



Introducing Statgraphics 18

Presented by
Dr. Neil W. Polhemus



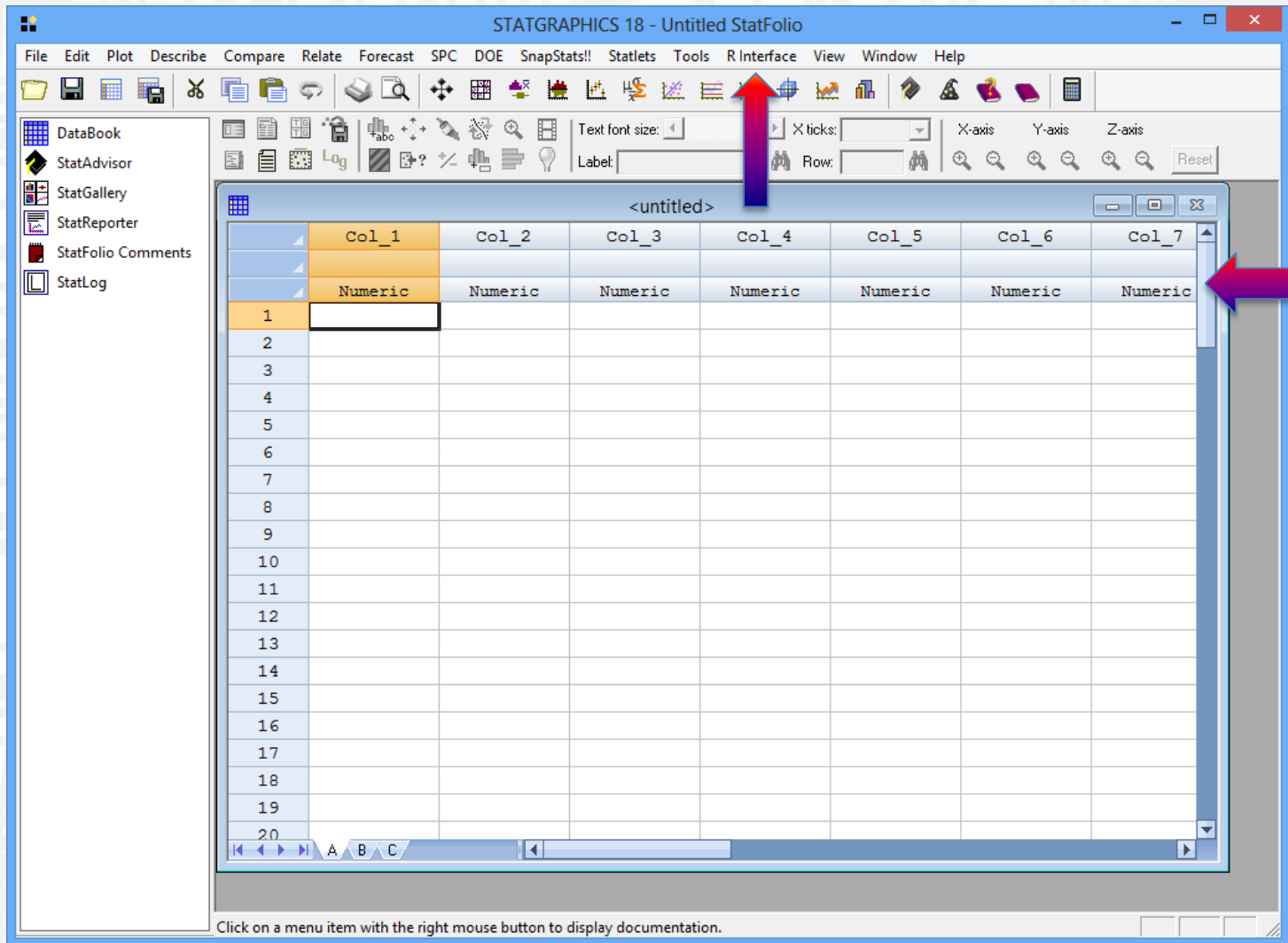
Statgraphics 18

- 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

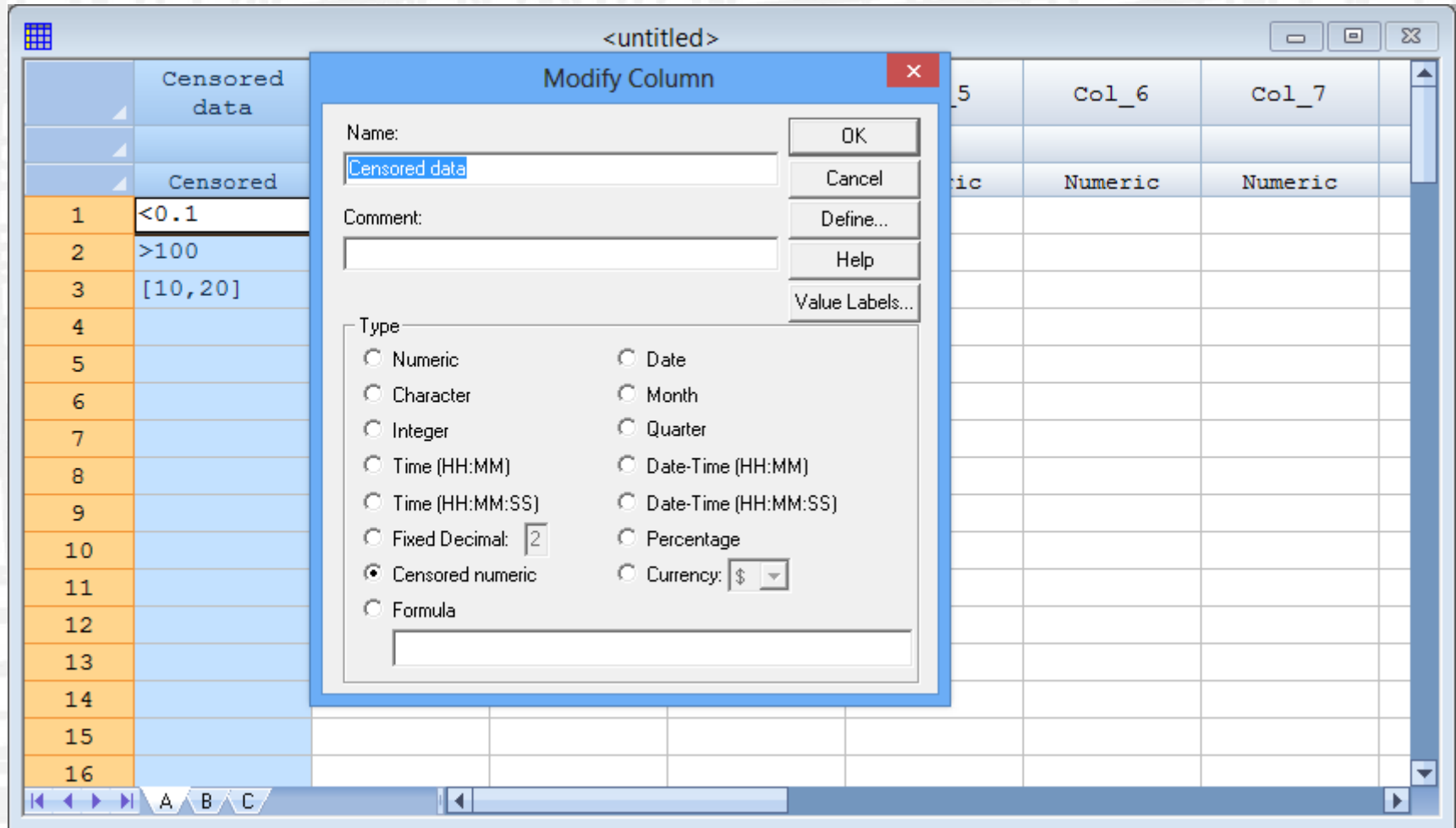
Today's Webinar

- 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



New Data Types



New Data Types

The screenshot displays the Statgraphics 18 software interface. A data table is visible in the background with columns 'Censored data' and 'Dollars'. A 'Modify Column' dialog box is open, allowing users to change the data type and format of a selected column. The dialog box includes fields for 'Name' (currently 'Dollars') and 'Comment'. The 'Type' section offers various data types: Numeric, Character, Integer, Time (HH:MM), Time (HH:MM:SS), Fixed Decimal (set to 2), Censored numeric, Formula, Date, Month, Quarter, Date-Time (HH:MM), Date-Time (HH:MM:SS), Percentage, and Currency (selected). The Currency dropdown menu is open, showing options for \$, €, £, and ¥. The background table shows data for rows 1 through 16, with the 'Dollars' column containing values like \$9.99, \$19.99, and \$5.50.

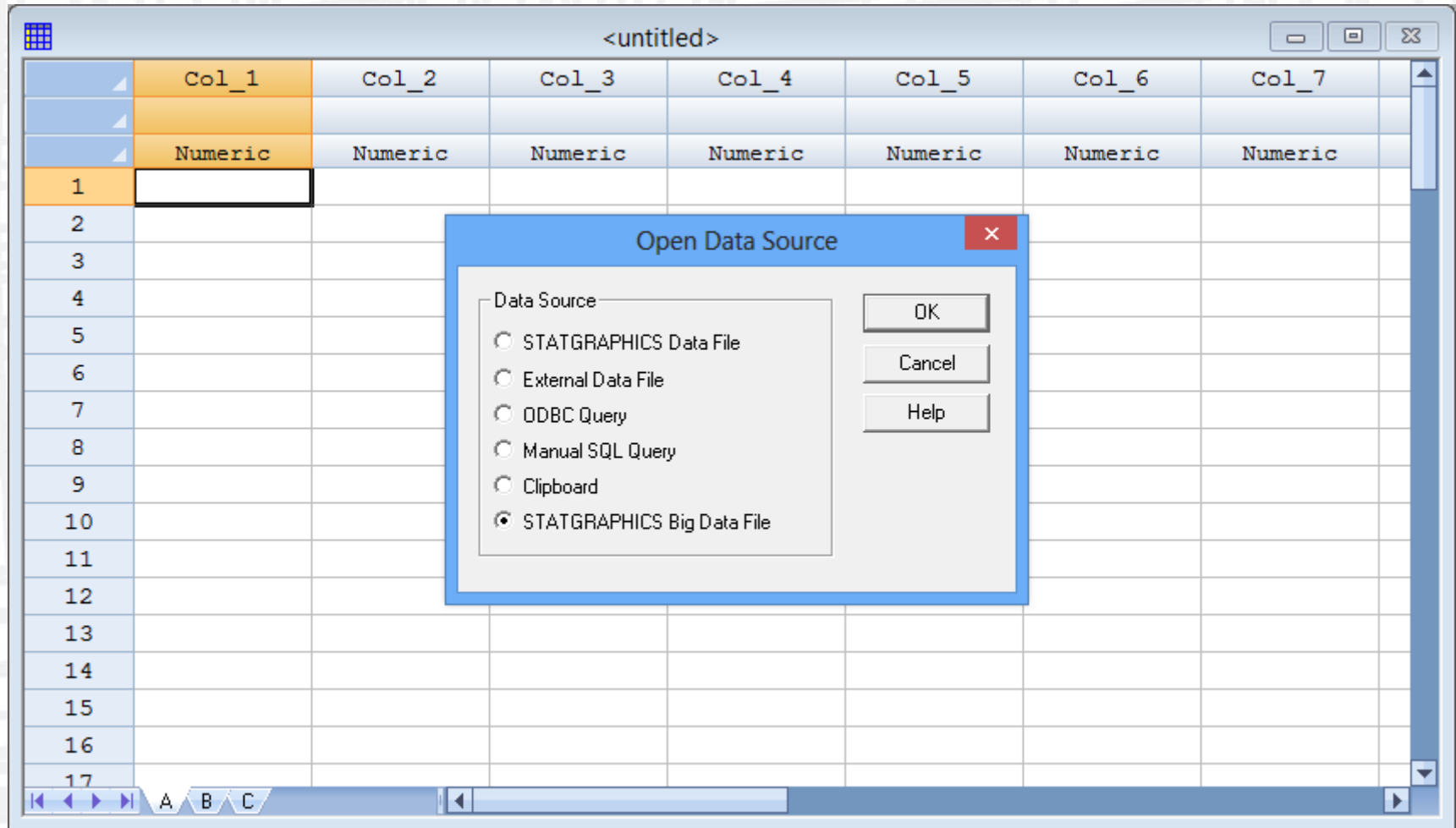
	Censored data	Dollars
1	<0.1	\$9.99
2	>100	\$19.99
3	[10,20]	\$5.50
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		
16		

R Interface Menu

The screenshot displays the STATGRAPHICS 18 - Untitled StatFolio window. The menu bar includes File, Edit, Plot, Describe, Compare, Relate, Forecast, SPC, DOE, SnapStats!!, Statlets, Tools, R Interface, View, Window, and Help. The R Interface menu is open, showing options: R - Installation and Configuration, Exchange Data, Execute Script, Classification and Regression Trees..., Distribution Fitting (Arbitrarily Censored Data)..., Multidimensional Scaling..., Text Mining..., and X-13ARIMA-SEATS Seasonal Adjustment... The main data table is visible with columns: Censored data, Dollars, Col_3, Censored, Currency, Numeric, Numeric, Numeric, and Numeric. The first three rows of data are highlighted in orange.

	Censored data	Dollars	Col_3					
	Censored	Currency	Numeric	Numeric	Numeric	Numeric	Numeric	Numeric
1	<0.1	\$9.99						
2	>100	\$19.99						
3	[10,20]	\$5.50						
4								
5								
6								
7								
8								
9								
10								
11								
12								
13								
14								
15								
16								

Big Data Files (.sgb)



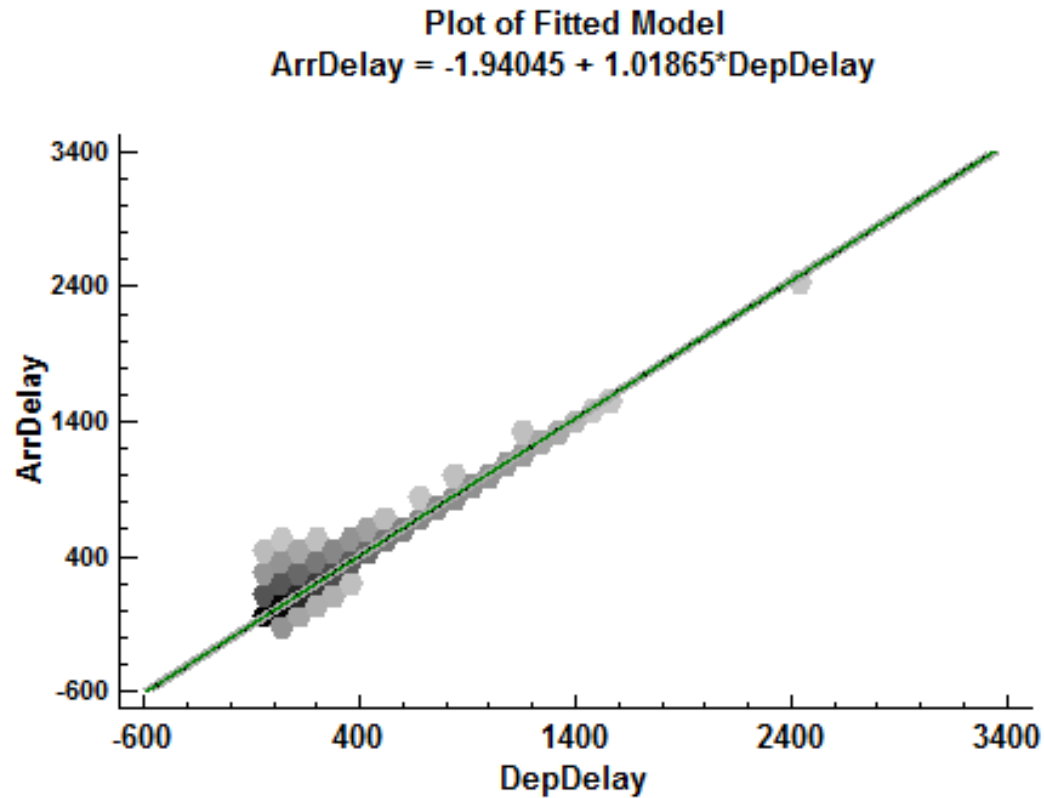
Big Data Files

- 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.

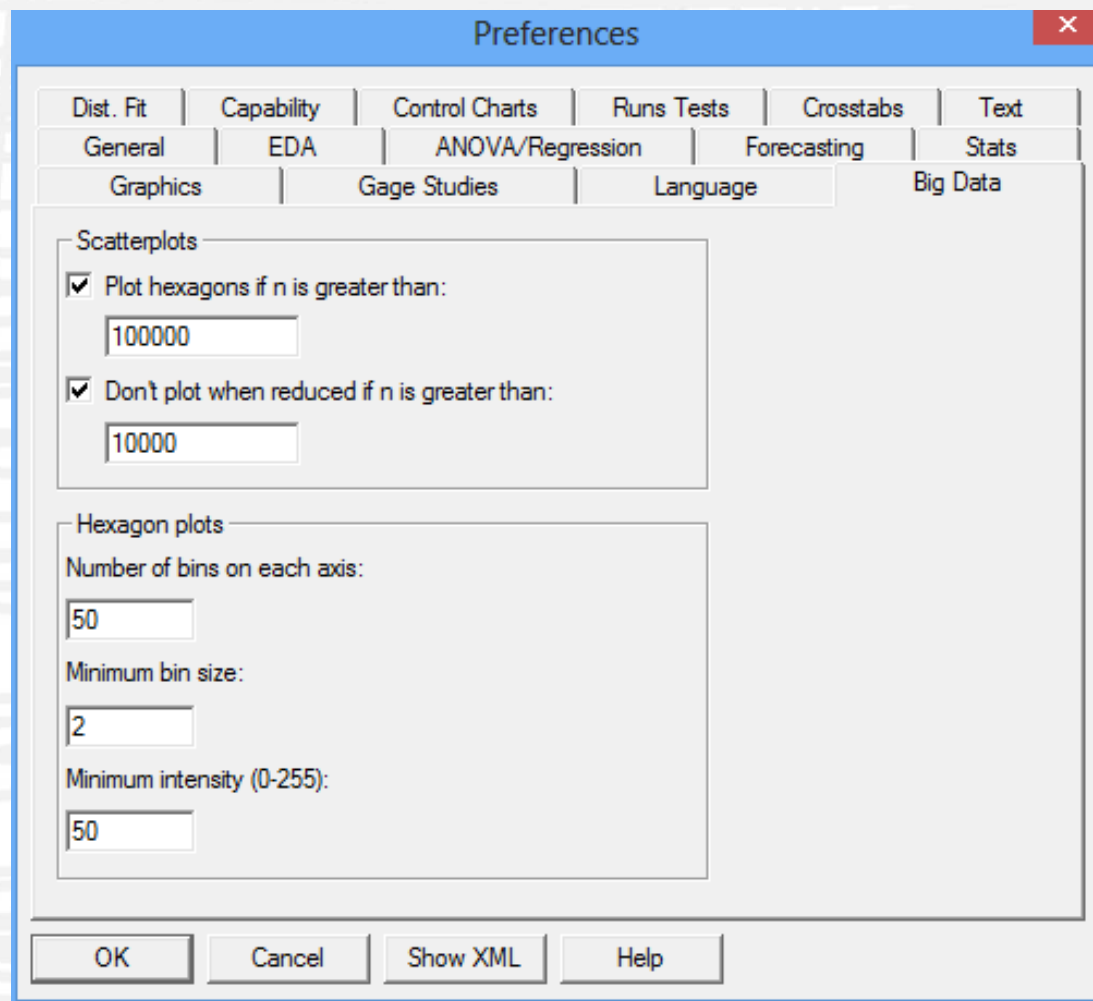
2008 Flight data

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

Hexagon Plots



Big Data Preferences



The image shows a screenshot of the Minitab 'Preferences' dialog box, with the 'Big Data' tab selected. The dialog box has a blue title bar with the text 'Preferences' and a close button. Below the title bar is a tabbed interface with the following tabs: 'Dist. Fit', 'Capability', 'Control Charts', 'Runs Tests', 'Crosstabs', 'Text', 'General', 'EDA', 'ANOVA/Regression', 'Forecasting', 'Stats', 'Graphics', 'Gage Studies', 'Language', and 'Big Data'. The 'Big Data' tab is active, showing two sections: 'Scatterplots' and 'Hexagon plots'. In the 'Scatterplots' section, there are two checked options: 'Plot hexagons if n is greater than:' with a text box containing '100000', and 'Don't plot when reduced if n is greater than:' with a text box containing '10000'. In the 'Hexagon plots' section, there are three settings: 'Number of bins on each axis:' with a text box containing '50', 'Minimum bin size:' with a text box containing '2', and 'Minimum intensity (0-255):' with a text box containing '50'. At the bottom of the dialog box are four buttons: 'OK', 'Cancel', 'Show XML', and 'Help'.

Preferences

Dist. Fit | Capability | Control Charts | Runs Tests | Crosstabs | Text
General | EDA | ANOVA/Regression | Forecasting | Stats
Graphics | Gage Studies | Language | **Big Data**

Scatterplots

☒ Plot hexagons if n is greater than:
100000

☒ Don't plot when reduced if n is greater than:
10000

Hexagon plots

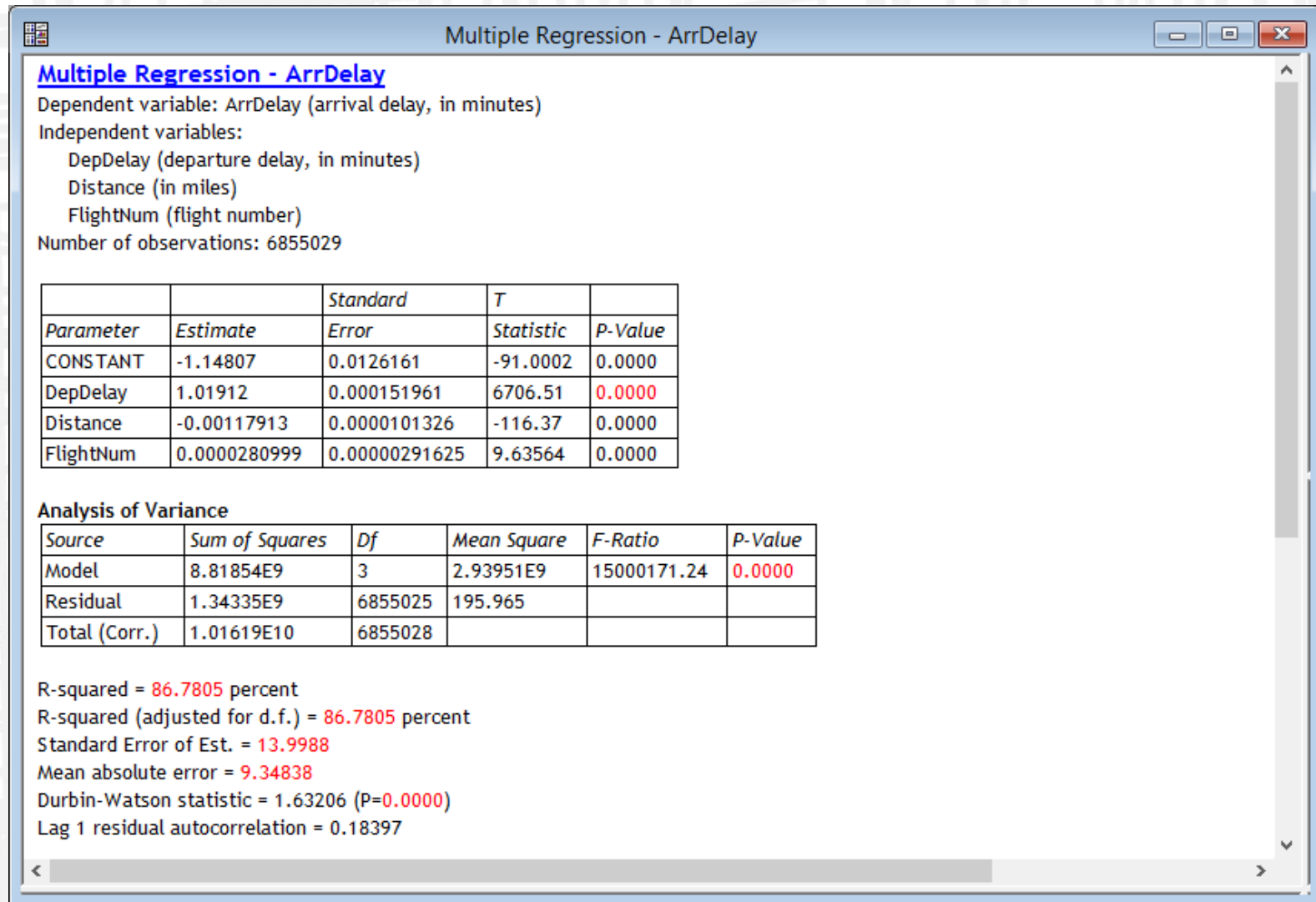
Number of bins on each axis:
50

Minimum bin size:
2

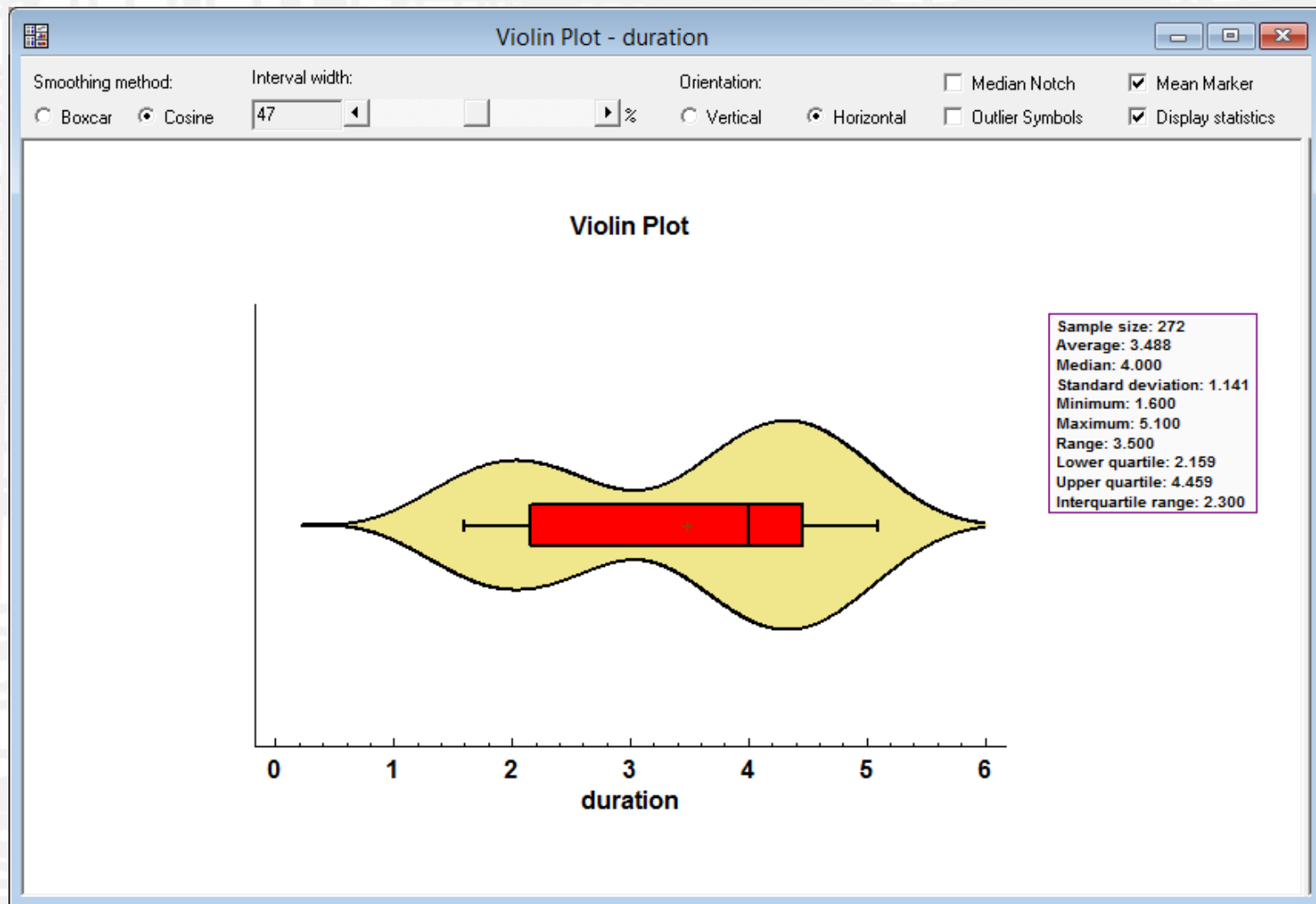
Minimum intensity (0-255):
50

OK Cancel Show XML Help

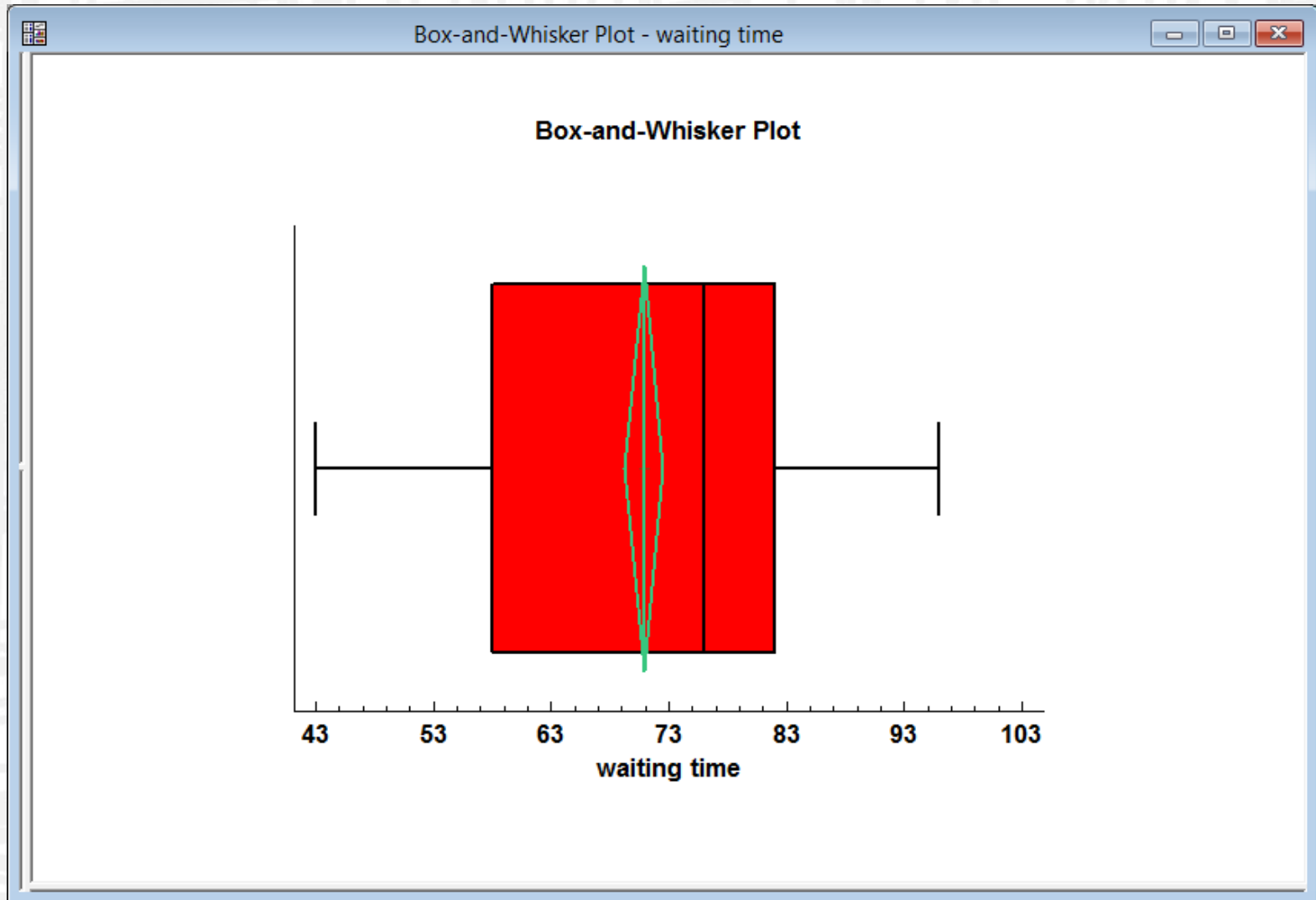
Judging Significance



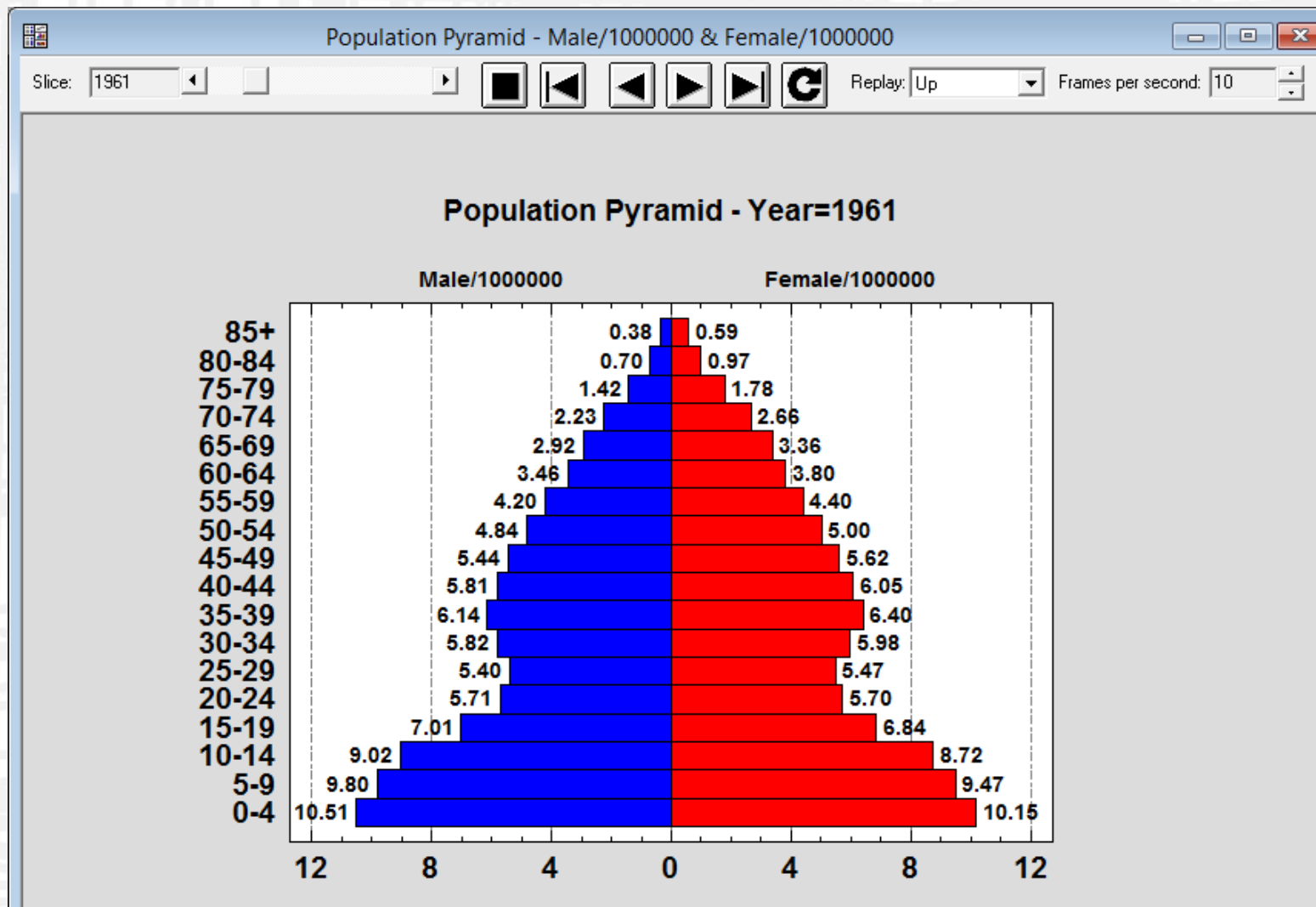
Data Visualization: Violin Plot



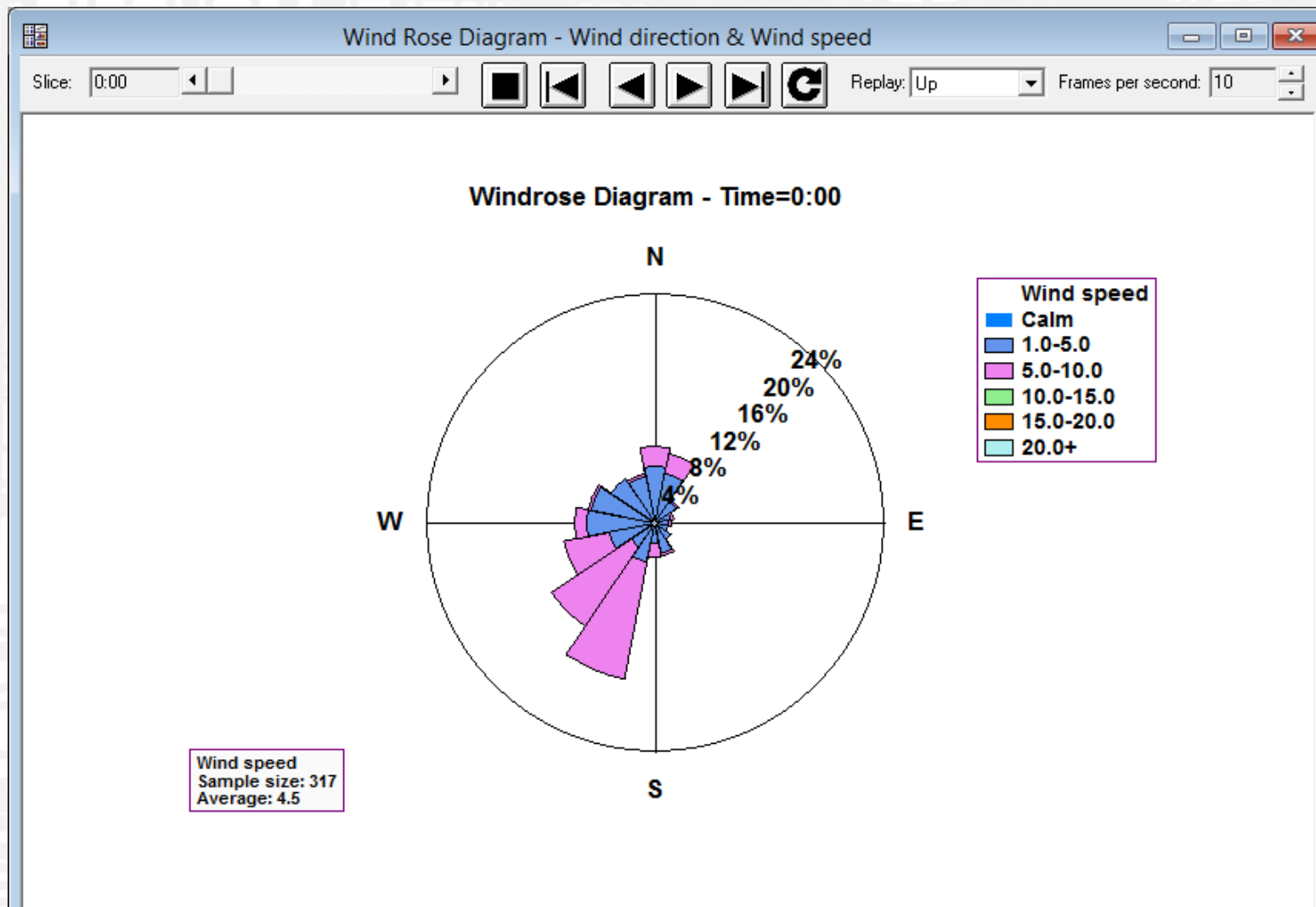
Diamond Plot



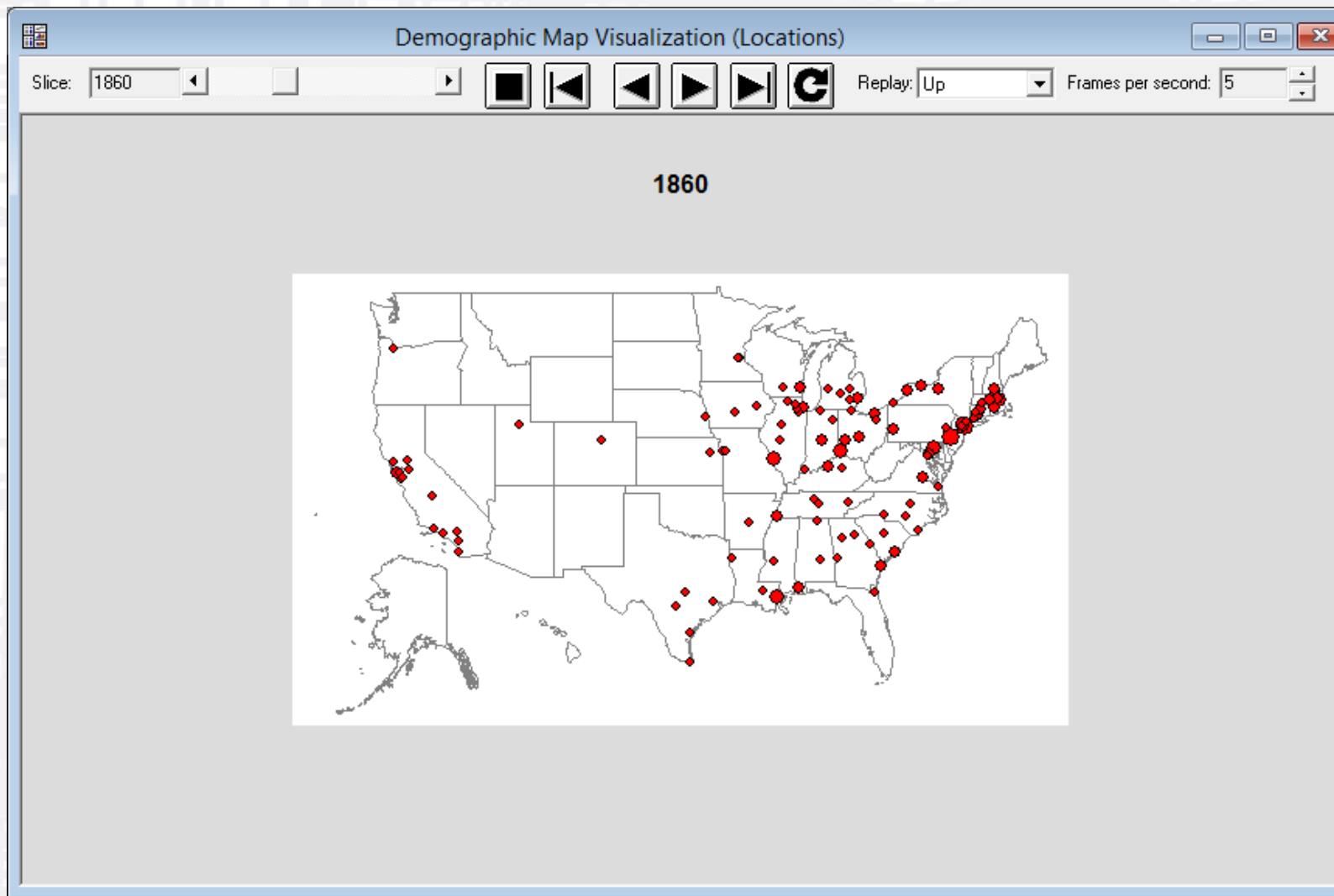
Population Pyramid



Wind Rose



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

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

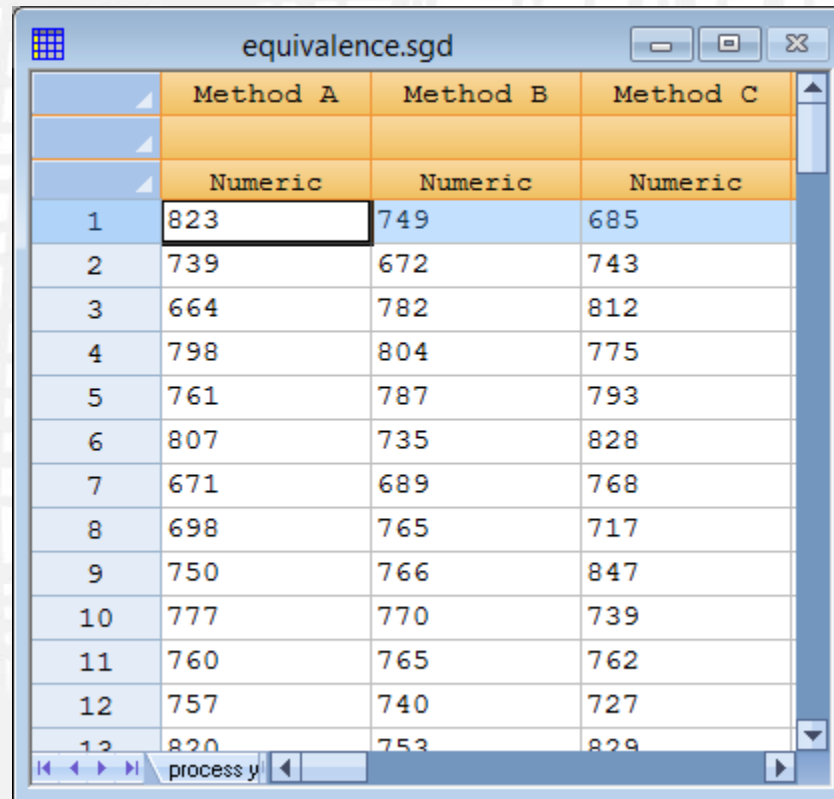
$$H_A: \Delta_L \leq \mu_1 - \mu_2 \leq \Delta_U$$

- Inferiority test

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

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

Example: 3 Methods



	Method A	Method B	Method C
	Numeric	Numeric	Numeric
1	823	749	685
2	739	672	743
3	664	782	812
4	798	804	775
5	761	787	793
6	807	735	828
7	671	689	768
8	698	765	717
9	750	766	847
10	777	770	739
11	760	765	762
12	757	740	727
13	820	753	829

Analysis Options

Equivalence/Noninferiority Tests Options

Null Hypothesis

- ☒ Not equivalent (two-sided)
- ☐ Inferior (less than)
- ☐ Inferior (greater than)

Standard error

- ☒ Pool 2 sample variances
- ☐ Pool all sample variances
- ☐ Allow for unequal variances
- ☐ Use z instead of t

Equivalence limits

Lower differential:

Upper differential:

Alpha: %

☐ Display 100(1-2alpha)% C.I.

OK Cancel Help

Two One-Sided Tests (TOST)

Equivalence & Noninferiority Tests - Comparison of Two Independent Sa...

Equivalence & Noninferiority Tests - Comparison of Two Independent Samples

Sample 1: Method A
Sample 2: Method B
Sample 3: Method C

Sample Statistics

Sample	n	Minimum	Maximum	Mean	Std. deviation
Method A	50	664.0	844.0	744.26	46.5586
Method B	50	672.0	844.0	752.64	40.3782
Method C	50	667.0	892.0	775.68	49.4981

Equivalence Analysis
Null hypothesis: Not equivalent (two-sided)
Lower equivalence differential: -25.0
Upper equivalence differential: 25.0

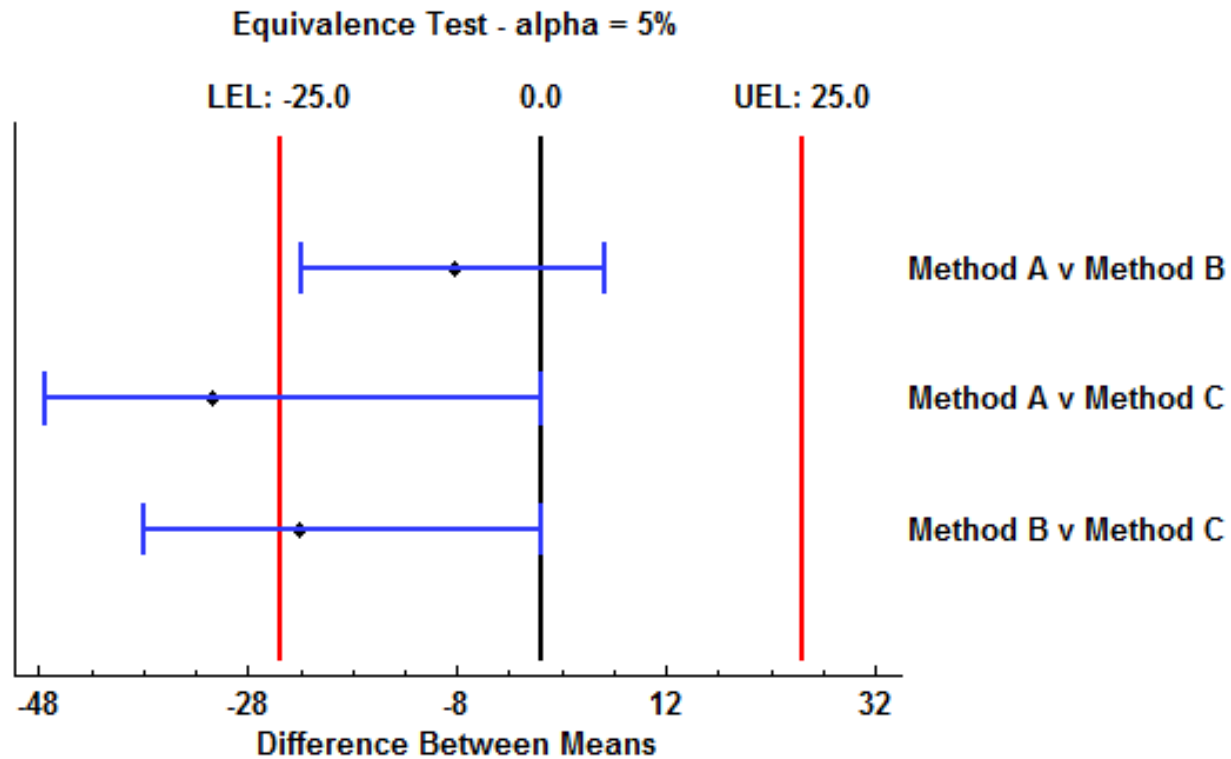
Comparison	Difference	Std. error	Lower 95% CL	Upper 95% CL
Method A v Method B	-8.38	8.71562	-22.8528	6.09277
Method A v Method C	-31.42	9.61017	-47.3782	0.0
Method B v Method C	-23.04	9.03378	-38.0411	0.0

Comparison	Lower t-value	Upper t-value	Lower P-value	Upper P-value
Method A v Method B	1.90692	-3.8299	0.0297	0.0001
Method A v Method C	-0.668043	-5.87087	0.7472	0.0000
Method B v Method C	0.216963	-5.31782	0.4143	0.0000

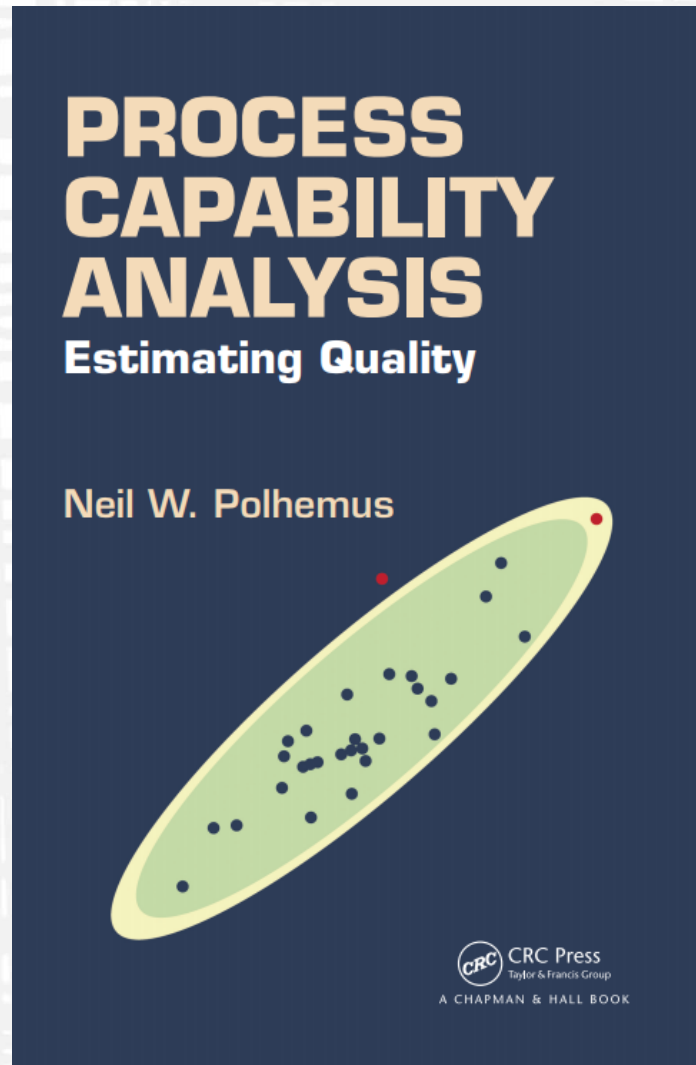
Comparison	Maximum P-value	Conclusion (alpha=5%)
Method A v Method B	0.0297	Equivalence has been demonstrated.
Method A v Method C	0.7472	Equivalence has not been demonstrated.
Method B v Method C	0.4143	Equivalence has not been demonstrated.

Note: The standard error was estimated by pooling 2 sample variances.

Confidence Intervals



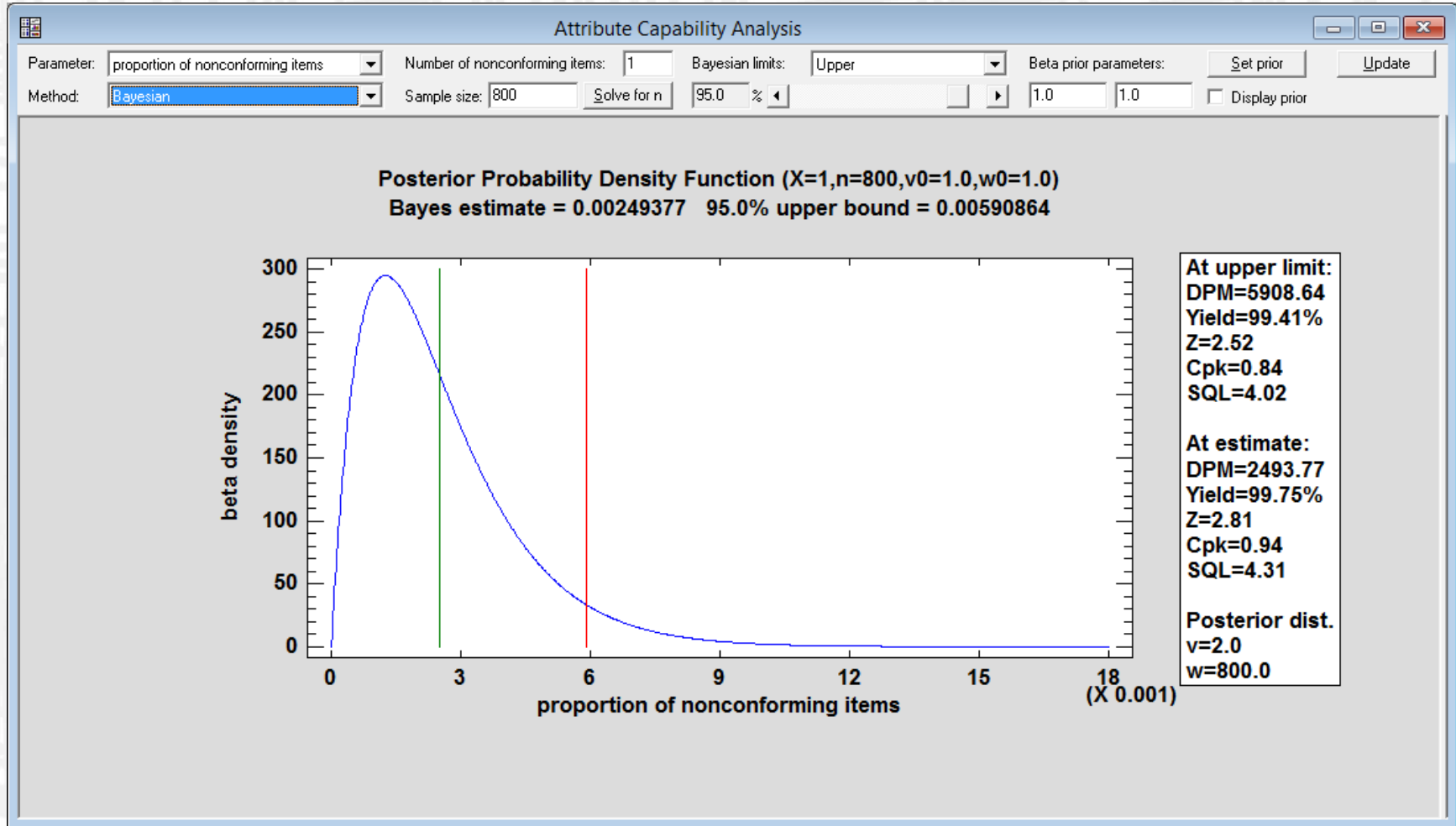
Process Capability Analysis



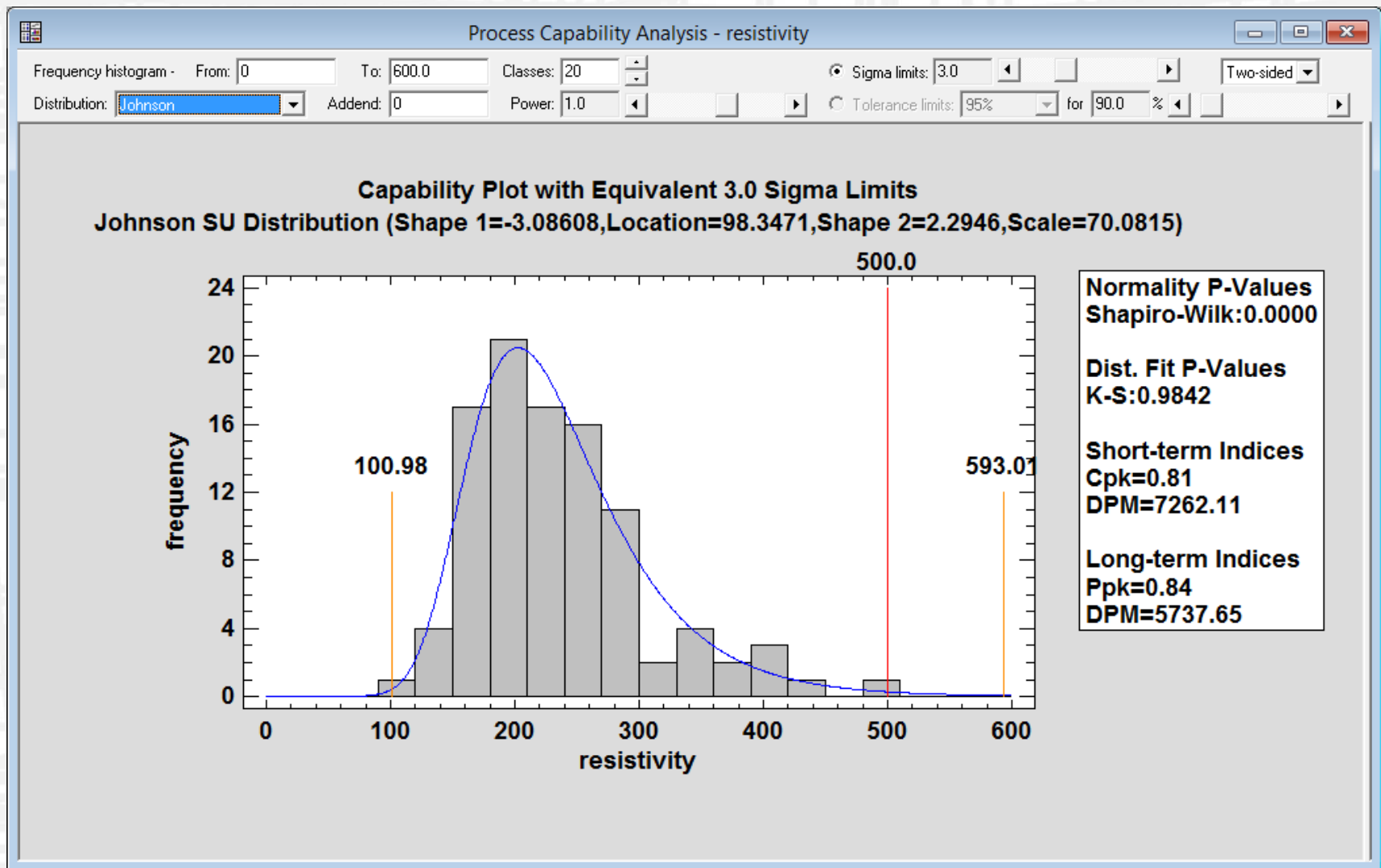
New SPC Procedures

- 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

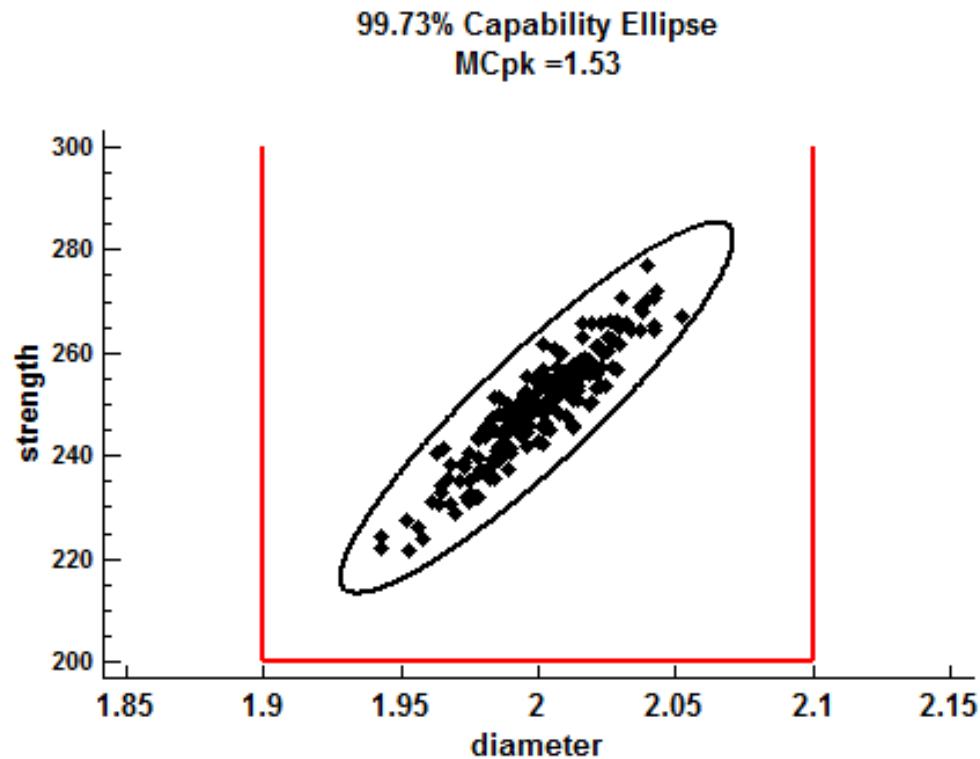
Attribute Capability Analysis Statlet



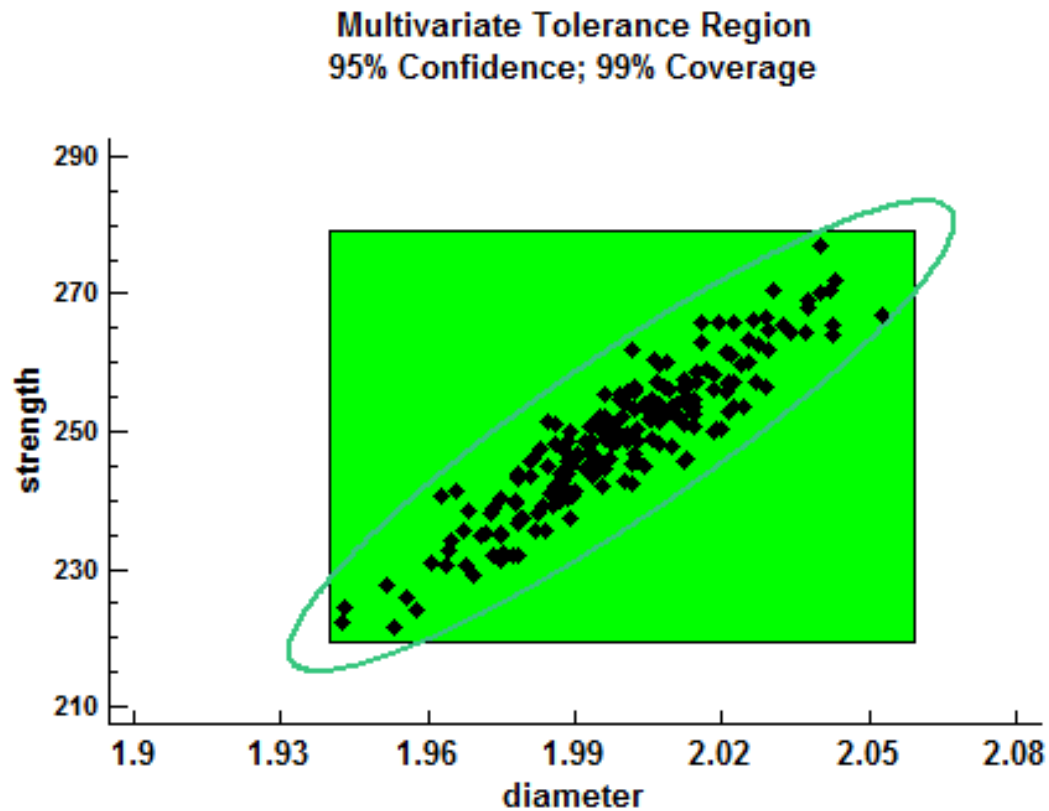
Variables Capability Analysis Statlet



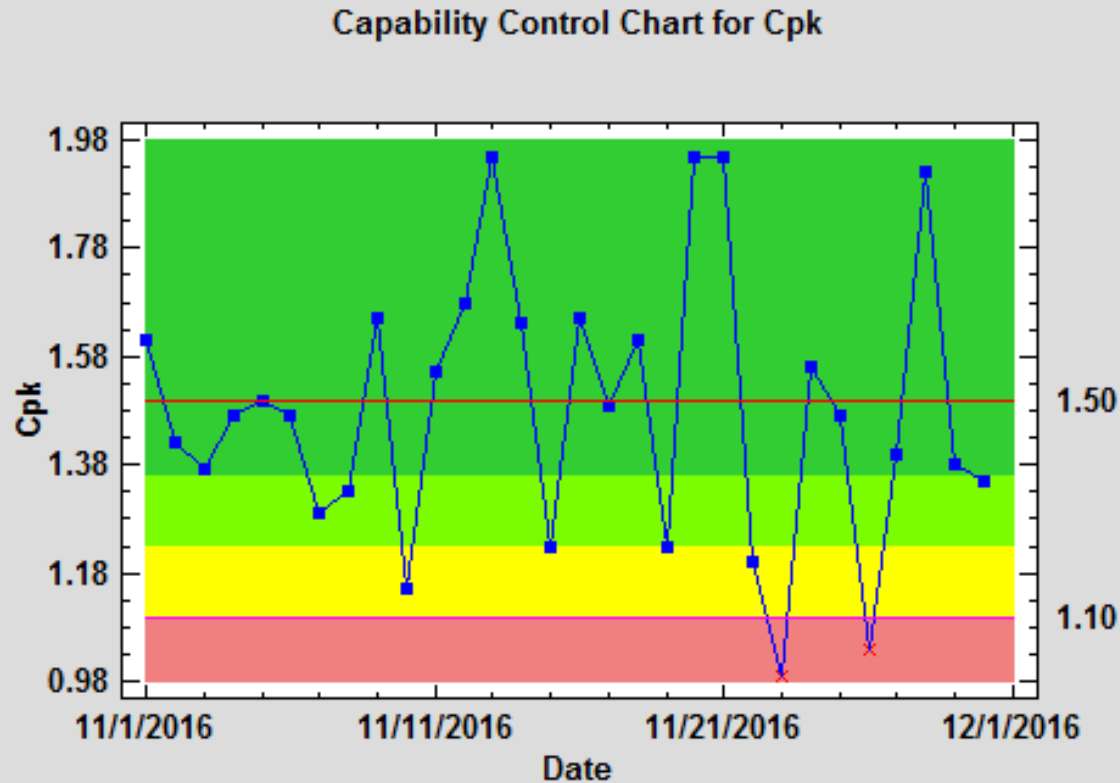
Multivariate Capability Analysis



Multivariate Tolerance Regions



Capability Control Charts



R Interface Enhancements

R - Installation and Configuration [X]

1. To install R, click the 'download R' link on the R-project website:

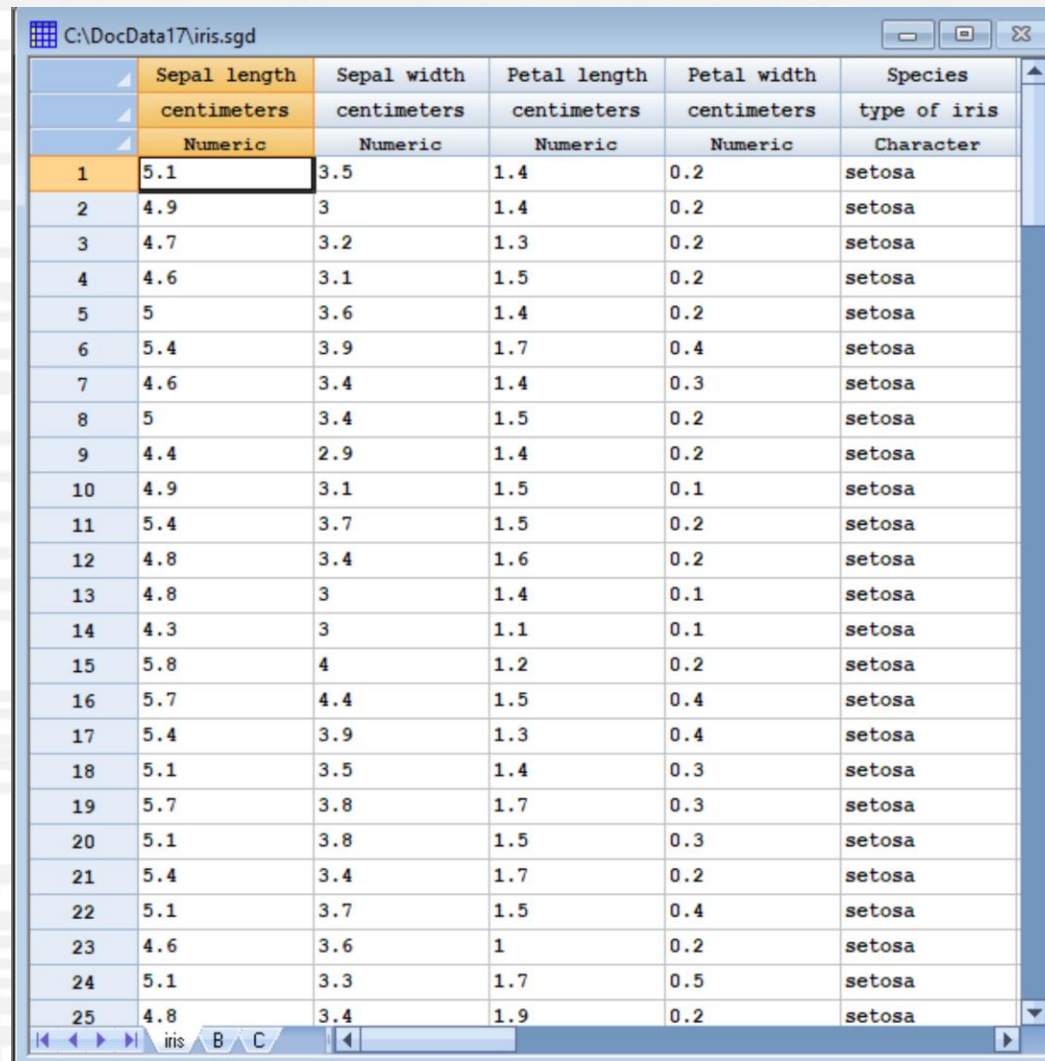
2. After installing R, enter the path to Rgui.exe in the field below:

3. Set the maximum time to wait for R to execute a set of commands: seconds

4. Install the R packages for the procedures you wish to use. After pressing a button, type Ctrl-V to copy and execute the required commands.

<input type="button" value="List installed packages"/>	
<input type="button" value="Install pandoc,rmarkdown,Rcpp,stringi"/>	Required by all procedures and for executing scripts.
<input type="button" value="Install ggplot2"/>	Required to create graphs.
<input type="button" value="Install seasonal"/>	For X-13ARIMA-SEATS Seasonal Adjustment.
<input type="button" value="Install interval,lens"/>	For nonparametric analysis of arbitrarily censored data.
<input type="button" value="Install tm,SnowballC,wordcloud,igraph"/>	For Text Mining.
<input type="button" value="Install MASS"/>	For Multidimensional Scaling.
<input type="button" value="Install tree"/>	For classification and regression trees.

Classification Trees - Iris Data



	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

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

CART Options ×

Type of Tree

☒ Classification

☐ Regression

Partitioning

Smallest allowed node size:

Minimum observations in each child:

Minimum within-node deviance to split:

Pruning

☒ None

☐ Specify number of leaves

☐ Crossvalidate

Number of leaves:

Training Set

☒ All rows

☐ First half of rows

☐ First

rows

☐ Every other row

R Tree Plot

Font size:

☒ Abbreviate labels

letters

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)

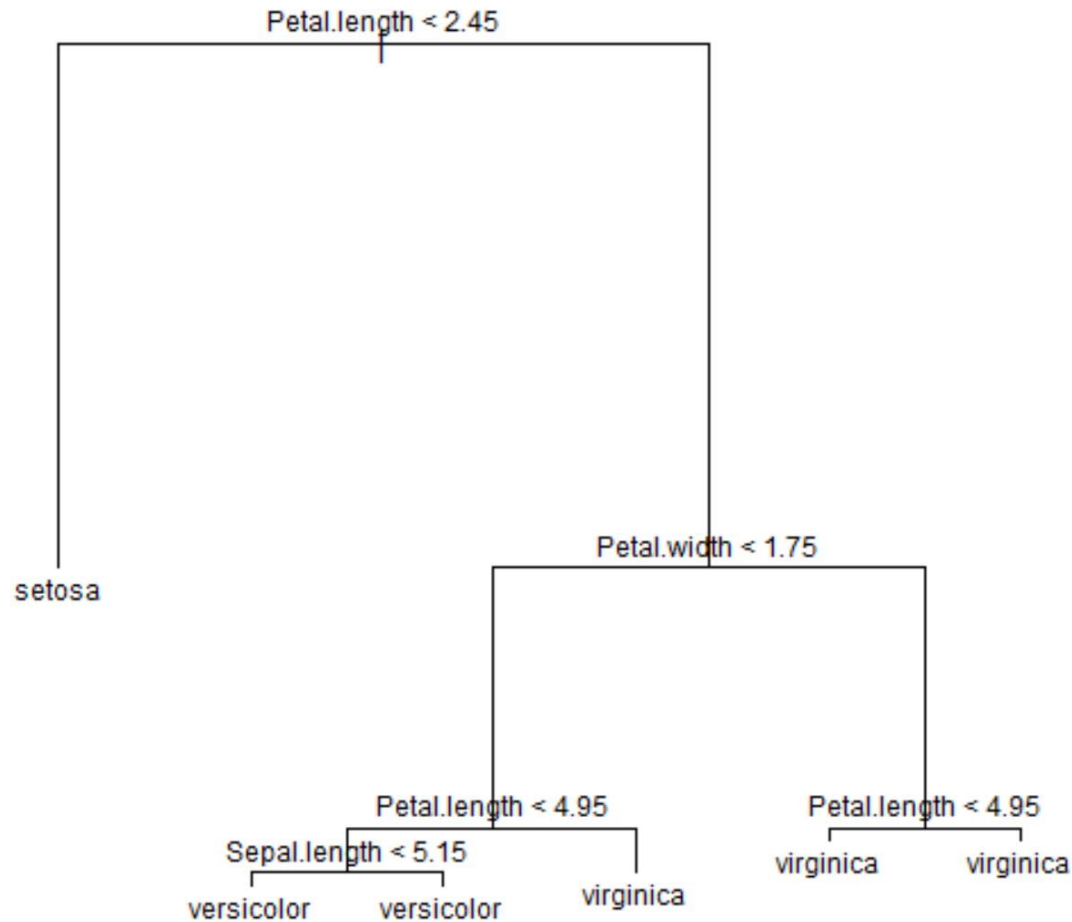
p<-prune.tree(treefit)
write.table(treefit$frame,file="C:\\\\Users\\\\NEIL~1.STA\\\\AppData\\\\Local\\\\Temp\\\\frame.csv",sep=".",)
write.table(treefit$where,file="C:\\\\Users\\\\NEIL~1.STA\\\\AppData\\\\Local\\\\Temp\\\\where.csv",sep=".",row.names=FALSE)
write.table(cbind(p$size,p$dev,p$sk),file="C:\\\\Users\\\\NEIL~1.STA\\\\AppData\\\\Local\\\\Temp\\\\prune.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

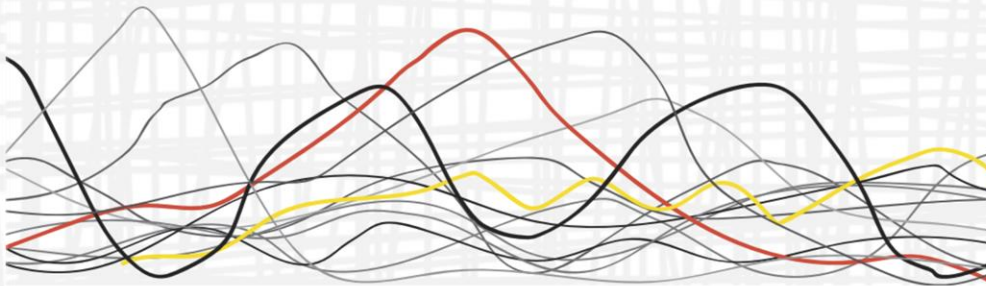
register.statgraphics.com



©Copyright 1982-2017 by Statgraphics Technologies, Inc.

This program is protected by U.S. and international copyright laws as described in the About dialog box.

Version 18 Registration Program



TRIAL PERIOD

REGISTERED SERIAL NUMBERS


ACTIVATIONS

ASSOCIATED SITE LICENSES



STATGRAPHICS.COM

Network Management Program

 Statgraphics Network Management Program ×

Serial number:

Product key:

Organization:

Administrator e-mail address:

License directory:

Browse

Status: NOT ACTIVATED

Activate

Deactivate

Upgrade

Manual activation code (if automatic activation fails):

Activate manually

Total number of seats:
Seats in use or checked out:
Number of seats checked out:

Check in/check out seat

Troubleshoot

Exit

More Information

- Recorded webinar will be posted at:

www.statgraphics.com/webinars

- Version 18 videos are available at:

www.statgraphics.com/instructional-videos