

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, ...)

#### 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

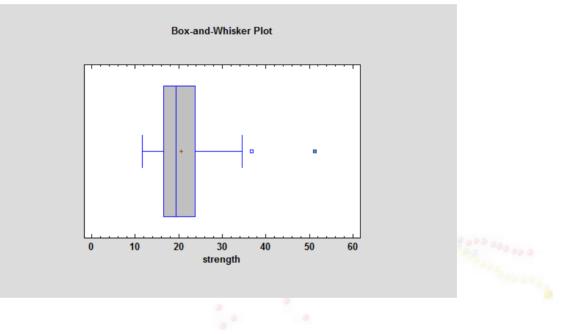




# Application: Transform data to satisfy assumptions of method to be used

- 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.

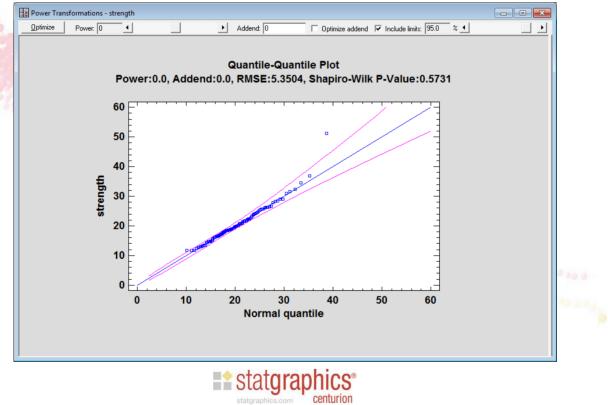
## **Example 1: Tensile Strength**





6

#### **Power Transformations Statlet**



#### **Outlier Identification**

Standar Winsori	ized 2	2.93211 3.03985 2.91209 3.04707					
Sorted V	Values	Studentized Values	Studentized Values	Modified	Г		
Row	Value	Without Deletion	With Deletion	MAD Z-Score	-		
	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		E	
 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			
	Test (assu	mes normality) 96					



#### 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: <u>Statistics for Experimenters</u> by Box, Hunter and Hunter

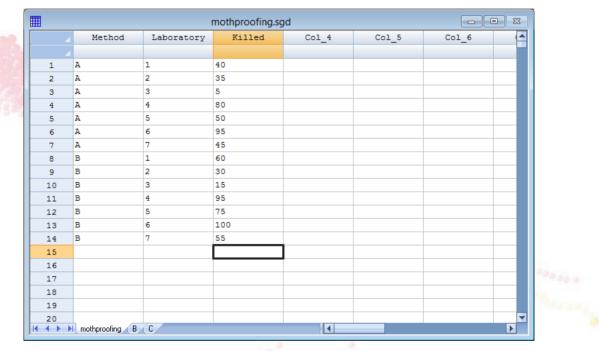


#### Data Setup

Generate Data		×		
KEP(Method,7)           Variables:         & 1         (         )           Method         7         8         9         +           Laboratory         4         5         6         -           1         2         3         *	Delete         Operators:           =         <	Generate Data	×	
0 . ^ /	Expression: RESHAPE(COUNT(1.7,1),14) Variables: Method Laboratory Killed	&       1       (       )       Delete         7       8       9       +       =       <>         4       5       6       -       >>         1       2       3       *       <=	Operators:           ABS(?)           ACDS(?)           ACDS(?)           ACDSR(?)           ASIN(?)           ASIN(?)	10 220 22 0 10 20 20 20 0 10 20 20 20 20 20 20 20 20 20 20 20 20 20
	OK	Cancel Display	Help	
	statorar	hics		

#### Data

1835 C





#### **Multifactor ANOVA**

#### Fisher's variance stabilizing transformation:

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

Method Laboratory Killed	Dependent Variable:	
	Factors:	
	Covariates:	
Sort column names	(Select:)           Delete         Transform         Help	
	Method Laboratory Killed	Laboratory         Killed         ASING(SQRT(Killed/100))         Factors:         Method         Laboratory         Covariates:         Soft column names

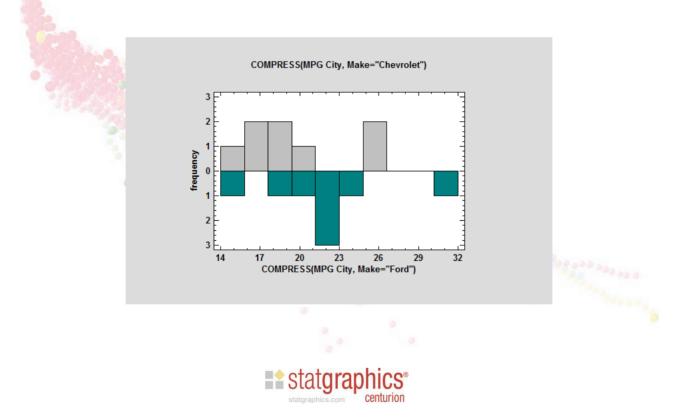
#### **Application: Boolean operators**

- 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?

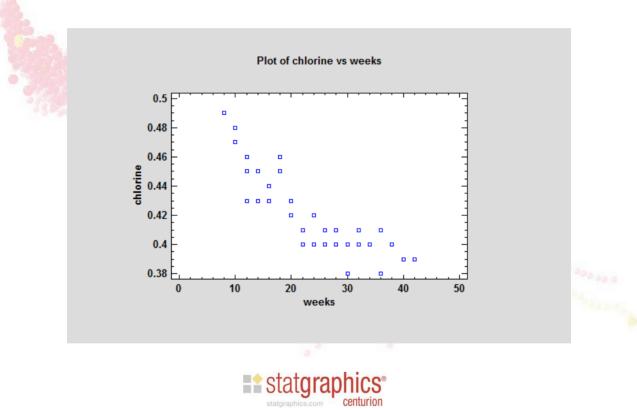
# Two Sample Comparison

Make Model Type Min Price Mid Price Max Price	Sample 1:	
MPG City MPG Highway Air Bags Drive Train Cylinders Engine Size Horsepower RPM Bow per Milo	Sample 2: COMPRESS(MPG City, Make="Ford") (Select:)	
Sort column names	Input     Two Data Columns     Data and Code Columns	
OK Cancel	Delete Transform Help	





## **Stability Study**



2/13/2015

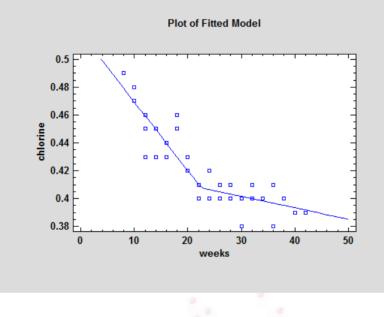
#### **Nonlinear Regression Model**

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

	Nonlinear Regression		
	weeks	Dependent Variable: Chlorine  Function:      A+b*weeks+c*(weeks-d)*(weeks>=d)	
· · · · · ·		(Weights:)	
	OK Cancel	Delete Transform Help	and a second second
	• 5	stat <b>oraphics</b> ®	

statgraphics.com Centurion

#### **Piecewise Linear Fit**



statgraphics®

#### **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**

Iodify Column		×	
Name:		OK	
horses per pound		Cancel	
Comment:		Define	
		Help	
		Value Labels	
O Numeric	C Date		
C Character	O Month		
C Integer	C Quarter		Cost a s
C Time (HH:MM)	🔿 Date-Time (HH:	MM)	
C Time (HH:MM:SS)	O Date-Time (HH:	MM:SS)	10 00 000
C Fixed Decimal: 2	C Percentage		1010 00 00 00 00 00 00 00 00 00 00 00 00
Formula			5
horsepower/weight			13.0.00
	horses per pound         Comment:         Oumeric         Character         Integer         Time (HH:MM)         Time (HH:MM:SS)         Fixed Decimal:         Formula	Improve provided         Comment:         Type         O Numeric       Date         O Character       Month         O Integer       Quarter         O Time (HH:MM)       Date-Time (HH:         O Time (HH:MM:SS)       Date-Time (HH:         O Fixed Decimal:       2       Percentage         Image: Promula       Image: Promula       Image: Promula	Impose per pound     Cancel       Comment:     Define       Help       Value Labels       Type       Numeric     Date       Character     Month       Integer     Quarter       Time (HH:MM)     Date-Time (HH:MM)       Time (HH:MM:SS)     Date-Time (HH:MM:SS)       Fixed Decimal:     2       Formula



#### **Generate Data**



## **Generating Data**



Generate Data Expression: JUXTAPOSE(Make,Model)			
Variables: Make Model Type Min Price Mid Price Max Price MPG City MPG Highway Air Bags Drive Train	&       I       [       ]       Delete         7       8       9       +       =       <>         4       5       6       -       <>       >         1       2       3       ×       <=	Operators: ABS(?) ACOS(?) ACOSG(?) ACOSR(?) ASIN(?) ASING(?) ASINR(?) ATANG(?) ATANG(?) ATANB(?)	<sup>0</sup> 222 23 3
ОК	Cancel Display	Help	-

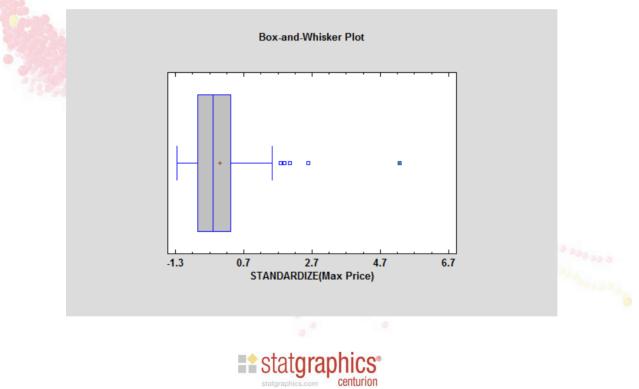
statgraphics® statgraphics.com

#### 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**



#### 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)



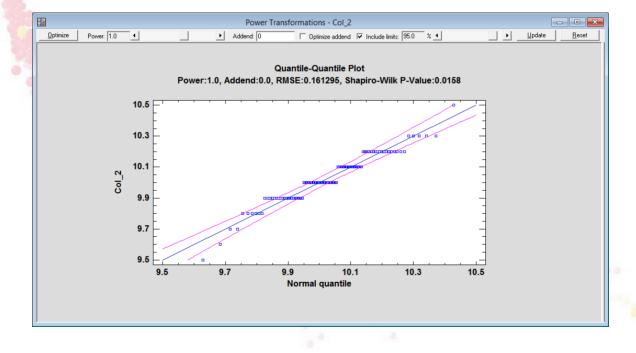
#### Data



			<untiled></untiled>					
	Col_1	Col_2	Col_3	Col_4	Col_5	Col_6		
	4							
1	9.87831789002	9.9	9.88					
2	9.86385795257	9.9	9.86					
3	9.91452151717	9.9	9.88					
4	10.0433132177	10	9.95					
5	10.1063153527	10.1	10.07					
6	10.1732957587	10.2	10.23					
7	10.0195983354	10	10.02					
8	10.3104340952	10.3	10.32					
9	9.92704519779	9.9	9.94					
10	10.2737112039	10.3	10.27					
11	9.72352260679	9.7	9.71					
12	10.0072611099	10	10.00					
13	10.1154713778	10.1	10.07					
14	10.1799583989	10.2	10.23					
15	9.91270999765	9.9	9.91					
16	10.1223093186	10.1	10.15				12.5	
17	9.90686572191	9.9	9.89					
18	10.1762600494	10.2	10.24					
19	10.0615148646	10.1	10.12					
20	10.2212017328	10.2	10.24				<b>T</b>	
<b>∢ ∢ ≻</b> .	A B C							



#### Q-Q Plot



statgraphics®

#### **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?

5/2002

6/2002

17

18

5.5

6

<b>#</b>	C:\Data\webinar\housing permits.sgd								
_	Month Unemployment Unemployment Housing permits Co (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					

**Housing Permits Data** 



1716

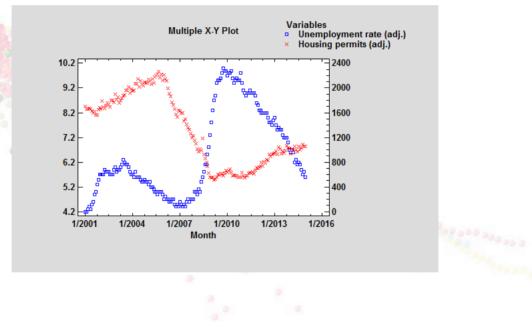
1758

 $\mathbf{F}$ 

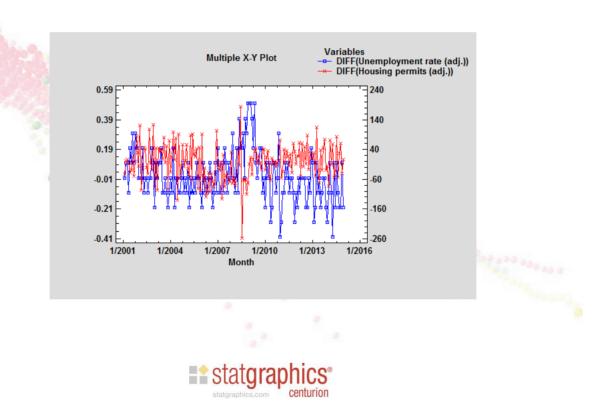
5.8

5.8

#### **Time Sequence Plot**

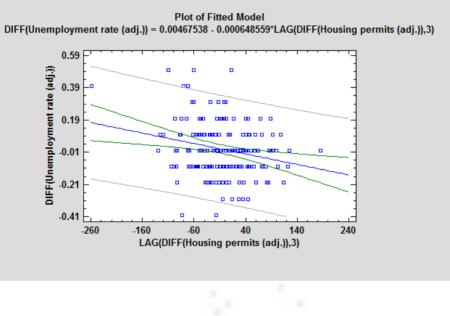


#### **First Differences**

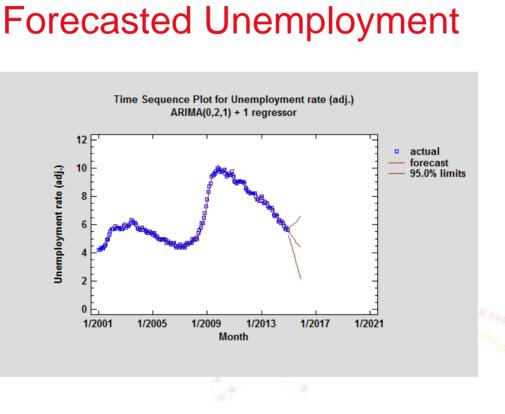


#### Simple Regression









Review

RESHAPE JOIN3 RUNIFORM

COUNT JUXTAPOSE DIFF

EXCLUDE L ASING S STANDARDIZE F LAG E

LOG SQRT RNORMAL Boolean REP COMPRESS ROUNDTO Algebraic