"Programming" and Other Features of the JMP® Calculator

Daniel J. Obermiller, The Dow Chemical Company, Midland, MI

Abstract

Many new users to JMP are sometimes confused by the JMP calculator. With all of the different functions available and various ways of building a formula, it is easy to be overwhelmed. However, the confusing aspects of the calculator also give JMP extra flexibility that one may not see at first glance. One example of this is that you can often build statistical tests that are not normally available within JMP. This paper briefly reviews some of the more esoteric JMP functions and how they may be used as well as ways of doing some simple "programming" with the JMP calculator.

Introduction

Many new users to JMP are sometimes confused by the JMP calculator. With all of the different functions available and various ways of building a formula, it is easy to be overwhelmed. However, the confusing aspects of the calculator also give JMP extra flexibility that one may not see at first glance. One example of this is that you can often build statistical tests that are not normally available within JMP. Although this feature will soon be available in version 4 of JMP, the techniques presented in this paper will work with the current and earlier versions of JMP as well as version 4.

Before utilizing the calculator to build "new" statistical tests, a few of the less commonly used features of the calculator will be introduced.

Calculator Features

Nested Do-Loops

As with most things in JMP, there is more than one way to create nested do-loops within JMP. For this example, one loop (called Inside Loop) will be created going from 1 to 4, within another loop (the outer loop) going from 1 to 4. The result will be 16 observations and two variables (Outer Loop and Inner Loop). To start, a JMP data table with 16 observations and two variables will be created. Enter a formula for the outer loop that looks like this:

$$\begin{cases} 1, & \text{if } i = 1 \\ \textit{Outer Loop}_{i-1} + 1, & \text{if } \textit{Inner Loop} = 1 \\ \textit{Outer Loop}_{i-1}, & \text{otherwise} \end{cases}$$

The above formula is created using the Conditional→If clause and inserting an additional if statement. A subscript of i-1 is used that will assign values using the previous row of Outer Loop. Notice that the values for Outer Loop will depend on the values for Inner Loop.

The Inner Loop formula utilizes the Numeric function Count. The function instructs JMP to count from 1 to 4 in 4 steps, 1 time.

count (from 1, to 4, in 4 steps , 1 time)

The final results are below.

2 Cols / 16 Rows	Outer Loop	Inner Loop
1	1	1
2	1	2
3	1	3
4	1	4
5	2	1
6	2	2
7	2	3
8	2	4
9	3	1
10	3	2
11	3	3
12	3	4
13	4	1
14	4	2
15	4	3
16	4	4

Statistics of Rows or Columns?

Many of the statistical functions that are provided in JMP are designed to work on a column of values. A small dataset of two different height values will be used to illustrate this (see Figure 1). The statistical function for the Mean will calculate the mean of a column of values as shown in the Mean Ht 1 column or the Mean Ht 2 column in Figure 1.

$$\overline{Ht\ 1}$$

But what if we wanted to compute a row mean? You could explicitly build the formula as below.

$$\frac{Ht\ 1\ + Ht\ 2}{2}$$

The results of this formula can be seen in the (Ht1 + Ht2)/2 column in Figure 1. Notice that if Ht 1 or Ht 2 have a missing value, this formula evaluates as missing. Using the Of statistical function will eliminate this potential problem. Choose the Mean function and then choose the Of function and put in Ht 1 and Ht 2 as the arguments.

$$\overline{\text{of}(Ht\ 1\ \ ,\ Ht\ 2\ \ )}$$

The results of this can be seen in the Mean of Ht1, Ht2 column of Figure 1.

6 Cols / 10 Rows	Ht 1	Ht 2	Mean Ht 1	Mean Ht 2	(Ht 1+Ht 2)/2	Mean of Ht1,Ht2
1	47	50	49.5	60	48.5	48.5
2	48	60	49.5	60	54	54
3	50	65	49.5	60	57.5	57.5
4	55	64	49.5	60	59.5	59.5
5	44	60	49.5	60	52	52
6	56	•	49.5	60	•	56
7	48	57	49.5	60	52.5	52.5
8	55	65	49.5	60	60	60
9	49	55	49.5	60	52	52
10	43	64	49.5	60	53.5	53.5

Figure 1: Example of Of Function

Temporary Variables

Utilizing a temporary variable can be very useful. A temporary variable is similar to other JMP variables, but it does not take up a column on the data spreadsheet.

To create a temporary variable, from the JMP calculator choose the New Variable choice from the Variables function section. You can set the variable name and type when you create the variable. Once created a temporary variable can be used in any JMP calculation by choosing it from the Variables function section. These variables are often used with a Conditions→Assignments function.

One example of using a temporary variable can be to generate numbers from a random normal distribution with a specific mean and standard deviation. Many users may not remember which statistic is multiplied by a random normal number and which one is added. To solve this, create two numeric temporary variables, one called Mean, the other called Std Dev. Now using the Assignments function, enter two assignments. In the results area, one can put in the desired function that does the correct algebra. The final formula will look as follows.

$$\begin{array}{|l} \boldsymbol{Mean} \ \ll= \ 5 \\ \boldsymbol{Std\ Dev} \ \ll= \ 2 \\ \text{results} \quad \text{?normal} \ \bullet \boldsymbol{Std\ Dev} \ + \boldsymbol{Mean} \end{array}$$

By utilizing the temporary variables, an infrequent user should be able to easily fill in the values to get the desired results.

Grubb's and Dixon's Outlier Tests

Implementing many of the above calculator techniques, one can "program" many statistical tests that are not currently available in JMP. These programs are not typical programs like a computer programmer would create. Instead these programs are typically called templates. Templates are standard formulas that are combined in a spreadsheet-like fashion to get the desired results. The example that will be gone through here will be creating the template to perform Grubb's outlier test and the Dixon outlier test.

Although JMP provides various means of detecting outliers, some current standard operating procedures may rely on the above-mentioned tests. An outlier is a value in a dataset that is unusually large or small compared to the other values. Because they differ significantly from the rest of the data, outliers can alter the results of an analysis. For this reason, the identification of outliers can be very important.

Grubb's Test

Grubb's test will take on one of two forms depending on whether the largest value in the dataset is suspected or if the smallest value is suspected.

If the largest value is suspected:

$$T_n = (x_n - \bar{x})/s$$

If the smallest value is suspected:

$$T_1 = (\bar{x} - x_1)/s$$

where x_n is the largest value
x_1 is the smallest value
$\bar{x}$ is the mean of all n values
s is the sample standard deviation of all n values

Compare T_i to a tabled critical value (Table 1) for sample size n and selected significance level.

Dixon's Test

The Dixon outlier test depends on how many data values are present and which observation is suspected.

n	Largest value suspected	Smallest value suspected
3 -7	$r_{10} = \frac{(x_n - x_{n-1})}{(x_n - x_1)}$	$r_{10} = \frac{(x_2 - x_1)}{(x_n - x_1)}$
8 - 10	$r_{11} = \frac{(x_n - x_{n-1})}{(x_n - x_2)}$	$r_{11} = \frac{(x_2 - x_1)}{(x_{n-1} - x_1)}$
11 - 13	$r_{21} = \frac{(x_n - x_{n-2})}{(x_n - x_2)}$	$r_{21} = \frac{(x_3 - x_1)}{(x_{n-1} - x_1)}$
14 - 25	$r_{22} = \frac{(x_n - x_{n-2})}{(x_n - x_3)}$	$r_{22} = \frac{(x_3 - x_1)}{(x_{n-2} - x_1)}$

where x_n is the largest value
x_1 is the smallest value
x_{n-1} is the second largest value, etc.

Compare r_{ij} to a tabled critical value (Table 2) for sample size n and selected significance level.

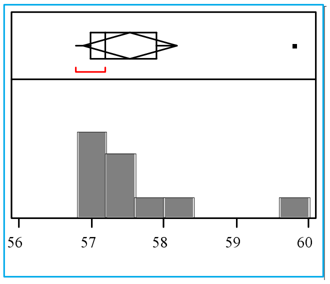
Example

Consider the following ten observations:

56.8, 57.0, 57.0, 57.0, 57.2,
57.2, 57.5, 57.8 58.2, 59.8

The observation in question is the high value, x_{10} = 59.8. At a significance level of 1%, is it possible that 59.8 is an outlier?

Looking at a histogram and outlier boxplot of the data, 59.8 does look suspicious.



Performing both outlier tests by hand yields the following results.

Grubbs:

$$T_{10} = \frac{x_{10} - \bar{x}}{s} = \frac{59.8 - 57.55}{0.898} = 2.504$$

The tabled value of T for n=10, significance level = 1% is 2.41. Since 2.504 > 2.41, 59.8 is suspect and should be investigated.

Dixon:

$$r_{11} = \frac{x_{10} - x_9}{x_{10} - x_2} = \frac{59.8 - 58.2}{59.8 - 57} = 0.571$$

The tabled value of r_{11} for n=10, significance level = 1% is 0.597. Because 0.571 < 0.597, there is insufficient evidence to suggest that 59.8 is an outlier at the 1% level. There is evidence that this point is an outlier at the 5% significance level.

This example shows how a borderline case may be considered as an outlier with one test but not with another. Now attention can be focused on creating the JMP template to perform both of these tests.

	Significance Level		
n	**5%**	**2.5%**	**1%**
3	1.15	1.15	1.15
4	1.46	1.48	1.49
5	1.67	1.71	1.75
6	1.82	1.89	1.94
7	1.94	2.02	2.10
8	2.03	2.13	2.22
9	2.11	2.21	2.32
10	2.18	2.29	2.41
11	2.23	2.36	2.48
12	2.29	2.41	2.55
13	2.33	2.46	2.61
14	2.37	2.51	2.66
15	2.41	2.55	2.71
16	2.44	2.59	2.75
17	2.47	2.62	2.79
18	2.50	2.65	2.82
19	2.53	2.68	2.85
20	2.56	2.71	2.88
21	2.58	2.73	2.91
22	2.60	2.76	2.94
23	2.62	2.78	2.96
24	2.64	2.80	2.99
25	2.66	2.82	3.01

Table 1: Grubbs Test Table of Critical Values of T

	Significance Level		
n	**10%**	**5%**	**1%**
3	0.886	0.941	0.988
4	0.679	0.765	0.889
5	0.557	0.642	0.780
6	0.482	0.560	0.698
7	0.434	0.507	0.736
8	0.479	0.554	0.683
9	0.441	0.512	0.635
10	0.409	0.477	0.597
11	0.517	0.576	0.679
12	0.490	0.546	0.642
13	0.467	0.521	0.615
14	0.492	0.546	0.641
15	0.472	0.525	0.616
16	0.454	0.507	0.595
17	0.438	0.490	0.577
18	0.424	0.475	0.561
19	0.412	0.462	0.547
20	0.401	0.450	0.535
21	0.391	0.440	0.524
22	0.382	0.430	0.514
23	0.374	0.421	0.505
24	0.367	0.413	0.497
25	0.360	0.406	0.489

Table 2: Dixon Criteria for Testing of Extreme Observation

“Programming” in JMP

The outlier tests will be limited to sample sizes of 25 or less (as indicated in Tables 1 and 2). The first step in creating the template will be to add 25 rows for the maximum dataset size. The first column will be called n and indicate the sample sizes. The values can be typed in or put in using the count function. The second column will be called Data and will hold the dataset to be examined. The Data column must be in increasing order.

To start performing the Grubb’s test, two columns will be added. The first column will be the Grubb’s T statistic for when the smallest value is suspected (Grubbs Min). The second column for when the largest value is suspected (Grubbs Max). The formulas for these columns are relatively straight forward, utilizing the statistical

Mean function, statistical # of Nonmissing function, and subscripts to specify specific observations such as the first (smallest) observation.

The formula for Grubbs Min is:

$$\frac{(\overline{Data} - Data_{1})}{\text{std}\ Data}$$

For Grubbs Max, the formula is:

$$\frac{(Data_{\text{N}\ Data} - \overline{Data})}{\text{std}\ Data}$$

Similar to Grubb's test two columns will be created for Dixon's test, one for the smallest value suspected, another for the largest value suspected. However, the formulas for Dixon's test change depending on the sample size. Thus, a Conditional→If function is required. To make the formula easier to follow, two temporary variables are created. The first, called Dixon Sample Size, will take on values 1, 2, 3, or 4 which will identify which Dixon formula to use. The other temporary variable is Sample Size which holds the number of data points in the dataset. A Choose function is then used to distinguish which Dixon formula to use, based on the value of Dixon Sample Size. The complete formula for the Dixon Min is below.

$$\textbf{\textit{Dixon Sample Size}} \;<\!<=\; \begin{cases} 1, & \text{if } 3 \le \text{N}\ Data \le 7 \\ 2, & \text{if } 8 \le \text{N}\ Data \le 10 \\ 3, & \text{if } 11 \le \text{N}\ Data \le 13 \\ 4, & \text{if } 14 \le \text{N}\ Data \le 25 \\ \square, & \text{otherwise} \end{cases}$$

$$\textbf{\textit{Sample Size}} \;<\!<=\; \text{N}\ Data$$

$$\text{results} \begin{cases} \text{choose} \quad \textbf{\textit{Dixon Sample Size}}: & \\ \dfrac{(Data_{2} - Data_{1})}{(Data_{\textbf{\textit{Sample Size}}} - Data_{1})}, & \text{choice 1} \\ \dfrac{(Data_{2} - Data_{1})}{(Data_{\textbf{\textit{Sample Size}} - 1} - Data_{1})}, & \text{choice 2} \\ \dfrac{(Data_{3} - Data_{1})}{(Data_{\textbf{\textit{Sample Size}} - 1} - Data_{1})}, & \text{choice 3} \\ \dfrac{(Data_{3} - Data_{1})}{(Data_{\textbf{\textit{Sample Size}} - 2} - Data_{1})}, & \text{choice 4} \\ \square, & \text{otherwise} \end{cases}$$

A similar approach is used for Dixon Max. An easier way to create the Dixon Max formula is to copy the Dixon Min formula. Paste it into the Dixon Max column and make the few changes as needed. The formula is:

$$\textbf{\textit{Dixon Sample Size}} \;<\!<=\; \begin{cases} 1, & \text{if } 3 \le \text{N}\ Data \le 7 \\ 2, & \text{if } 8 \le \text{N}\ Data \le 10 \\ 3, & \text{if } 11 \le \text{N}\ Data \le 13 \\ 4, & \text{if } 14 \le \text{N}\ Data \le 25 \\ \square, & \text{otherwise} \end{cases}$$

$$\textbf{\textit{Sample Size}} \;<\!<=\; \text{N}\ Data$$

$$\text{results} \begin{cases} \text{choose} \quad \textbf{\textit{Dixon Sample Size}}: & \\ \dfrac{(Data_{\textbf{\textit{Sample Size}}} - Data_{\textbf{\textit{Sample Size}} - 1})}{(Data_{\textbf{\textit{Sample Size}}} - Data_{1})}, & \text{choice 1} \\ \dfrac{(Data_{\textbf{\textit{Sample Size}}} - Data_{\textbf{\textit{Sample Size}} - 1})}{(Data_{\textbf{\textit{Sample Size}}} - Data_{2})}, & \text{choice 2} \\ \dfrac{(Data_{\textbf{\textit{Sample Size}}} - Data_{\textbf{\textit{Sample Size}} - 2})}{(Data_{\textbf{\textit{Sample Size}}} - Data_{2})}, & \text{choice 3} \\ \dfrac{(Data_{\textbf{\textit{Sample Size}}} - Data_{\textbf{\textit{Sample Size}} - 2})}{(Data_{\textbf{\textit{Sample Size}}} - Data_{3})}, & \text{choice 4} \\ \square, & \text{otherwise} \end{cases}$$

With the above formulas entered, the JMP spreadsheet will have the calculated statistics for both Grubbs and the Dixon test. As can be seen in the spreadsheet below, the Grubbs Max and Dixon Max columns match the by-hand calculations presented earlier.

16 Cols / 25 Rows	n	Data	Grubbs Min	Grubbs Max	Dixon Min	Dixon Max
1	?	56.8	0.834766	2.504298	0.142857	0.571429
2	?	57	0.834766	2.504298	0.142857	0.571429
3	3	57	0.834766	2.504298	0.142857	0.571429
4	4	57	0.834766	2.504298	0.142857	0.571429
5	5	57.2	0.834766	2.504298	0.142857	0.571429
6	6	57.2	0.834766	2.504298	0.142857	0.571429
7	7	57.5	0.834766	2.504298	0.142857	0.571429
8	8	57.8	0.834766	2.504298	0.142857	0.571429
9	9	58.2	0.834766	2.504298	0.142857	0.571429
10	10	59.8	0.834766	2.504298	0.142857	0.571429

Finally, we need to compare the calculated statistic to the tabled value to determine significance. The tabled values are entered in the JMP table according to the sample size indicated in the n variable. The tabled values are named as G5%, G2.5%, and G1% for the Grubbs significance values and D10%, D5%, and D1% for the Dixon values.

The comparison to the table values are accomplished using four temporary variables and an assignment function. Three temporary variables are needed to store the tabled value for each significance level. The fourth temporary variable holds the lowest significance level that the statistic is greater than. This fourth temporary variable is then combined with some text in the final result. The final result will state "Significant at" the lowest possible significance level of the test if the test is significant. If the test is not significant, it will return the phrase "Not Significant". This approach is used for both the Grubbs Min and Max tests as well as the Dixon Min and Max tests. The full formula for the Grubbs Max test is in Figure 2.

5% Sig <<= { *G 5%* , if *n* = N *Data* ; ☐ , otherwise }

2.5% Sig <<= { *G 2.5%* , if *n* = N *Data* ; ☐ , otherwise }

1% Sig <<= { *G 1%* , if *n* = N *Data* ; ☐ , otherwise }

Sig Level <<= { { "5%" , if *Grubbs Max* < ***2.5% Sig*** ; { "1%" , if *Grubbs Max* >= ***1% Sig*** ; "2.5%" , otherwise } , otherwise } , if *Grubbs Max* >= ***5% Sig*** and *n* = N *Data* ; "0" , otherwise }

results { { "Not Significant" , if *n* = N *Data* ; ☐ , otherwise } , if ***Sig Level*** = "0" ; "Significant at " ‖ ***Sig Level*** , otherwise }

Figure 2. Grubbs Test Table Look-up

16 Cols / 25 Rows	n	Data	Grubbs Min	Grubbs Max	Dixon Min	Dixon Max	Grubbs Min Sig	Grubbs Max Sig	Dixon Min Sig	Dixon Max Sig
1	?	56.8	0.834766	2.504298	0.142857	0.571429				
2	?	57	0.834766	2.504298	0.142857	0.571429				
3	3	57	0.834766	2.504298	0.142857	0.571429				
4	4	57	0.834766	2.504298	0.142857	0.571429				
5	5	57.2	0.834766	2.504298	0.142857	0.571429				
6	6	57.2	0.834766	2.504298	0.142857	0.571429				
7	7	57.5	0.834766	2.504298	0.142857	0.571429				
8	8	57.8	0.834766	2.504298	0.142857	0.571429				
9	9	58.2	0.834766	2.504298	0.142857	0.571429				
10	10	59.8	0.834766	2.504298	0.142857	0.571429	Not Significant	Significant at 1%	Not Significant	Significant at 5%

Figure 3: Final Result of Outlier Template

The final complete spreadsheet with the results is seen in Figure 3.

The templates can be very helpful, but often some extra guidance or instructions may be needed to help a new user. Fortunately, JMP does offer you some capability of putting in some instructions. First, information about the entire template can be stored in the table information notes (Tables→Table Info). By using the column notes (from the Column Information dialog), one can put in column specific instructions that are saved with the template. The following table shows the comments that are part of the outlier detection template.

Column Name	Column Notes
n	Various sample sizes for which a critical value is available.
Data	Paste your sorted data into this column. Be sure to have more than 2 samples, but less than 26.
Grubbs Min	This column computes the Grubbs test if the smallest data value is suspect.
Grubbs Max	This column computes the Grubbs test if the largest data value is suspect.
Dixon Min	This column computes the Dixon test if the smallest data value is suspect.
Dixon Max	This column computes the Dixon test if the largest data value is suspect
Grubbs Min Sig	Significance of Grubbs test when smallest value is suspect. Note that the result will only appear in the row corresponding to the sample dataset size.
Grubbs Max Sig	Significance of Grubbs test when largest value is suspect. Note that the result will only appear in the row corresponding to the sample dataset size.
Dixon Min Sig	Significance of Dixon's test when smallest value is suspect. Note that the result will only appear in the row corresponding to the sample dataset size.
Dixon Max Sig	Significance of Dixon's test when largest value is suspect. Note that the result will only appear in the row corresponding to the sample dataset size.

Another tip that can make using a template easier is to use the Hide Columns command. For the outlier detection template, notice in Figure 3 that none of the critical table values from Tables 1 and 2 can be seen. Those values are part of the spreadsheet, but are in a hidden column since the user has no need to see those values. Of course, one can easily unhide the columns as needed.

Conclusion

The JMP calculator, although it may be confusing at first, is a very powerful and flexible tool. Utilizing the calculator, many different analyses not currently available within JMP can still be performed. The approach of creating JMP templates as a means of "programming" can still be useful even when version 4 of JMP provides a true programming feature as these techniques will work with all previous versions of JMP as well as version 4. This paper presented several calculator techniques that may be useful in creating JMP templates and demonstrated a template that performed Grubbs outlier test as well as the Dixon outlier test. Some helpful hints in creating JMP templates was also provided.

References

Barnett, V. And Lewis, T., *Outliers in Statistical Data*, John Wiley & Sons, 1978.

Best, Al M., "Writing Programs for JMP Software", *Proceedings of the Twentieth Annual SAS Users Group International Conference*, 1995, pp. 1188-1193.

Dean, R.B. and Dixon, W.J., "Simplified Statistics for Small Numbers of Observations", *Analytical Chemistry*, April 1951, Vol. 23, No. 4, pp. 636-638.

Duncan, A.J., *Quality Control and Industrial Statistics, Fifth Edition*, Homewood, IL, Richard D. Irwin, Inc., 1986, pp. 760-761.

Grubbs, F.E., "Procedures for Detecting Outlying Observations in Samples", *Technometrics*, February 1969, Vol. 11, No. 1, pp. 1-21.

SAS Institute, Inc. (1995). *JMP User's Guide*, Cary, NC.

SAS Institute, Inc. (1995). *JMP Statistics and Graphics Guide*, Cary, NC.

Acknowledgments

I would like to thank Jeff Sweeney of Dow Chemical for his comments and help in preparing this paper.

Dan Obermiller may be reached through the internet at djobermiller@dow.com or by mail at The Dow Chemical Company, 1702 Building, Midland, MI 48674. Phone 517-636-1000. Fax 517-638-9623.