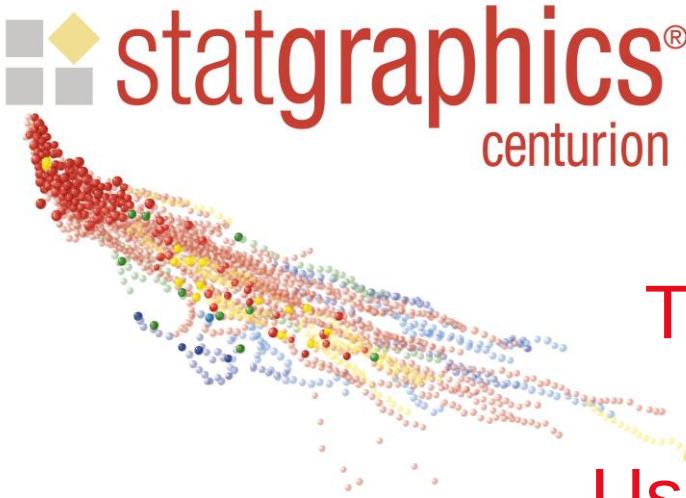


Tips, Tricks and Time-Savers: Using Statgraphics Operators

Presented by Dr. Neil W. Polhemus

Statgraphics Operators

- Operators are predefined functions that perform specific tasks.

- They are used in:
  - Data fields on analysis input dialog boxes.
  - "Select" field on analysis input dialog boxes.
  - Data editor when using "Generate Data".
  - Procedures such as *Nonlinear Regression* to specify statistical models.

Types of Operators

- Algebraic operators (+, -, /, *, ^)
- Mathematical transformations (LOG, SQRT, ABS, ROUNDTO, …)
- Sequential operators (RUNTOT, DIFF, SDIFF, …)
- Random number generators (RNORMAL, RUNIFORM, …)
- Statistical summaries (MEAN, SD, MEDIAN, …)
- Distribution functions (NORMAL, INVNORMAL, …)
- Boolean operators (<, >, =, <=, >=, <>, |, &)
- Data selectors (TAKE, TAKELAST, DROP, …)
- Pattern generators (COUNT, REP, RESHAPE, …)
- Utility functions (REPLACE, JOIN, JUXTAPOSE, ENDSWITH, …)

statgraphics® centurion
statgraphics.com

Reference

- Select *Help – Procedure Documentation* from the main menu. Click on the PDF file titled "STATGRAPHICS Operators".

PROPER(x)
Purpose: converts each string in a character column to a proper name by capitalizing the first letter of each word in the string.
Type: utility function
Argument: data column
Example: PROPER(*make*)
Result: column of strings

RNORMAL(n,mu,sigma)
Purpose: generates random numbers from a normal distribution
Type: random number generator
Argument: sample size, mean, standard deviation
Example: RNORMAL(3,10,3)
Result: 13.4892 9.85616 11.9911

statgraphics® centurion
statgraphics.com

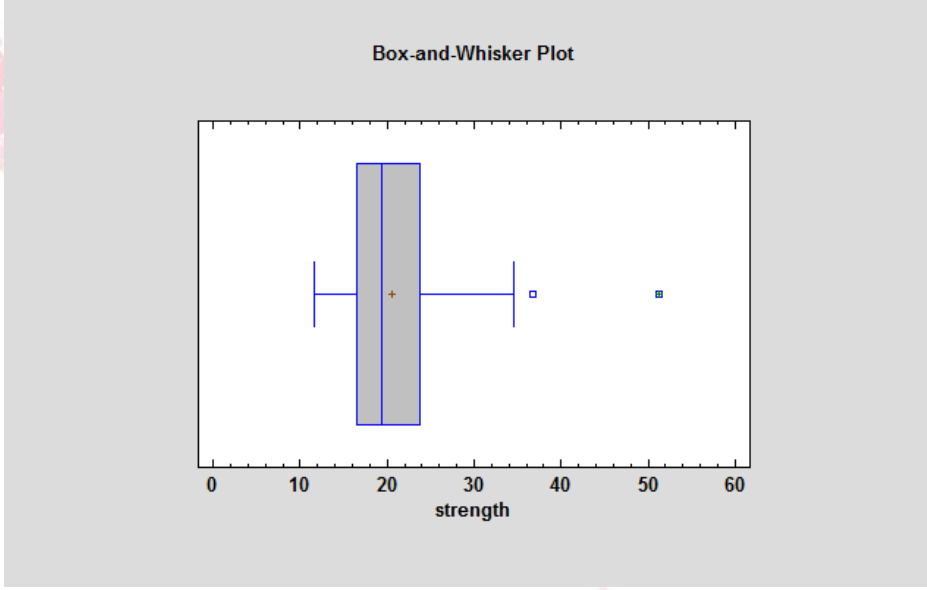
Application: Transform data to satisfy assumptions of method to be used

- Example #1: I have a set of data that may contain outliers. I'd like to use Grubbs' test, which assumes the data come from a normal distribution. My data appear to be skewed.

- Example #2: I've created a designed experiment. The data I wish to analyze are counts, which are likely to follow a binomial distribution in which the variance is a function of the mean.

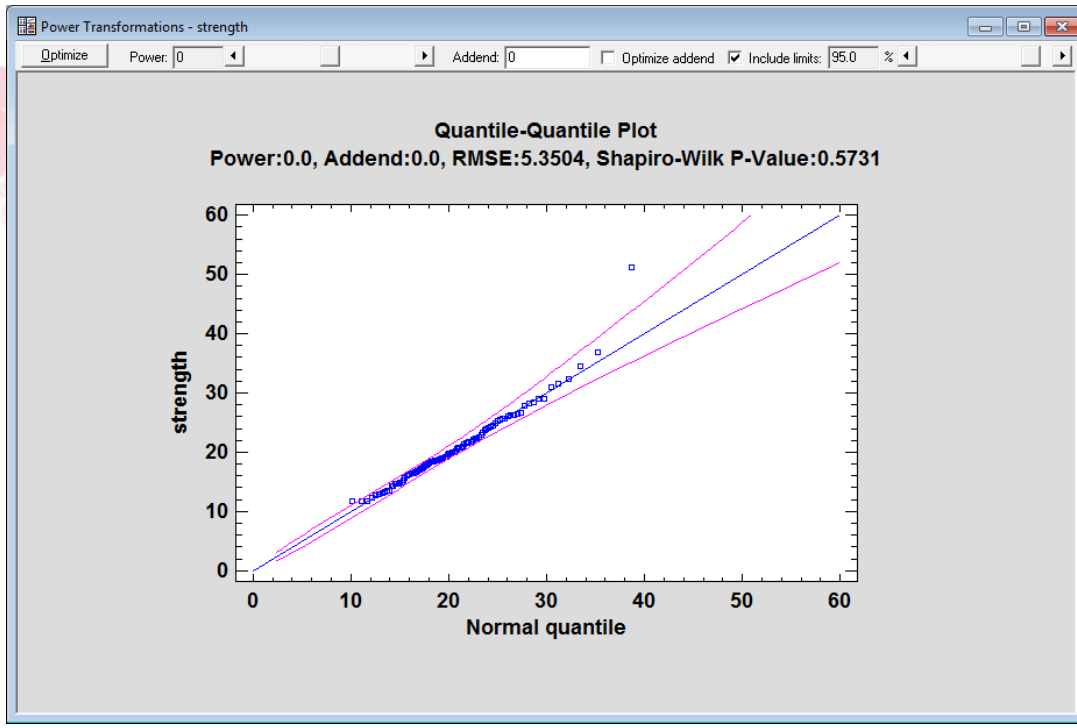
statgraphics® centurion
statgraphics.com

Example 1: Tensile Strength



statgraphics® centurion
statgraphics.com

Power Transformations Statlet



statgraphics® centurion
statgraphics.com

Outlier Identification

Outlier Identification - LOG(strength)

95.0% confidence intervals for the mean

	Lower Limit	*Upper Limit*
Standard	2.93211	3.03985
Winsorized	2.91209	3.04707

Sorted Values

		Studentized Values	*Studentized Values*	*Modified*
Row	*Value*	*Without Deletion*	*With Deletion*	*MAD Z-Score*
51	2.45531	-1.95458	-2.00377	-2.04871
12	2.45616	-1.95142	-2.0004	-2.04526
58	2.4681	-1.90746	-1.95357	-1.99721
49	2.51447	-1.73668	-1.77283	-1.81056
52	2.53845	-1.64835	-1.68002	-1.71401
...				
20	3.45062	1.71137	1.74619	1.95807
11	3.47538	1.80255	1.84233	2.05772
85	3.54241	2.04944	2.10526	2.32756
93	3.60631	2.28481	2.35991	2.58482
18	3.93613	3.4996	3.75979	3.91254

Grubbs' Test (assumes normality)
Test statistic = 3.4996
P-Value = 0.0308909

statgraphics® centurion
statgraphics.com

Example 2: Mothproofing

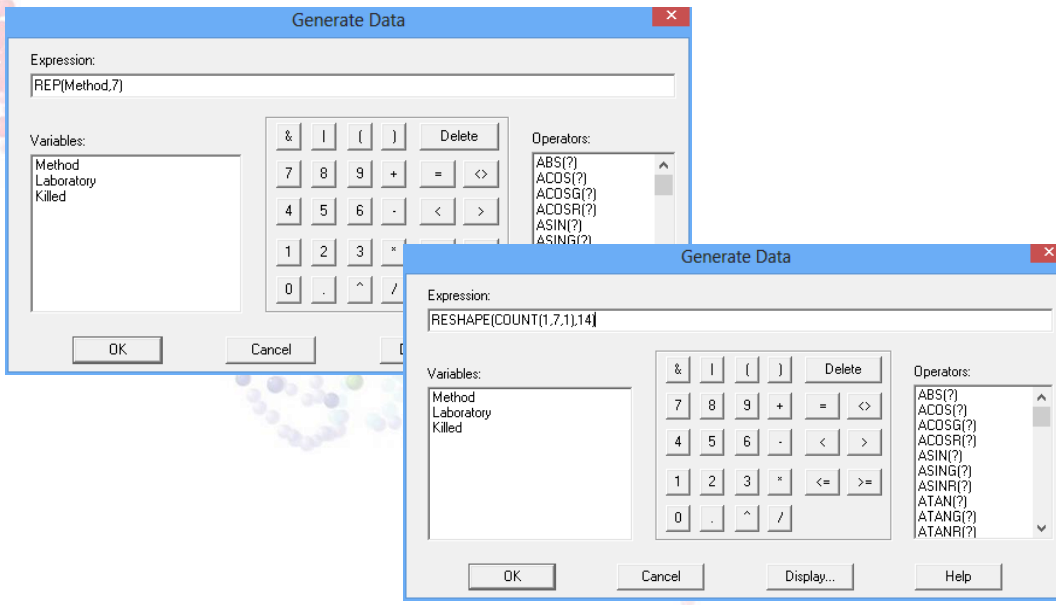
- Data: percentage of 20 moth larvae killed

	Lab 1	Lab 2	Lab 3	Lab 4	Lab 5	Lab 6	Lab 7
Method A	40	35	5	80	50	95	45
Method B	60	30	15	95	75	100	55

- Goal: compare the 2 methods

- Source: Statistics for Experimenters by Box, Hunter and Hunter

Data Setup



statgraphics® centurion
statgraphics.com

Data

mothproofing.sgd

	Method	Laboratory	Killed	Col_4	Col_5	Col_6
1	A	1	40			
2	A	2	35			
3	A	3	5			
4	A	4	80			
5	A	5	50			
6	A	6	95			
7	A	7	45			
8	B	1	60			
9	B	2	30			
10	B	3	15			
11	B	4	95			
12	B	5	75			
13	B	6	100			
14	B	7	55			
15						
16						
17						
18						
19						
20						

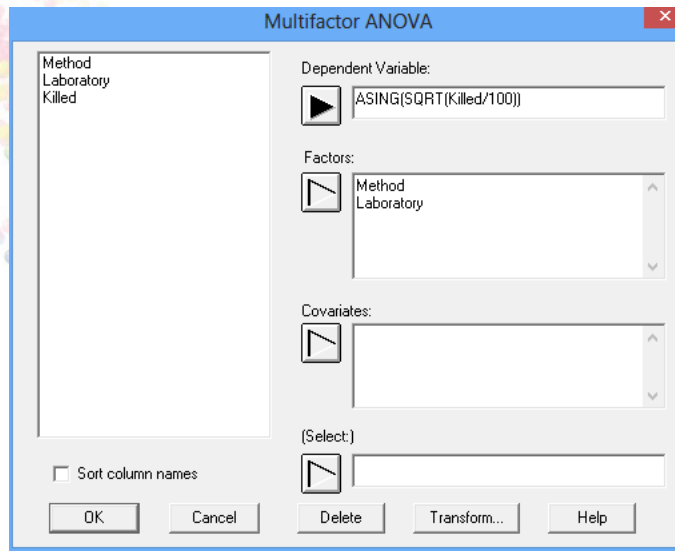
mothproofing | B | C

statgraphics® centurion
statgraphics.com

Multifactor ANOVA

- Fisher's variance stabilizing transformation:

$$x = sin^{-1}\sqrt{\hat{p}}$$



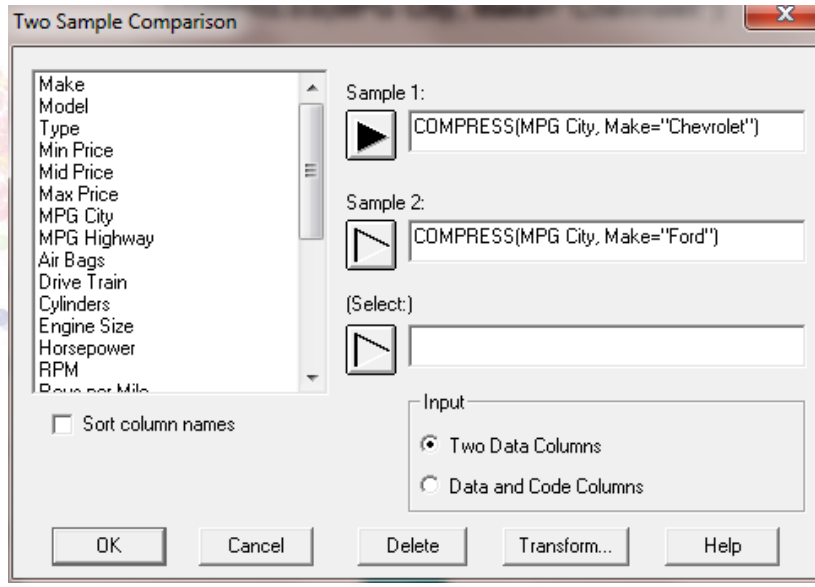
statgraphics centurion
statgraphics.com

Application: Boolean operators

- Example #1: I have a data set containing information on 93 makes and models of automobiles. How can I compare Fords and Chevrolets?

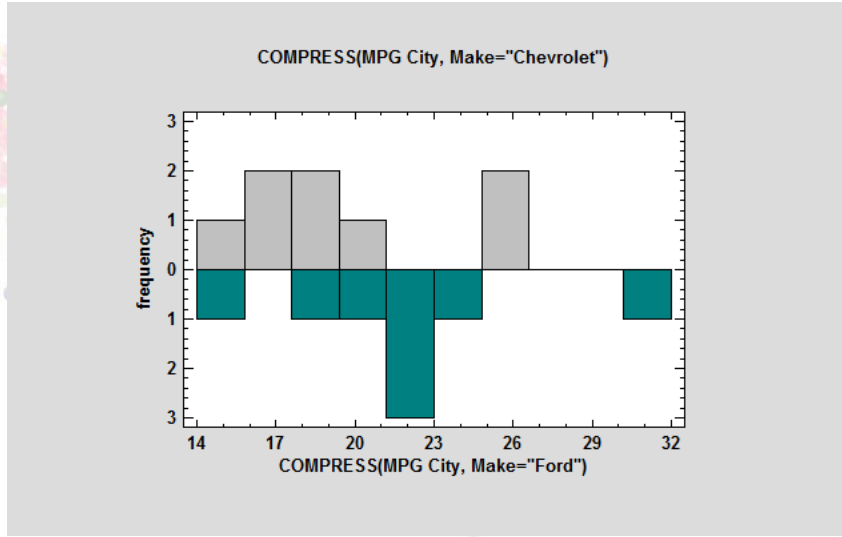
- Example #2: I've collected data for a stability study regarding the loss of chlorine over time in a product I produce. How can I fit a piecewise linear model to the data?

Two Sample Comparison



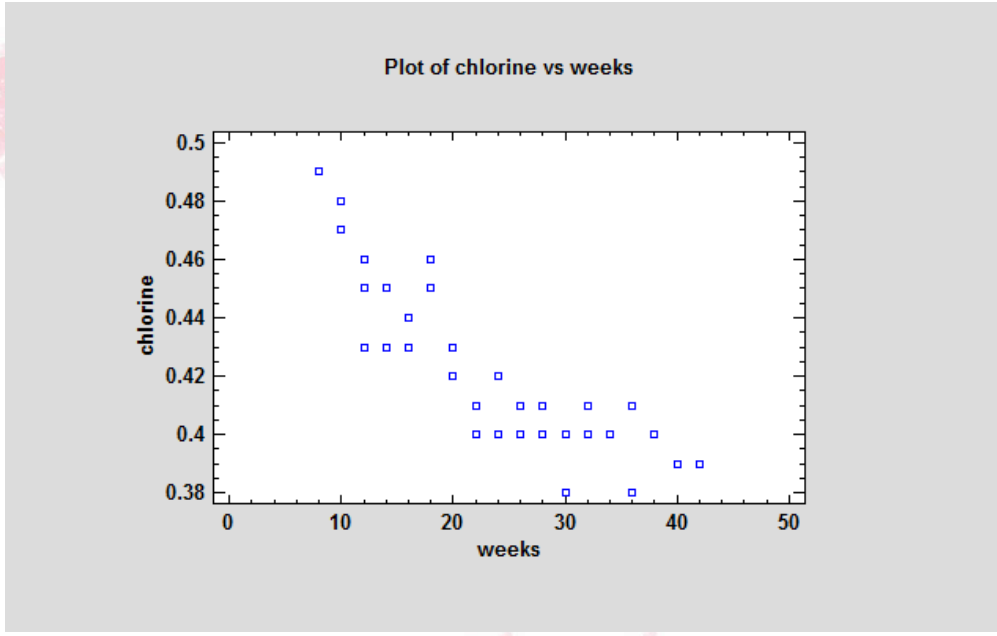
statgraphics centurion
statgraphics.com

Histogram



statgraphics® centurion
statgraphics.com

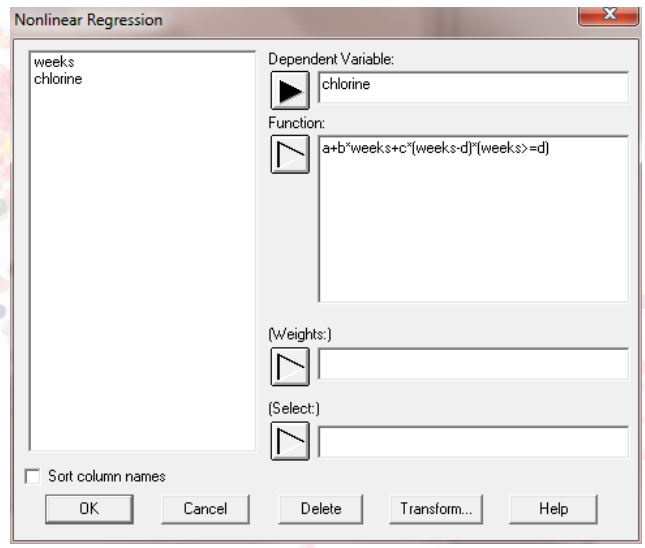
Stability Study



statgraphics centurion
statgraphics.com

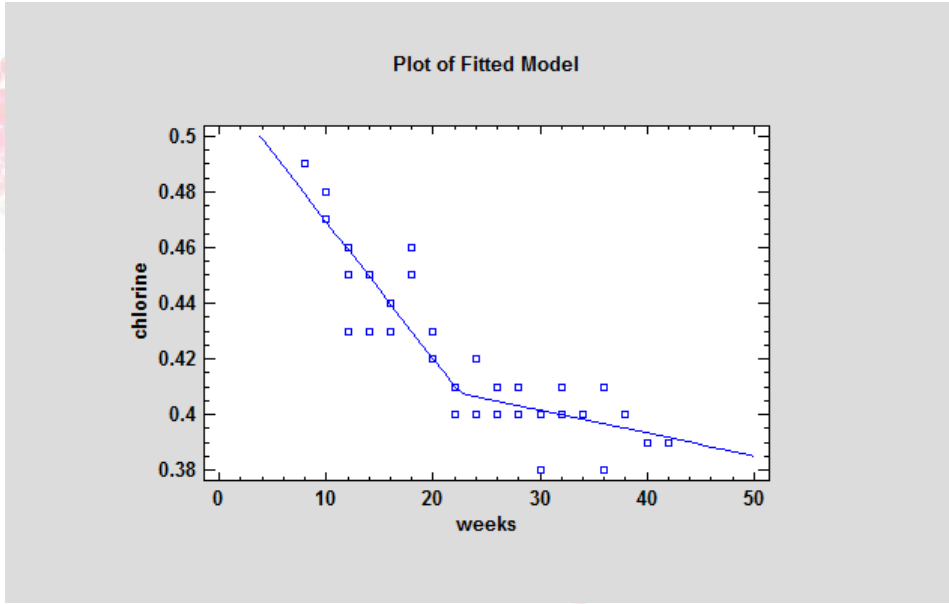
Nonlinear Regression Model

chlorine = a + b*weeks + c*(weeks-d)*(weeks>=d)



statgraphics centurion
statgraphics.com

Piecewise Linear Fit

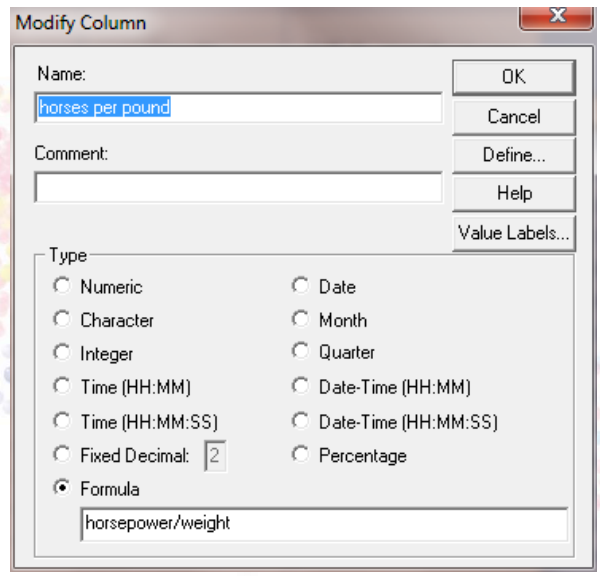


statgraphics®
statgraphics.com centurion

Application: Combining Columns

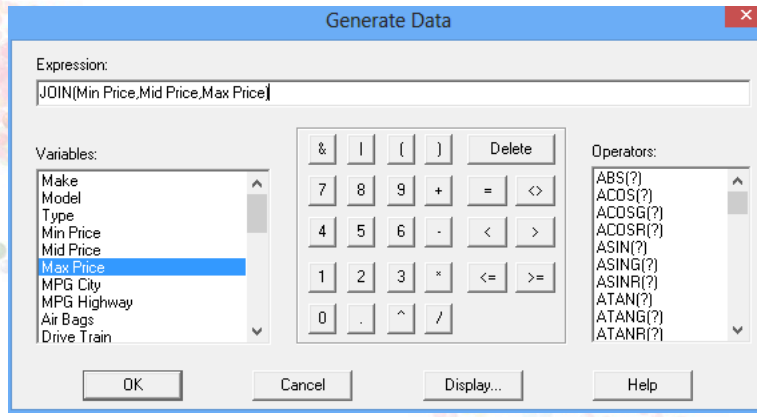
- Example #1: I will take my automobile data and construct a column consisting of the ratio of horsepower to weight.
- Example #2: I will join end to end my 3 columns with price.
- Example #3: I will combine “make” and “model” into a single column.

Defining a Function



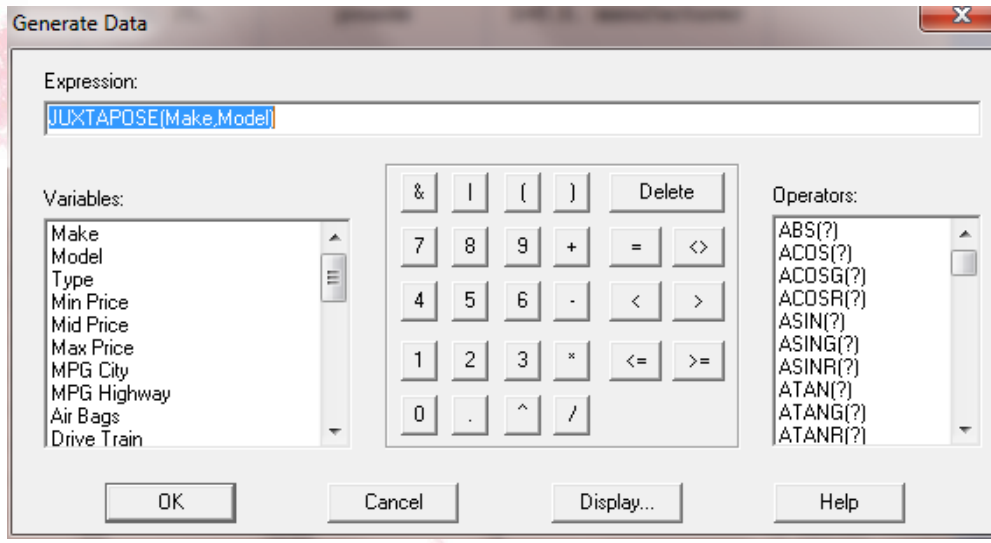
statgraphics® centurion
statgraphics.com

Generate Data



statgraphics® centurion
statgraphics.com

Generating Data

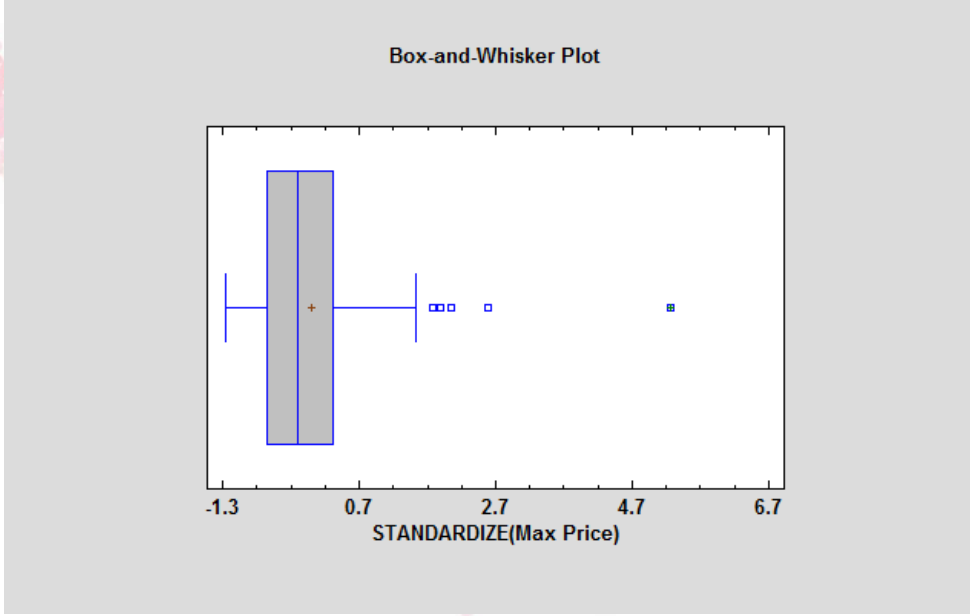


statgraphics centurion
statgraphics.com

Application: Converting data to Z-scores

- Example: I've created a box-and-whisker plot for a sample of *n* observations and wish to determine how many standard deviations the outside points are from their mean.

Box-and-Whisker Plot



statgraphics centurion
statgraphics.com

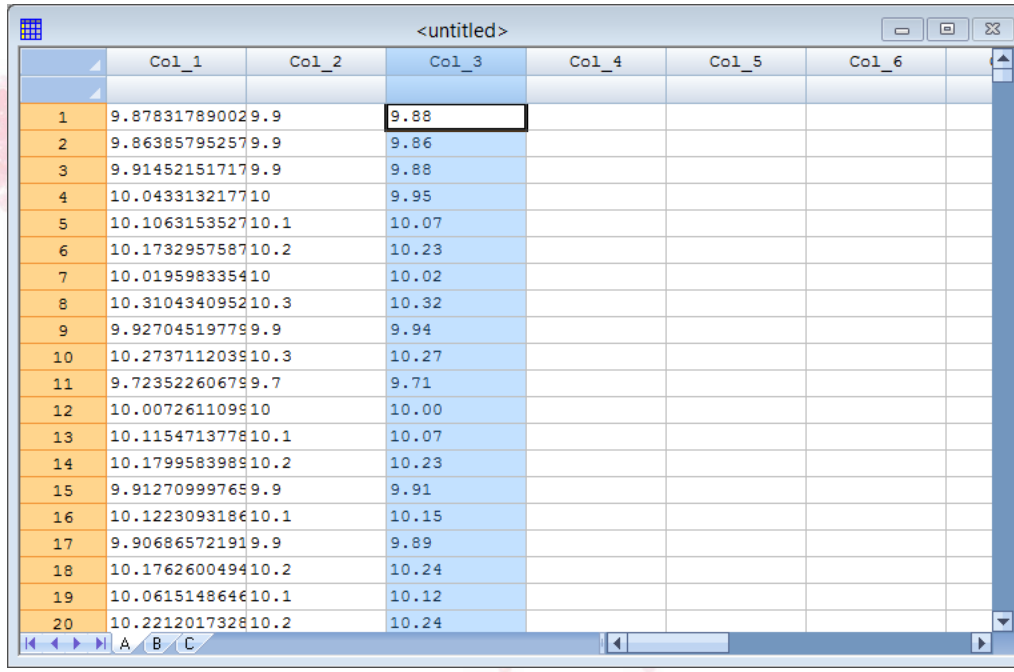
Application: Adding decimal places

- Example: Suppose I have a set of data that I wish to test for normality. It fails the test, but I suspect the failure is due to the fact that the measurements were only recorded to one decimal place. How can I test that suspicion?

- I'll create 3 columns:
  - Col_1: RNORMAL(100,10,0.15)
  - Col_2: ROUNDTO(Col_1,1)
  - Col_3: Col_2+RUNIFORM(100,-.05,.05)

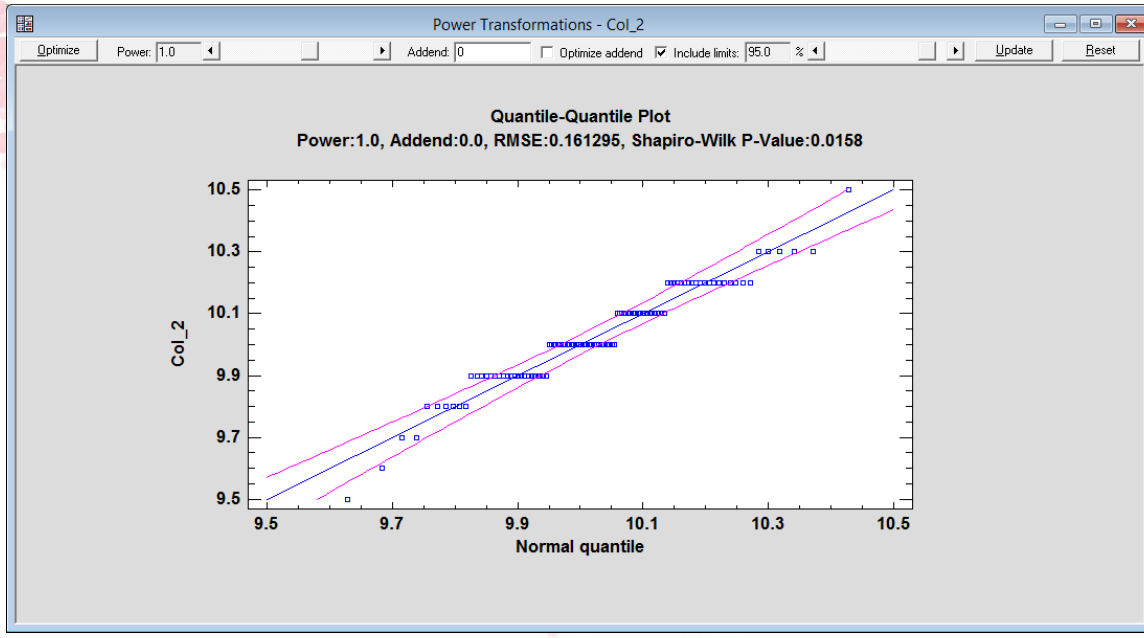
statgraphics® centurion
statgraphics.com

Data



statgraphics® centurion
statgraphics.com

Q-Q Plot



statgraphics® centurion
statgraphics.com

Application: Leading Indicators

- Example: I'd like to forecast the monthly U.S. unemployment rate. I've heard that the number of permits issued for new housing construction is a leading indicator for the economy. Can I use housing permits to improve my forecasts for unemployment?

Housing Permits Data

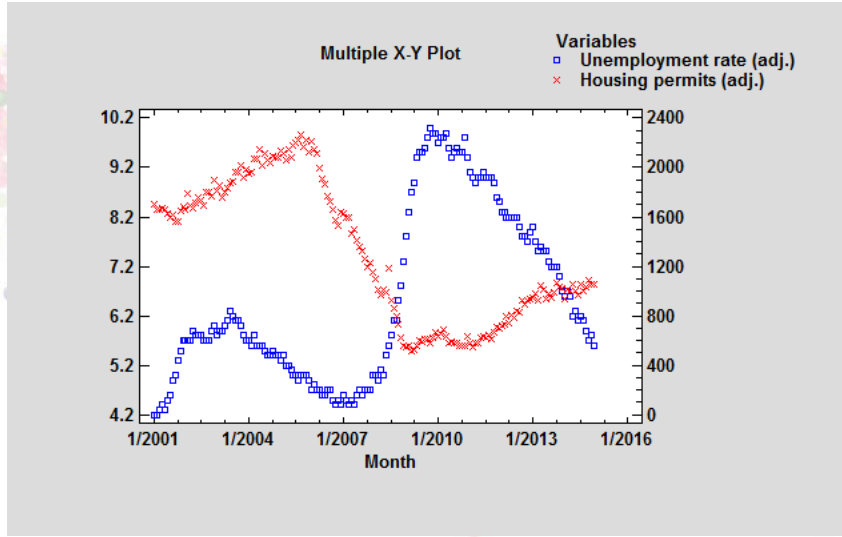
C:\Data\webinar\housing permits.sgd

	Month	Unemployment rate (unadj.)	Unemployment rate (adj.)	Housing permits (adj.)	Col_5	Col_6
1	1/2001	4.7	4.2	1699		
2	2/2001	4.6	4.2	1656		
3	3/2001	4.5	4.3	1659		
4	4/2001	4.2	4.4	1666		
5	5/2001	4.1	4.3	1665		
6	6/2001	4.7	4.5	1626		
7	7/2001	4.7	4.6	1598		
8	8/2001	4.9	4.9	1615		
9	9/2001	4.7	5.0	1565		
10	10/2001	5	5.3	1566		
11	11/2001	5.3	5.5	1651		
12	12/2001	5.4	5.7	1680		
13	1/2002	6.3	5.7	1665		
14	2/2002	6.1	5.7	1787		
15	3/2002	6.1	5.7	1691		
16	4/2002	5.7	5.9	1669		
17	5/2002	5.5	5.8	1716		
18	6/2002	6	5.8	1758		

housing permits | B | C

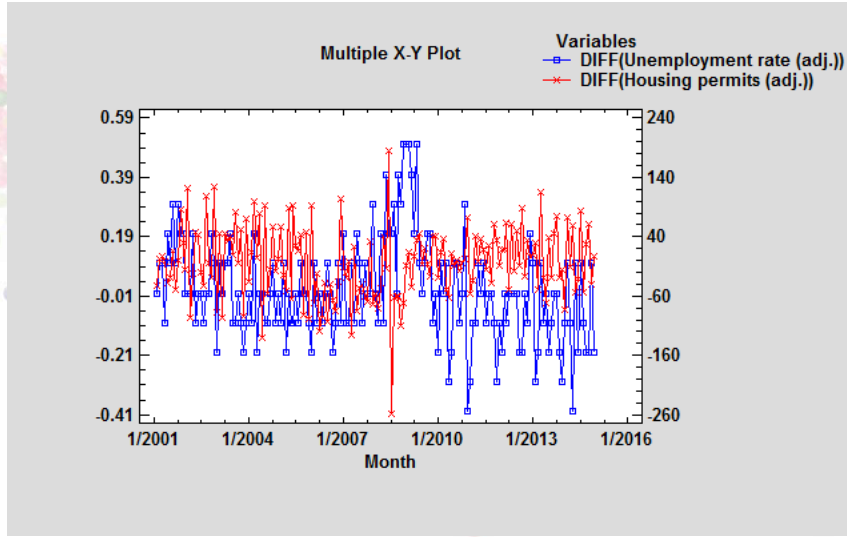
statgraphics® centurion
statgraphics.com

Time Sequence Plot



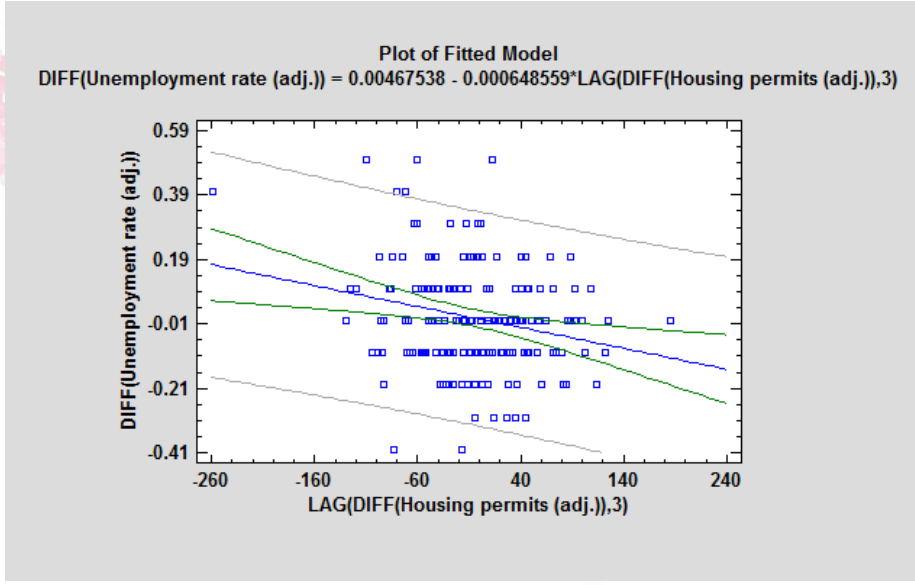
statgraphics®
statgraphics.com centurion

First Differences



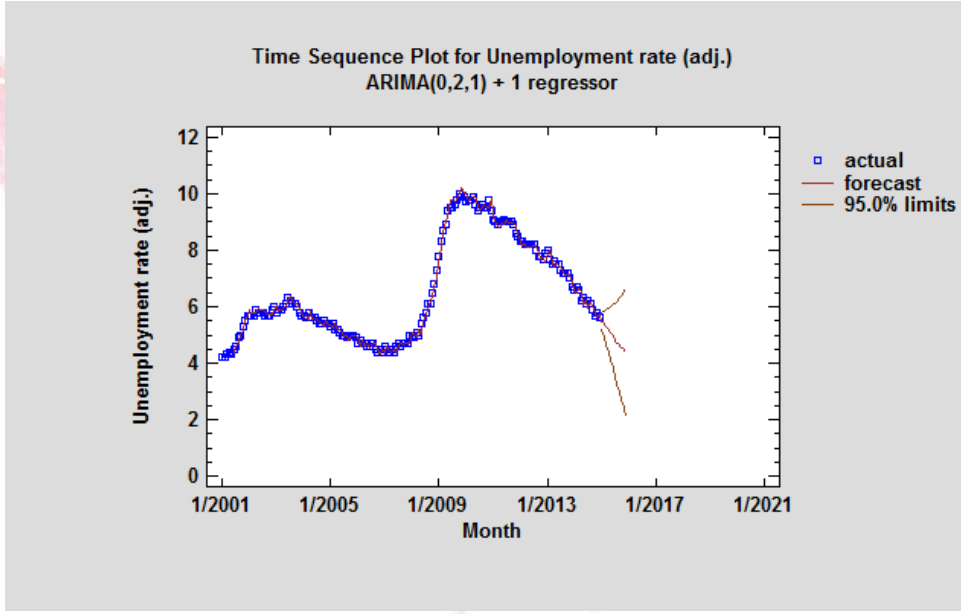
statgraphics®
statgraphics.com centurion

Simple Regression



statgraphics® centurion
statgraphics.com

Forecasted Unemployment



statgraphics centurion
statgraphics.com

Review

EXCLUDE	LOG	REP	RESHAPE	COUNT
ASING	SQRT	COMPRESS	JOIN3	JUXTAPOSE
STANDARDIZE	RNORMAL	ROUNDTO	RUNIFORM	DIFF
LAG	Boolean	Algebraic