Statgraphics

Stratus:
Data Analysis in the Cloud

Presented by Dr. Neil W. Polhemus
Statgraphics Product Line

• Statgraphics *Centurion* – flagship product for Windows with over 250 procedures.

• Statgraphics *Sigma Express* – add-on for Microsoft Excel with emphasis on Six Sigma techniques.

• Statgraphics *Stratus* – streamlined version of Centurion for use within web browsers.
Stratus

• Runs on PCs, MACs, iPads and other tablets.

• Available over the Internet at www.statpoint.net.

• May also be placed on an enterprise server. Requires a Windows server running ASP.Net.

• Unlike Statgraphics Centurion, all computation is done on the server.
### Current data file: Untitled

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comment</th>
<th>Nonmissing Values</th>
<th>Numeric Values</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>Col_1</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_2</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_3</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_4</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_5</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_6</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_7</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_8</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_9</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Col_10</td>
<td></td>
<td>0</td>
<td>0</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>
Accessing Statgraphics *Stratus*

1. Go to: **www.statpoint.net**.

2. Create username and password.

3. Click on link in automated email to approve account.

4. Login using your username.
Usernames

- Used to identify your datasets.

- Types of users:
  - **Guests**: can only analyze sample datasets.
  - **Registered users**: free access for small datasets (100 rows by 10 columns).
  - **Subscriber**: monthly access fee to analyze larger datasets.
  - **Academic site license users**: access for all individuals at an institution.
Ways to Access Data

1. Type data into the data editor.
2. Open a sample dataset.
3. Open a dataset from your computer.
4. Open a dataset you have saved on the server.
5. Paste data from the clipboard.
Example: Old Faithful Geyser

Yellowstone National Park
Example: Opening an Excel File

Open User Dataset from Client

Step 1: Select the file to open: Browse... old faithful.xlsx

Step 2: Select the file attributes (except for SGD files):

- Header: ○ Column names ○ Column names and comments ○ None
- Delimiter (text files only): ○ Tab ○ Comma () ○ Space ○ Semicolon (;)
- Decimal symbol: ○ Period (.) ○ Comma (,)
- Date separator: ○ Slash (/) ○ Period (.) ○ Dash (-)
- Date format: ○ M D Y ○ D M Y ○ Y M D

Step 3: Press the Open File button to read the file:
- Open File
- Open in readonly mode
- Cancel
## File Summary

### Current data file: old faithful.xlsx

<table>
<thead>
<tr>
<th>Variable</th>
<th>Comment</th>
<th>Nonmissing Values</th>
<th>Numeric Values</th>
<th>Minimum</th>
<th>Maximum</th>
</tr>
</thead>
<tbody>
<tr>
<td>duration</td>
<td>minutes</td>
<td>272</td>
<td>272</td>
<td>1.6</td>
<td>5.1</td>
</tr>
<tr>
<td>waiting time</td>
<td>minutes between eruptions</td>
<td>272</td>
<td>272</td>
<td>43</td>
<td>96</td>
</tr>
</tbody>
</table>
## Display Data

Current data file: old faithful.xlsx

<table>
<thead>
<tr>
<th>duration (minutes)</th>
<th>waiting time (minutes between eruptions)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>79</td>
</tr>
<tr>
<td>2</td>
<td>54</td>
</tr>
<tr>
<td>3</td>
<td>74</td>
</tr>
<tr>
<td>4</td>
<td>62</td>
</tr>
<tr>
<td>5</td>
<td>85</td>
</tr>
<tr>
<td>6</td>
<td>85</td>
</tr>
<tr>
<td>7</td>
<td>88</td>
</tr>
<tr>
<td>8</td>
<td>85</td>
</tr>
<tr>
<td>9</td>
<td>51</td>
</tr>
<tr>
<td>10</td>
<td>85</td>
</tr>
<tr>
<td>11</td>
<td>54</td>
</tr>
<tr>
<td>12</td>
<td>84</td>
</tr>
<tr>
<td>13</td>
<td>78</td>
</tr>
<tr>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>15</td>
<td>83</td>
</tr>
<tr>
<td>16</td>
<td>52</td>
</tr>
<tr>
<td>17</td>
<td>62</td>
</tr>
<tr>
<td>18</td>
<td>84</td>
</tr>
<tr>
<td>19</td>
<td>52</td>
</tr>
<tr>
<td>20</td>
<td>79</td>
</tr>
<tr>
<td>21</td>
<td>51</td>
</tr>
<tr>
<td>22</td>
<td>47</td>
</tr>
<tr>
<td>23</td>
<td>78</td>
</tr>
<tr>
<td>24</td>
<td>69</td>
</tr>
<tr>
<td>25</td>
<td>74</td>
</tr>
</tbody>
</table>

1 2 3 4 5 6 7 8 9 10 ...
## Edit Data

### Current data file: old faithful.xlsx

<table>
<thead>
<tr>
<th>A</th>
<th>B</th>
</tr>
</thead>
<tbody>
<tr>
<td>Name</td>
<td>duration (minutes)</td>
</tr>
<tr>
<td>Comment</td>
<td>minutes (minutes between eruptions)</td>
</tr>
<tr>
<td>1</td>
<td>3.6</td>
</tr>
<tr>
<td>2</td>
<td>1.8</td>
</tr>
<tr>
<td>3</td>
<td>3.333</td>
</tr>
<tr>
<td>4</td>
<td>2.283</td>
</tr>
<tr>
<td>5</td>
<td>4.533</td>
</tr>
<tr>
<td>6</td>
<td>2.883</td>
</tr>
<tr>
<td>7</td>
<td>4.7</td>
</tr>
<tr>
<td>8</td>
<td>3.6</td>
</tr>
<tr>
<td>9</td>
<td>1.95</td>
</tr>
<tr>
<td>10</td>
<td>4.35</td>
</tr>
<tr>
<td>11</td>
<td>1.833</td>
</tr>
<tr>
<td>12</td>
<td>3.917</td>
</tr>
<tr>
<td>13</td>
<td>4.2</td>
</tr>
<tr>
<td>14</td>
<td>1.75</td>
</tr>
<tr>
<td>15</td>
<td>4.7</td>
</tr>
<tr>
<td>16</td>
<td>2.167</td>
</tr>
<tr>
<td>17</td>
<td>1.75</td>
</tr>
<tr>
<td>18</td>
<td>4.8</td>
</tr>
<tr>
<td>19</td>
<td>1.6</td>
</tr>
<tr>
<td>20</td>
<td>4.25</td>
</tr>
</tbody>
</table>
Save File on Server

**SAVE DATA FILE**

User files on server: unemployment_rate.xlsx

Save data file as: old_faithful

- STATGRAPHICS file (.sgd)
- Excel file (.xls)
- XML file without comments (.xml)
- Tab delimited text file
- Comma delimited text file
- Blank delimited text file
- Semicolon delimited text file

[Save File] [Cancel]
One Variable Analysis – Data Input

This procedure calculates statistics for a single column of numeric data. 

Data variable (Y): **duration**

(Select:)
One Variable Analysis – Analysis Options

This procedure calculates statistics for a single column of numeric data. [Documentation]

Tabulation Intervals
(Leave fields blank to use system defaults.)

Number of classes: 

Lower limit: 

Upper limit:
One Variable Analysis – Tables and Graphs

The procedure calculates statistics for a single column of numeric data. Documentation

Select All | Unselect All
---
Average | Winsorized sigma | 1/6 sextile
Median | MAD | 5/6 sextile
Mode | Sbi | Interquartile range
Geometric mean | Minimum | Skewness
Trimmed mean | Maximum | Kurtosis
Winsorized mean | Range | Std. kurtosis
Variance | Lower quartile | Std. skewness
Standard deviation | Upper quartile | Sum
Coefficient of variation | Interquartile range | Sum of squares
Standard error

Trimming: 5 %

Box-and-Whisker Plot | Horizontal | Vertical
---
Next plot | median notch | Outlier symbols | Mean marker
Frequency Tabulation
Frequency Histogram | Relative | Cumulative | Display as polygon
One Variable Analysis – Results to Save

This procedure calculates statistics for a single column of numeric data. [Documentation]
One Variable Analysis

This procedure calculates statistics for a single column of numeric data. Documentation

Data file: stratussupport_old_faithful.sgd

**One Variable Analysis - duration**
Data variable: duration (minutes)
272 values ranging from 1.6 to 5.1

The StatAdvisor
This procedure is designed to summarize a single sample of data. It will calculate various statistics and graphs. Also included in the procedure are confidence intervals and hypothesis tests.

**Summary Statistics for duration**

<table>
<thead>
<tr>
<th>Statistic</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>272</td>
</tr>
<tr>
<td>Average</td>
<td>3.48778</td>
</tr>
<tr>
<td>Standard deviation</td>
<td>1.14137</td>
</tr>
<tr>
<td>Coeff. of variation</td>
<td>32.7248%</td>
</tr>
<tr>
<td>Minimum</td>
<td>1.6</td>
</tr>
<tr>
<td>Maximum</td>
<td>5.1</td>
</tr>
<tr>
<td>Range</td>
<td>3.5</td>
</tr>
<tr>
<td>Stnd. skewness</td>
<td>-2.81541</td>
</tr>
<tr>
<td>Stnd. kurtosis</td>
<td>-5.07051</td>
</tr>
</tbody>
</table>
Frequency Histogram

[Image of Frequency Histogram]

- The x-axis represents the duration, ranging from 1.4 to 5.4.
- The y-axis represents the frequency, ranging from 0 to 40.
- The histogram displays two peaks, indicating two distinct frequency distributions within the given duration range.
Changing Analysis Options

One Variable Analysis

This procedure calculates statistics for a single column of numeric data. Documentation

Data Input | Analysis Options | Tables and Graphs | Results to Save | Output

Tabulation Intervals
(Leave fields blank to use system defaults)

Number of classes: 40

Lower limit: 1.5
Upper limit: 5.5
# Changing Titles and Scaling

<table>
<thead>
<tr>
<th>Item</th>
<th>Text</th>
<th>Scale From</th>
<th>Scale To</th>
<th>Scale By</th>
<th>Skip Increment</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>Title</td>
<td>Old Faithful Geyser</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Subtitle</td>
<td>272 eruptions</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>X-axis</td>
<td>duration in minutes</td>
<td>1.5</td>
<td>5.5</td>
<td>0.5</td>
<td>0</td>
<td>Log scale, Powers of 10, No power, Skip repeats, Rotate tickmarks</td>
</tr>
<tr>
<td>Y-axis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Log scale, Powers of 10, No power, Skip repeats</td>
</tr>
<tr>
<td>Z-axis</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td>Log scale, Powers of 10, No power, Skip repeats</td>
</tr>
</tbody>
</table>
# Changing Fills and Fonts

## Set Preferences

<table>
<thead>
<tr>
<th>Fill set #1</th>
<th>Fill set #2</th>
<th>Fill set #3</th>
<th>Fill set #4</th>
<th>Fill set #5</th>
<th>Fill set #6</th>
<th>Fill set #7</th>
<th>Fill set #8</th>
<th>Fill set #9</th>
<th>Fill set #10</th>
<th>Fill set #11</th>
<th>Fill set #12</th>
<th>Fill set #13</th>
<th>Fill set #14</th>
<th>Fill set #15</th>
<th>Fill set #16</th>
<th>Fill set #17</th>
<th>Fill set #18</th>
<th>Fill set #19</th>
<th>Fill set #20</th>
</tr>
</thead>
<tbody>
<tr>
<td>SpringGreen</td>
<td>CornflowerBlue</td>
<td>Violet</td>
<td>LightGreen</td>
<td>DarkOrange</td>
<td>PaleTurquoise</td>
<td>Tan</td>
<td>Olive</td>
<td>Khaki</td>
<td>SteelBlue</td>
<td>Tomato</td>
<td>DarkOrchid</td>
<td>SlateGray</td>
<td>Goldenrod</td>
<td>Pink</td>
<td>Chocolate</td>
<td>SpringGreen</td>
<td>Coral</td>
<td>Yellow</td>
<td>LawnGreen</td>
</tr>
</tbody>
</table>

Select new color: **SpringGreen**  
Select new type: **Solid**  

Save  | Cancel  | Color Table:  | Foreground:  | Background:  |
<statgraphics>
   <globals>
      <DataEditorType Value="0"/>
      <FunctionResolution Value="101"/>
      <ContourResolution Value="51"/>
      <BorderColor Value="SkyBlue"/>
      <InputDateFormat Value="M/D/YYYY"/>
      <OutputDateFormat Value="M/D/YYYY"/>
      <GraphTextSize Value="12"/>
      <FillColor1 Value="SpringGreen"/>
      <FillColor10 Value="SteelBlue"/>
      <PointType1 Value="Filled Square"/>
      <PointType17 Value="Square"/>
      <PointType18 Value="X"/>
      <PointType19 Value="Circle"/>
      <PointType20 Value="Plus"/>
   </globals>
   <data Alias="stratussupport_old faithful.sgd" Source="datafiles/stratussupport_old faithful.sgd">
      <DecimalSeparator Value="."/>
      <DateFormat Value="M/D/YYYY"/>
      <MissingValue Value=""/>
   </data>
   <proc name="ONEVAR">
      <input>
         <Y Value="duration"/>
      </input>
      <options>
         <Classes Value="40"/>
         <From Value="1.5"/>
         <To Value="5.5"/>
      </options>
      <output>
         <graph Name="Histogram">
            <TopTitle Value="Old Faithful Geyser"/>
            <Subtitle Value="272 eruptions"/>
            <XAxisTitle Value="duration in minutes"/>
            <XAxisFrom Value="1.5"/>
            <XAxisTo Value="5.5"/>
         </graph>
      </output>
   </proc>
</statgraphics>
Rerun Script

Statgraphics Stratus
Generated: 9/28/2013 11:45:10 AM
Data file: stratussupport_old faithful.sgd

One Variable Analysis - duration
Frequency Histogram

Old Faithful Geyser
272 eruptions
Statlets

- *Stratus* also contains implementations of the interactive Statlets introduced in Centurion XVII.

- Statlets use HTML5 canvases and Javascript to let users interact with the graphs.

- Very useful for rotating 3D graphs and visualizing the effect of changing procedure options.
**Interactive Histogram Statlet**

This applet displays a frequency histogram for a column of numeric data. You may change the definition of the intervals into which the data are tabulated. You may also display a fitted normal curve or a nonparametric density trace. [Documentation](#)  

**Data:** duration  

(Select:)  

- [x] Show histogram  
  - From: 0  
  - To: 0  
  - Classes: 0  
  - Use default scaling  

- [ ] Add normal curve  
- [x] Add density trace with width: 50%  

Update Plot  
Titles and scaling
Statlets - Output

Frequency Histogram
Mean=3.48778 Sigma=1.14137
Add to Session Logfile

• Push button to add contents of output button to the session logfile.

• View by selecting File – Manage Files – Manage Session Logfiles.
Session Logfile
Bivariate Histogram
Linear Regression

Waiting time = 33.4744 + 10.7296*duration
RMSE: 5.91401, R-squared: 81.15%, P-value: 0.0000
Entering Expressions

**X-Y Scatterplot**

This procedure plots data contained in two data columns. [Documentation]

**Data Input**

**Y:** duration

**X:**

(Point Codes)

(Standard Errors for X:)

(Standard Errors for Y:)

(Select:)

[Help]

[Log In]
X-Y Scatterplot
Interactive Map Statlet

Median Age

States highlighted in blue have a lower median age compared to the rest of the states, which are highlighted in red.
Sample Size Determination Statlet

This Statlet determines the sample size needed to estimate or test values of various parameters. The size may be based on either the width of a confidence interval or the power of a hypothesis test. Documentation

Parameter to be estimated: Capability index Cpk

Null hypothesis: 1.33  Alpha risk: 5 %

Alt. hypothesis: 1  Type: Two-sided

Base sample size on: Relative error  Relative error: 10 %

Update Plot  Titles and scaling  Plot alternative hypothesis  Plot confidence limits
Visualizers
Stock Price Plot

Candlestick Plot for FB
Smother: MA(20) Trading Bands: Bollinger(2 sigma)

Date: 2015-09-25
Open: 95.79
High: 95.85
Low: 92.06
Close: 92.77

Upper band: 97.66
Smooth: 91.75
Lower band: 85.84

%b: 0.59
BW: 0.13
Volume: 28803200.00
Recorded Webinar

• You may find the recorded webinar, PowerPoint slides and sample data at:
  
  www.statgraphics.com

• Look for “Instructional Videos”.

• Try Stratus at www.statpoint.net