

---

# STATGRAPHICS *Plus* 5

---

## Web-Enabled Statistical Software For Six Sigma

Dr. Neil W. Polhemus

January, 2001

STATGRAPHICS *Plus* is published by Manugistics, Inc.  
Distributeur en France: SIGMA PLUS, 29 rue Lauriston, 75116 Paris.



---

## Executive Summary

STATGRAPHICS *Plus 5* is the only full-featured statistical package for Six Sigma which is designed to interact seamlessly with the Internet. It includes all of the statistical capabilities taught in Six Sigma training programs, plus many unique features which make it much more versatile and easy to use than other programs once that training is over.

Important features of STATGRAPHICS *Plus* which are not found in other programs include:

**StatAdvisor** – takes the numeric results from each statistical analysis and incorporates them into a paragraph of plain English. Analysts may paste these paragraphs into their reports and feel confident that the important conclusions of the analysis have been properly conveyed.

**StatLink** – links a collection of statistical analyses to a data source, which may be an Excel file, a centralized database, or an XML file at a remote site.

**StatFolio** – saves analyses in a format which can be run repeatedly whenever the data in the data source changes.

**StatReporter** – integrates the results of one or more analyses into a report prepared by the analyst. Tables and graphs may be linked back to the analyses so that the report updates automatically whenever the analyses are rerun.

**StatPublish** – uploads the results of each analysis (text and graphics) to a web server for internal and external access. When used together with StatLink and StatReporter, this allows analysts to share up-to-the-minute results with customers or others throughout the company.

**World Class Graphics** – includes hundreds of pre-defined graphs which may be easily modified by the user. For special needs, the **StatGallery** provides overlay facilities to place several graphs on the same set of axes.

Designed for companies with a serious commitment to Six Sigma, STATGRAPHICS *Plus* provides the tools necessary to make quality improvement an integral part of corporate data systems.

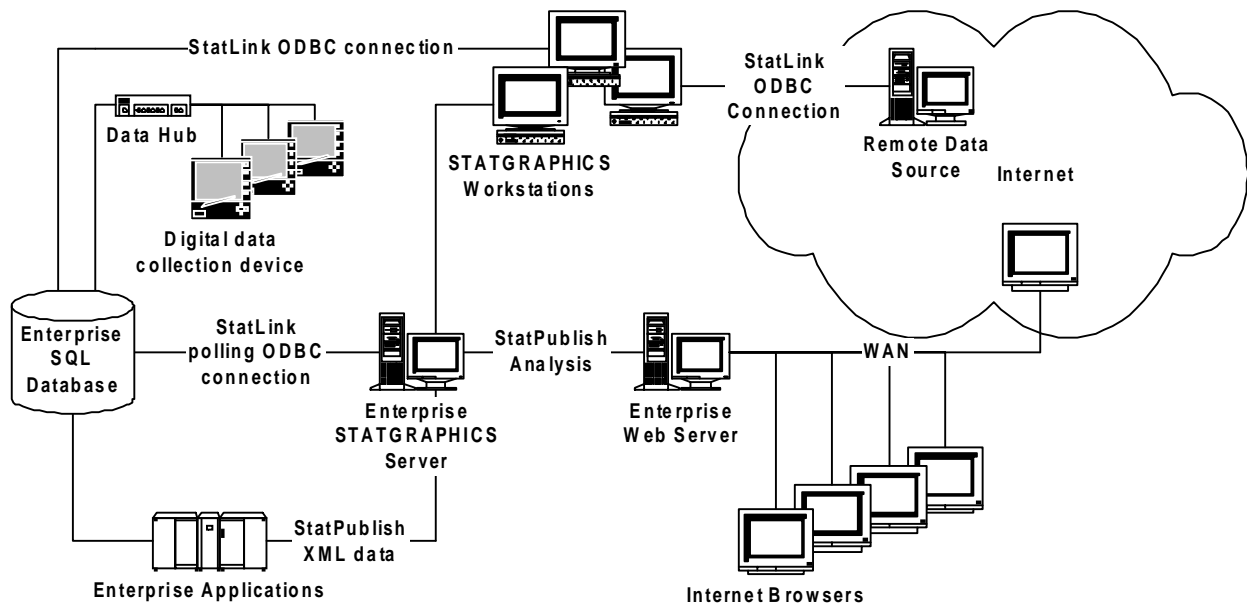
# Statistical Computing in the 21<sup>st</sup> Century

In the 1990's, engineers who wished to analyze data would typically sit down at their computer and:

1. Load a statistical package.
2. Enter or read in some data.
3. Run one or more analyses.
4. Print the results or save them to files.
5. Send copies of the results to whomever they thought might be interested.

There was little integration of the analyses with other corporate-wide systems, either in terms of automated data queries or routine dissemination of results.

With the advent of the Internet, it has now become possible to continuously poll data from sites around the world, and to post the results of statistical analyses to web sites so that corporate management and interested customers can view up-to-date information about product quality. Statistical packages such as STATGRAPHICS are now important parts of the overall enterprise computing environment, as illustrated below.



In the 21st century, data is likely to come from a remote source in another part of the world, and the parties interested in the statistical results are likely to be geographically dispersed.

Statistical analysis in the 21<sup>st</sup> century is therefore much more likely to proceed by:

1. Creating a statistical document.
  2. Linking a data source to the document.
  3. Defining the analyses to be performed on that data.
  4. Publishing the results to a web server.
  5. Polling the data source at regular intervals to automatically update the web site.
- 

## A Capability Analysis Example

As an example of this new computing paradigm, suppose we interested in supplying information to our customers about the capability of a manufacturing process in some overseas location which produces glass bottles. We will suppose that bottles are sampled from the production line every 15 minutes, measurements made, and the results entered into a centralized database at that facility. We will also suppose that the database is structured so that it periodically generates a summary of the most recent 24 hours worth of production in the form of an XML document, which is becoming a standard method for transferring information across the Internet. A fragment of a typical XML document might take the following form:

```
<DATAFILE_sample>
  <OBSERVATION>
    <Date>1/10/01</Date>
    <Time>0:15</Time>
    <Strength>255</Strength>
  </OBSERVATION>
  <OBSERVATION>
    <Date>1/10/01</Date>
    <Time>0:30</Time>
    <Strength>232</Strength>
  </OBSERVATION>
</DATAFILE_sample>
```

A modern web-enabled statistical package must be able to capture that data, do an assessment of the capability of the process, and publish the results to a web site for viewing by a web browser.

With STATGRAPHICS *Plus* 5, such a task is easy to accomplish.

## Step 1: Load STATGRAPHICS and create a StatFolio.

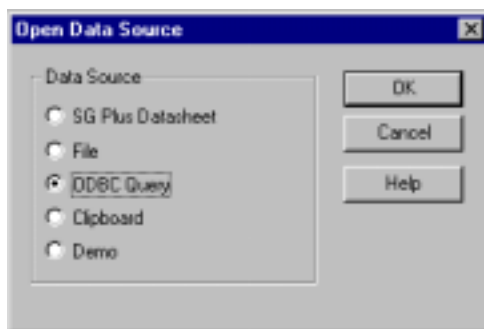
Like Excel, Word, PowerPoint, and other Windows programs, STATGRAPHICS Plus is document centric. Upon entering the program, the user creates a statistical document called a StatFolio to contain the analyses to be performed on the data of interest.

**StatFolio**  
(main STATGRAPHICS document)

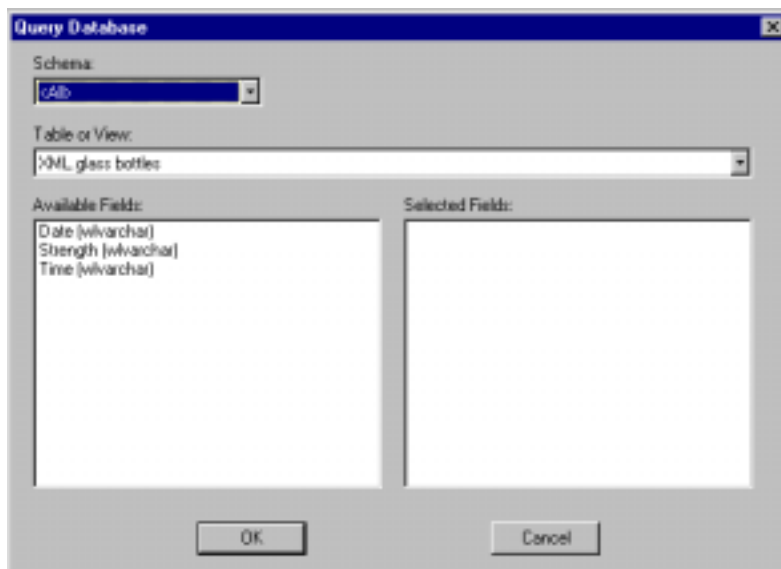
Each StatFolio points to a data source from which it retrieves the data to be analyzed. It also defines the analyses to be run and the location in which the output will be saved.

## Step 2: Define the data source.

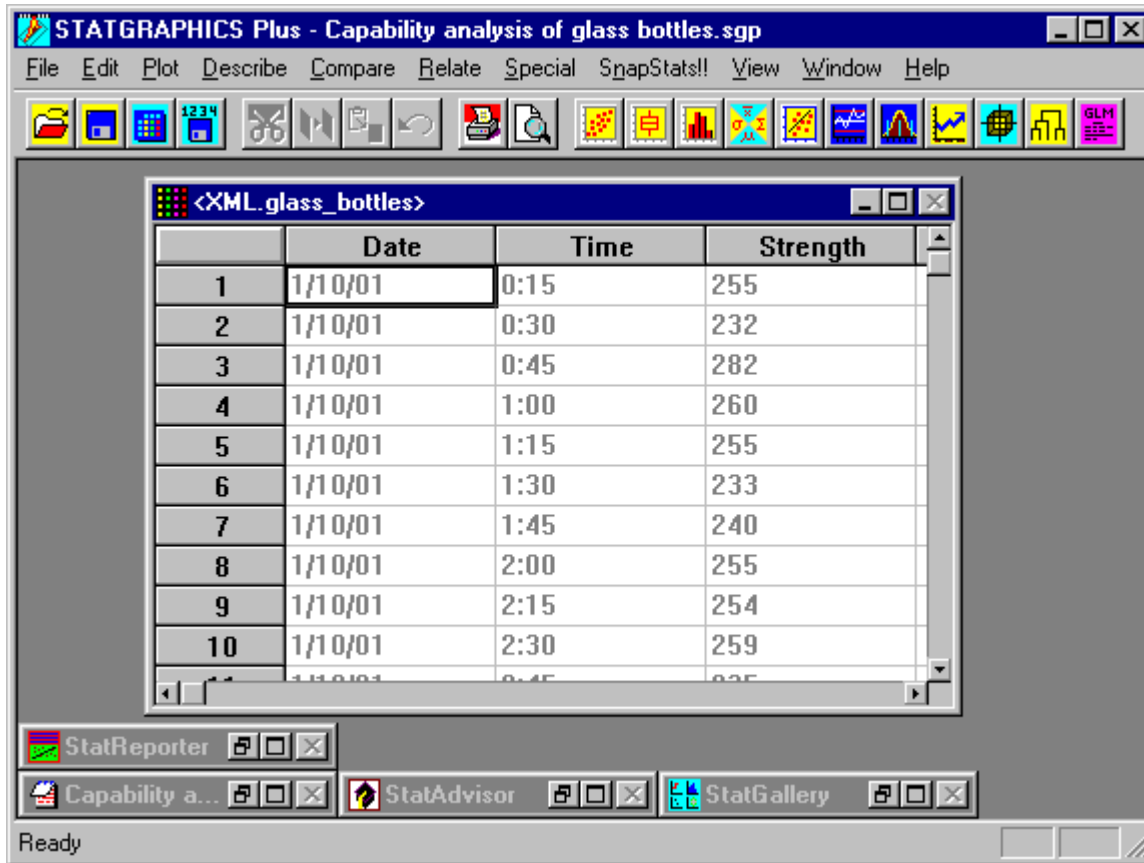
When a new StatFolio is created, the first step is to define the source of the data to be analyzed. Through a feature called StatLink, the analyst can specify a data source on his or her computer, elsewhere on the corporate network, or on any server connected to the Internet. In our example, we might define a link to an ODBC-compliant database:



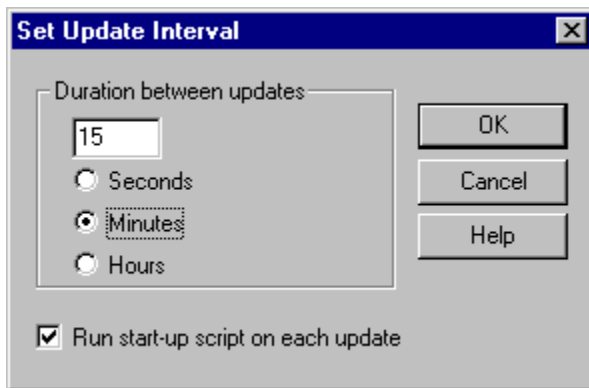
As part of the process, we can specify a set of rows to read and a SQL query if we wish to limit the analysis to certain records:



The data is then read and placed into the STATGRAPHICS data sheet:



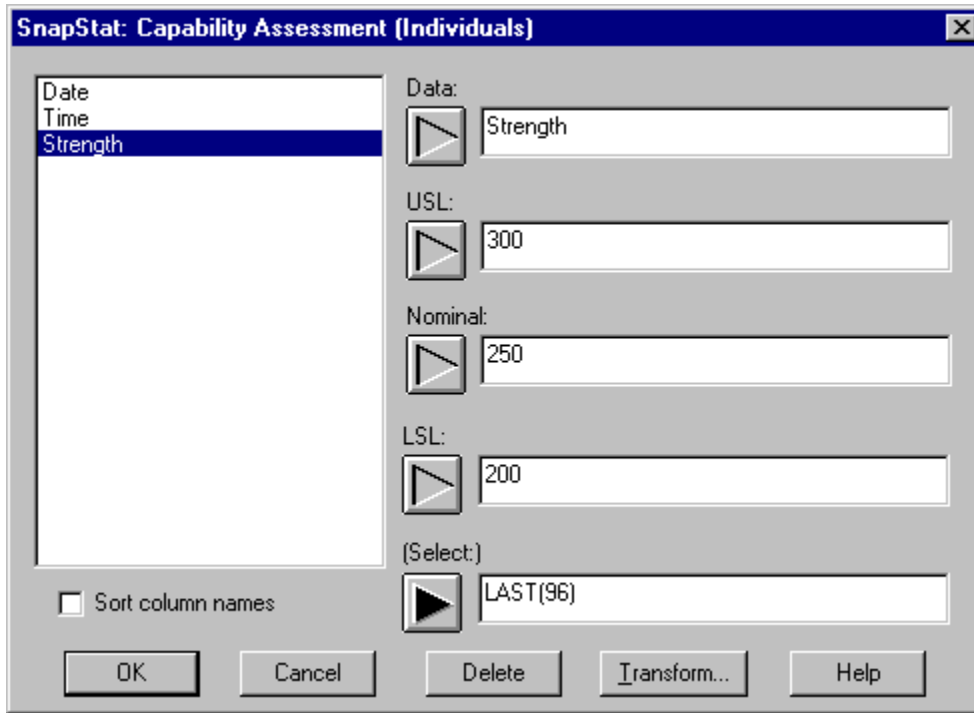
If we want, we can also define an update interval:



This instructs the StatFolio to query the data source every 15 minutes and update its analyses using any new data. It also tells the StatFolio to rerun its start-up script every time the data is polled (see Step 6 below).

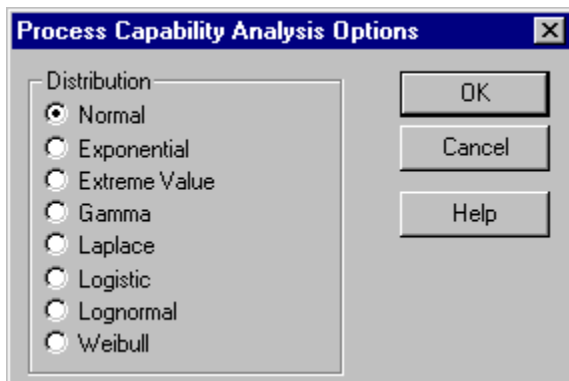
### Step 3: Specify the analyses to be performed.

We now need to specify the analyses we wish to perform on the data. In the current example, a natural choice is a Capability Assessment SnapStat, which produces a one-page summary comparing the data to the specification:

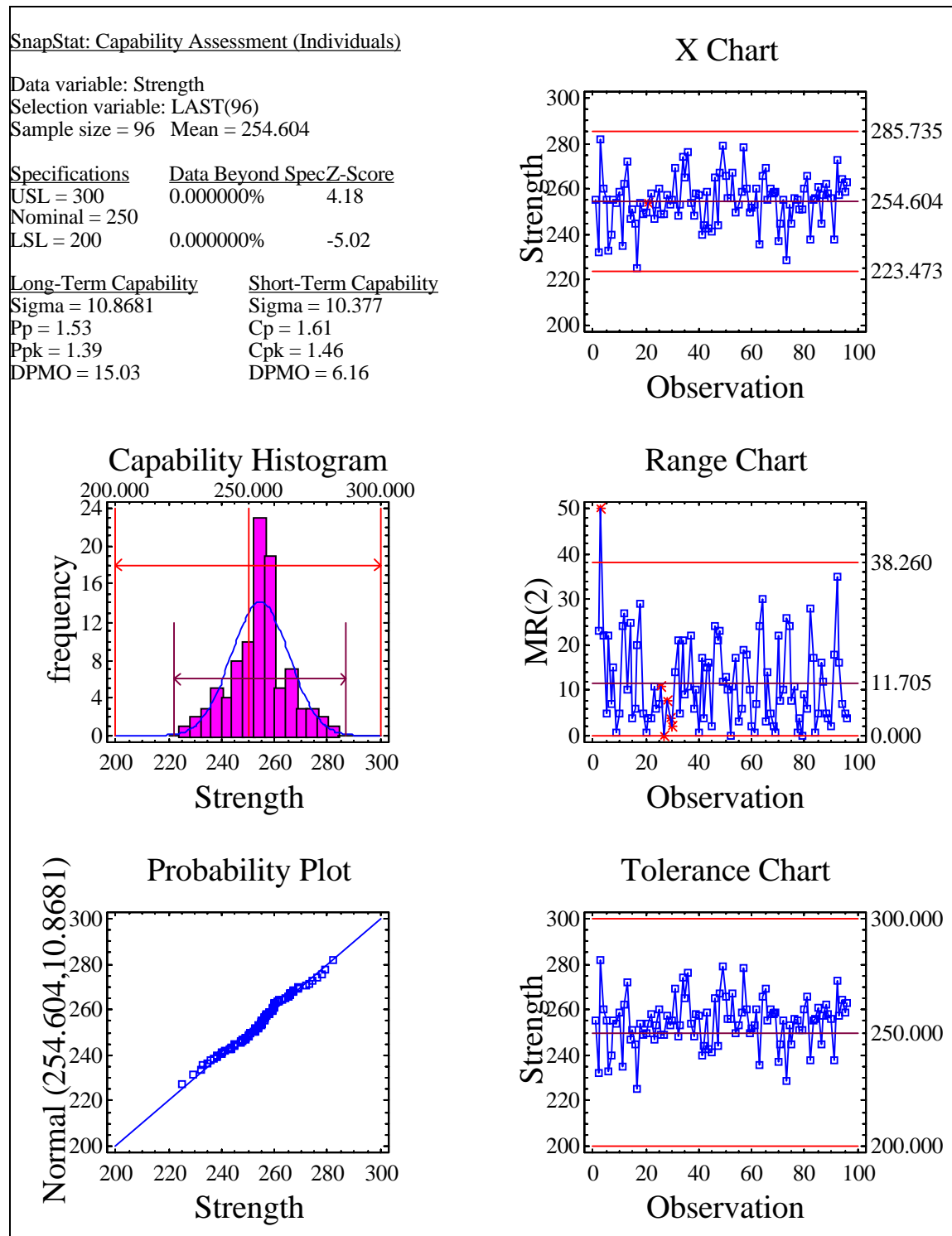


Using the "Select" field, we can restrict the analysis to the last 24 hours.

Defining a new analysis opens an analysis window. By default, the analysis assumes a normal distribution for the data, although we may select another distribution instead:



The output of the Capability Assessment SnapStat is shown below:

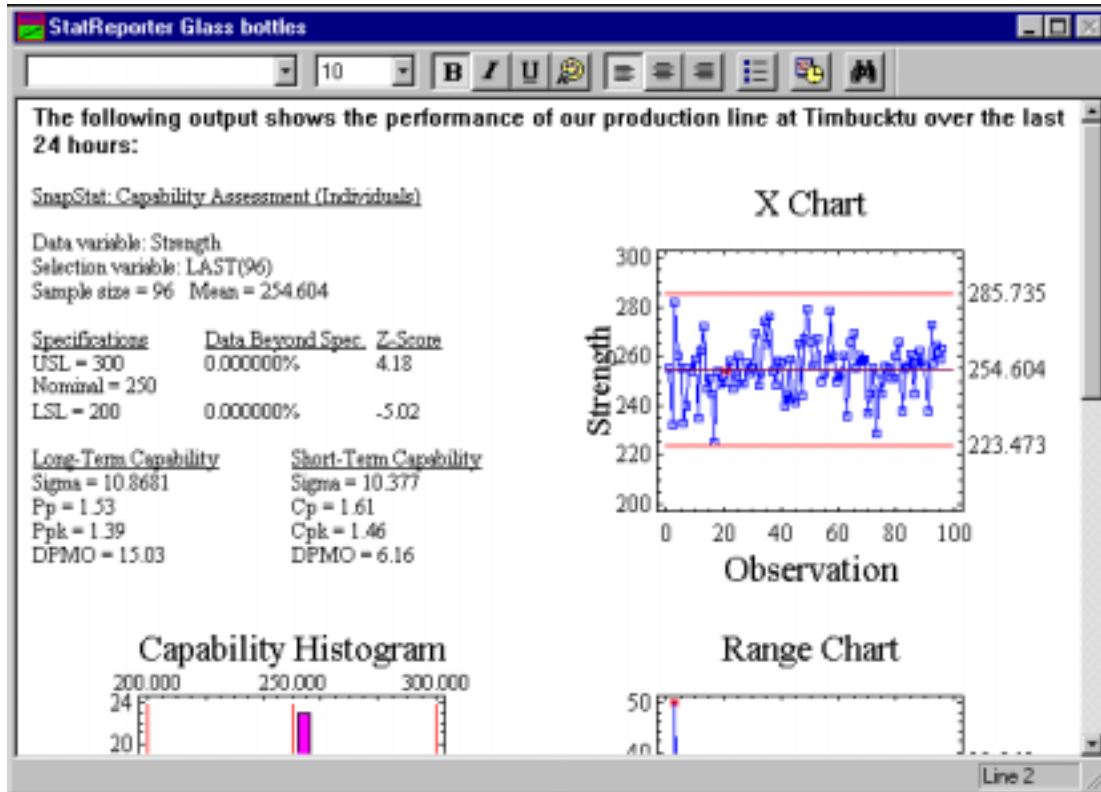


The analysis computes both short-term and long-term capability indices. It also displays control charts, a capability histogram, a normal probability plot, and a tolerance chart.



#### Step 4: Create a report using the StatReporter.

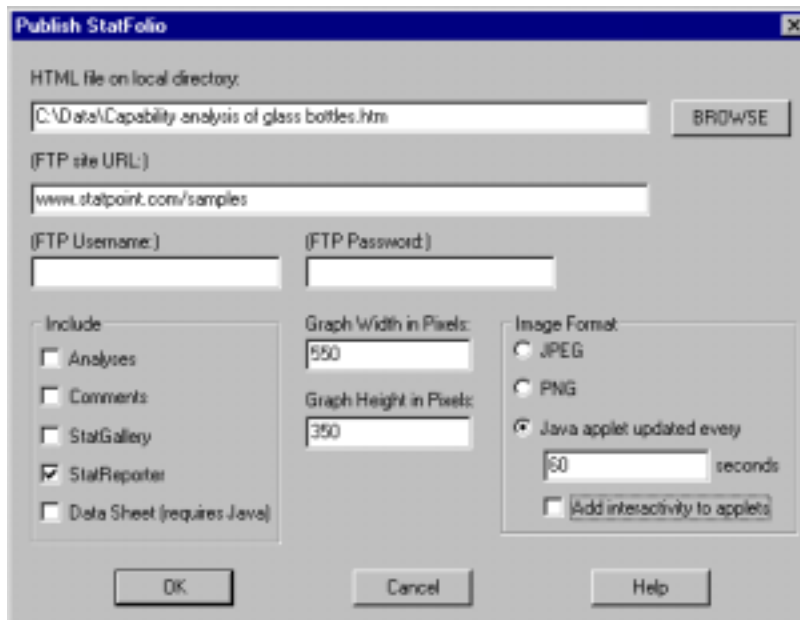
The results of the Capability Assessment may be integrated into a report by first copying the output to the Windows clipboard and then using “Paste Link” to place the analysis in the StatReporter. The StatReporter is a separate window within STATGRAPHICS where users may enter text or paste results:



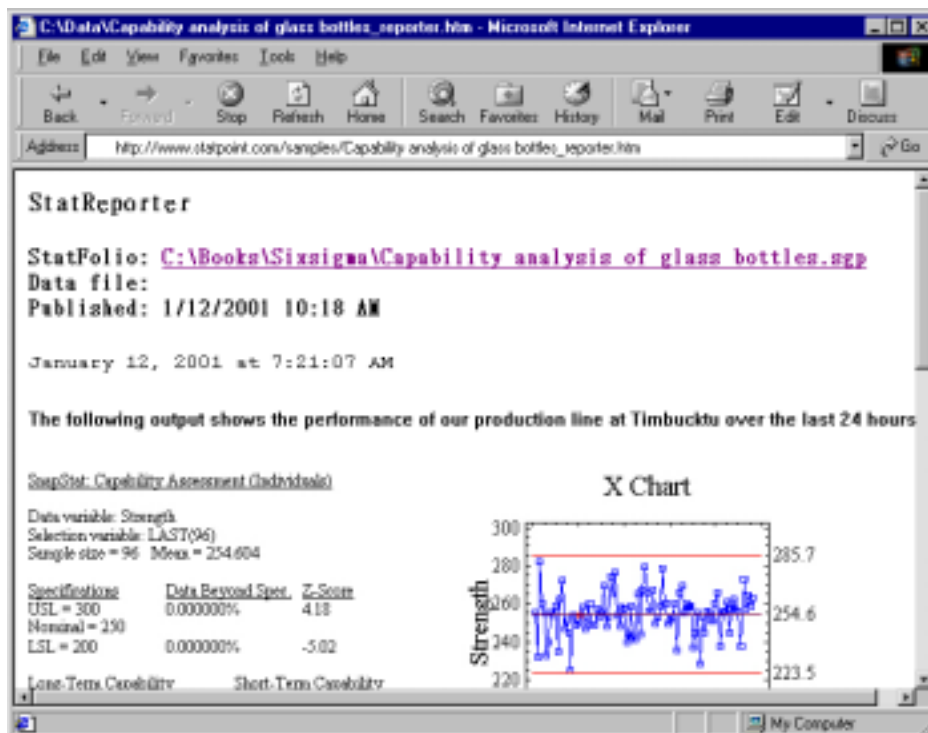
StatReporter output is stored in rtf (Rich Text Format) files which can easily be manipulated in Word, PowerPoint, and many other programs. Since the results were placed in the StatReporter using “Paste Link”, the contents of the report will update whenever the output of the analysis changes, which will be every 15 minutes when the data source is polled.

## Step 5: Publish the results to a web site.

To make the output available to others, the output can be directed to a web site using StatPublish. The location of the web site is defined on the following dialog box:

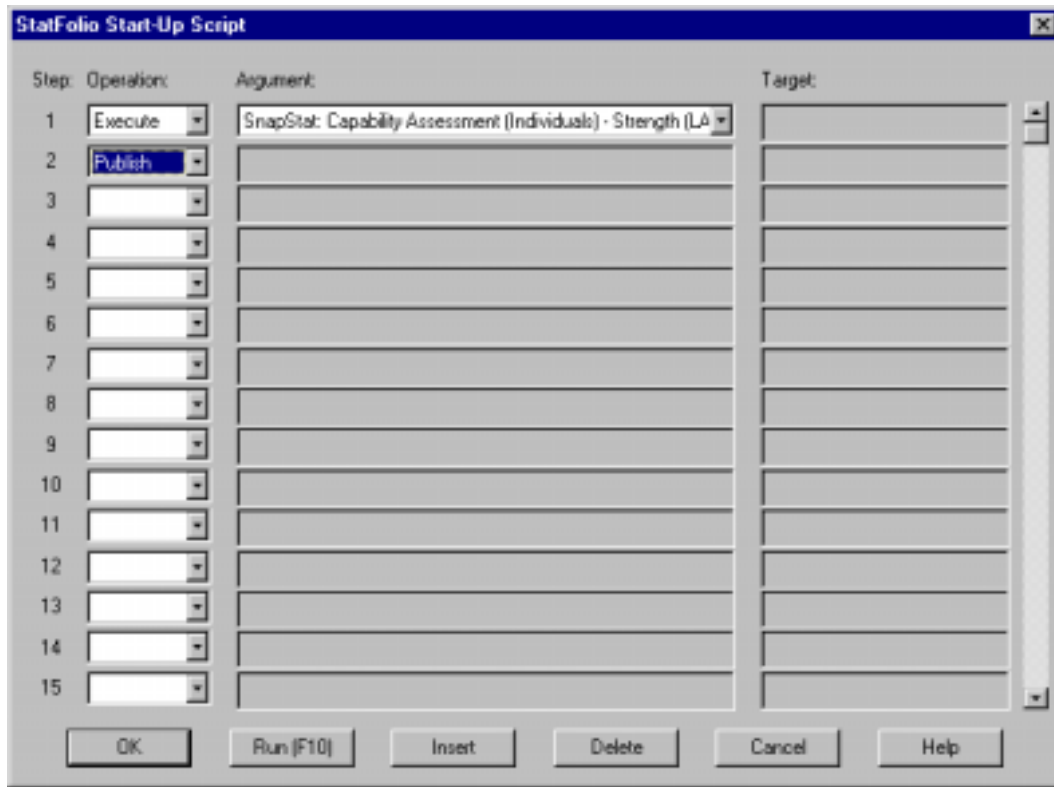


In this case, we have asked for the output of the StatReporter to be placed at [www.statpoint.com/samples](http://www.statpoint.com/samples), in the form of Java applets which will be updated every 60 seconds. This means that an interested party could view our output from within a web browser and expect the results to be automatically refreshed every 60 seconds:



## Step 6: Define a script to be executed whenever the data changes.

Finally, we will define a script to be executed each time the data source is polled:



The script requests that the Capability Assessment be reexecuted and the StatFolio republished whenever the data changes. Since the StatReporter is linked to the Capability Assessment, the web site will automatically display the results of the new analysis.

## Step 7: Save the results.

When the StatFolio is resaved, we now have a document which will:

1. Query the data source every 15 minutes.
2. Rerun the capability assessment on the most recent data.
3. Automatically publish the results to our web site.

Obviously, much more complicated scenarios are possible. For example, the computed statistics could be saved to an XML file and passed to another application for further processing. What makes STATGRAPHICS *Plus* unique is its ability to retrieve data from anywhere in the world, compute the desired statistics, and automatically share the results with anyone who might be interested.

---

# The StatAdvisor

One of the unique features of STATGRAPHICS Plus which makes it ideal for individuals who don't use statistical methods every day is the StatAdvisor. The StatAdvisor takes the statistical output and translates it into simple English in a way that everyone can understand. For example, the output below shows both the standard output and the StatAdvisor output for a typical two-sample comparison

```
Comparison of Means
-----

95.0% confidence interval for mean of Current Design: 0.955071 +/- 0.112711
95.0% confidence interval for mean of New Design: 0.631286 +/- 0.0591332
95.0% confidence interval for the difference between the means
    assuming equal variances: 0.323786 +/- 0.121104    [0.202681,0.44489]

t test to compare means

    Null hypothesis: mean1 = mean2
    Alt. hypothesis: mean1 NE mean2
    assuming equal variances: t = 5.4957    P-value = 0.00000912916

The StatAdvisor
-----

    This option runs a t-test to compare the means of the two samples.
    It also constructs confidence intervals or bounds for each mean and
    for the difference between the means. Of particular interest is the
    confidence interval for the difference between the means, which
    extends from 0.202681 to 0.44489. Since the interval does not contain
    the value 0.0, there is a statistically significant difference between
    the means of the two samples at the 95.0% confidence level.

    A t-test may also be used to test a specific hypothesis about the
    difference between the means of the populations from which the two
    samples come. In this case, the test has been constructed to
    determine whether the difference between the two means equals 0.0
    versus the alternative hypothesis that the difference does not equal
    0.0. Since the computed P-value is less than 0.05, we can reject the
    null hypothesis in favor of the alternative.

    NOTE: these results assume that the variances of the two samples
    are equal. In this case, that assumption is questionable since the
    results of an F-test to compare the standard deviations suggests that
    there may be a significant difference between them. You can see the
    results of that test by selecting Comparison of Standard Deviations
    from the Tabular Options menu.
```

The StatAdvisor highlights the important conclusions, warns of any assumptions which may be violated, and makes suggestions about other procedures to run. Many analysts like to cut and paste StatAdvisor output into their own reports.

# World-Class Graphics

Over 20 years ago, STATGRAPHICS Plus broke new ground as the first statistical package to integrate high-resolution graphics into each of its statistical procedures. We continue to believe that graphical statistics are as important as numerical statistics, since they are unmatched in their ability to communicate information.

Each statistical procedure in STATGRAPHICS has several graphs integrated directly into that procedure. Some examples are shown below:

