statgraphics 18®
centurion

Introducing Statgraphics 18

Presented by

Dr. Neil W. Polhemus

Statgraphics 18

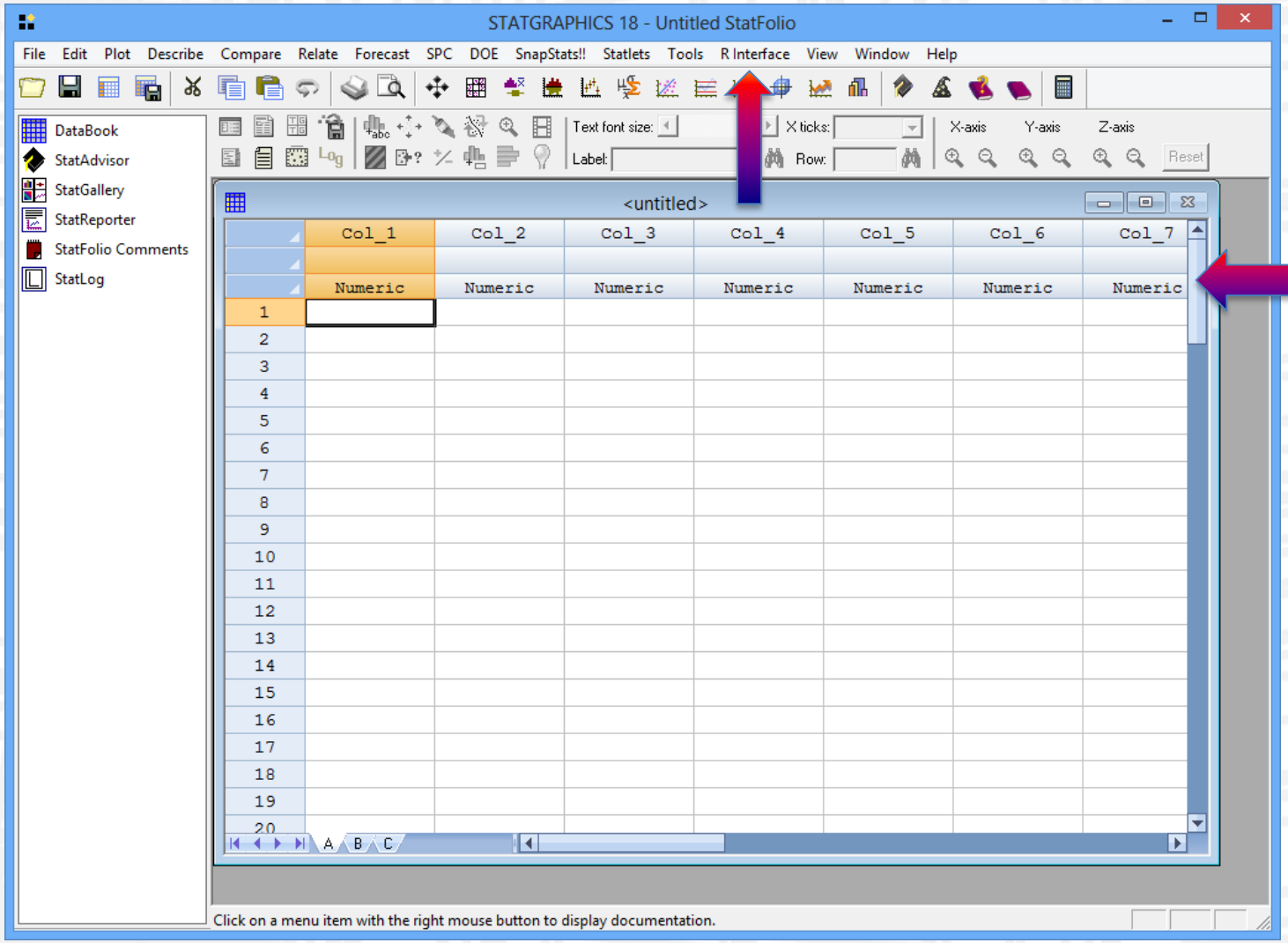
- The 18th version of Statgraphics for PCs.

- Featuring:
  - 30 new statistical procedures
  - Significant enhancements to 18 existing procedures
  - New file formats and improved methods for handling “Big Data”
  - Expanded dialog-based interface to R procedures
  - Streamlined activation
  - Concurrent-user network license management with “check-out” feature

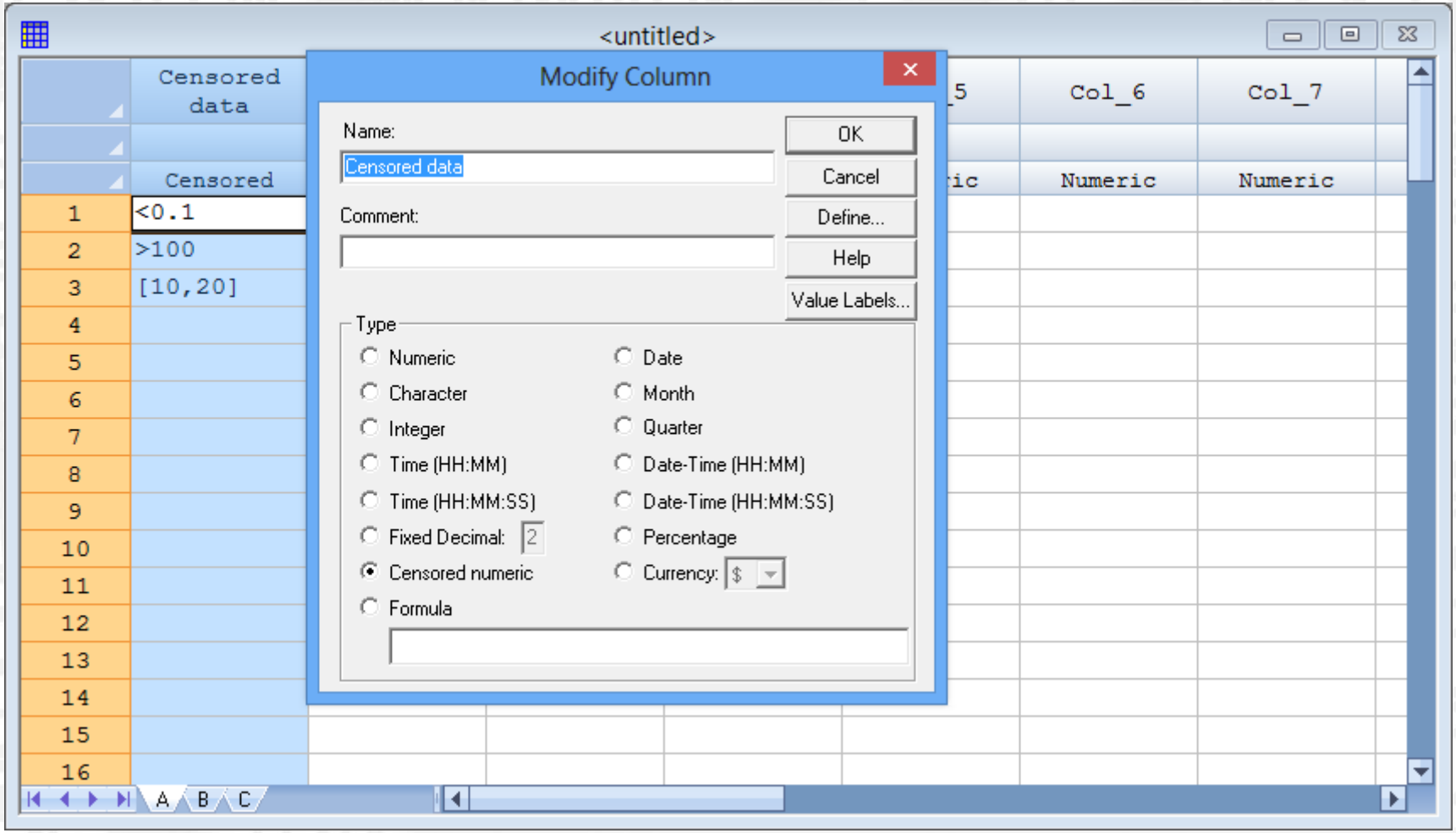
Today's Webinar

- A general overview of what's new or different in Statgraphics 18 compared to version 17
  - Interface
  - Big data
  - Data visualization
  - Equivalence and noninferiority tests
  - New process capability analysis features
  - Enhancements to the R interface
  - Installation changes
- Over 50 videos covering each new feature (including installation and activation) at www.statgraphics.com/instructional-videos

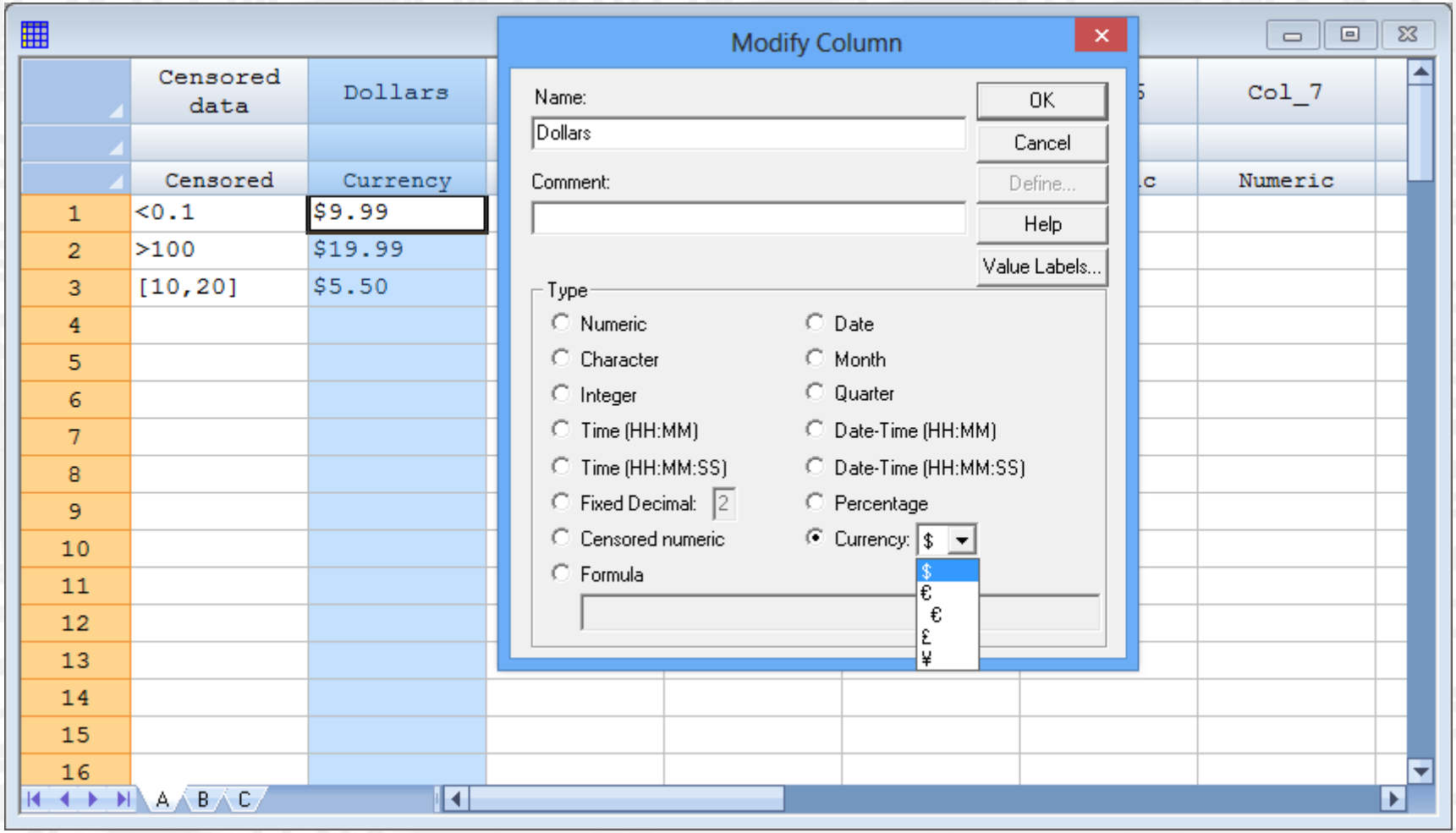
Interface


STATGRAPHICS 18 - Untitled StatFolio
File Edit Plot Describe Compare Relate Forecast SPC DOE SnapStats!! Statlets Tools R Interface View Window Help
DataBook
StatAdvisor
StatGallery
StatReporter
StatFolio Comments
StatLog
Text font size:
X ticks:
X-axis
Y-axis
Z-axis
Label:
Row:
Reset
<untitled>
Col_1 Col_2 Col_3 Col_4 Col_5 Col_6 Col_7
Numeric Numeric Numeric Numeric Numeric Numeric Numeric
A B C
Click on a menu item with the right mouse button to display documentation.

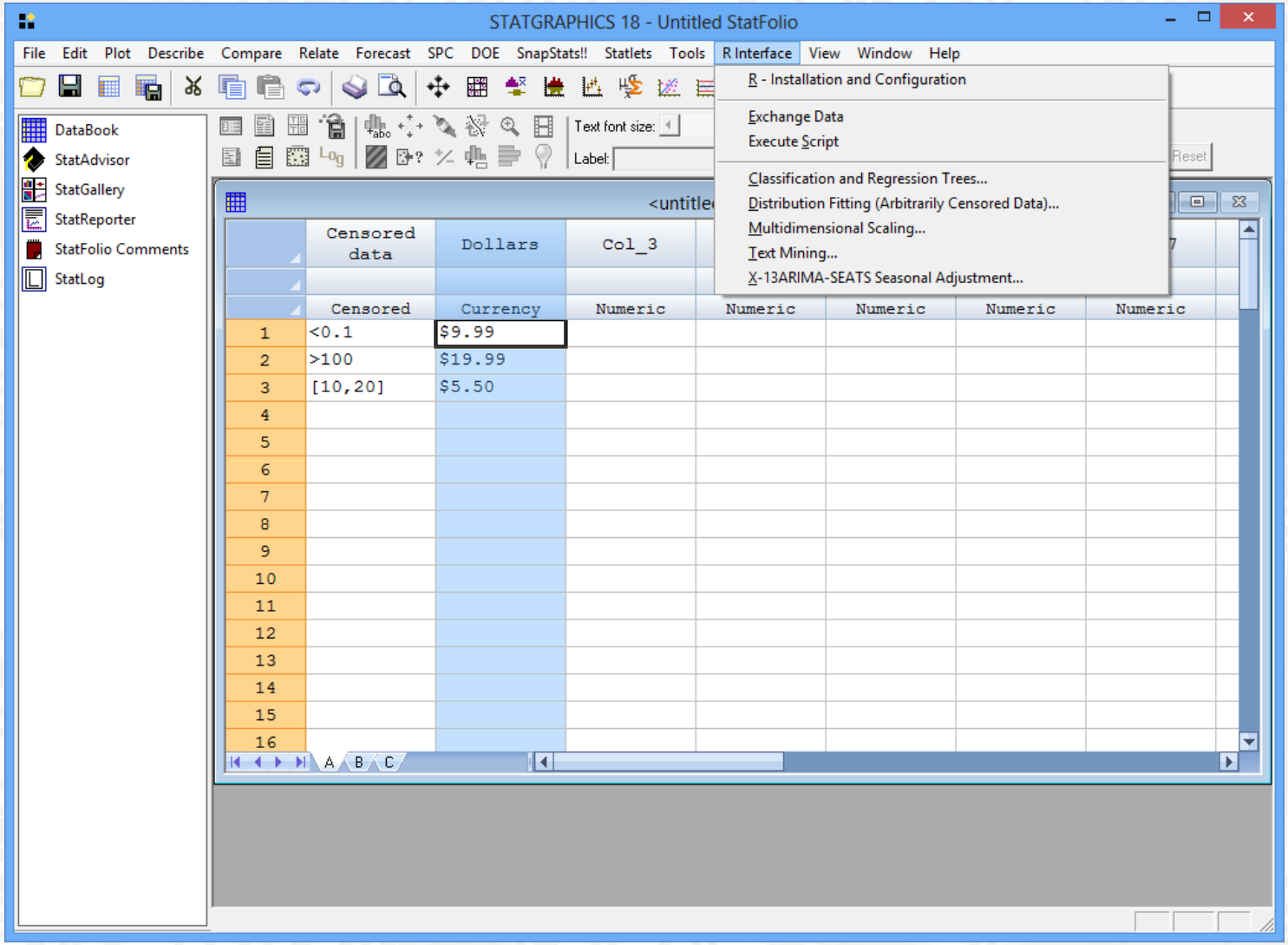
New Data Types



New Data Types

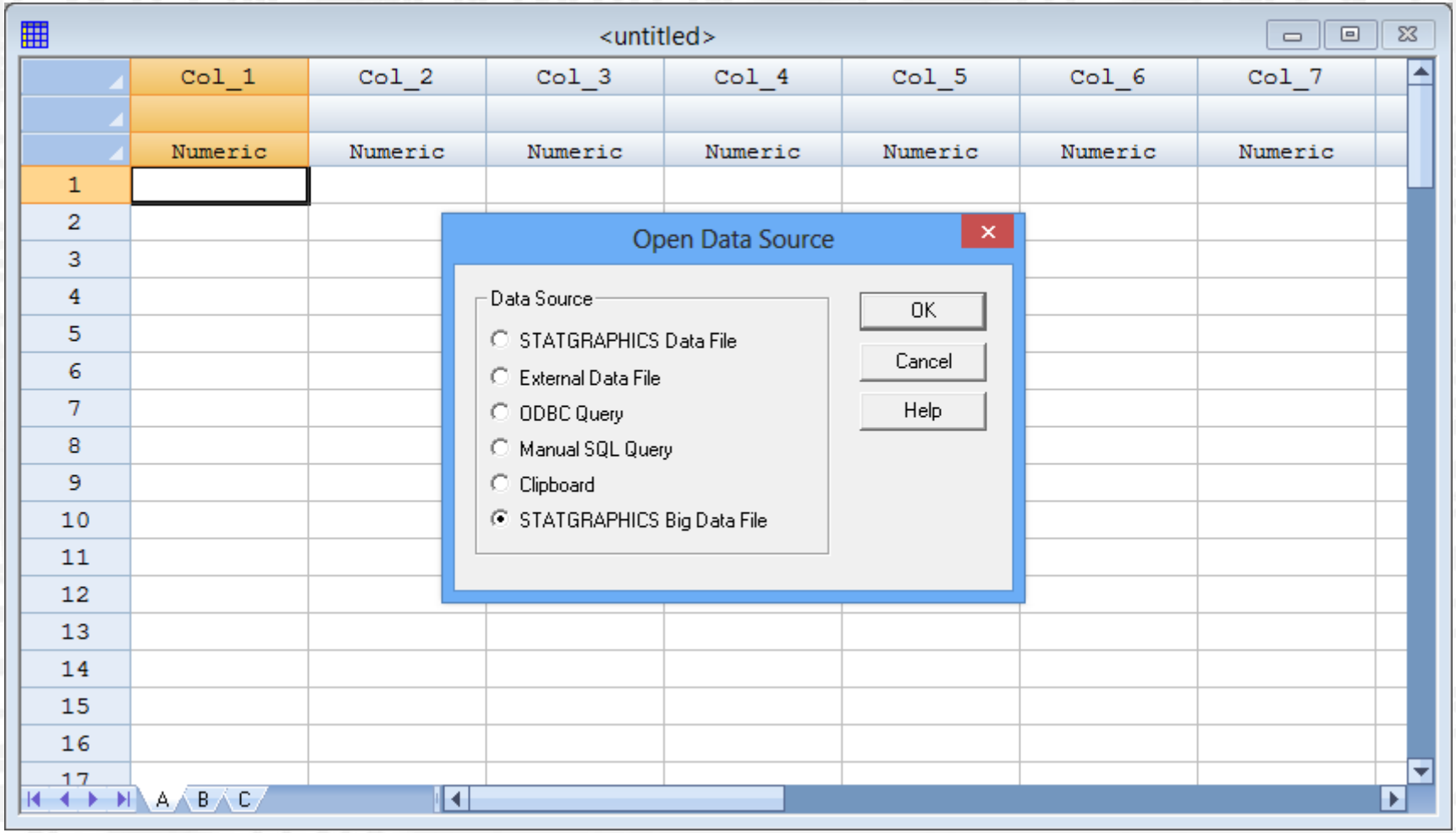


R Interface Menu



statgraphics 18®
centurion

Big Data Files (.sgb)



Big Data Files

- Hold numeric data in binary format rather than as text
- Files are organized by column rather than by row.
- Created by converting a text file to an SGB file.

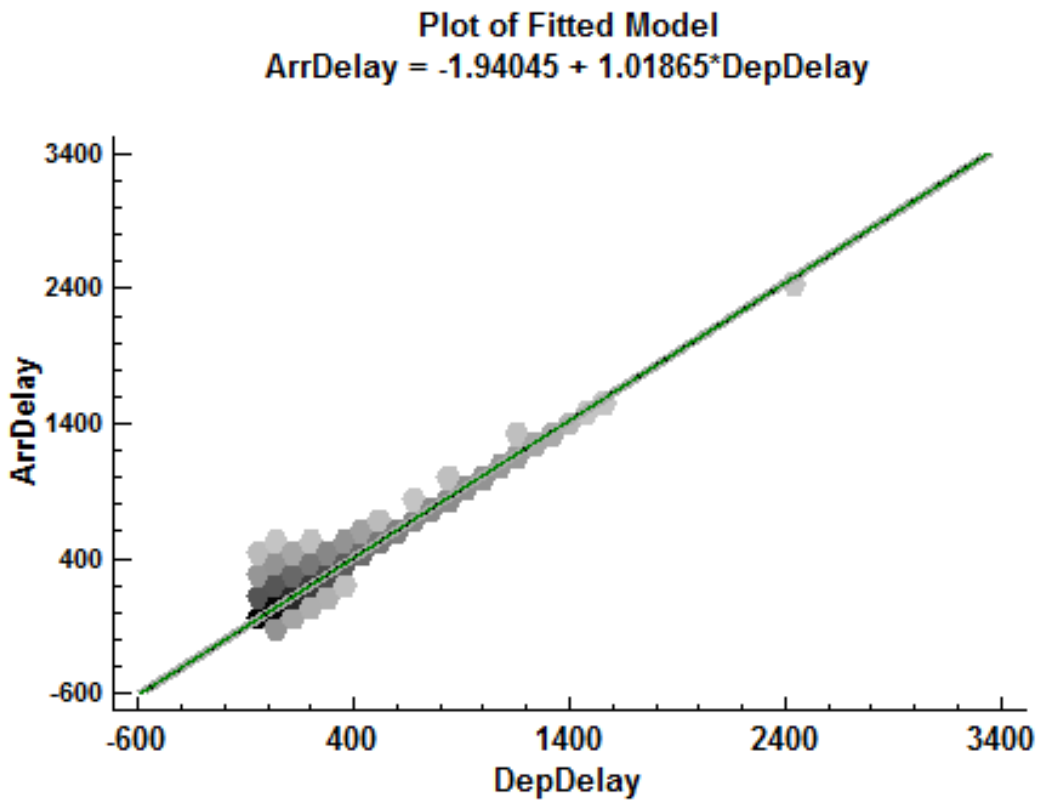
2008 Flight data

C:\DocData18\2008 Flight Data.sgb

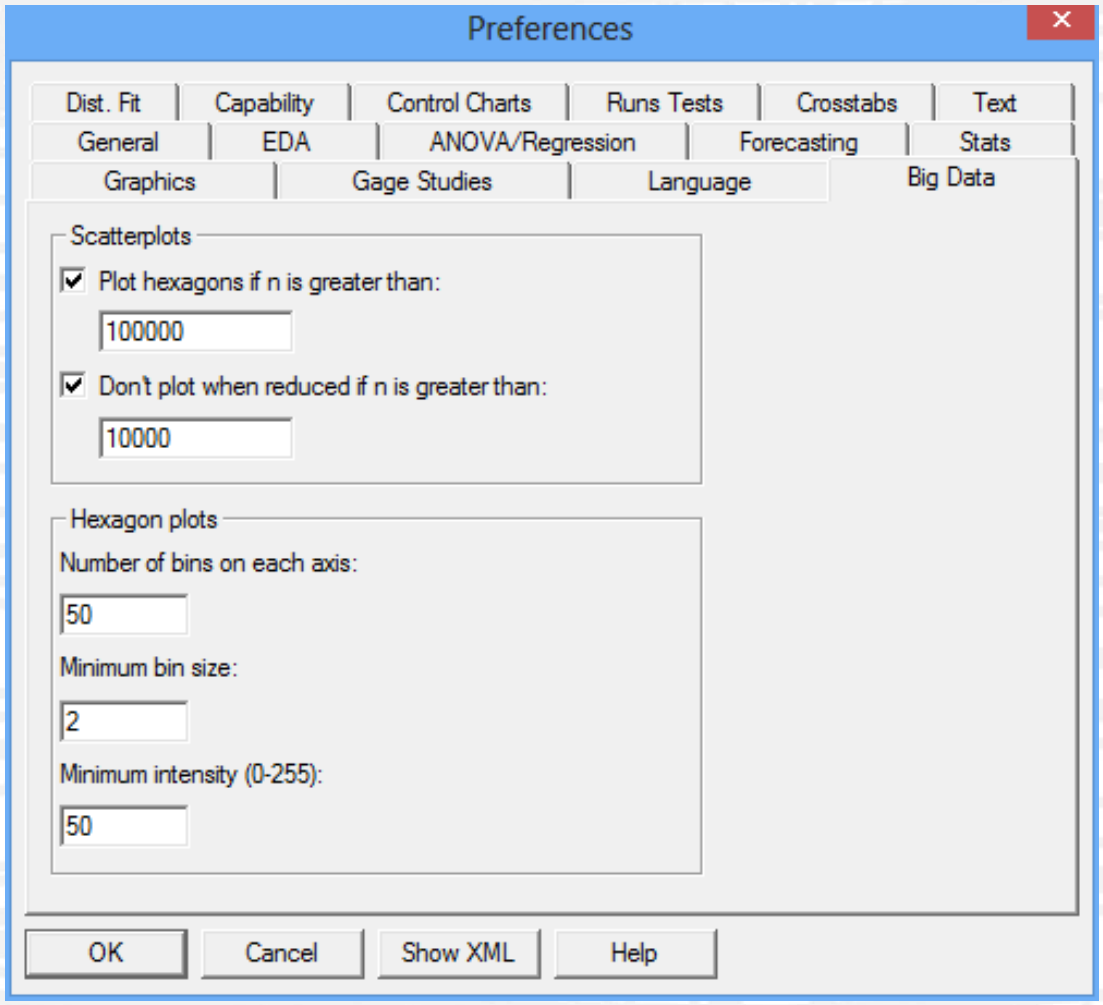
	Year	Month	DayofMonth	DayOfWeek	DepTime	CRSDepTime	ArrTime	C
	1987-2008	Month (1-12)	Day (1-31)	1 (Monday) - 7 (Sunday)	actual departure time (local, hhmm)	scheduled departure time (local, hhmm)	actual arrival time (local, hhmm)	s
	Integer	Integer	Integer	Integer	Integer	Integer	Integer	
1	2008	1	3	4	2003	1955	2211	22
2	2008	1	3	4	754	735	1002	10
3	2008	1	3	4	628	620	804	75
4	2008	1	3	4	926	930	1054	11
5	2008	1	3	4	1829	1755	1959	19
6	2008	1	3	4	1940	1915	2121	21
7	2008	1	3	4	1937	1830	2037	19
8	2008	1	3	4	1039	1040	1132	11
9	2008	1	3	4	617	615	652	65
10	2008	1	3	4	1620	1620	1639	16
11	2008	1	3	4	706	700	916	91

2008 Flight Data | B | C

Hexagon Plots



Big Data Preferences



statgraphics 18®
centurion

Judging Significance

Multiple Regression - ArrDelay

Multiple Regression - ArrDelay

Dependent variable: ArrDelay (arrival delay, in minutes)
Independent variables:
DepDelay (departure delay, in minutes)
Distance (in miles)
FlightNum (flight number)
Number of observations: 6855029

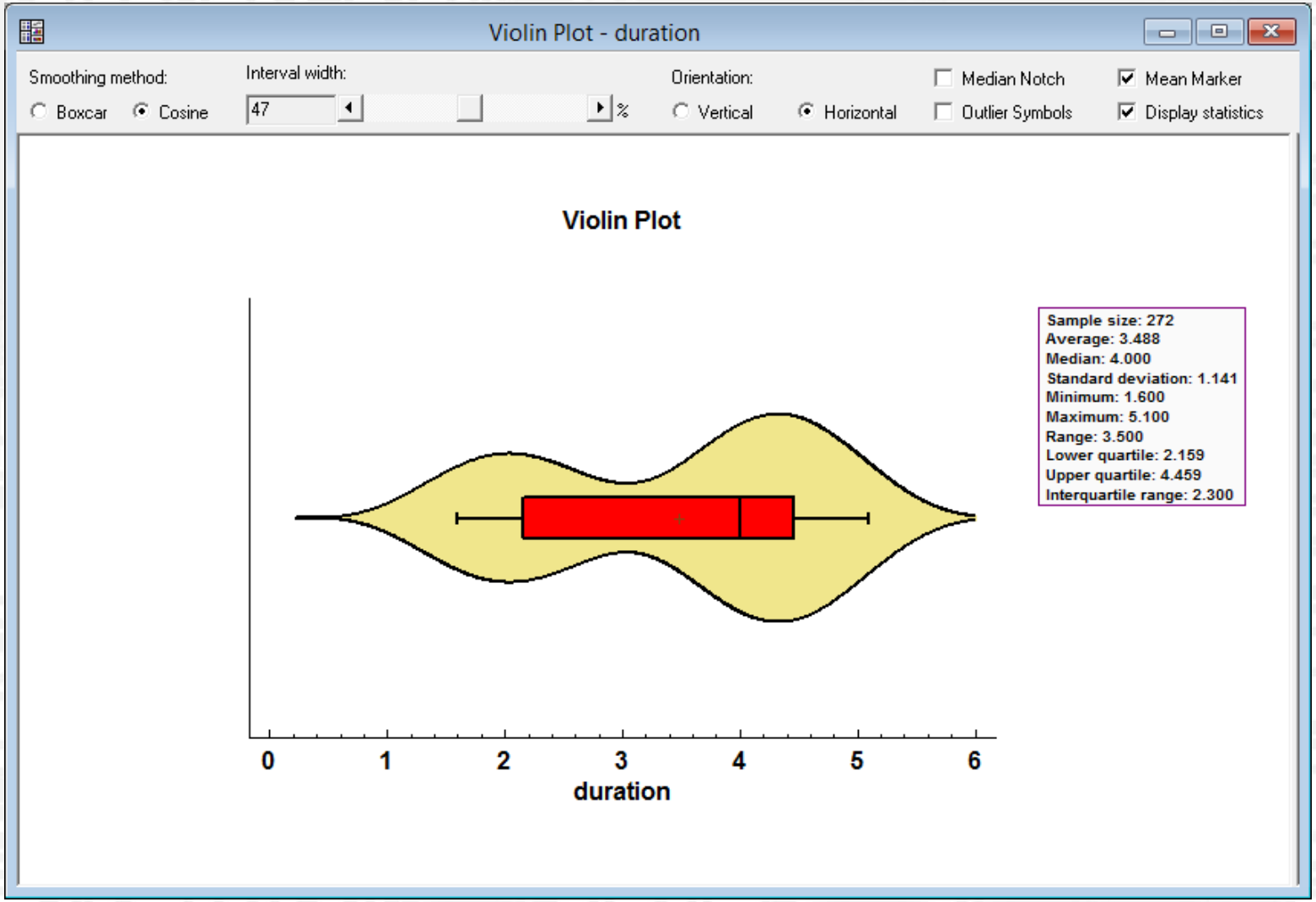
		Standard	*T*	
Parameter	*Estimate*	*Error*	*Statistic*	*P-Value*
CONSTANT	-1.14807	0.0126161	-91.0002	0.0000
DepDelay	1.01912	0.000151961	6706.51	0.0000
Distance	-0.00117913	0.0000101326	-116.37	0.0000
FlightNum	0.0000280999	0.00000291625	9.63564	0.0000

Analysis of Variance

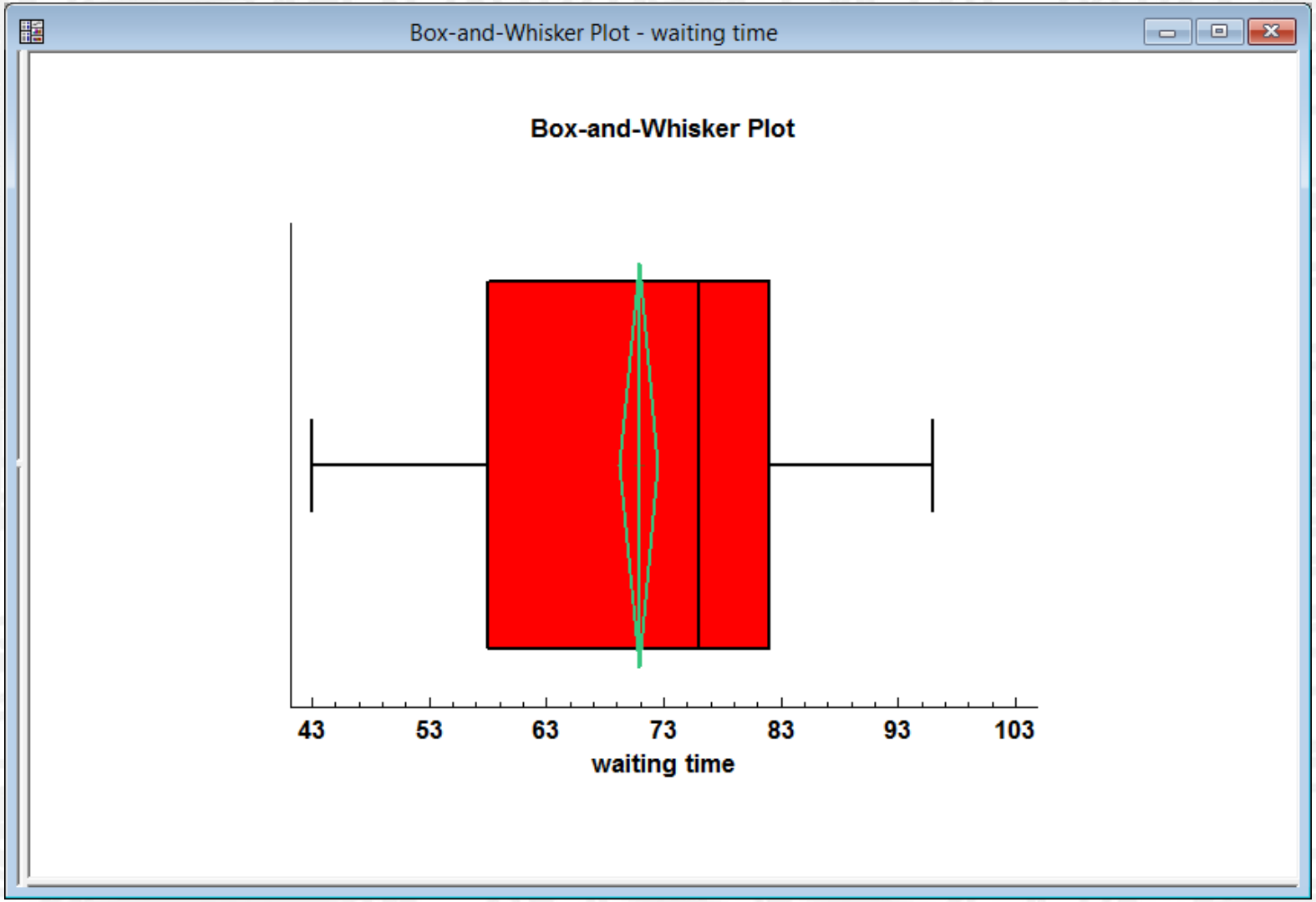
Source	*Sum of Squares*	*Df*	*Mean Square*	*F-Ratio*	*P-Value*
Model	8.81854E9	3	2.93951E9	15000171.24	0.0000
Residual	1.34335E9	6855025	195.965		
Total (Corr.)	1.01619E10	6855028			

R-squared = 86.7805 percent
R-squared (adjusted for d.f.) = 86.7805 percent
Standard Error of Est. = 13.9988
Mean absolute error = 9.34838
Durbin-Watson statistic = 1.63206 (P=0.0000)
Lag 1 residual autocorrelation = 0.18397

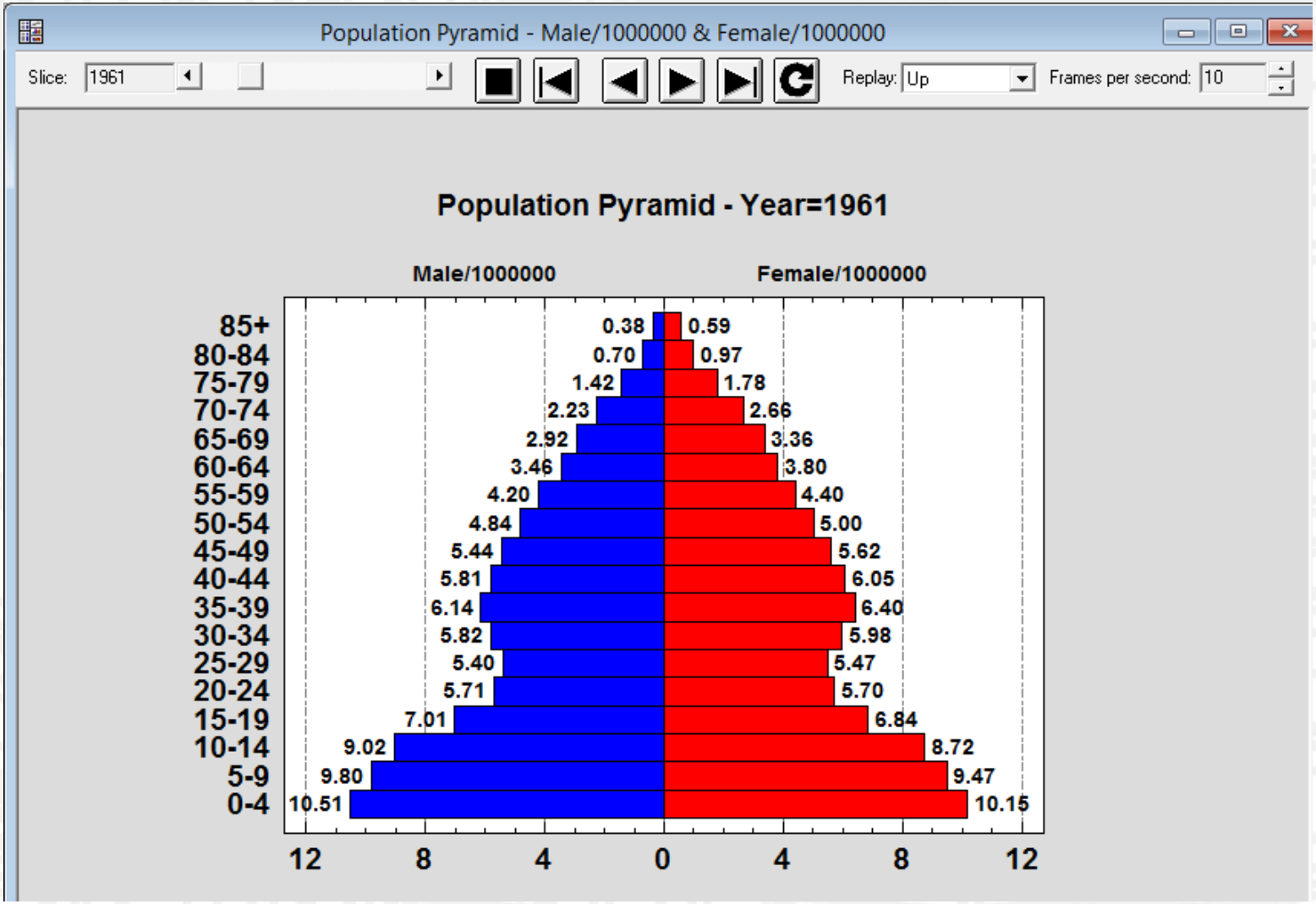
Data Visualization: Violin Plot



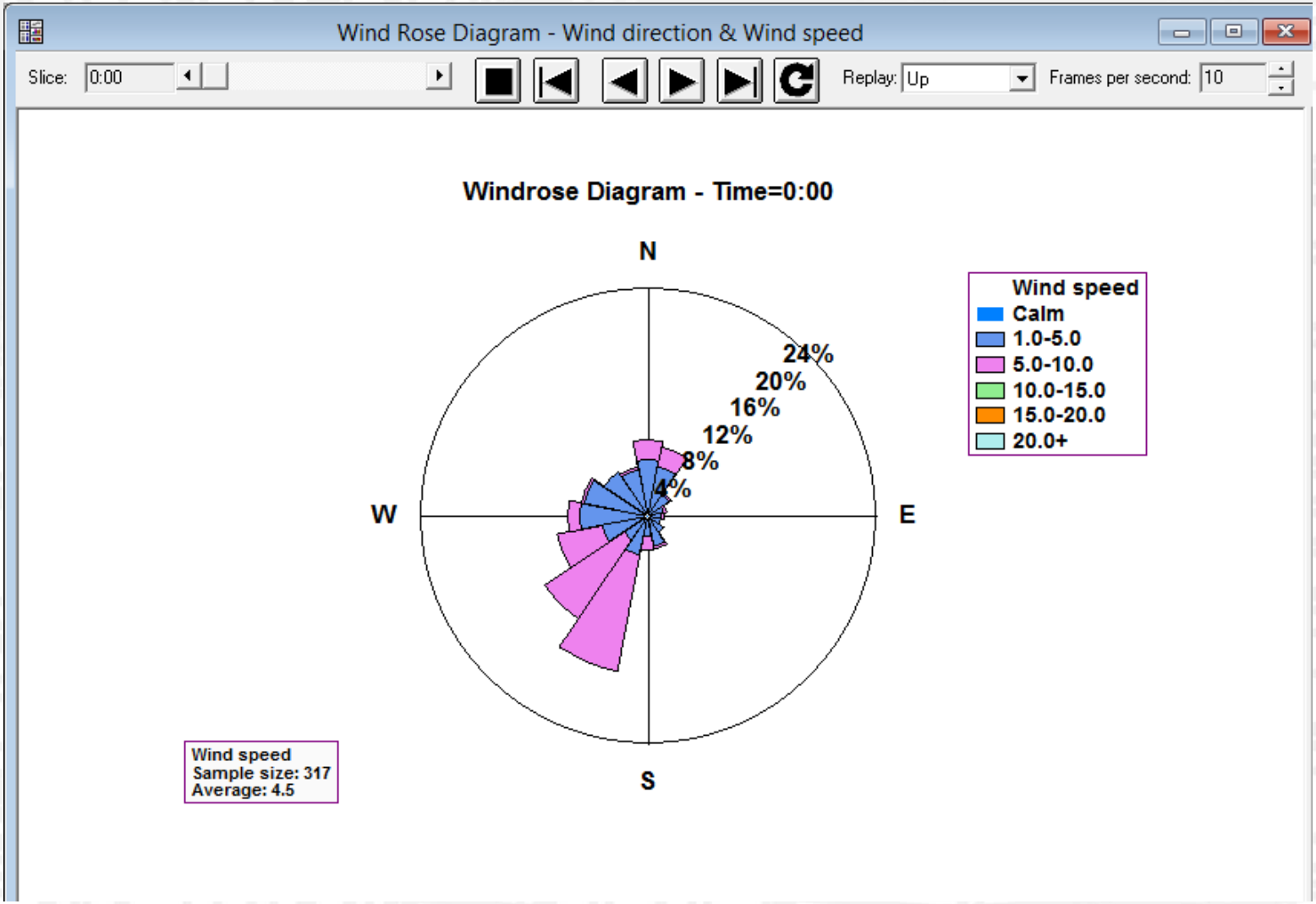
Diamond Plot



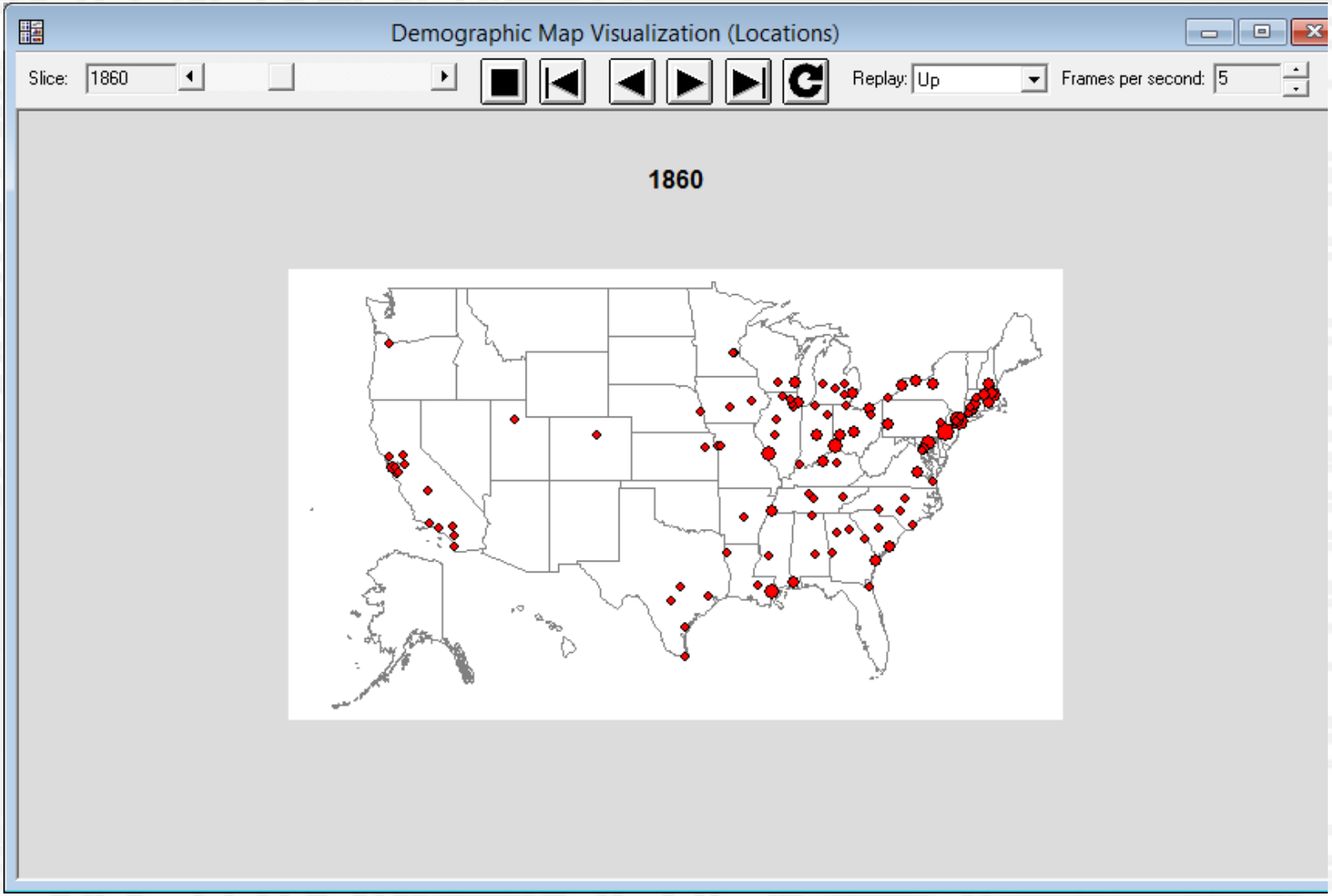
Population Pyramid



Wind Rose


Wind Rose Diagram - Wind direction & Wind speed
Slice: 0:00
Replay: Up
Frames per second: 10
Windrose Diagram - Time=0:00
N
W
E
S
24%
20%
16%
12%
8%
4%
Wind speed
Calm
1.0-5.0
5.0-10.0
10.0-15.0
15.0-20.0
20.0+
Wind speed
Sample size: 317
Average: 4.5

Demographic Maps



Equivalence and Noninferiority Tests

- Designed to demonstrate that:
  - 2 population means are equivalent; or
  - 1 mean is not inferior to the other
- Differs from standard two-sample hypothesis tests which are designed to demonstrate that 2 population means are different
- May also compare 1 mean against a target value

Comparison of 2 Means

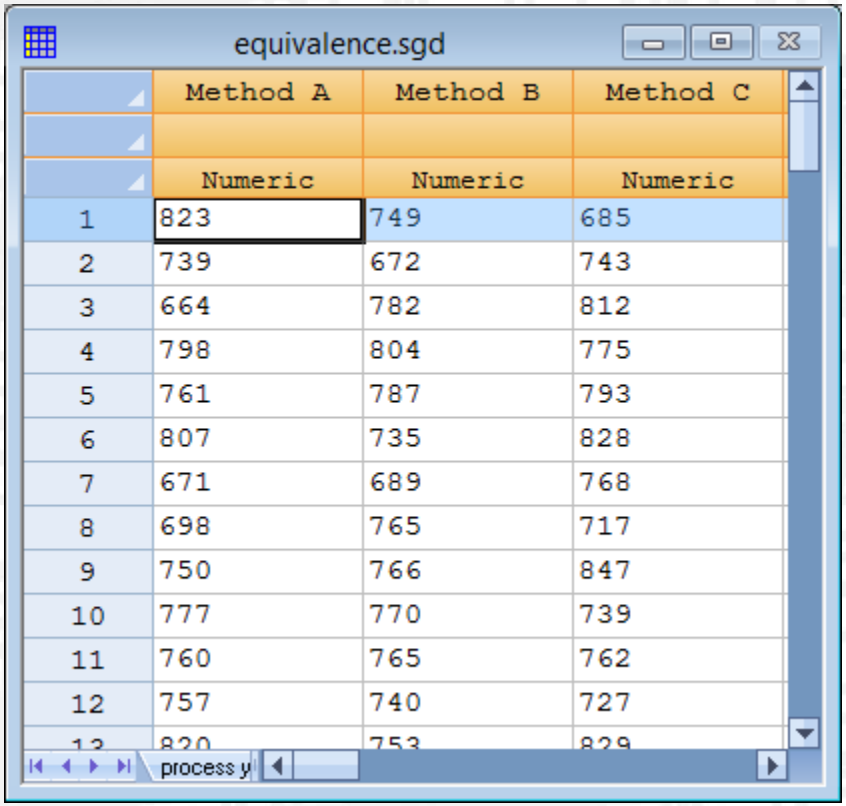
- Equivalence test

$$H_0: \mu_1 - \mu_2 < \Delta_L \text{ or } \mu_1 - \mu_2 > \Delta_U$$
$$H_A: \Delta_L \leq \mu_1 - \mu_2 \leq \Delta_U$$

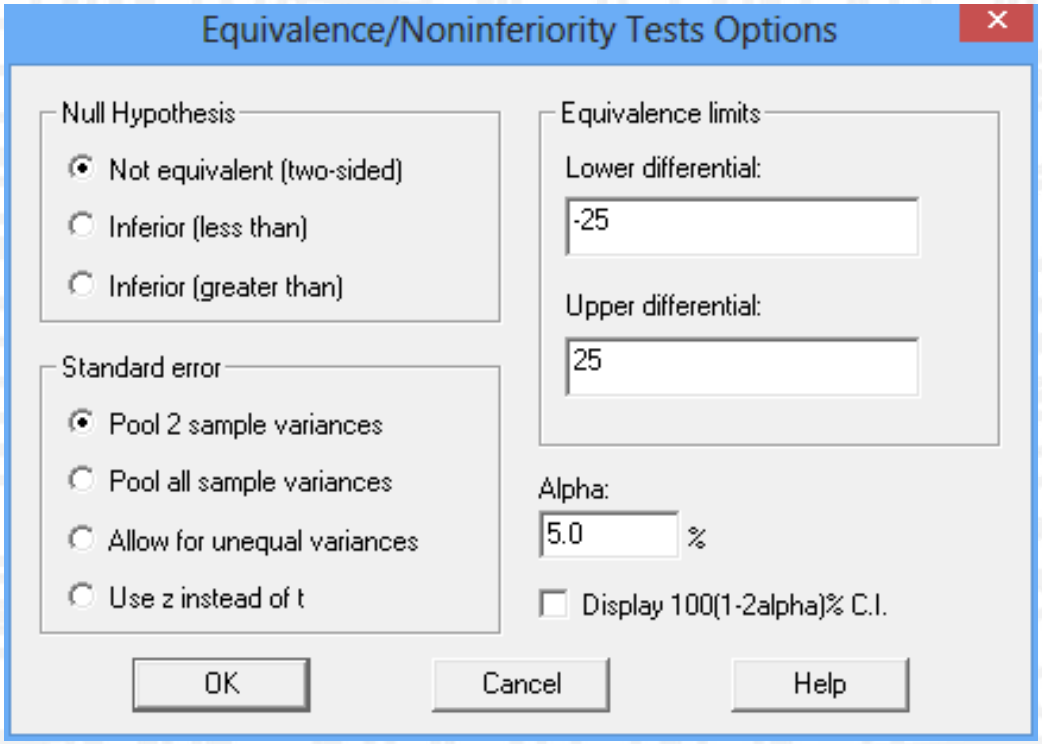
- Inferiority test

$$H_0: \mu_1 - \mu_2 < \Delta_L$$
$$H_A: \Delta_L \leq \mu_1 - \mu_2$$

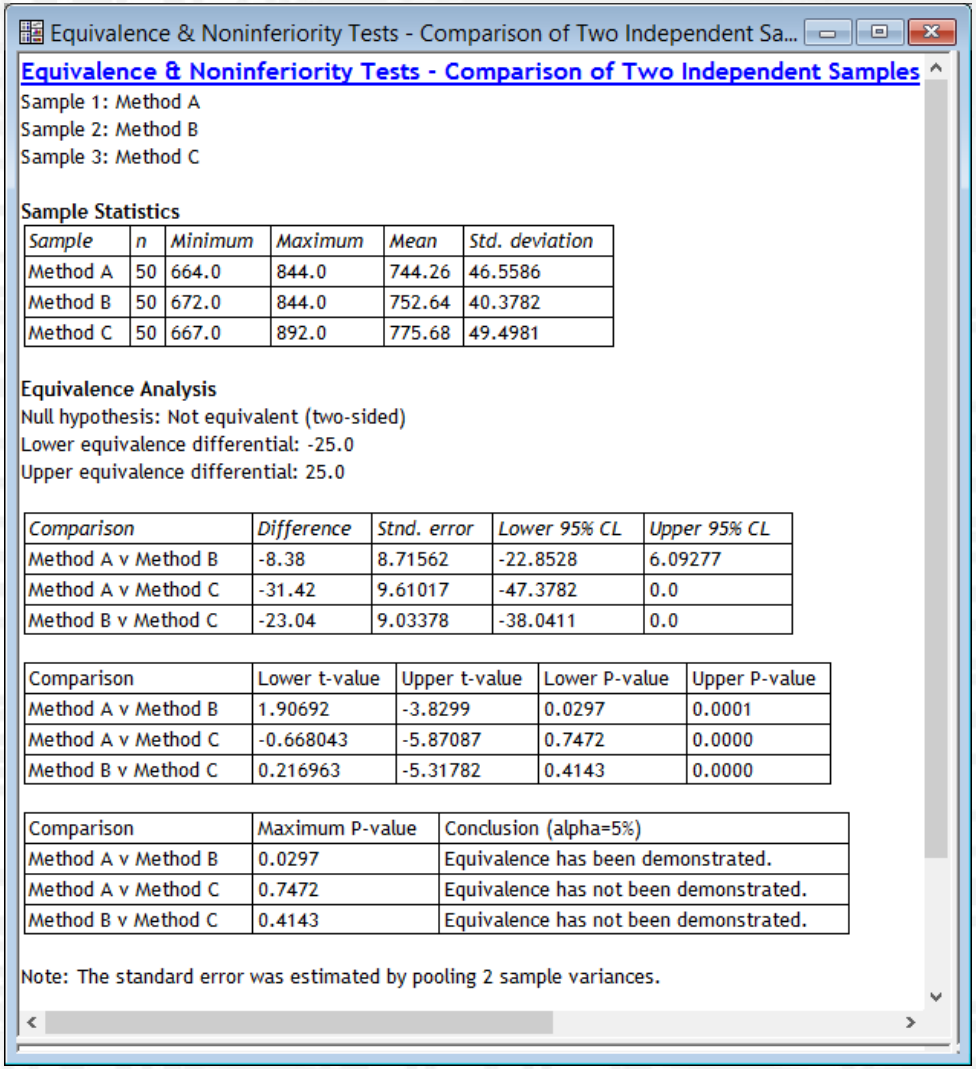
Example: 3 Methods



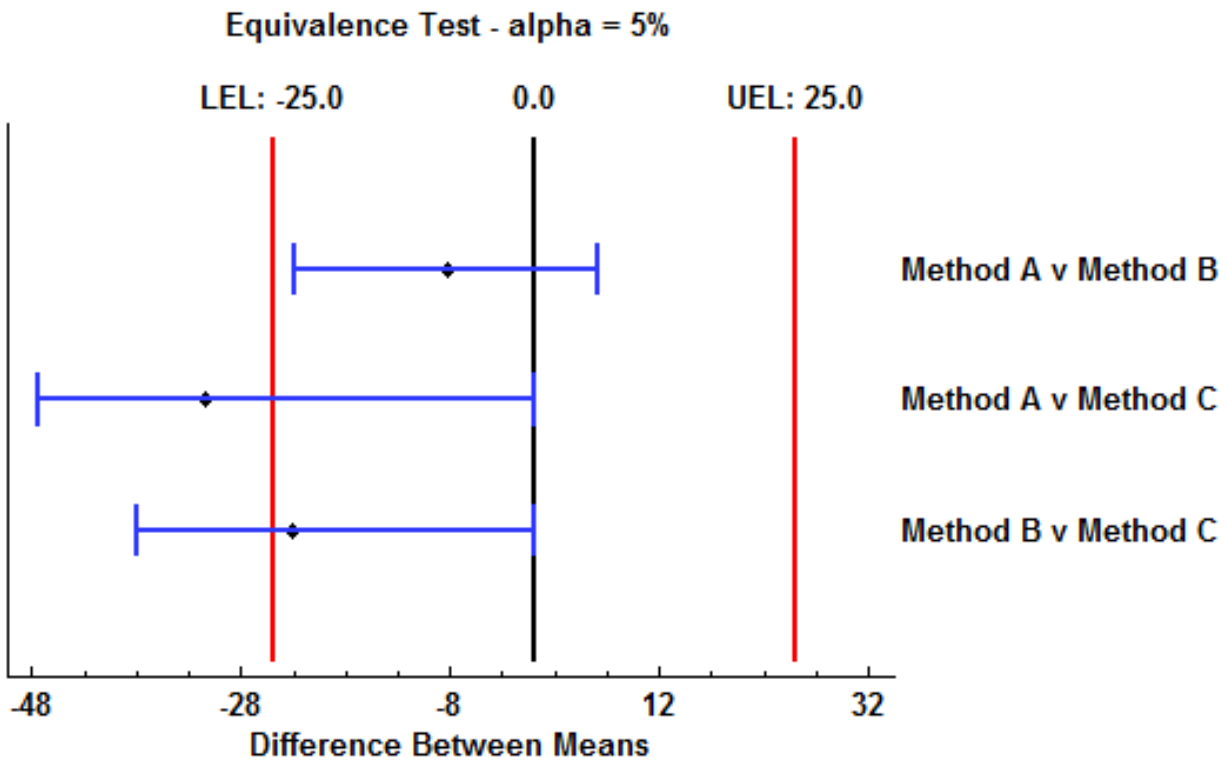
Analysis Options



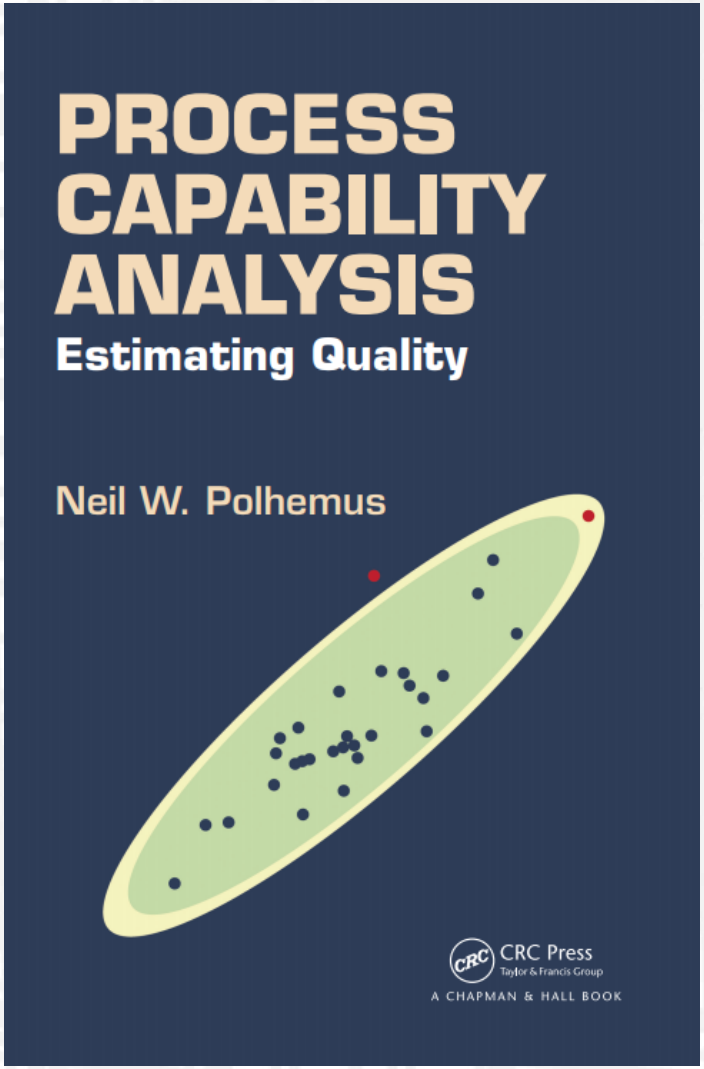
Two One-Sided Tests (TOST)



Confidence Intervals



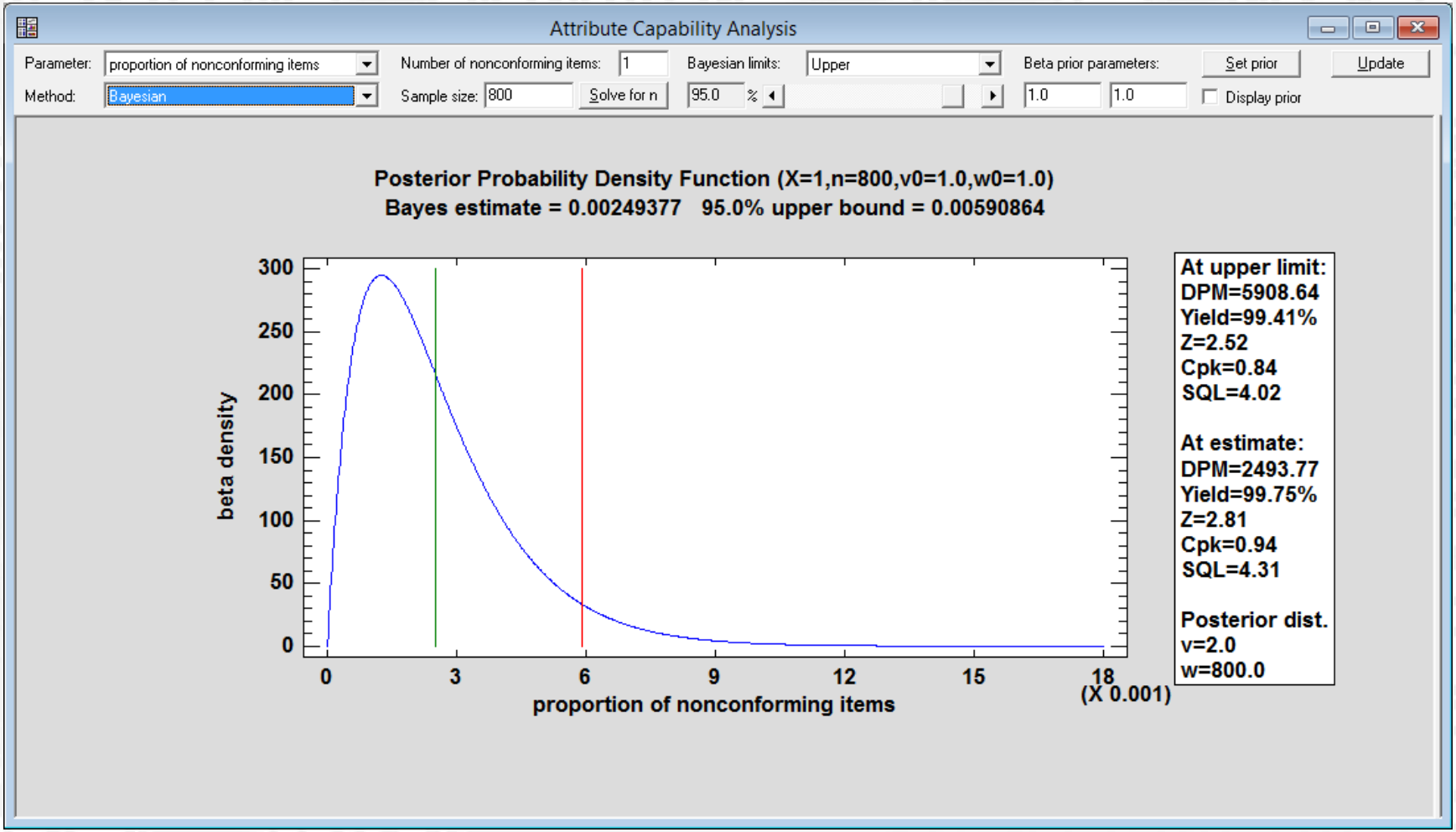
Process Capability Analysis



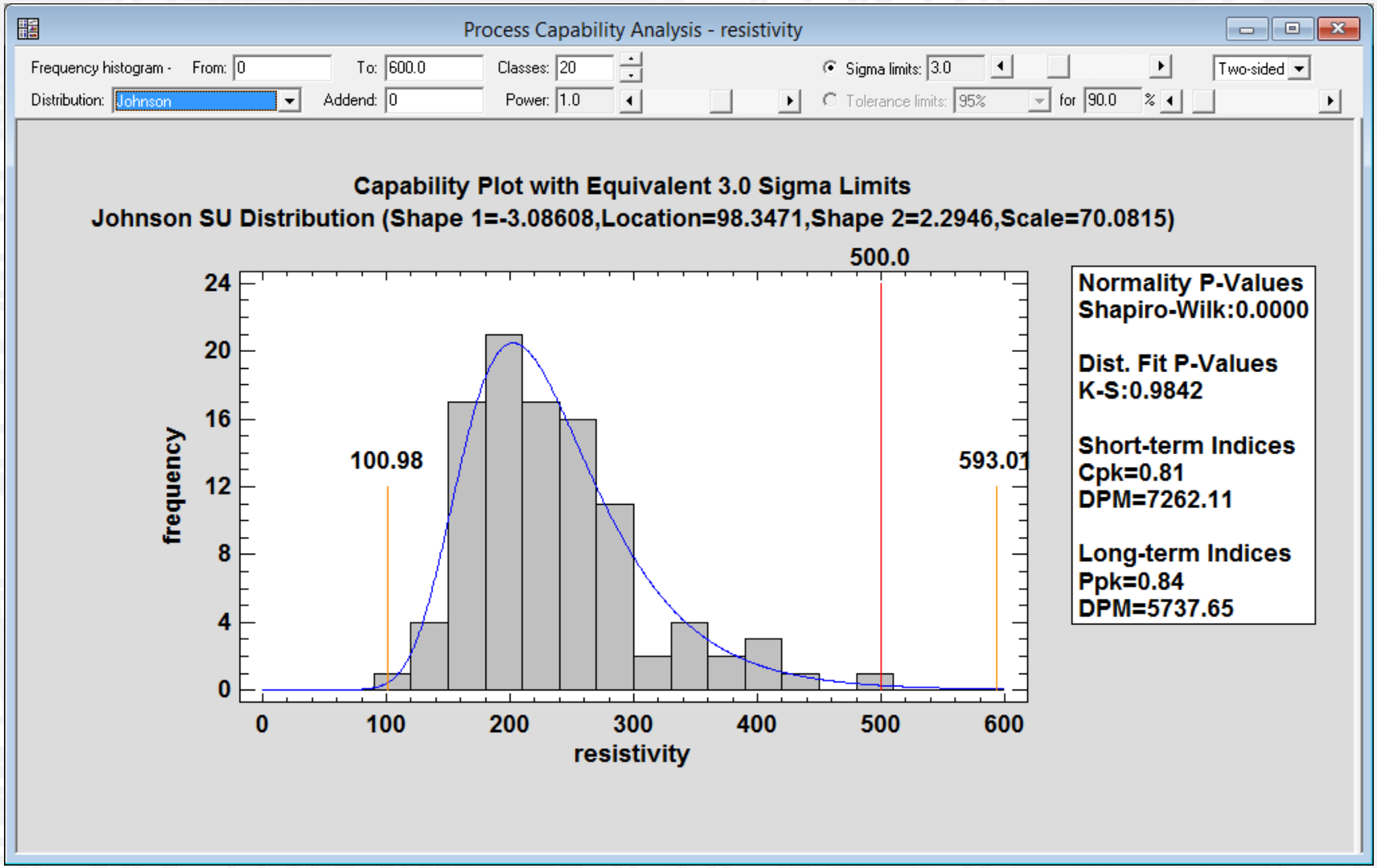
New SPC Procedures

- Attribute capability analysis Statlet (with Bayesian methods)
- Variables capability analysis Statlet (with Johnson curves)
- Multivariate capability analysis
- Multivariate statistical tolerance regions
- Capability control charts
- Sample size determination

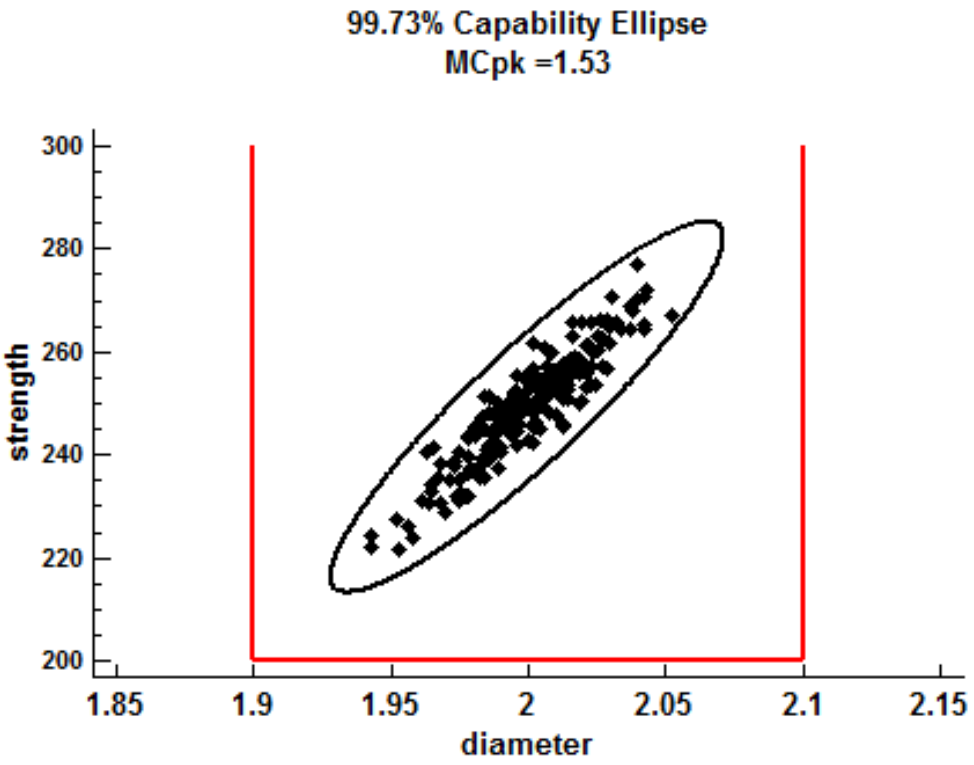
Attribute Capability Analysis Statlet



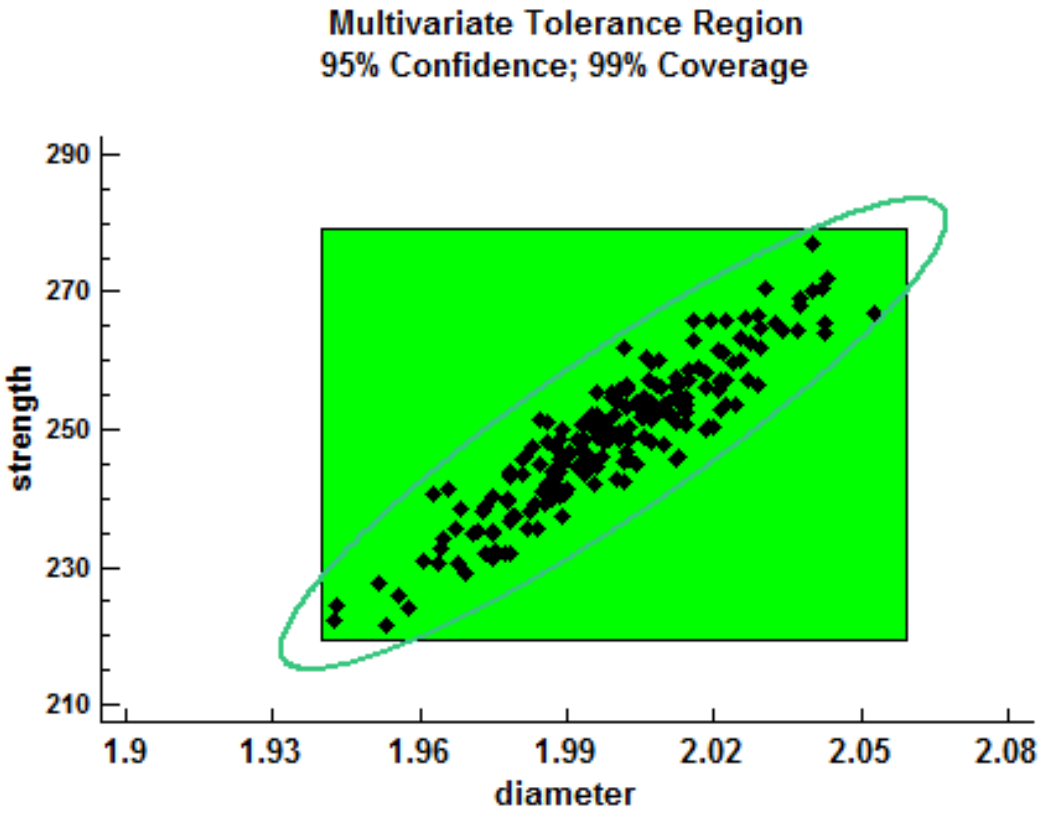
Variables Capability Analysis Statlet



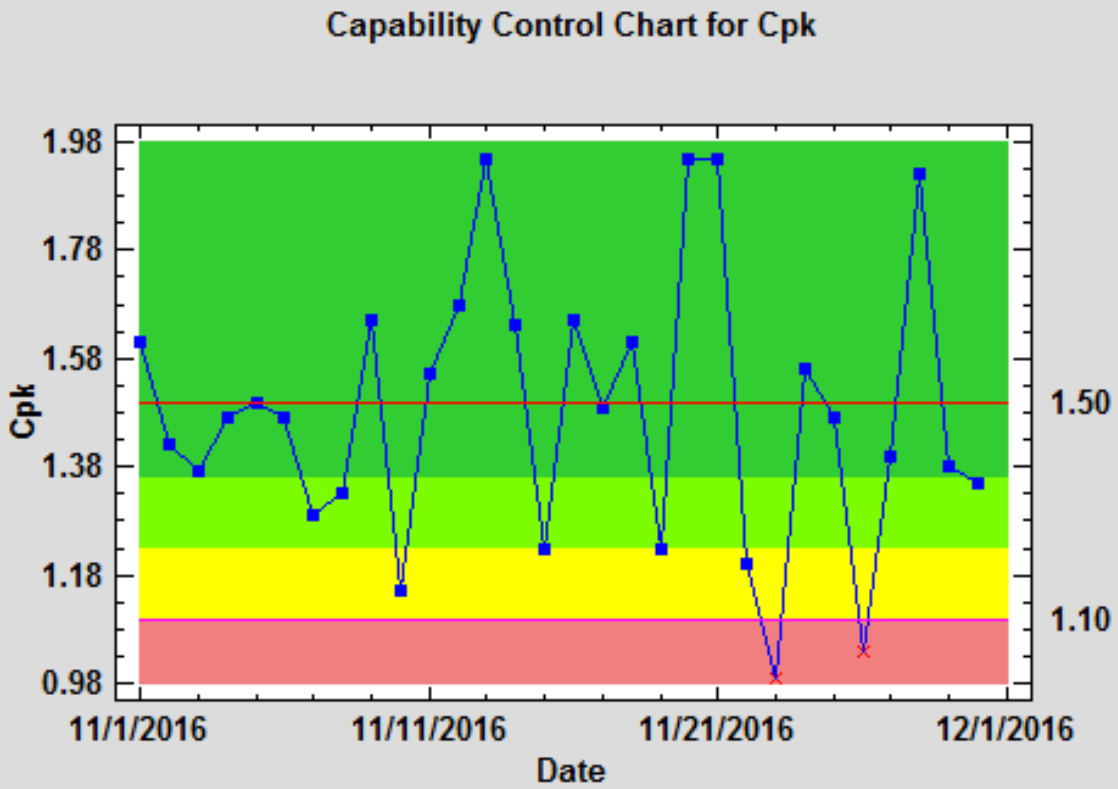
Multivariate Capability Analysis



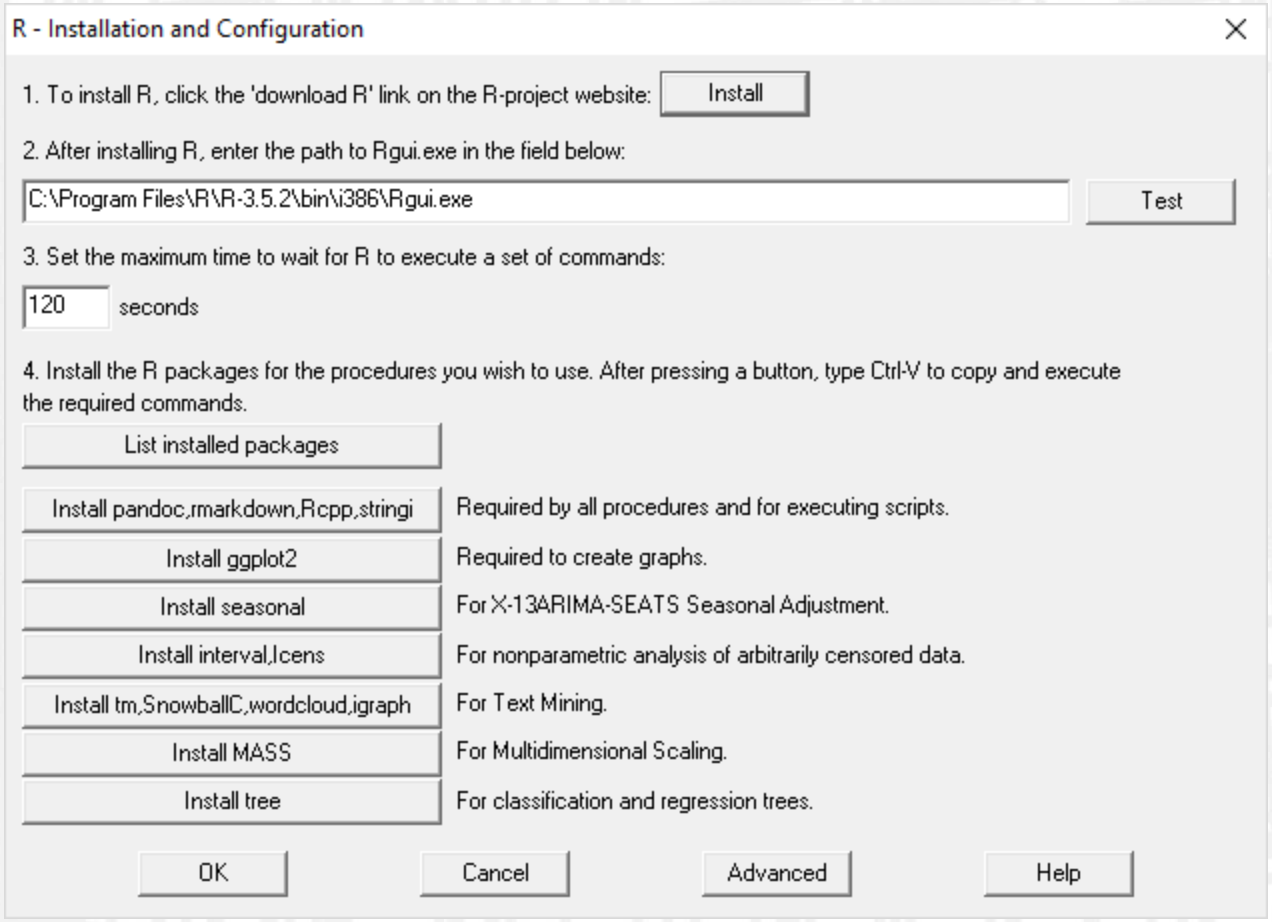
Multivariate Tolerance Regions



Capability Control Charts



R Interface Enhancements



Classification Trees - Iris Data

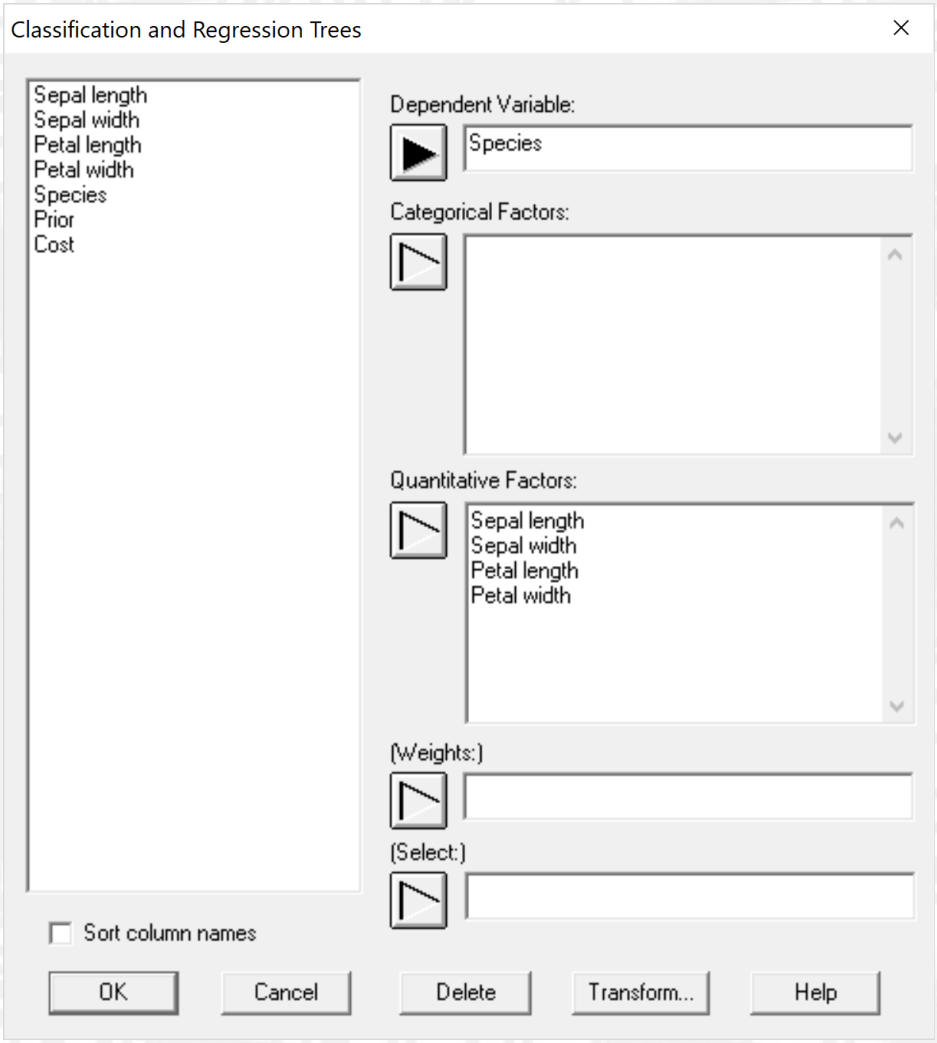
C:\DocData17\iris.sgd

	Sepal length	Sepal width	Petal length	Petal width	Species
	centimeters	centimeters	centimeters	centimeters	type of iris
	Numeric	Numeric	Numeric	Numeric	Character
1	5.1	3.5	1.4	0.2	setosa
2	4.9	3	1.4	0.2	setosa
3	4.7	3.2	1.3	0.2	setosa
4	4.6	3.1	1.5	0.2	setosa
5	5	3.6	1.4	0.2	setosa
6	5.4	3.9	1.7	0.4	setosa
7	4.6	3.4	1.4	0.3	setosa
8	5	3.4	1.5	0.2	setosa
9	4.4	2.9	1.4	0.2	setosa
10	4.9	3.1	1.5	0.1	setosa
11	5.4	3.7	1.5	0.2	setosa
12	4.8	3.4	1.6	0.2	setosa
13	4.8	3	1.4	0.1	setosa
14	4.3	3	1.1	0.1	setosa
15	5.8	4	1.2	0.2	setosa
16	5.7	4.4	1.5	0.4	setosa
17	5.4	3.9	1.3	0.4	setosa
18	5.1	3.5	1.4	0.3	setosa
19	5.7	3.8	1.7	0.3	setosa
20	5.1	3.8	1.5	0.3	setosa
21	5.4	3.4	1.7	0.2	setosa
22	5.1	3.7	1.5	0.4	setosa
23	4.6	3.6	1	0.2	setosa
24	5.1	3.3	1.7	0.5	setosa
25	4.8	3.4	1.9	0.2	setosa

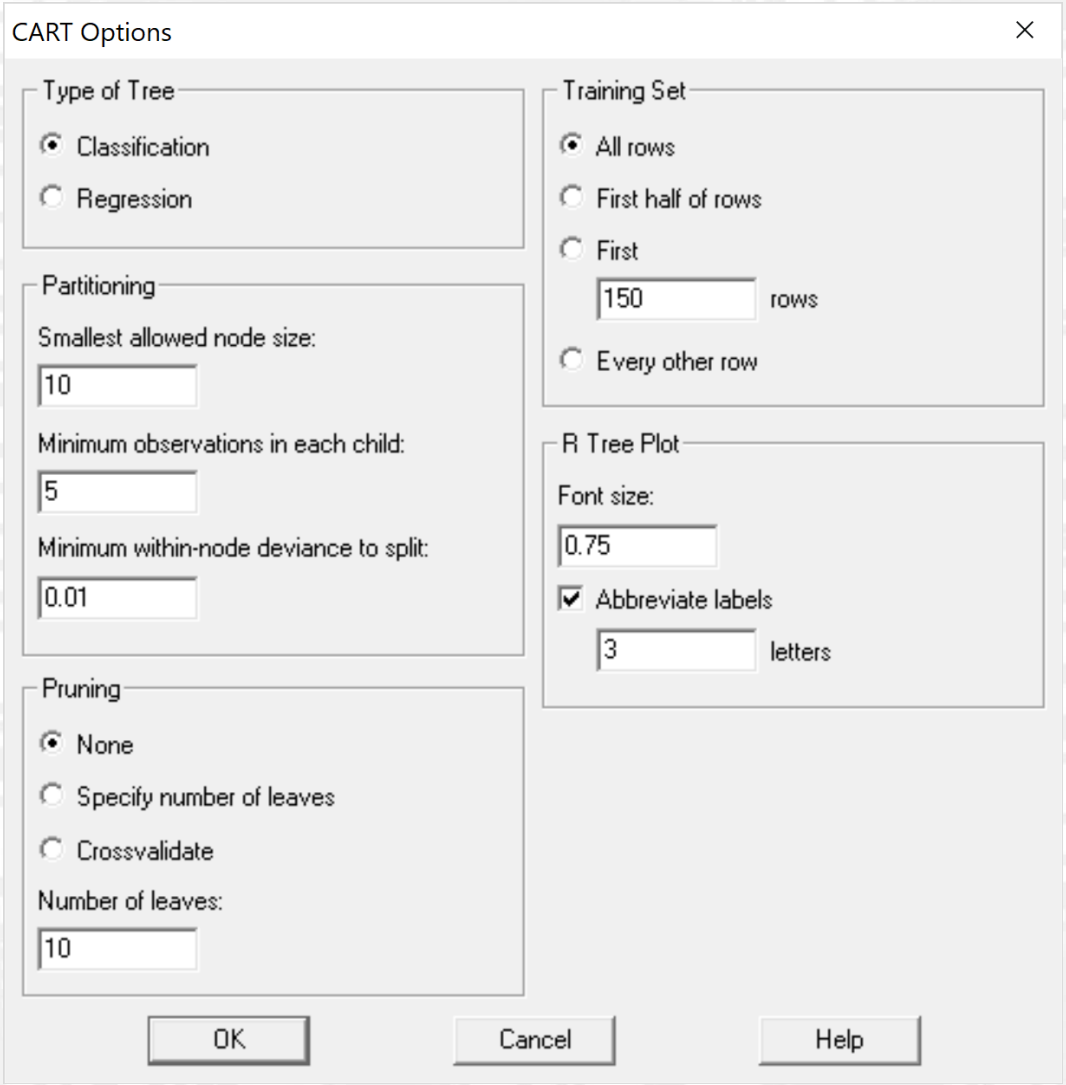
iris B C

statgraphics 18 centurion

Data Input


Classification and Regression Trees
Sepal length
Sepal width
Petal length
Petal width
Species
Prior
Cost
Dependent Variable:
Species
Categorical Factors:
Quantitative Factors:
Sepal length
Sepal width
Petal length
Petal width
(Weights:)
(Select:)
Sort column names
OK
Cancel
Delete
Transform...
Help

Analysis Options



Analysis Summary

Classification and Regression Trees - Species

Classification and Regression Trees

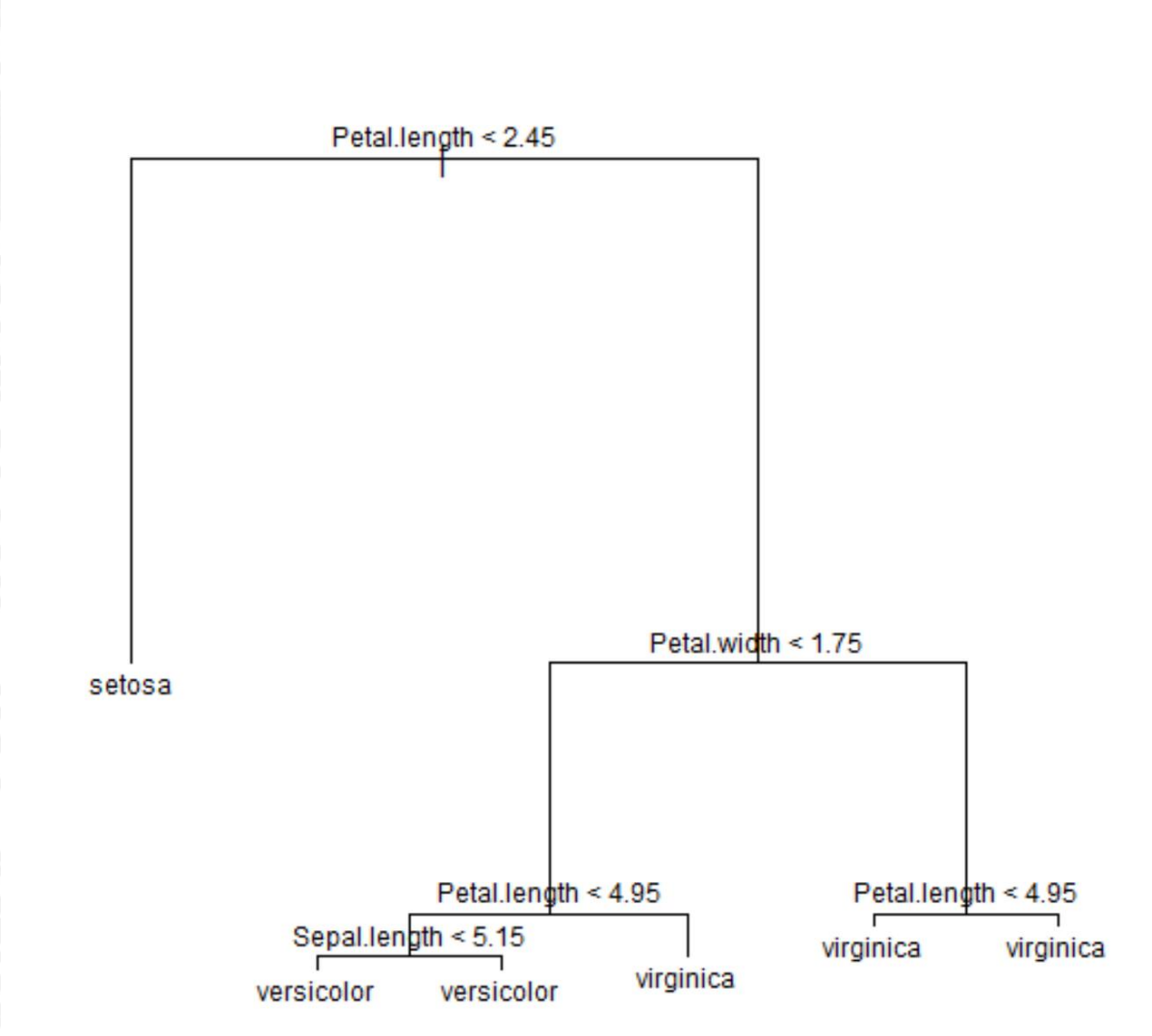
```
d<-read.csv("C:\\\\Users\\\\NEIL~1.STA\\\\AppData\\\\Local\\\\Temp\\\\data.csv",dec=".",sep=",",stringsAsFactors=TRUE)
setwd("C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\")
library("tree")
treefit=tree(Species~Sepal.length+Sepal.width+Petal.length+Petal.width,control=tree.control(nobs=150,mincut=5,minsize=10,minc
)
summary(treefit)

##
## Classification tree:
## tree(formula = Species ~ Sepal.length + Sepal.width + Petal.length +
##     Petal.width, data = d, control = tree.control(nobs = 150,
##     mincut = 5, minsize = 10, mindev = 0.01))
## Variables actually used in tree construction:
## [1] "Petal.length" "Petal.width"  "Sepal.length"
## Number of terminal nodes:  6
## Residual mean deviance:  0.1253 = 18.05 / 144
## Misclassification error rate: 0.02667 = 4 / 150

plot(treefit)
text(treefit,pretty=3,cex=0.75)



p<-prune.tree(treefit)
write.table(treefit$frame,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\frame.csv",sep=",")
write.table(treefit$where,file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\where.csv",sep=",",row.names=FALSE)
write.table(cbind(p$size,p$dev,p$k),file="C:\\Users\\NEIL~1.STA\\AppData\\Local\\Temp\\prune.csv",sep=",",row.names=FALSE)
```

The StatAdvisor

The output above shows the results of instructing the "tree" package in R to construct a classification tree to predict the values

of Species. Use the Analysis Options dialog box to control how large a tree is created.

Tree Diagram



Tree Diagram

Node Probabilities

Node	*Label*	*Size*	*yprob.setosa*	*yprob.versicolor*	*yprob.virginica*
1	Petal.length	150	0.333333	0.333333	0.333333
2	<leaf>	50	1.0	0.0	0.0
3	Petal.width	100	0.0	0.5	0.5
4	Petal.length	54	0.0	0.907407	0.0925926
5	Sepal.length	48	0.0	0.979167	0.0208333
6	<leaf>	5	0.0	0.8	0.2
7	<leaf>	43	0.0	1.0	0.0
8	<leaf>	6	0.0	0.333333	0.666667
9	Petal.length	46	0.0	0.0217391	0.978261
10	<leaf>	6	0.0	0.166667	0.833333
11	<leaf>	40	0.0	0.0	1.0

The StatAdvisor

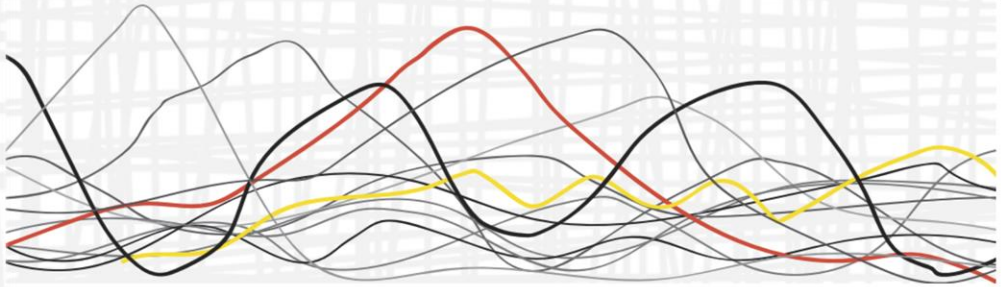
This table shows the probability distribution for Species at each of the nodes in the tree. The probabilities are based on the number of members of the training set that reach the node.

Installation Changes

- Online registration program for attaching user email to serial number
- Activation no longer requires administrative rights
- New deactivation option for moving license to different machine

register.statgraphics.com

statgraphics 18®



Version 18 Registration Program

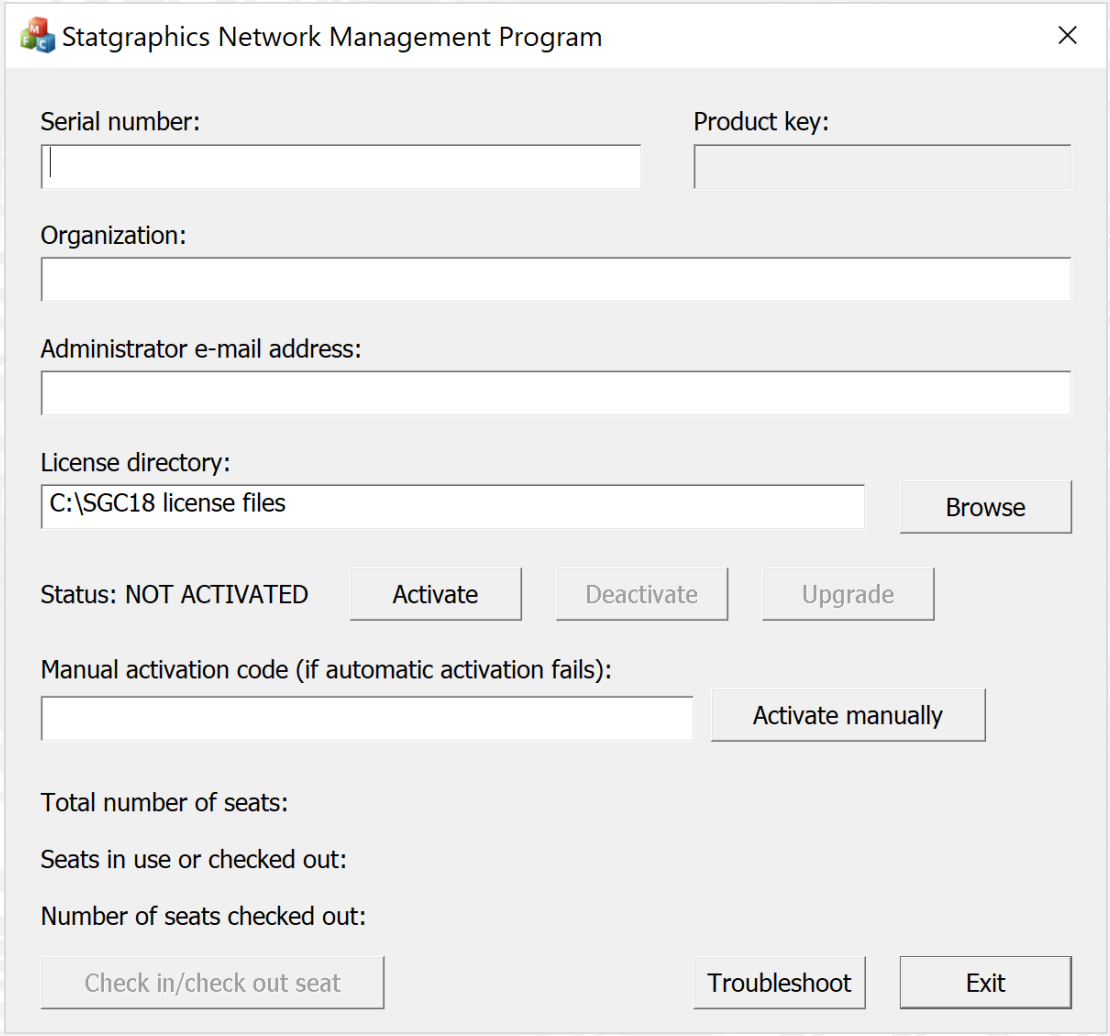
TRIAL PERIOD

REGISTERED SERIAL NUMBERS

ACTIVATIONS

ASSOCIATED SITE LICENSES

Network Management Program



More Information

- Recorded webinar will be posted at:

  www.statgraphics.com/webinars

- Version 18 videos are available at:

  www.statgraphics.com/instructional-videos