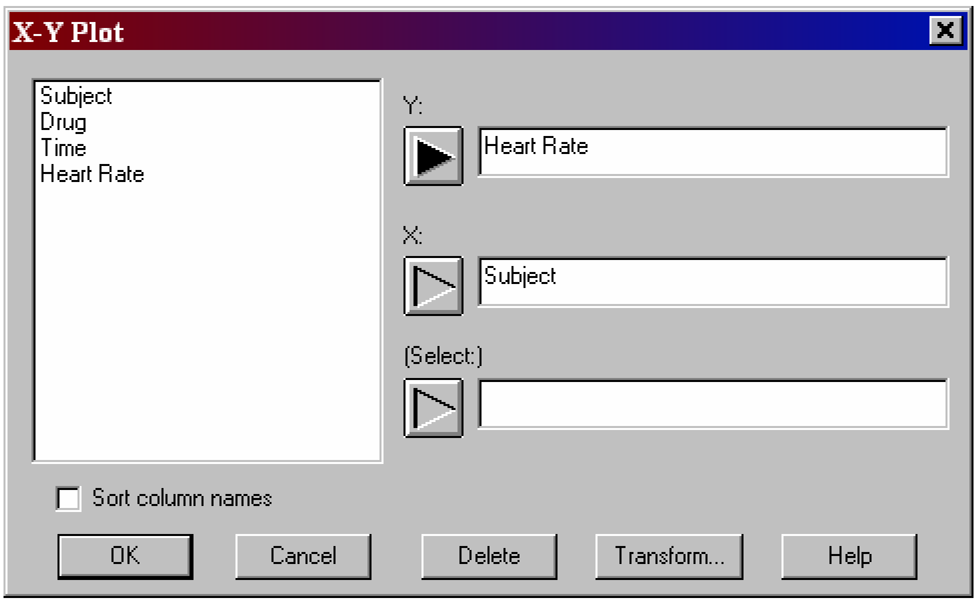
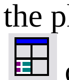


Figure 1: Data Input Dialog Box for X-Y Scatterplot

When the plot is displayed, double-click on the graph to enlarge it and press the *Pane Options* button  on the analysis toolbar. This will display the following dialog box:

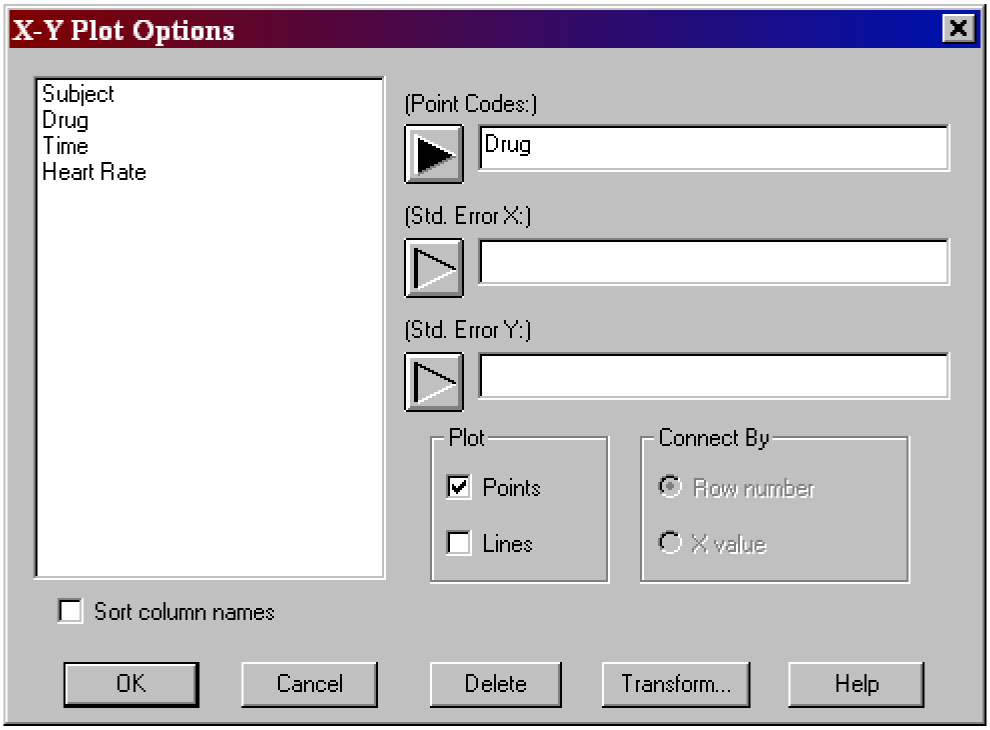


Figure 2: X-Y Plot Options Dialog Box

Enter *Drug* in the *Point Codes* field to generate a coded scatterplot:

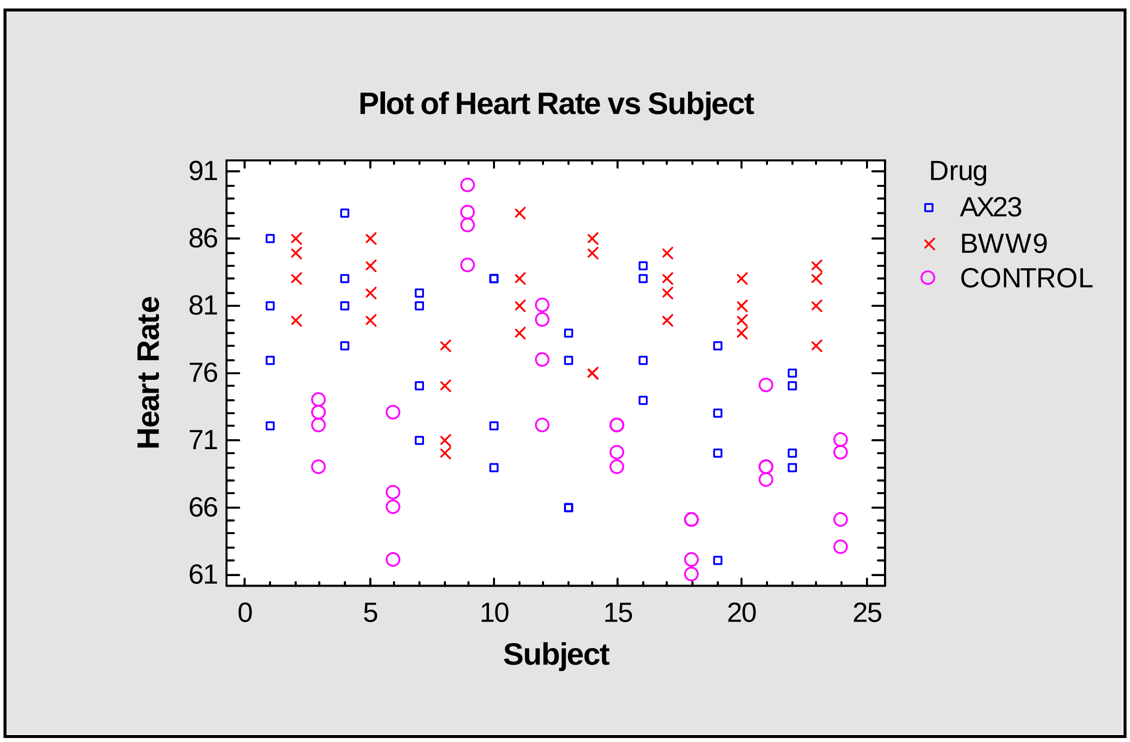


Figure 3: Coded X-Y Scatterplot

The four observations for each subject were taken at different points in time. Each type of point symbol represents a different treatment. The most noticeable aspect of the data is probably the relatively large amount of between subject variability compared to the within-subject variability.

Step 2: Analyze the Data

As with split plot designs, it is important to distinguish between the factors varied across subjects and the factors varied within subjects. In this case, *Drug* and *Subject* form one experiment, with *Drug* nested within *Subject* (since each subject got only one drug). *Time* and *Subject* form a second experiment, with *Time* crossed by *Subject* (since a measurement was taken at each level of *Time* for each *Subject*).

Procedure: General Linear Models

To analyze this data, we will use the *General Linear Models* procedure. This is accessed from the main STATGRAPHICS Centurion menu by selecting:

- If using the Classic menu: *Compare – Analysis of Variance – General Linear Models.*
- If using the Six Sigma menu: *Improve – Analysis of Variance – General Linear Models.*

The data input dialog box is shown below:

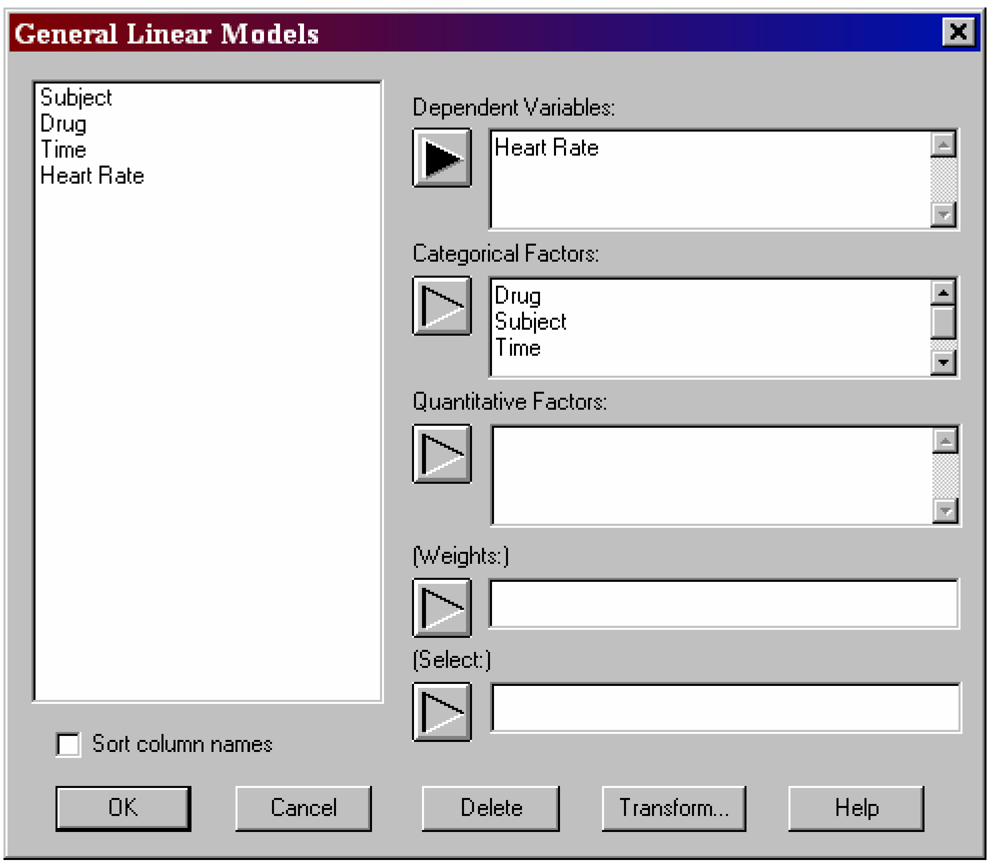


Figure 4: Data Input Dialog Box for General Linear Models

After completing the first dialog box, a second dialog box is displayed on which to specify the statistical model. It should be completed as shown below:

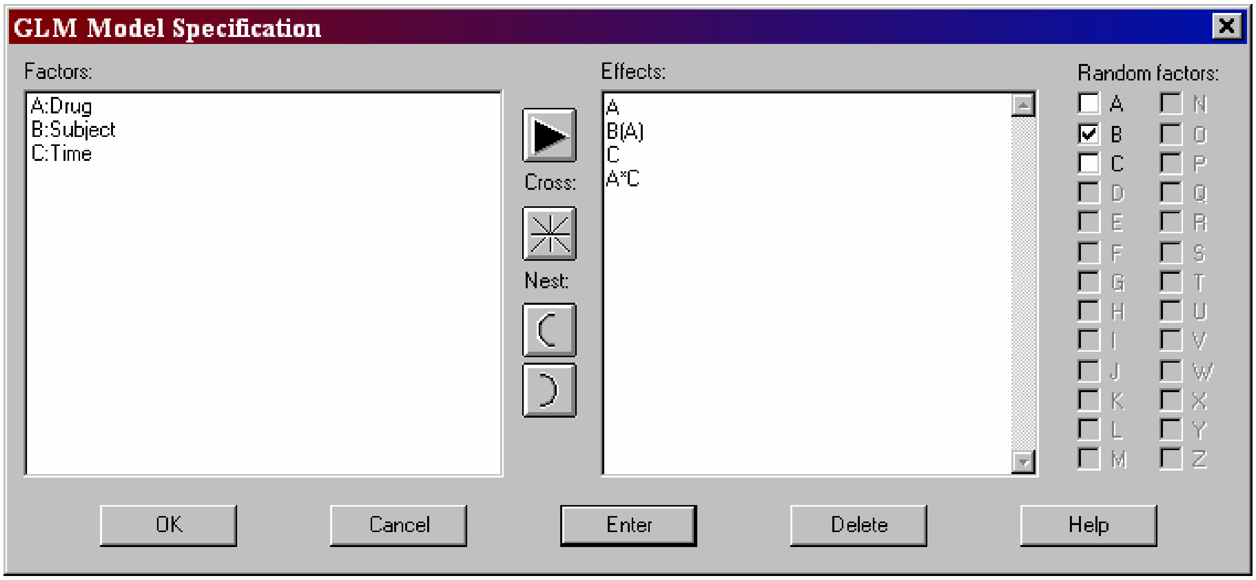


Figure 5: Model Specification Dialog Box for General Linear Models

Note the following:

1. The main effect of the factor *Drug* is specified by placing the single letter A in the *Effects* field. Since specific drugs were being studied, it is a fixed rather than a random factor.

2. *Subject* is entered using the notation B(A). This indicates that subjects (Factor B) are nested within drugs (Factor A). Factor B is also specified to be a *random* factor, since the subjects are assumed to be a random sample of many individuals who might take the drugs in the future.

3. *Time* is entered using the single letter C to represent its main effect and the notation A*C to represent an interaction between *Time* and *Drug*. Since the time intervals were the same for all subjects, it is also considered to be a fixed factor.

Pressing OK causes the specified model to be fit. The *Analysis Summary* pane summarizes the fitted model. The top section of that summary is shown below:

General Linear Models
Number of dependent variables: 1
Number of categorical factors: 3
A=Drug
B=Subject
C=Time
Number of quantitative factors: 0

Analysis of Variance for Heart Rate

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	4487.94	32	140.248	18.83	0.0000
Residual	469.219	63	7.44792		
Total (Corr.)	4957.16	95			

Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Drug	1333.0	2	666.5	5.99	0.0088
Subject(Drug)	2337.91	21	111.329	14.95	0.0000
Time	289.615	3	96.5382	12.96	0.0000
Drug*Time	527.417	6	87.9028	11.80	0.0000
Residual	469.219	63	7.44792		
Total (corrected)	4957.16	95			

Figure 6: GLM Analysis Summary – Top Section

The most important information in the above table is in the section labeled *Type III Sums of Squares*. The rightmost column of that table contains a P-Value for each term in the model. P-Values less than 0.05 indicate effects that are statistically significant at the 5% significance level, while values less than 0.01 indicate statistical significance at the 1% level. In this case, all effects are highly significant.

Also included in the *Analysis Summary* is the table shown below:

F-Test Denominators

Source	*Df*	*Mean Square*	*Denominator*
Drug	21.00	111.329	(2)
Subject(Drug)	63.00	7.44792	(5)
Time	63.00	7.44792	(5)
Drug*Time	63.00	7.44792	(5)

Variance Components

Source	*Estimate*
Subject(Drug)	25.9702
Residual	7.44792

Figure 7: GLM Analysis Summary – Bottom Section

The *F-Test denominators* indicate which line in the ANOVA table has been used to test the significance of each effect. For *Drug*, the (2) indicates that it has been compared against the *Subject(Drug)* mean square on the second line of the ANOVA table. The other factors have been compared to the *Residual* error in line 5.

Also shown are the estimates of the error components:

Between subject variance: $\hat{\sigma}_B^2 = 25.97$

Residual variance: $\hat{\sigma}_R^2 = 7.45$

Notice that the residual error is considerably smaller than the variance between the subjects.

Step 3: Display the Results

Once the important factors have been identified, it is useful to display the estimated effects graphically. Since there is a significant interaction between *Drug* and *Time*, the two factors need to be considered together. To create an *Interaction Plot*, use the *Graphs* button  on the analysis toolbar. You can then use *Pane Options* to indicate that *Time* (the "second factor") should be plotted along the horizontal axis. This displays the following plot:

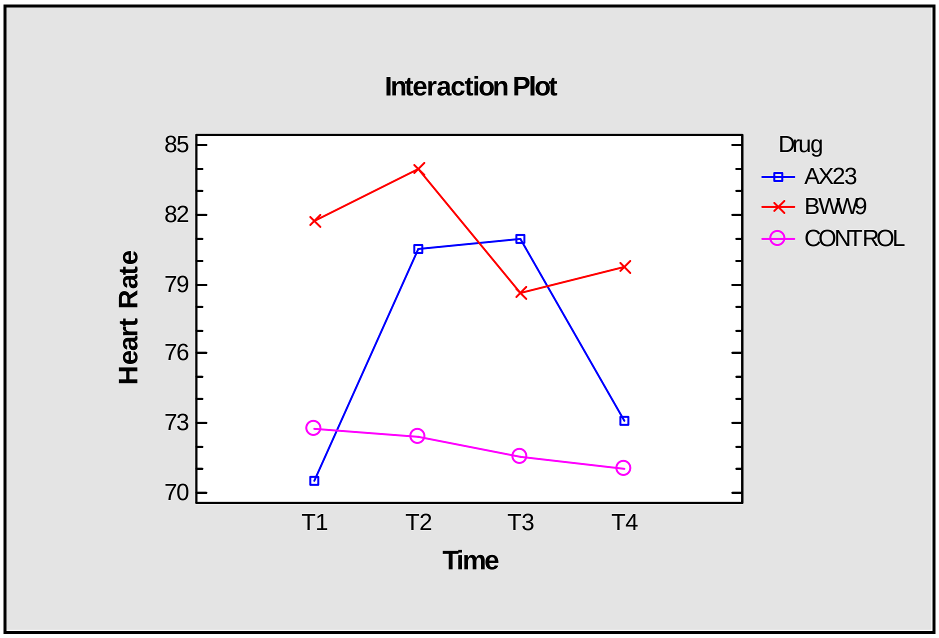


Figure 8: Interaction Plot for Drug by Time

In the above plot, the 12 points represent the average heart rate of the subjects at each combination of *Time* and *Drug*. Notice that drug BWW9 appears to act more quickly and last longer than drug AX23.

You can also add uncertainty intervals around each mean using *Pane Options* in order to compare the means heart rate of a selected drug at any two points in time. The following plot shows LSD intervals around each of the 12 means:

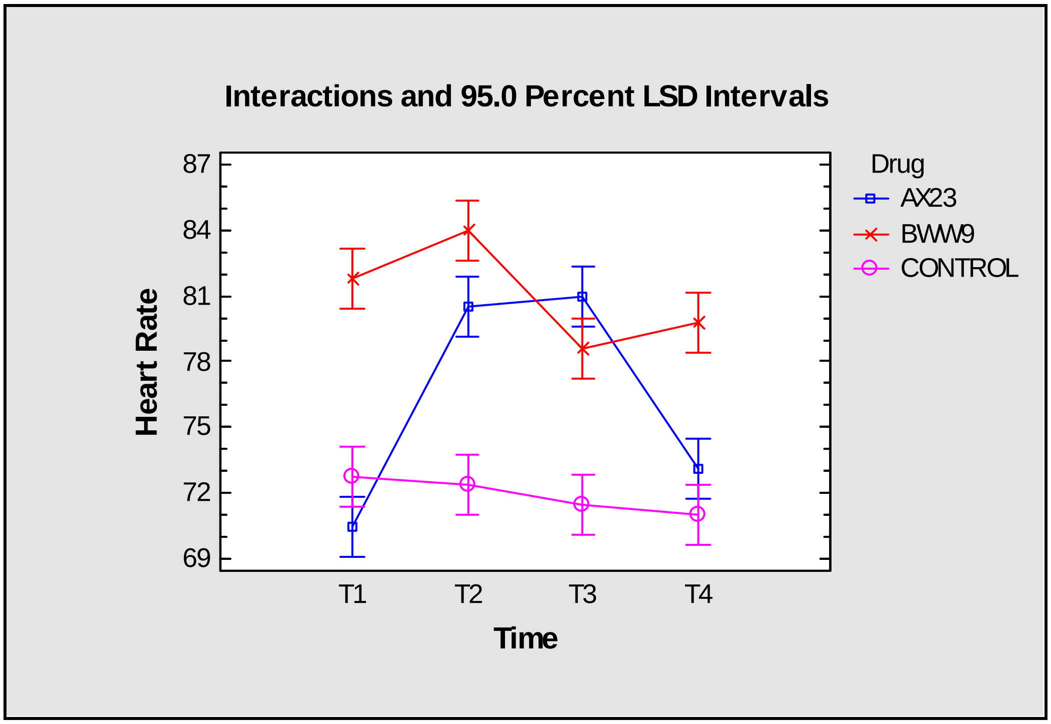


Figure 9: Interaction Plot with LSD Intervals

Non-overlapping intervals indicate statistically significant differences between two time periods for the same drug. For example, drug AX23 shows significant differences between heart rates at the following pairs of times: T1 and T2, T1 and T3, T4 and T2, T4 and T3. Note: these intervals should not be used to compare the means of different drugs, since the intervals are scaled to include only the residual variability.

Example #2

The second example comes from Environmental Statistics: Methods and Applications by Vic barnett (Wiley, 2004). It is an example of an experiment performed to compare the effect of different doses of a drug on the plasma fluoride concentration in litters of baby rats. For each litter, the concentration was measured at 15 minutes after the drug was injected, after 30 minutes, and after 60 minutes. The measured concentrations are shown below:

Litter	Age in days	Dose in μg	Concentration after 15 mins.	Concentration after 30 mins.	Concentration after 60 mins.
1	6	0.50	4.1	3.9	3.3
2	6	0.50	5.1	4.0	3.2
3	6	0.50	5.8	5.8	4.4
4	6	0.25	4.8	3.4	2.3
5	6	0.25	3.9	3.5	2.6
6	6	0.25	5.2	4.8	3.7
7	6	0.10	3.3	2.2	1.6
8	6	0.10	3.4	2.9	1.8
9	6	0.10	3.7	3.8	2.2
10	11	0.50	5.1	3.5	1.9
11	11	0.50	5.6	4.6	3.4
12	11	0.50	5.9	5.0	3.2
13	11	0.25	3.9	2.3	1.6
14	11	0.25	6.5	4.0	2.6
15	11	0.25	5.2	4.6	2.7
16	11	0.10	2.8	2.0	1.8
17	11	0.10	4.3	3.3	1.9
18	11	0.10	3.8	3.6	2.6

Figure 10: Plasma Fluoride Experiment

The three important factors in the above experiment are *Age*, *Dose*, and *Time*. It is important to note that data is taken from each litter at all three values of *Time*, but at only one *Age* and one *Dose*. Comparisons of different ages and different doses is thus such to a larger experimental error that comparisons of different times. This type of repeated measures experiment is an example of a *split-plot* design, in which each litter is a “whole-plot” and times within litters are “sub-plots”.

Step 1: Construct the Design

The above experiment consists of 9 replicates of each of the 2 x 3 x 3 = 18 combinations of the three factors. It is most easily constructed by creating a spreadsheet with the following structure:

Age	Dose	Litter	Time	Concentration
6	0.5	1	15	4.1
6	0.5	1	30	3.9
6	0.5	1	60	3.3
6	0.5	2	15	5.1
6	0.5	2	30	4.0

6	0.5	2	60	3.2
6	0.5	3	15	5.8
6	0.5	3	30	5.8
6	0.5	3	60	4.4
6	0.25	4	15	4.8
6	0.25	4	30	3.4
6	0.25	4	60	2.3
6	0.25	5	15	3.9
6	0.25	5	30	3.5
6	0.25	5	60	2.6
6	0.25	6	15	5.2
6	0.25	6	30	4.8
6	0.25	6	60	3.7
6	0.1	7	15	3.3
6	0.1	7	30	2.2
6	0.1	7	60	1.6
6	0.1	8	15	3.4
6	0.1	8	30	2.9
6	0.1	8	60	1.8
6	0.1	9	15	3.7
6	0.1	9	30	3.8
6	0.1	9	60	2.2
11	0.5	10	15	5.1
11	0.5	10	30	3.5
11	0.5	10	60	1.9
11	0.5	11	15	5.6
11	0.5	11	30	4.6
11	0.5	11	60	3.4
11	0.5	12	15	5.9
11	0.5	12	30	5.0
11	0.5	12	60	3.2
11	0.25	13	15	3.9
11	0.25	13	30	2.3
11	0.25	13	60	1.6
11	0.25	14	15	6.5
11	0.25	14	30	4.0
11	0.25	14	60	2.6
11	0.25	15	15	5.2
11	0.25	15	30	4.6
11	0.25	15	60	2.7
11	0.1	16	15	2.8
11	0.1	16	30	2.0
11	0.1	16	60	1.8
11	0.1	17	15	4.3
11	0.1	17	30	3.3
11	0.1	17	60	1.9
11	0.1	18	15	3.8
11	0.1	18	30	3.6
11	0.1	18	60	2.6

As with all data to be analyzed in STATGRAPHICS, there is one column for each experimental factor and one for the response. The data is stored in the file *howto12b.sf6*.

Step 2: Analyze the Results

To analyze this data, we will again use the *General Linear Models* procedure. This is accessed from the main STATGRAPHICS Centurion menu by selecting:

- If using the Classic menu: *Compare – Analysis of Variance – General Linear Models.*
- If using the Six Sigma menu: *Improve – Analysis of Variance – General Linear Models.*

The data input dialog box is shown below:

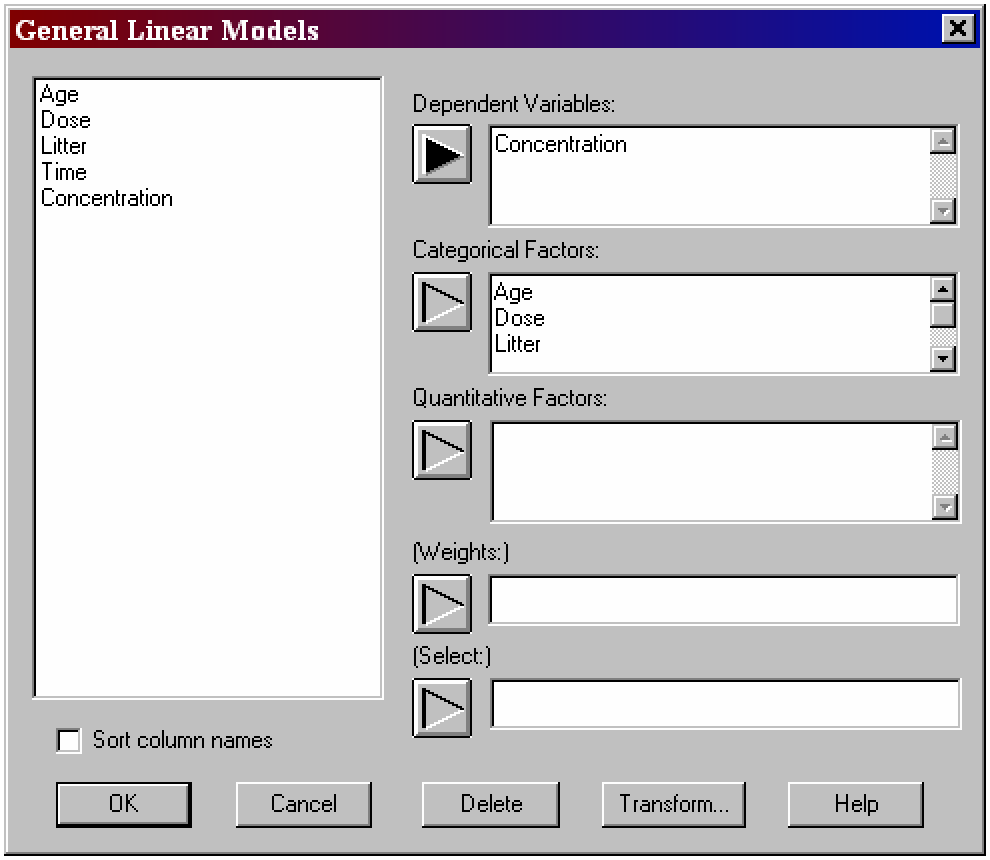


Figure 11: GLM Data Input Dialog Box

Note: the *Categorical Factors* field contains *Age*, *Dose*, *Litter* and *Time*, although the latter is not visible.

After completing the first dialog box, a second dialog box is displayed on which to specify the statistical model. It should be completed as shown below:

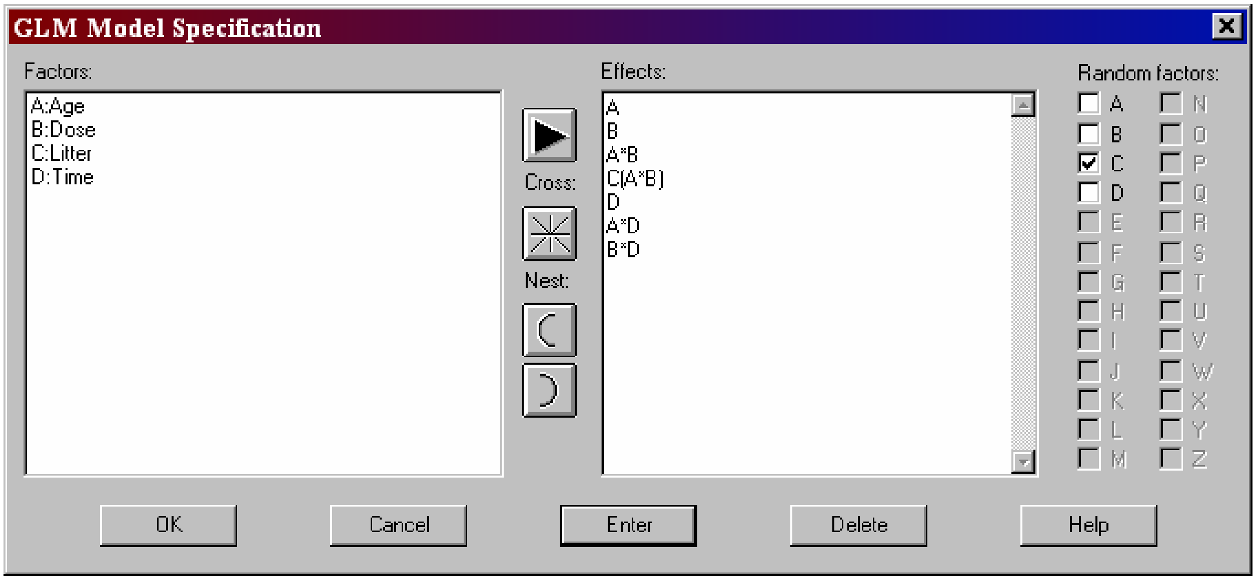


Figure 12: GLM Model Specification Dialog Box

The factors in the whole plot experiment, *Age* and *Dose*, are listed first, together with their interaction. The column representing their experimental unit, *Litter*, comes next. Note that factor C is entered as a nested factor *C(A*B)*, which indicates that each litter corresponds to only one combination of *Age* and *Dose*. *Litter* is also identified as a random factor, since many litters could have been chosen. The factor varied within litters, *Time*, is then listed, together with its interactions.

The top section of the *Analysis Summary* is shown below:

General Linear Models

Number of dependent variables: 1
Number of categorical factors: 4
A=Age
B=Dose
C=Litter
D=Time
Number of quantitative factors: 0

Analysis of Variance for Concentration

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	76.8563	25	3.07425	18.05	0.0000
Residual	4.76963	28	0.170344		
Total (Corr.)	81.6259	53			

Type III Sums of Squares

Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Age	0.0185185	1	0.0185185	0.01	0.9141
Dose	20.3304	2	10.1652	6.66	0.0113
Age*Dose	0.205926	2	0.102963	0.07	0.9351
Litter(Age*Dose)	18.3178	12	1.52648	8.96	0.0000
Time	35.4548	2	17.7274	104.07	0.0000
Age*Time	1.53481	2	0.767407	4.51	0.0201
Dose*Time	0.994074	4	0.248519	1.46	0.2412
Residual	4.76963	28	0.170344		
Total (corrected)	81.6259	53			

Figure 13: GLM Analysis Summary – Top Section

The output indicates a significant main effect for *Dose* and a significant interaction between *Age* and *Time*.

The second half of the *Analysis Summary* contains additional information about the analysis:

Expected Mean Squares

Source	*EMS*
Age	(8)+3.0(4)+Q1
Dose	(8)+3.0(4)+Q2
Age*Dose	(8)+3.0(4)+Q3
Litter(Age*Dose)	(8)+3.0(4)
Time	(8)+Q4
Age*Time	(8)+Q5
Dose*Time	(8)+Q6
Residual	(8)

F-Test Denominators

Source	*Df*	*Mean Square*	*Denominator*
Age	12.00	1.52648	(4)
Dose	12.00	1.52648	(4)
Age*Dose	12.00	1.52648	(4)
Litter(Age*Dose)	28.00	0.170344	(8)
Time	28.00	0.170344	(8)
Age*Time	28.00	0.170344	(8)
Dose*Time	28.00	0.170344	(8)

Variance Components

Source	*Estimate*
Litter(Age*Dose)	0.452046
Residual	0.170344

Figure 14: GLM Analysis Summary – Bottom Section

There are two variance components: *Litter(Age*Dose)*, which represents variability between litters, and *Residual*, which represents variability within litters. When F tests are performed, the whole plot factors and their interaction are compared to the between-litter component. The subplot factors and its interactions are compared to the within-litter component.

Step 3: Display the Results

Since *Dose* does not interact with any other factors, its effect can be displayed using a *Means Plot*:

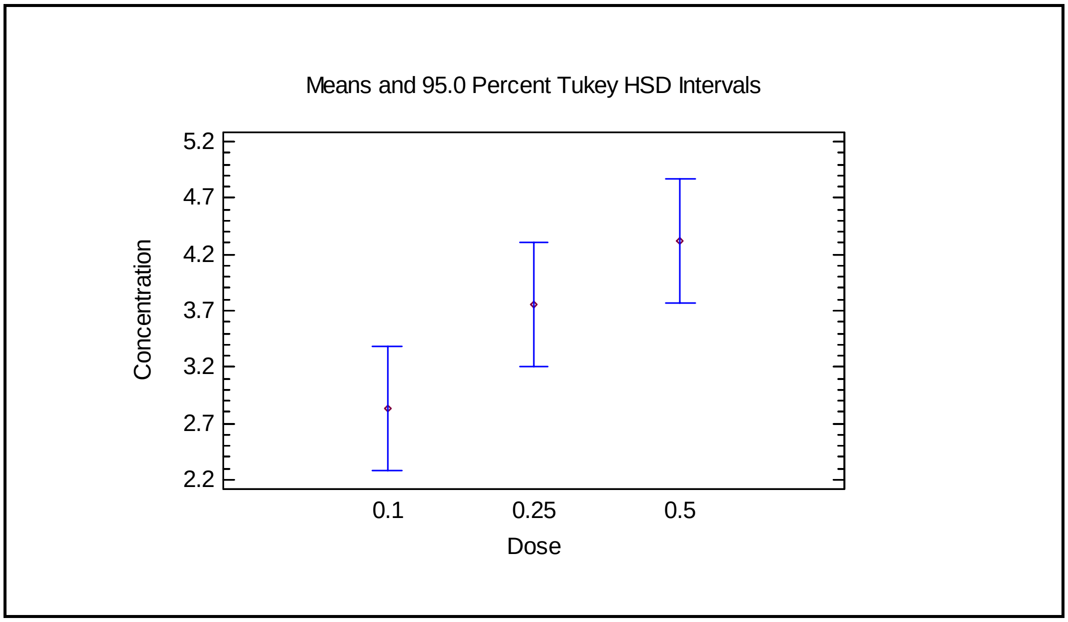


Figure 15: Means Plot for Dose

Note that concentration increases with increasing dose.

Since *Age* and *Time* interact, they must be viewed using an *Interaction Plot*:

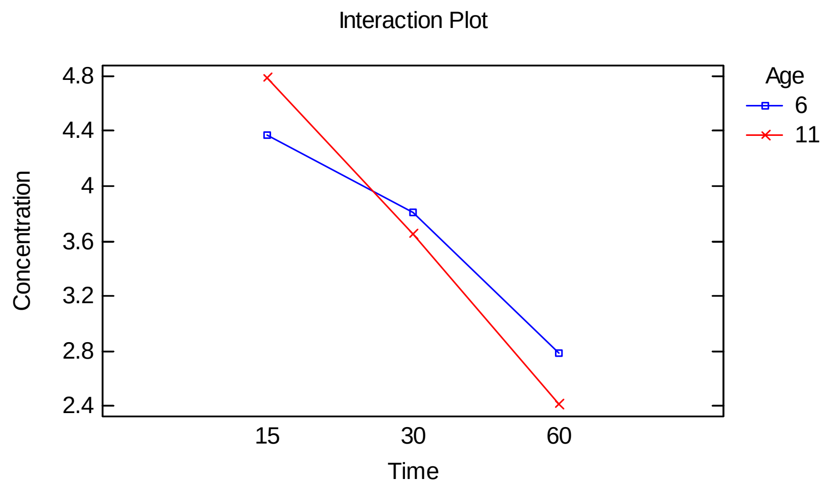


Figure 16: Interaction Plot for Age and Time

For the 11-day old litter, concentration starts at a higher level than for the 6-day old litter but does not decrease as rapidly.

Conclusion

The *General Linear Models* procedure will analyze the results of repeated measures experiments, in which measurements are taken from subjects at different points in time. As illustrated by the examples in this guide, it is important to distinguish between those factors that are varied between subjects and those that are varied within subjects.