



User Experience, Not Metrics

by:

R. Scott Barber

Part 6: What is an Outlier and how do I account for one?

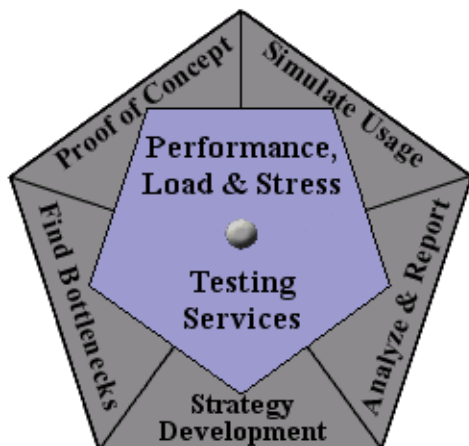
Everyone who uses the Internet on a regular basis knows that every once in a while, a web page that normally loads quickly takes longer than expected. Experienced web users will simply hit the refresh button and the page will normally load as expected. When this happens, long time web users think, "hmmm, that was odd," if they even consciously notice at all, and immediately forget the incident. So, what do you do when this happens during your performance test? That's exactly what we are going to discuss this month.

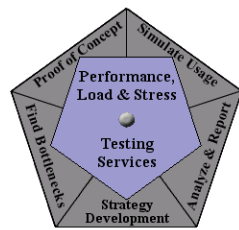
Last month we discussed placing timers in scripts and what those timers measure. This month we begin to discuss what to do with the resulting times received after executing a test. For the next few articles, I will assume that we know how to create scripts with the delays and timers to represent your user community model and focus on what to do with the results of the tests. This month's edition of the "User Experience, Not Metrics" article series discusses how to use Rational TestStudio and Microsoft Excel to determine whether or not a collected measurement is an outlier, and how to account for identified outliers.

If you have not yet read the other articles in this series, the topics presented in this article do not directly depend on topics presented previously. The intent of this article is to discuss how to handle what can be a complex mathematical and statistical topic in a way that non-math majors will understand. We will do this by combining statistics with practical experience based on years of conducting web-based Performance Testing. The article is intended for all levels of TestStudio users, and may also prove useful to managers of projects where performance testing will occur.

What is an Outlier?

If we were to ask statisticians what an outlier is they would tell you that any measurement that falls outside of three standard deviations, or 99% of all collected measurements, is an outlier. By definition one standard deviation is the variance in measurements that encompasses 68% of all measurements. That means if your sample of 100 measurements has an average of 10 seconds and a standard deviation of 3 seconds, that 68% of all measurements are between 7 seconds and 13 seconds. Following this example, three standard deviations, would encompass a range of 1 second to 19 seconds. Any measurement of less than 1 second or greater than 19 seconds is an outlier. The problem with this definition is that it assumes that your collected measurements are distributed normally, which is rarely





the case for web-based page response times. If you are interested in the mathematical equations for this type of analysis you can visit <http://www.health.uottawa.ca/biomech/courses/hss2381/means.pdf> published by the faculty of health sciences from the University of Ottawa, Canada for a biomedical data analysis and measurements course.

A more applicable definition of an outlier for our purposes can be found in the glossary of terms at the following URL <http://www.statsoftinc.com/textbook/glosfra.html> published by StatSoft Inc.

“Outliers are atypical (by definition), infrequent observations; data points which do not appear to follow the characteristic distribution of the rest of the data. These may reflect genuine properties of the underlying phenomenon (variable), or be due to measurement errors or other anomalies which should not be modeled.”

While that definition sounds rather technical, it can be depicted easily by the following graphs.

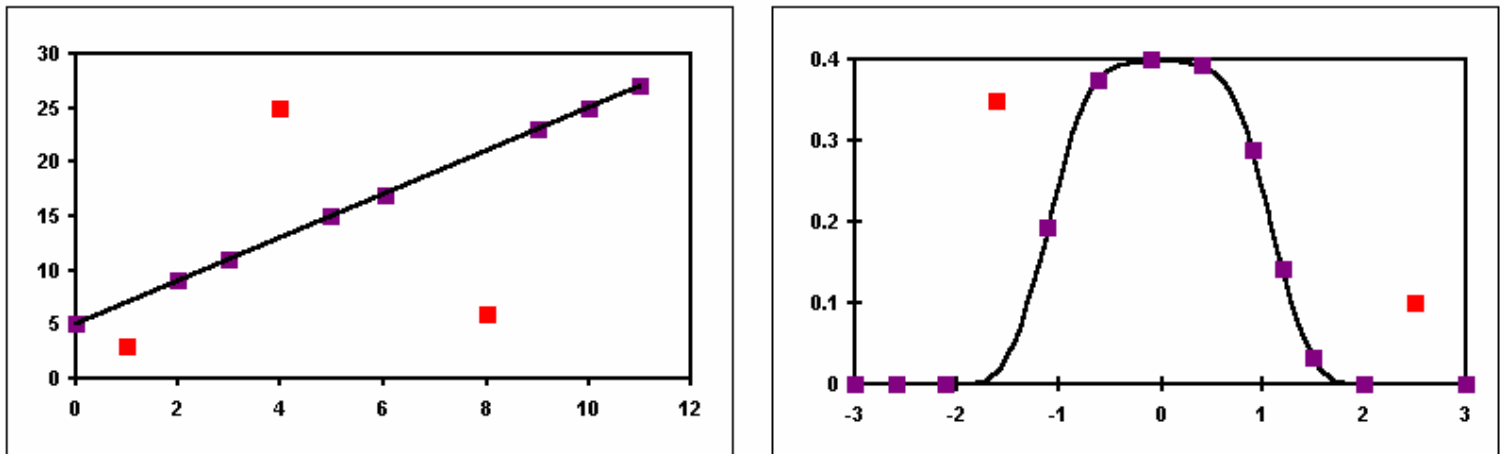


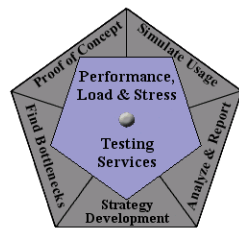
Figure 1 – Visual examples of outliers

The datapoints that do not fall along the trend lines, depicted in red, are outliers by this definition.

There is still a problem with this definition for our purposes, and that is statistical significance. In an effort to keep this discussion from getting overly mathematical, let's take a common sense approach. In the first graph of figure 1, we see that three of eleven datapoints qualify as outliers. Any of you that have ever taken a statistics course, will immediately agree that nearly 30% of all collected datapoints cannot be dismissed as outliers. So, how do we determine how many datapoints we can reasonable dismiss as outliers? That is where common sense comes in.

How many Outliers can I have?

So how many outliers can we dismiss as “atypical infrequent observations”? To answer this question, we must first determine if the sample of measurements are statistically significant. Mathematically calculating statistical significance or reliability, based on sample size is well beyond the scope of this article. Luckily, we have a way around this problem. First, it is easy for us to add iterations to our tests to increase the total number of measurements collected. Whenever possible, ensure that before attempting to determine if there are outliers, that you have at least 100 measurements of each timer.



Then, if you see evidence of outliers, re-execute the tests and compare them. If the majority of the measurements are the same, plus or minus the outliers, then the results are likely to be statistically significant. In article 8 of this series we will discuss more about what types of tests are most useful throughout the testing cycle and how to ensure statistical significance.

Now, assuming we have a statistically significant sample of measurements, we can continue our discussion on outliers. I submit that there is no set number of outliers that can be dismissed, but a maximum percentage of the total observations. If we apply the spirit of the two definitions that we have discussed, we come to the conclusion that up to 1% of the total measurements beyond the third standard deviation are significantly outside the rest of the measurements and can be considered outliers.

Identifying Outliers

Now that we agree on a working definition of an outlier, let's discuss how to identify them. To identify outliers, we are going to use a version of the Response vs. Time charts available in TestManger. To display the chart we will be using, execute a test, then upon completion, press the Resp vs. Time button. Once the chart is displayed, on the menu bar, select View and then Settings. Doing that will display the window shown in figure 2.

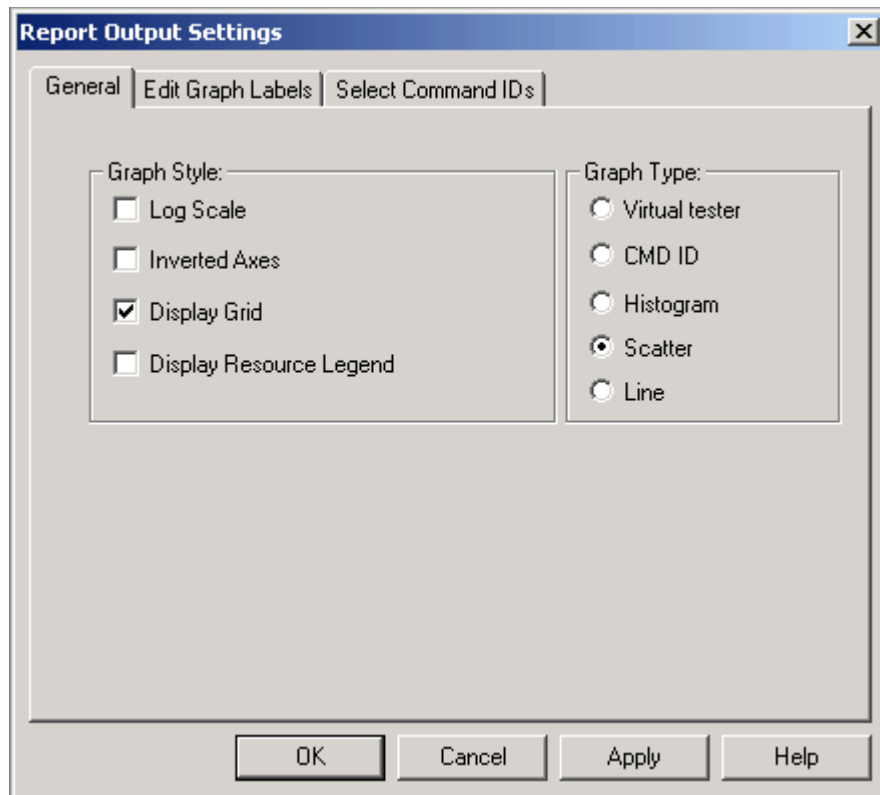


Figure 2 – Response vs. Time Options window

The original settings on this window is for the Virtual tester radio button on the graph type section to be selected. You can see in figure 2 that the Scatter option is selected. Scatter charts make it easy to identify outliers.

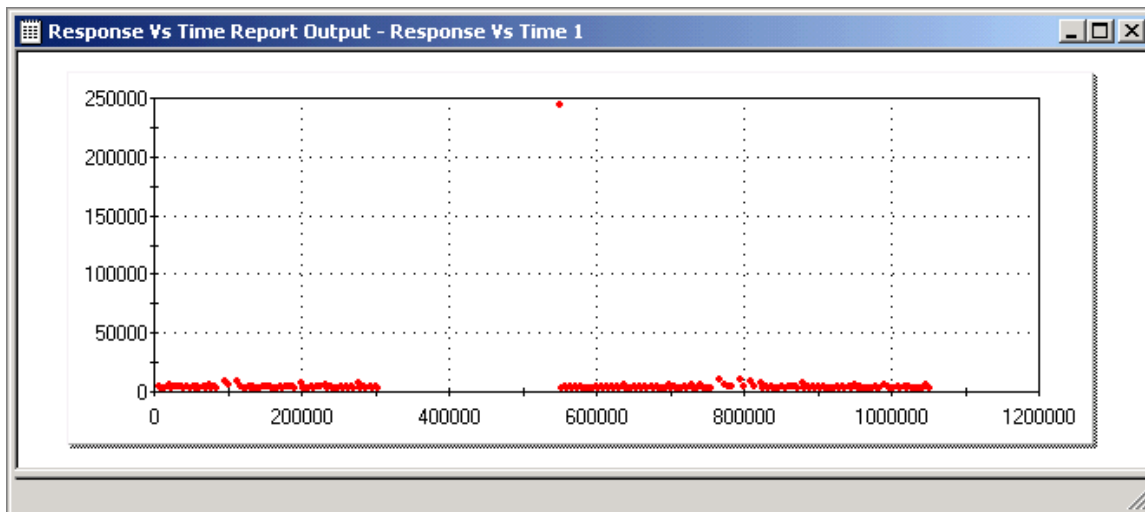
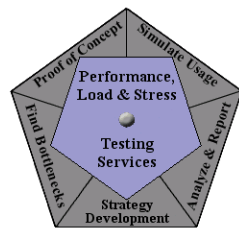


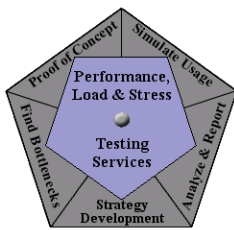
Figure 3 – Response vs. Time Scatter Chart

For those of you who are unfamiliar with the scatter chart, allow me to describe what you see in figure 3. Vertically on the left hand side is the response time in milliseconds. Across the bottom of the graph is the time since the start of the test run, also in milliseconds. Each red dot represents a command ID or timer measurement. Each measurement is placed on the graph according to the intersection of the time the timer ends, and the length of time it takes to receive a response. This performance test represents 100 measurements each of two timers against the Noblestar.com web site (Home Page and Page1). According to the 1% part of our definition of an outlier, we may identify 1 individual measurement of each timer as an outlier. According to the “significantly outside the rest of the measurements” part of our definition, there is actually 1 outlier on the chart. With all other transactions taking under 15 seconds, the one that took nearly 250 seconds definitely qualifies as significantly outside the rest of our measurements.

Recreating Performance Response and Scatter Charts in Excel

Since our goal is to eliminate this outlier from our results, we need to discuss a way to do that. TestManager does not have the ability to delete individual measurements from our results. What we can do, is copy the raw data into Excel and recreate our graphs and tables there, without the outliers. If you are familiar with Excel, the next few sections may be obvious to you. If you are new to Excel, I encourage you to follow along with results of your own. We will be expanding on the use of Excel in later articles when we discuss reporting. If you use a spreadsheet program other than Excel, the concepts in the next few paragraphs will still apply, even if the specific steps do not.

Before we start recreating the charts and graphs that we use in TestManager, we must first get the raw data that we want into Excel. Once we launch Excel, we will need to go back to TestManager and get the data we want. First, since we are only interested in the data related to timers, and not all of the individual command IDs, we will limit the results in the table to just timers. To do this, we will once again go to View on the menu bar then Settings. This time we will go to the Select Command IDs tab. By default all timers and command IDs appear in the selected window on the right hand side. We want to select only timers. Press the << button to move all of the timers and command IDs to the Available



window on the left hand side from the Selected window. After doing that select each timer that you want to include, in this case Home Page and Page1, and press the > button to move them back to the Selected window. See Figure 4.

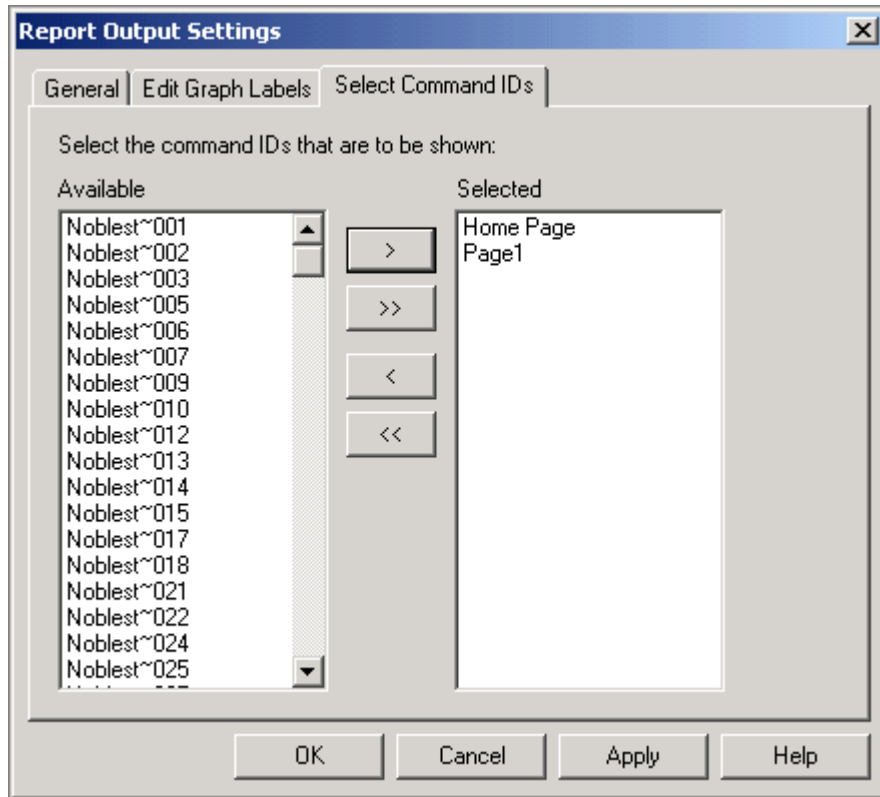


Figure 4 – Response vs. Time Select Command IDs window

When we return to the scatter chart, we see that only the timers are included in the table below the graph. Finally, to copy the data, highlight the first three columns of the table, right click and select copy.

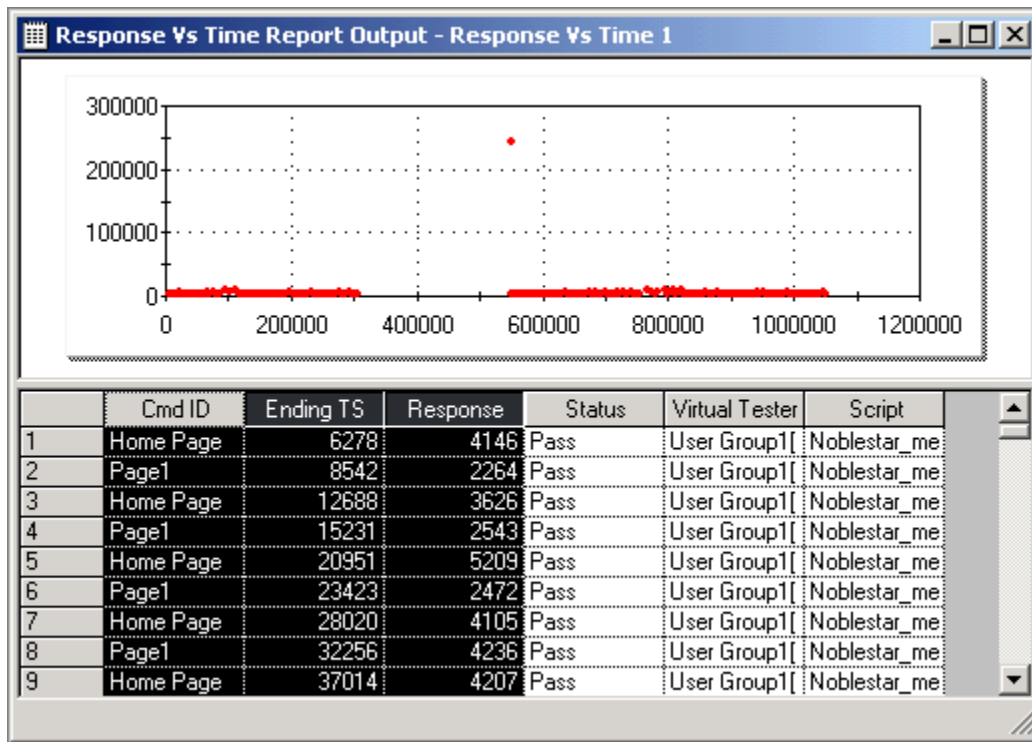
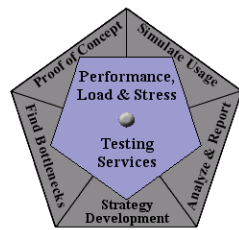


Figure 5 – Scatter Chart data selected

The three columns that we need to recreate this chart are Cmd ID, Ending TS and Response. The Cmd ID column is simply the timer or Command ID name. Ending TS is the Ending Time Stamp or the point in time that the response is completely received as measured from the start of the test execution. The Response column is the total response time of the timer as we discussed in Article 5.

Once copy is selected, simply highlight the top left hand cell in Excel, right-click and select paste.

Creating the Scatter Chart

Since the point of this exercise was to remove the outlier, then recreate our charts and graphs, let's start by removing the outlier. The easiest way to find the outlier is to sort the data by response time. To do this highlight the three columns, then select Data from the Menu bar then Sort. We want to select to sort descending by column C.

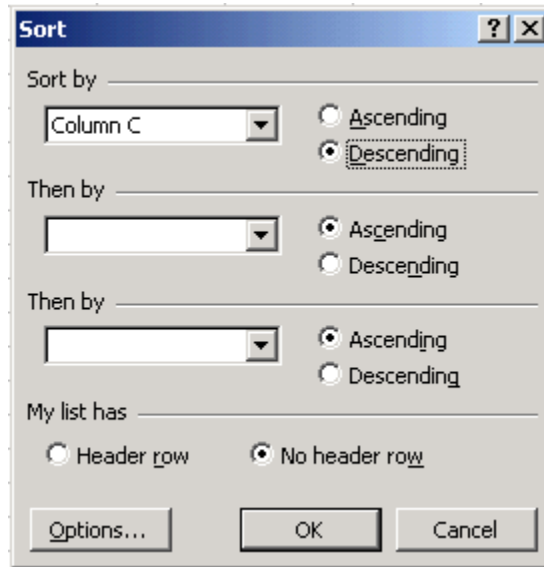
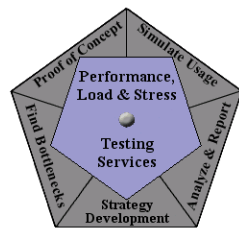


Figure 6 – Excel Sort Window

When I executed this sort, I immediately found the outlier from figure 3.

	A	B	C
1	Home Page	549521	244513
2	Home Page	767373	11045
3	Home Page	796004	10735
4	Home Page	112522	9433
5	Home Page	95977	8732
6	Home Page	808472	8211
7	Home Page	822653	7961
8	Home Page	277730	6961
9	Home Page	879575	6950
10	Home Page	198825	6759

Figure 7 – Excel Sorted Data

Now, delete the row containing the outlier, and it's time to create the new scatter chart. Creating the scatter chart is really quite easy. Highlight columns B and C, then select Insert on the menu bar, and select Chart. On the Standard Types tab, under Chart Types, select XY (Scatter). I won't talk you through all of the possible options and customizations for this chart. Figure 8 shows the chart created by pressing finish after selecting XY (Scatter).

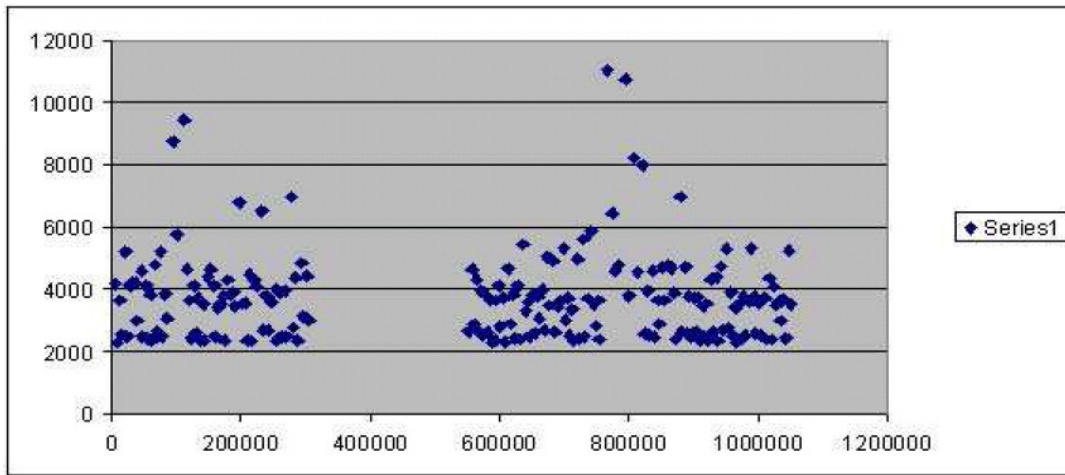
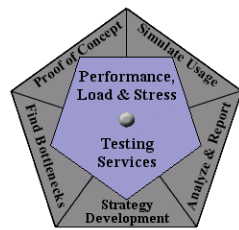


Figure 8 – Scatter Chart in Excel without the Outlier

Comparing this chart to figure 3, you see that the scale on the left hand side only goes up to 12 seconds instead of 250 seconds, that the dots are blue instead of red, and there is a box on the right hand side that says “Series 1”. Otherwise these charts are the same. Figure 9 is a preview of how we will clean this chart up and make it easier to understand during future articles.

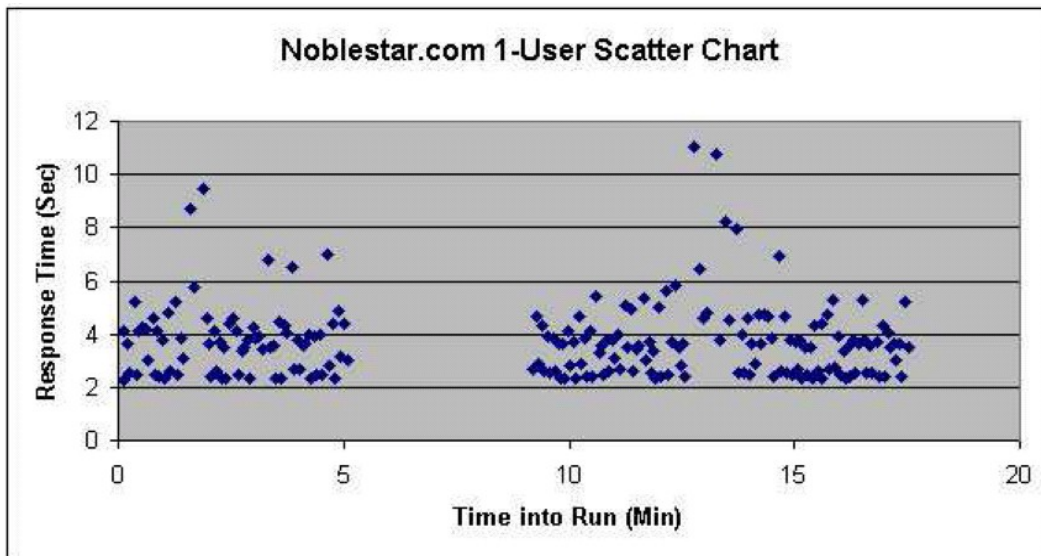


Figure 9 – Scatter Chart in Excel without the Outlier, Customized

Calculating Average, Standard Deviation and Percentiles

The next thing we need to do is recalculate the average, standard deviation, minimum, maximum and percentile values after eliminating the outlier. Figure 10 shows the Performance Report Output from the test run as it appears in TestManager with the outlier. I have modified the settings of this chart to include just the two timers we have been discussing in the same way the Scatter chart settings were modified.

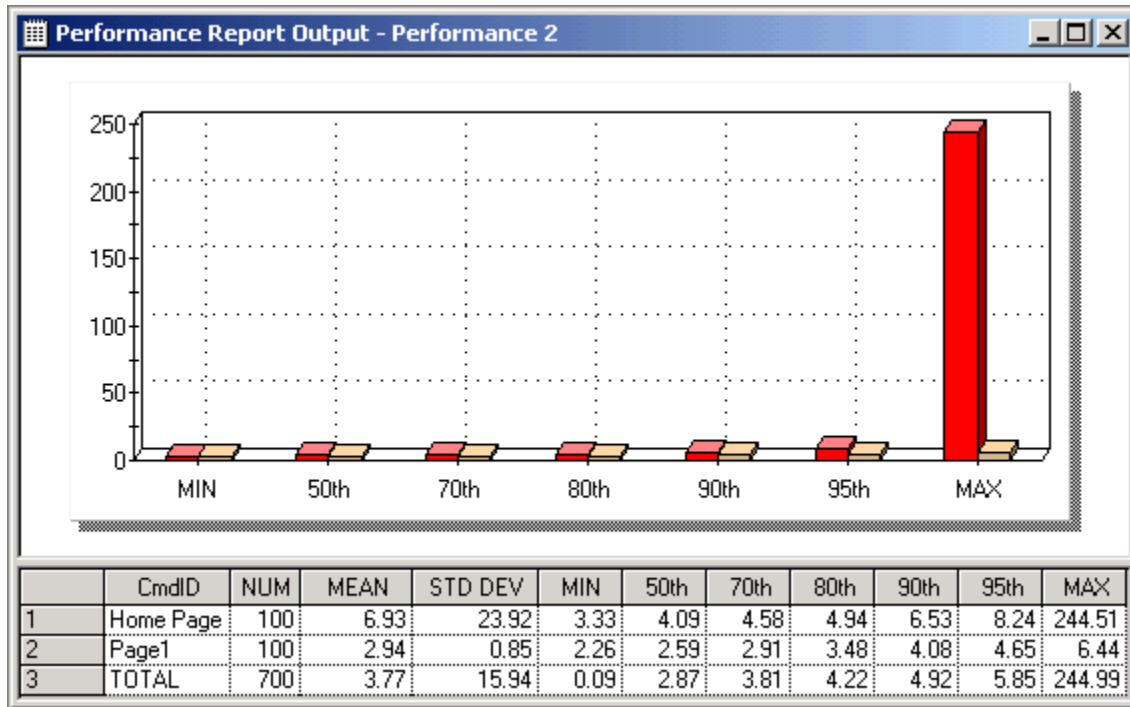
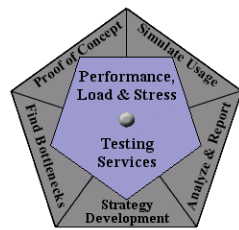


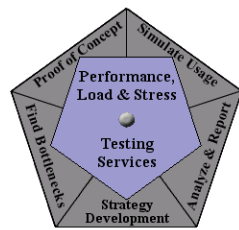
Figure 10 – Performance Report Output in TestManager

Since we have not eliminated any of the Page1 measurements, we won't be doing any calculations with these measurements. To calculate the new Home Page values, we use the values that we have already pasted into Excel. We start by highlighting the three columns of data, copying them and pasting them onto Sheet2. The next step is to separate the Home Page measurements from the Page1 measurements. To do this we will once again highlight all three columns, go to the Data menu option and select sort. This time we will sort ascending on Column A. Once this is done, scroll down the page, highlight all of the rows that start with Page1 and delete them, since we won't be using them for this set of calculations. I find it helps me stay organized to now add my column headers and row labels before starting the calculations. Once you have done this, the first 10 rows of the spreadsheet should look like this.

Timer	Start Time	Resp Time
Home Page	767373	11045
Home Page	796004	10735
Home Page	112522	9433
Home Page	95977	8732
Home Page	808472	8211
Home Page	822653	7961
Home Page	277730	6961
Home Page	879575	6950
Home Page	198825	6759
Home Page	231954	6510

CmdID	NUM	MEAN	STD DEV	MIN	50th	70th	80th	90th	95th	MAX
Home Page										
Page1										

Figure 11 – First Ten Rows of Excel Worksheet



Now we just have to add the calculations. The same basic process applies for determining all 10 values. Place your cursor in the cell next to Home Page below NUM. Next select Insert on the Menu bar and select Function. All of the functions that we will need are in the Statistical Function category. The first function we will need is COUNT.

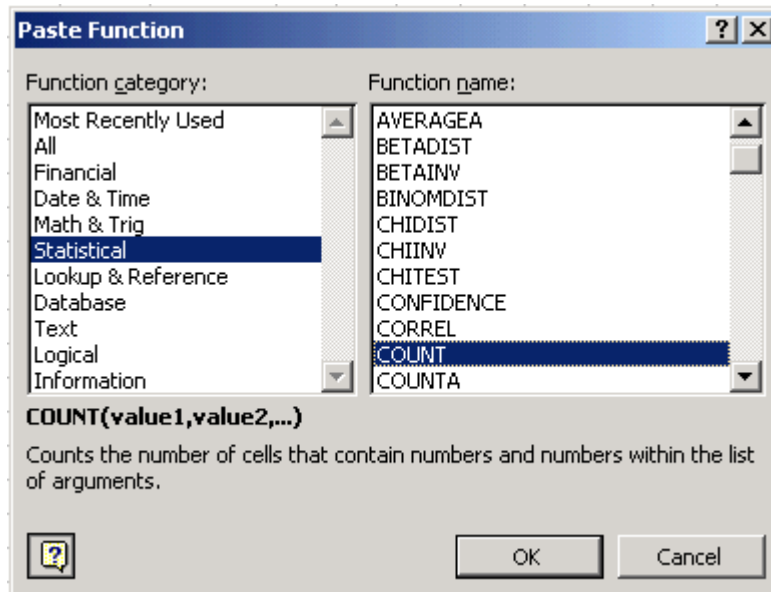


Figure 12 – Excel Paste Function Wizard - Part 1

When the next screen appears, all you have to do is highlight all of the values in Column C and press Enter or click on the OK button.

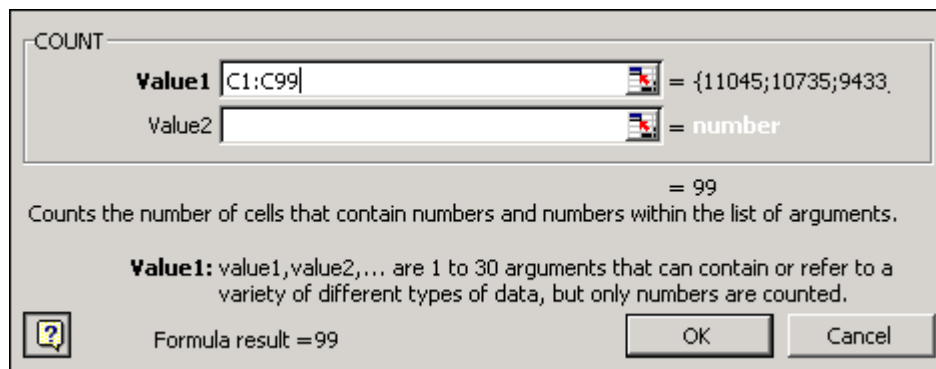
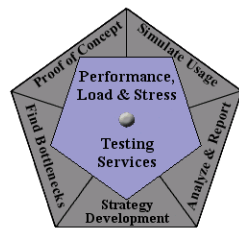


Figure 12 – Excel Paste Function Wizard – Part 2

The same procedure is followed exactly for calculating the MEAN, STD DEV, MIN, and MAX. The function names for finding those values are AVERAGE, STDEV, MIN, and MAX respectively. If you are following along, you will notice that all of the values in the spreadsheet are in milliseconds. To convert to seconds as we see in TestManager, we simply need to divide the results by 1000. To do that, first put your cursor in the cell with the value calculated under the MEAN heading. In the formula bar the function will appear rather than the value. On this spreadsheet the formula currently is



“=AVERAGE(C1:C99)”. To convert the value to milliseconds, simply add “/1000” to the formula so it looks like “=AVERAGE(C1:C99)/1000” then press enter. We will make this same modification for every function other than COUNT.

Calculating the percentile values follows much the same procedure, except this time use the PERCENTILE function. In the Array section, highlight the values once again (in this case C1:C99), and in the K section enter the decimal number representing the percentile we want to calculate. For example, the K value for a 50th Percentile is 0.5. See figure 13.

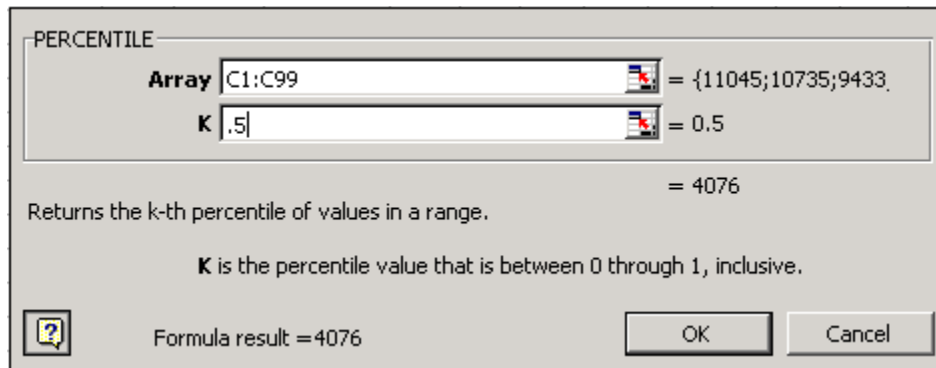


Figure 13 – Excel Paste Function Wizard – Part 2, Percentiles

In case you are not familiar with what the percentile calculations represent, I will pause to explain here, using the 95th percentile as an example. If you have 100 measurements ordered from greatest to least, and counted down the 5 largest measurements, then the next largest measurement represents the 95th percentile of those measurements. For our purposes this measurement would be read “95% of all users would experience a response time of this value or less under the same conditions as the test execution.”

To complete this chart, all we have left to do is copy the results for Page1 and paste them into our new chart and format the cells to the proper number of decimal places. To format the cells to two decimal places, highlight the cells and select Format on the menu bar then select Cells. In the window that appears, under Category, select Number and ensure the number 2 is in the Decimal places section. See figure 14.

The resulting table is below.

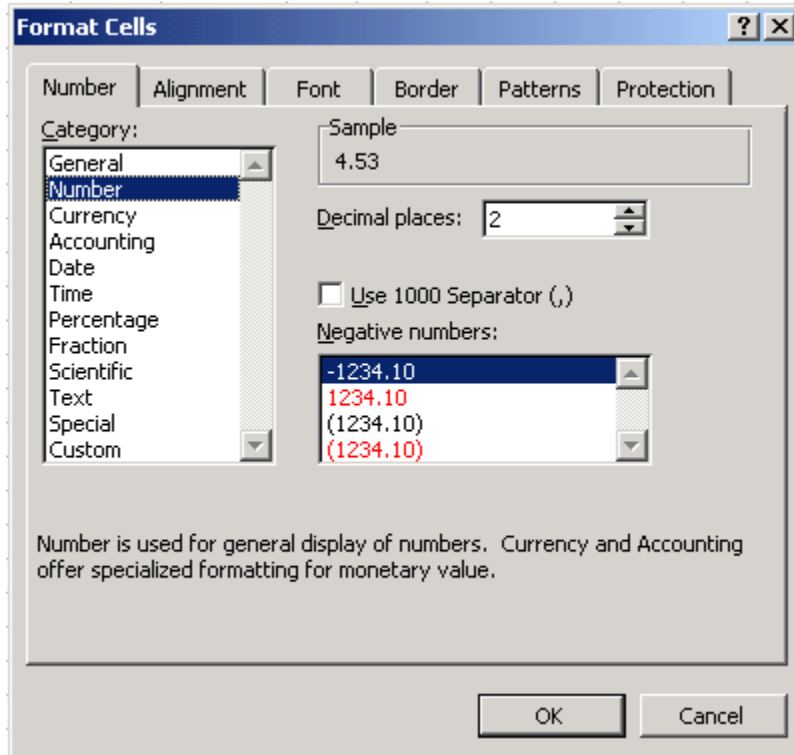
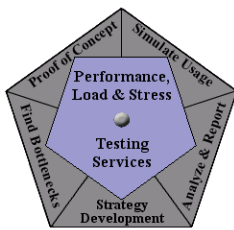


Figure 14 – Excel Format Cells

Notice the difference in the values for the Home Page timer by eliminating just one value. For example, the average is now 4.53 seconds where it was 6.93 before. You will also notice that the Standard Deviation is 1.47 seconds, rather than 23.92. If we were to return to the statistician that gave us the definition about 3 standard deviations, that statistician would tell us that we now have a statistically valid sample. Any time the standard deviation of a sample is greater than the average, the sample is considered to be mathematically skewed, and deserves a closer look to determine validity.

CmdID	NUM	MEAN	STD DEV	MIN	50th	70th	80th	90th	95th	MAX
Home Page	99	4.53	1.47	3.33	4.08	4.57	4.87	5.77	7.99	11.05
Page1	100	2.94	0.85	2.26	2.59	2.91	3.48	4.08	4.65	6.44

Figure 15 – Performance Results Output in Excel

Creating Performance Response Chart

Finally we have reached the last step in our journey to recreating the charts and graphs automatically generated charts by TestManager. To create the graph from the top part of figure 10, we first need to copy the table without the NUM, MEAN, and STD DEV columns as shown below.

Now highlight this entire section and select Insert then Chart from the Menu bar. This time we select the Column Chart type option and press Finish. This results in the chart in figure 16.

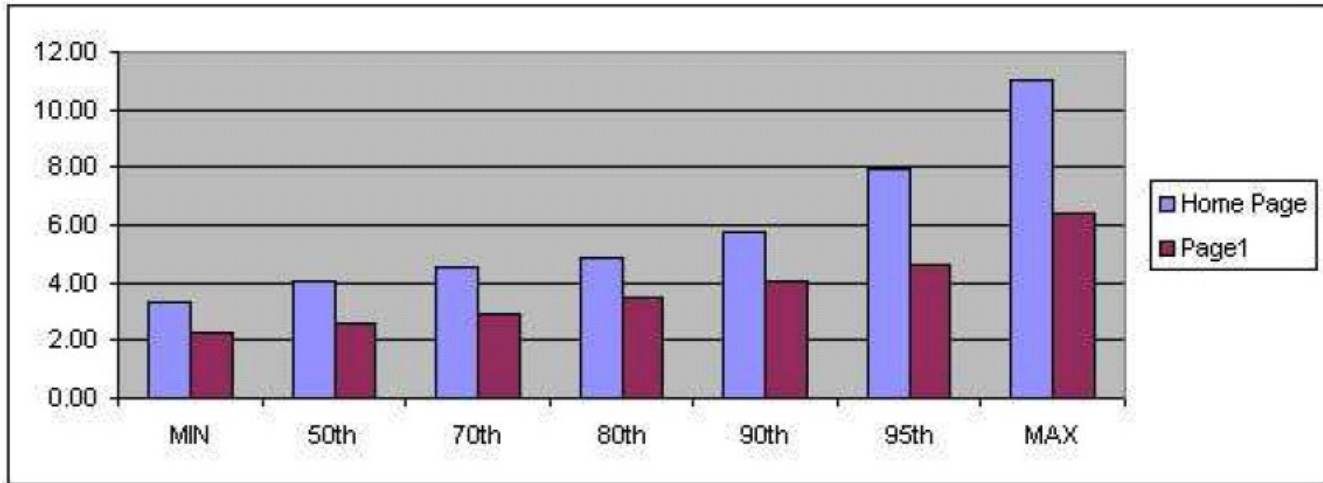
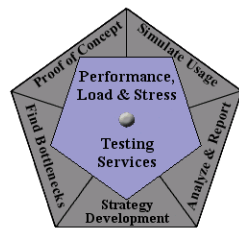


Figure 16 – Performance Report Output in Excel

Again, aside from different colors and values (based on the removal of the outlier), the only difference we see is the box on the right hand side with a key to the colors.

Now You Try It

As I am sure you have noticed, nearly this entire article has been written as a set of instructions to follow. To try these concepts yourself, simply look at any scatter chart from a previous test execution and see if you have any outliers. When you find one that does, and if your testing experience is anything like mine, it won't take long to find one, simply follow the steps in this article to eliminate the outlier and recreate the table and graphs.

Summing It Up

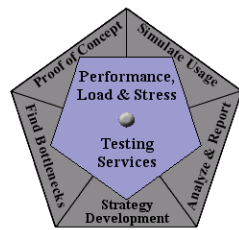
Hopefully the point of this article wasn't lost in the extensive instructions on how to create tables and graphs in Excel. The intent was to show you just how much a single, unrepresentative, measurement can skew your results, and then show you how to eliminate that outlier, rather than having to re-execute the test repeatedly until you get lucky enough to execute a test with no outliers what-so-ever.

Acknowledgments

- The original version of this article was written on commission for IBM Rational and can be found on the [IBM DeveloperWorks](http://www.ibm.com/developerworks) web site

About the Author

Scott Barber is the CTO of PerfTestPlus (www.PerfTestPlus.com) and Co-Founder of the Workshop on Performance and Reliability (WOPR – www.performance-workshop.org). Scott's particular specialties are testing and analyzing performance for complex systems, developing customized testing methodologies, testing embedded systems, testing biometric identification and security systems, group



facilitation and authoring instructional or educational materials. In recognition of his standing as a thought leading performance tester, Scott was invited to be a monthly columnist for Software Test and Performance Magazine in addition to his regular contributions to this and other top software testing print and on-line publications, is regularly invited to participate in industry advancing professional workshops and to present at a wide variety of software development and testing venues. His presentations are well received by industry and academic conferences, college classes, local user groups and individual corporations. Scott is active in his personal mission of improving the state of performance testing across the industry by collaborating with other industry authors, thought leaders and expert practitioners as well as volunteering his time to establish and grow industry organizations.

His tireless dedication to the advancement of software testing in general and specifically performance testing is often referred to as a hobby in addition to a job due to the enjoyment he gains from his efforts.

About PerfTestPlus

PerfTestPlus was founded on the concept of making software testing industry expertise and thought-leadership available to organizations, large and small, who want to push their testing beyond "state-of-the-practice" to "state-of-the-art." Our founders are dedicated to delivering expert level software-testing-related services in a manner that is both ethical and cost-effective. PerfTestPlus enables individual experts to deliver expert-level services to clients who value true expertise. Rather than trying to find individuals to fit some pre-determined expertise or service offering, PerfTestPlus builds its services around the expertise of its employees. What this means to you is that when you hire an analyst, trainer, mentor or consultant through PerfTestPlus, what you get is someone who is passionate about what you have hired them to do, someone who considers that task to be their specialty, someone who is willing to stake their personal reputation on the quality of their work - not just the reputation of a distant and "faceless" company.