Using JMP to Develop a Model Specification for Daily Energy Consumption

Thomas A. Fuchs, LG&E Energy Corp., Louisville, KY

Abstract

Recently, there has been a lot of discussion on the SAS-L SAS news list about SAS Institute's JMP product. What is it? What does it do? Is this a SAS *Lite*? SAS describes JMP as "the statistical discovery software." Visual and tabular information are linked together to facilitate discovery. One who uses the SAS System will find themselves "over the rainbow" with JMP. This paper will provide a simple example of how the discovery software was used to interactively develop an Ordinary Least Squares (OLS) model specification to explain daily electric energy consumption. The presentation was prepared using JMP 4.0 on a Windows 2000 machine.

Introduction

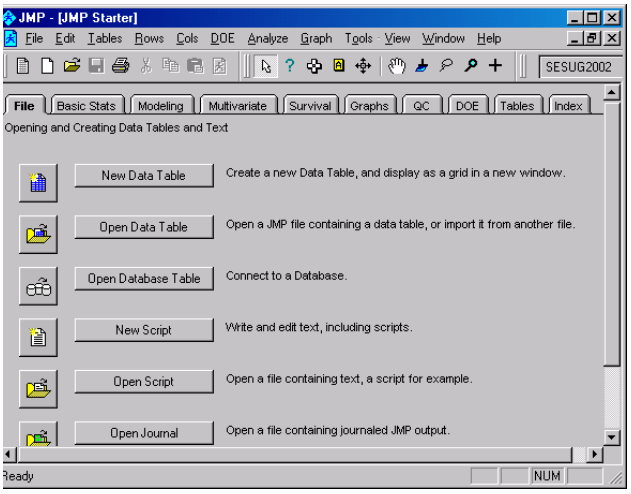
The Forecasting and Market Analysis Department of LG&E Energy Corp. uses a number of computer tools to prepare forecasts and perform customer research. Some examples are SAS, Eviews and MetrixND. Each tool has its own strengths and weaknesses. Last year, the department purchased a copy of JMP 4.0 from SAS Institute. The group wanted to demonstrate capabilities of JMP. They desired to test SAS Institute's claim that JMP was "the statistical discovery software." The demonstration would illustrate how the product compares with other tools.

Recently, there has been an interest in JMP among posters to the SAS-L SAS news list. Many wish to learn if JMP is a SAS *Lite*. The remainder of this paper will show that JMP is not SAS, but it has been developed with some powerful capabilities. JMP's strength is in its eye-catching visuals.

A True Windows Application

SAS was developed in the days of the mighty mainframe, and it has maintained backward compatibility of its procedural coding as the product has migrated to graphical user interface environments like Windows. JMP was developed when SAS Institute recognized a need for statistical software on Apple Macintosh platforms.

JMP was designed to appeal to this unique group of clients by integrating visual representations with the tabular numeric information typically used in statistical analysis. Click a point on a graph, and JMP will highlight the corresponding row in the data table. Select rows in the data table, and the graph highlights the points. JMP abandoned DATA and PROC steps for "point and click" and "drag and drop." In time, JMP moved from the Mac to PC Windows. This move allowed JMP to grow to use the new functionality offered by the Windows environment. In version 4.0, JMP employs object linking and embedding (OLE) and open database connectivity (ODBC). It has also added scripting capability. Scripts allow users to repeat an analysis among different data sets. This demonstration was performed on a Windows 2000 workstation. The picture below shows the JMP Starter screen which begins a JMP session.



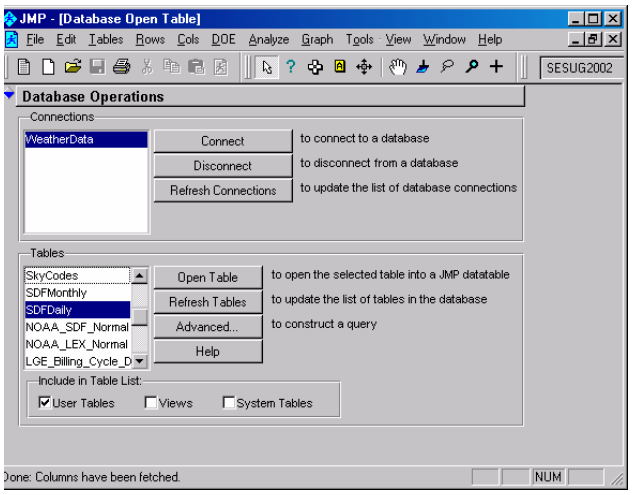
Can JMP process SAS data? Yes, it can perform analyses on SAS Datasets and SAS Transport files.

Can an analyst use his/her SAS code in JMP? No. But, for some SAS programs, a few mouse clicks and keystrokes in JMP can reproduce the SAS analysis.

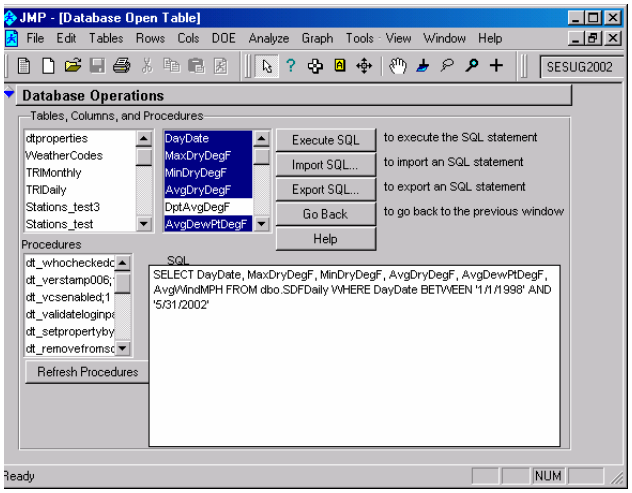
Creating a JMP Data Table

The first step in any analysis is to enter or load the data. In this demonstration, the data were located in a series of MS SQL Server and MS Access databases. The data were successfully loaded using the JMP *Open Database*

Table button to retrieve via an ODBC connection. In the following picture, the analyst retrieves the data by selecting an ODBC data source and a table or view.



Using the *Advanced* button, the columns can be specified. The JMP SQL statement can be edited to customize the retrieval.



JMP can create scripts for most operations. A script file to perform the data gathering shown above looks like:

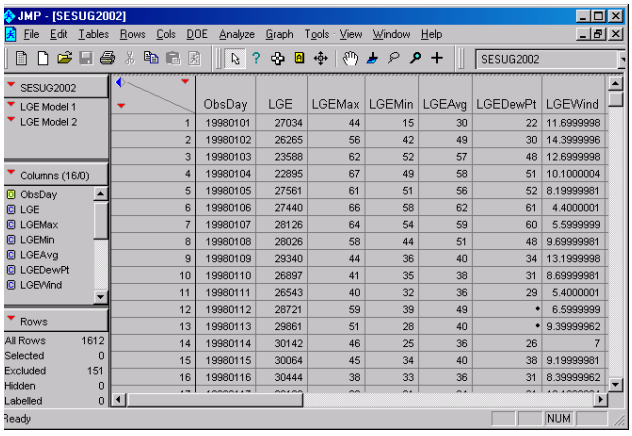
```
// Step01-LoadJMP.jsl
// Load JMP data tables
// from database tables
// T A Fuchs
// 6/2002

open database("DSN=WeatherData",
 "SELECT W.DayDate, W.MaxDryDegF AS LGEMax,
W.MinDryDegF AS LGEMin, W.AvgDryDegF AS
LGEAvg, W.AvgDewPtDegF AS LGEDewPt,
W.AvgWindMPH AS LGEWind FROM dbo.SDFDaily
As W WHERE W.DayDate BETWEEN '1/1/1998' AND
'5/31/2002'",
 "LGEWeather");
```

Energy usage is highly dependent on the weather, but it also is influenced by behavioral factors. Some examples are the day of the week, holidays and the temperature level at which most people switch from using their furnace to their air conditioner. JMP allows the user to create data columns and use formulas to provide the data values. The creation of columns and entering formulas is performed very well by using a script. The script used for this part of the demonstration is shown below.

```
// Step03-AddFormulas.jsl
// Add date formulas
// T A Fuchs
// 6/2002

dt = open("DailyLoads.jmp");

dc = dt << New Column("ObsYr", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula(Floor( :ObsDay * 0.0001));

dc = dt << New Column("ObsMon", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula(Floor( :ObsDay * 0.01) -
:ObsYr * 100);

dc = dt << New Column("MMDD", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula( :ObsDay - 10000 *  :ObsYr);

dc = dt << New Column("DOW", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula(Day Of Week(Date MDY(
:ObsMon,
  :MMDD - 100 *  :ObsMon,  :ObsYr)));

dc = dt << New Column("Holiday", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula( :MMDD == 101 |  :MMDD == 704
|
  :MMDD == 1225 |  :ObsDay == 19980412 |
  :ObsDay == 19990404 |  :ObsDay ==
20000423 |
  :ObsDay == 20010415 |  :ObsDay ==
20020331 |
  :ObsDay == 20030420 | ( :MMDD == 102 |
  :MMDD == 705 |  :MMDD == 1226 |
  524 <  :MMDD <= 531 | 900 <  :MMDD <=
907) &
  :DOW == 2 | 1121 <  :MMDD <= 1128 &  :DOW
== 5 |
```

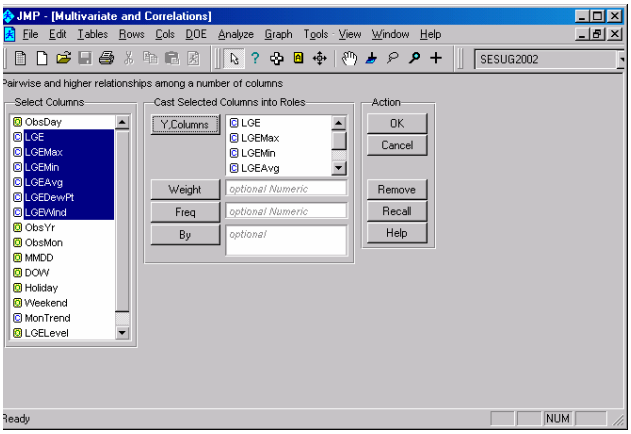
```
  ( :MMDD == 1231 |  :MMDD == 703 |  :MMDD
== 1224) &
  :DOW == 6);

dc = dt << New Column("Weekend", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula(( :DOW == 1 |  :DOW == 7) & !
:Holiday);

dc = dt << New Column("MonTrend", Numeric,
Best);
dc << Set Modeling Type("Continuous");
dc << Formula(( :ObsYr - 1998) * 12 +
:ObsMon);

dc = dt << New Column("LGELevel", Numeric,
Best);
dc << Set Modeling Type("Ordinal");
dc << Formula( :LGEAvg > 65);

close(dt);
```

A snapshot of the resulting JMP data file looks like

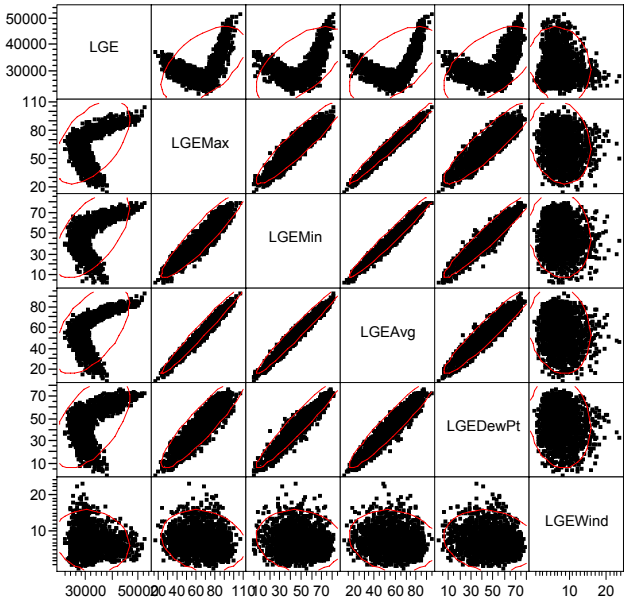


JMP Data Analysis

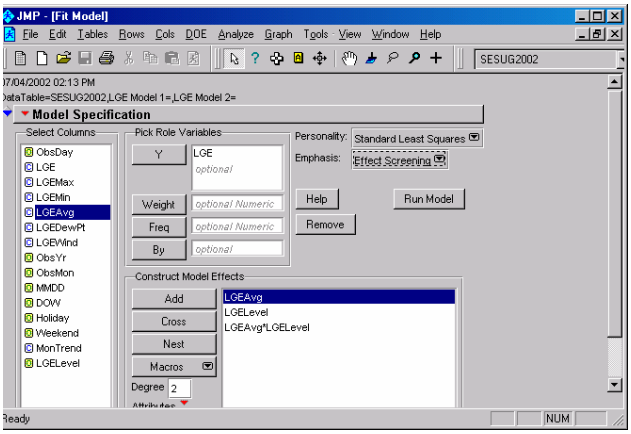
A scatter plot is a tool to introduce an analyst to his/her data. JMP makes this easy. Click the *Multivariate* tab and the *Multivariate* button. Next, select the desired data columns and click the *Y, columns* button, as shown below.



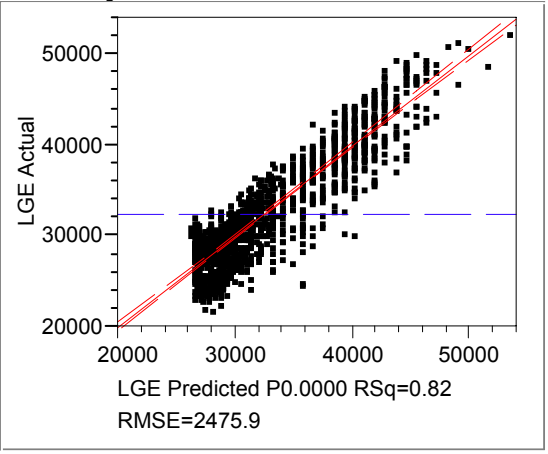
Scatter plots are created by default.



Regression models are prepared in JMP using the *Fit Model* button of the *Modeling* tab. A simple model and results are shown below.



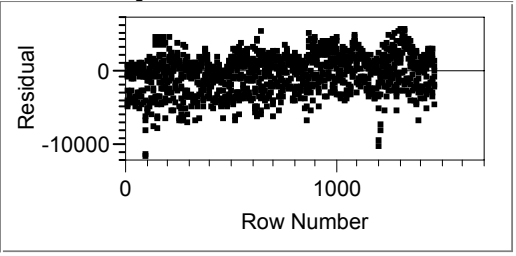
Response LGE
Actual by Predicted Plot



Summary of Fit

RSquare	0.82418
RSquare Adj	0.823818
Root Mean Square Error	2475.938
Mean of Response	32462.56
Observations (or Sum Wgts)	1461

Residual by Row Plot



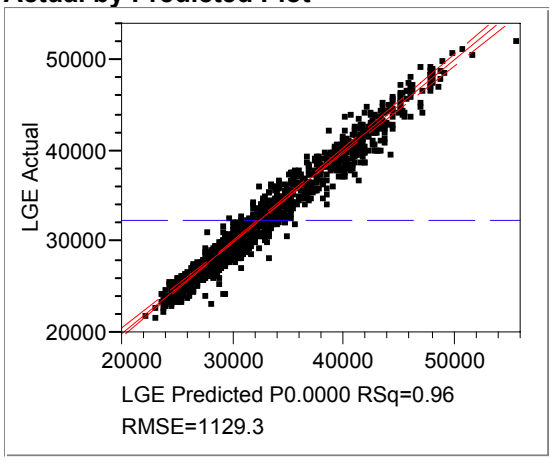
Another helpful feature of JMP is the *Save Script to Dataset* item of the Fit Model analysis. This feature saves a script of the analysis as a property of the dataset. For the simple example, the script is

```
Fit Model(Y( :LGE), Effects( :LGELevel,
:LGEAvg,  :LGEAvg *  :LGELevel),
Personality(Standard Least Squares),
Emphasis(Effect Screening), Run
Model(Profiler(Confidence Intervals(1)),
:LGE << {Scaled Estimates(1), Plot Actual
by Predicted(1), Plot Residual by Row(1)}))
```

The *lasso* tool can be used to drill down to see if outliers have any characteristics in common. This technique was used to identify the effects of months, weekends and holidays.

The final model is shown below.

Response LGE
Actual by Predicted Plot



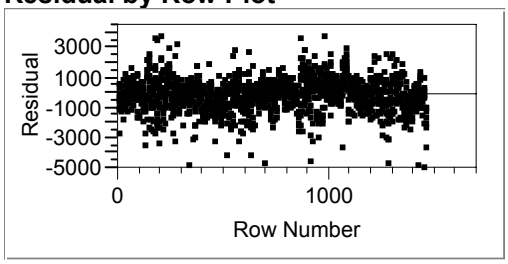
Summary of Fit

RSquare	0.964374
RSquare Adj	0.963345
Root Mean Square Error	1129.343
Mean of Response	32462.56
Observations (or Sum Wgts)	1461

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
LGEAvg	1	1	21843106.1	17.1263	<.0001
LGEAvg*LGEAvg	1	1	8993029.16	7.0511	0.0080
Holiday	1	1	779167595	610.9125	<.0001
Weekend	1	1	4108327322	3221.166	0.0000
LGEAvg*Weekend	1	1	35804599	28.0729	<.0001
MonTrend	1	1	168231171	131.9030	<.0001
LGEAvg*MonTrend	1	1	25537810.9	20.0231	<.0001
LGEAvg*LGEAvg*MonTrend	1	1	16891728.1	13.2441	0.0003
ObsMon	11	11	43685210.6	3.1138	0.0004
LGEAvg*ObsMon	11	11	254716740	18.1557	<.0001
LGEAvg*LGEAvg*ObsMon	11	11	241626624	17.2227	<.0001

Residual by Row Plot



Conclusion

JMP is not SAS, but it is a very good analysis tool with many unique features. JMP's strength is in its data visualization. Information in charts and tables are linked.

Anything highlighted in one will be highlighted in the other. This helps the analyst to better understand and analyze his/her data.

Contact Information

Any comments or experiences are appreciated. Information or questions can be directed to the author at:

Thomas A. Fuchs
LG&E Energy Corp.
P O Box 32010
Louisville, KY 40232
(502) 627-2225
tom.fuchs@lgeenergy.com