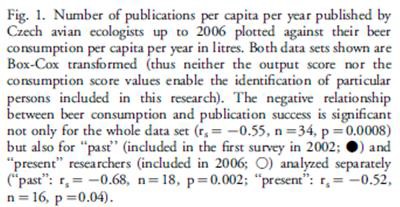
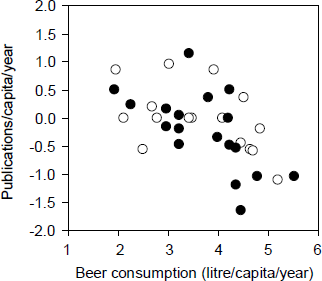


**Biostatistics Manual**

**(BSC 490 & 420.27)**





**FROM: Grim, T. 2008. A possible role of social activity to explain differences in publication output among ecologists.**

***Oikos* 117:484-487**

**Steven A. Juliano August 2018**

**TABLE OF CONTENTS**

|  |  |  |
| --- | --- | --- |
| Part 1. | Using *SAS* for Windows | 2 |
|  | Using *SAS* for Windows | 3 |
| Part 2. | Homework Assignments | 8 |
|  | Assignment 1: Summary Statistics | 9 |
|  | Assignment 2: Probability & Generating Random Numbers | 13 |
|  | Assignment 3: One and two sample tests | 14 |
|  | Assignment 4: One way Fixed Effects Analysis of Variance & Contrasts | 16 |
|  | Assignment 5: One way Random Effects Analysis of Variance | 17 |
|  | Assignment 6: Two way Factorial Analysis of Variance | 19 |
|  | Assignment 7: Mixed Model Analysis of Variance: Comparing procedures | 20 |
|  | Assignment 8: Two stage Nested Analysis of Variance | 22 |
|  | Assignment 9: Linear & Multiple Regression | 23 |

Assignment 10: Analysis of Covariance 24

[Part 3. Examples 25](#_TOC_250009)

[ANALYSIS OF VARIANCE: Simulations of the ANOVA model 26](#_TOC_250008)

[ANOVA, MULTIPLE COMPARISONS, & CONTRASTS 28](#_TOC_250007)

[UNBALANCED ANOVA EXAMPLE 34](#_TOC_250006)

[LEAST SQUARES MEANS 41](#_TOC_250005)

[ANOVA MODELS WITH RANDOM EFFECTS 47](#_TOC_250004)

[NESTED ANOVA 53](#_TOC_250003)

[SIMPLE LINEAR REGRESSION 56](#_TOC_250002)

[LINEAR & MULTIPLE REGRESSION 65](#_TOC_250001)

[ANALYSIS OF COVARIANCE (ANCOVA) 75](#_TOC_250000)

**PART 1. Using *SAS* FOR WINDOWS**

### INTRODUCTION:

*SAS* (Statistical Analysis System) is a statistical package designed for data base management and statistical analysis. Biology uses *SAS 9.4 for Windows,* which is available on computers in SLB 121 and many of the laboratories in the department. It can be set up to run on any IBM- compatible computer in the department (see me if you need to get it set up in your laboratory).

Nearly all internal statements, commands, and procedures are identical to those used in every other version of *SAS*. The main differences are in the interface, that is, the mechanisms for submitting, calling files, and printing output, and in the on-screen viewing of input and output. The interface for *SAS* works much like other programs that run in Windows (e.g., Word, Excel). You will find that the Windows version of *SAS* has a user-friendly interface and program editor. Possibilities for using *SAS* and interfacing *SAS* programming with other Windows programs (e.g., Excel, Word, PowerPoint) are numerous and complex, and we will only deal with the simpler aspects of such programming.

This handout is designed to introduce you to the use of *SAS* in the Windows environment. This handout will NOT teach you to use fully the *SAS* graphic interface and command language. *SAS* consists of hundreds of separate procedures, each with 10-200 subcommands and options, plus the DATA step, with hundreds of subcommands, functions, and many variable formats. There are numerous *SAS* windows for specialized jobs. It therefore takes much more than a single handout to introduce *SAS*. The handout gives the *minimum* information necessary for using *SAS* on the departmental computers in the computer room (SLB 121). Once you know how to start *SAS*, use the basic windows, submit, save, and print input and output, and how to get on-line help, you can quickly teach yourself *SAS* windows, language, and procedures by using on-line help. More information on manuals is given below - sections **E.**

**STARTING *SAS***:

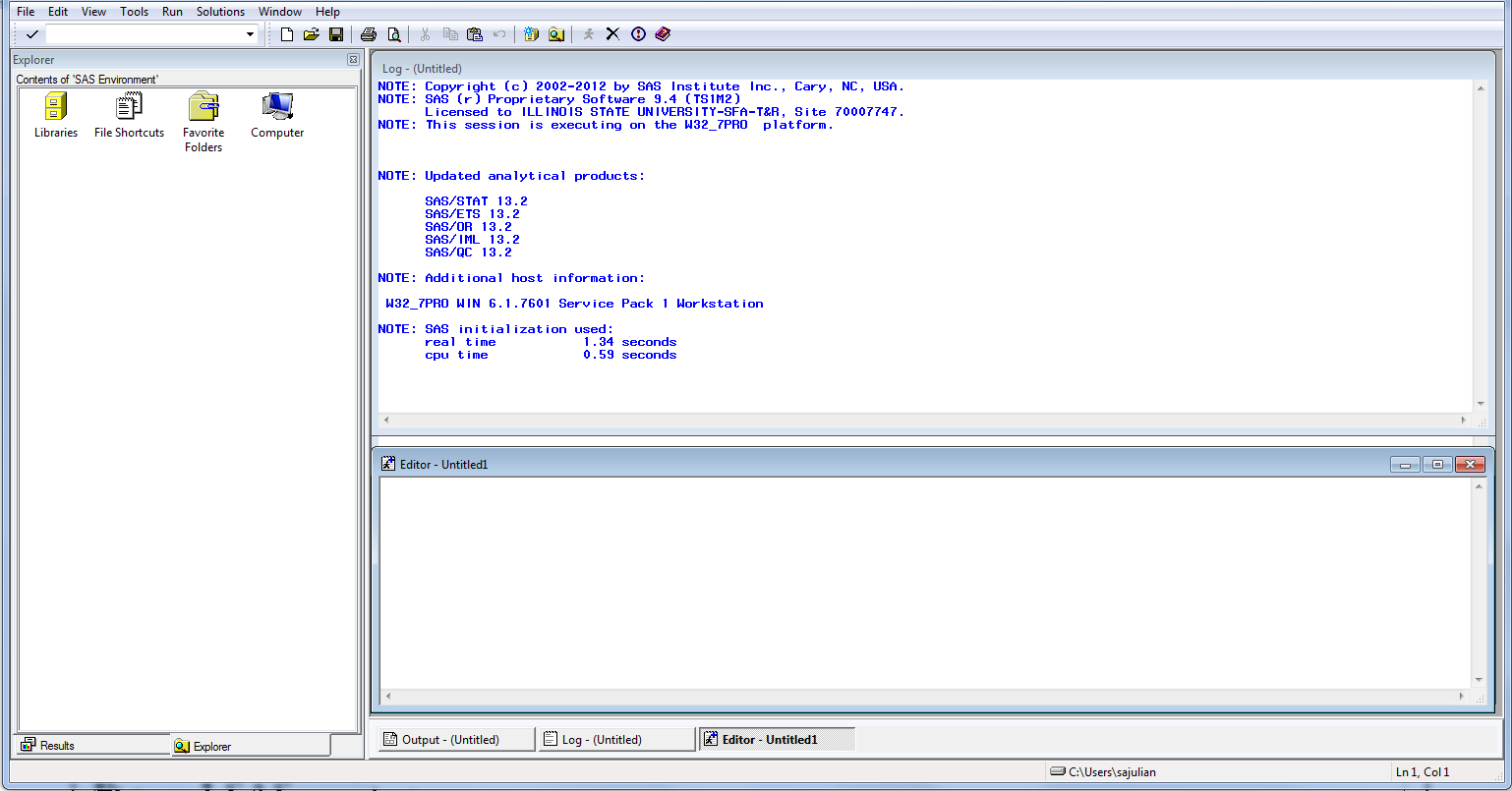
To start *SAS* on computers in the Biological Sciences computer room (SLB 121):

1. The computers should be on, with Windows running.
2. In Windows, locate the *SAS* icon on the **Desktop or Start menu**
3. Double click

**USING *SAS***

1. **The screen**. The *SAS* screen (Fig. 1, next page) is controlled by the DISPLAY MANAGER. There are three windows on the starting screen: the OUTPUT window, in which output of procedures appears; the LOG window, in which statements submitted for processing, along with error messages appear; and the ENHANCED PROGRAM EDITOR or EPE window, in which you type input statements. Other windows appear as commands are given. *SAS* for Windows operates much like other Windows programs (e.g., Word, Excel). To move between windows, simply left click on the destination window. You can also enter window names in the command slot (upper left, Fig. 1) and hit **ENTER** (throughout this handout, keys to be hit will be shown in **bold Arial type**, *SAS* screen buttons or menu items to be clicked will be shown in **bold Times Roman type**). You can tell which window is active by the location of the cursor (which can be

hard to see) and by the color of the window banner (top). Any window may be expanded to fill the entire screen by clicking on the icons in the upper right hand corner of each window (Fig. 1) You can shrink a window to an icon by clicking on the little icon with a ‘—’. You can also go to different windows via the **VIEW** menu.



**Figure 1.** *SAS* opening screen. Red arrows = HELP

1. **Ways of using *SAS*.** We will cover only one way of using *SAS* for data entry and analysis: *SAS* commands and programming, which is not automated and driven mostly by writing code (as opposed to point-and-click menus). That makes it almost infinitely flexible, but also with a steep learning curve. There are alternative ways of using *SAS*, however our license does not include them.

***SAS* commands**. *SAS* is a command driven program. *SAS* statements are typed in a PROGRAM EDITOR window. They are submitted by hitting the button next to the command slot (the one with the little running man on it…. get it?). You type in the PROGRAM EDITOR, which is a full screen editor (much like Word for Windows). The cursor is moved around the PROGRAM EDITOR window using the mouse, or by using the cursor control keys (Arrows). The PROGRAM EDITOR window allows full-screen editing. Simply type what you want on each line at any point on the screen

**Punctuation.** *SAS* has some very specific, very important rules of punctuation.

|  |  |
| --- | --- |
| All *SAS* statements must end with a semicolon [ ; ]. | **Errors with semicolons will be the most common mistakes you will make** |
| All data lines must NOT end with a semicolon [ ; ]. |
| An asterisk [ \* ] at the beginning of a line renders the  line a comment, which has no effect on the input. | Useful for removing a line from a program  without deleting it. |
| Surrounding text like this: /\* <text here> \*/ Also renders it a comment that is visible but has no  effect on the program. | Useful for reminding yourself what a particular line does. |

Type the sample job in the box below in ENHANCED PROGRAM EDITOR (EPE) (**Fig. 2**). You may find typing easiest if you expand the EPE window to fill the screen.

The 1st three lines form the DATA STEP. They create the data set named SAMPLE. The next nine lines contain data to be read (paired observations of variables X and Y). The semicolon marks the end of the data. The procedure UNIVARIATE computes a large number of summary

statistics (e.g., mean, SD, SE, median, variance, etc.) for each variable. The RUN statements indicates to end of a set of submitted statements.

data Sample;

input X Y; datalines; 6 1

7 2

8 3

9 4

10 5

11 6

12 7

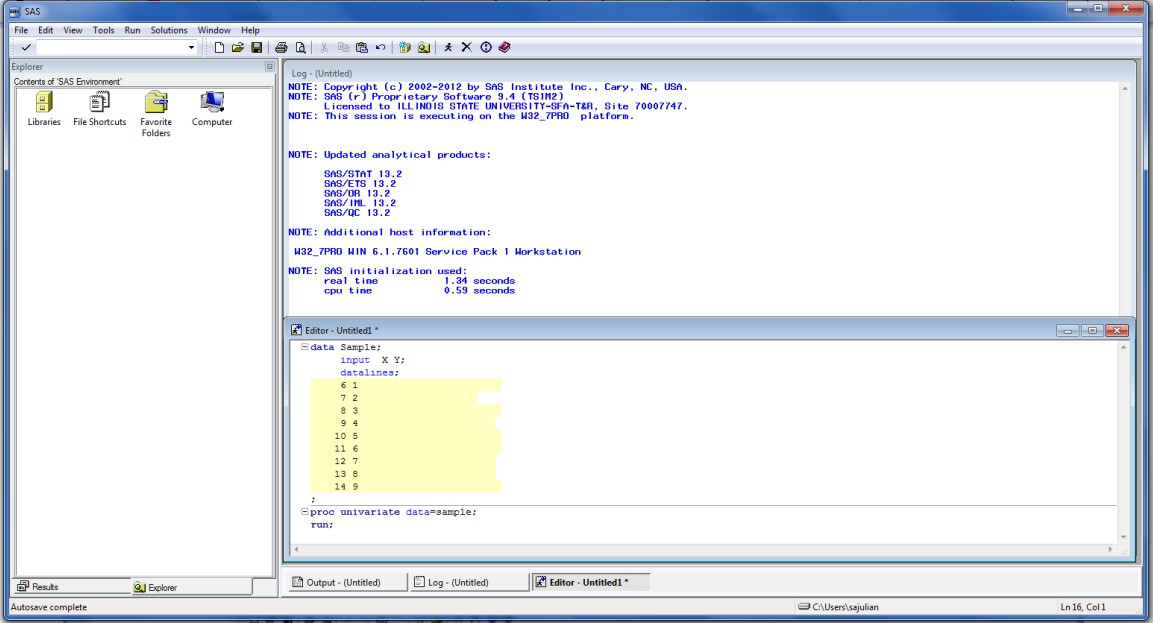
13 8

14 9

;

proc univariate data=sample; run;

Note that in the EPE, data are highlighted in pale yellow, *SAS* statements (procedures, data step) are highlighted in **bold blue** text, and *SAS* keywords are in blue text. A useful feature of EPE is that errors appear in ***red.***

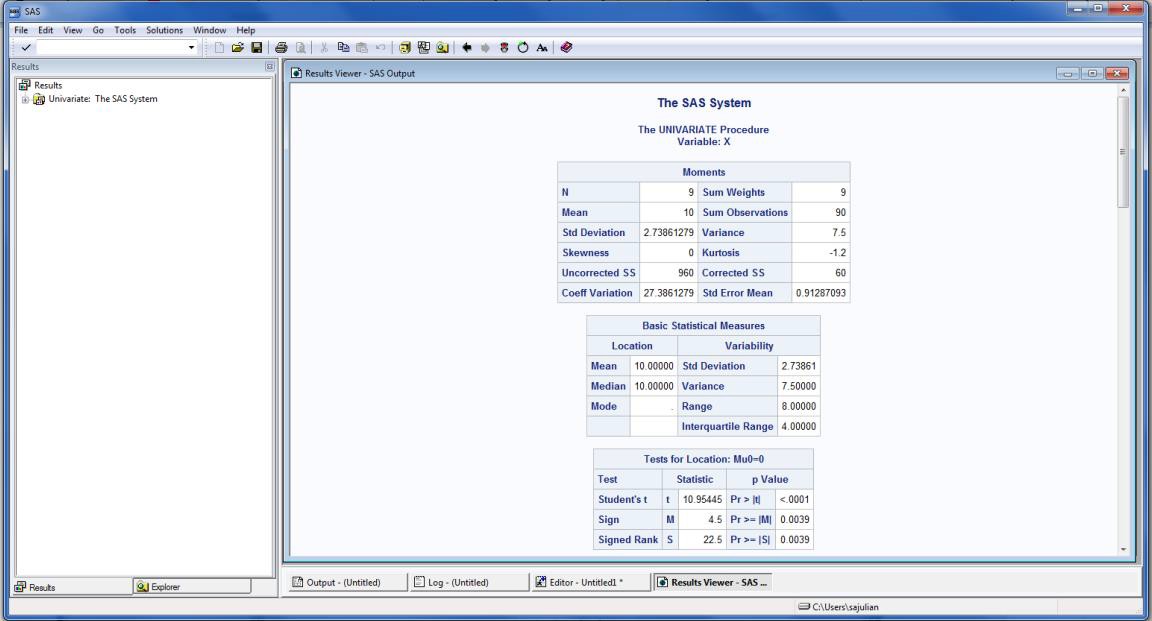
After typing, return the EPE window to its original size, and hit the **SUBMIT** button (that is the button with the little running man … cute, huh?). Output will appear in the RESULTS VIEWER aswhat *SAS* calls the “Output Delivery System”, which has relatively nice graphics and is organized to look good on screen (Fig. 3)and

**Figure 2.** Commands and data in the Enhanced Program Editor.

input statements along with any error messages will appear in the LOG window. When processing is done, the cursor returns to the EPE window. The output may now be viewed by moving to the RESULTS VIEWER window (Fig. 3), and using **PAGEUP** and **PAGEDOWN** keys or the window scroll bar (at right) to scroll through the output. The LOG may be viewed in a similar fashion.

If you find errors (you will) they will be noted in the LOG window. The EPE window allows for easy correction. The most recently submitted commands remain in the EPE until you erase them. You may now make corrections by typing over lines or by adding, moving, or deleting

lines, and resubmitting. Full screen editing, cutting, pasting, and copying works just like other Windows programs (e.g., Word). You can also cut and paste from any text file.



**Figure 3.** Ouput in the RESULTS VIEWER

window.

**Clearing, printing, and saving window contents.** To get rid of the

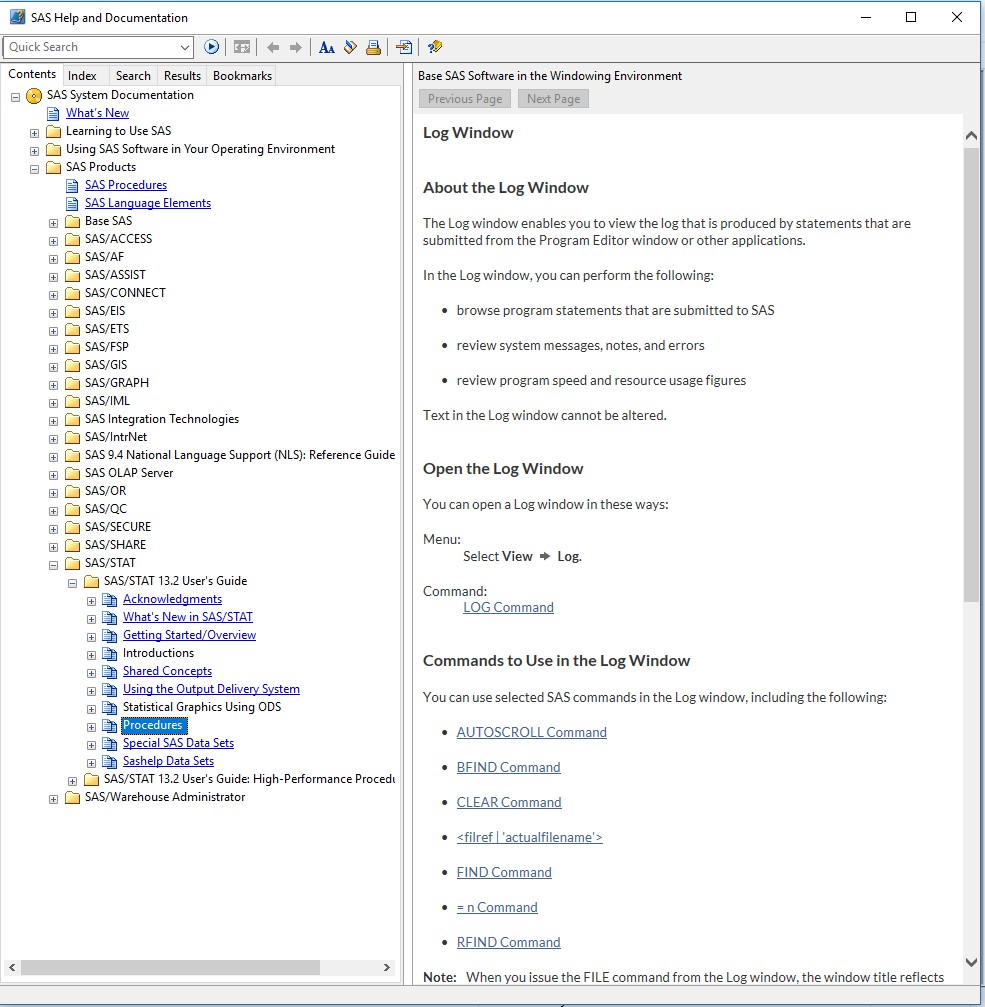
contents of the EDITOR and LOG windows, choose the **EDIT** menu (Fig. 1), and click **CLEAR ALL**. You can save the contents of the EDITOR, LOG, and most other windows. Simply choose the **FILE** menu (Fig. 1), and click **SAVE** (or **SAVE AS** if you wish to change a file’s name). The save dialog looks a lot like those of other Windows programs. You can choose the disk, the directory, the file name, and the extension. **An important exception to this is the RESULTS VIEWER window.** Because it contains graphics, it is in HTML language and there are limited options for saving it – basically web page format or plain text. There are two good ways to save the output from the RESULTS VIEWER so that it is useful to you: 1. Select all, then cut and paste into a Word document. You can then change the font, margins, delete parts you don’t want, etc., to make it fit and look good. 2. Choose print, and print as a PDF, which usually looks good but is not as easy to edit.

The RESULTS VIEWER has another quirk. You can delete things the RESULTS window on the left (Fig. 3) but for reasons that are unclear the cursor will be “stuck” in the RESULTS window. Happily, the solution is simple: clicking again on the RESULTS window “unsticks” the cursor.

Contents of most windows can be cut and pasted into another Windows document (e.g., into a Word for Windows document, which facilitates writing reports). In addition, saved files can be opened in other Windows programs (e.g., Word for Windows). The EPE window gets saved by default as a *SAS* file in ASCII text (no formatting characters except carriage return and page eject). If you intend to save output for later use, I suggest youcut and paste it to a Word file. To get lines that have been stored on disk, choose the **FILE** menu, and click **OPEN**. Again, you will get a dialog box that operates much like every other Windows program.

To print the LOG, RESULTS VIEWER, PROGRAM EDITOR, or any other windows, choose the **FILE** menu (Fig. 1) and click **PRINT**. The active window will then be printed, via the

typical Windows dialog box. You should, as much as possible, work with input and output electronically (saves paper).

1. **Getting help within *SAS***. *SAS* provides extensive html format help with commands, procedures, and windows. These are available by choosing the **HELP** menu or the little book icon on the toolbar (**Fig. 1,** Red arrows), and click on ***SAS* PRODUCTS > *SAS/STAT* >*SAS/STAT* USERS GUIDE 12.2>PROCEDURES (Fig. 4).**
2. **Reference manuals**. *SAS* produces many useful references. ALL manuals for *SAS* version

9.4 or later for Windows can be found in the *SAS* HELP menu or on line.

#### *S. A. Juliano*

**Fig. 4.** Screen shot of *SAS* System Documentation.

(August 2018)

**PART 2. Homework assignments**

**BIOSTATISTICS 490 ASSIGNMENT 1**

**Summary Statistics**

This assignment will introduce you to *SAS*, give you first-hand experience with entering, manipulating data, and help you to learn *SAS*'s capabilities for summarizing data.

Input the data on the attached sheet into *SAS*. These data represent geographic ranges within England and Scotland for water beetles species (Dytiscidae), by subfamily. 'Pre 1960' is the number of 10X10 km squares in which the species was recorded before 1960 but not after 1960, and 'Post 1960' is the number of 10X10 km squares in which the species has been recorded since 1960, including squares with records both before and after 1960. 'Minimum' is the minimum recorded body length (mm) and 'Maximum' is the maximum recorded body length (mm). Once the data are in:

1. Compute the total geographic range (number of squares) by adding the number of old and recent squares.
2. Compute the average body size of the species, based on the maximum and minimum recorded.
3. Transform total geographic range and average body size to logs (base 10).
4. Obtain histograms of total geographic range and average body size and their transformations. also obtain normal probability plots.
5. Obtain summary statistics for total geographic range and average body size and their transformations.
6. Plot total geographic range vs. average body size on natural and logarithmic scales. Label the points to identify the subfamily (note: from the **LOCAL** menu for the graph, choose **VARIABLES**, then **GROUPING VARIABLE**). Then obtain correlation coefficients for the relationship between the two variables, and between the two transformed variables.
7. Obtain summary statistics and histograms for log(average body size) and log(total geographic range) by subfamily (there are 5 subfamilies) (note: subset the data set).

Once you have done this, you should answer the following questions:

* 1. Examine the histograms of the data on the natural and logarithmic scales. Which scale comes closest to producing symmetrical distributions?
  2. Are geographic range and average body size related? How? Does transformation alter the relationship?
  3. Do the subfamilies differ in any substantial way in average body size or in geographic range (both log transformed)?

### THE REPORT

In your report, you should include the relevant parts of your *SAS* output, appropriately labeled so that I can tell what you are presenting. The output should be integrated into the report. Answers to questions should be physically near the relevant bits of computer output to make it easy for the reader to understand what you are talking about. If you submit raw output, I won't grade it. I will **NOT** accept homework late.

**DATA**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subfamily** | **Pre 1960** | **Post 1960** | **Minimum** | **Maximum** |
| HYDROPOR | 109 | 59 | 3.0 | 3.0 |
| HYDROPOR | 7 | 8 | 2.3 | 2.7 |
| HYDROPOR | 22 | 10 | 2.5 | 2.5 |
| HYDROPOR | 61 | 40 | 2.0 | 2.3 |
| HYDROPOR | 129 | 126 | 2.0 | 2.4 |
| HYDROPOR | 102 | 85 | 3.0 | 3.0 |
| HYDROPOR | 43 | 12 | 4.5 | 4.5 |
| HYDROPOR | 139 | 130 | 4.0 | 4.3 |
| HYDROPOR | 28 | 35 | 4.5 | 4.5 |
| HYDROPOR | 60 | 65 | 4.0 | 4.0 |
| HYDROPOR | 153 | 265 | 3.0 | 3.0 |
| HYDROPOR | 22 | 23 | 4.0 | 4.0 |
| HYDROPOR | 10 | 19 | 5.0 | 5.0 |
| COLYMBET | 149 | 208 | 7.5 | 8.5 |
| COLYMBET | 64 | 95 | 7.0 | 8.0 |
| COLYMBET | 73 | 74 | 7.0 | 7.5 |
| COLYMBET | 61 | 44 | 9.0 | 9.0 |
| COLYMBET | 6 | 4 | 9.0 | 9.5 |
| COLYMBET | 47 | 99 | 7.0 | 7.0 |
| COLYMBET | 86 | 81 | 7.0 | 8.0 |
| COLYMBET | 43 | 31 | 6.0 | 6.0 |
| COLYMBET | 3 | 18 | 8.0 | 8.5 |
| COLYMBET | 18 | 23 | 7.0 | 7.0 |
| COLYMBET | 13 | 7 | 7.0 | 8.0 |
| COLYMBET | 66 | 53 | 11.5 | 11.5 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subfamily** | **Pre 1960** | **Post 1960** | **Minimum** | **Maximum** |
| COLYMBET | 19 | 61 | 9.5 | 10.0 |
| COLYMBET | 66 | 82 | 10.5 | 11.5 |
| COLYMBET | 16 | 45 | 11.0 | 11.0 |
| COLYMBET | . | . | 13.0 | 15.0 |
| COLYMBET | . | . | 10.0 | 11.0 |
| LACCOPHI | 75 | 75 | 4.0 | 4.5 |
| LACCOPHI | 132 | 158 | 4.0 | 4.5 |
| LACCOPHI | 6 | 4 | 3.0 | 3.0 |
| HYDROPOR | 19 | 5 | 2.0 | 2.5 |
| HYDROPOR | 107 | 154 | 4.5 | 5.0 |
| HYDROPOR | 44 | 19 | 2.0 | 2.0 |
| HYDROPOR | 10 | 2 | 1.5 | 1.5 |
| HYDROPOR | 13 | 1 | 2.0 | 2.0 |
| HYDROPOR | 10 | 20 | 2.0 | 2.0 |
| HYDROPOR | 165 | 253 | 3.0 | 3.0 |
| HYDROPOR | 29 | 18 | 4.0 | 4.0 |
| HYDROPOR | 67 | 58 | 4.0 | 4.0 |
| HYDROPOR | 73 | 31 | 3.0 | 3.5 |
| HYDROPOR | 70 | 96 | 5.0 | 5.0 |
| HYDROPOR | 49 | 22 | 4.0 | 4.0 |
| HYDROPOR | . | . | 3.2 | 3.5 |
| HYDROPOR. | . | 204 | 4.5 | 5.2 |
| COLYMBET | . | . | 8.0 | 8.5 |
| COLYMBET | . | . | 7.5 | 8.0 |
| COLYMBET | . | . | 6.0 | 6.5 |
| COLYMBET | . | . | 7.0 | 7.0 |
| COLYMBET | . | . | 16.0 | 18.0 |
| HYDROPOR | . | . | 3.4 | 11.0 |
| HYDROPOR | . | . | 3.8 | 5.2 |
| HYDROPOR | 76 | 244 | 2.8 | 3.3 |
| HYDROPOR | 86 | 185 | 3.0 | 3.5 |
| HYDROPOR | 0 | 7 | 3.0 | 3.8 |
| HYDROPOR | 188 | 420 | 3.7 | 4.5 |
| HYDROPOR | 33 | 47 | 3.5 | 4.2 |
| HYDROPOR | 1 | 9 | 3.0 | 3.4 |
| HYDROPOR | 159 | 345 | 3.5 | 4.1 |
| HYDROPOR | 67 | 277 | 3.3 | 3.8 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Subfamily** | **Pre 1960** | **Post 1960** | **Minimum** | **Maximum** |
| HYDROPOR | 31 | 19 | 3.5 | 4.5 |
| HYDROPOR | 61 | 147 | 3.0 | 3.5 |
| HYDROPOR | 1 | 3 | 3.0 | 3.5 |
| HYDROPOR | 62 | 113 | 3.3 | 3.5 |
| HYDROPOR | 11 | 33 | 2.2 | 2.6 |
| HYDROPOR | 135 | 384 | 2.8 | 3.5 |
| HYDROPOR | 119 | 233 | 2.5 | 3.1 |
| HYDROPOR | 41 | 25 | 3.3 | 4.2 |
| HYDROPOR | 228 | 804 | 3.3 | 3.8 |
| HYDROPOR | 145 | 494 | 4.0 | 4.6 |
| HYDROPOR | 23 | 25 | 4.0 | 5.3 |
| HYDROPOR | 5 | 7 | 1.7 | 2.2 |
| HYDROPOR | 81 | 251 | 3.0 | 3.4 |
| HYDROPOR | . | 388 | 3.4 | 3.8 |
| HYDROPOR | 111 | 280 | 2.8 | 3.3 |
| HYDROPOR | 84 | 254 | 2.5 | 2.9 |
| HYDROPOR | 70 | 119 | 4.5 | 5.5 |
| COLYMBET | 3 | 1 | 9.5 | 9.5 |
| COLYMBET | 111 | 122 | 9.0 | 10.0 |
| COLYMBET | 85 | 161 | 9.0 | 10.0 |
| COLYMBET | 30 | 35 | 10.0 | 11.0 |
| COLYMBET | 29 | 60 | 10.0 | 11.0 |
| COLYMBET | 81 | 93 | 10.0 | 12.0 |
| DYTISCIN | 22 | 38 | 13.0 | 14.5 |
| DYTISCIN | 20 | 14 | 12.0 | 13.0 |
| DYTISCIN | 1 | 9 | 14.0 | 16.0 |
| DYTISCIN | 0 | 1 | 14.0 | 16.0 |
| DYTISCIN | 13 | 29 | 14.0 | 16.0 |
| DYTISCIN | 117 | 184 | 16.0 | 18.0 |
| DYTISCIN | 11 | 28 | 22.0 | 25.0 |
| DYTISCIN | 29 | 19 | 27.0 | 32.0 |
| DYTISCIN | 26 | 41 | 26.0 | 32.0 |
| DYTISCIN | 8 | 18 | 32.0 | 38.0 |
| DYTISCIN | 134 | 291 | 26.0 | 32.0 |
| NOTERIDA | 81 | 141 | 4.0 | 5.0 |
| NOTERIDA | 23 | 39 | 3.5 | 4.0 |

**BIOSTATISTICS 490 ASSIGNMENT 2**

**Probability & generating random numbers**

This assignment will continue your introduction to *SAS*, and give you firsthand experience with simulating data using random number generators. Through this modeling exercise you will learn more about probability theory

*SAS* can simulate random sampling from a population. Use the RANUNI (uniform random number function) to simulate random draws from a binomial population.

1. Consider a population of organisms consisting of 0.55 females and 0.45 males. Simulate 10, 50, 100, and 1000 random repetitions of 2 draws from such a population. This gives you a total of 4 experiments. Determine the frequencies of 0, 1, and 2 females in 2 draws for each of the 4 experiments. Determine the actual mean and SD of number of females in each of the 4 experiments.
2. Consider sampling from a normal distribution where *Y*~N(0,1). How does N, the sample size, influence estimates of the mean and SD derived from theses samples? Use the function RANNOR to simulate repeated samples of N=4, 16, 64, 100. Draw 100 repeated samples for each N. For each sample determine mean and SD. Then for each sample size N, determine the mean estimate of the distribution mean (i.e., the mean of the 100 means) and the associated SD of those 100 sample means.

Answer the following questions:

1. In a), how do the observed distributions, means, and SD's for the 4 experiments compare with the theoretical expectation based on the binomial distribution? What do your results suggest regarding samples from binomial populations and the expectations generated by theory?
2. In b), How close do the individual samples come to the true mean, and how does that depend on N? How is the **mean of the 100 sample means** related N? What pattern do you detect? Consider the **SD of the set of 100 sample means**. How is that SD related to N? How is it related to the underlying SD of the population (~N(0,1)). **Hint:** consider the formula for the Standard Error of the mean (SE) [which you might have to look up].

**BIOSTATISTICS 490 ASSIGNMENT 3**

**One & two sample tests**

This assignment will teach you about one and two sample *t-*tests in *SAS* and the nonparametric equivalents of these *t*- tests. You will also learn how to test assumptions of these tests.

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

1. **One sample tests.** An investigator wants to test whether herbivory induces increases in the concentration of condensed tannins (plant defense chemicals) in the leaves of birch. She measures tannin concentration of leaves on 30 trees before and after herbivore attack. The results (in ppm tannin) are:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| TREE | BEFORE | AFTER | TREE | BEFORE | AFTER | TREE | BEFORE | AFTER |
| 1 | 249.226 | 268.637 | 11 | 250.174 | 261.866 | 21 | 245.350 | 250.517 |
| 2 | 250.111 | 259.269 | 12 | 246.937 | 259.834 | 22 | 245.051 | 266.954 |
| 3 | 248.442 | 264.517 | 13 | 245.818 | 247.879 | 23 | 247.605 | 256.993 |
| 4 | 248.864 | 253.011 | 14 | 246.596 | 258.382 | 24 | 245.509 | 244.026 |
| 5 | 252.536 | 253.978 | 15 | 247.370 | 249.798 | 25 | 250.488 | 264.173 |
| 6 | 249.411 | 253.891 | 16 | 249.602 | 256.302 | 26 | 245.429 | 252.435 |
| 7 | 247.676 | 259.680 | 17 | 246.705 | 264.113 | 27 | 252.686 | 247.669 |
| 8 | 245.540 | 262.027 | 18 | 252.351 | 248.684 | 28 | 246.900 | 258.216 |
| 9 | 252.962 | 250.786 | 19 | 250.911 | 264.232 | 29 | 252.403 | 255.929 |
| 10 | 252.251 | 255.430 | 20 | 251.460 | 266.409 | 30 | 249.247 | 266.287 |

1. Do an appropriate parametric test on these data. Give summary statistics. What do you conclude?
2. Assess the assumptions of the test using a normal probability plot and a test for fit to a normal distribution. What do you conclude?
3. Perform an appropriate nonparametric test on these data. Report summary statistics. Is the conclusion the same as that given by the parametric test? Which test do you believe?
4. **Two sample tests.** A researcher is interested in the whether different diets result in differential growth in *Aedes triseriatus* (a mosquito). He tests liver powder vs. powdered dog food, both provided at a rate of 0.1 mg larva-1 day-1. The mosquitoes used were from a single population, and assuming that all started the experiment at the same mass, the investigator measures mass at pupation as an index of growth. The results (in mg) are:

|  |  |  |
| --- | --- | --- |
| DOG FOOD |  | LIVER |
| 0.758 0.818 0.922 0.924 1.020 0.742 0.778 | 0.805 0.815 | 0.858 2.475 1.246 1.193 1.223 |
| 1.201 0.978 1.174 0.830 0.915 0.739 0.686 | 0.992 | 1.148 1.039 1.216 1.800 1.346 |

* 1. Test the assumption of equal variances. What do you conclude?
  2. Perform a two sample *t*-test on these data. What conclusion do you reach? Are the 7conclusions the same for the "equal variances" and the "unequal variances" *t*- tests? Which test do you believe? Report appropriate summary statistics.
  3. Consider transformations of the data. Evaluate whether log, square root, or reciprocal transformations lead to homogeneous variance. Which is the best transformation, and why?
  4. Do the two sample *t*-test on the transformed data. Are the conclusions the same as those for the untransformed data? Report appropriate summary statistics (for the best transformation).

Using the data from problem 1, do a comparable nonparametric test. Report summary statistics.

Are the conclusions the same? Compare the results of all the tests done.

**BIOSTATISTICS 490 ASSIGNMENT 4**

**One way Fixed effects Analysis of Variance & Contrasts**

This assignment will introduce you to fixed effects analysis of variance. First you will learn how to execute and how to interpret a fixed effects analysis of variance, and how to do one kind of follow up test - linear contrasts, in an analysis of variance. In addition, you will learn how to conduct and to interpret a residual analysis to check assumptions in an analysis of variance.

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion.**

An investigator is interested in the effects of soil nutrients on growth of *Aster* in a particular prairie soil. She is specifically interested in two categories of nutrients: Macronutrients, of which Potassium and Phosphorus are thought to be important; and Micronutrients, of which Boron and Copper are thought to be important. She does an experiment with 5 treatments: CONTROL (no nutrients added); Added Boron; Added Copper; Added Potassium; and Added Phosphorus.

Below are the results (mg growth).

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **TREATMENT** | CONTROL | +Boron | +Copper | +Potassium | +Phosphorus |
|  | 45 | 58 | 52 | 54 | 56 |
|  | 51 | 62 | 49 | 52 | 64 |
|  | 46 | 61 | 47 | 53 | 57 |
|  | 53 | 59 | 55 | 60 | 59 |
|  | 48 | 63 | 55 | 50 | 62 |
|  | 43 | 55 | 45 | 53 | 59 |
|  | 51 | 66 | 50 | 53 | 61 |
|  | 47 | 57 | 53 | 60 | 56 |

1. Do a one way ANOVA on these data. Present the results as an ANOVA table, along with statements of the null and alternative hypotheses and assumptions.
2. Perform follow up analyses by doing orthogonal contrasts to determine: whether added nutrients enhance growth; whether macronutrients and micronutrients have different effects on growth; whether single micronutrients produce different growth effects; and whether single macronutrients have different growth effects.
3. Test the assumption of homogeneous variance using Levene's test. Test residuals for normality.
4. Analyze residuals graphically for normality and homogeneous variance.
5. Report treatment means graphically, including SE's.

### QUESTIONS

1. Are the assumptions of ANOVA reasonably well met by the raw data? If not, is there a transformation that improves the situation? If so, do the transformation and reanalyze. When you are satisfied, report **both** the analysis of raw data and of the **best** transformation (if any).
2. Write a paragraph or two summarizing the conclusions from the ANOVA an follow up tests.

**BIOSTATISTICS 490 ASSIGNMENT 5**

**One way Random effects Analysis of Variance**

This assignment will introduce you to random effects analysis of variance. You will learn how to execute and how to interpret a random effects analysis of variance, and how to conduct and interpret residual analyses to check assumptions in an analysis of variance.

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

The data on the next page were obtained in a survey designed to determine whether size (length in mm) of carabid beetles varies among local populations. Five population out of 42 were selected at random, and a random sample of beetles from each population were collected and measured. The ORDER variable represents the order in which the measurements were made.

1. Do a one way ANOVA on these data. Present the results as an ANOVA table, along with statements of the null and alternative hypotheses and assumptions.
2. Perform follow up analyses and estimate variance components and intraclass correlation.
3. Analyze residuals to check for normality, homogeneous variance, and autocorrelation. Also check for homogeneous variance using Levene’s test.

**QUESTIONS**

1. Are the assumptions of ANOVA reasonably well met by the raw data? If not, is there a transformation that is likely to improve the situation? If so, do the transformation and reanalyze. When you are satisfied, report **both** the analysis of raw data and of the **best** transformation (if any).
2. Does size vary among populations? Write a paragraph or two summarizing the conclusions from the ANOVA.

|  |  |  |
| --- | --- | --- |
| POPULATION | LENGTH | ORDER |
| 1 | 13 | 1 |
| 1 | 16 | 48 |
| 1 | 12 | 68 |
| 1 | 10 | 6 |
| 1 | 17 | 94 |
| 1 | 15 | 62 |
| 1 | 13 | 96 |
| 1 | 17 | 70 |
| 1 | 16 | 7 |
| 1 | 16 | 2 |
| 1 | 14 | 36 |
| 1 | 17 | 63 |
| 1 | 14 | 26 |
| 1 | 11 | 97 |
| 1 | 14 | 98 |
| 1 | 13 | 4 |
| 1 | 18 | 30 |
| 1 | 19 | 77 |
| 1 | 16 | 43 |
| 1 | 15 | 34 |
| 2 | 23 | 16 |
| 2 | 29 | 61 |
| 2 | 24 | 83 |
| 2 | 20 | 85 |
| 2 | 21 | 37 |
| 2 | 20 | 69 |
| 2 | 25 | 82 |
| 2 | 27 | 81 |
| 2 | 28 | 8 |
| 2 | 27 | 3 |
| 2 | 27 | 71 |
| 2 | 23 | 45 |
| 2 | 21 | 27 |
| 2 | 28 | 57 |
| 2 | 23 | 64 |
| 2 | 24 | 67 |
| 2 | 21 | 31 |
| 2 | 24 | 99 |
| 2 | 24 | 44 |
| 2 | 22 | 66 |
| 3 | 28 | 95 |
| 3 | 31 | 84 |
| 3 | 33 | 13 |
| 3 | 27 | 23 |
| 3 | 30 | 18 |
| 3 | 31 | 35 |
| 3 | 28 | 87 |
| 3 | 29 | 53 |
| 3 | 34 | 55 |
| 3 | 21 | 56 |

|  |  |  |
| --- | --- | --- |
| POPULATION | LENGTH | ORDER |
| 3 | 36 | 14 |
| 3 | 29 | 46 |
| 3 | 31 | 22 |
| 3 | 32 | 17 |
| 3 | 34 | 80 |
| 3 | 31 | 58 |
| 3 | 33 | 32 |
| 3 | 34 | 65 |
| 3 | 36 | 100 |
| 3 | 33 | 60 |
| 4 | 20 | 50 |
| 4 | 18 | 11 |
| 4 | 23 | 21 |
| 4 | 24 | 86 |
| 4 | 29 | 38 |
| 4 | 22 | 92 |
| 4 | 22 | 51 |
| 4 | 28 | 54 |
| 4 | 26 | 10 |
| 4 | 17 | 72 |
| 4 | 21 | 40 |
| 4 | 20 | 47 |
| 4 | 20 | 28 |
| 4 | 18 | 41 |
| 4 | 24 | 88 |
| 4 | 25 | 76 |
| 4 | 21 | 33 |
| 4 | 17 | 59 |
| 4 | 20 | 78 |
| 4 | 24 | 91 |
| 5 | 10 | 5 |
| 5 | 8 | 74 |
| 5 | 12 | 15 |
| 5 | 14 | 49 |
| 5 | 11 | 93 |
| 5 | 12 | 73 |
| 5 | 18 | 52 |
| 5 | 13 | 25 |
| 5 | 17 | 20 |
| 5 | 12 | 75 |
| 5 | 11 | 24 |
| 5 | 14 | 19 |
| 5 | 11 | 29 |
| 5 | 14 | 42 |
| 5 | 16 | 9 |
| 5 | 8 | 90 |
| 5 | 9 | 39 |
| 5 | 10 | 12 |
| 5 | 12 | 79 |
| 5 | 11 | 89 |

**BIOSTATISTICS 490 ASSIGNMENT 6**

**Two-way factorial Analysis of Variance**

This assignment will introduce you to factorial analysis of variance. You will learn how to execute and how to interpret a fixed effects factorial analysis of variance.

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

**For simplicity, you may assume *without testing* that the data meet the assumptions of ANOVA.**

The following data resulted from an experiment designed to test for the effects of growing conditions and fruit removal on production of flowers by morning glory vines. Individual vines were grown in plots with one of 4 soil treatments: 1) control; 2) Nitrogen fertilizer added; and 3) grown with clover (a Nitrogen fixing legume). The fruit produced by each flower was either: 1) removed immediately after fruit set; or 2) allowed to mature. The data represent number of flowers produced by a vine over the season.

|  |  |  |  |
| --- | --- | --- | --- |
|  | Soil treatment | | |
| Control | Nitrogen | Legume |
| Fruit | 41 46 37 48 | 49 52 47 53 | 61 55 62 |
| Not | 43 47 41 45 | 46 56 47 58 | 59 57 65 |
| Removed | 45 42 39 45 | 50 54 45 50 | 56 57 |
| Fruit | 48 56 50 | 49 59 50 | 60 60 57 |
| Removed | 49 49 52 | 55 53 54 | 53 55 57 |
|  | 50 54 46 | 55 55 47 | 55 61 59 |
|  | 46 | 53 57 51 | 58 |

The biological hypotheses of interest are: a) Resources invested in maturing fruit cannot be invested in producing more flowers, hence removal of fruits prior to maturation enables the plant to invest more in flowers (i.e., produce more flowers) than it would if fruit were allowed to mature. b) Flower production is Nitrogen limited, and addition of Nitrogen (by artificial additions, or by the presence of legumes) increases flower production. c) The effect of fruit removal may depend on availability of Nitrogen.

1. Do a two way ANOVA on these data, and test for main effects and interactions.
2. Do appropriate follow up tests to assess the biological hypotheses described above. Report means and SE's graphically.

**BIOSTATISTICS 490 ASSIGNMENT 7**

**Mixed Model Analysis of Variance: Comparing procedures**

This assignment will introduce you to mixed model ANOVA, and the several *SAS* procedures for doing mixed model ANOVA. Comparisons of output and conclusions for these procedures is the center of this assignment

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

### For simplicity, you may assume *without testing* that the data meet the assumptions of ANOVA.

**PART 1.** The following data are from Box 9.5 of your text. An investigator sampled limpets (snails) living on oyster shells. He sampled 4 zones running up the shore: **Seaward shore – mangroves; Seaward shore + mangroves; Middle zone + trees; Landward zone.** He is interested in the hypothesis that distance-vegetation combination affects limpet numbers. He did this sampling at two randomly chosen sites (chosen from among many). He then counted the number of limpets / 100 oysters in 5 quadrats at each site, in each zone.

The data have been square root transformed:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SITE | ZONE | SQLIM100 | SITE | ZONE | SQLIM100 |
| A | SZ(-TR) | 4 | B | SZ(-TR) | 4 |
| A | SZ(-TR) | 3.317 | B | SZ(-TR) | 3.464 |
| A | SZ(-TR) | 3.162 | B | SZ(-TR) | 4.123 |
| A | SZ(-TR) | 4 | B | SZ(-TR) | 2.828 |
| A | SZ(-TR) | 3.873 | B | SZ(-TR) | 2.828 |
| A | SZ(+TR) | 3.464 | B | SZ(+TR) | 3.742 |
| A | SZ(+TR) | 0 | B | SZ(+TR) | 4 |
| A | SZ(+TR) | 1.732 | B | SZ(+TR) | 2.646 |
| A | SZ(+TR) | 2.236 | B | SZ(+TR) | 3.873 |
| A | SZ(+TR) | 6.557 | B | SZ(+TR) | 1.414 |
| A | MZ | 0 | B | MZ | 0 |
| A | MZ | 0 | B | MZ | 5 |
| A | MZ | 4.123 | B | MZ | 7.071 |
| A | MZ | 0 | B | MZ | 4.123 |
| A | MZ | 0 | B | MZ | 7.071 |
| A | LZ | 0 | B | LZ | 0 |
| A | LZ | 0 | B | LZ | 0 |
| A | LZ | 0 | B | LZ | 2.828 |
| A | LZ | 0 | B | LZ | 0 |
| A | LZ | 5.745 | B | LZ | 0 |

**Analyze these data** using PROC GLM, PROC VARCOMP, and PROC MIXED. Compare the outputs, and compare these results to the results in Box 9.5.

**PART 2.** Consider the effect of unbalanced data on the analysis of mixed models. What would be the result if the previous data set had not been balanced? Would the agreement/disagreement among the 3 *SAS* procedures be greater or less?

The modified data have again been square root transformed:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SITE | ZONE | SQLIM100 | SITE | ZONE | SQLIM100 |
| A | SZ(-TR) | 4 | A | SZ(-TR) | 4 |
| A | SZ(-TR) | 3.317 | B | SZ(-TR) | 3.464 |
| A | SZ(-TR) | 3.162 | B | SZ(-TR) | 4.123 |
| A | SZ(-TR) | 4 | B | SZ(-TR) | 2.828 |
| A | SZ(-TR) | 3.873 | B | SZ(-TR) | 2.828 |
| A | SZ(+TR) | 3.464 | B | SZ(+TR) | 3.742 |
| A | SZ(+TR) |  | B | SZ(+TR) | 4 |
| A | SZ(+TR) | 1.732 | B | SZ(+TR) | 2.646 |
| A | SZ(+TR) | 2.236 | B | SZ(+TR) | 3.873 |
| A | SZ(+TR) | 6.557 | B | SZ(+TR) | 1.414 |
| A | MZ | 0 | B | MZ | 0 |
| A | MZ | 0 | B | MZ | 5 |
| A | MZ | 4.123 | B | MZ | 7.071 |
| A | MZ |  | B | MZ | 4.123 |
| A | MZ |  | B | MZ | 7.071 |
| A | LZ |  | B | LZ | 0 |
| A | LZ | 0 | B | LZ |  |
| A | LZ | 0 | B | LZ | 2.828 |
| A | LZ | 0 | B | LZ |  |
| A | LZ | 5.745 | B | LZ |  |

**Analyze these data** using PROC GLM, PROC VARCOMP, and PROC MIXED. Compare the outputs. What, if anything, can you conclude about the effects of unbalanced designs on mixed model ANOVA?

**BIOSTATISTICS 490 ASSIGNMENT 8**

**Two-stage nested Analysis of Variance**

This assignment will introduce you to nested analysis of variance. You will learn how to execute and how to interpret a two-stage, nested, random effects analysis of variance.

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

### For simplicity, you may assume *without testing* that the data meet the assumptions of ANOVA.

The data below are from a sampling plan designed to estimate variance components for phosphate concentration (mg/L) in water. Three sites are to be compared. Within each site, 4 sample points are randomly chosen. At each sample point, 5 water replicate samples are taken at random. The results:

|  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| SITE |  | Industrial | |  |  | Suburban | |  |  | Rural | |  |
| SAMPLE  POINT | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 | 1 | 2 | 3 | 4 |
| rep. 1 | 1.0 | 1.3 | 1.1 | 1.0 | 1.3 | 1.5 | 1.6 | 1.2 | 1.8 | 2.1 | 2.2 | 2.1 |
| rep. 2 | 1.2 | 1.2 | 1.3 | 0.9 | 1.5 | 1.4 | 1.8 | 1.5 | 1.7 | 2.1 | 1.9 | 2.3 |
| rep. 3 | 1.0 | 1.5 | 1.4 | 1.1 | 1.7 | 1.4 | 1.4 | 1.6 | 1.5 | 1.8 | 1.8 | 1.8 |
| rep. 4 | 1.3 | 1.1 | 1.0 | 0.8 | 1.2 | 1.8 | 1.5 | 1.4 | 1.4 | 1.9 | 2.0 | 1.8 |
| rep. 5 | 0.9 | 1.2 | 1.0 | 1.1 | 1.4 | 1.6 | 1.6 | 1.1 | 1.9 | 2.0 | 2.0 | 2.0 |

Perform an analysis of variance to test for site differences and variance components. Obtain **point** estimates of the variance components and obtain estimates of relative amounts of variation due to each component. Perform follow-up tests for site means. (NOTE: You may want to examine expected mean squares - see notes or text - and think about how to do follow-up tests for site).

Answer the following questions:

1. Explain your reasons for doing your follow-up tests in the way you did them. Be very specific.
2. What is the practical importance of your results for investigators planning to sample these lakes?

**BIOSTATISTICS 490 ASSIGNMENT 9**

**Linear & Multiple Regression**

This assignment will introduce you to linear and multiple regression. You will learn how to run and to interpret regressions, to test the regression assumptions, and to take remedial measures. You will about alternative ways of evaluating the suitability of regression models.

**For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

An investigator is attempting to determine the relationship of mass and age for laboratory mice. He obtains from a sample of mice of **known** birth date and **known** number of cage mates during rearing, and determines their masses at specific times. The data are given below. There are two parts to this assignment.

### Simple linear regression

* + 1. Do a linear regression of mass vs. age.
    2. Do the assumptions seem to be met?
    3. Test for nonlinearity using a lack of fit test. Does a linear regression fit the data?
    4. If the data do not meet the assumptions of regression, take appropriate remedial measures and re-do the analysis (include both complete analyses in the report). What conclusions do you obtain?
    5. Write a paragraph or two summarizing the results of the regression analysis, including all information that you think is useful.

### Multiple regression

* + 1. Do a regression of mass vs. both age and number of cage mates.
    2. Do the assumptions seem to be met?
    3. If the data do not meet the assumptions of regression, take appropriate remedial measures and re-do the analysis (include both complete analyses in the report). Are both variables important?
    4. Write a paragraph or two summarizing the results of the regression analysis, including all information that you think is useful.

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Cage | Mass | Age | Cage | Mass | Age | Cage | Mass |
| (days) | mates | (g) | (days) | mates | (g) | (days) | mates | (g) |
|  |  |  |  |  |  |  |  |  |
| 20 | 0 | 38.785 | 35 | 0 | 74.081 | 50 | 2 | 105.174 |
| 20 | 0 | 36.025 | 35 | 2 | 67.812 | 50 | 2 | 103.255 |
| 20 | 2 | 32.858 | 35 | 2 | 73.214 | 50 | 4 | 102.763 |
| 20 | 2 | 36.085 | 35 | 4 | 59.243 | 50 | 5 | 103.219 |
| 20 | 5 | 34.418 | 35 | 4 | 69.036 |  |  |  |
| 30 | 0 | 63.981 | 40 | 0 | 86.660 |  |  |  |
| 30 | 0 | 65.864 | 40 | 1 | 83.584 |  |  |  |
| 30 | 1 | 61.121 | 40 | 3 | 84.631 |  |  |  |
| 30 | 3 | 58.216 | 50 | 0 | 108.195 |  |  |  |
| 30 | 3 | 64.055 | 50 | 0 | 108.472 |  |  |  |

**BIOSTATISTICS 490 ASSIGNMENT 10**

**Analysis of Covariance**

This assignment will introduce you to analysis of covariance. You will learn how to test the assumptions of analysis of covariance, and how to interpret the results of an analysis of covariance. **For all tests, state assumptions, null & alternative hypotheses, test statistic, rejection region, and conclusion**.

An investigator is interested in the effect of temperature on growth and development of larvae of a mosquito. He raises mosquitoes individually at three temperatures. He is not able to control food supply (leaf material, which provides a substrate for the bacteria the larvae eat) precisely, but he records the amount of leaf material given to each individual mosquito. It is likely that amount of leaf material also affects growth and development. He records the number of days to adulthood and the mass at adulthood for each individual female mosquito. The data are below.

1. Do an analysis of covariance on these data in order to determine how the temperature influences size at and time to adulthood. Analyze both these variables. Use temperature as a **categorical variable.** Be sure to test assumptions, and to provide plots and summary statistics.
2. Use LSMEANS to compare the three temperatures.
3. Does the amount of leaf material added affect the outcome? Describe is its effect.
4. Write a few paragraphs summarizing your conclusions.

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **mosquito** | **temp** | **Leafmass** | **days** | **Mass** |
| 5 | 15 | 29 | 31 | 0.1761 |
| 7 | 15 | 33 | 34 | 0.1410 |
| 8 | 15 | 35 | 28 | 0.2876 |
| 11 | 15 | 40 | 24 | 0.3636 |
| 12 | 15 | 42 | 26 | 0.3036 |
| 16 | 15 | 49 | 26 | 0.3194 |
| 18 | 15 | 53 | 25 | 0.4123 |
| 19 | 15 | 54 | 26 | 0.4799 |
| 22 | 15 | 62 | 27 | 0.4518 |
| 25 | 15 | 66 | 26 | 0.5223 |
| 26 | 15 | 68 | 24 | 0.6478 |
| 27 | 15 | 69 | 26 | 0.5087 |
| 30 | 15 | 78 | 24 | 0.5642 |
| 31 | 15 | 81 | 25 | 0.4769 |
| 33 | 15 | 85 | 31 | 0.4703 |
| 34 | 15 | 88 | 26 | 0.6492 |
| 35 | 15 | 89 | 24 | 0.8107 |
| 38 | 15 | 95 | 25 | 0.6655 |
| 40 | 15 | 100 | 25 | 0.6585 |
| 41 | 20 | 23 | 22 | 0.1806 |
| 42 | 20 | 25 | 24 | 0.1069 |
| 45 | 20 | 29 | 24 | 0.2024 |
| 46 | 20 | 32 | 16 | 0.3265 |
| 50 | 20 | 38 | 16 | 0.3728 |
| 51 | 20 | 41 | 17 | 0.4049 |
| 52 | 20 | 42 | 18 | 0.5641 |
| 57 | 20 | 52 | 17 | 0.4097 |
| 60 | 20 | 57 | 16 | 0.5211 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **mosquito** | **temp** | **Leafmass** | **days** | **Mass** |
| 63 | 20 | 63 | 15 | 0.6903 |
| 64 | 20 | 64 | 15 | 0.6159 |
| 66 | 20 | 68 | 16 | 0.6923 |
| 67 | 20 | 69 | 15 | 0.6825 |
| 72 | 20 | 83 | 16 | 0.5059 |
| 74 | 20 | 89 | 16 | 0.6785 |
| 75 | 20 | 91 | 15 | 0.8204 |
| 76 | 20 | 92 | 15 | 0.6973 |
| 77 | 20 | 95 | 17 | 0.6576 |
| 80 | 20 | 100 | 16 | 0.7709 |
| 82 | 25 | 25 | 14 | 0.2092 |
| 84 | 25 | 28 | 14 | 0.1812 |
| 88 | 25 | 36 | 12 | 0.2948 |
| 89 | 25 | 37 | 12 | 0.2379 |
| 94 | 25 | 46 | 11 | 0.3927 |
| 95 | 25 | 49 | 12 | 0.3773 |
| 99 | 25 | 55 | 12 | 0.3954 |
| 102 | 25 | 63 | 12 | 0.4337 |
| 103 | 25 | 64 | 11 | 0.4574 |
| 104 | 25 | 65 | 12 | 0.5045 |
| 105 | 25 | 67 | 12 | 0.4445 |
| 106 | 25 | 68 | 12 | 0.6317 |
| 110 | 25 | 80 | 12 | 0.6498 |
| 113 | 25 | 88 | 12 | 0.7181 |
| 114 | 25 | 89 | 12 | 0.7217 |
| 117 | 25 | 95 | 12 | 0.7942 |
| 118 | 25 | 97 | 11 | 0.7115 |
| 120 | 25 | 102 | 12 | . |

# PART 3. Examples

## ANALYSIS OF VARIANCE: Simulations of the ANOVA model

The following two examples make use of the ANOVA model described in class to show by example what happens in an analysis when **H0** is false (example 1) and when **H0** is true (example 2). The purpose of the handout is to make more concrete the abstract description of the model that was given in lecture. Recall that the ANOVA model is:

(1) *Yij* = µ +*j* +*ij*

where *Yij* is the *i*th observation for treatment *j*, µ is the true overall mean for the population in question, *j* is the additive effect of treatment *j* on the mean, and *ij* is error, and is distributed normally with mean=0 and standard deviation=.

EXAMPLE 1. Suppose µ=10, and *ij* ~ ***N*** (0,0.5). Let there be *a*=3 treatments, with *j*'s=3.33, - 2.00, -1.33. Thus, the assumptions of the model are met, and **H0** is false (i.e., all *j*'s are not 0). Let there be *nj*=3 observations per treatment. We create observations by making a random draw from the distribution of *ij*, and determining *Yij* using eq. (1). The values obtained are:

Treatment (*j*)

|  |  |  |  |
| --- | --- | --- | --- |
| Obs. (*i*) | 1 | 2 | 3 |
| 1 | 10+3.33+0.035=13.365 | 10-2.00+0.178= 8.178 | 10-1.33-0.088= 8.582 |
| 2 | 10+3.33-0.021=13.309 | 10-2.00+0.786= 8.786 | 10-1.33+0.065= 8.735 |
| 3 | 10+3.33-0.214=13.116 | 10-2.00-0.747= 7.253 | 10-1.33+1.295= 9.965 |

|  |  |  |  |
| --- | --- | --- | --- |
| Sum | 39.790 | 24.217 | 27.282 |
| Est. mean (*Yj*) | 13.26 | 8.07 | 9.09 |
| True mean (µ + *j*) | 13.33 | 8.00 | 8.67 |

=

Overall est. mean (*Y* )=10.14 Overall true mean (µ)=10.00

Notice first that the estimates based on the data are close to the true values. An ANOVA table for these data is given below:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| SOURCE | DF | SS | MS (obs.) | *F* | MS (expected) |
| Betw. Treats. | 2 | *nj*(*Yj* - *Y*)2 = 45.375 | SS/DF = 22.687 | 56.860 | 25.54 |
| Within (error) | 6 |   (*Yij* -*Yj*)2 = 2.376 | SS/DF = 0.399 |  | 0.25 |
|  |  | *j i*  = |  |  |  |
| Total | 8 |   (*Yij* - *Y*)2 = 47.749 |  |  |  |
|  |  | *j i* |  |  |  |

Recall that expected mean squares are:

 *nj j*2

FOR TREATMENTS EMS = 2 +------------ FOR ERROR EMS = 2

*a*-1

You should verify that the expected mean squares reported in the table are correct. Note that observed and expected mean squares are in rather good agreement. Note also that an *F* test would lead to rejection of **H0** at a very low significance level, which is the correct conclusion.

EXAMPLE 2. Suppose µ=10, and *ij* ~ N(0,0.5), as before. Let there be *a*=3 treatments, with all *j*'s=0. Thus, the assumptions of the model are met, and **H0** is true. Again let there be *nj*=3 observations per treatment. We create observations by using the same random draws from the distribution of *ij*, and determining *Yij* using eq. (1). The values obtained are:

Treatment (*j*)

|  |  |  |  |
| --- | --- | --- | --- |
| Obs. (i) | 1 | 2 | 3 |
| 1 | 10+0+0.035=10.035 | 10+0+0.178= 10.178 | 10+0-0.088= 9.912 |
| 2 | 10+0-0.021= 9.979 | 10+0+0.786= 10.786 | 10+0+0.065=10.065 |
| 3 | 10+0-0.214= 9.786 | 10+0-0.747= 9.253 | 10+0+1.295=11.295 |

Notice again that the estimates based on the data are close to the true values. An ANOVA table for these data is given below:

|  |  |  |  |
| --- | --- | --- | --- |
| Sum | 29.800 | 30.211 | 31.272 |
| Est. mean (*Yj* ) | 9.93 | 10.07 | 10.42 |
| True mean (µ + *j* ) | 10.00 | 10.00 | 10.00 |
| = |  |  |  |
| Overall est. mean (*Y*) = 10.14 |  | Overall true mean | (µ)=10.00 |

DF SS MS (obs.) *F* MS (expected)

SOURCE

=

Betw. Treats. 2 *nj*(*Yj* – *Y* )2 = 0.385 SS/DF = 0.193 0.489 0.25

Within (error) 6   (*Yij* - *Yj* )2 = 2.376 SS/DF = 0.399 0.25

*j i*

=

Total 8   (*Yij* – *Y* )2 = 2.752

*j i*

Again, expected mean squares are:

 *nj* *j*2

FOR TREATMENTS EMS = 2 + ---------- FOR ERROR EMS = 2

*a*-1

Again, you should verify the values for expected mean squares. Note that observed and expected mean squares are in good agreement. Note also that an *F* test would lead to failure to reject H0, which is the correct conclusion.

What does this have to do with real life? The model, eq. (1), is an abstraction of how you suspect the experimental treatments + random errors influence the observations that you obtain. You **assume** that the observations are a product of a processes operating like eq. (1), and then construct the analysis to detect the presence (or absence) of the *j*'s, the treatment effects. These examples show that when the model is an accurate description of how the observations were generated, the analysis will work in the way we expect.

## ANOVA, MULTIPLE COMPARISONS, & CONTRASTS

(note: Boxes drawn around **Editor** and **Log** window contents

ods html close; /\* close previous \*/ ods html; /\* open new \*/

ods graphics on; DATA ANOVAEX001; INPUT TREAT $ Y; CARDS;

**/\***

**ANOVA Input \*/**

PROC GLM DATA=ANOVAEX001; CLASS TREAT;

MODEL Y = TREAT;

**/\* PROC GLM with multiple comparisons \*/**

MEANS TREAT / REGWQ TUKEY BON LINES

HOVTEST=BF; **/\* Test for homogeneous variance\*/**

RUN;

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| CONTROL | 5.5 |  | MEDIUM | 7.3 |
| CONTROL | 4.5 | | | MEDIUM | 6.8 |
| CONTROL | 5.6 | | | MEDIUM | 7.5 |
| CONTROL | 4.9 | | | MEDIUM | 6.9 |
| LOW | 5.9 | | | MEDIUM | 7.7 |
| LOW | 5.9 | | | HIGH | 10.0 |
| LOW | 6.1 | | | HIGH | 9.7 |
| LOW | 5.2 | | | HIGH | 8.9 |
| LOW | 6.6 | | | HIGH | 9.7 |
|  |  |  | HIGH | 8.9 |
|  |  |  | HIGH | 8.2 |
|  |  |  | ; |  |

### ANOVA

The GLM Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **TREAT** | 4 | CONTROL HIGH LOW MEDIUM |

|  |  |
| --- | --- |
| **Number of Observations Read** | 20 |
| **Number of Observations Used** | 20 |

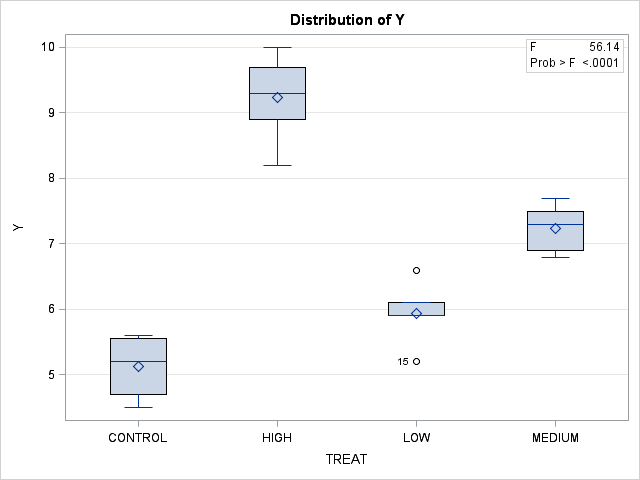
Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 49.73316667 | 16.57772222 | 56.14 | <.0001 |
| **Error** | 16 | 4.72483333 | 0.29530208 |  |  |
| **Corrected Total** | 19 | 54.45800000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913239 | 7.664556 | 0.543417 | 7.090000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **TREAT** | 3 | 49.73316667 | 16.57772222 | 56.14 | <.0001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **TREAT** | 3 | 49.73316667 | 16.57772222 | 56.14 | <.0001 |



The GLM Procedure

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Brown and Forsythe's Test for Homogeneity of Y Variance ANOVA of Absolute Deviations from Group Medians** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **TREAT** | 3 | 0.2492 | 0.0831 | 1.12 | 0.3712 |
| **Error** | 16 | 1.1888 | 0.0743 |  |  |

Ryan-Einot-Gabriel-Welsch Multiple Range Test for Y

|  |  |  |
| --- | --- | --- |
| Note: | This test controls the Type I experimentwise error rate. | |
|  | **Alpha** | 0.05 |
|  | **Error Degrees of Freedom** | 16 |
|  | **Error Mean Square** | 0.295302 |
|  | **Harmonic Mean of Cell Sizes** | 4.897959 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Note: | Cell sizes are not equal. | | | |
| **Number of Means** | | **2** | **3** | **4** |
| **Critical Range** | 0.8564688 | | 0.8960167 | 0.9934776 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Means with the same letter are not significantly different.** | | | |
| **REGWQ Grouping** | **Mean** | **N** | **TREAT** |
| A | 9.2333 | 6 | HIGH |
|  |  |  |  |
| B | 7.2400 | 5 | MEDIUM |
|  |  |  |  |
| C | 5.9400 | 5 | LOW |
| C |  |  |  |
| C | 5.1250 | 4 | CONTROL |

Tukey's Studentized Range (HSD) Test for Y

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Note: | This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ. | | | |
|  | | **Alpha** | 0.05 |  |
| **Error Degrees of Freedom** | 16 |
| **Error Mean Square** | 0.295302 |
| **Critical Value of Studentized Range** | 4.04606 |
| **Minimum Significant Difference** | 0.9935 |
| **Harmonic Mean of Cell Sizes** | 4.897959 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
|  | Note: | Cell sizes are not equal. | | | |  |
| **Means with the same letter are not significantly different.** | | | | | | |
| **Tukey Grouping** | | | **Mean** | **N** | **TREAT** | |
| A | | | 9.2333 | 6 | HIGH | |
|  | | |  |  |  | |
| B | | | 7.2400 | 5 | MEDIUM | |
|  | | |  |  |  | |
| C | | | 5.9400 | 5 | LOW | |
| C | | |  |  |  | |
| C | | | 5.1250 | 4 | CONTROL | |

Bonferroni (Dunn) t Tests for Y

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Note: | This test controls the Type I experimentwise error rate, but it generally has a higher Type II error rate than REGWQ. | | | |
|  | | **Alpha** | 0.05 |  |
| **Error Degrees of Freedom** | 16 |
| **Error Mean Square** | 0.295302 |
| **Critical Value of t** | 3.00833 |
| **Minimum Significant Difference** | 1.0446 |
| **Harmonic Mean of Cell Sizes** | 4.897959 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Note: | Cell sizes are not equal. | | | |
| **Means with the same letter are not significantly different.** | | | | |
| **Bon Grouping** | | **Mean** | **N** | **TREAT** |
| A | | 9.2333 | 6 | HIGH |
|  | |  |  |  |
| B | | 7.2400 | 5 | MEDIUM |
|  | |  |  |  |
| C | | 5.9400 | 5 | LOW |
| C | |  |  |  |
| C | | 5.1250 | 4 | CONTROL |

RUN;

CONTRAST 'CONTROL VS. TREAT' TREAT -3 1 1 1;

CONTRAST 'CONTROL VS. TREAT' TREAT -1 0.333333 0.333333 0.333333;

ESTIMATE 'CONTROL VS. TREAT' TREAT -3 1 1 1 / DIVISOR=3;

**/\* PROC GLM with contrasts \*/**

PROC GLM DATA=ANOVAEX001; CLASS TREAT;

MODEL Y = TREAT; MEANS TREAT;

The GLM Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **TREAT** | 4 | CONTROL HIGH LOW MEDIUM |

|  |  |
| --- | --- |
| **Number of Observations Read** | 20 |
| **Number of Observations Used** | 20 |

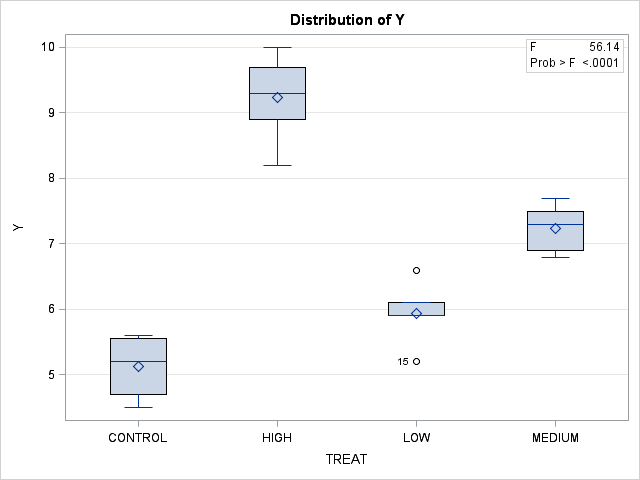
Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 49.73316667 | 16.57772222 | 56.14 | <.0001 |
| **Error** | 16 | 4.72483333 | 0.29530208 |  |  |
| **Corrected Total** | 19 | 54.45800000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913239 | 7.664556 | 0.543417 | 7.090000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **TREAT** | 3 | 49.73316667 | 16.57772222 | 56.14 | <.0001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **TREAT** | 3 | 49.73316667 | 16.57772222 | 56.14 | <.0001 |



|  |  |  |  |
| --- | --- | --- | --- |
| **Level of TREAT** | **N** | **Y** | |
| **Mean** | **Std Dev** |
| **CONTROL** | **4** | 5.12500000 | 0.51881275 |
| **HIGH** | **6** | 9.23333333 | 0.68019605 |
| **LOW** | **5** | 5.94000000 | 0.50299105 |
| **MEDIUM** | **5** | 7.24000000 | 0.38470768 |

The GLM Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Contrast** | **DF** | **Contrast SS** | **Mean Square** | **F Value** | **Pr > F** |
| **CONTROL VS. TREAT** | 1 | 17.58750394 | 17.58750394 | 59.56 | <.0001 |
| **CONTROL VS. TREAT** | 1 | 17.58750006 | 17.58750006 | 59.56 | <.0001 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **CONTROL VS. TREAT** | 2.34611111 | 0.30400430 | 7.72 | <.0001 |

## UNBALANCED ANOVA EXAMPLE

PROC ANOVA DATA=UNBAL001; CLASS ROW COLUMN;

MODEL Y=ROW COLUMN ROW\*COLUMN; PROC ANOVA DATA=UNBAL001;

CLASS ROW COLUMN;

MODEL Y=COLUMN ROW\*COLUMN ROW; PROC GLM DATA=UNBAL001;

CLASS ROW COLUMN;

MODEL Y=COLUMN ROW ROW\*COLUMN / SS1 SS2 SS3 SS4; PROC GLM DATA=UNBAL001;

CLASS ROW COLUMN;

MODEL Y=ROW ROW\*COLUMN COLUMN / SS1 SS2 SS3 SS4; PROC GLM DATA=UNBAL001;

CLASS ROW COLUMN;

MODEL Y=ROW\*COLUMN COLUMN ROW / SS1 SS2 SS3 SS4;

Run; Quit;

17

15

18

16

21

19

22

2 1

2 2

2 2

2 2

2 3

2 3

2 3

;

1 1 10

1 1 11

1 1 9

1 1 12

1 2 15 |

1 2 16 |

1 2 17 |

1 3 18 |

1 3 19 |

2 1 15 |

**/\* Unbalanced ANOVA: Input \*/**

DATA UNBAL001

INPUT ROW COLUMN Y; CARDS;



1. CLASS ROW COLUMN;
2. MODEL Y=ROW COLUMN ROW\*COLUMN;

WARNING: PROC ANOVA has determined that the number of observations in each cell is not equal.

PROC GLM may be more appropriate.

NOTE: PROCEDURE ANOVA used (Total process time): real time 0.05 seconds

cpu time 0.01 seconds

**Unbalanced ANOVA: Log**

117 PROC ANOVA DATA=UNBAL001;

The ANOVA Procedure

### Unbalanced ANOVA: Output

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **ROW** | 2 | 1 2 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |

|  |  |
| --- | --- |
| **Number of Observations Used** | 17 |

The ANOVA Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 198.9313725 | 39.7862745 | 23.24 | <.0001 |
| **Error** | 11 | 18.8333333 | 1.7121212 |  |  |
| **Corrected Total** | 16 | 217.7647059 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913515 | 8.238581 | 1.308480 | 15.88235 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.0008170 | 60.0008170 | 35.04 | 0.0001 |
| **COLUMN** | 2 | 152.7980392 | 76.3990196 | 44.62 | <.0001 |
| **ROW\*COLUMN** | 2 | 0.0000000 | 0.0000000 | 0.00 | 1.0000 |

The ANOVA Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **ROW** | 2 | 1 2 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

The ANOVA Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 198.9313725 | 39.7862745 | 23.24 | <.0001 |
| **Error** | 11 | 18.8333333 | 1.7121212 |  |  |
| **Corrected Total** | 16 | 217.7647059 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913515 | 8.238581 | 1.308480 | 15.88235 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Anova SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 152.7980392 | 76.3990196 | 44.62 | <.0001 |
| **ROW\*COLUMN** | 2 | 0.0000000 | 0.0000000 | 0.00 | 1.0000 |
| **ROW** | 1 | 60.0008170 | 60.0008170 | 35.04 | 0.0001 |

The GLM Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **ROW** | 2 | 1 2 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

The GLM Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 198.9313725 | 39.7862745 | 23.24 | <.0001 |
| **Error** | 11 | 18.8333333 | 1.7121212 |  |  |
| **Corrected Total** | 16 | 217.7647059 |  |  |  |

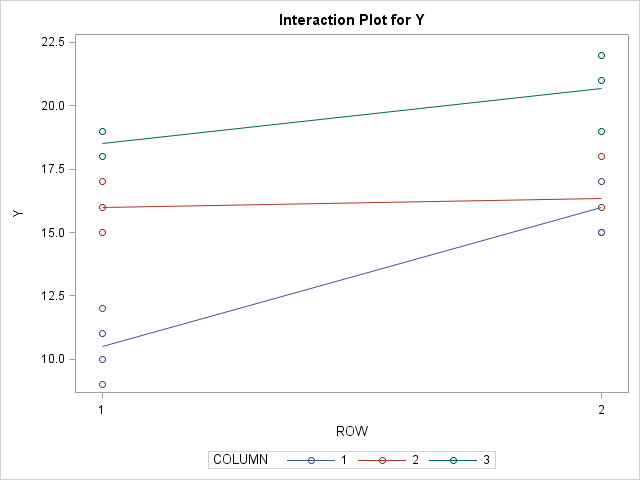
|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913515 | 8.238581 | 1.308480 | 15.88235 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 152.7980392 | 76.3990196 | 44.62 | <.0001 |
| **ROW** | 1 | 26.9887052 | 26.9887052 | 15.76 | 0.0022 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type II SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 119.7859275 | 59.8929637 | 34.98 | <.0001 |
| **ROW** | 1 | 26.9887052 | 26.9887052 | 15.76 | 0.0022 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 101.3595041 | 50.6797521 | 29.60 | <.0001 |
| **ROW** | 1 | 28.4444444 | 28.4444444 | 16.61 | 0.0018 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type IV SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 101.3595041 | 50.6797521 | 29.60 | <.0001 |
| **ROW** | 1 | 28.4444444 | 28.4444444 | 16.61 | 0.0018 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |



The GLM Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **ROW** | 2 | 1 2 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

The GLM Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 198.9313725 | 39.7862745 | 23.24 | <.0001 |
| **Error** | 11 | 18.8333333 | 1.7121212 |  |  |
| **Corrected Total** | 16 | 217.7647059 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913515 | 8.238581 | 1.308480 | 15.88235 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.0008170 | 60.0008170 | 35.04 | 0.0001 |
| **ROW\*COLUMN** | 4 | 138.9305556 | 34.7326389 | 20.29 | <.0001 |
| **COLUMN** | 0 | 0.0000000 | . | . | . |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type II SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 26.9887052 | 26.9887052 | 15.76 | 0.0022 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |
| **COLUMN** | 2 | 119.7859275 | 59.8929637 | 34.98 | <.0001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 28.4444444 | 28.4444444 | 16.61 | 0.0018 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |
| **COLUMN** | 2 | 101.3595041 | 50.6797521 | 29.60 | <.0001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type IV SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 28.4444444 | 28.4444444 | 16.61 | 0.0018 |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |
| **COLUMN** | 2 | 101.3595041 | 50.6797521 | 29.60 | <.0001 |

The GLM Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **ROW** | 2 | 1 2 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |

|  |  |
| --- | --- |
| **Number of Observations Used** | 17 |

The GLM Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 198.9313725 | 39.7862745 | 23.24 | <.0001 |
| **Error** | 11 | 18.8333333 | 1.7121212 |  |  |
| **Corrected Total** | 16 | 217.7647059 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913515 | 8.238581 | 1.308480 | 15.88235 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW\*COLUMN** | 5 | 198.9313725 | 39.7862745 | 23.24 | <.0001 |
| **COLUMN** | 0 | 0.0000000 | . | . | . |
| **ROW** | 0 | 0.0000000 | . | . | . |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type II SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |
| **COLUMN** | 2 | 119.7859275 | 59.8929637 | 34.98 | <.0001 |
| **ROW** | 1 | 26.9887052 | 26.9887052 | 15.76 | 0.0022 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |
| **COLUMN** | 2 | 101.3595041 | 50.6797521 | 29.60 | <.0001 |
| **ROW** | 1 | 28.4444444 | 28.4444444 | 16.61 | 0.0018 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type IV SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW\*COLUMN** | 2 | 19.1446281 | 9.5723140 | 5.59 | 0.0211 |
| **COLUMN** | 2 | 101.3595041 | 50.6797521 | 29.60 | <.0001 |
| **ROW** | 1 | 28.4444444 | 28.4444444 | 16.61 | 0.0018 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| 1 | 1 | 10 |  | | |
| 1 | 1 | 11 | 1 | 3 | 17 |
| 1 | 1 | 9 | | 1 | 3 | 19 |
| 1 | 1 | 12 | | 2 | 2 | 15 |
| 1 | 2 | 15 | | 2 | 2 | 18 |
| 1 | 2 | 16 | | 2 | 2 | 16 |
| 1 | 2 | 17 | | 2 | 3 | 21 |
|  |  |  | 2 | 3 | 19 |
|  |  |  | 2 | 3 | 22 |
|  |  |  | ; |  |  |

### Type III vs. IV Sums of Squares: Output

DATA UNBAL002; **/\* Type III vs. IV Sums of Squares: Input \*/**

INPUT ROW COLUMN Y;

CARDS; **/\* Note: combination row 2 column 1 completely absent \*/**

PROC GLM DATA=UNBAL002; CLASS ROW COLUMN;

MODEL Y=COLUMN ROW ROW\*COLUMN / SS3 SS4;

The GLM Procedure Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 4 | 194.0666667 | 48.5166667 | 26.46 | <.0001 |
| **Error** | 10 | 18.3333333 | 1.8333333 |  |  |
| **Corrected Total** | 14 | 212.4000000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.913685 | 8.569661 | 1.354006 | 15.80000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 110.0544218 | 55.0272109 | 30.01 | <.0001 |
| **ROW** | 1 | 6.0000000 | 6.0000000 | 3.27 | 0.1006 |
| **ROW\*COLUMN** | 1 | 3.6296296 | 3.6296296 | 1.98 | 0.1897 |

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Source** | **DF** |  | **Type IV SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | \* | 76.52380952 | 38.26190476 | 20.87 | 0.0003 |
| **ROW** | 1 | \* | 6.00000000 | 6.00000000 | 3.27 | 0.1006 |
| **ROW\*COLUMN** | 1 |  | 3.62962963 | 3.62962963 | 1.98 | 0.1897 |

\* NOTE: Other Type IV Testable Hypotheses exist which may yield different SS

## LEAST SQUARES MEANS

PROC GLM DATA=LSMEANS; CLASS FAC1 FAC2;

MODEL Y=FAC1|FAC2 / SS3; MEANS FAC1 FAC2 FAC1\*FAC2;

LSMEANS FAC1 FAC2 FAC1\*FAC2 / STDERR PDIFF;

LSMEANS FAC1 FAC2 FAC1\*FAC2 / STDERR PDIFF ADJUST=TUKEY;

RUN;

B A 3

B A 4

B B 5

B B 6

B B 7

;

A B 5 |

INPUT FAC1 $ FAC2 $ Y; CARDS;

A A 2

A A 3 |

A A 4 |

A B 4 |

**/\* Least Squares means – input \*/**

DATA LSMEANS;

The GLM Procedure **Least Squares means - output**

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **FAC1** | 2 | A B |
| **FAC2** | 2 | A B |

|  |  |
| --- | --- |
| **Number of Observations Read** | 10 |
| **Number of Observations Used** | 10 |

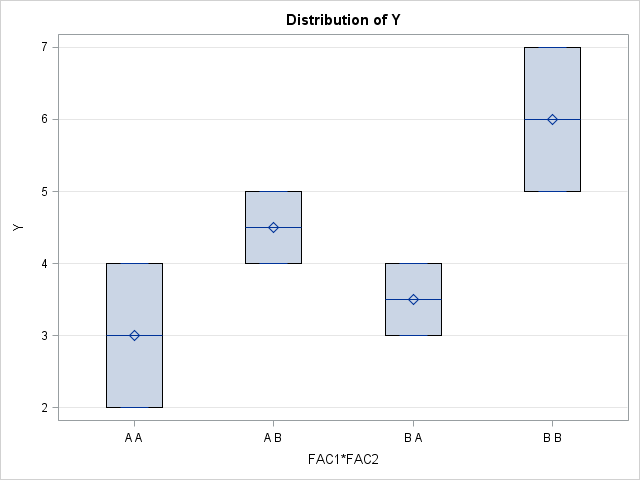
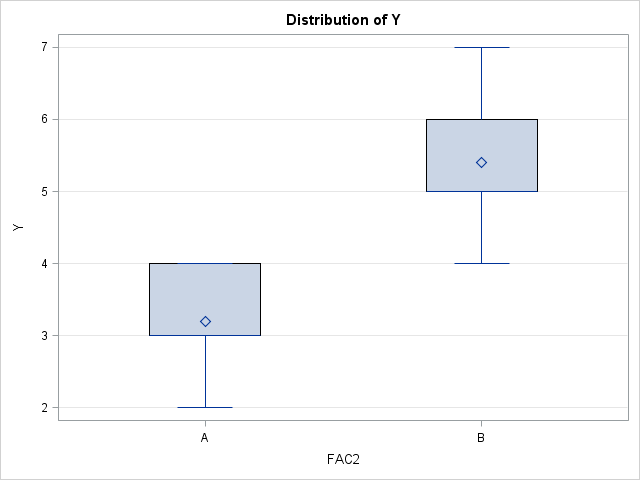
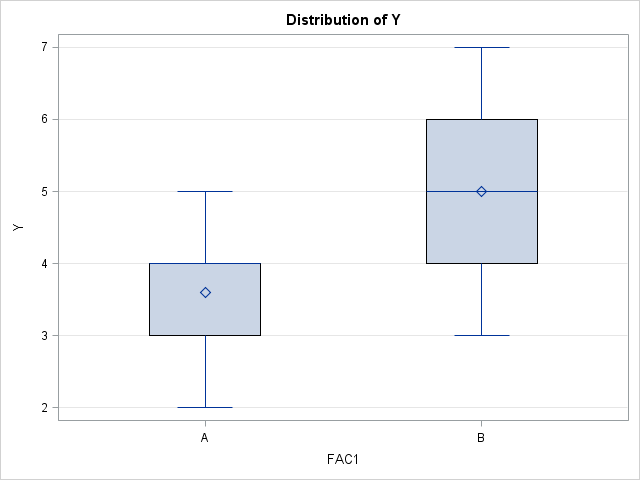
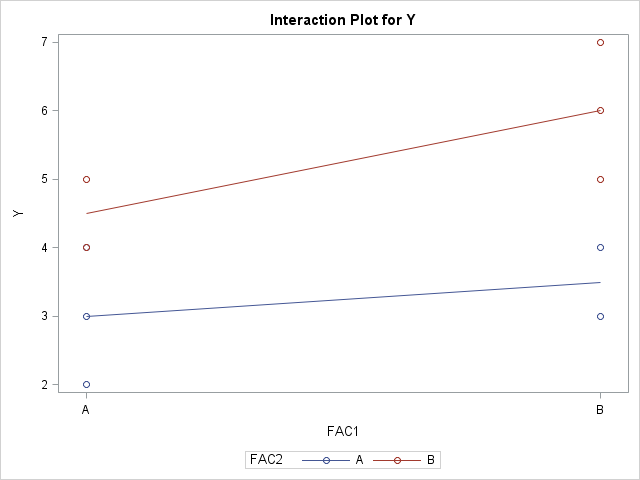
Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 15.10000000 | 5.03333333 | 6.04 | 0.0304 |
| **Error** | 6 | 5.00000000 | 0.83333333 |  |  |
| **Corrected Total** | 9 | 20.10000000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.751244 | 21.22956 | 0.912871 | 4.300000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **FAC1** | 1 | 2.40000000 | 2.40000000 | 2.88 | 0.1406 |
| **FAC2** | 1 | 9.60000000 | 9.60000000 | 11.52 | 0.0146 |
| **FAC1\*FAC2** | 1 | 0.60000000 | 0.60000000 | 0.72 | 0.4287 |

### Arithmetic Means



|  |  |  |  |
| --- | --- | --- | --- |
| **Level of FAC1** | **N** | **Y** | |
| **Mean** | **Std Dev** |
| **A** | **5** | 3.60000000 | 1.14017543 |
| **B** | **5** | 5.00000000 | 1.58113883 |

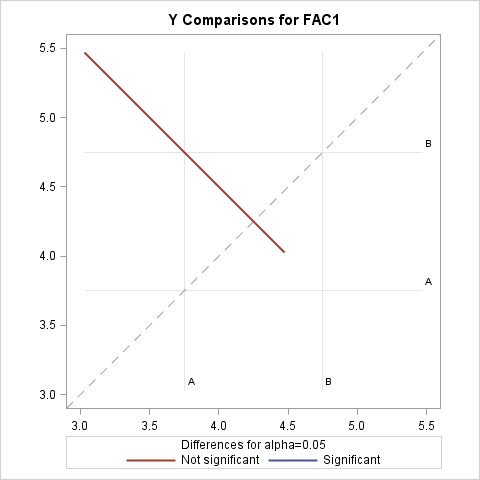
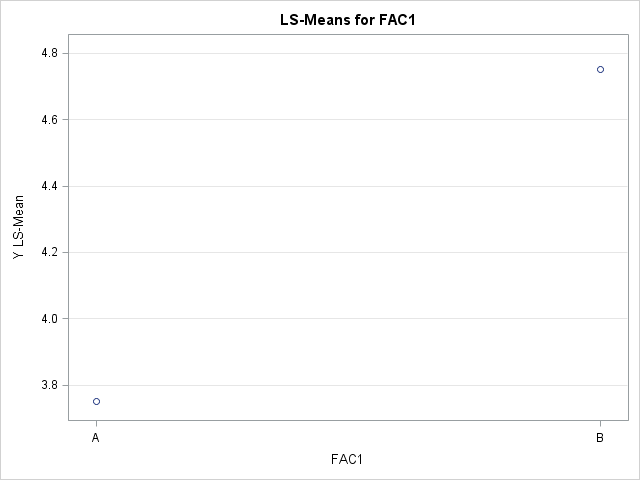
|  |  |  |  |
| --- | --- | --- | --- |
| **Level of FAC2** | **N** | **Y** | |
| **Mean** | **Std Dev** |
| **A** | **5** | 3.20000000 | 0.83666003 |
| **B** | **5** | 5.40000000 | 1.14017543 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Level of FAC1** | **Level of FAC2** | **N** | **Y** | |
| **Mean** | **Std Dev** |
| **A** | **A** | **3** | 3.00000000 | 1.00000000 |
| **A** | **B** | **2** | 4.50000000 | 0.70710678 |
| **B** | **A** | **2** | 3.50000000 | 0.70710678 |
| **B** | **B** | **3** | 6.00000000 | 1.00000000 |

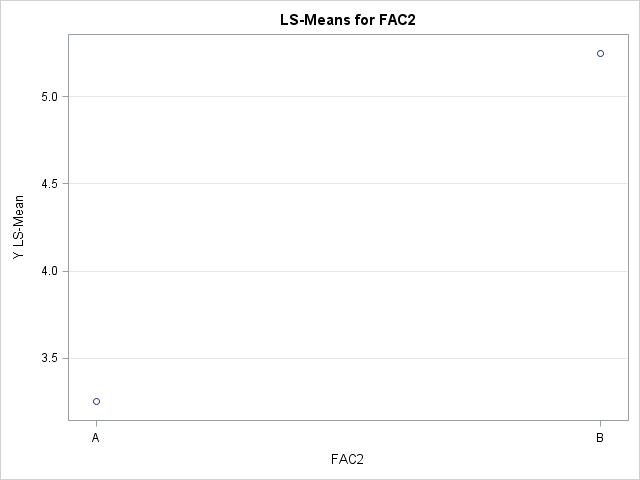
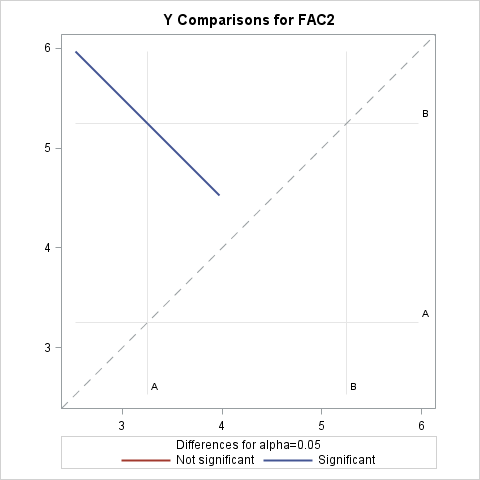
The GLM Procedure **Least squares Means**

Least Squares Means

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FAC1** | **Y LSMEAN** | **Standard**  **Error** | **H0:LSMEAN=0** | **H0:LSMean1=LSMean2** |
| **Pr > |t|** | **Pr > |t|** |
| **A** | 3.75000000 | 0.41666667 | 0.0001 | 0.1406 |
| **B** | 4.75000000 | 0.41666667 | <.0001 |  |

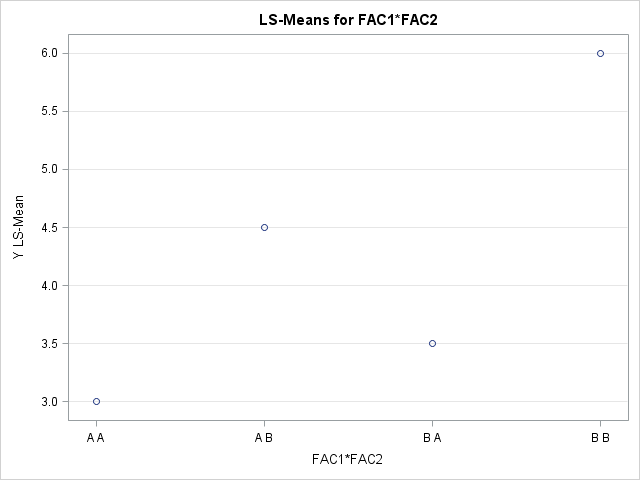
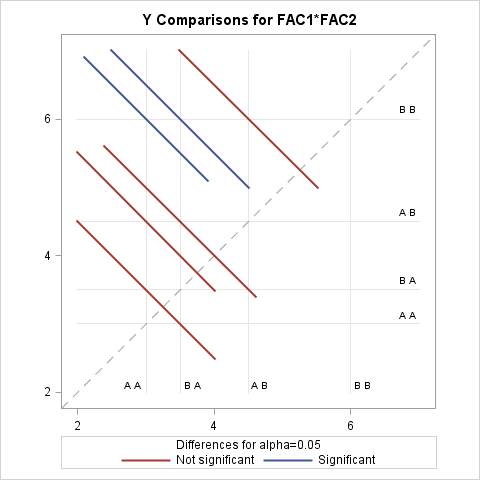


|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **FAC2** | **Y LSMEAN** | **Standard**  **Error** | **H0:LSMEAN=0** | **H0:LSMean1=LSMean2** |
| **Pr > |t|** | **Pr > |t|** |
| **A** | 3.25000000 | 0.41666667 | 0.0002 | 0.0146 |
| **B** | 5.25000000 | 0.41666667 | <.0001 |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **FAC1** | **FAC2** | **Y LSMEAN** | **Standard**  **Error** | **Pr > |t|** | **LSMEAN Number** |
| **A** | **A** | 3.00000000 | 0.52704628 | 0.0013 | 1 |
| **A** | **B** | 4.50000000 | 0.64549722 | 0.0004 | 2 |
| **B** | **A** | 3.50000000 | 0.64549722 | 0.0016 | 3 |
| **B** | **B** | 6.00000000 | 0.52704628 | <.0001 | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Least Squares Means for effect FAC1\*FAC2 Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: Y** | | | | |
| **i/j** | **1** | **2** | **3** | **4** |
| **1** |  | 0.1220 | 0.5705 | 0.0069 |
| **2** | 0.1220 |  | 0.3153 | 0.1220 |
| **3** | 0.5705 | 0.3153 |  | 0.0240 |
| **4** | 0.0069 | 0.1220 | 0.0240 |  |

|  |  |
| --- | --- |
| Note: | To ensure overall protection level, only probabilities associated with pre-planned comparisons should be used. |

### Tukey adjustment for

### Mulitple tests

Least Squares Means

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Adjustment for Multiple Comparisons: Tukey | | | | | | |
|  | **FAC1** | **Y LSMEAN** | **Standard**  **Error** | **H0:LSMEAN=0** | **H0:LSMean1=LSMean2** |  |
| **Pr > |t|** | **Pr > |t|** |
| **A** | 3.75000000 | 0.41666667 | 0.0001 | 0.1406 |
| **B** | 4.75000000 | 0.41666667 | <.0001 |  |

Least Squares Means

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Adjustment for Multiple Comparisons: Tukey | | | | | | |
|  | **FAC2** | **Y LSMEAN** | **Standard**  **Error** | **H0:LSMEAN=0** | **H0:LSMean1=LSMean2** |  |
| **Pr > |t|** | **Pr > |t|** |
| **A** | 3.25000000 | 0.41666667 | 0.0002 | 0.0146 |
| **B** | 5.25000000 | 0.41666667 | <.0001 |  |

Least Squares Means

Adjustment for Multiple Comparisons: Tukey-Kramer

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **FAC1** | **FAC2** | **Y LSMEAN** | **Standard**  **Error** | **Pr > |t|** | **LSMEAN Number** |
| **A** | **A** | 3.00000000 | 0.52704628 | 0.0013 | 1 |
| **A** | **B** | 4.50000000 | 0.64549722 | 0.0004 | 2 |
| **B** | **A** | 3.50000000 | 0.64549722 | 0.0016 | 3 |
| **B** | **B** | 6.00000000 | 0.52704628 | <.0001 | 4 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Least Squares Means for effect FAC1\*FAC2 Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: Y** | | | | |
| **i/j** | **1** | **2** | **3** | **4** |
| **1** |  | 0.3578 | 0.9284 | 0.0266 |
| **2** | 0.3578 |  | 0.7050 | 0.3578 |
| **3** | 0.9284 | 0.7050 |  | 0.0860 |
| **4** | 0.0266 | 0.3578 | 0.0860 |  |

## ANOVA MODELS WITH RANDOM EFFECTS

2 3 15 |

\*\*\* FIXED EFFECTS MODEL \*\*\*; PROC GLM DATA=RANDOM;

CLASS ROW COLUMN;

MODEL Y=ROW COLUMN ROW\*COLUMN / SS3;

\*\*\* RANDOM EFFECTS MODEL \*\*\*; PROC GLM DATA=RANDOM;

CLASS ROW COLUMN;

MODEL Y=ROW COLUMN ROW\*COLUMN / SS3; RANDOM ROW COLUMN ROW\*COLUMN / TEST; TEST H=ROW COLUMN E=ROW\*COLUMN;

\*\*\* MIXED MODEL \*\*\*; PROC GLM DATA=RANDOM;

CLASS ROW COLUMN;

MODEL Y=ROW COLUMN ROW\*COLUMN / SS3; RANDOM ROW ROW\*COLUMN / TEST;

TEST H=COLUMN E=ROW\*COLUMN; PROC MIXED DATA=RANDOM;

CLASS ROW COLUMN; MODEL Y=COLUMN;

RANDOM ROW ROW\*COLUMN; RUN;QUIT;

15

16

2 3

2 3

;

|

|

1 3 15

2 1 15

2 1 16

2 1 17

2 2 18

2 2 18

2 2 17

15 |

|

14

|

|

|

|

10

11

12

12

12

13

1 1

1 1

1 1

1 2

1 2

1 2

1 3

1 3

**/\*Random and Mixed 2-way ANOVA: Input\*/**

DATA RANDOM;

INPUT ROW COLUMN Y; CARDS;

### Random and Mixed 2-way ANOVA: Output

The GLM Procedure **Fixed effects**

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **ROW** | 2 | 1 2 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 18 |
| **Number of Observations Used** | 18 |

Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 89.83333333 | 17.96666667 | 32.34 | <.0001 |
| **Error** | 12 | 6.66666667 | 0.55555556 |  |  |
| **Corrected Total** | 17 | 96.50000000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.930915 | 5.140386 | 0.745356 | 14.50000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.50000000 | 60.50000000 | 108.90 | <.0001 |
| **COLUMN** | 2 | 9.00000000 | 4.50000000 | 8.10 | 0.0059 |
| **ROW\*COLUMN** | 2 | 20.33333333 | 10.16666667 | 18.30 | 0.0002 |

Dependent Variable: Y **Random effects**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 89.83333333 | 17.96666667 | 32.34 | <.0001 |
| **Error** | 12 | 6.66666667 | 0.55555556 |  |  |
| **Corrected Total** | 17 | 96.50000000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.930915 | 5.140386 | 0.745356 | 14.50000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.50000000 | 60.50000000 | 108.90 | <.0001 |
| **COLUMN** | 2 | 9.00000000 | 4.50000000 | 8.10 | 0.0059 |
| **ROW\*COLUMN** | 2 | 20.33333333 | 10.16666667 | 18.30 | 0.0002 |

### From the RANDOM statement

The GLM Procedure

|  |  |
| --- | --- |
| **Source** | **Type III Expected Mean Square** |
| **ROW** | Var(Error) + 3 Var(ROW\*COLUMN) + 9 Var(ROW) |
| **COLUMN** | Var(Error) + 3 Var(ROW\*COLUMN) + 6 Var(COLUMN) |
| **ROW\*COLUMN** | Var(Error) + 3 Var(ROW\*COLUMN) |

Tests of Hypotheses for Random Model Analysis of Variance Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.500000 | 60.500000 | 5.95 | 0.1349 |
| **COLUMN** | 2 | 9.000000 | 4.500000 | 0.44 | 0.6932 |
| **Error** | 2 | 20.333333 | 10.166667 |  |  |
| **Error: MS(ROW\*COLUMN)** | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW\*COLUMN** | 2 | 20.333333 | 10.166667 | 18.30 | 0.0002 |
| **Error: MS(Error)** | 12 | 6.666667 | 0.555556 |  |  |

Dependent Variable: Y **From the TEST statement**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tests of Hypotheses Using the Type III MS for ROW\*COLUMN as an Error Term** | | | | | |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.50000000 | 60.50000000 | 5.95 | 0.1349 |
| **COLUMN** | 2 | 9.00000000 | 4.50000000 | 0.44 | 0.6932 |

Dependent Variable: Y **Mixed model**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 89.83333333 | 17.96666667 | 32.34 | <.0001 |
| **Error** | 12 | 6.66666667 | 0.55555556 |  |  |
| **Corrected Total** | 17 | 96.50000000 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.930915 | 5.140386 | 0.745356 | 14.50000 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.50000000 | 60.50000000 | 108.90 | <.0001 |
| **COLUMN** | 2 | 9.00000000 | 4.50000000 | 8.10 | 0.0059 |
| **ROW\*COLUMN** | 2 | 20.33333333 | 10.16666667 | 18.30 | 0.0002 |

### From the RANDOM statement

|  |  |
| --- | --- |
| **Source** | **Type III Expected Mean Square** |
| **ROW** | Var(Error) + 3 Var(ROW\*COLUMN) + 9 Var(ROW) |
| **COLUMN** | Var(Error) + 3 Var(ROW\*COLUMN) + Q(COLUMN) |
| **ROW\*COLUMN** | Var(Error) + 3 Var(ROW\*COLUMN) |

Tests of Hypotheses for Mixed Model Analysis of Variance Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW** | 1 | 60.500000 | 60.500000 | 5.95 | 0.1349 |
| **COLUMN** | 2 | 9.000000 | 4.500000 | 0.44 | 0.6932 |
| **Error** | 2 | 20.333333 | 10.166667 |  |  |
| **Error: MS(ROW\*COLUMN)** | | | | | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **ROW\*COLUMN** | 2 | 20.333333 | 10.166667 | 18.30 | 0.0002 |
| **Error: MS(Error)** | 12 | 6.666667 | 0.555556 |  |  |

### From the TEST statement

Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tests of Hypotheses Using the Type III MS for ROW\*COLUMN as an Error Term** | | | | | |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 9.00000000 | 4.50000000 | 0.44 | 0.6932 |

### Mixed Model, PROC MIXED

The Mixed Procedure

|  |  |
| --- | --- |
| **Model Information** | |
| **Data Set** | WORK.RANDOM |
| **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 |
| **COLUMN** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Dimensions** | |
| **Covariance Parameters** | 3 |
| **Columns in X** | 4 |
| **Columns in Z** | 8 |
| **Subjects** | 1 |
| **Max Obs per Subject** | 18 |

|  |  |
| --- | --- |
| **Number of Observations** | |
| **Number of Observations Read** | 18 |
| **Number of Observations Used** | 18 |
| **Number of Observations Not Used** | 0 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Iteration History** | | | |
| **Iteration** | **Evaluations** | **-2 Res Log Like** | **Criterion** |
| **0** | 1 | 74.39726329 |  |
| **1** | 1 | 49.63086658 | 0.00000000 |

Convergence criteria met.

|  |  |
| --- | --- |
| **Covariance Parameter Estimates** | |
| **Cov Parm** | **Estimate** |
| **ROW** | 5.5926 |
| **ROW\*COLUMN** | 3.2037 |
| **Residual** | 0.5556 |

|  |  |
| --- | --- |
| **Fit Statistics** | |
| **-2 Res Log Likelihood** | 49.6 |
| **AIC (Smaller is Better)** | 55.6 |
| **AICC (Smaller is Better)** | 57.8 |
| **BIC (Smaller is Better)** | 51.7 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Type 3 Tests of Fixed Effects** | | | | |
| **Effect** | **Num DF** | **Den DF** | **F Value** | **Pr > F** |
| **COLUMN** | 2 | 2 | 0.44 | 0.6932 |

## NESTED ANOVA

IF SUBGP=1 THEN K=0.9\*RANUNI(SEED);

ELSE IF SUBGP=2 THEN K=1.7\*RANUNI(SEED); ELSE IF SUBGP=3 THEN K=1.3\*RANUNI(SEED); Y=K+RANNOR(SEED);

OUTPUT; END;

END; END;;

PROC GLM DATA=NESTED;

CLASS MAINGP SUBGP SUBSUBGP;

MODEL Y=MAINGP SUBGP(MAINGP) SUBSUBGP(MAINGP SUBGP); RANDOM MAINGP SUBGP(MAINGP) SUBSUBGP(MAINGP SUBGP); TEST H=MAINGP E=SUBGP(MAINGP);

TEST H=SUBGP(MAINGP) E=SUBSUBGP(MAINGP SUBGP); PROC SORT DATA=NESTED;

BY MAINGP SUBGP SUBSUBGP; PROC NESTED DATA=NESTED;

CLASS MAINGP SUBGP SUBSUBGP; VAR Y;

PROC VARCOMP DATA=NESTED;

CLASS MAINGP SUBGP SUBSUBGP;

MODEL Y=MAINGP SUBGP(MAINGP) SUBSUBGP(MAINGP SUBGP) / FIXED=1;

RUN;

**/\*Nested ANOVA Example – Input\*/**

DATA NESTED;

RETAIN SEED 12345; DO MAINGP=1 TO 3;

DO SUBGP=1 TO 3;

DO SUBSUBGP=1 TO 3;

The GLM Procedure **PROC GLM - Nested ANOVA - Output**

Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 26 | 19.49247883 | 0.74971072 | . | . |
| **Error** | 0 | 0.00000000 | . |  |  |
| **Corrected Total** | 26 | 19.49247883 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 1.000000 | . | . | 0.569242 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type I SS** | **Mean Square** | **F Value** | **Pr > F** |
| **MAINGP** | 2 | 0.54620795 | 0.27310397 | . | . |
| **SUBGP(MAINGP)** | 6 | 5.65238884 | 0.94206481 | . | . |
| **SUBSUB(MAINGP\*SUBGP)** | 18 | 13.29388204 | 0.73854900 | . | . |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **MAINGP** | 2 | 0.54620795 | 0.27310397 | . | . |
| **SUBGP(MAINGP)** | 6 | 5.65238884 | 0.94206481 | . | . |
| **SUBSUB(MAINGP\*SUBGP)** | 18 | 13.29388204 | 0.73854900 | . | . |

|  |  |
| --- | --- |
| **Source** | **Type III Expected Mean Square** |
| **MAINGP** | Var(Error) + Var(SUBSUB(MAINGP\*SUBGP)) + 3 Var(SUBGP(MAINGP)) + 9 Var(MAINGP) |
| **SUBGP(MAINGP)** | Var(Error) + Var(SUBSUB(MAINGP\*SUBGP)) + 3 Var(SUBGP(MAINGP)) |
| **SUBSUB(MAINGP\*SUBGP)** | Var(Error) + Var(SUBSUB(MAINGP\*SUBGP)) |

Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tests of Hypotheses Using the Type III MS for SUBGP(MAINGP) as an Error Term** | | | | | |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **MAINGP** | 2 | 0.54620795 | 0.27310397 | 0.29 | 0.7583 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Tests of Hypotheses Using the Type III MS for SUBSUB(MAINGP\*SUBGP) as an Error Term** | | | | | |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **SUBGP(MAINGP)** | 6 | 5.65238884 | 0.94206481 | 1.28 | 0.3170 |

### PROC NESTED - Nested ANOVA - Output

The NESTED Procedure

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Coefficients of Expected Mean Squares** | | | | |
| **Source** | **MAINGP** | **SUBGP** | **SUBSUBGP** | **Error** |
| **MAINGP** | 9 | 3 | 1 | 1 |
| **SUBGP** | 0 | 3 | 1 | 1 |
| **SUBSUBGP** | 0 | 0 | 1 | 1 |
| **Error** | 0 | 0 | 0 | 1 |

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Nested Random Effects Analysis of Variance for Variable Y** | | | | | | | | |
| **Variance Source** | **DF** | **Sum of Squares** | **F**  **Value** | **Pr > F** | **Error Term** | **Mean Square** | **Variance Component** | **Percent of Total** |
| **Total** | 26 | 19.492479 |  |  |  | 0.749711 | 0.806388 | 100.0000 |
| **MAINGP** | 2 | 0.546208 | 0.29 | 0.7583 | SUBGP | 0.273104 | -0.074329 | 0.0000 |
| **SUBGP** | 6 | 5.652389 | 1.28 | 0.3170 | SUBSUBGP | 0.942065 | 0.067839 | 8.4127 |
| **SUBSUBGP** | 18 | 13.293882 | . | . |  | 0.738549 | 0.738549 | 91.5873 |
| **Error** | 0 | . |  |  |  | . | 0 | 0.0000 |

|  |  |
| --- | --- |
| **Y Mean** | 0.56924222 |
| **Standard Error of Y Mean** | 0.10057317 |

### PROC VARCOMP - Nested ANOVA - Output

Variance Components Estimation Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **MAINGP** | 3 | 1 2 3 |
| **SUBGP** | 3 | 1 2 3 |
| **SUBSUBGP** | 3 | 1 2 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Number of Observations Read** | 27 | **Number of Observations Used** | 27 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **MIVQUE(0) SSQ Matrix** | | | | |
| **Source** | **SUBGP(MAINGP)** | **SUBSUB(MAINGP\*SUBGP)** | **Error** | **Y** |
| **SUBGP(MAINGP)** | 54.00000 | 18.00000 | 18.00000 | 16.95717 |
| **SUBSUB(MAINGP\*SUBGP)** | 18.00000 | 24.00000 | 24.00000 | 18.94627 |
| **Error** | 18.00000 | 24.00000 | 24.00000 | 18.94627 |

|  |  |
| --- | --- |
| **MIVQUE(0) Estimates** | |
| **Variance Component** | **Y** |
| **Var(SUBGP(MAINGP))** | 0.06784 |
| **Var(SUBSUB(MAINGP\*SUBGP))** | 0.73855 |
| **Var(Error)** | 0 |

## SIMPLE LINEAR REGRESSION

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| 20 | 32.858 |  | 35 | 59.243 |
| 20 | 36.025 | | | 35 | 67.812 |
| 20 | 28.785 | | | 35 | 74.081 |
| 20 | 36.085 | | | 35 | 73.214 |
| 20 | 34.418 | | | 35 | 69.036 |
| 30 | 58.216 | | | 40 | 84.631 |
| 30 | 65.864 | | | 40 | 83.584 |
| 30 | 61.121 | | | 40 | 86.660 |
| 30 | 63.981 | | | 50 | 108.195 |
| 30 | 64.055 | | | 50 | 105.174 |
|  | | | 50 | 108.472 |
| 50 | 103.255 |
| 50 | 102.763 |
| 50 | 103.219 |
| ; |  |

The REG Procedure **Regression Procedure Output**

DATA RATS;

INPUT AGE MASS; SQRTAGE=SQRT(AGE); LOGAGE=LOG10(AGE); LOGMASS=LOG10(MASS);

CARDS;

**/\*Regression & transformations – input\*/**

PROC REG DATA=RATS;

\*\*\* LINEAR REGRESSION WITH TRANSFORMATIONS ; RAW: MODEL MASS=AGE;

SQRT: MODEL MASS=SQRTAGE; LOG: MODEL MASS=LOGAGE; LOG2: MODEL LOGMASS=LOGAGE;

RUN;

Model: RAW Dependent Variable: MASS

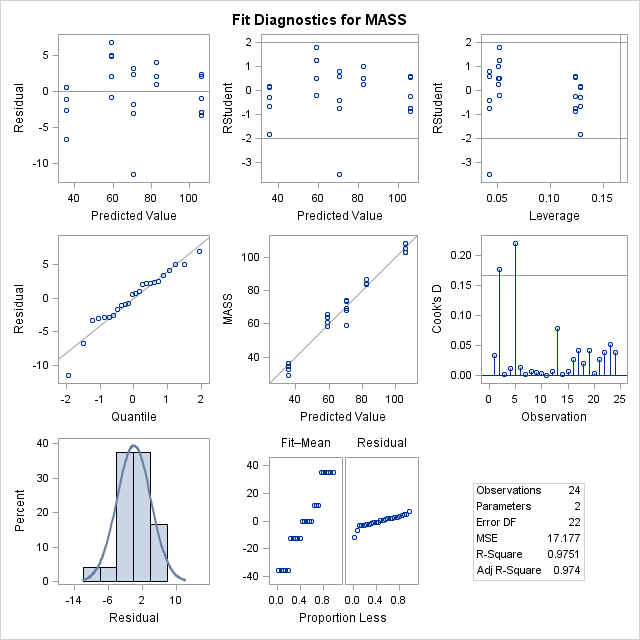
|  |  |
| --- | --- |
| **Number of Observations Read** | 24 |
| **Number of Observations Used** | 24 |

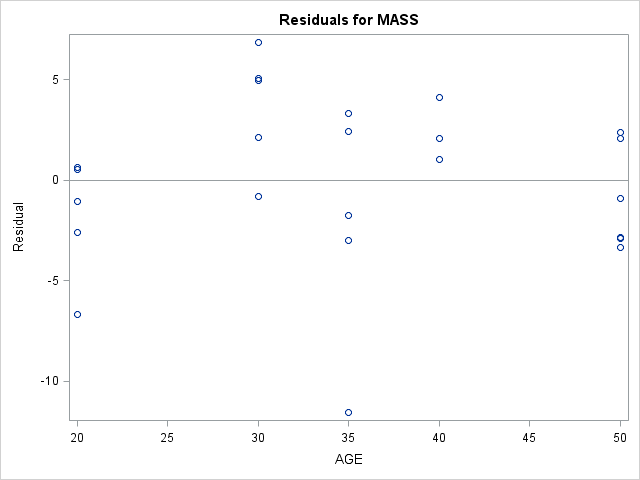
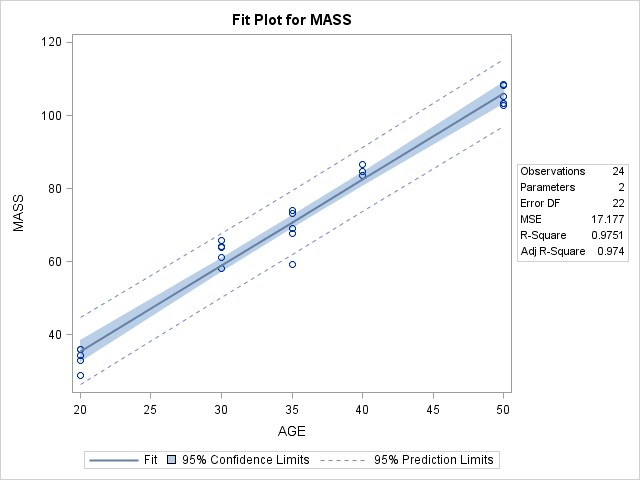
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 14821 | 14821 | 862.88 | <.0001 |
| **Error** | 22 | 377.88712 | 17.17669 |  |  |
| **Corrected Total** | 23 | 15199 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 4.14448 | **R-Square** | 0.9751 |
| **Dependent Mean** | 71.28113 | **Adj R-Sq** | 0.9740 |
| **Coeff Var** | 5.81427 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -11.61103 | 2.94596 | -3.94 | 0.0007 |
| **AGE** | **1** | 2.35433 | 0.08015 | 29.37 | <.0001 |

Dependent Variable: MASS



Model: SQRT Dependent Variable: MASS

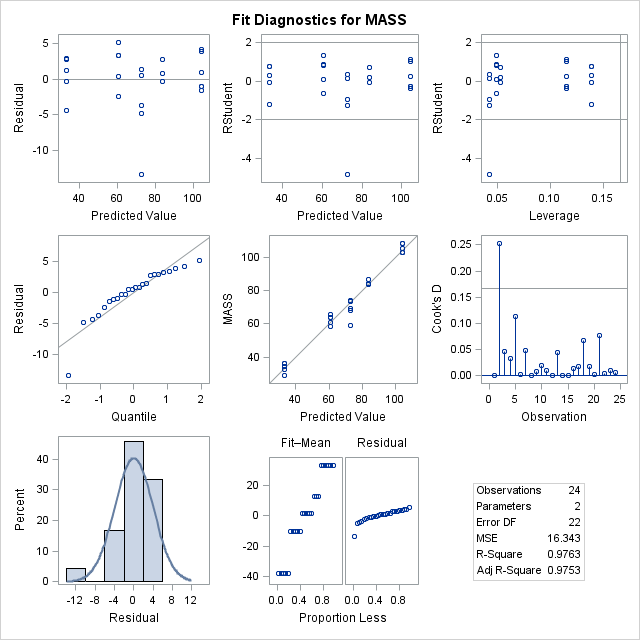
|  |  |
| --- | --- |
| **Number of Observations Read** | 24 |
| **Number of Observations Used** | 24 |

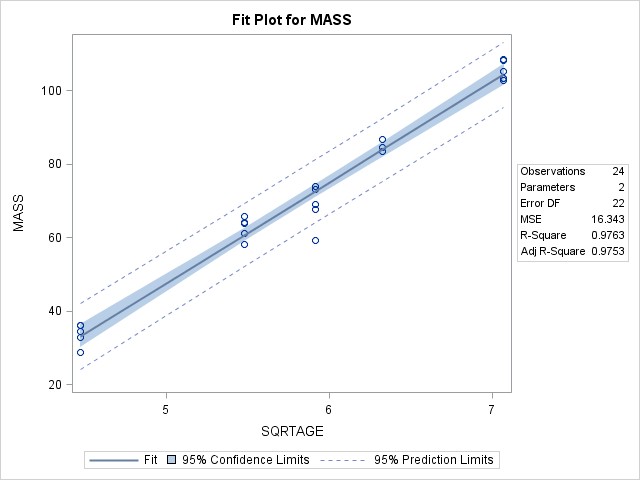
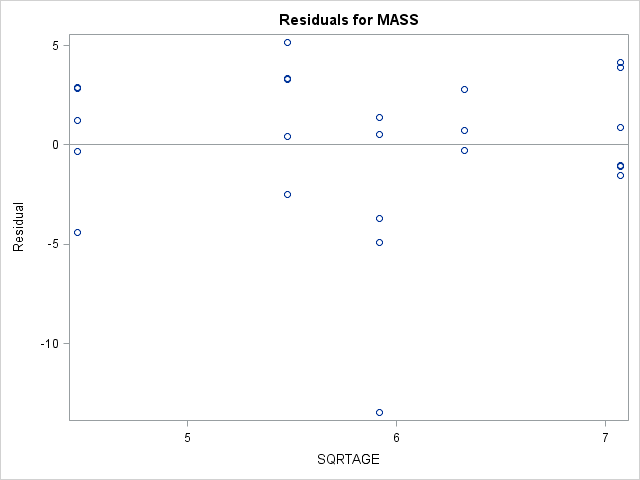
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 14840 | 14840 | 908.05 | <.0001 |
| **Error** | 22 | 359.53566 | 16.34253 |  |  |
| **Corrected Total** | 23 | 15199 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 4.04259 | **R-Square** | 0.9763 |
| **Dependent Mean** | 71.28113 | **Adj R-Sq** | 0.9753 |
| **Coeff Var** | 5.67133 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -89.13922 | 5.38717 | -16.55 | <.0001 |
| **SQRTAGE** | **1** | 27.35851 | 0.90790 | 30.13 | <.0001 |

Model: SQRT Dependent Variable: MASS





Model: LOG Dependent Variable: MASS

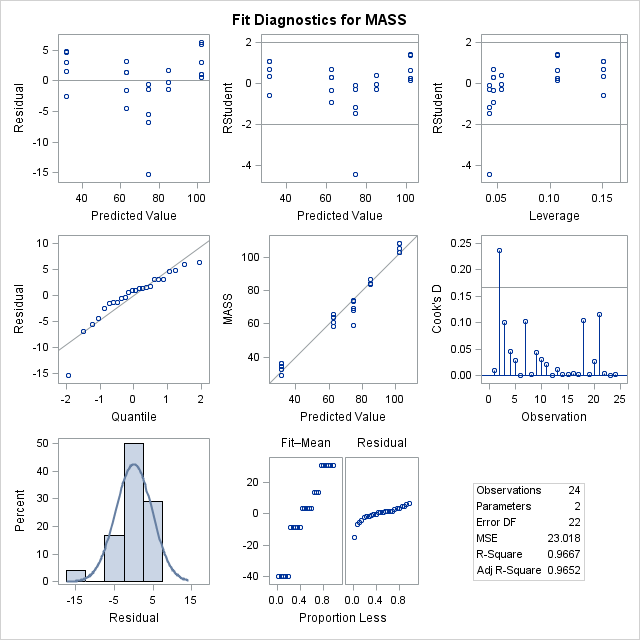
|  |  |
| --- | --- |
| **Number of Observations Read** | 24 |
| **Number of Observations Used** | 24 |

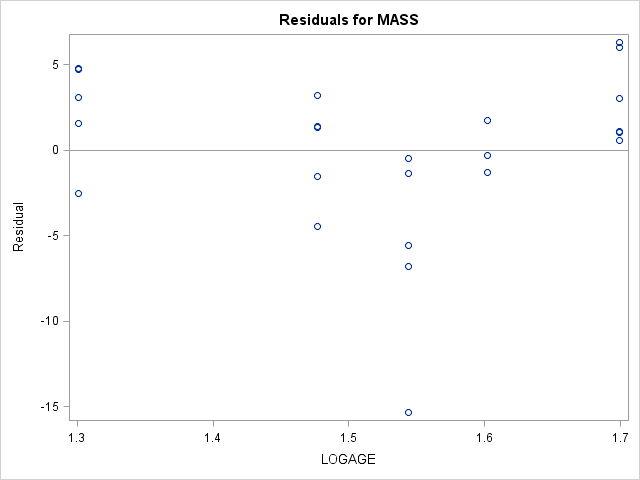
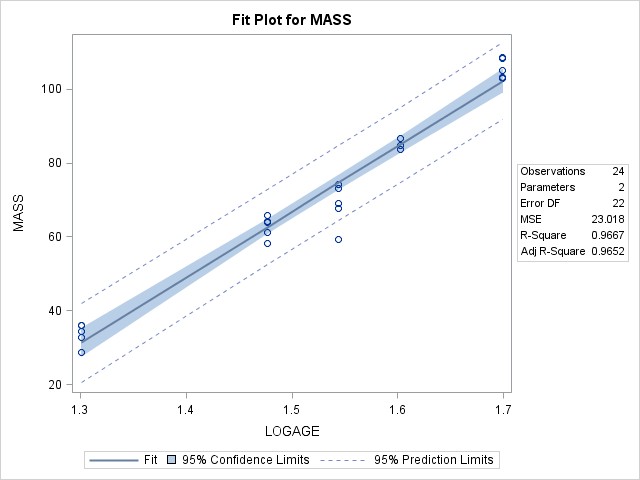
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 14693 | 14693 | 638.32 | <.0001 |
| **Error** | 22 | 506.40170 | 23.01826 |  |  |
| **Corrected Total** | 23 | 15199 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 4.79773 | **R-Square** | 0.9667 |
| **Dependent Mean** | 71.28113 | **Adj R-Sq** | 0.9652 |
| **Coeff Var** | 6.73072 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -200.32083 | 10.79468 | -18.56 | <.0001 |
| **LOGAGE** | **1** | 178.04567 | 7.04715 | 25.26 | <.0001 |

Model: LOG Dependent Variable: MASS



Model: LOG2 Dependent Variable: LOGMASS

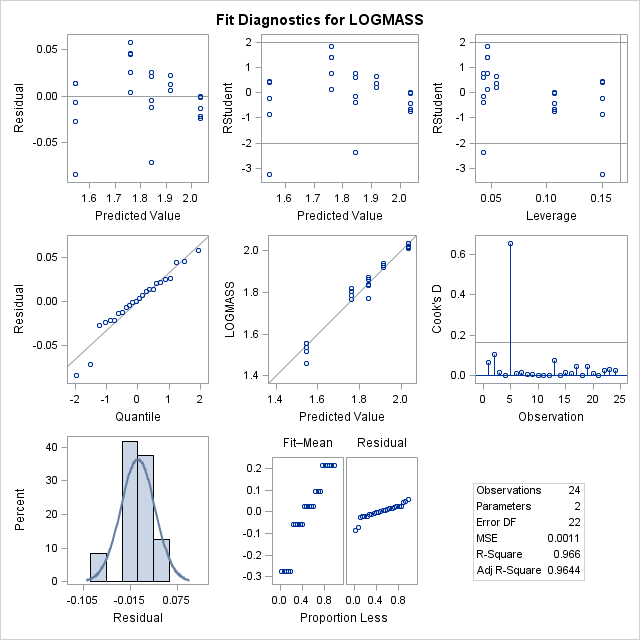
|  |  |
| --- | --- |
| **Number of Observations Read** | 24 |
| **Number of Observations Used** | 24 |

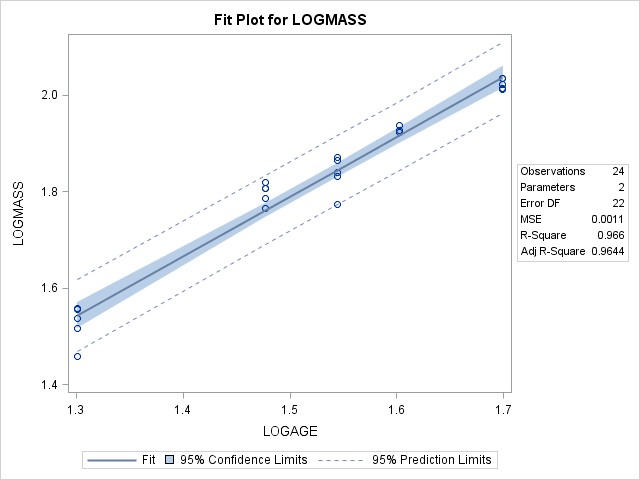
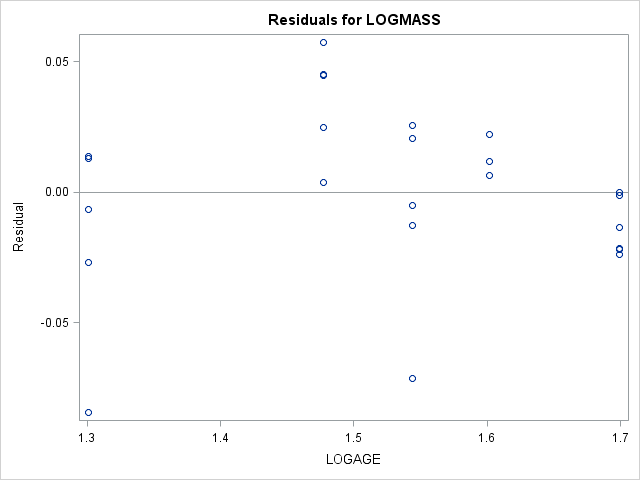
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 0.70839 | 0.70839 | 624.30 | <.0001 |
| **Error** | 22 | 0.02496 | 0.00113 |  |  |
| **Corrected Total** | 23 | 0.73335 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.03369 | **R-Square** | 0.9660 |
| **Dependent Mean** | 1.82103 | **Adj R-Sq** | 0.9644 |
| **Coeff Var** | 1.84979 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | -0.06485 | 0.07579 | -0.86 | 0.4014 |
| **LOGAGE** | **1** | 1.23627 | 0.04948 | 24.99 | <.0001 |

Model: LOG2 Dependent Variable: LOGMASS





## LINEAR & MULTIPLE REGRESSION

DATA REGRESS;

INPUT GROUP X1 X2 Y; CARDS;

**/\*Regression Input\*/**

PROC GLM DATA=REGRESS;

**/\*PROC GLM for Regression\*/**

\*\*\* LINEAR REGRESSION WITH RESIDUAL PLOT; MODEL Y=X1 / SS3 SOLUTION;

OUTPUT OUT=RESIDS PREDICTED=P RESIDUAL=R; PROC PLOT DATA=RESIDS;

PLOT R\*P;

RUN;

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 10 | 3 | 12 |  | | | | |
| 1 | 15 | 2 | 15 |  | 2 | 30 | 2 | 18 |
| 1 | 20 | 2 | 16 | | | 2 | 35 | 1 | 22 |
| 1 | 25 | 1 | 18 | | | 2 | 40 | 1 | 25 |
| 1 | 30 | 1 | 21 | | | 3 | 15 | 5 | 11 |
| 1 | 35 | 0 | 23 | | | 3 | 20 | 5 | 15 |
| 2 | 15 | 4 | 13 | | | 3 | 25 | 4 | 17 |
| 2 | 20 | 4 | 16 | | | 3 | 30 | 3 | 20 |
| 2 | 25 | 3 | 17 | | | 3 | 40 | 1 | 23 |
|  |  |  |  |  | ; |  |  |  |

The GLM Procedure

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

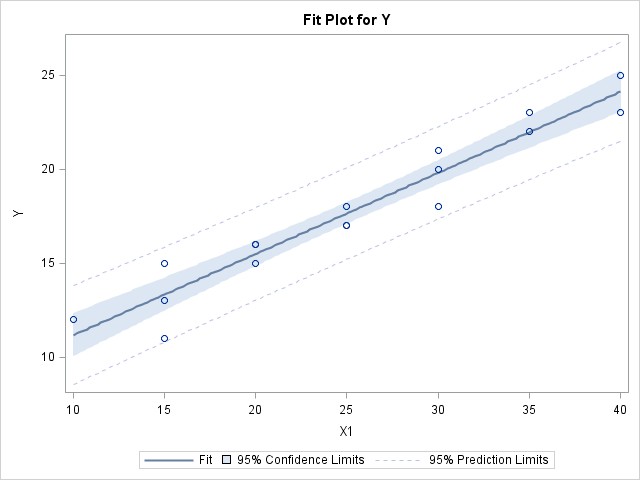
Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 246.4943791 | 246.4943791 | 199.17 | <.0001 |
| **Error** | 15 | 18.5644444 | 1.2376296 |  |  |
| **Corrected Total** | 16 | 265.0588235 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.929961 | 6.262350 | 1.112488 | 17.76471 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **X1** | 1 | 246.4943791 | 246.4943791 | 199.17 | <.0001 |

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 6.848888889 | 0.81918880 | 8.36 | <.0001 |
| **X1** | 0.431555556 | 0.03057937 | 14.11 | <.0001 |



|  |  |  |
| --- | --- | --- |
| Plot of R\*P. Legend: A = 1 obs, B = 2 obs, etc.  R ‚  ‚  2.0 ˆ  ‚  ‚  ‚ A  1.5 ˆ  ‚  ‚ A  ‚  1.0 ˆ A  ‚ A A  ‚  ‚  0.5 ˆ B  ‚ A  ‚ A  ‚  0.0 ˆ A  ‚  ‚  ‚ A  -0.5 ˆ A  ‚ B | | |
| ‚ |  |  |
| ‚ |  |  |
| -1.0 ˆ |  |  |
| ‚ |  | A |
| ‚ |  |  |
| ‚ |  |  |
| -1.5 ˆ |  |  |
| ‚ |  |  |
| ‚ |  | A |
| ‚ |  |  |
| -2.0 ˆ |  |  |
| ‚ |  |  |
| ‚ |  |  |
| ‚ | A |  |
| -2.5 ˆ |  |  |
| ‚  Šƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒƒƒƒƒƒƒƒƒƒƒƒˆƒƒ 11.164 13.322 15.480 17.638 19.796 21.953 24.111  P | | |

\*\*\* LINEAR AND MULTIPLE REGRESSION WITH RESIDUAL PLOTS; X1: MODEL Y=X1 ;

X2: MODEL Y=X2 ; X1X2: MODEL Y=X1 X2 ;

PLOT RESIDUAL.\*PREDICTED.;

RUN;

**/\*PROC REG for Regression\*/**

PROC REG DATA=REGRESS;

The REG Procedure Model: X1

Dependent Variable: Y

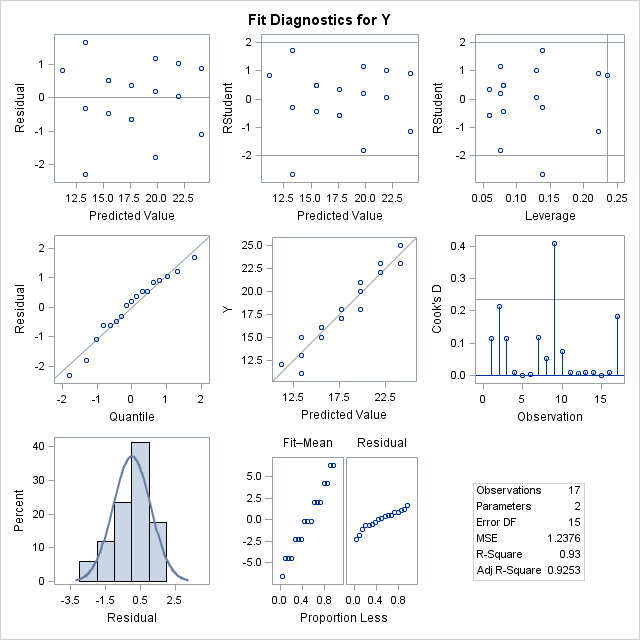
|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

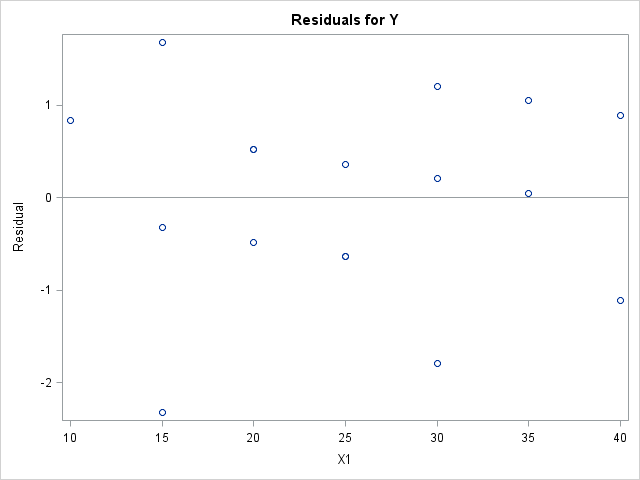
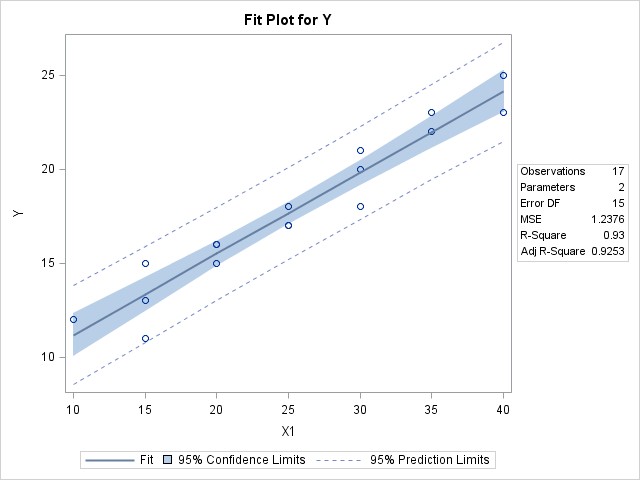
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 246.49438 | 246.49438 | 199.17 | <.0001 |
| **Error** | 15 | 18.56444 | 1.23763 |  |  |
| **Corrected Total** | 16 | 265.05882 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 1.11249 | **R-Square** | 0.9300 |
| **Dependent Mean** | 17.76471 | **Adj R-Sq** | 0.9253 |
| **Coeff Var** | 6.26235 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 6.84889 | 0.81919 | 8.36 | <.0001 |
| **X1** | **1** | 0.43156 | 0.03058 | 14.11 | <.0001 |

Model: X1 Dependent Variable: Y



The REG Procedure Model: X2

Dependent Variable: Y

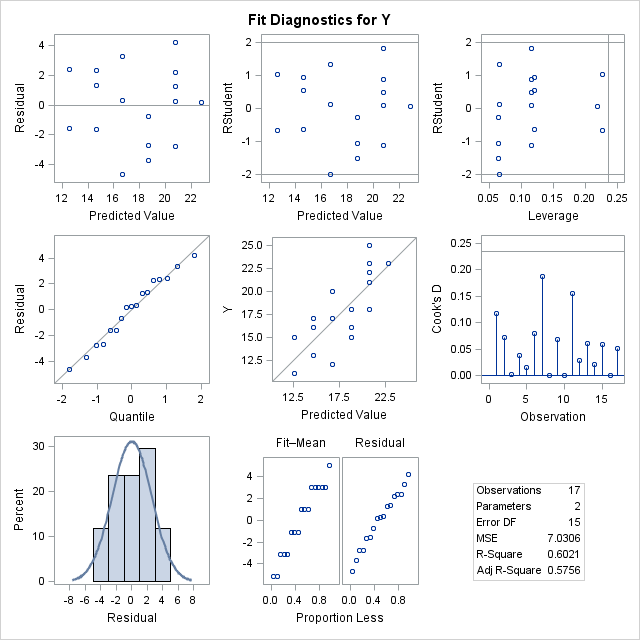
|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

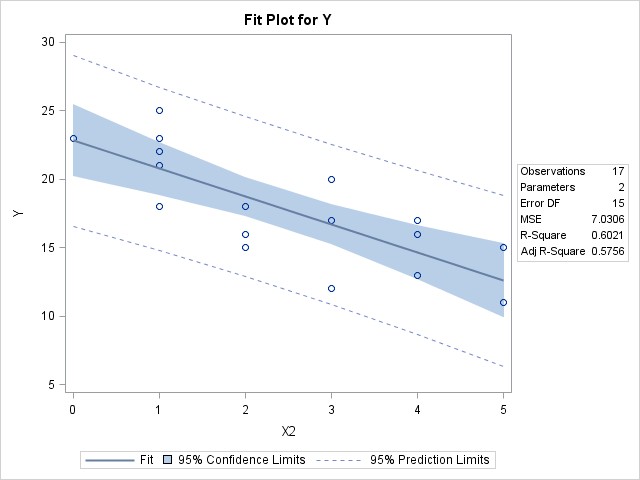
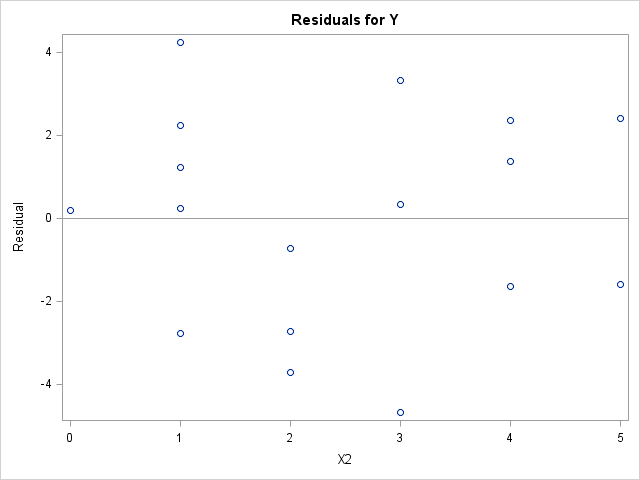
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 1 | 159.60036 | 159.60036 | 22.70 | 0.0003 |
| **Error** | 15 | 105.45846 | 7.03056 |  |  |
| **Corrected Total** | 16 | 265.05882 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 2.65152 | **R-Square** | 0.6021 |
| **Dependent Mean** | 17.76471 | **Adj R-Sq** | 0.5756 |
| **Coeff Var** | 14.92578 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 22.81231 | 1.23932 | 18.41 | <.0001 |
| **X2** | **1** | -2.04308 | 0.42881 | -4.76 | 0.0003 |

Model: X2 Dependent Variable: Y





The REG Procedure Model: X1X2

Dependent Variable: Y

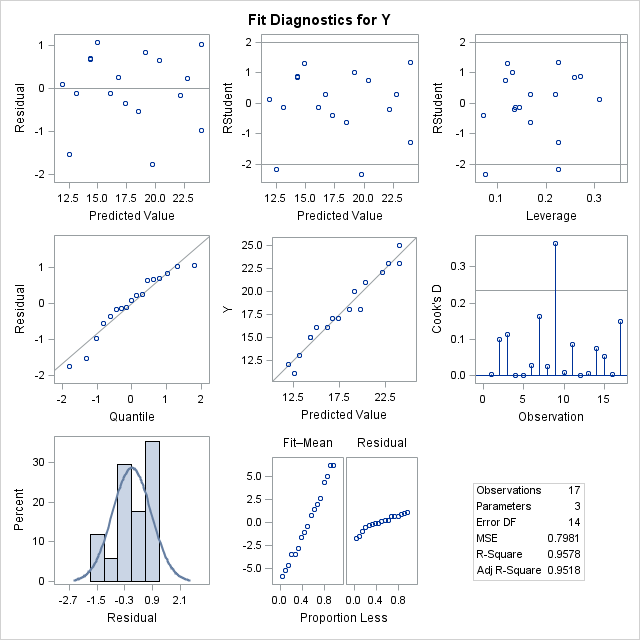
|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

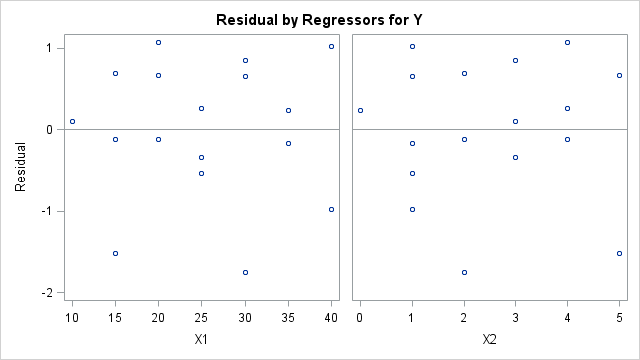
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Analysis of Variance** | | | | | |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 2 | 253.88597 | 126.94298 | 159.06 | <.0001 |
| **Error** | 14 | 11.17286 | 0.79806 |  |  |
| **Corrected Total** | 16 | 265.05882 |  |  |  |

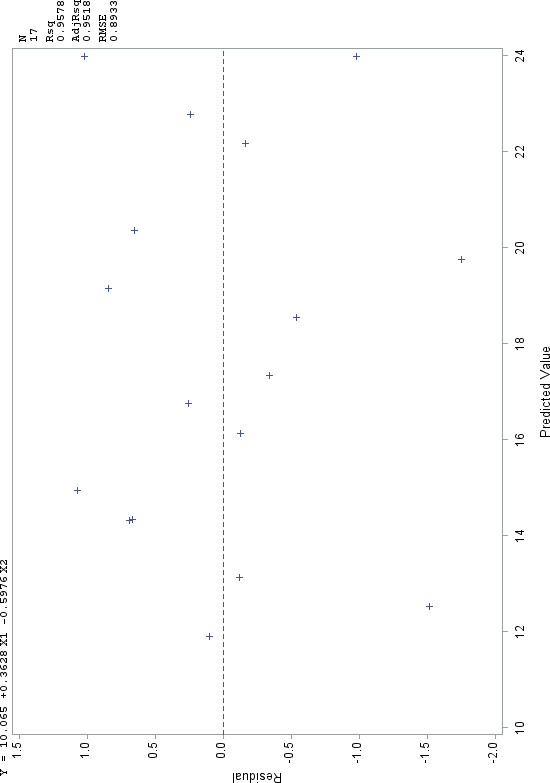
|  |  |  |  |
| --- | --- | --- | --- |
| **Root MSE** | 0.89334 | **R-Square** | 0.9578 |
| **Dependent Mean** | 17.76471 | **Adj R-Sq** | 0.9518 |
| **Coeff Var** | 5.02875 |  |  |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter Estimates** | | | | | |
| **Variable** | **DF** | **Parameter Estimate** | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | **1** | 10.06528 | 1.24486 | 8.09 | <.0001 |
| **X1** | **1** | 0.36277 | 0.03338 | 10.87 | <.0001 |
| **X2** | **1** | -0.59760 | 0.19636 | -3.04 | 0.0088 |

Model: X1X2 Dependent Variable: Y







## ANALYSIS OF COVARIANCE (ANCOVA)

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| 1 | 10 | 3 | 12 |  | | | | |
| 1 | 15 | 2 | 15 |  | 2 | 30 | 2 | 18 |
| 1 | 20 | 2 | 16 | | | 2 | 35 | 1 | 22 |
| 1 | 25 | 1 | 18 | | | 2 | 40 | 1 | 25 |
| 1 | 30 | 1 | 21 | | | 3 | 15 | 5 | 11 |
| 1 | 35 | 0 | 23 | | | 3 | 20 | 5 | 15 |
| 2 | 15 | 4 | 13 | | | 3 | 25 | 4 | 17 |
| 2 | 20 | 4 | 16 | | | 3 | 30 | 3 | 20 |
| 2 | 25 | 3 | 17 | | | 3 | 40 | 1 | 23 |
|  |  |  |  |  | ; |  |  |  |

### Test for equal slopes

DATA REGRESS;

INPUT GROUP X1 X2 Y; CARDS;

**/\*ANCOVA Input\*/**

PROC GLM DATA=REGRESS; **/\*PROC GLM for ANCOVA\*/**

\*\*\* ANCOVA STEP 1 -- TEST FOR EQUAL SLOPES; CLASS GROUP;

MODEL Y=GROUP X1 GROUP\*X1 / SS3 SOLUTION; PROC GLM DATA=REGRESS;

\*\*\* ANCOVA STEP 2 -- TEST COMMON-SLOPE MODEL; CLASS GROUP;

MODEL Y=GROUP X1 / SS3 SOLUTION;

LSMEANS GROUP / STDERR PDIFF ADJUST=TUKEY; PROC GLM DATA=REGRESS;

\*\*\* LINEAR MODEL FOR ESTIMATES OF UNEQUAL SLOPES; CLASS GROUP;

MODEL Y=GROUP X1(GROUP) / SS3 SOLUTION;

RUN;

The GLM Procedure

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **GROUP** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

Dependent Variable: Y

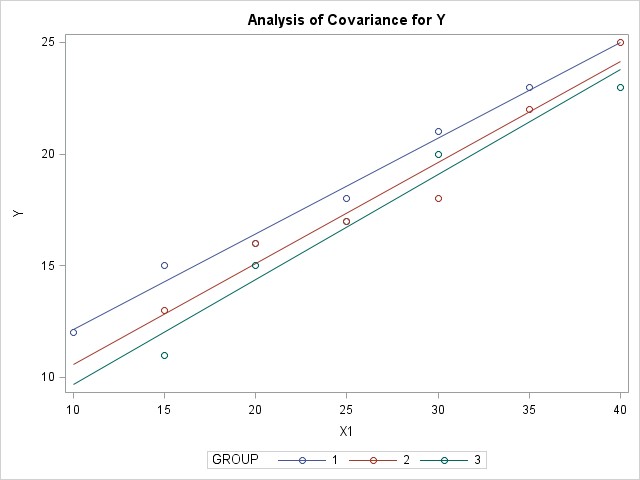
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 256.6001363 | 51.3200273 | 66.74 | <.0001 |
| **Error** | 11 | 8.4586873 | 0.7689716 |  |  |
| **Corrected Total** | 16 | 265.0588235 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.968088 | 4.936250 | 0.876910 | 17.76471 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **GROUP** | 2 | 2.6306393 | 1.3153197 | 1.71 | 0.2255 |
| **X1** | 1 | 250.6457107 | 250.6457107 | 325.95 | <.0001 |
| **X1\*GROUP** | 2 | 0.3525577 | 0.1762789 | 0.23 | 0.7988 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** |  | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 4.972972973 | B | 1.24848955 | 3.98 | 0.0021 |
| **GROUP 1** | 2.884169884 | B | 1.60520966 | 1.80 | 0.0999 |
| **GROUP 2** | 1.112741313 | B | 1.73669524 | 0.64 | 0.5348 |
| **GROUP 3** | 0.000000000 | B | . | . | . |
| **X1** | 0.470270270 | B | 0.04558839 | 10.32 | <.0001 |
| **X1\*GROUP 1** | -0.041698842 | B | 0.06193505 | -0.67 | 0.5147 |
| **X1\*GROUP 2** | -0.018841699 | B | 0.06193505 | -0.30 | 0.7666 |
| **X1\*GROUP 3** | 0.000000000 | B | . | . | . |

|  |  |
| --- | --- |
| Note: | The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable. |



The GLM Procedure T**est common-slope model**

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **GROUP** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

Dependent Variable: Y

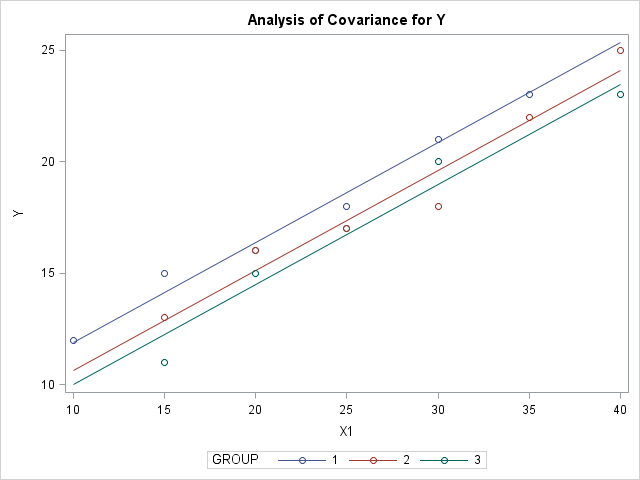
|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 3 | 256.2475785 | 85.4158595 | 126.02 | <.0001 |
| **Error** | 13 | 8.8112450 | 0.6777881 |  |  |
| **Corrected Total** | 16 | 265.0588235 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.966757 | 4.634351 | 0.823279 | 17.76471 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **GROUP** | 2 | 9.7531995 | 4.8765997 | 7.19 | 0.0079 |
| **X1** | 1 | 250.9887550 | 250.9887550 | 370.31 | <.0001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** |  | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 5.526104418 | B | 0.70963194 | 7.79 | <.0001 |
| **GROUP 1** | 1.871485944 | B | 0.50516462 | 3.70 | 0.0026 |
| **GROUP 2** | 0.626506024 | B | 0.49974715 | 1.25 | 0.2320 |
| **GROUP 3** | 0.000000000 | B | . | . | . |
| **X1** | 0.448995984 |  | 0.02333255 | 19.24 | <.0001 |

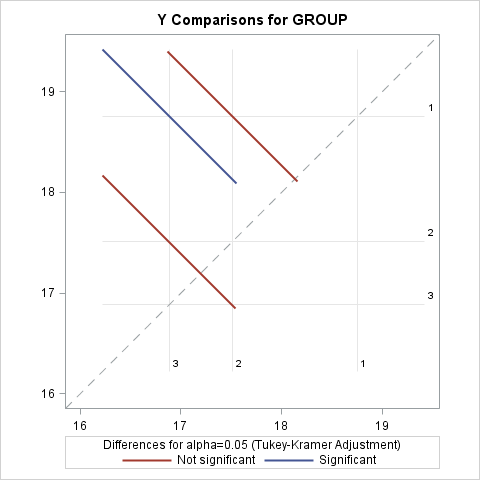
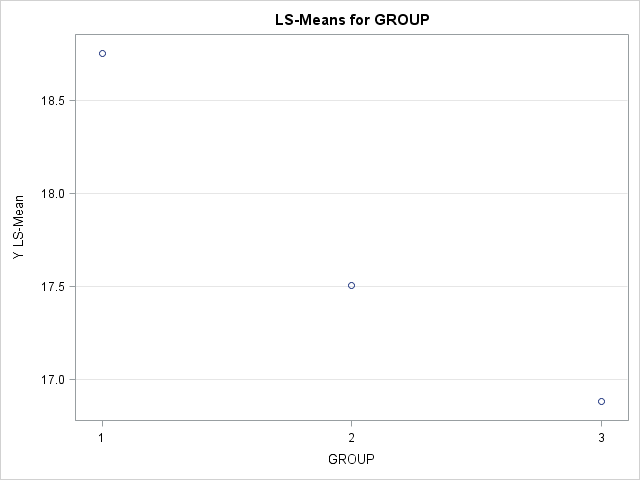
|  |  |
| --- | --- |
| Note: | The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable. |

The GLM Procedure Least Squares Means

Adjustment for Multiple Comparisons: Tukey-Kramer

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **GROUP** | **Y LSMEAN** | **Standard**  **Error** | **Pr > |t|** | **LSMEAN Number** |
| **1** | 18.7545476 | 0.3423667 | <.0001 | 1 |
| **2** | 17.5095677 | 0.3400202 | <.0001 | 2 |
| **3** | 16.8830617 | 0.3685497 | <.0001 | 3 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Least Squares Means for effect GROUP Pr > |t| for H0: LSMean(i)=LSMean(j) Dependent Variable: Y** | | | |
| **i/j** | **1** | **2** | **3** |
| **1** |  | 0.0595 | 0.0069 |
| **2** | 0.0595 |  | 0.4445 |
| **3** | 0.0069 | 0.4445 |  |



The GLM Procedure **Linear model for estimates**

### of unequal slopes

|  |  |  |
| --- | --- | --- |
| **Class Level Information** | | |
| **Class** | **Levels** | **Values** |
| **GROUP** | 3 | 1 2 3 |

|  |  |
| --- | --- |
| **Number of Observations Read** | 17 |
| **Number of Observations Used** | 17 |

Dependent Variable: Y

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Sum of Squares** | **Mean Square** | **F Value** | **Pr > F** |
| **Model** | 5 | 256.6001363 | 51.3200273 | 66.74 | <.0001 |
| **Error** | 11 | 8.4586873 | 0.7689716 |  |  |
| **Corrected Total** | 16 | 265.0588235 |  |  |  |

|  |  |  |  |
| --- | --- | --- | --- |
| **R-Square** | **Coeff Var** | **Root MSE** | **Y Mean** |
| 0.968088 | 4.936250 | 0.876910 | 17.76471 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Source** | **DF** | **Type III SS** | **Mean Square** | **F Value** | **Pr > F** |
| **GROUP** | 2 | 2.6306393 | 1.3153197 | 1.71 | 0.2255 |
| **X1(GROUP)** | 3 | 251.3413127 | 83.7804376 | 108.95 | <.0001 |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Parameter** | **Estimate** |  | **Standard**  **Error** | **t Value** | **Pr > |t|** |
| **Intercept** | 4.972972973 | B | 1.24848955 | 3.98 | 0.0021 |
| **GROUP 1** | 2.884169884 | B | 1.60520966 | 1.80 | 0.0999 |
| **GROUP 2** | 1.112741313 | B | 1.73669524 | 0.64 | 0.5348 |
| **GROUP 3** | 0.000000000 | B | . | . | . |
| **X1(GROUP) 1** | 0.428571429 |  | 0.04192433 | 10.22 | <.0001 |
| **X1(GROUP) 2** | 0.451428571 |  | 0.04192433 | 10.77 | <.0001 |
| **X1(GROUP) 3** | 0.470270270 |  | 0.04558839 | 10.32 | <.0001 |

|  |  |
| --- | --- |
| Note: | The X'X matrix has been found to be singular, and a generalized inverse was used to solve the normal equations. Terms whose estimates are followed by the letter 'B' are not uniquely estimable. |