A Technical Publication for JMP® Users
Winter 2003 Issue 11

JMPer Cable

Inside This Newsletter

Comparing Intra-Laboratory Assays p. 2

How Are You Using JMP? p. 5

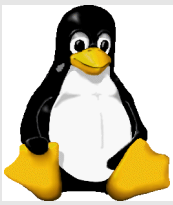
Contrasts and Custom Tests p. 6

Arranging Character Values p. 9

IT Provides a Little Animal Magic p. 10

JMP Named to 'Top 5 Products of 2002' List p. 11

Breaking News



JMP for Linux is coming! www.jmp.com/linux

Web Classes for JMP 5

Our live Web classes specifically for JMP 5 customers encourage you to interact with a JMP expert, learn and apply new JMP software skills, and receive feedback. The next scheduled class dates are Mon., April 28 and Fri., June 13 from 1-4 pm EDT. For more dates and details, see http://support.sas.com/courses/lwjgs5.html.

Working with a new SAS Solution

John Sall, Executive Vice President, SAS Institute

JMP has recently accepted a participating role as a visualization client component in the SAS Microarray solution. Microarrays (gene chips) are slides that have thousands of spots, each spot printed with a fragment of DNA ready to bind to expressed DNA from a sample.

For example, one recent study had 9600 genes on the chip, and several dozen chips. The model for this is much too big for JMP, which holds all the analyses in memory. For SAS this is fairly easy, running it as 9000 By-groups. The SAS Proc Mixed output is then stored by SAS ODS (Output Delivery System). To print the analysis would require a small forest of trees to supply the paper, and no one wants to see all this output. The values of interest are the 600,000 (BLUP) estimates across the levels and genes. To solve this problem, the values are extracted into a SAS data sets by ODS. Then JMP reads these results and provides exploratory visualization.

So what are the issues in JMP when looking at tens of thousands or hundreds of thousands of points?

- Speed—First we worked on speed. JMP was speedy enough for most users, but bogged down for huge problems such as microarray data. We knew we could image fast in the spinning plot because it is imaged many times per second. So we adapted this technology in other plots so that they could image more than a million points in one second.
- Point Density—The resulting plots were so heavily laden with points that they appeared as a black cloud. You couldn't tell which areas were denser than others. So we developed the transparency option, where the points are semitransparent, and darken as they are overlaid many times. Areas with more overlaid points appear as darker areas.
- Presentation Graphics—Typically, after initial visualization, the data is reduced to the genes that represent strong differences. This much smaller set needs to be imaged in a high-quality presentation suitable for publication. So we developed a marker option called Framed Points that surrounds each point by a black enclosure.
- Clustering—Clustering is important in gene expression studies. The goal is to use co-expression patterns to provide clues as to which genes are part of the same pathways of expression. In Version 5.0 of JMP we had improved the speed of

hierarchical clustering so that it could handle data sets up to 10,000 rows (provided you had sufficient). Although it is now feasible to cluster this many rows, it is difficult to see details in a dendrogram with many thousands of branches. We improved this by adding support for a magnifier to the dendrogram so that you can zoom in to an area of interest and see it expanded enough to read the gene names.

K-Means clustering has been improved too with the addition of parallel coordinate plots for each cluster.

- Parallel Coordinate Plots—Parallel coordinate plots and heat maps are also provided as stand-alone platforms for exploring results graphically.
- Data Tables—The data table has also been improved for use in the Microarray solution. When an experimental level participates in a number of points you want to find all the points that are similar. When you select cells of interest, the Select Similar command finds and selects all cells with the same values.

All of this is part of JMP 5.0.2, which is tied to the SAS Microarray Solution, but not part of the current general release. However these features will be part of the JMP 5.1.

The result of this effort is a fluid environment that handles larger problems gracefully so that researchers can get better insight into their data.

From the Trainer

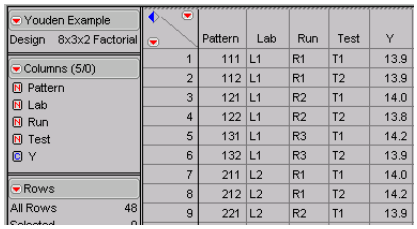
Comparing Intra-Laboratory Assays

Mark Bailey, SAS Statistical Training & Technical Services

A common task for analytical, clinical, and medical laboratories is checking for consistency in assays. This procedure involves testing duplicate control or standard samples in each of several runs in each of several laboratories. (Some readers may recognize this study as a nested design.) The runs may span several days.

For example, many clinical laboratories subscribe to a national survey[1]. This program distributes samples to laboratories around the United States and collects and compiles the test results to compare accuracy and precision. Participants are informed of where they place compared to the rest of the population of subscribers.

You can set up your own study using JMP's full factorial design. Figure 1 shows part of the table that was created in this way for the following hypothetical example.



	Pattern	Lab	Run	Test	Y
1	111	L1	R1	T1	13.9
2	112	L1	R1	T2	13.9
3	121	L1	R2	T1	14.0
4	122	L1	R2	T2	13.8
5	131	L1	R3	T1	14.2
6	132	L1	R3	T2	13.9
7	211	L2	R1	T1	14.0
8	212	L2	R1	T2	14.2
9	221	L2	R2	T1	13.9

Figure 1: Factorial Design for Assay Variability

The assay in our example is hemoglobin measured in units of grams per decilitre (g/dl). The sample was tested twice (T1, T2) in each of three assay runs (R1-R3) conducted in parallel in each of eight laboratories (L1-L8). The sample assay is recorded in Y. The level for this sample is known to be 14 g/dl.

Youden Plot

What is the best way to understand the variation and potential inconsistency in these assays? There are several graphs that display the data in a way that helps to assess the accuracy and precision of the assay. One such graph is the Youden plot[2]. This graph is not built into JMP but it is easily constructed.

The Youden plot visually separates the within-laboratory variability from the between-laboratory variability. In this plot:

- Duplicated assays form a pair of coordinates (T1, T2) in a scatterplot.
- Points are plotted with the laboratory designator.

You can form these coordinates in JMP by following these steps:

1. Select **Tables > Split**.

2. Select Y as the **Split Columns** variable, Test as the **Split Label Col** and both Lab and Run as grouping variables, as shown in Figure 2. Name the new table Youden Pairs.

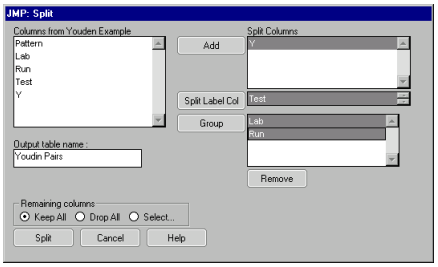
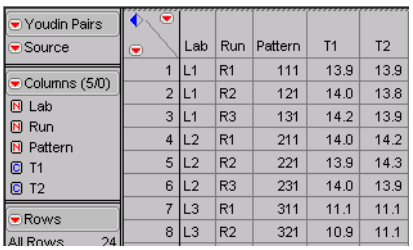


Figure 2: Split Columns Window

3. Click the **Split** button, and the new table appears. The values of the Y column split to form two columns, and the original 48 rows combine to form 24 rows (Figure 3).



	Lab	Run	Pattern	T1	T2
1	L1	R1	111	13.9	13.9
2	L1	R2	121	14.0	13.8
3	L1	R3	131	14.2	13.9
4	L2	R1	211	14.0	14.2
5	L2	R2	221	13.9	14.3
6	L2	R3	231	14.0	13.9
7	L3	R1	311	11.1	11.1
8	L3	R2	321	10.9	11.1

Figure 3. Tests in Separate Columns

Next, compare the two paired test values. To do this:

1. Select **Analyze > Fit Y by X**.
2. Assign T1 as *y* and T2 as *x*, as shown in Figure 4.
3. Click **OK** to see the Youden plot.

Enhance the Plot with Labels and Color

You can enhance the standard Youden plot so it looks like Figure 5. By doing this, you will be able to understand the structure of the study just by viewing the enhanced graph. You can:

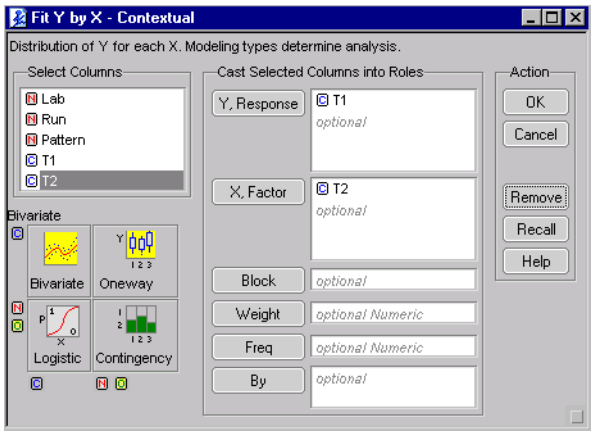


Figure 4. Fit Y by X Launch Dialog

- Use the label variable, Lab, to identify points.
- Indicate the runs with different colors.
- Add vertical and horizontal reference lines at 14 to show the location of the known value.
- Add a medium gray, diagonal identity line with a graphics script to show agreement between duplicate assays.

To label the points:

1. Select **Rows > Row Selection > Select All Rows**.
2. Select **Rows > Label/Unlabel**.
3. Click the upper-left corner of the data table to deselect all the rows.

To color the points:

1. Select **Rows > Color or Mark by Column**.
2. Choose the Lab variable in the dialog that appears and click **OK**.

To add reference lines:

1. Right-click the vertical axis and select **Axis Settings**.
2. Type 14 in the text box to the left of the **Add Ref Line** button, then click the **Add Ref Line** button. Click **OK**.
3. Repeat for the horizontal axis.

To add an identity line to the plot:

1. Right-click the graph and select **Add Graphics Script**.
2. Enter these script commands and click **OK**:

```
Pen Color(1); Y Function(x,x)
```

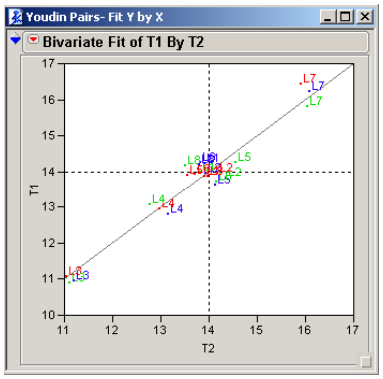


Figure 5: Enhanced Youden Plot for Assay Data

Interpret the results

By looking at the Youden plot, you can see where testing has gone astray. All of the data should cluster at the point where the reference and identity lines cross. From Figure 5 you can see that in this sample:

- Labs 3 and 4 are consistently low by about 3 and 1 g/dl, respectively (inaccurate).

- Lab 7 is consistently high by about 2 g/dl (inaccurate).
- Labs 1, 2, 5, 6 and 8 cluster in the way that indicates that these laboratories are generally accurate.

The inaccuracies suggest that three labs have a calibration or reagent problem. To see this more clearly, zoom in on the center of the graph. To do this, select the magnifier tool and click once where the reference and identity lines cross.

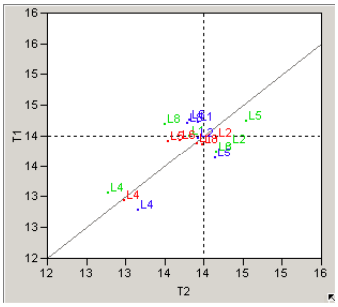


Figure 6: Zooming on Youden Plot Center

Separation of the points perpendicular to the diagonal line indicates consistent runs but inconsistent duplicates. Vertical or horizontal separation indicates inconsistent runs. With the new ordinate and abscissa ranges you can see that:

- The variability between runs for Lab 5 is large.
- The variability between duplicate assays for Labs 5 and 8 is large.

Variability Charts

Another method specifically designed to address different sources of variability is to use a variability chart. The variability chart presents a hierarchical display of the data and summary statistics. To generate a variability chart of the example table:

1. Return to the original data table (before the Y column split).
2. Select **Graph > Variability Chart**.
3. In the window that appears, use Y as **Y, Response**, and use Lab and Run as **X, Grouping**.
4. On the chart that appears, click the red triangle icon on the title bar and select **Connect Cell Means**.
5. Click the red triangle icon on the title bar again and **Show Group Means**.
6. Double-click the vertical axis of the top graph and add a gray reference line at 14.

Figure 7 shows the resulting Variability and Standard Deviation charts. Notice the hierarchical grouping of the data that facilitates finding patterns and anomalies. You can see that

- Labs 1, 2, 5, 6 and 8 are accurate on the average but Labs 3, 4 and 7 are not, as noted before from the Youden plot.
- Labs 5, 7 and 8 exhibit noticeable variability between runs. The variability within runs (between duplicate assays) is indicated by the range bars for each run or plotted in the Standard Deviation Chart.

Variance Components

To estimate the amount of variability attributed to Lab, Run and Test, click the red triangle icon in the title bar and select **Variance Components > Nested**. The factors are nested because the runs are unique to each laboratory.

Figure 8 shows the window that appears. The first part of the report is an analysis of variance, which is used to determine if the sources are significant. In this case, Lab is a significant source of variability. The second part of the report includes the variance component and the

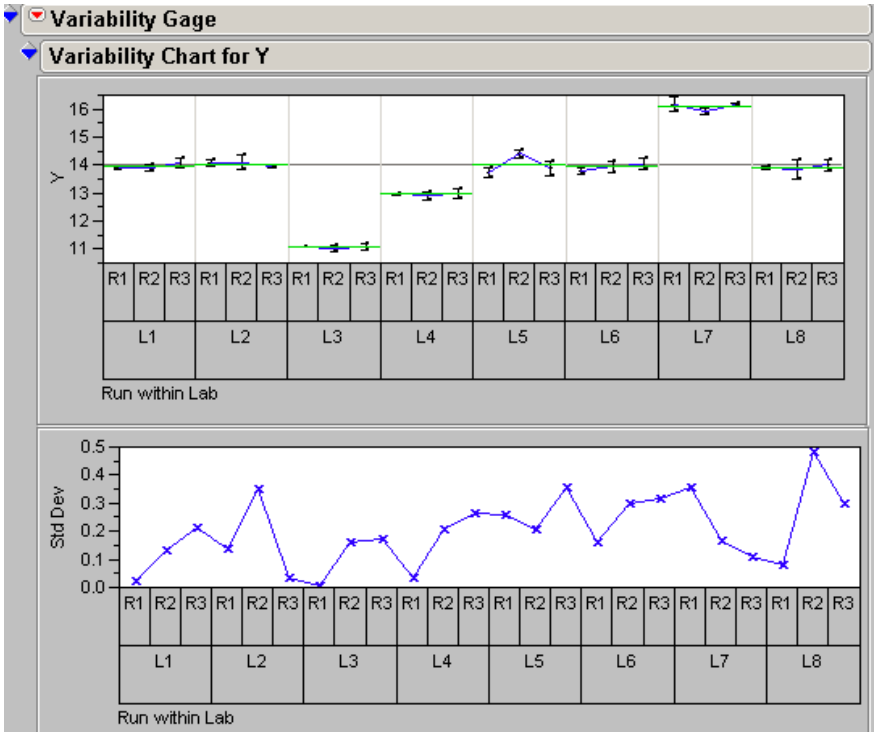


Figure 7: Variability and Standard Deviation Charts

corresponding percent of the total variability. This analysis indicates that between laboratory variability is about 97% of the total.

Scripting

If so desired, you can create a short JMP Scripting Language (JSL) script to build the Youden plot for this data, as shown in Figure 9. The script can be saved as a table property so the data table can serve as a template for a variability study.

Conclusion

This example highlights some important JMP tools you can use to analyze sources of variation in analytical test methods.

You can produce the Youden plot easily with the commands in the **Tables** and **Rows** menus. Or, you can create a short script.

Additionally, you can use a variability chart to organize data for detecting important patterns. Using this chart, you can plot two or three factors at a time and analyze any combination of nested or crossed factors for variance components.

References

1. College of American Pathologists http://www.cap.org/html/lip.html, September 5, 2002.
2. NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/eda/section3/eda3331.htm, September 5, 2002.

Script 1

```
Dlg = Dialog(
      "Known sample value", val = Edit Number(14),
      H List( Button("OK"), Button("Cancel"))
);
If(dlg["Button"] == -1, Throw("User cancelled")); Eval(dlg[1]);
dt = Current Data Table() << Split(
      Split Label Col(:Test),
      Split(:Y),
      Group(:Lab, :Run)
);
Column(dt,"Lab") << Label;
dt << {Select All Rows, Label, Invert Row Selection, Color by Column(:Run)};
biv = dt << Bivariate(Y(:T1), X(:T2));
Report(biv)[Frame Box(1)] << {Frame Size(450,450), Add Graphics Script(Pen
Color(1); Y Function(x,x))};
Report(biv)[Axis Box(1)] << Add Ref Line(val,1,1);
Report(biv)[Axis Box(2)] << Add Ref Line(val,1,1);
```

Figure 9: Script for Youden Plot

How Are You Using JMP?

Robert Reid, Sr. Project Engineer, CTS Microelectronics

I've been a JMP user since version 1 on the Macintosh. I've used JMP for everything from a paraxial ray trace simulation of precision fiber optic components to automated data gathering from automated electrical test machines.

Currently, our company uses JMP as a statistical client to connect to a company-wide SPC database. Data is harvested from CMMs, electrical testers, micrometers, weigh scales, viscometers, etc. by Rockwell Automation DataMyte data collectors and sent to our SQL server database. Scripts have been developed in-house to query this database to provide everything from failure analysis reports to process capability analysis reports.

JMP has become the "Swiss Army Knife" tool in our process maintenance/improvement efforts. With every day that passes, we depend on its robust feature set more and more.

Contrasts and Custom Tests

Gunter Hartel, Research and Development, CSL Limited

Lee Creighton, JMP Development

JMP implements tests on linear contrasts with its **Custom Test** command, which is found in Fit Model analysis reports. You will understand the custom tests' full power once you understand the way JMP parameterizes the linear model.

SAS users should note that the parameterization in JMP is different than the one implemented in PROC GLM.

Consider the following data set from a medical study, as shown in Figure 10. Three treatments (group), which are a vitamin supplement (Supplement), a medication (Medication), and a placebo (Placebo), were examined for their effectiveness in reducing the levels of a blood protein (change). The data include a baseline measure (initial).

When there are three groups, there are only three estimable parameters. However, you might be interested in four parameters: the overall mean and offsets from that mean to the levels of each group. Stated mathematically, you might be interested in:

- μ, the overall mean
- $\mu - d_1$, where d_1 is the offset to the first group mean
- $\mu - d_2$, where d_2 is the offset to the second group mean
- $\mu - d_3$, where d_3 is the offset to the third group mean

Because only three are estimable, one of the four must be defined in terms of the other three. In SAS, PROC GLM sets the third of the three offsets equal to zero ($d_3 = 0$). JMP, however, sets the sum of the three offsets to zero ($d_1 + d_2 + d_3 = 0$).

To see this in action, choose **Analyze > Fit Model**, and specify change as *y*, and both group and initial as effects in the model. Some of the output from the report is shown in Figure 11.

Note that in the parameter estimates, only the first two levels of group are listed. The estimate of the supplement group is the negative of the sum of the other two: –[–4.095+(–13.400)] = 17.495881. If you want to see this estimate without doing the arithmetic, click the red triangle icon on the report title bar and select **Estimates > Expanded Estimates**. This adds a report showing estimates for all levels of the nominal variable (see Figure 12).

However, although you can see estimates for all three levels, there are only two parameters being estimated, which can be verified by noting there are two degrees of freedom for the group effect in the Effects Test portion of the report.

	group	initial	change
1	Supplement	207	9
2	Supplement	196	3
3	Supplement	217	39
4	Supplement	210	24
5	Supplement	202	1
6	Supplement	201	13
7	Supplement	214	11
8	Supplement	223	32
9	Supplement	190	-8
10	Medication	220	5
11	Medication	201	-24
12	Medication	229	23
13	Medication	212	2
14	Medication	210	-8
15	Medication	217	20
16	Medication	240	38
17	Medication	232	27
18	Medication	199	-15
19	Medication	223	21
20	Medication	230	16
21	Medication	231	35
22	Placebo	239	29
23	Placebo	226	3
24	Placebo	235	24
25	Placebo	226	14
26	Placebo	208	-26
27	Placebo	236	17
28	Placebo	218	-4
29	Placebo	225	21
30	Placebo	207	-13
31	Placebo	238	34
32	Placebo	219	3

Figure 10: Drug Data Example

Response change

Summary of Fit

RSquare	0.850043
RSquare Adj	0.833977
Root Mean Square Error	7.198616
Mean of Response	11.4375
Observations (or Sum Wgts)	32

Analysis of Variance

Source	DF	Sum of Squares	Mean Square	F Ratio
Model	3	8224.9129	2741.64	52.9069
Error	28	1450.9621	51.82	Prob > F
C. Total	31	9675.8750		<.0001

Lack Of Fit

Source	DF	Sum of Squares	Mean Square	F Ratio
Lack Of Fit	27	1390.4621	51.4986	0.8512
Pure Error	1	60.5000	60.5000	Prob > F
Total Error	28	1450.9621		0.7120
				Max RSq
				0.9937

Parameter Estimates

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-298.2977	24.78222	-12.04	<.0001
group[Medication]	-4.095492	1.787297	-2.29	0.0297
group[Placebo]	-13.40039	1.999845	-6.70	<.0001
initial	1.425385	0.113844	12.52	<.0001

Effect Tests

Source	Nparm	DF	Sum of Squares	F Ratio	Prob > F
group	2	2	3377.9427	32.5930	<.0001
initial	1	1	8123.4420	156.7625	<.0001

Effect Details

Figure 11: Fit Model Analysis

Expanded Estimates

Nominal factors expanded to all levels

Term	Estimate	Std Error	t Ratio	Prob>|t|
Intercept	-298.2977	24.78222	-12.04	<.0001
group[Medication]	-4.095492	1.787297	-2.29	0.0297
group[Placebo]	-13.40039	1.999845	-6.70	<.0001
group[Supplement]	17.495881	2.248041	7.78	<.0001
initial	1.425385	0.113844	12.52	<.0001

Figure 12: Expanded Estimates Table

The parameter estimates given by Fit Model are generally useful. However, sometimes you need to test for something other than the standard comparison to zero. For example, you might need to:

- Test that an estimate is equal to 1 instead of 0.
- Test the difference between two estimates.
- Test each estimate against a control group (the placebo group, in this example).

These kinds of tests are done with *custom tests*. Custom tests produce a parameter estimate and *t*-test for a particular linear contrast, just as the ones in the pre-defined tests shown in the Parameter Estimates section of the report (Figure 11). They also provide you with an overall *F*-test for the set of contrasts you specify.

To continue with this example, suppose you want to compare group[Medication] with group [Placebo]. To do this:

1. Click the red triangle icon on the report title bar and choose **Estimates > Custom Test**. The custom test area appears in the report.
2. Enter 1 next to group[Medication] and –1 beside group[Placebo].
3. (Optional) Enter a description of the test hypothesis in the small text entry box at the top of the custom test report.
4. Click **Done** to see the results shown in Figure 13.

Response change

Custom Test

compare medication to placebo

Parameter	
Intercept	0
group[Medication]	1
group[Placebo]	-1
initial	0
=	0

Value	9.304897027
Std Error	3.0551479022
t Ratio	3.0456453583
Prob>|t|	0.0050159899
SS	480.68071212

Sum of Squares	480.68071212
Numerator DF	1
F Ratio	9.2759556482
Prob > F	0.0050159899

Figure 13: Custom Test to Compare Medication to Placebo

The custom test report shows that the difference between the two groups is significant with a Prob > *F* of 0.005.

Suppose you want to compare group[Supplement] with group[Placebo]. That is, you want to test the hypothesis group[Supplement]=group[Placebo] or, equivalently, group[Supplement]–group[Placebo] = 0.

The custom test dialog doesn't have a listing for group[Supplement] because it is a function of the other group levels, so you must consider the parameterization in constructing the test. Because JMP makes the sum of the three offsets zero, you know group[Medication] + group[Placebo] + group[Supplement] = 0 therefore group[Medication] – group[Placebo] = group[Supplement].

Into the desired test, substitute

group[Supplement] – group[Placebo] = 0

(–group[Medication] – group[Placebo]) – –group[Placebo] = 0

2*group[Placebo] + group[Medication] = 0

To do this custom test:

1. Click the red triangle icon and select **Estimates > Custom Test**.
2. Enter 2 beside group[Placebo], and enter 1 beside group[Medication].
3. Click **Done**. The custom test area appears in the report, as shown in Figure 14.

Response change

Custom Test

compare supplement to placebo

Parameter	
Intercept	0
group[Medication]	1
group[Placebo]	2
initial	0
=	0

Value	-30.89626997
Std Error	3.8615680198
t Ratio	-8.000964843
Prob>|t|	1.0311079e-8
SS	3317.2847839

Sum of Squares	3317.2847839
Numerator DF	1
F Ratio	64.015438424
Prob > F	1.0311079e-8

Figure 14: Custom Test to Compare Supplement to Placebo

On the other hand, there is a simpler way to construct the same test:

1. Open the Effect Details section of the report.
2. Click the red triangle icon on the group title bar and select **LS Means Contrast**.
3. Click the minus sign (–) beside Supplement and the plus sign (+) beside Placebo. Then click **Done**.

The Contrast Test Detail results (Figure 15) are the same as those given from the custom test in Figure 14.

However, this second method of comparing groups does not work when the necessary comparison is more complex. For example, suppose you wanted to test that the supplement

group had twice as much change as the placebo group. Mathematically, you want to test group[Supplement] = 2*group[Placebo]. There is no way to specify this in the LS Means Contrast dialog shown in Figure 15. Therefore, you must construct a custom test.

Because an equivalent test is group[Supplement] – 2 group[Placebo] = 0, you can use the same substitution as before to write (– group[Medication] – group[Placebo]) – 2 group[Placebo] = 0.

Therefore, group[Medication] + 3*group[Placebo] = 0 is the appropriate hypothesis test.

Or, suppose you wanted to test that the results from medication were half as great as those from the placebo. The test is written group[Medication] – .5*group[Placebo] = 0. The custom test dialog allows you to test that the difference between model effects is a specific amount. You could test that group[Medication] – group[Placebo] = 2.

The custom test dialog can be used to simultaneously run these and any other tests. Figure 16 shows results for tests described above.

Note that in this case, there are more tests performed than there are degrees of freedom. There are *t*-tests for each contrast, but the *F*-test is for an untestable combined set of contrasts.

Using the custom test techniques illustrated in this article, you can build complicated contrasts. The key is understanding the way JMP parameterizes the estimates. Once you grasp this, you can utilize the power of the custom test.

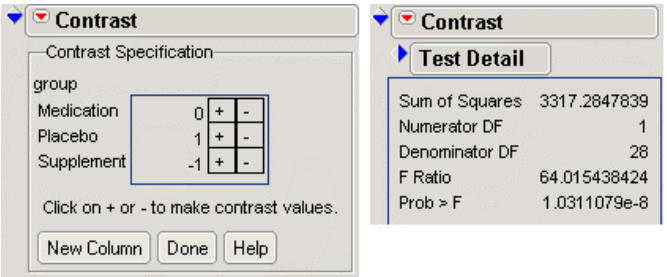


Figure 15: Contrast Dialog and Results

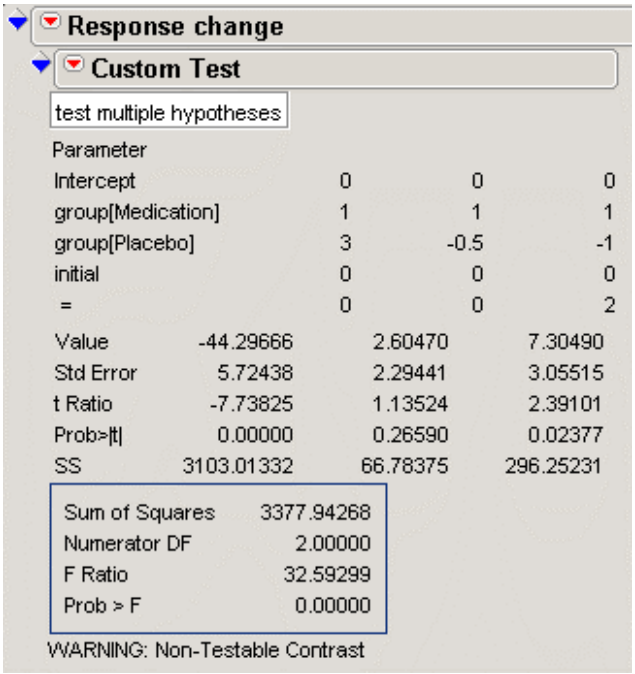


Figure 16: Custom Test Dialog for Multiple Tests

Look for JMP at these Conferences and Trade Shows

April 15-17, 2003	Quality Expo International in Rosemont, IL
May 4-7, 2003	PharmaSUG in Miami, FL
May 19-21, 2003	ASQ's AQC in Kansas City, MO
June 8-11, 2003	ACS Mid-Atlantic Meeting in Princeton, NJ
June 15-19, 2003	DIA 2003 in San Antonio, TX

Arranging Character Values

Annie Dudley Zangi, JMP Development

Have you found that when you run an analysis on character data, the data is displayed either alphabetically or in the order encountered in the data table? For those who want the data displayed in a different order, this default arrangement is a problem. It is particularly frustrating when the character data are month names because the character axes in plots and charts will start with Apr and proceed with Aug, Dec, Feb, and so on.

As an example, let's use the draft lottery data from 1971. Figure 17 shows a partial listing of the data in a JMP table.

Each day of the year is ranked by the number of draftees whose birthday fell on that date. The table has the following variables: Rank, Month (character name), Day (day of the month), and Month #. This article shows a way to obtain results where the order of months is in chronological order (without using the chronological month number).

	Rank	Month	Day
1	1	Sep	14
2	2	Apr	24
3	3	Dec	30
4	4	Feb	14
5	5	Oct	18

Figure 17: Partial Listing of Draft Lottery Data

Suppose you wanted to run a oneway ANOVA on Rank by Month. To do this:

1. Select **Analyze > Fit Y by X**.
2. Choose Rank as *y* and Month as *x*.
3. Click **OK**. Figure 18 shows the results.
4. After the scatterplot appears, click the red triangle icon on the Oneway title bar and select **Means/ANOVA/t test**.
5. Again click the red triangle icon on the Oneway title bar and select **Display Options > Connect Means**.

The ANOVA, which has a *p*-value of .0055, does not support the hypothesis that at the end of this first draft year, the draft occurred randomly across the months. Although it is easy to see that there were fewer drafts in the month of December than any other, it is not customary to see months displayed in alphabetical order. It would be better if the months were presented in chronological order.

JMP provides an easy way to fix the ordering. You do not have to add characters to month names that force chronological ordering, and you do not have to order months with a number. To order the month names chronologically:

1. Right-click the Month column name and select **Column Info** from the menu that appears. Or, highlight the Month column and select **Cols > Column Info**. The Column Info window appears.
2. Click the **New Properties** button and select **List Check**. The values in the Month column are displayed in the window in alphabetical order. You want to reorder the values in the list so analyses will recognize this new order.
3. Highlight Jan and click the **Move up** button until Jan is first in the list.
4. Highlight Feb and click the **Move Up** button until Feb is second in the list.
5. Continue arranging the list by clicking **Move Up** or **Move Down**, as shown in Figure 19.
6. Click **OK**.

In the column panel, you should now see the list check symbol (⇅) next to the column name.

Now rerun the oneway analysis by clicking the red triangle icon on the report's Oneway title bar and selecting **Script > Redo Analysis**. The results in Figure 20 show that birthdays later in the year were less likely to be selected for the draft. The months are listed in chronological order.

Although this works well with a oneway analysis, not all types of analyses recognize the list check order. For example, a Chart analysis arranges bars on the *x*-axis in the order it encounters values in the data table. If you run a Chart analysis using the Draft Lottery data table, the order of month is neither alphabetical nor chronological.

However, to remedy this problem, you can use the Sort procedure, which does recognize the list check order. So, to produce the bar chart that corresponds to the bivariate analysis in Figure 20, follow these steps:

1. Select **Tables > Sort**.
2. In the window that appears, highlight Month and click the By button to sort the table by the Month variable, which has the ordered list check property in effect.
3. Click **Sort**.
4. Select **Graph > Chart**.
5. In the window that appears, click Month, then click **X, Level**.
6. Highlight Rank, then click the Statistics button. Select **Mean** from the menu for the *y*-axis.
7. Click **OK**, and you will see the results in Figure 21.

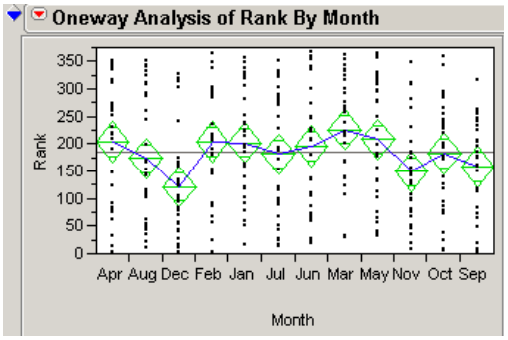


Figure 18: Oneway Anova of Rank by Month

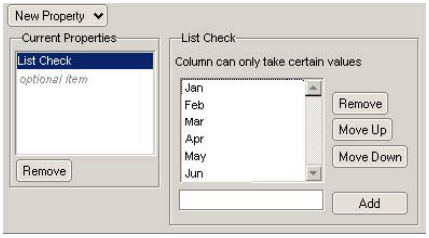


Figure 19: Column Info Window with List Check

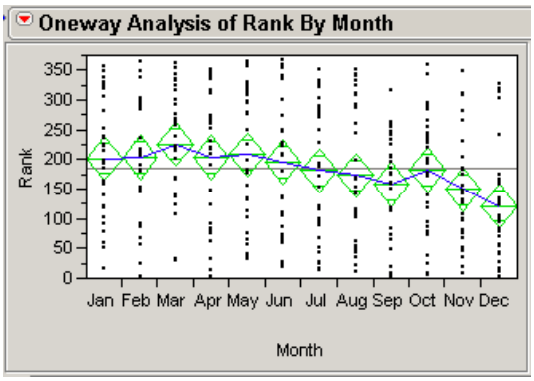


Figure 20: ANOVA with Chronological Month Names

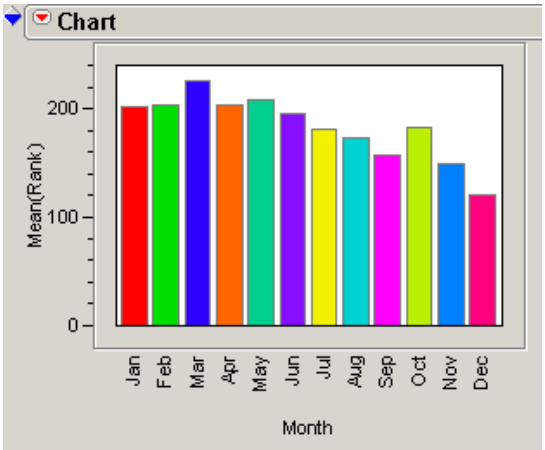


Figure 21: Chart of Mean(Rank) by Month Sorted by List Check Order Names

Case Study

IT Provides A Little Animal Magic

Bryan Glick, Computing Magazine

Conservationists are using specialist statistical analysis software to track rhinos in Africa. Now the technology could help save the tiger too.

There can't be many IT initiatives that have been inspired by Kalahari bushmen. Nor would business users of statistical analysis software immediately think that the northern hairy-nosed wombat might benefit from such complex technology.

But a pioneering project to help the survival of endangered rhinos in southern Africa demonstrates that the influence and benefits of IT are spreading to the most unlikely locations.

Wildlife biologists Zoë Jewell and her husband Sky Alibhai are behind the idea. Next month they are off to India to test out technology originally developed for their Rhinowatch project on a new initiative to help save wild tigers (http://www.assisi.demon.co.uk/rhinowatch).

Rhinowatch was first established in 1991 to help monitor black rhino populations in Zimbabwe at a time when poaching was seriously threatening the species.

Over the next five years, Jewell and Alibhai began to realise that some of the existing methods of monitoring had a worrying downside.

Rhinos had to be immobilised to fit radio collars for tracking, and the couple discovered that this process

was inhibiting the fertility of female rhinos. The attempts to protect the species were actually slowing reproduction rates.

A new method was needed that was sustainable but that did not invade the rhinos' lives. The biologists had watched local tribesmen identify individual rhinos by their unique footprints.

Jewell explained that it was a skill based on generations of experience that she could never hope to learn.

"But if it was possible to translate that technique using modern technology, we would have a cost-effective and non-invasive way to find out how many rhinos there are and their location. That's where we began," she said.

In 1996, they began work with supplier SAS, using its specialised statistical software to analyse data they collected by measuring rhino footprints. This information was then used to follow the movement of each animal.

[Rhinowatch used software that was a combination of SAS and JMP. They used a SAS application written by Nigel Law to enter and process the landmark points. They used JMP for statistical analysis.]

"It sounds easy but it took a very long time," said Jewell. "We started by tracing footprints on to paper and measuring them with a ruler.

"But as technology such as digital cameras and scanners evolved, so did our methods. Now the whole thing is running smoothly with all the technology, and we're comparable with the bushmen."

The SAS software has proved to be a vital conservation tool: the method delivers accuracy rates of 90 per cent in identifying individual animals.

And Jewell explained that the package has had further benefits. "Many common monitoring techniques are invasive to the animals and quite often not reliable," she said.

"They can also be extremely expensive and therefore not cost-effective, especially for developing countries."

Projects no longer need teams of scientific researchers to bring in radio receivers and helicopters to follow the herds.

Now local patrols can take digital photos, which Rhinowatch uses to build a picture of the animals' movements, meaning that anti-poaching teams can be deployed more accurately.

"We can now use the indigenous skills of people who were brought up around these animals, which is key to sustainability in the long term," said Jewell.

After the technique was published for the first time in 2001, Jewell and Alibhai were contacted by conservationists working to protect tigers in India.

Together with SAS, which by now was also providing extra support and funding for the project, they adapted the software for a trial at an animal sanctuary in North Carolina. US conservationists took photos of tiger paw prints and emailed them to Jewell.

The pictures are optimised using Adobe Photoshop software, then processed through SAS software to obtain the raw data and compare the characteristics of every print to identify each animal. Graphical geographic information systems help to map their movements.

(continued on page 12)

JMP Named to 'Top 5 Products of 2002' List

Amy Smith, JMP Sales and Marketing

JMP™, a business unit of SAS, announced its election to the "Top 5 Products of 2002" list of *Scientific Computing & Instrumentation* (February, 2003) magazine in two categories. The publication, with circulation exceeding 45,000, has created a list of the five best products for each of their wide-ranging solutions categories, based on the number of inquiries from readers during the past year. JMP excelled among readers in the categories of "data analysis" and "statistical software." JMP was also touted as a technology leader—developing new products and technologies applicable to the needs of tomorrow.

(continued from page 11)

The test with captive tigers was a success, and the next step, in March, is to find out whether the terrain in India will be suitable for photographing the detail in paw prints so that the software could be considered for use with wild tigers. Field trips are also planned to Bangladesh and Siberia.

If sufficient funding can be found, it will be good news for tigers, and potentially for other animals too.

"Conservationists are very interested in whether the software could be used for other species," said Jewell. "There is a team in Australia working on the northern hairy-nosed wombat, which is extremely endangered. But it is mainly for large animals that are difficult or dangerous to track."

Just as in business, the increasing power and capability of IT has meant huge boosts in productivity. IT has revolutionised the conservation process for wildlife biologists.

"Using digital cameras has enormously increased the number of footprint photographs we can take and process, enabling us to deliver rapid feedback on rhino movements to those field managers responsible for their protection," said Jewell. "What we do would be impossible without technology."

This article was originally published in the February 19, 2003, issue of Computing Magazine.

About JMPer Cable
Issue 11 Winter 2003

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

Contributors
John Sall
Mark Bailey
Gunter Hartel
Lee Creighton
Robert Reid
Annie Dudley Zangi
Brian Glick
Amy Smith

Editors
Ann Lehman
Meredith Blackwelder

Designer
Meredith Blackwelder

Printing
SAS Institute Print Center

Questions or comments
jmp@sas.com

To order JMP software
1-877-594-6567

For more information on JMP
1-877-594-6567
www.jmp.com

Copyright 2003, SAS Institute. All rights reserved. JMPer Cable SAS, JMP, JMPer Cable, 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 registration. Other brand and product names are trademarks of their respective companies.