A Technical Publication for JMP® Users
Issue 21 Fall 2006

JMPer Cable

Inside This Newsletter

JMP is a business division of SAS, the leader in business intelligence and analytics. SAS software answers strategic business questions no one else can—enabling you to control costs, drive revenue, achieve capital efficiency, and lead with confidence. For more than a quarter of a century, SAS has continued an uninterrupted commitment to solving world class business problems.

jmp

Driving a Designed Experiment

By Diana Levey, SAS Institute

The world's fastest-growing sport just got faster. That's because AIM Autosport and DMH Racing are now using JMP statistical discovery software from SAS to fine-tune open-wheel race cars.

The ability to visually interact with racing data has enabled engineers to understand more clearly, and even quantify, the effect of various parameters on the car's performance. That information is empowering these engineers to prepare a car for specific track conditions. The result is optimized performance, which translates to quicker lap times.

With JMP on the team, rookie driver Daniel Herrington staked his claim to fame in just a couple of months in the Star Mazda Championship Series: a first place in Houston, a second place at Sebring and Milwaukee, and, at Mid-Ohio, the fastest single lap time in the May 2006 race. "You have to know that you're the fastest guy out there, that you have the best team, and that you have the best equipment," says the 19-year-old driver. "We do. And we have JMP software."

A Golden Opportunity

The idea to use statistical software came to Dr. David Herrington, owner of DMH Racing and Daniel's father, when he was looking for a way to help his son gain a performance advantage in such a competitive field. In addition to being a racing enthusiast, Dr. Herrington is a medical doctor with about 20 years of experience using SAS software for research.

"I recognized that this was a golden opportunity to do some statistical data analysis that's not typically done in racing," says Dr. Herrington. "A lot of factors influence a race car's performance. And many of these factors are interrelated and influence one another to produce the aggregate of the lap time. However, it's not always easy to discern the magnitude of the effect of each parameter or how each parameter might interact with another."

That's where JMP came in. Dr. Herrington starts with exploratory data analysis to get a sense of what factors are best modeled using simple linear models—and what factors have more complex relationships that should be explored with the response surface features of JMP.

"We attempted initially to model tire pressure as a function of lap time with a simple linear model, but we realized that it might be more appropriate as a response surface model," says Dr. Herrington. "The analysis indicated that there was an optimal pressure that yielded the best lap time under certain conditions, and higher pressures or lower pressures were suboptimal. We learned that even half a pound of pressure can have a big impact on the tire performance—knowing the right tire pressure is very important."

"We look at quadratic relationships, and we also do a fair amount of looking for plausible interactions," says Dr. Herrington. "For example, it's plausible to expect that the front wing setting and rear wing setting could influence each other. In those instances, we routinely build an interaction term into our models to see if there is sufficient data to confirm our suspicions."

Adds Dr. Herrington, "Using JMP, we can understand the data and the relationships among different factors in an efficient and intuitive way. The graphical nature of JMP is extremely helpful because it allows us to quickly interpret the results of the models and incorporate what we've learned into how we set up the car."

Ian Willis, Formula Mazda Team Manager and partner of AIM Autosport, the team for which Daniel drives, explains that all drivers in this type of racing have the same kind of car and are limited in how they can set up the chassis. "When you're working with spec-car racing, which is so restrictive, the performance advantage comes in the details," says Willis. "So when we use JMP for design of experiments, we're able to optimize those settings beyond any other race team."

When the car was set to the optimal settings predicted by JMP, Willis says, Daniel started cutting time off laps. "Cutting a 121-second lap time to 119 seconds is huge." Even in the preliminary laps of a race weekend, it's important for the car to be set up as perfectly as possible, says Daniel Herrington. "On the first session out, we can be quick right away instead of having to chase our tails trying to find the right setup. The predictive nature of the software allows us to start the weekend off up front—then everyone else is trying to catch you."

The team's use of JMP saves money, too. "Every piece of information about the performance of the race car costs money to obtain," says Dr. Herrington. He estimates the preparation and operation of the car costs about $60 per minute of track time. "That translates into enormous amounts of money for testing and development of the car. If we can be even 20 percent more efficient, that translates into hundreds of thousands of dollars of savings."

So Many Influential Factors

As the team gets better at predicting the best setups for its open-wheel racers, it is working to incorporate as much data into the process as possible, which includes information on the tracks themselves. In Milwaukee, for example, the track is oval, which requires a completely different setup than the road courses on which the team usually races. It makes sense to consider even differences between road courses, says Willis, because the car could be configured differently for different track characteristics, such as the number and length of straightaways or the amount of low-, medium-, and high-speed corners.

Even weather needs to be factored in, Dr. Herrington says. "We are collecting and making use of other co-variants outside of our control—but that still have an impact on our performance. Being able to add air and track temperature, wind direction and speed, and humidity to our models is very important. The fuel load and the number of miles that the tires have on them—these, too, are treated as co-variants in all of our models."

In the long term, the team will continue gathering data to create a massive database that spans racing seasons and tracks across North America. "I see that data gathering is critical," says Willis. "You know, garbage in, garbage out. People grossly underestimate how important good data-record keeping is. Every little factor counts. And nobody has really maximized how to make the best use of that data."

Of course, Willis' more than 20 years of racing experience, gut instinct, and anecdotal information will continue to be a part of the decision-making process at AIM Autosport. "This will not only verify what I think; it will quantify it," says Willis. "Instead of acting on instincts, we can act on instincts backed up by real facts."

Quote: "*Instead of acting on instincts, we can act on instincts backed up by real facts.*"

Ian Willis, Partner, AIM Autosport

Analysis of Carryover Effects

By Duane Hayes, SAS Institute

A common experimental design in clinical trials is the crossover study, where each treatment is administered one after another to all participating subjects. This design is also used in industrial and consumer research settings. The challenge with this type of design is that residual effects of one treatment can carry over to influence the effect of the next treatment (carryover effect).

An example of a crossover study based on balanced Latin Square designs is given in Cochran and Cox (1957). Using balanced Latin squares, residual effects can be estimated, and true treatment effects can be adjusted for any observed carryover effects. Littell, *et al.* (1991) show how to use SAS software to estimate and test for carryover effects, and how to estimate and test treatment effects adjusted for carryover effects for this study. This article shows how to do the same analysis using JMP.

The study in these references was conducted to determine the direct and residual (carryover) effects of three diets (treatments A, B, and C) on milk yields of dairy cows. Two balanced Latin squares were used to assign treatments to cows as shown in *Figure 1*.

An intuitive way to set up a JMP data table for this crossover design is shown in *Figure 1*. The residual effects variable (Carryover Effect) has a level for each treatment. Its values are the lag of the Treatment values, except that an additional "Null" level is assigned to the first period for each cow, where there is no carryover effect.

The interest lies in the residual carryover effects for the three treatment levels, A, B, and C, not for the "Null" level. JMP recognizes the fact that a singularity was introduced with the inclusion of the "Null" level and automatically zeros one of the parameter estimates. Because there are only two degrees of freedom given to this carryover effect, JMP's effect coding parameterization estimates the first two levels and zeros the third ordered level. The effect of the fourth level will be the negative sum of the parameter estimates for the first two levels (since the third is zeroed).

Therefore, you need to order the levels of the carryover effect such that the "Null" level is the zeroed effect.

You can use either the **List Check** or the **Value Ordering** Column Property to order column values:

- Highlight the Carryover Effect column in the data table.
- Choose **Cols** > **Column Info**.
- In the resulting Column Info dialog, click the New Property button and select **List Check**.
- Reorder the levels, as shown in *Figure 2*.

Figure 2 Use the List Check Column Property to Order the Values of Residual Effect

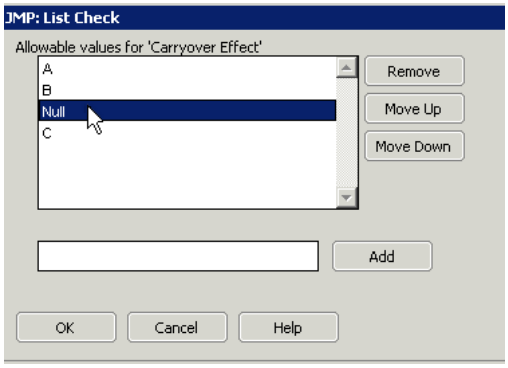


Figure 1 Schematic and Arrangement of JMP Table for Design with Carryover Effects

Schematic of Balanced Latin Squares for Crossover Experiment

Cow	1	2	3	4	5	6
Period 1	A	B	C	A	B	C
Period 2	B	C	A	C	A	B
Period 3	C	A	B	B	C	A

Crossover
JMP Carryover Model

Columns (5/0)
Cow
Period
Treatment
Carryover Effect
Milk Yield

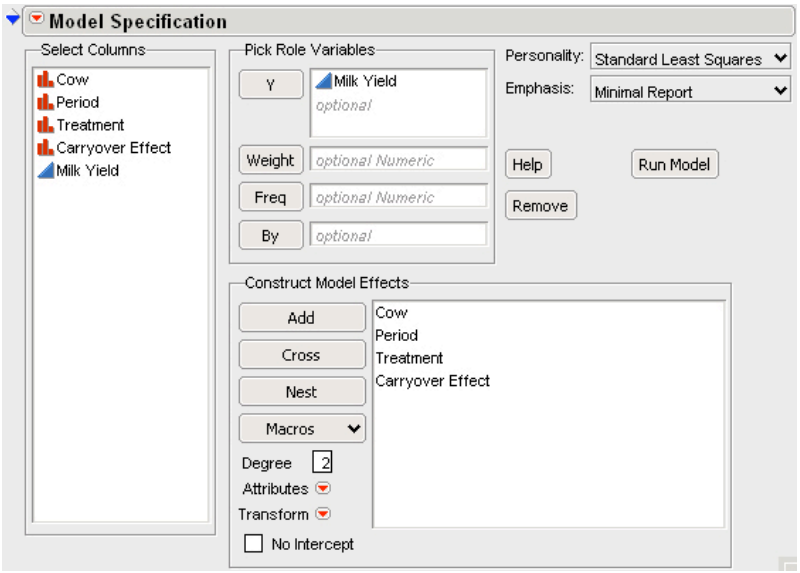
Rows
All rows 18
Selected 0
Excluded 0
Hidden 0
Labelled 0

	Cow	Period	Treatment	Carryover Effect	Milk Yield
1	1	1	A	Null	38
2	1	2	B	A	25
3	1	3	C	B	15
4	2	1	B	Null	109
5	2	2	C	B	86
6	2	3	A	C	39
7	3	1	C	Null	124
8	3	2	A	C	72
9	3	3	B	A	27
10	4	1	A	Null	86
11	4	2	C	A	76
12	4	3	B	C	46
13	5	1	B	Null	75
14	5	2	A	B	35
15	5	3	C	A	34
16	6	1	C	Null	101

To analyze this crossover design with JMP use the following steps:

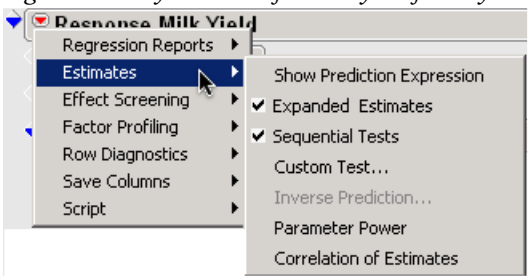
1) Select **Analyze** > **Fit Model**
2) Select the Response, Model Terms, Personality, and Emphasis as shown in *Figure 3*.
3) Run the model to generate effects tests for residual and treatment effects.
4) Select **Estimates** > **Sequential Tests** to generate ANOVA tests for subject and period effects.
5) Select **Estimates** > **Expanded Estimates** to generate estimates of residual and direct treatment effects.

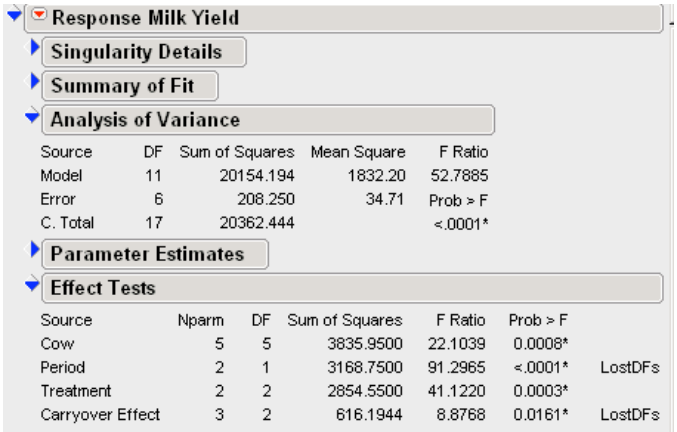
Figure 3 Fit Model Dialog for Analysis of Crossover Effects

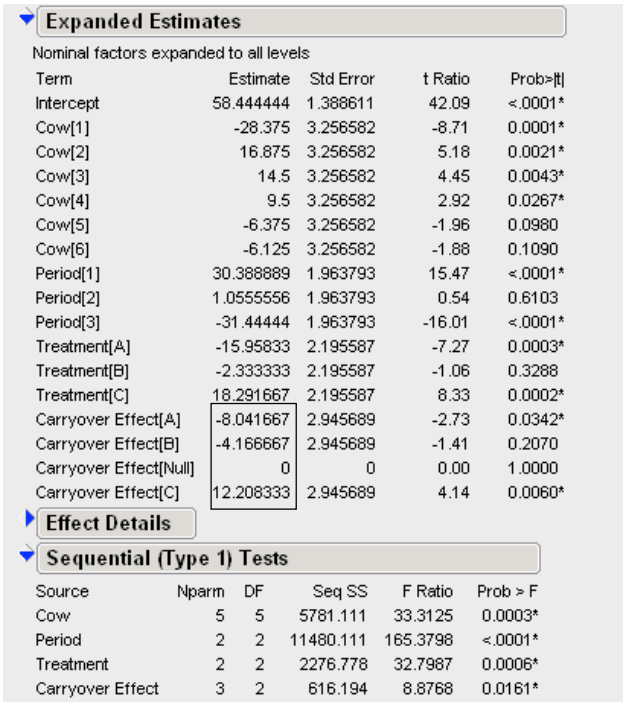


The residual effects are given by the estimates for the corresponding level of the Carryover variable. So the residual for Treatment A is –8.041667, for Treatment B it is –4.166667, and the residual effect for Treatment C is 12.208333. Estimates of the direct effect treatment means adjusted for residual effects are found by adding the expanded treatment effect estimates to the expanded model intercept.

Figure 4 Analysis Results for Analysis of Carryover Effects







As many would conclude, analysis of carryover effects is an important issue in crossover designs, and JMP Software can easily handle the assignment.

References

Cochran and Cox (1957), *Experimental Designs*, New York: J. Wiley and Sons.

Littell, Freund, Spector (1991), *SAS System for Linear Models*, 3rd Ed., Cary, NC: SAS Institute.

JMP Developer Wins Visualization Presentation Competition

Diana Levey and Ann Lehman, SAS Institute

Congratulations to **Xan Gregg**, principal systems developer on the JMP Team at SAS Institute, for his winning entries in two of five categories and for top prize as the comprehensive winner of the 2006 Data Visualization competition sponsored by the Business Intelligence Network™.

The Data Visualization Challenge

The Business Intelligence Network (*b•eye*) offers comprehensive resources for business intelligence and data warehousing professionals. Recently, the Business Intelligence Network announced the winners of its 2006 Data Visualization Competition. The competition invited contestants to submit visual representations of data that solved business problems identified by five different scenarios. Steven Few, visualization expert and author, provided the scenarios, judged all submissions, and selected the winners based on the effectiveness of their solutions.

www.b-eye-network.com

Known for its ability to present complex data in visually interesting and easy-to-understand ways, JMP statistical discovery software from SAS has been the desktop data analysis tool of choice for scientists, engineers, and quality control professionals since 1989.

Said Gregg, "They [*b•eye*] provided the raw data in Excel format, which I imported into JMP. From there, I used JMP's data manipulation and graphical exploration abilities to look for the important features of the data and experiment with various visualizations."

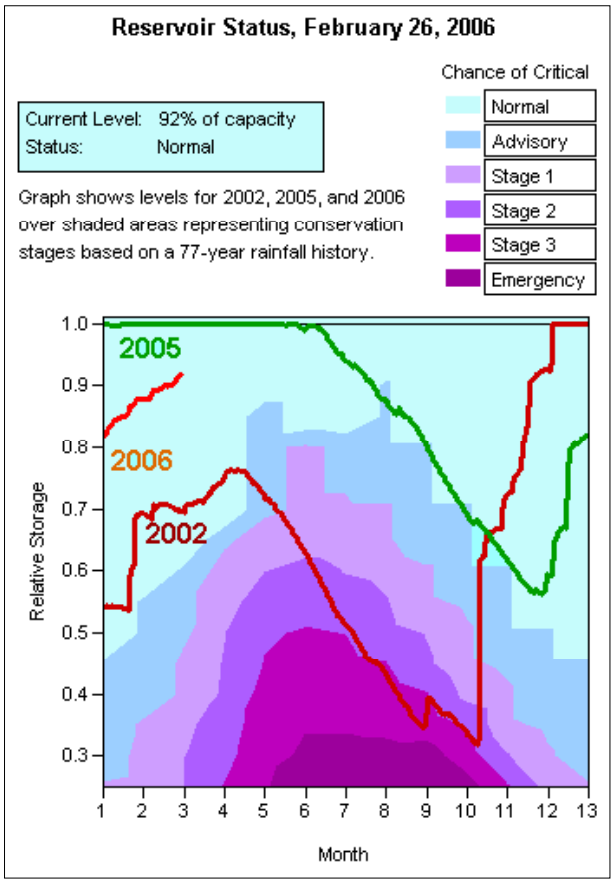
Gregg took first place for one of the business scenarios, in which contestants designed a visualization that could be used by bank customers to track their checking account activities during a single month.

Gregg also tied for first place for the freestyle scenario in which contestants were able to submit any data visualization that displayed quantitative business data in a manner that supports people in their efforts to understand the data in meaningful ways. For this scenario, Gregg used water reservoir data. The challenge was to design a visual presentation for a town's water authority to use on its Web page to show residents the current reservoir status and conservation stage, including comparisons to previous years of interest (drought years).

Figure 1 shows Gregg's winning entry produced by JMP. The graph shows levels for different years presented over shaded areas that represented conservation stages—from normal to emergency based on a 77-year rainfall history.

"The important thing is that it shows the status by month," added Gregg. "From a conservation standpoint, you can really see how a 60 percent reservoir level in June means something very different than the same level in November."

Figure 1 Water Reservoir Data Visualization Entry



Contours and Overlays in JMP

The problem uses two tables, Flow.jmp and Reservoir.jmp. These JMP tables have JSL scripts attached that generate overlay and contour plots, and are available with this issue of JMPer Cable at:

www.jmp.com/about/newsletters

The Flow.jmp has daily flow data from 1996 to present (February 28, 2006). Each day shows the relative reservoir storage (Relative Storage) where 1.0 = 100% full

and the inflow rate (Inflow) from streams in millions of gallons per day. You can see in the data table (*Figure 2*) that during October of 2002, Relative Storage was only about 33% and Inflow was zero. May of 2005 showed storage at 100% and a healthy inflow of water.

Figure 2 Example of Flow Data Table

	Year	Month Name	Day	Relative Storage	Inflow	MonthDay
2470	2002	OCT	5	0.33	0.00	10.1147541
2471	2002	OCT	6	0.33	0.00	10.147541
2472	2002	OCT	7	0.32	0.00	10.1803279
2473	2002	OCT	8	0.32	0.00	10.2131148
2474	2002	OCT	9	0.32	0.00	10.2459016
2475	2002	OCT	10	0.33	0.13	10.2786885
3432	2005	MAY	24	1.00	7.94	5.72131148
3433	2005	MAY	25	1.00	6.88	5.75409836
3434	2005	MAY	26	1.00	5.82	5.78688525
3435	2005	MAY	27	1.00	5.13	5.81967213
3436	2005	MAY	28	1.00	4.66	5.85245902
3437	2005	MAY	29	0.99	4.23	5.8852459

Note that Flow.jmp contains daily data but the final contour plot in *Figure 1* shows monthly Relative Storage. To make the two tables compatible for plotting together, the Flow table has a computed variable, MonthDay, that maps the daily information into an appropriate month value.

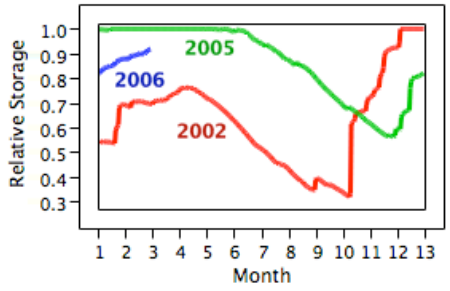
The MonthDay variable approximates a month value using a column formula. The formula, shown below, creates the JMP date (number of seconds since January 1, 1904) for each observation, divides it by 30.5 (rough approximation of number of days in a month) and adds 1, to give monthly categories numbered 1 to 13. The Reservoir.jmp table has monthly data plus an extra "End" category, giving 13 monthly categories to plot.

Day Of Year

$$\frac{\left[\text{Informat}\left(\text{Char}\left(\textit{Day}_{\_}\right) \,||\, \textit{Month Name} \,||\, \text{Char}\left(\textit{Year}_{\_}\right), \text{"ddMonYYYY"}\right)\right]}{30.5} + 1$$

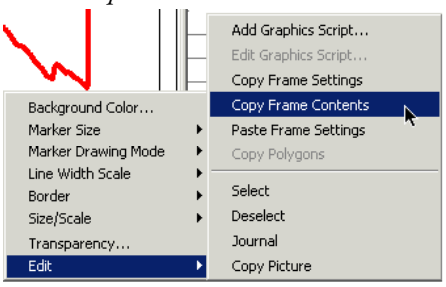
Figure 3 is an overlay plot of monthly Relative Storage for 2002, 2005, and 2006, created by the Overlay Platform. To complete the competition entry, these lines need to be superimposed on the contour plot of conservation stages.

Figure 3 Overlay of Relative Storage for Three Years



The Reservoir.jmp table was used to make the initial contour plot in *Figure 1* that shows conservation stages. To combine the overlay plot (*Figure 3*) and the contour plot, right-click in a graphics frame to see a context menu with an **Edit** submenu (*Figure 4*). Use the **Copy Frame Contents** (in the overlay plot) and **Paste Frame Contents** (in the contour plot) to superimpose one graph on another.

Figure 4 Right-Click in Plot Frame for Additional Commands and Options



Final Remarks

On some entries, Gregg submitted the graphs just as they were produced in JMP. In other cases, he used additional tools to fine-tune the layout of the graphs.

Gregg noted that the experience of creating these business visualizations has provided firsthand knowledge that is being used to improve the data visualization features of JMP, including, for instance, better legends.

Gregg's reservoir data visualization and checking account scenario presentations will be featured in upcoming issues of the data visualization edition of the *Business Intelligence Newsletter* as well as in the "Articles" section of Stephen Few's expert channel on the Business Intelligence Network web site.

Few says, "Data visualization extends our ability to make sense of and communicate business information. It provides the means to fulfill the promise of business intelligence (BI), but only if BI software vendors develop the tools and business people develop the skills needed to apply it effectively. The winners of the Business Intelligence Network's 2006 Data Visualization Competition show us what data visualization can achieve when it's done right. I hope this competition and the follow-up articles that feature the winners will help to expand the reach of effective data visualization beyond the domain of a few to 'business as usual' for many."

www.b-eye-network.com

b·eye Business Intelligence Network™ The Vision for BI and Beyond

Summer Time

Mark Bailey, SAS Statistical Training and Technical Services

During the cold and wet spring time in the northeast United States, everyone awaits the warm, dry days of summer. Summer officially arrives on June 21, but why is this day the start of summer? The data in *Figure 1* contains the time of day that the sun rises and sets for a four-week period around the start of summer.

The Sun Rise and Sun Set information is stored as the number of seconds since midnight of the date showing in the Date column, and displayed using the h:m:s time format. Since these values are just numbers of seconds, the Length of Day is found by subtracting them using a column formula. This value is displayed using the h:m:s duration format.

Figure 1 Sunrise and Sunset Times

	Date	Sun Rise	Sun Set	Length of Day
1	06/07/2003	5:31:45 AM	8:26:55 PM	14:55:10
2	06/08/2003	5:31:32 AM	8:27:30 PM	14:55:58
3	06/09/2003	5:31:22 AM	8:28:04 PM	14:56:42
4	06/10/2003	5:31:13 AM	8:28:36 PM	14:57:23
5	06/11/2003	5:31:06 AM	8:29:07 PM	14:58:01
6	06/12/2003	5:31:01 AM	8:29:37 PM	14:58:36
7	06/13/2003	5:30:59 AM	8:30:04 PM	14:59:05
8	06/14/2003	5:30:58 AM	8:30:30 PM	14:59:32
9	06/15/2003	5:30:59 AM	8:30:55 PM	14:59:56
10	06/16/2003	5:31:01 AM	8:31:18 PM	15:00:17
11	06/17/2003	5:31:06 AM	8:31:39 PM	15:00:33
12	06/18/2003	5:31:13 AM	8:31:58 PM	15:00:45
13	06/19/2003	5:31:21 AM	8:32:16 PM	15:00:55
14	06/20/2003	5:31:31 AM	8:32:32 PM	15:01:01
15	06/21/2003	5:31:43 AM	8:32:46 PM	15:01:03
16	06/22/2003	5:31:57 AM	8:32:58 PM	15:01:01
17	06/23/2003	5:32:13 AM	8:33:09 PM	15:00:56
18	06/24/2003	5:32:30 AM	8:33:18 PM	15:00:48
19	06/25/2003	5:32:49 AM	8:33:24 PM	15:00:35
20	06/26/2003	5:33:10 AM	8:33:29 PM	15:00:19
21	06/27/2003	5:33:32 AM	8:33:31 PM	14:59:59
22	06/28/2003	5:33:56 AM	8:33:33 PM	14:59:37
23	06/29/2003	5:34:21 AM	8:33:31 PM	14:59:10
24	06/30/2003	5:34:48 AM	8:33:28 PM	14:58:40
25	07/01/2003	5:35:17 AM	8:33:24 PM	14:58:07
26	07/02/2003	5:35:46 AM	8:33:16 PM	14:57:30
27	07/03/2003	5:36:18 AM	8:33:07 PM	14:56:49
28	07/04/2003	5:36:51 AM	8:32:57 PM	14:56:06

You can see that the length of day increases as the date approaches June 21 and then begins to decrease. It appears that spring concludes and summer begins on June 21, which is the longest day of the year. The first day of summer is also known as the summer solstice. (The winter solstice is the shortest day of the year.)

You might guess that the longest day occurs when the sun rises at the earliest time and sets at the latest time. However, the flaw in that guess is the assumption that the earliest sunrise and the latest sunset occur on the same day. Inspecting the data table again shows that the earliest sunrise occurs on June 14th, at 5:30:58 AM, seven days before the solstice. The latest sunset occurs on June 28th at 8:33:33 PM, which is seven days after the solstice. The solstice, the day with the greatest difference between sunrise and sunset, is June 21.

The data can be found on the Web page for this issue of JMPer Cable:

www.jmp.com/about/newsletters

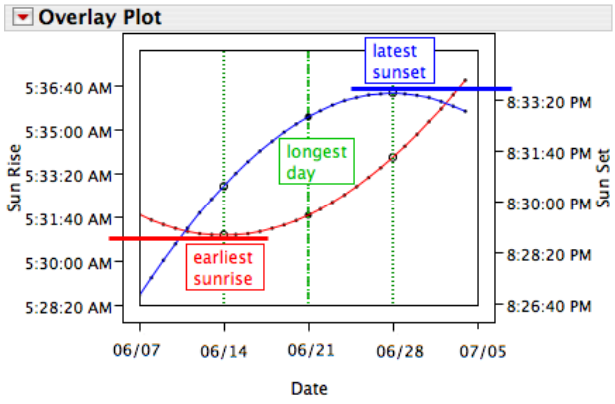
It is easier to see these relationships with a graph of sunrise and sunset data. Use an overlay plot with the Sun Rise on the left Y axis, Sun Set on the right Y axis, and Date on the horizontal axis. This way you can see them together for comparison but separate axes show them each with an optimal individual scale.

The following steps show how to create this plot.

- Choose **Graph** > **Overlay Plot**.
- Select Sun Rise and Sun Set, then click **Y**.
- Select Sun Set in the **Y** list and click Left Scale/Right Scale.
- Click **OK** to see the overlay plot in *Figure 2*.

The plot clearly shows that the peak in the sunset and the trough in the sunrise do not coincide with the solstice. Instead, the minimum sunrise leads the solstice by one week and the maximum sunset lags by one week. The vertical displacement of the two curves is the length of the day. The greatest vertical displacement occurs on June 21. (The actual length of day would have to include the offset between the left and right scales.)

Figure 2 Plot of Sun Rise and Sun Set Times by Date



The reason for the offset of sunrise and sunset seen in *Figure 2* is due to the earth's elliptical orbit around the sun and the tilt of the earth's axis (obliquity) with respect to its orbit. The earth's elliptical orbit and tilt of the axis causes the sun to be at a different position horizontally each day, while the tilt of the earth's axis also causes the sun to be at a different position vertically.

Reference:

Data computed with Sol! II, version 5.2.0 (Palm OS), by M. Edward Wilborne III, Star Pilot Technologies

JMP Across the Border to Canada

Carl J. Schwarz and Ian J. Bercovitz
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada V5A 1S6

JMP is used at Simon Fraser University (SFU), Canada, in roles ranging from undergraduate education, statistical consulting, and outreach to scientists outside the university. This article describes the use of JMP in four roles and gives links to public domain tutorial and course notes using JMP at the introductory and intermediate level.

JMP Conquers Fear and Loathing

To many students, a course in statistics is rarely a course that is eagerly anticipated. Rather, statistics is often synonymous with "sadistics." This is particularly true for new students starting at Simon Fraser University who must complete at least two courses in both quantitative reasoning and writing skills. For many non-science students, at least one of the quantitative courses is in statistics. Students can be overwhelmed by the mechanical computational details that so often occupy introductory courses in statistics. They cannot see the big ideas because of the computational hurdles. Carl Schwarz, of the Department of Statistics and Actuarial Science at SFU, is using JMP to give these students a more intuitive, visually rich statistical experience.

www.stat.sfu.ca

Carl Schwarz (left) and Student in Statistics Lab

One of the introductory courses in the department is Stat-201, Statistics for the Life Sciences

www.stat.sfu.ca/~cschwarz/Stat-201

Here, Schwarz uses JMP as the package of choice for several reasons. First, says Schwarz, "I can get students up and running JMP in very short order—often 30 minutes or less." Second, students appreciate the "intuitive approach to analyses featured in JMP." Third, the "automatic, high-quality graphics produced by most platforms in JMP make it easier to understand the results." Also, JMP is available for Macintosh, Windows, and Linux operating systems via a site license at SFU so students can quickly download and install the software for their courses.

JMP For More Advanced Topics

But the advantages of JMP don't stop at introductory courses. Indeed, one of the reasons that Schwarz uses JMP is the powerful analytics available for more advanced courses.

For example, most biology, environmental science, and resource managers take a three-course sequence in statistics starting with Stat-201 (the introductory course above), Stat-302 (ordinary, logistic, and poisson regression), and Stat-403 (more advanced experimental design and analysis). This course sequence again relies upon JMP to remove the computational burden and concentrates on the interpretation of results.

Schwarz has an extensive set of course notes that demonstrate the use of JMP (and/or SAS) in the analysis of these intermediate topics. These notes include screen shots of the JMP steps and annotated examples of output. These pages are always available to anyone with Web access.

www.stat.sfu.ca/~cschwarz/CourseNotes.html

These courses take advantage of the many powerful analytics of JMP. For example, it is rare for courses at this level to do much with power analyses in the design of experiments, but JMP makes this easy with its **DOE > Sample Size and Power** platform.

Using JMP with Clients of the Statistical Consulting Service

The department also runs a successful Statistical Consulting Service dealing with research projects both within and external to the SFU campus. "We are essentially running at full capacity," according to the director, Ian Bercovitz, "and JMP is our tool of choice for many clients." You can find out more at

www.stat.sfu.ca/people/consulting

"Ten years ago, we would get clients who needed help in doing the computations of a two-sample *t*-test. Now they need help not with statistical computations, but help in understanding the output from packages or help with non-standard problems such as how to handle extensive missing values or analysis of an unusual experimental design. Ironically, the goal of our consulting service is to avoid repeat business by showing the clients how to analyze their own data quickly and efficiently."

"Clients are amazed with the capabilities of JMP and appreciate its powerful and intuitive interface. It is

essential that a friendly-to-use software package such as JMP is available because many clients won't return to a repeat analysis until many months down the road."

Ian Berkovitz (right) and Client

Our approach has been successful, and we find plenty of repeat customers who are quickly building analytical capabilities and are coming back with more and more advanced problems. For example, clients now arrive suspecting they need some sort of generalized linear modeling, and JMP 6's generalized linear models platform allows clients to do these analyses themselves with expert guidance from our service.

Using JMP Outside the University

The department also has an extremely active outreach program. For example, the Government of British Columbia approached the department (through the Co-operative Research Management Institute) for help with continuing education for its regional wildlife offices. A three-day short course has been offered 20 times in the last two years to wildlife biologists and resource managers. The course reviews many of the topics in intermediate statistics, such as regression, survey sampling, and experimental design. The demo version of JMP has been used for these training sessions because we can get this installed quickly on everyone's machines and everyone has access to the same facilities.

JMP is quickly becoming the package of choice for these professionals in their work force.

If you would like further information, please don't hesitate to contact:

Carl Schwartz
Department of Statistics and Actuarial Science
Simon Fraser University
Burnaby, BC, Canada
cschwarz@stat.sfu.ca

Simon Fraser University

www.sfu.ca

Upcoming JMP Events

JMP is scheduled to attend the following conferences.

October 15-17, 2006
SCSUG Irving, TX

October 22-24, 2006
MWSUG Dearborn, MI

October 25-27, 2006
IQPC European Lean Six Sigma Summit
Amsterdam, The Netherlands

October 29-31, 2006
PNWSUG Seaside, OR

October 30-31, 2006
California Advancement Researchers Association (CARA) Costa Mesa, CA

October 31-November 1, 2006
NC Conference for Excellence: Battling for Tomorrow's Business Greensboro, NC

November 4-8, 2006
INFORMS Annual Meeting Pittsburgh, PA

November 14-15, 2006
7th Annual National Six Sigma in Healthcare Forum Orlando, FL

US Webinars

October 13, 2006
JMP Software for Microarray, SNP and Proteomics Expression Analysis JMP for Genomics
1 PM EDT (10 AM PDT)

October 16, 2006
Analyzing High-Throughput Data with JMP Genomics JMP for Genomics
10 AM EDT (7 AM PDT)

October 20, 2006
Getting a JMP Start JMP
1 PM EDT (10 AM PDT)

October 20, 2006
Analyzing High-Throughput Data with JMP Genomics JMP for Genomics
3 PM EDT (12 PM PDT)

October 27, 2006
Getting a JMP Start JMP
1 PM EDT (10 AM PDT)

November 3, 2006
Getting a JMP Start JMP
1 PM EDT (10 AM PDT)

A Script Quickie: Rename Multiple Columns

Jeff Perkinson, SAS Institute

Cleaning and reorganizing data are often a large part of pre-analysis, especially for survey data. For example, suppose a survey has many questions of the type shown here:

Please check **Yes** or **No** for the following questions:

QUESTION	YES	NO
Q1. I have a great job.	☐	☐
Q2. My company treats employees well.	☐	☐
Q3. Benefits are a major concern.	☐	☐
Q4. Salaries are not competitive.	☐	☐

•
•
•

For convenience, the result of each question was entered into a JMP table with columns named Q1, Q2, and so forth *(see Figure 1)*.

It would be helpful if the questions were labeled with an informative phrase on analysis results. For this to happen, the columns need to have names that reflect the content of the questions.

Figure 1 Survey Example.jmp Data Table

Survey Example

Survey Example

Columns (4/0)
Q1 ✱
Q2 ✱

Rows
All rows 20
Selected 0

	Q1	Q2	Q3	Q4
1	Yes	No	Yes	No
2	Yes	Yes	No	No
3	No	No	Yes	Yes
4	No	Yes	No	Yes
5	Yes	No	Yes	Yes
6	No	Yes	Yes	No
7	No	No	Yes	Yes

One way to modify column names requires a JMP table with columns that contain the original names (Q1, Q2, etc.) and the names you want to use in the analysis. Then, a simple script can replace the column names in original data table with the new names. The table needed to update the original table looks like the one in *Figure 2*.

Figure 2 Data Table List of Old and New Column Names

New Names

New Names

Columns (2/0)
Q names
long name

Rows

	Q names	long name
1	Q1	I have a great job
2	Q2	My company treats employees well
3	Q3	Benefits are a major concern
4	Q4	Salaries are not competitive

The short script shown below replaces the 'Q' column names in the original table with the long name values.

```
x=datatable("New Names");
y=datatable("Survey Example");

current data table(x);
newcolnames=:long name<<get values;
oldcolnames=:Q names<<get values;

current data table(y);
ni = nitems(newcolnames);
for(I = 1, I <= ni, i++,
      eval(
         substitute(
               expr(column(xxx)<<setname(yyy)),
               expr(xxx),oldcolnames[i],
               expr(yyy),newcolnames[i]
               )
         )
   );
```

Figure 3 Survey Example.jmp with New Column Names

	I have a great job	My company treats employees well	Benefits are a major concern	Salaries are not competitive
1	Yes	No	Yes	No
2	Yes	Yes	No	No
3	No	No	Yes	Yes
4	No	Yes	No	Yes
5	Yes	No	Yes	Yes
6	No	Yes	Yes	No
7	No	No	Yes	Yes
8	Yes	Yes	Yes	No

Note: The list of question names in the New Names table does not have to be in the same order as the columns in the survey data table. The Substitute command finds each column name in the list of old names and replaces it in the data table with the corresponding name (new name) from the list of new names.

Importing Data With Column Labels From SAS

Column names in a SAS data set can have descriptive labels up to 40 bytes long. In SAS, survey data like the data described in this article could have the questions recorded as column name labels. If you read the SAS data from JMP, those labels are automatically used as JMP column names. JMP column names can be as long as you want, within practical limits.

TIPS & TECHNIQUES

Summarizing Survey Data with the Tabulate Platform

The article on page 10, "Rename Multiple Columns," shows how to use long, descriptive column names. These kinds of column names are especially helpful with survey data, where the names can reflect the survey questions. Naming columns this way also lends itself to concise summarization using the Tabulate platform.

Suppose you have survey data similar to that shown in *Figure 1*. There are many questions with yes/no answers and you want to see counts and percents for each question. The Tabulate platform can produce a report that lists the questions and the number of yes and no answers for each.

Figure 1 Survey Results with Descriptive Column Names

Employee Survey.jmp

Columns (26/0)

- Division Code
- I feel free to give my opinions
- I am sometimes unhappy with assignments
- I am frustrated by micromanagement
- Schedules in my work group are reasonable
- I have adequate support from management
- My job responsibilities are not made clear
- Work groups are too large to be effective
- My job is frequently tedious and boring
- Morale at my company is generally high
- Employees are given adequate time off
- My manager explains things clearly
- Creativity is not a plus at my company
- Company policies are enforced fairly
- I do not trust upper management
- My performance is evaluated fairly
- The mission of my company is a mystery
- My work group is self-motivated
- My job interferes with my personal life
- People I work with bicker and backstab
- I do not understand company goals
- I get unnecessary information daily
- My ideas are often implemented
- I believe I should be paid more money
- My manager listens to what I have to say
- Some in my work group are laggards

	I feel free to give my opinions
1	Yes
2	Yes
3	No
4	No
5	Yes
6	No
7	No
8	Yes
9	No
10	Yes
11	Yes
12	Yes
13	Yes
14	Yes
15	No
16	Yes
17	Yes
18	Yes
19	Yes
20	Yes
21	Yes
22	Yes
23	Yes
24	Yes
25	No
26	Yes
27	Yes

To produce a Tabulate summary table,

- Choose **Table** > **Tabulate**.
- In the Tabulate Control Panel, select all the columns (questions).
- Drag the columns to Drop zone for rows drop area (see *Figure 2*).
- When you drop the columns, a context menu appears. Choose **Add Columns by Categories**.
- Drag **% of Total** statistic next to **N** at the top right of the table.

The Tabulate report in *Figure 3* now appears.

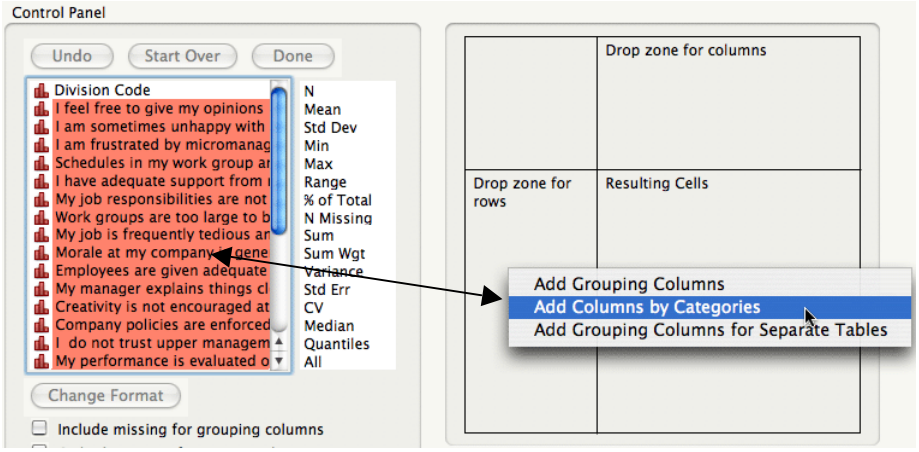
The Annotate tool was used to place titles at the top of the report.

Figure 3 Tabulate Report of Survey Results

Results of Employee Survey

	N		% of Total	
Question	No	Yes	No	Yes
I feel free to give my opinions	113	235	32.01	66.57
I am sometimes unhappy with assignments	271	75	76.77	21.25
I am frustrated by micromanagement	259	87	73.37	24.65
Schedules in my work group are reasonable	126	221	35.69	62.61
I have adequate support from management	122	221	34.56	62.61
My job responsibilities are not made clear	87	253	24.65	71.67
Work groups are too large to be effective	200	144	56.66	40.79
My job is frequently tedious and boring	303	38	85.84	10.76
Morale at my company is generally high	126	215	35.69	60.91
Employees are given adequate time off	261	83	73.94	23.51
My manager explains things clearly	99	245	28.05	69.41
Creativity is not a plus at my company	90	250	25.50	70.82
Company policies are enforced fairly	90	251	25.50	71.10
I do not trust upper management	318	21	90.08	5.95
My performance is evaluated fairly	126	211	35.69	59.77
The mission of my company is a mystery	85	253	24.08	71.67
My work group is self-motivated	94	243	26.63	68.84
My job interferes with my personal life	295	43	83.57	12.18
People I work with bicker and backstab	43	295	12.18	83.57
I do not understand company goals	62	274	17.56	77.62
I get unnecessary information daily	233	103	66.01	29.18
My ideas are often implemented	293	43	83.00	12.18
I believe I should be paid more money	130	207	36.83	58.64
My manager listens to what I have to say	56	281	15.86	79.60
Some in my work group are laggards	288	48	81.59	13.60

Figure 2 Drag Selected Columns to Drop zone for rows *area and select* **Add Columns by Categories**



JMP
SAS Campus Drive
Cary, NC 27513 USA
Tel: (919) 677-8000

About JMPer Cable®
Issue 20 Summer 2006

JMPer Cable is mailed to JMP® users who are registered users with SAS Institute. It is also available online at www.jmp.com

Contributors
Mark Bailey, Duane Hayes, Diana Levey, Jeff Perkinson, Carl Schwarz and Ian Bercovitz

Editing and Layout
Anne Bullard, Melanie Drake

Editor
Ann Lehman

Printing
SAS Institute Print Center

Questions, comments, or for more information about JMP, call
1-877-594-6567
or visit us online at
www.jmp.com

To Order JMP Software
1-877-594-6567