SAS Tutorial

Contents

1. SAS Windows Environment

Upon entering into a SAS session, you will see a screen similar to Figure 1. The window at the bottom of the screen is the SAS Editor window. The SAS program commands for creating and analyzing data are specified in this window. The window at the top of the screen is the SAS Log window, which logs whether or not each command line has been successfully executed. Once a program is run, a third window appears—the SAS Output window. This window will show the results of the analysis. The SAS printouts shown throughout this text appear in the SAS Output window.

2. Creating a SAS Data Set Ready for Analysis

In the SAS Editor window, three basic types of instructions are utilized:

1. *DATA entry commands*: instructions on how the data will be entered
2. *INPUT data values*: the values of the variables in the data set
3. *Statistical procedural (PROC) commands*: instructions on what type of analysis is to be conducted on the data

The commands shown in Figure 2 create and print a SAS data set, called FUEL, that contains the data for Exercise 1.2 in Chapter 1. The names of the SAS variables (e.g., MODEL) are listed on the INPUT command. (*Note*: Qualitative variable names are followed by a dollar sign.) The input data values must be typed (or copied) directly into the Editor window.

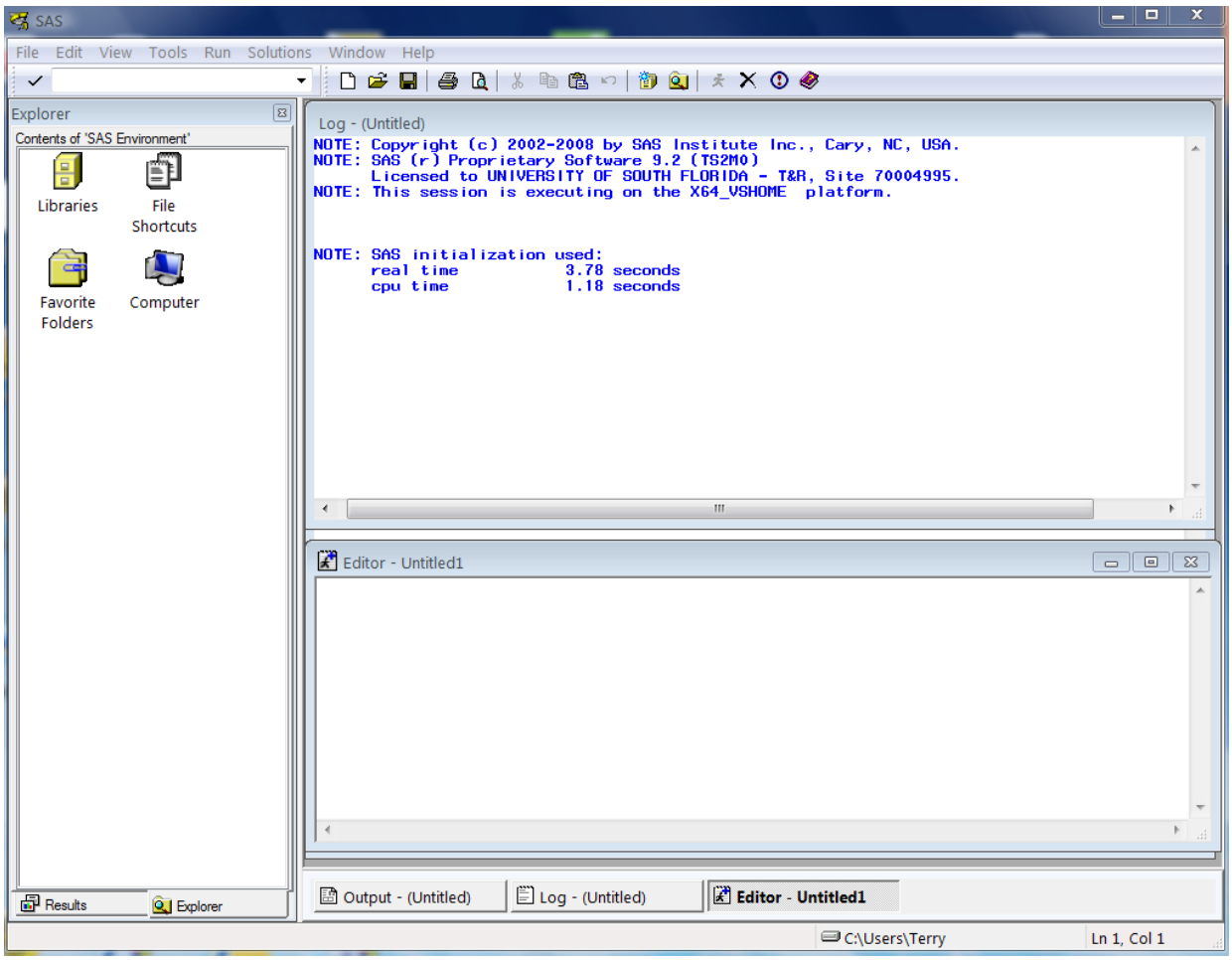


Figure 1 Initial screen viewed by the SAS user

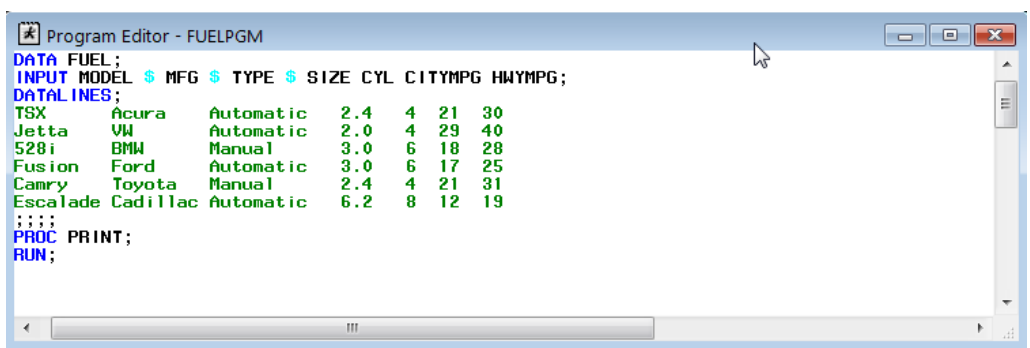
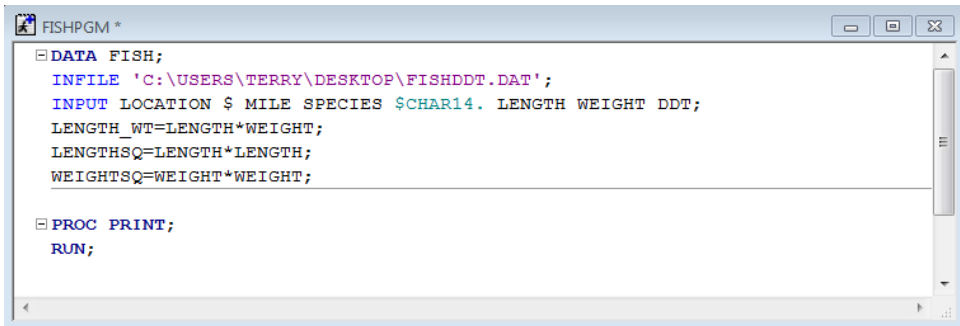


Figure 2 SAS Editor commands for creating and listing a data set

If the data are saved in an external data file (e.g., the FISHDDT data set in Example 1.1, in Chapter 1), you can access it using the INFILE command. This is

illustrated in Figure 3. (*Note*: The program in Figure 3 also shows how to create interaction and squared terms using the standard symbol, *, for multiplication.)

Figure 3 SAS Editor commands for accessing and listing a data set from an external file



To submit the SAS program (and obtain results in the Output window), you must click on the Run button shown on the menu bar at the top of the SAS screen (see Figure 1).

3. Using SAS Analyst

For SAS users who are not familiar with SAS procedure commands, SAS has available a "user-friendly" menu interface called SAS Analyst. In SAS Analyst, you do not need to know any SAS commands. You obtain results by simply clicking on the appropriate menu options.

Enter into an Analyst session by clicking on the Solutions button on the SAS menu bar, then click on Analysis, then on Analyst (see Figure 4). The resulting screen appears as in Figure 5.

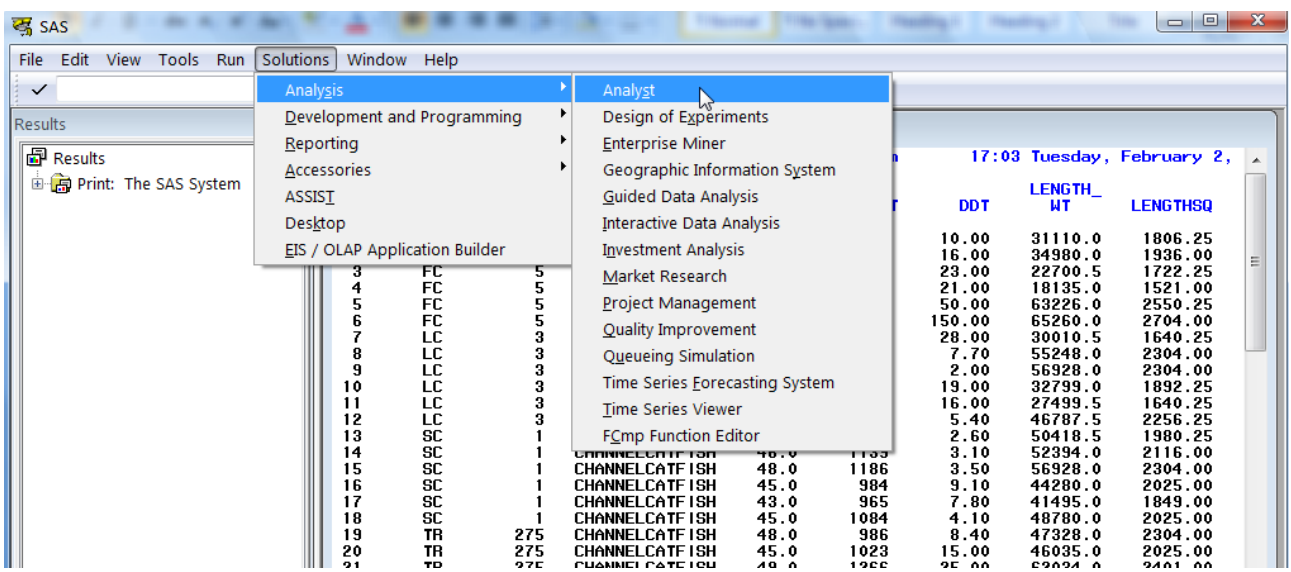


Figure 4 Entering into a SAS Analyst session

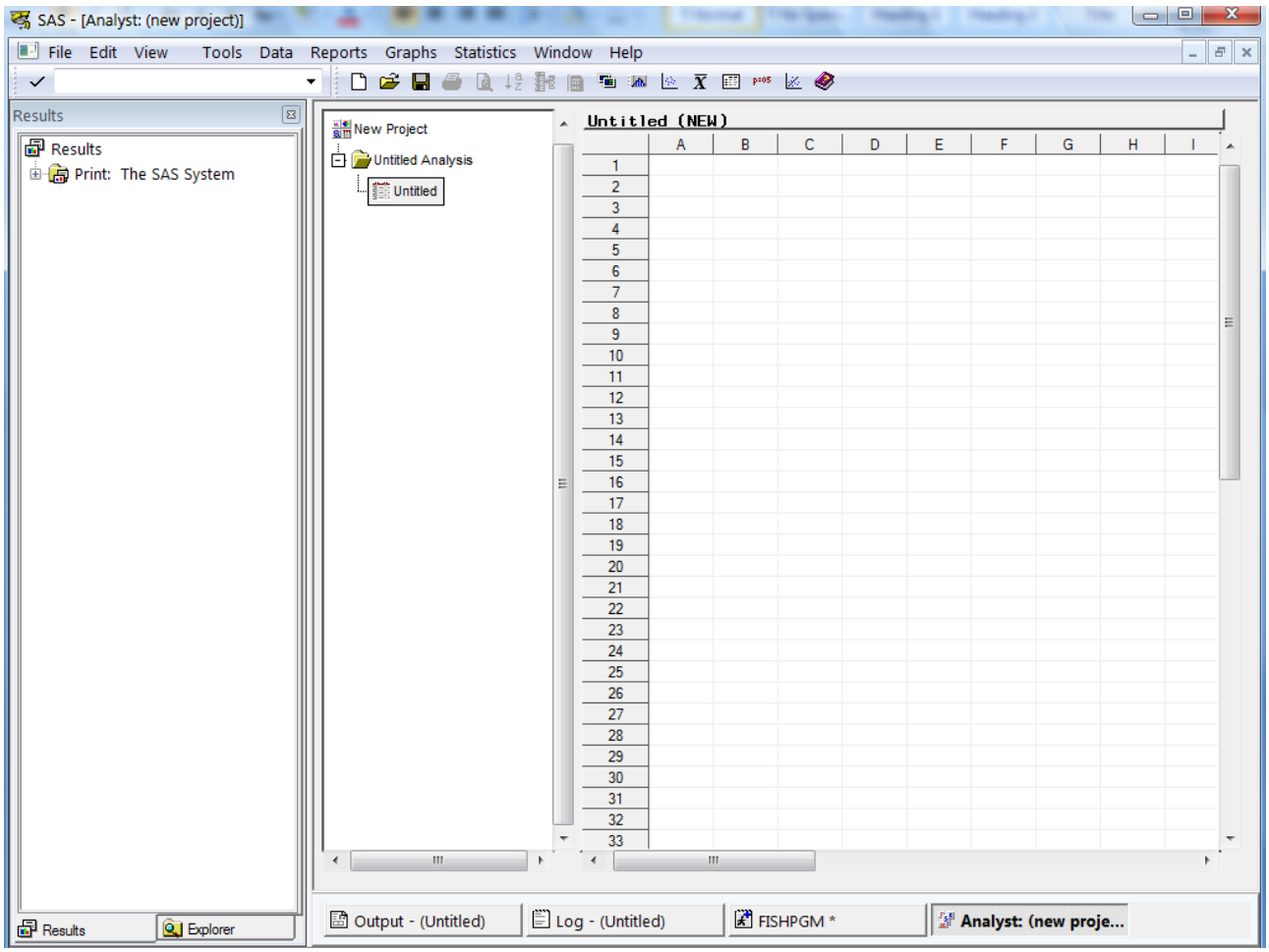


Figure 5 Initial screen in a SAS Analyst session

Notice that Figure 5 shows an empty "data table." To put the data you want to analyze with SAS Analyst in the table, you must click on the File button on the SAS menu bar, then click on Open by SAS Name, then click on the folder that contains the SAS data set, and finally click on the name of the SAS file (e.g., FISHDDT). Now the data table will contain the values of the data set (see Figure 6).

Once you access the data in this fashion, you are ready to analyze it using the menu-driven features of SAS Analyst.

4. Listing Data

To obtain a listing (printout) of your data using SAS Analyst, click on the Reports button on the Analyst menu bar, then click on List Data (see Figure 7). The resulting menu, or dialog box, appears as in Figure 8.

Enter the names of the variables you want to print in the Print box (you can do this by simply clicking on the variables), then click OK. The printout will show up on your screen.

Figure 6 Data table in SAS Analyst

Fishddt (Browse)

	RIVER	MILE	SPECIES	LENGTH	WEIGHT	DDT
1	FCM	5	CHANNELC	42.5	732	10
2	FCM	5	CHANNELC	44	795	16
3	FCM	5	CHANNELC	41.5	547	23
4	FCM	5	CHANNELC	39	465	21
5	FCM	5	CHANNELC	50.5	1252	50
6	FCM	5	CHANNELC	52	1255	150
7	LCM	3	CHANNELC	40.5	741	28
8	LCM	3	CHANNELC	48	1151	7.7
9	LCM	3	CHANNELC	48	1186	2
10	LCM	3	CHANNELC	43.5	754	19
11	LCM	3	CHANNELC	40.5	679	16
12	LCM	3	CHANNELC	47.5	985	5.4
13	SCM	1	CHANNELC	44.5	1133	2.6
14	SCM	1	CHANNELC	46	1139	3.1
15	SCM	1	CHANNELC	48	1186	3.5
16	SCM	1	CHANNELC	45	984	9.1
17	SCM	1	CHANNELC	43	965	7.8
18	SCM	1	CHANNELC	45	1084	4.1
19	TRM	275	CHANNELC	48	986	8.4
20	TRM	275	CHANNELC	45	1023	15
21	TRM	275	CHANNELC	49	1266	25
22	TRM	275	CHANNELC	50	1086	5.6
23	TRM	275	CHANNELC	46	1044	4.6
24	TRM	275	CHANNELC	52	1770	8.2
25	TRM	280	CHANNELC	48	1048	6.1
26	TRM	280	CHANNELC	51	1641	13
27	TRM	280	CHANNELC	48.5	1331	6
28	TRM	280	CHANNELC	51	1728	6.6
29	TRM	280	CHANNELC	44	917	5.5
30	TRM	280	CHANNELC	51	1398	11
31	TRM	280	SMALLMO	49	1763	4.5
32	TRM	280	SMALLMO	46	1459	4.2
33	TRM	280	SMALLMO	52	2302	3

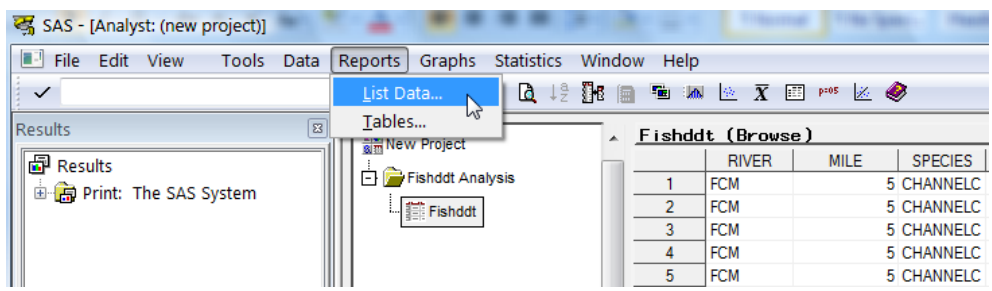
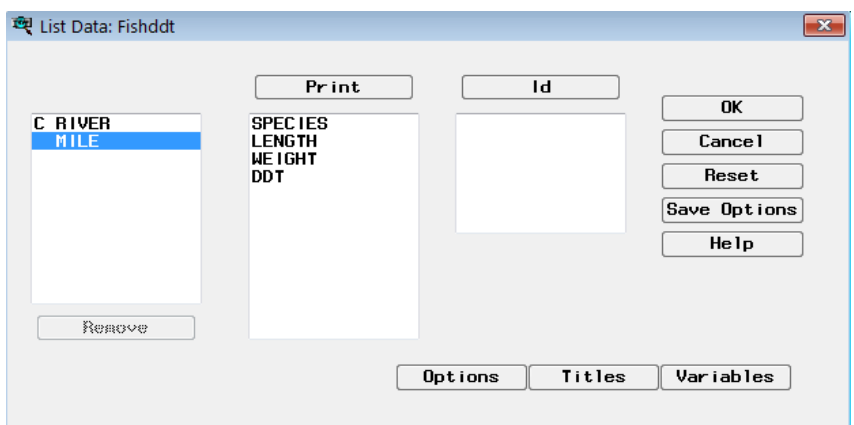


Figure 7 SAS Analyst options for list data

Figure 8 SAS Analyst List Data dialog box



5. Graphing Data

To obtain graphical descriptions of your data (e.g., bar charts, histograms, scatterplots, etc.) using SAS Analyst, click on the Graphs button on the Analyst menu bar (see Figure 9). The resulting menu list appears as shown in Figure 9. Four of the options covered in this text are Bar Chart, Histogram, (Normal) Probability Plot, and Scatterplot. Click on the graph of your choice to view the appropriate dialog box. For example, the dialog boxes for a vertical bar chart and a scatterplot are shown, respectively, in Figures 10 and 11. Make the appropriate variable selections and click OK to view the graph.

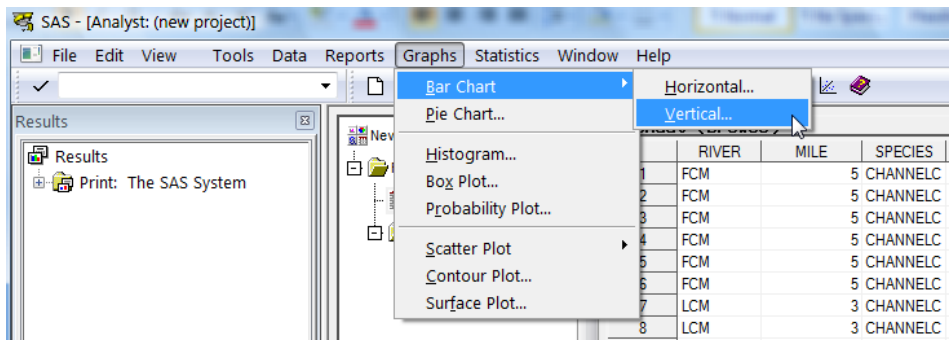


Figure 9 SAS Analyst options for graphing data

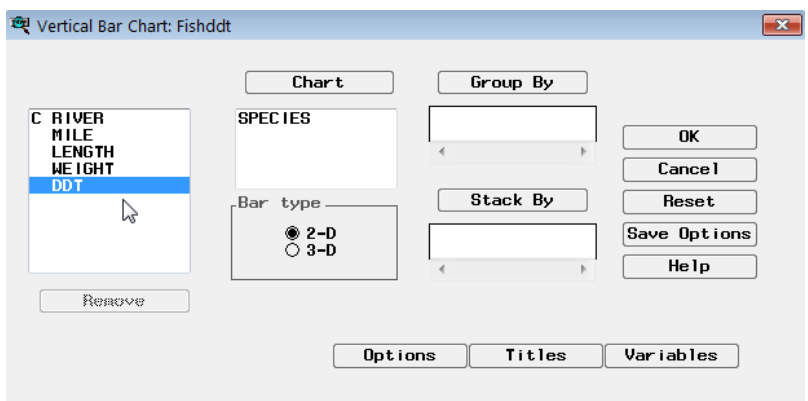


Figure 10 SAS Analyst Bar Chart dialog box

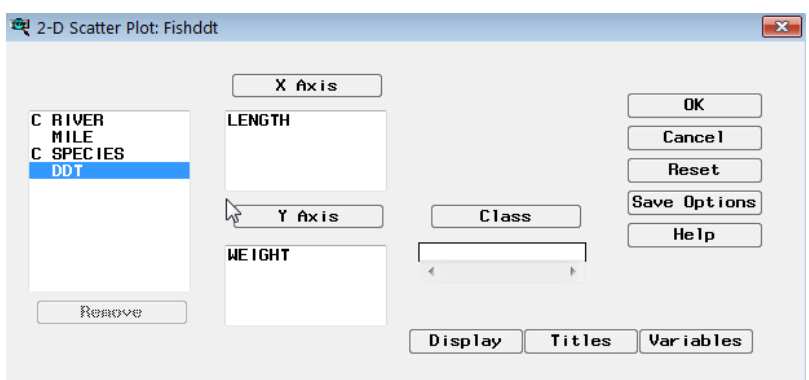


Figure 11 SAS Analyst scatterplot dialog box

6. Descriptive Statistics and Correlations

To obtain numerical descriptive measures for a quantitative variable (e.g., mean, standard deviation, etc.) using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Descriptive and finally click on Summary Statistics (see Figure 12). The resulting dialog box appears in Figure 13.

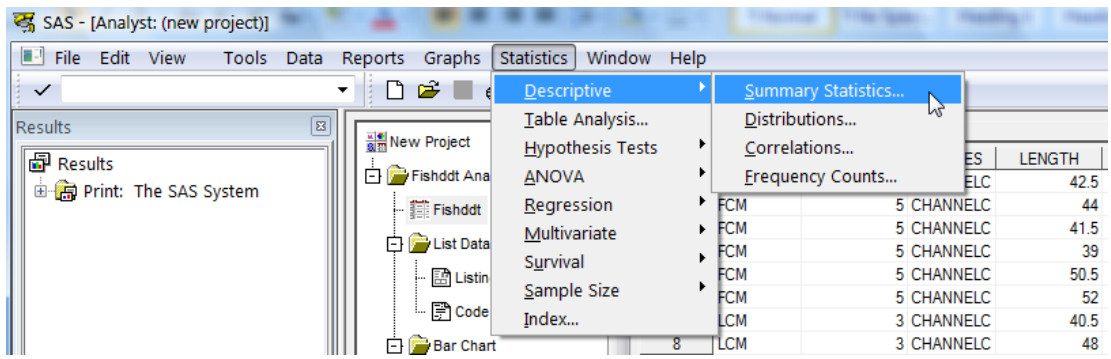
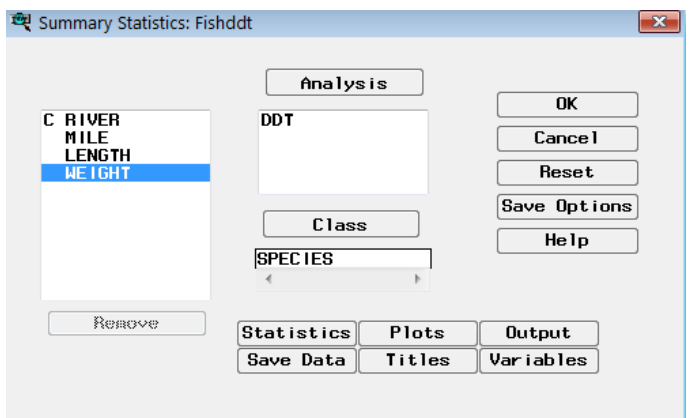


Figure 12 SAS Analyst options for descriptive statistics

Figure 13 SAS Analyst Summary Statistics dialog box



Select the quantitative variable you want to analyze and place it in the Analysis box. (As an option, you can obtain summary statistics on this quantitative variable for different levels of a qualitative variable by placing the qualitative variable in the Class box. Also, you can control which particular descriptive statistics appear by clicking the Statistics button on the dialog box and making your selections.) Click on OK to view the descriptive statistics printout.

To obtain Pearson product moment correlations for pairs of quantitative variables using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Descriptive and finally click on Correlation (see Figure 12). The resulting dialog box appears in Figure 14.

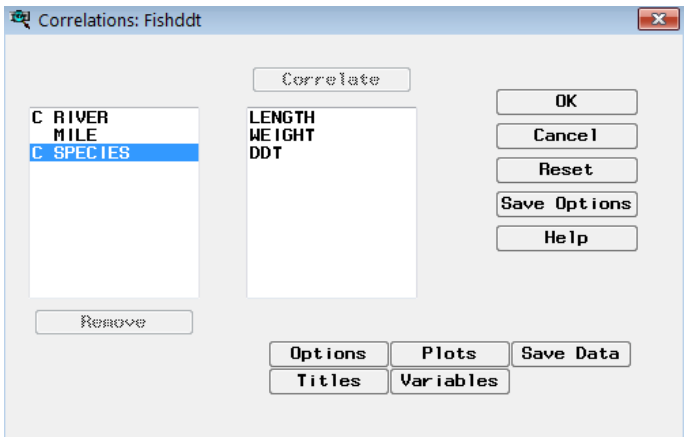


Figure 14 SAS Analyst Correlations dialog box

Enter the variables of interest in the Correlate box, then click OK to obtain a printout of the correlations.

7. Hypothesis Tests

To conduct tests of hypotheses on population parameters (e.g., mean, variance) for quantitative variables using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Hypothesis Tests. The resulting menu appears as shown in Figure 15.

Click on the test of interest to view the appropriate dialog box. For example, the dialog boxes for a One-Sample t-Test for a Mean and a Two-Sample t-Test for Means are shown, respectively, in Figures 16 and 17. Specify the quantitative variable to be tested, the null hypothesis value, and form of the alternative hypothesis in the appropriate boxes. For a two-sample test, you must also specify the qualitative variable that represents the two samples (groups). Click OK to view the test results.

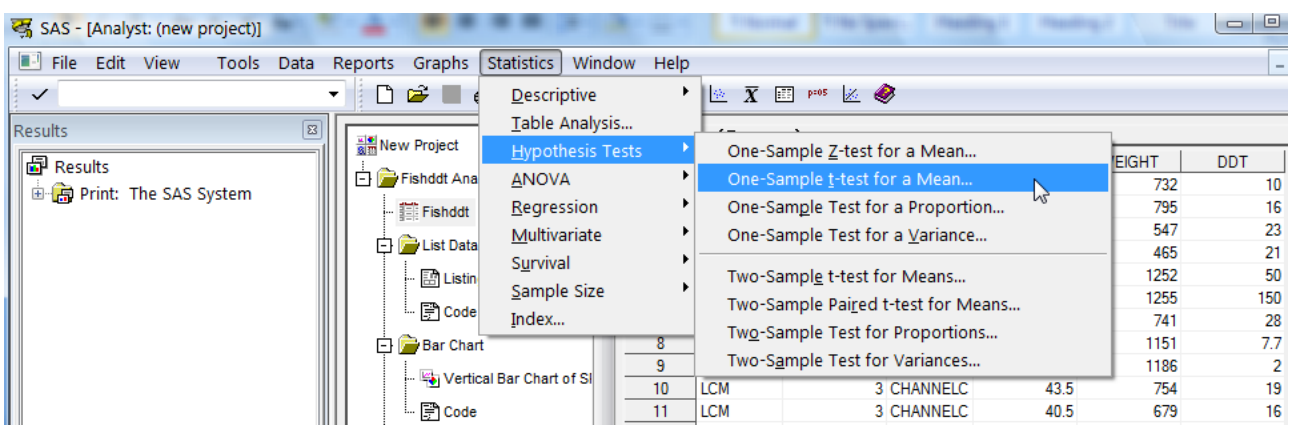


Figure 15 SAS Analyst options for Hypothesis Tests

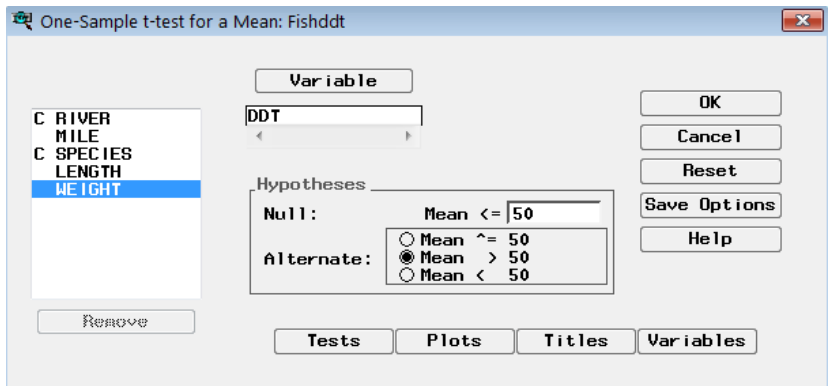


Figure 16 SAS Analyst One-Sample *t*-Test for Mean dialog box

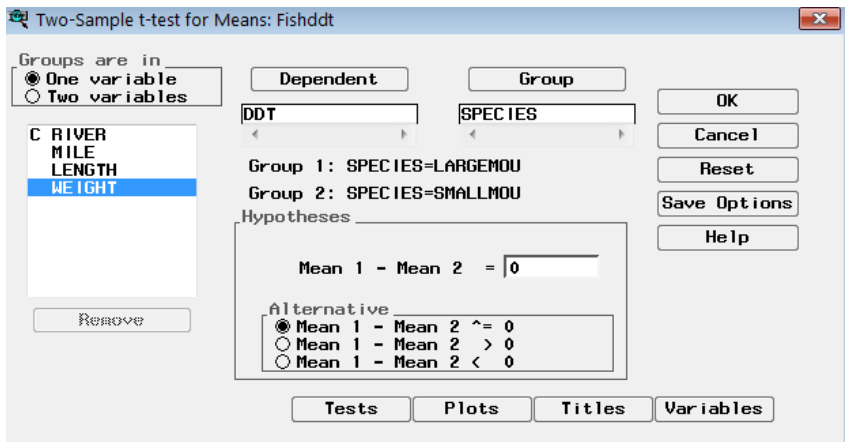


Figure 17 SAS Analyst Two-Sample *t*-Test for Means dialog box

8. Simple Linear Regression

To conduct a simple linear regression analysis using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Regression, and finally click on Simple, as shown in Figure 18. The resulting dialog box appears as shown in Figure 19.

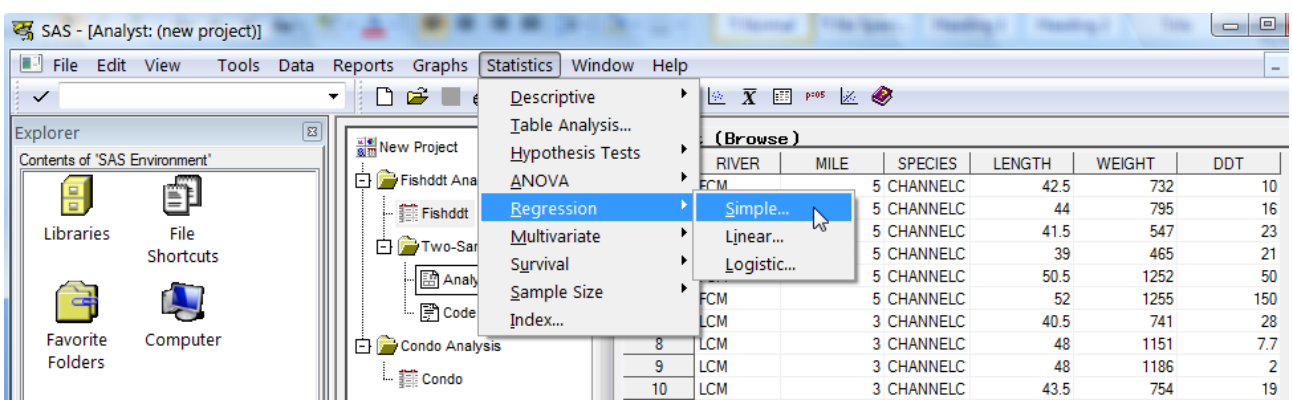


Figure 18 SAS Analyst options for Simple Linear Regression

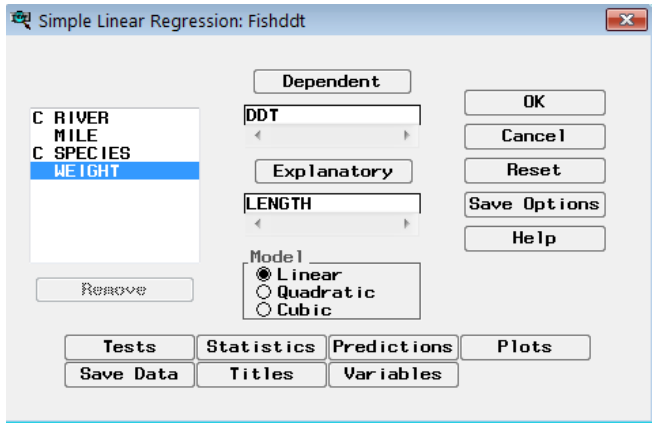


Figure 19 SAS Analyst Simple Linear Regression dialog box

Specify the quantitative dependent variable in the Dependent box and the quantitative independent variable in the Explanatory box. Be sure to select Linear in the Model box. Optionally, you can get SAS to produce confidence intervals for the model parameters by clicking the Statistics button and checking the appropriate menu item in the resulting menu list. Also, you can obtain prediction intervals and residual plots by clicking the Predictions button and Plots button, respectively, and making the appropriate selections on the resulting menus. Click OK to view the simple linear regression results.

9. Multiple Regression

To conduct a multiple regression analysis of a general linear model using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Regression, and finally click on Linear, as shown in Figure 20. The resulting dialog box appears in Figure 21.

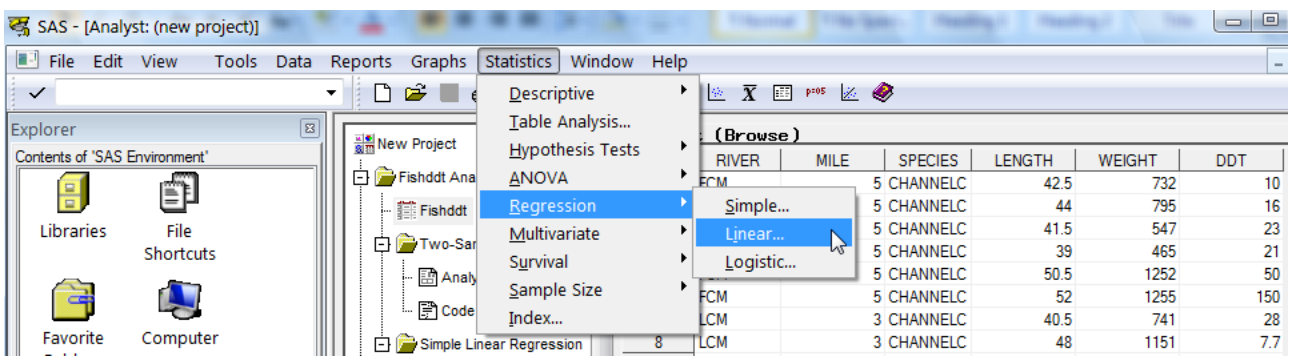


Figure 20 SAS Analyst options for multiple regression

Specify the dependent variable in the Dependent box and the independent variables in the model in the Explanatory box. Optionally, you can get SAS to produce confidence intervals for the model parameters by clicking the Statistics button and checking the appropriate menu item in the resulting menu list. To produce variance inflation factors, click the Statistics button, then click the Tests button, and

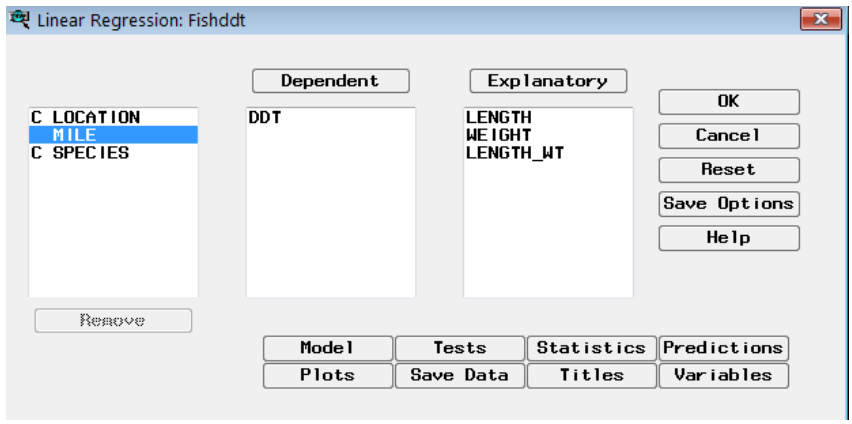


Figure 21 SAS Analyst Linear regression dialog box

then select the Variance Inflation Factors option. Also, you can obtain prediction intervals and residual plots by clicking the Predictions button and Plots button, respectively, and making the appropriate selections on the resulting menus. When all the options you desire have been checked, click OK to view the multiple regression results.

[*Note*: If your model includes interaction and/or squared terms, you must create these higher-order variables in the DATA command lines in your SAS program before entering into a SAS Analyst session. See Figure 3 for an example.]

As an alternative, you can fit general linear models using the ANOVA option available in SAS Analyst. To do this, click on the Statistics button on the Analyst menu bar, then click on ANOVA, and finally click on Linear Models, as shown in Figure 22. The resulting dialog box appears in Figure 23.

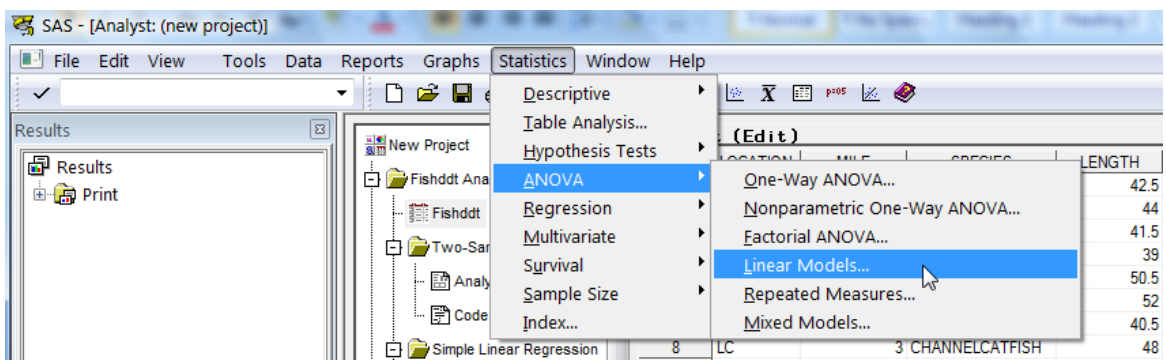


Figure 22 SAS Analyst options for general linear models

Specify the dependent variable in the Dependent box, the quantitative independent variables in the model in the Quantitative box, and the qualitative independent variables in the Class box. (*Note*: SAS will automatically create the appropriate number of dummy variables for each qualitative variable specified.) After making the variable selections, click the Model button to view the dialog box shown in Figure 24.

Specify the terms in the model using the Cross button (for interactions) and the Polynomial button (for higher-order terms). The model terms will appear in the Effects in Model box. Click OK to return to the Linear Models dialog box

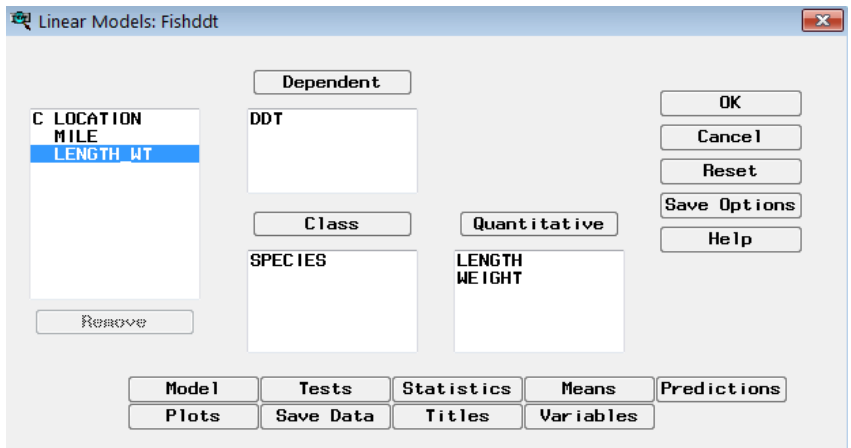


Figure 23 SAS Analyst Linear Models dialog box

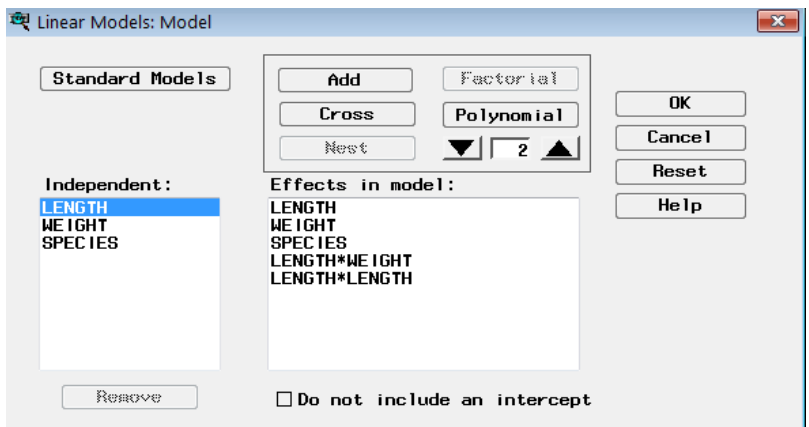


Figure 24 SAS Analyst dialog box for selecting model terms

(Figure 23). Click the Statistics button and check Calculate Parameter Estimates on the resulting menu to produce the estimates of the model parameters. Also, you can obtain prediction intervals and residual plots by clicking the Predictions button and Plots button, respectively, and by making the appropriate selections on the resulting menus. When all the options you desire have been checked, click OK to view the multiple regression results.

10. Stepwise Regression

To conduct a stepwise regression analysis using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Regression, and click on Linear, as shown in Figure 20. The resulting dialog box appears in Figure 21. Specify the dependent variable in the Dependent box and the independent variables in the stepwise model in the Explanatory box. Now click on the Model button. The resulting menu appears as shown in Figure 25.

For the stepwise regression method, choose Stepwise Selection. (The default method is Full Model.) For the all-possible-regressions-selection method, choose Mallows's Cp, R-Square, or Adjusted R-Square. As an option, you can select the value of α to use in the analysis by clicking on the Criteria button and specifying the value. (The default is $\alpha = .05$.) Click OK to view the stepwise regression results.

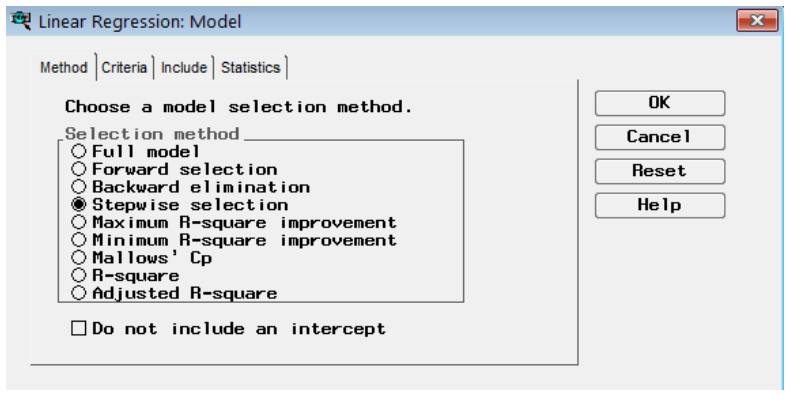


Figure 25 SAS Analyst Linear Regression dialog box for stepwise regression

11. Residual Analysis and Influence Diagnostics

To conduct a residual analysis using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Regression, and then click on Linear, as shown in Figure 20. The resulting dialog box appears in Figure 21. Specify the dependent variable in the Dependent box and the independent variables in the model in the Explanatory box. To produce residual scatterplots, click on the Plots button, then click on the Residual button. The resulting menu appears as shown in Figure 26.

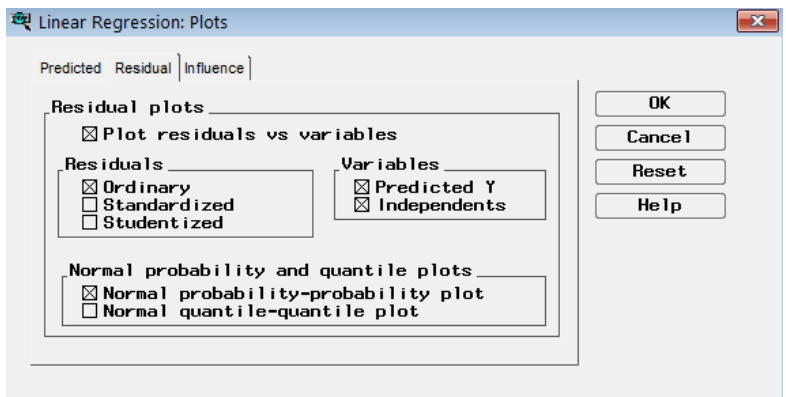


Figure 26 SAS Analyst Linear Regression dialog box for residual plots

Select the residual scatterplot you want to produce by clicking on the open box next to the type of residual (vertical axis) and variable (horizontal axis) on the menu, as shown in Figure 26. A normal probability plot of the residuals can also be produced by checking the appropriate box. Click OK to view the residual plots.

To produce influence diagnostics, click on the Save Data button in the Multiple Regression dialog box (Figure 21). The resulting menu appears as shown in Figure 27.

Click the box next to Create and Save Diagnostics Data, then select the diagnostics you want to produce by double clicking each of your choices. These diagnostics will appear in the Add box, as shown in Figure 27. Click OK to return to the Multiple Regression dialog box (Figure 21), then click OK to run the analyses.

To view the influence diagnostics, double click on the Diagnostics Table option, which now appears on the SAS Analyst main menu screen, as shown in Figure 28.

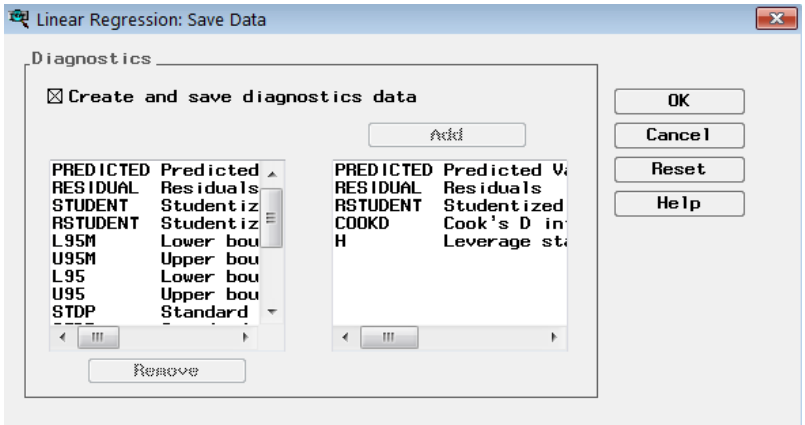


Figure 27 SAS Analyst Linear Regression dialog box for influence diagnostics

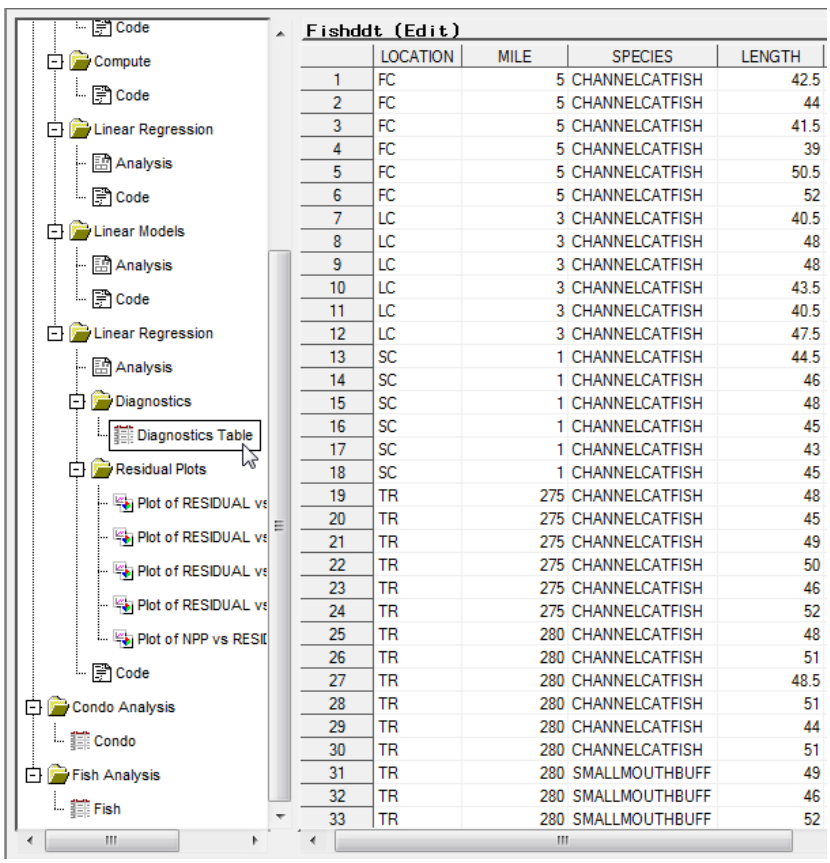


Figure 28 SAS Analyst options for listing multiple regression diagnostics

12. Logistic Regression

To conduct a logistic regression analysis for a two-level dependent qualitative variable using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on Regression, and finally click on Logistic, as shown in Figure 29. The resulting dialog box appears in Figure 30.

Specify the dependent variable in the Dependent box and the level of the dependent variable that will be modeled as π in the Model Pr{} box. Qualitative

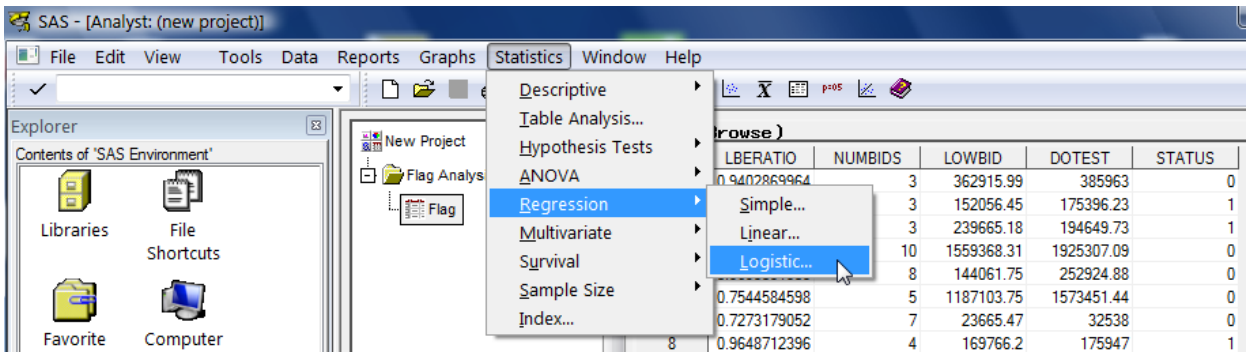
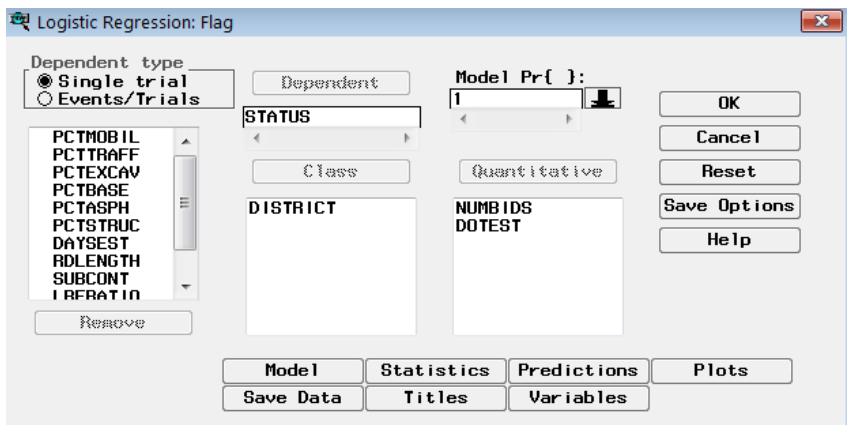


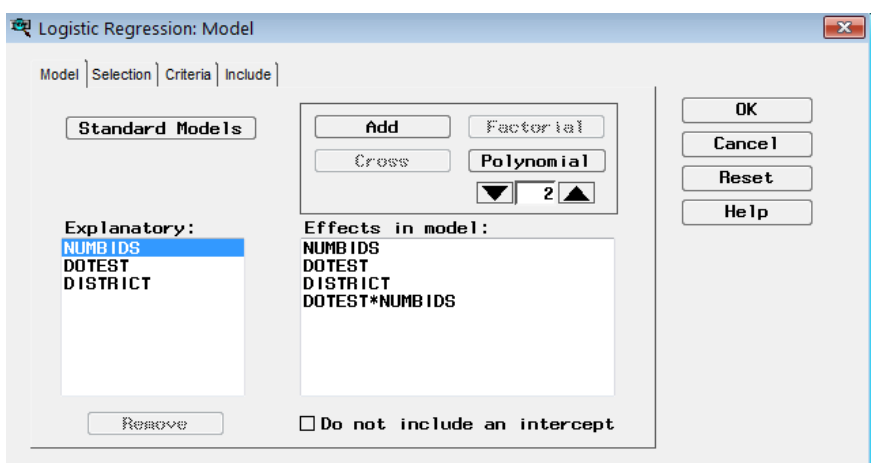
Figure 29 SAS Analyst options for Logistic Regression

Figure 30 SAS Analyst Logistic Regression dialog box



independent variables in the model are specified in the Class box, while quantitative independent variables are specified in the Quantitative box. (*Note*: SAS will automatically create the appropriate number of dummy variables for each qualitative variable specified.) After making the variable selections, click the Model button to view the dialog box shown in Figure 31.

Figure 31 SAS Analyst dialog box for selecting logit model terms



Specify the terms in the model using the Cross button (for interactions) and the Polynomial button (for higher-order terms). The model terms will appear in the Effects in Model box. Click OK to return to the Logistic Regression dialog box (Figure 30). Click the Statistics button and check Classification table on the resulting menu to produce a classification table for the analysis. Also, you can obtain prediction intervals for π by clicking the Predictions button and making the appropriate selections on the resulting menu. When all the options you desire have been checked, click OK to view the logistic regression results.

13. One-Way Analysis of Variance

To conduct a one-way ANOVA for a completely randomized design using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on ANOVA, and finally click on One-Way ANOVA, as shown in Figure 32. The resulting dialog box appears in Figure 33.

Specify the dependent variable in the Dependent box and the qualitative variable that represents the single factor in the Independent box.

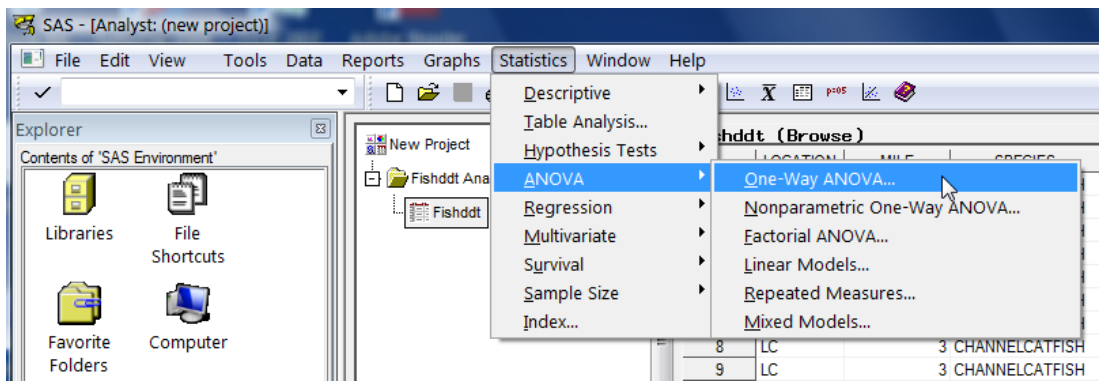


Figure 32 SAS Analyst options for ANOVA

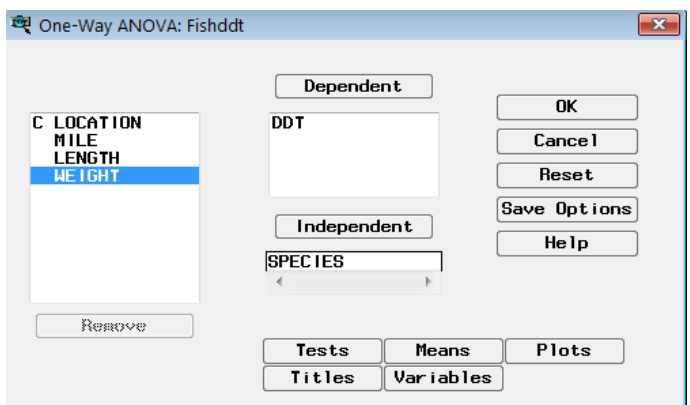


Figure 33 SAS Analyst dialog box for One-Way ANOVA

To perform multiple comparisons of treatment means, click the Means button to obtain the dialog box shown in Figure 34. On this box, select the comparison method

(e.g., Bonferroni's method) and the comparison-wise error rate (e.g., significance level of .05), and specify the main effect to be tested. Click OK to return to the One-Way ANOVA dialog box (Figure 33).

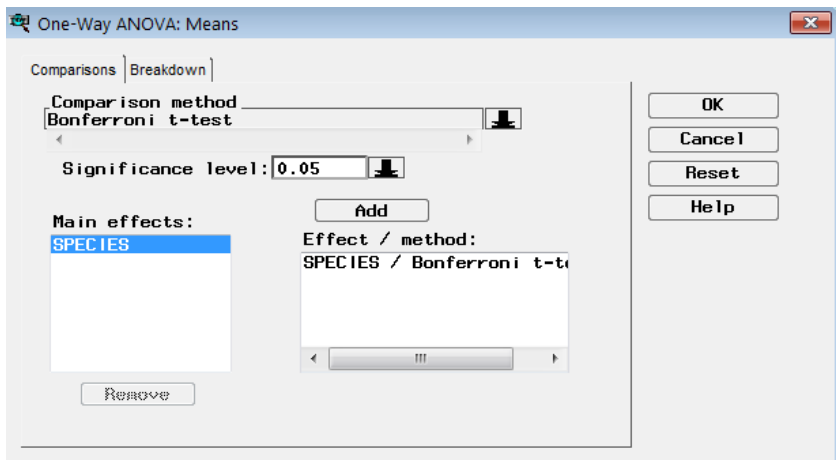


Figure 34 SAS Analyst dialog box for multiple comparisons of means

To perform a test of equality of variances, click the Tests button on the One-Way ANOVA dialog box to obtain the menu shown in Figure 35. From this menu, select the test to be performed (e.g., Levene's test), then click OK to return to the One-Way ANOVA dialog box (Figure 33). Click OK to view the ANOVA results.

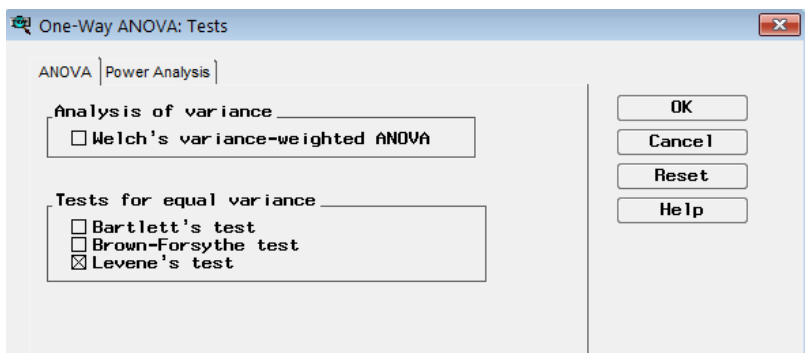


Figure 35 SAS Analyst dialog box for testing equal variances

14. Analysis of Variance for Factorial and Other Designs

To conduct an ANOVA for a factorial design using SAS Analyst, click on the Statistics button on the Analyst menu bar, then click on ANOVA, and finally click on Factorial ANOVA (see Figure 32). The resulting dialog box appears in Figure 36.

Specify the dependent variable in the Dependent box and the qualitative variables that represent the factors in the Independent box. To specify the factorial model, click on the "Model" button. The dialog box shown in Figure 37 will appear. Specify the terms in the model using the Cross button (for interactions). The model terms will appear in the Effects in Model box. Click OK to return to the Factorial ANOVA dialog box (Figure 36).

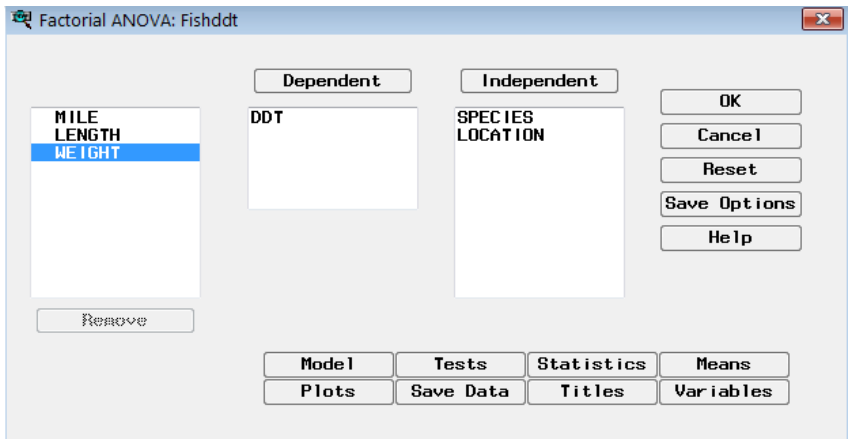


Figure 36 SAS Analyst dialog box for Factorial ANOVA

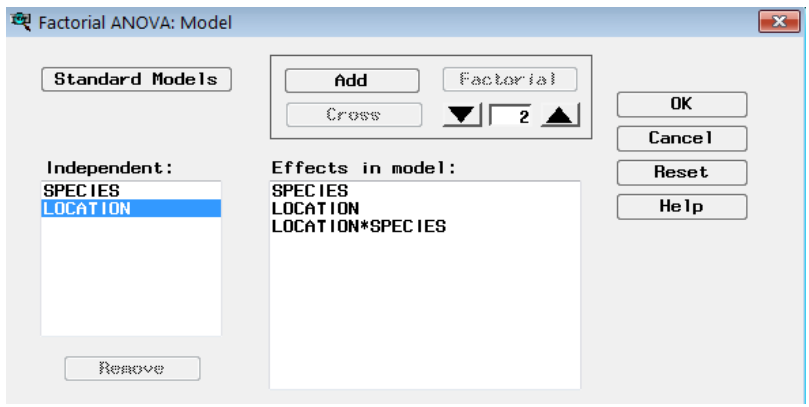


Figure 37 SAS Analyst dialog box for factorial model selection

To run multiple comparisons of means for all treatment combinations, click the Means button on the Factorial ANOVA dialog box and then click the LS Means button. The dialog box shown in Figure 38 appears. Specify the interaction effect of interest in the LS Mean box and select the comparison method (e.g., Bonferroni's method), then click OK to return to the Factorial ANOVA dialog box (Figure 36). Click OK to run and view the ANOVA results.

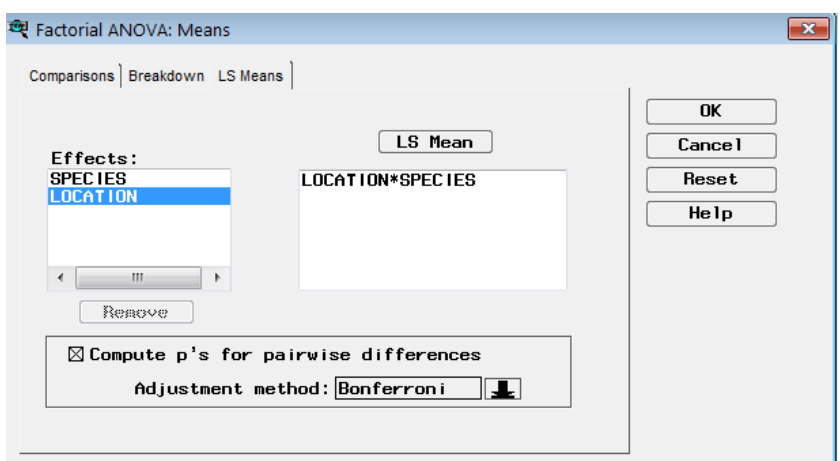


Figure 38 SAS Analyst dialog box for multiple comparisons of factorial means

15. Time Series Forecasting Models

To conduct the Durbin–Watson test for autocorrelated errors in a model for time series data using SAS Analyst, first specify the model to be fit. That is, click on the Statistics button on the Analyst menu bar, then click on Regression, and click on Linear, as shown in Figure 20. Specify the dependent and independent variables in the model on the resulting dialog box (see Figure 21). Once the model has been specified, click on the Statistics button on the Multiple Regression dialog box, then click on Tests to obtain the menu shown in Figure 39.

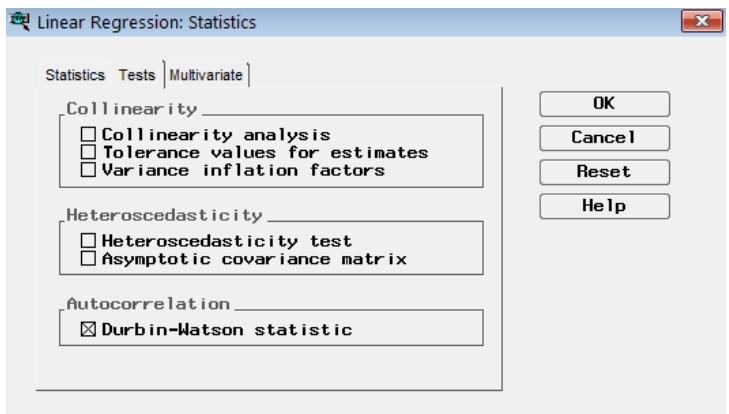


Figure 39 SAS Analyst dialog box for Linear Regression Tests

Check the Durbin–Watson Statistic box, then click OK to return to the Multiple Regression dialog box. Click OK to view the results.

Currently, you cannot fit a time series model with autoregressive errors using SAS Analyst. Consequently, you must return to the Editor window of SAS and type in the program commands to do this. The commands for fitting a time series autoregressive forecasting model with AR(1) errors are shown in Figure 40. After entering the commands, click the Run button to submit the program.

Figure 40 SAS program to fit time series autoregressive error model