

Paper 9-28

Advanced Analytics with Enterprise Guide®

Catherine Truxillo, Ph.D., Stephen McDaniel, and David McNamara,
SAS Institute Inc., Cary, NC

ABSTRACT

From SAS/STAT® to SAS/ETS® to SAS/QC® to SAS/GRAPH®, Enterprise Guide provides a powerful graphical interface to access the depth and breadth of analytic capabilities in SAS. This paper provides an overview of analytic methods available via SAS/Enterprise Guide task wizards. Custom coded analyses using the insert code feature are demonstrated. The paper also describes how you can use Enterprise Guide to "package" automated analyses, custom analyses, graphs, notes, and relevant documents into one comprehensive project file that can be easily scheduled and delivered to others.

BACKGROUND

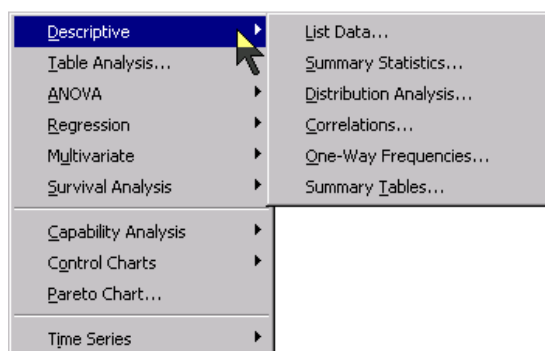
If you are not familiar with Enterprise Guide software, the SUGI 28 beginning tutorial *SAS® Enterprise Guide® -- Getting the Job Done* provides helpful background information about Enterprise Guide and getting started. All demonstrations and examples in this paper are relevant to Enterprise Guide 2.0 on a Microsoft Windows client with SAS Release 8.2 or SAS System 9 as the SAS engine.

The paper assumes that you are already familiar with Enterprise Guide software and some of the analytic procedures implemented in base SAS, SAS/STAT, SAS/QC, and/or SAS/ETS. The paper focuses on modeling tasks in Enterprise Guide, although the information in this tutorial can be applied to many of the tasks in Enterprise Guide software.

This paper first presents a brief overview of the analytic tasks available through Enterprise Guide tasks. Next, it gives examples of how to use Enterprise Guide for statistical modeling tasks including general linear models such as regression, analysis of variance (ANOVA), and analysis of covariance (ANCOVA) and for generalized linear models such as logistic regression and Poisson regression. Finally, it teaches through example about several custom tools and options such as inserted task code, custom documents, project automation, creating custom tasks, and bundling projects.

AN OVERVIEW OF STATISTICAL TASKS

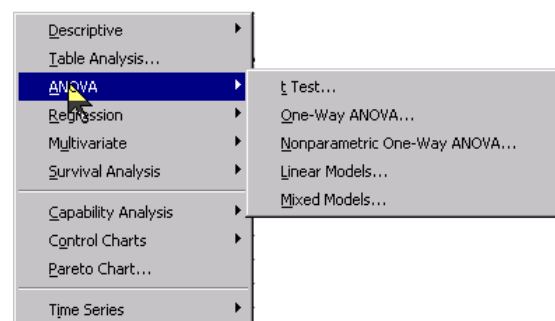
The following discussion describes the usage of many of the analytic tasks in Enterprise Guide and some of their potential applications. Most of the tasks can, in fact, be used for a much broader range of analyses than those described here.

THE BASE SAS TASKS:**THE DESCRIPTIVE MENU**

The tasks in the Descriptive portion of the Analysis menu call

base SAS procedures such as PRINT, MEANS, FREQ, CORR, UNIVARIATE, and TABULATE. These tasks are useful for basic statistical reports with descriptive statistics, simple plots of data, distribution analysis, and correlation analysis.

The Table Analysis task performs two-way table analysis using the FREQ procedure. Various tests of association and cross-tabulation statistics are available in the Table Analysis task.

THE SAS/STAT TASKS:**THE ANOVA MENU**

The t-Test task performs one-sample, paired, and two-sample t-tests by calling the TTEST procedure and produces plots of the means with SAS/GRAPH. Use the one-sample t-test to compare a group mean to a known (or hypothesized) value (one-sample t-test). Use the paired t-test to compare the means of two related samples, or the same sample at two points in time. Use the two-sample t-test to compare the means of two independent groups. The output from the two-sample t-test includes a test for equality of variances and the Satterthwaite adjusted degrees of freedom in addition to the t-test assuming equal variances.

The One-Way ANOVA task compares the means of 2 or more groups that are defined by a single independent (or classification) variable by calling the ANOVA procedure. The ANOVA procedure is generally only appropriate for completely balanced data or one-way models. Multiple comparisons can be performed using a variety of methods to adjust for experimentwise and comparisonwise type-I error. This task offers Bartlett's, Levene's, and Brown & Forsythe's tests for homogeneity of variances, and offers Welch's ANOVA for comparing group means with unequal variances. The task also produces plots of the means using SAS/GRAPH.

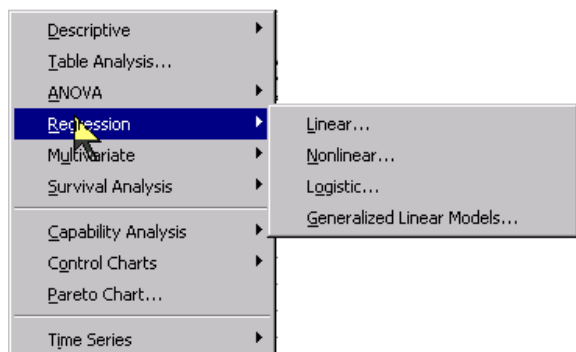
The Nonparametric One-Way ANOVA task performs nonparametric group comparisons (an underlying distribution is not assumed for the data). Group comparisons using the Wilcoxon, Median, Savage, and Van der Waerden tests are available, using asymptotic or exact p-values. The analysis also produces an empirical distribution function from the data. This task calls the NPAR1WAY procedure.

The Linear Models task uses the GLM procedure to perform a variety of linear models including factorial ANOVA, ANCOVA, simple linear regression, multiple regression, polynomial regression, and many customized models. A variety of least-squares means comparisons are available to control for type-I error, and the task offers many diagnostic plots for exploring patterns in your data.

The Mixed Models task uses the MIXED procedure to analyze models with a variety of nested and factorial effects. The options available in the Mixed Models task are similar to the options available in the Linear Models task. As with other tasks, you can easily specify additional options and statements for the MIXED

procedure with code.

THE REGRESSION MENU



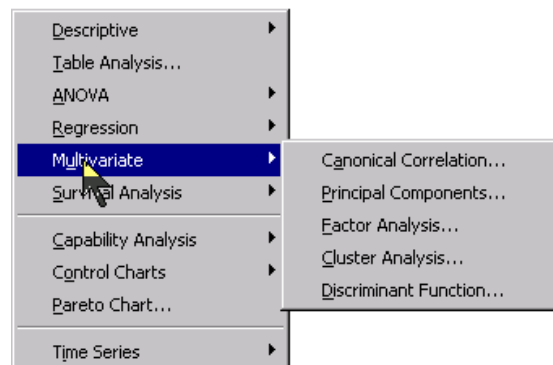
The first task under the Regression menu is the Linear regression task, which uses the REG procedure to perform linear regression analysis with a variety of useful options for model selection, diagnostic statistics, and output data sets. The task also uses SAS/GRAPH to produce a variety of diagnostic, predictive, and influence plots. You can create plots for residuals, influential points, the least-squares regression line and confidence intervals, and many other graphics to help you understand the relationships in your data. You can perform analysis of variance by using dummy-coded categorical variables, although the Linear Models task is generally preferable to the Linear regression task for ANOVA because the Linear Models task allows classification variables. The primary distinguishing feature of the Linear regression task is the availability of several stepwise and all-regressions options for model selection.

The Nonlinear regression task performs nonlinear regression analysis by fitting a variety of power, inverse, log base e , and exponential models. These are models in which the response is not adequately predicted by a simple linear combination of predictors and constants, but by a nonlinear combination of predictors and constants. The Nonlinear regression task uses the NLIN procedure to specify a variety of iteration and step-size search methods, and uses SAS/GRAPH to create plots for prediction, residuals, and influential observations. You can save many of the statistics from the nonlinear regression analysis.

The Logistic regression task calls the LOGISTIC procedure to perform logistic regression on dichotomous or multi-level response variables. The task allows for continuous or categorical predictor variables and offers the same model-building options as the Linear models task, as well as automatic model-selection options. You can specify the logit, the probit, or the complementary log-log link functions for logistic regression, a variety of diagnostic statistics, and odds ratios based on profile likelihoods and/or Wald tests. The task uses SAS/GRAPH to create plots for prediction, influential points, and receiver operator characteristic (ROC) curves.

The Generalized Linear Models task uses the GENMOD procedure to fit linear models with continuous or discrete responses with a variety of distributions. You can perform linear regression, ANOVA, and logistic regression with the Generalized Linear Models task, as well as many other models such as log-linear models and Poisson regression. The task provides the same familiar model-building interface you saw in the Linear Models task as well as several others. As with other modeling tasks, the Generalized Linear Models task uses SAS/GRAPH to create graphs of observed, predicted, and residual values.

THE MULTIVARIATE MENU



The Multivariate menu offers a variety of commonly used multivariate statistics.

The Canonical Correlation Analysis Task allows you to investigate relationships (correlation) among two sets of numeric variables. Canonical correlation analysis is a multivariate extension of multiple correlation analysis, which is an extension of correlation analysis. Canonical correlation analysis creates canonical variates that are linear combinations of the variables within each set of variables and are maximally correlated with one another. The Canonical Correlation Analysis task calls the CANCELL procedure.

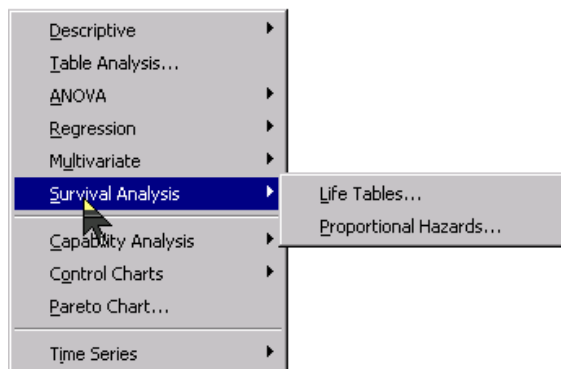
The Principal Components Analysis task creates a set of uncorrelated variables (principal components) from a set of correlated continuous (or dummy-coded discrete) variables. In principal components analysis, the first component accounts for the maximum shared variability among the input variables, the second accounts for the next most variability, and so on. Principal components analysis is a popular method for reducing a large number of correlated variables to a smaller number of uncorrelated variables.

The Factor Analysis task performs common and canonical factor analysis as well as several methods of components analysis including principal components analysis. The task creates factor scores and saves useful statistics, and creates plots helpful for interpreting latent factors. The Factor Analysis task uses the FACTOR procedure.

The Cluster Analysis task uses the CLUSTER and FASTCLUS procedures to perform hierarchical cluster analysis. Cluster analysis allows you to find groups of observations that are similar to one another on a set of input variables. The task also uses the TREE procedure to create tree diagrams using output from the cluster analysis.

The Discriminant Function task uses the DISCRIM procedure to perform several types of parametric and nonparametric discriminant function analysis. The task allows you to specify options for prior probability estimates and produce error tables for classification. You can also use leave-one-out cross-validation or specify a second data set to perform empirical validation of the discriminant functions.

THE SURVIVAL ANALYSIS MENU



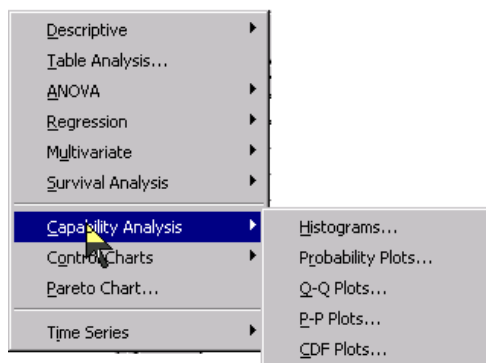
The two tasks on the Survival Analysis menu perform analyses on time-to-event data, which can include survival data, relapse data, recidivism data, failure data, warranty repair data, or any of a variety of other time-to-event data types. Time-to-event data often has censoring, which means that the event occurred before (left-censored), after (right-censored), or between (interval-censored) measurement periods.

The Life Tables task uses the LIFETEST procedure to perform one of two nonparametric forms of survival analysis, including the Kaplan-Meier method, which models the empirical distribution of the data, and the Life Tables method, which evaluates events per unit time where time is divided into equal-spaced units. The Life Tables task uses SAS/GRAPH to create useful plots for evaluating the distribution of time. Linearized plots for evaluating the exponential, Weibull, and lognormal distributions are available. Only uncensored or right-censored data should be used with this task.

The Proportional Hazards task uses the PHREG procedure to perform semi-parametric survival analysis using the Cox Proportional Hazards model. The task offers a variety of options including stepwise model selection and methods for handling ties in event times.

THE SAS/QC MENUS:

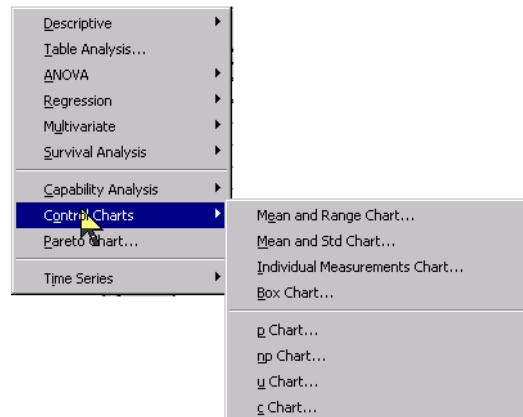
CAPABILITY ANALYSIS



Capability analysis in Enterprise Guide allows you to evaluate the distribution of a variable, enter target values and specification limits, and evaluate process capability using statistical and graphical methods. There are five tasks for capability analysis, each of which calls the CAPABILITY procedure and is named for the type of graph used to evaluate the specified distribution.

There are many options for customizing the graph in capability analysis. By default, the plots display the target, upper specification limit, and lower specification limit when specified as well as a reference line for the specified distribution.

CONTROL CHARTS



The Control Charts tasks use the SHEWHART procedure to create eight different types of control charts appropriate for individual or subgrouped, measurement or attribute data.

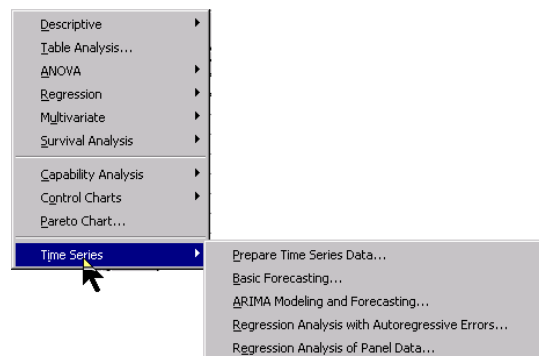
The Control Chart tasks share many features in common, although each creates a different control chart. You can specify methods for calculating control limits. You can specify a multiple of sigma (the within-group standard error of the estimate), an input data set with limits, or specific user-entered control limits. Block variables can identify groups of observations in the process without affecting control limit calculations, and you can apply standard tests for special causes to find nonrandom run patterns in your process. Many graphical options for customizing the appearance and usefulness of your control charts are available in the tasks.

PARETO CHARTS

The Pareto Chart task uses the PARETO procedure to create a horizontal or vertical chart (similar to a bar chart) to allow you to identify commonly occurring categories in your data. Typically, these categories include causes of process or product failure, although they can include nearly any type of categorical variable that is counted. For example, you could identify offices with the most frequent repair calls, sales territories with the most customer inquiries, or retail locations with the greatest number of sales.

The Pareto Chart task allows you to customize the chart in many ways. You can create two-way charts, label bars with frequencies or percentages, show cumulative percentages on the chart, use special colors to identify the largest or smallest n groups, and so on.

THE SAS/ETS TIME SERIES MENU



The Time Series menu offers several tasks for time series analysis. Time series analysis allows you to identify autoregressive, seasonal, or cyclical patterns in a time series.

The Prepare Time Series Data task prepares time series data for use by other Time Series tasks. It is also very useful for performing transformations on data for use in other Enterprise Guide tasks. For example, you can apply mathematical

operations, functions, and other transformations to variables in a data set. The Prepare Time Series Data task uses the EXPAND procedure.

The Basic Forecasting task uses the FORECAST procedure to generate forecasts for a time series in a single step using a variety of forecasting methods. The task also uses SAS/GRAPH to create plots of the observed, forecast, and residual values.

The ARIMA Modeling and Forecasting task analyzes and forecasts equally spaced univariate time series data, transfer function data, and intervention data using the autoregressive integrated moving-average (ARIMA) or autoregressive moving-average (ARMA) model. An ARIMA model predicts a value in a response time series as a linear combination of its own past values, past errors (also called shocks or innovations), and current and past values of other time series. The task uses the ARIMA procedure.

The Regression Analysis with Autoregressive Errors task uses the AUTOREG procedure to perform linear regression analysis on time series data or data with autoregressive errors.

The Regression Analysis of Panel Data task uses the TSCSREG procedure to fit linear econometric models for data in which time series and cross-sectional measurements are combined. This task deals with panel data sets that consist of time series observations on each of several cross-sectional units.

FOCUS ON THE MODELING TASKS

Perhaps the most commonly used statistical tasks in Enterprise Guide are the modeling tasks in the ANOVA and the Regression menus.

The following discussion compares four of the modeling tasks, provides suggested uses of each, and demonstrates through examples how to specify various models using the tasks. The following modeling tasks are presented: Linear Models, Linear regression, Logistic regression, and Generalized Linear Models.

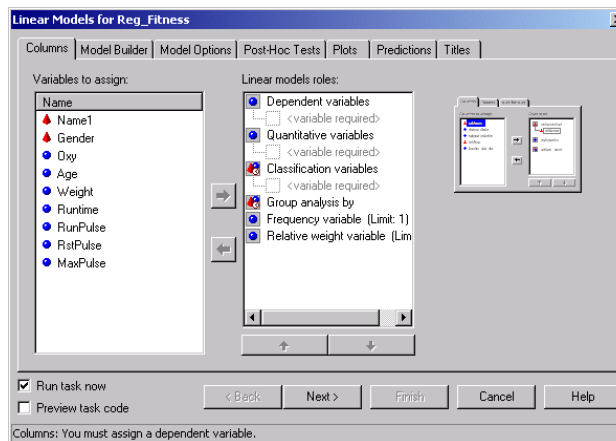
ASSIGNING COLUMNS TO ROLES

With the exception of the Linear regression task, each of the featured modeling tasks allows any combination of quantitative and classification variables. If you would like to use a classification variable in the Linear regression task, you should use $k-1$ dummy-coded indicator variables, where k is the number of levels of the classification variable.

The Linear Models and Linear regression tasks are intended for use with normally distributed (and hence quantitative) response variables. The Logistic regression task is intended for use with categorical response variables. The Generalized Linear Models task is appropriate for any combination of discrete or continuous responses and predictors, where the distribution of the random component is one of those available in the task (PROC GENMOD handles distributions from the exponential family).

Each of the modeling tasks allows you to specify a frequency variable, which can simplify your data-entry system in many cases, particularly for data sets with count or discrete response variables.

The Columns tab of the Linear Models task window is displayed below:

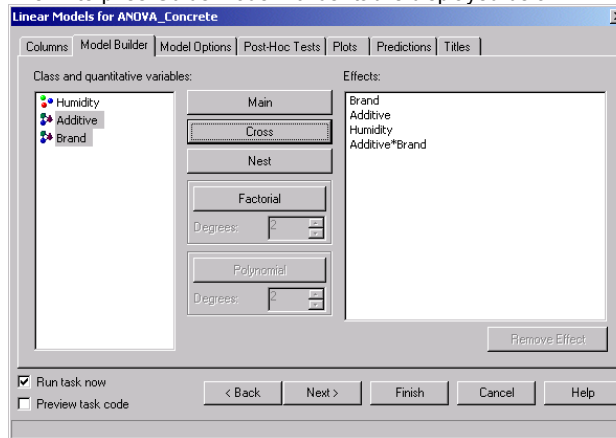


THE MODEL BUILDER

Using the Linear Models, Logistic, and Generalized Linear Models tasks, you can specify your model using the Model Builder tab.

This model builder is similar to model builders in other SAS products including the Analyst application in SAS/STAT software.

The Enterprise Guide Model Builder tab is displayed below:



The panel on the left shows the predictor variables in the analysis. The panel on the right shows the effects in the model. Main effects are displayed by default. You can specify interactions by selecting variables and selecting the Cross button, or you can specify all effects up to a certain degree with the Factorial button. To specify all main effects and two-way interactions, simply select Factorial with Degrees set to a value of 2. Similarly, to specify all main effects, two-factor interactions, and three-factor interactions, select Factorial with Degrees set to 3.

You can use the model builder to specify polynomial effects for quantitative variables. If only one variable is selected, you can select the Cross button to get the quadratic term for that variable, or you can specify polynomial terms for one or more quantitative variables at a time with the Polynomial button. As with the Factorial button, if you select Polynomial, Degrees n , you will get all terms to the power of n and lower. For example, Polynomial, Degree 3 produces all cubic, quadratic, and linear terms.

The model builder also features a Nest button, which allows you to specify effects that are nested within other effects-- for example, patients within a clinic, customers within a territory, or test units within a purchase lot.

MODEL SELECTION

The Linear regression and Logistic regression tasks offer model-selection features that allow you to automatically generate candidate models, compare them, and determine the most useful model for your situation.

In the Linear regression task, there are nine model-selection options, including the full model (no selection), three stepwise methods, and five all-possible regressions techniques. You can specify effects to lock into the model, and you can specify the

alpha for terms entering or leaving the model in stepwise selection. You can also specify which model fit statistics you want displayed in the final report.

In the Logistic regression task, there are five model-selection options available, including the full model (no selection), three stepwise methods, and the best-subset technique, which selects the model with the number of terms that maximizes the likelihood score (chi-square) for all possible model sizes. It is important to note that the best subset option cannot be used with classification variables.

STATISTICS, MODEL OPTIONS, AND POST-HOC TESTS

The Linear and Logistic regression tasks feature a Statistics tab in which you can select many output statistics and options for the analysis. The Linear Models task and the Generalized Linear Models task feature a Model Options tab, which serves a similar purpose to the Statistics tab in the Linear and Logistic regression tasks.

The Statistics tab and the Model Options tab typically allow you to specify the type of sums of squares to display, any transformation and link functions you wish to apply to the model, diagnostics for model adequacy, and the type of confidence intervals to display. The Generalized Linear Models task is the most flexible of the modeling tasks, as it allows you to fit models with any combination of discrete and continuous predictor and response variables. For this reason, there are many more options in the Generalized Linear Models task than in any of the other modeling tasks. For example, in the Statistics tab, you can specify one of nine different link functions (which relate to the transformation applied to the data in the model) and one of seven different distributions from the exponential family (which relate to the assumed distribution of the random errors in the model) using point-and-click options. Generalized Linear Models are highly versatile and flexible, and you should be as educated about them as you can be when using them. For more information about generalized linear models, see Nelder and Wedderburn (1972) and McCullach and Nelder (1989).

In addition to the statistics and model options tabs, the Linear Models task and the Generalized Linear Models task feature a Post-Hoc Tests tab that allows you to specify class effects (main effects and interactions) for least-squares means comparisons, estimates, and means estimates. A variety of options are available for adjusting comparisonwise and experimentwise type-I error, displaying additional statistics, and displaying p-values for means comparisons.

If you run a model, and go back to the task dialog to change your model (for example, to add or remove a term), you should always verify that your post-hoc tests are set up as you intended before rerunning the model.

PLOTS

Each of the modeling tasks features a Plots tab that allows you to create plots of observed, predicted, and residual values from the analysis. Additional plots are available as appropriate within each modeling task. For example, the Linear Models task allows you to plot group means, while the Linear regression task allows you to plot your least-squares regression line with 95% confidence curves. There are also plots available that allow you to evaluate influential and outlying points in the analysis.

PREDICTIONS

Each of the modeling tasks contains a Predictions tab on which you can specify statistics, predictions, residuals, and other values to be stored in an output data set. These output tables are useful for evaluating the assumptions of your models. For example, you can create an output data table containing the residuals from linear regression and perform distribution analysis on them to evaluate the regression assumption that the response variable, conditional on the predictor variable(s), is normally distributed.

GRAPHS TO AID INTERPRETATION

After you have performed your analysis, you might want graphics beyond those available in the task to understand the effects in your data. There are many different types of graphs available in

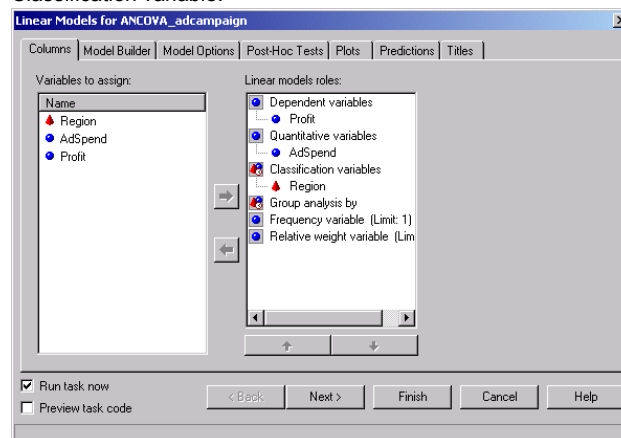
the Graph menu to help you do this.

EXAMPLE: ANALYSIS OF COVARIANCE

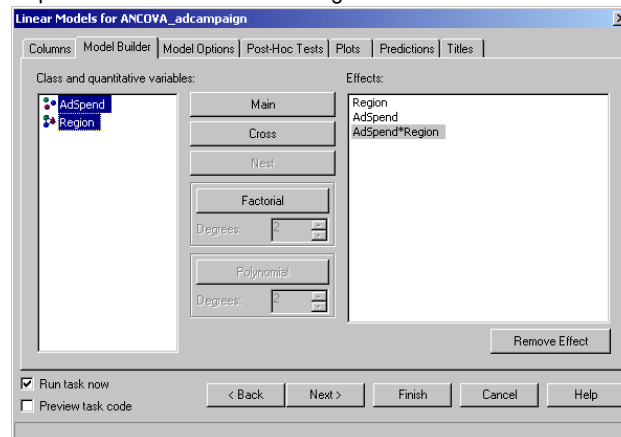
This example uses a data set called AdCampaign that contains profit figures from stores in two regions where different amounts of money were allocated for advertising campaigns. Use Enterprise Guide for analysis of covariance with separate slopes and a plot to help interpret the effects in the model.

First, specify the model using the Linear Models task.

On the Columns tab, assign Profit as the Dependent Variable, AdSpend as the Quantitative variable, and Region as the Classification variable.



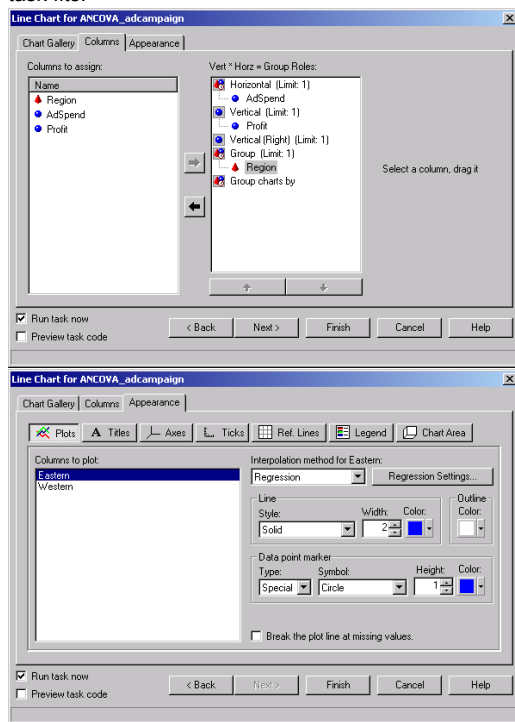
On the Model Builder tab, specify the AdSpend by Region interaction using the Cross button. This specifies that separate slopes should be fitted for each region.



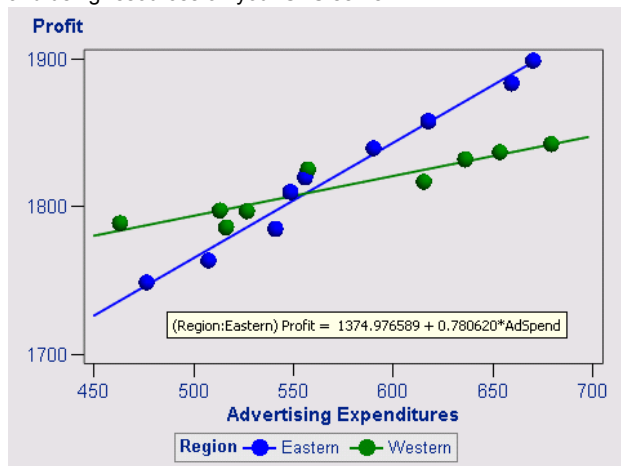
You will use graphics from the Graph menu. Select Finish to run the ANCOVA. Partial output is displayed below:

Source	DF	Type III SS	Mean Square	F Value	Pr > F
Region	1	4732.83147	4732.83147	69.69	<.0001
AdSpend	1	21596.82102	21596.82102	318.00	<.0001
AdSpend*Region	1	5106.43413	5106.43413	75.19	<.0001

To help you interpret the significant interaction, use the *Multiple line plots by group column* graph. Specify Profit as the Vertical axis variable, AdSpend as the Horizontal axis variable, and Region as the Group variable. On the Appearance tab, make sure you specify Regression as the Interpolation method for each group separately. This will create two separate regression lines on the plot so that you can see the model that the Linear Models task fits.



If you have ActiveX or Java controls as your graphical output format, you can move the mouse pointer over the lines to see the regression equation for each. These controls also permit a high level of interaction with your graph without rerunning any code and using resources on your SAS server.



CUSTOM ANALYSES

In Enterprise Guide, you can write your own SAS programs using the code window or you can edit code that is produced by a task. With Version 2.0 of Enterprise Guide, you also have the ability to directly customize task code while you are setting up an analysis. There are many situations in which you may want to customize code. The following examples present several scenarios in which you would want to customize the code that is created by the task. There are two primary ways of inserting SAS code within the task dialogs: You can insert statements preceding specific SAS statements, or you can insert options (and other SAS code) within

specific SAS statements.

If you insert code before or within a specific statement, that statement must be used in order for the custom code to be executed.

Note that the Insert Code features in Enterprise Guide are considered advanced features and do not include any code validation, so make sure your code is correct!

EXAMPLE: MULTIVARIATE ANALYSIS OF VARIANCE

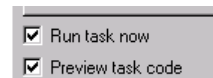
The GLM procedure performs traditional multivariate analysis of variance (MANOVA). However, the MANOVA statement is not called by the Linear Models task in Enterprise Guide. Nonetheless, it is relatively straightforward to insert the appropriate SAS code to perform MANOVA.

This example uses a data set called Concrete that contains two response variables, Strength1 and SetTime, and two independent variables, Brand and Additive.

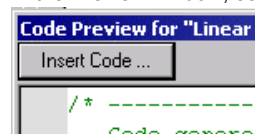
Set up the model using the Linear Models task as you normally would for 2-way ANOVA. Specify both responses in the Dependent Variables role.

Specify the two main effects (Brand, Additive) and the interaction (Brand * Additive) in the model builder.

To insert SAS code to perform MANOVA instead of ANOVA, select the Preview Task Code checkbox.



In the Preview window, select Insert Code.

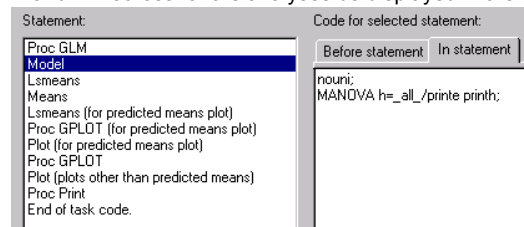


There are several ways to insert the code that would perform equally well. For this example, select the Model statement and the In Statement tab to insert code within the Model statement.

When you select In Statement, the custom code is always inserted at the end of the statement. Therefore, insert the NOUNI option to suppress the univariate ANOVA output, and end the Model statement with a semicolon. Then type the following statement:

```
MANOVA h=_all_ /printe printh;
```

This statement specifies that MANOVA be performed using all the effects specified in the MODEL statement, and requests that the H and E matrices for the analyses be displayed in the output.



Select OK and then Finish to run the task. Partial output from the MANOVA is displayed below:

MANOVA Test Criteria and F Approximations for the Hypothesis of No Overall Brand* Additive Effect					
H = Type III SSCP Matrix for Brand*Additive					
E = Error SSCP Matrix					
S=2 M=0.5 N=10.5					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.11549000	22.34	4	46	<.0001
Pillai's Trace	0.89283754	9.68	4	46	<.0001
Hotelling-Lawley Trace	7.58665227	43.07	4	26.595	<.0001
Roy's Greatest Root	7.57713599	90.93	2	24	<.0001

NOTE: F Statistic for Roy's Greatest Root is an upper bound.

NOTE: F Statistic for Wilks' Lambda is exact.

EXAMPLE: CONTRASTS

Many of the modeling tasks allow you to perform pairwise means

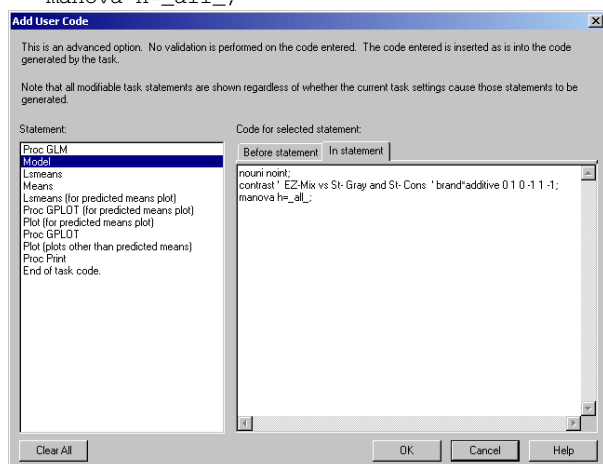
comparisons on the Post-Hoc Tests tab. You can use one of a variety of p-value adjustments to control type-I error.

However, sometimes you are interested in comparing only one pair of groups, or perhaps you are interested in comparing one group to the mean of several other groups. In these situations, contrasts are more appropriate. To perform contrasts in Enterprise Guide, you insert CONTRAST statements in the code. This example uses the data set Concrete to compare one brand, EZ-Mix (both additive groups, Standard and Reinforced), to the Standard group for the other two brands (Standard-Graystone, Standard-Consolidated). You will use multivariate contrasts and specify the model using the cell means model. This is not the same model you fit with MANOVA earlier, as you will see. The cell means model allows you to set up and test contrasts more easily than your full model of interest. You can test any contrast you are interested in by using the cell means model, and it requires fewer coefficients.

Set up the same analysis as the MANOVA by opening the task window for the previous Linear Model. Select the Preview Task Code check box. Select the Model statement and the In Statement tab.

Insert the following code:

```
nouni nouni;
contrast 'EZ-Mix vs St-Gray and St-Cons'
  brand*additive 0 1 0 -1 1 -1;
manova h=_all_;
```



Select OK and then Finish to run the task. Enterprise Guide asks whether you want to replace the results from the previous run. Select No. The MANOVA model you fit in this example is different from the model you fit in the previous example. The model you fit simply allowed you to simplify the process of specifying contrasts. The results for the multivariate contrast are shown below:

MANOVA Test Criteria and Exact F Statistics for the Hypothesis of No Overall EZ-Mix vs St- Gray and St- Cons Effect					
H = Contrast SSCP Matrix for EZ-Mix vs St- Gray and St- Cons					
E = Error SSCP Matrix					
S=1 M=0 N=10.5					
Statistic	Value	F Value	Num DF	Den DF	Pr > F
Wilks' Lambda	0.20078402	45.78	2	23	<.0001
Pillai's Trace	0.79921598	45.78	2	23	<.0001
Hotelling-Lawley Trace	3.98047593	45.78	2	23	<.0001
Roy's Greatest Root	3.98047593	45.78	2	23	<.0001

BUNDLING YOUR WORK

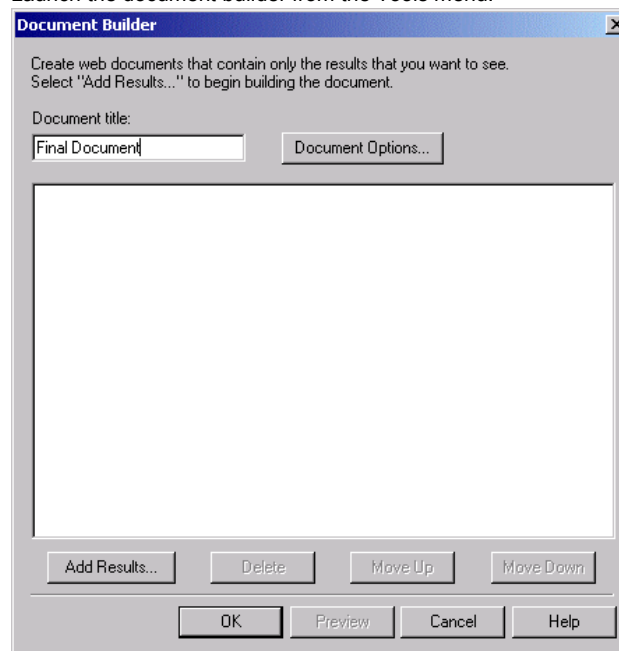
BUILDING DOCUMENTS

After you perform your analyses, you can create a final document that can be published on the Web or to a channel, emailed to others, printed, or placed into another document or presentation. There are several ways that you can select portions of output from Enterprise Guide for presentation. One elegant way to create

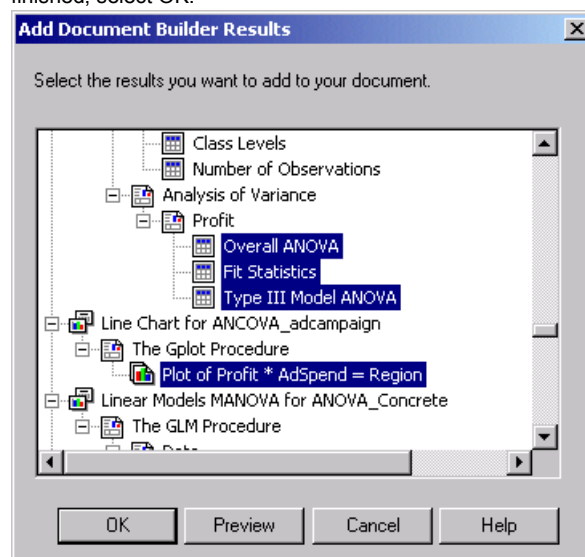
a single report is with the document builder. The document builder organizes and combines HTML reports into a single HTML document.

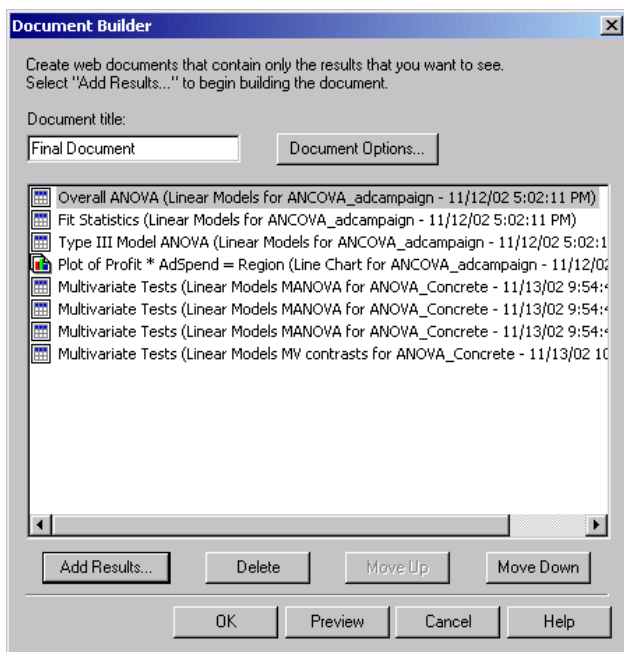
In the previous examples, you created three reports. Suppose you are not interested in all the results from all three analyses, but you would like to combine the relevant portions of each report into a single HTML document.

Launch the document builder from the Tools menu.

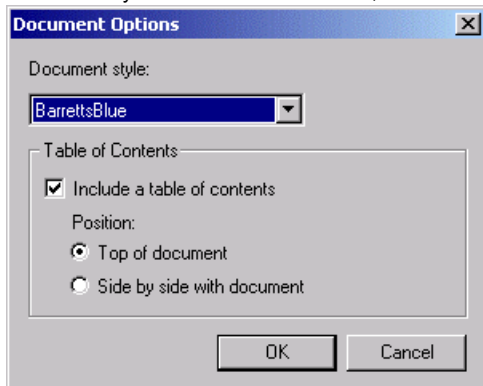


Select the Add Results button. Choose the portions of the output you want in your final document by selecting the titles from the list that appears. You can use Control-click to select more than one nonconsecutive portion of output at a time. When you have finished, select OK.





To customize the report, select Document Options. Select a document style and a table of contents, if desired.



To view and save the document, select Preview. Your default HTML browser will open and display the report. You can save the document to a specific location from the File menu. Click OK in the Document Builder window to return to the project. The document builder is now a node in your project.

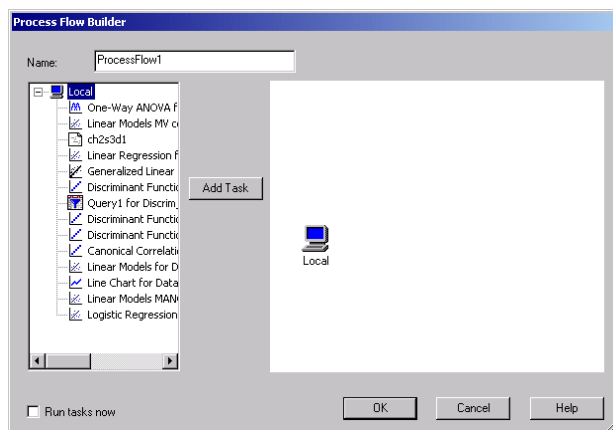
AUTOMATING THE PROCESS

You put a great deal of care and effort into specifying your models, creating your custom analyses, making your custom graphics, and building a final document. Ideally, you would like to automate these analyses to run them in a desired order the next time your data sets are updated rather than respecify all of your analyses.

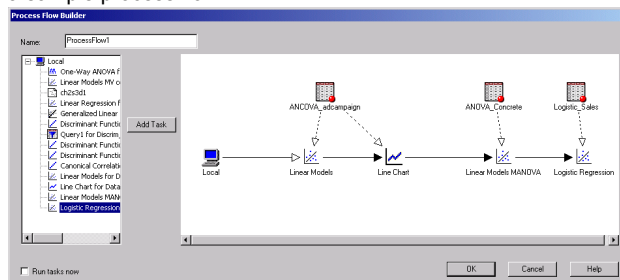
Projects in Enterprise Guide allow you to rerun tasks individually without setting up the options a second time. However, if you are running the same analyses frequently, regardless of whether they are complex statistical reports or simple frequency tables, even rerunning individual tasks can be cumbersome. And re-creating your document repeatedly can be frustrating.

Your analyses and queries can be rerun with a single task in your project. This task is called the process flow builder (PFB). The PFB can also be scheduled to run automatically at a regular time. After you have run the PFB, simply Preview your document in the document builder to see the updated report.

Launch the PFB from the Tools menu. The process flow builder shows all the tasks in your project in the left pane and your SAS server(s) in the right pane.

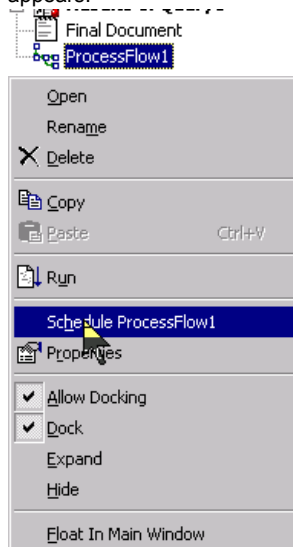


Select the tasks you wish to automate and select the Add Task button to add them to the process flow. The picture below shows a sample process flow:



Select OK to create the process flow.

You can schedule your PFB to run regularly or at a specified time. To activate the Enterprise Guide scheduler, right-click on the PFB in the project window and select Schedule from the menu that appears.



Enterprise Guide automatically creates a VB script that runs your PFB. Select your schedule from the Schedule tab. Note that the project code section is scheduled on (via a Windows system called "AT") and launched from the machine it was scheduled on, but all code processing occurs on the last servers selected in the project document.

Select Apply and then OK to finalize your scheduled PFB.

CREATING YOUR OWN TASKS

Perhaps there are analyses you perform regularly that are not yet available in the Enterprise Guide task list, or outstanding macro libraries you would like to surface in a friendly interface to analysts, or maybe you would like to simplify existing tasks for less experienced users. Using Microsoft Visual Basic, C++, or .NET development environments, you can create your own tasks to suit your organization's needs and share those add-in tasks with other Enterprise Guide users.

The SUGI 28 advanced tutorial *Developing Custom Analytic Tasks for Enterprise Guide* provides useful information and a demonstration of creating custom applications and tasks using Visual Basic and the open EG API.

For further information about custom tasks and to view and download sample code and add-ins, visit:

<http://www.sas.com/eguide>

STORED PROCESS AUTHORIZING in Release 2.1

In Release 2.1 of Enterprise Guide (forthcoming) you will be able to publish your work from Enterprise Guide as a stored process to the Stored Process Server for parameterized consumption by end users via other applications, including Analytic and Web Report Center reports (new in 9.1), Office Integration (new in 9.1), Information Delivery Portal, and SAS/IntrNet®, as well as Enterprise Guide.

SAVING YOUR WORK

Saving and keeping track of your work in Enterprise Guide is made simple with the use of *project (*.seg) files*. Project files organize data, tasks, code, SAS logs, results, notes, document builders, queries, and process flows in a single location.

Data files are not actually part of the project file. Instead, project files contain references to data locations through *file references* and *SAS libraries*. If a data location changes, the reference to the data must be updated in your project file.

If your data sources reside on a remote server, or in a location that is accessible by others, an easy way to share your projects with others is to define a library name for each user's Enterprise

Guide installation. Define data locations with SAS libraries in Enterprise Guide. You can save the *.seg file to a shared location or send the project file to others. Your colleagues will be able to see the work you have done, continue work of their own, add notes, and so forth. Using the Enterprise Guide repository facilitates sharing of projects and usage of common libraries.

If your data sources are *not* in an accessible location for others, you can send the data sets as well as the *.seg file to others. The recipient should specify the path to the data sets in the Properties (from the right-click menu) for each data set.

CONCLUSIONS

The modeling tasks in Enterprise Guide are flexible enough to handle a wide variety of statistical models, and the graphing tasks allow you to investigate your findings graphically.

With the customization features in Enterprise Guide 2.0, you can make your analyses fit your business problems in a variety of ways. You can add SAS code within a task, organize your results using the document builder, automate your analyses using the process flow builder and the scheduler. You can even create your own tasks to perform analyses that are not in the Enterprise Guide task list.

Enterprise Guide project files organize your work and allow you to know exactly what has been done, keep track of notes, and even write your own SAS programs.

REFERENCES

Nelder, J.A., and R.W.M. Wedderburn(1972). "Generalized Linear Models." *Journal of the Royal Statistical Society A*135: 370-384.
McCullach, P., and J.A. Nelder(1989). *Generalized Linear Models. 2nd ed.* London: Chapman and Hall.

CONTACT INFORMATION

Your comments and questions are valued and encouraged.

Contact the author at:

Catherine Truxillo
SAS Institute Inc.
SAS Campus Drive
Cary, NC 27513
Work phone: 919-531-4641
E-mail: Catherine.Truxillo@sas.com

SAS and all other SAS Institute Inc. product or service names are registered trademarks or trademarks of SAS Institute Inc. in the USA and other countries. ® indicates USA registration.

Other brand and product names are trademarks of their respective companies.

Copyright © 2002 by SAS Institute Inc.