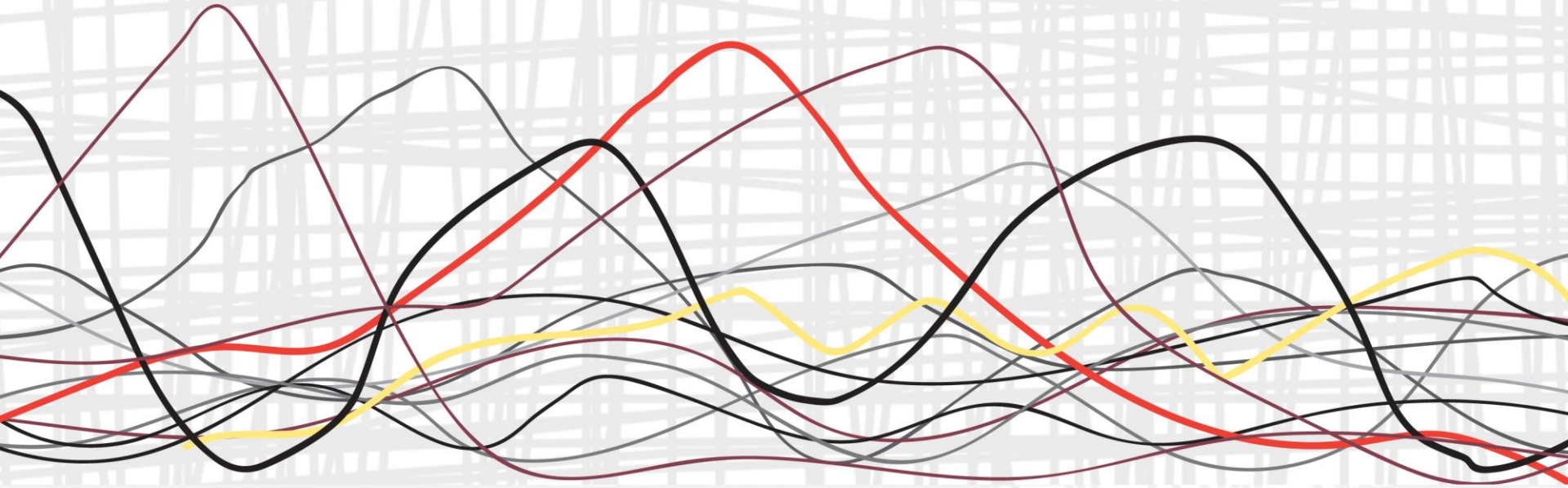
statgraphics 19®

Introduction to Statgraphics 19



Statgraphics 19

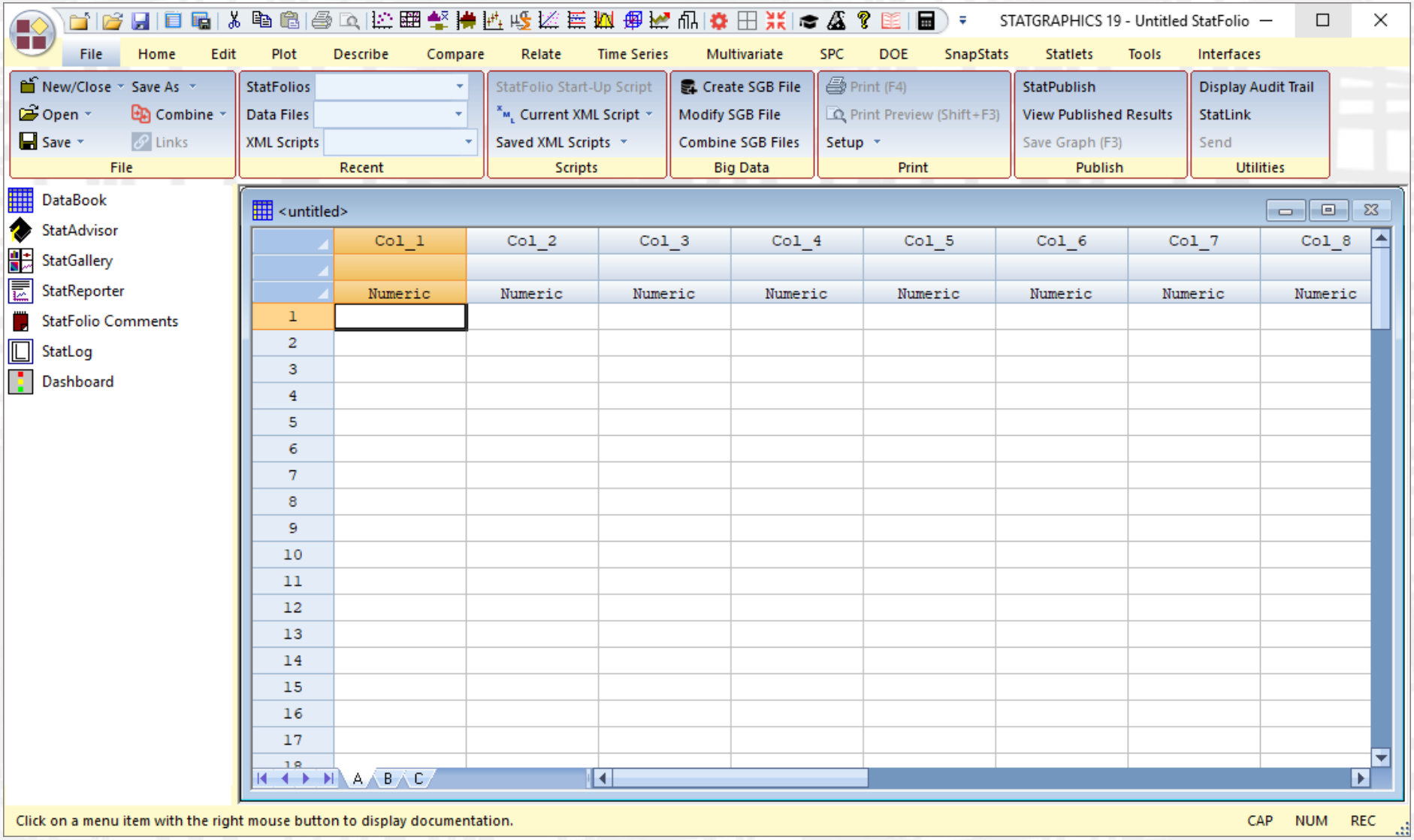
UI changes:

- New ribbon menu
- Relocated quick access toolbar
- Modified tabbed dialog boxes
- Single pane option in analysis windows
- Dashboard for displaying tables and graphs with alerts

Procedural enhancements:

- 15 new statistical procedures
- Enhancements to 40 existing procedures
- Modifiable table headers
- Transparent background for graphs
- Saving images to file from StatGallery
- Interface to Python

Ribbon Menu and QAT



Covid-19 Hospitilizations

Automatic Forecasting - hospitalizedCurrently

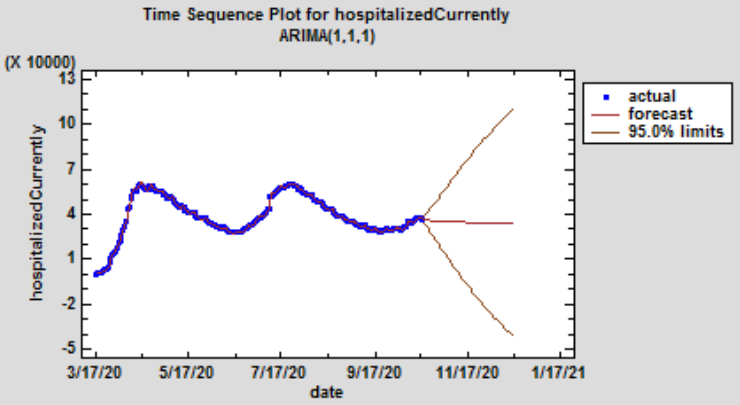
Automatic Forecasting - hospitalizedCurrently

Data variable: hospitalizedCurrently

Number of observations = 216
Time indices: date

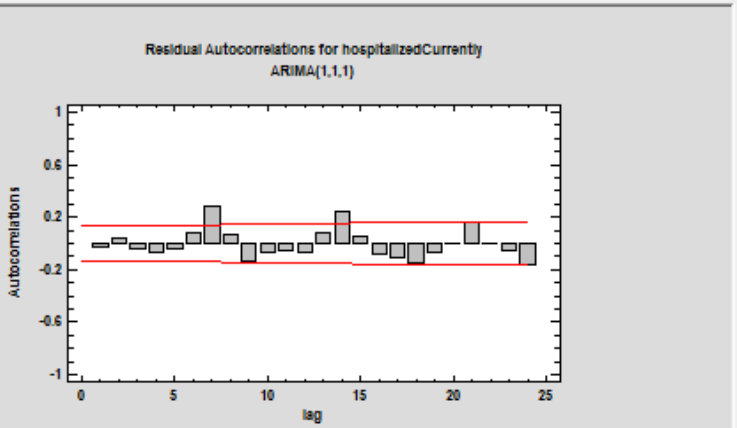
Forecast Summary
Forecast model selected: ARIMA(1,1,1)
Number of forecasts generated: 60
Number of periods withheld for validation: 0

Statistic	*Estimation Period*	*Validation Period*
RMSE	1030.14	
MAE	685.38	
MAPE	2.39008	



Model Comparison
Data variable: hospitalizedCurrently
Number of observations = 216

Models
(A) Random walk
(B) Random walk with drift = 167.921
(C) Constant mean = 38861.2
(D) Linear trend = 37048.0 + 11.09 t
(E) Quadratic trend = -5095.73 + 614.347 t + -1.84482 t^2
(F) Exponential trend = exp(9.86631 + 0.00339805 t)
(G) S-curve trend = exp(11.2645 + -114.64 /t)
(H) Simple moving average of 2 terms
(I) Simple exponential smoothing with alpha = 0.9999
(J) Brown's linear exp. smoothing with alpha = 0.9999
(K) Holt's linear exp. smoothing with alpha = 0.8133 and beta = 0.9244
(L) Brown's quadratic exp. smoothing with alpha = 0.6107



STATGRAPHICS.COM

Analyses – Optional Single Pane Layout

Automatic Forecasting - hospitalizedCurrently

Automatic Forecasting - hospitalizedCurrently

Data variable: hospitalizedCurrently

Number of observations = 216
Time indices: date

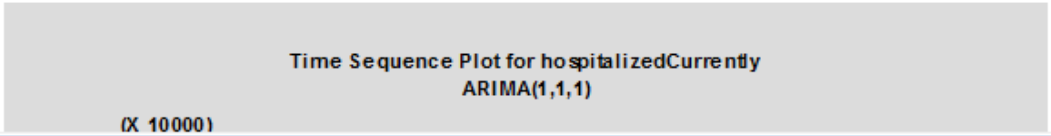
Forecast Summary
Forecast model selected: ARIMA(1,1,1)
Number of forecasts generated: 60
Number of periods withheld for validation: 0

Statistic	*Estimation Period*	*Validation Period*
RMSE	1030.14	
MAE	685.38	
MAPE	2.39008	
ME	23.8128	
MPE	0.744956	

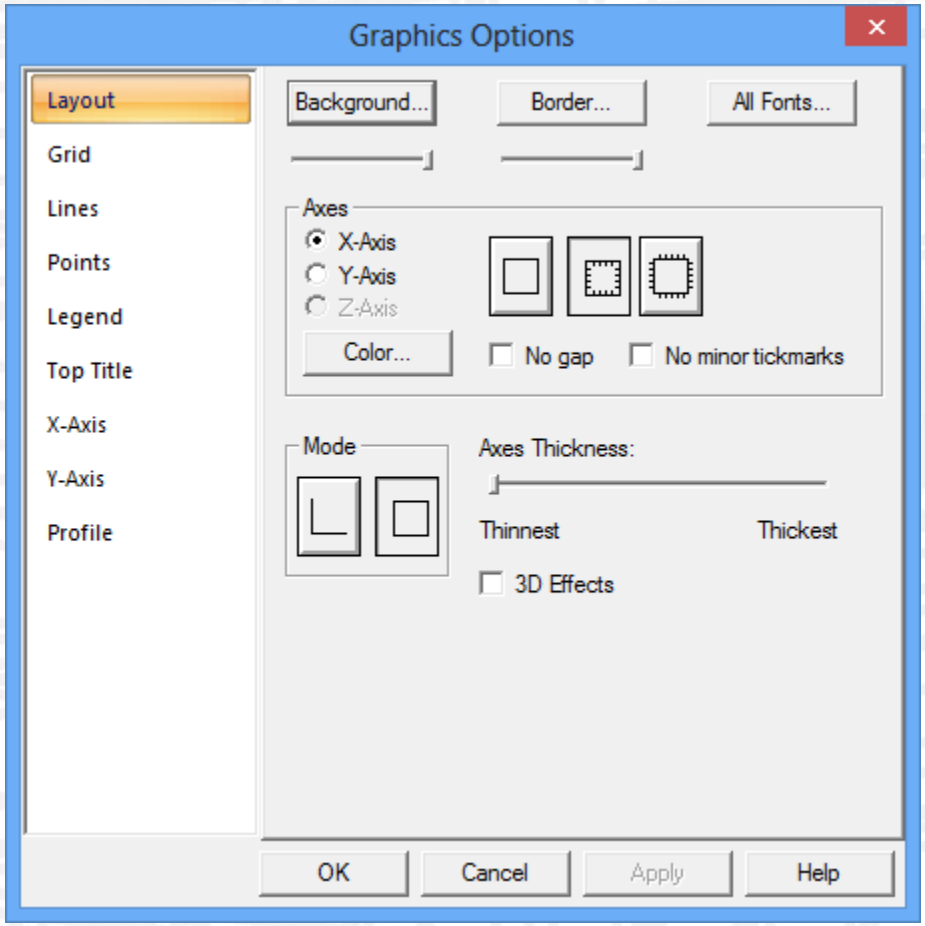
ARIMA Model Summary

Parameter	*Estimate*	*Stnd. Error*	*t*	*P-value*
AR(1)	0.944227	0.0295187	31.9874	0.000000
MA(1)	0.65951	0.0671045	9.82811	0.000000

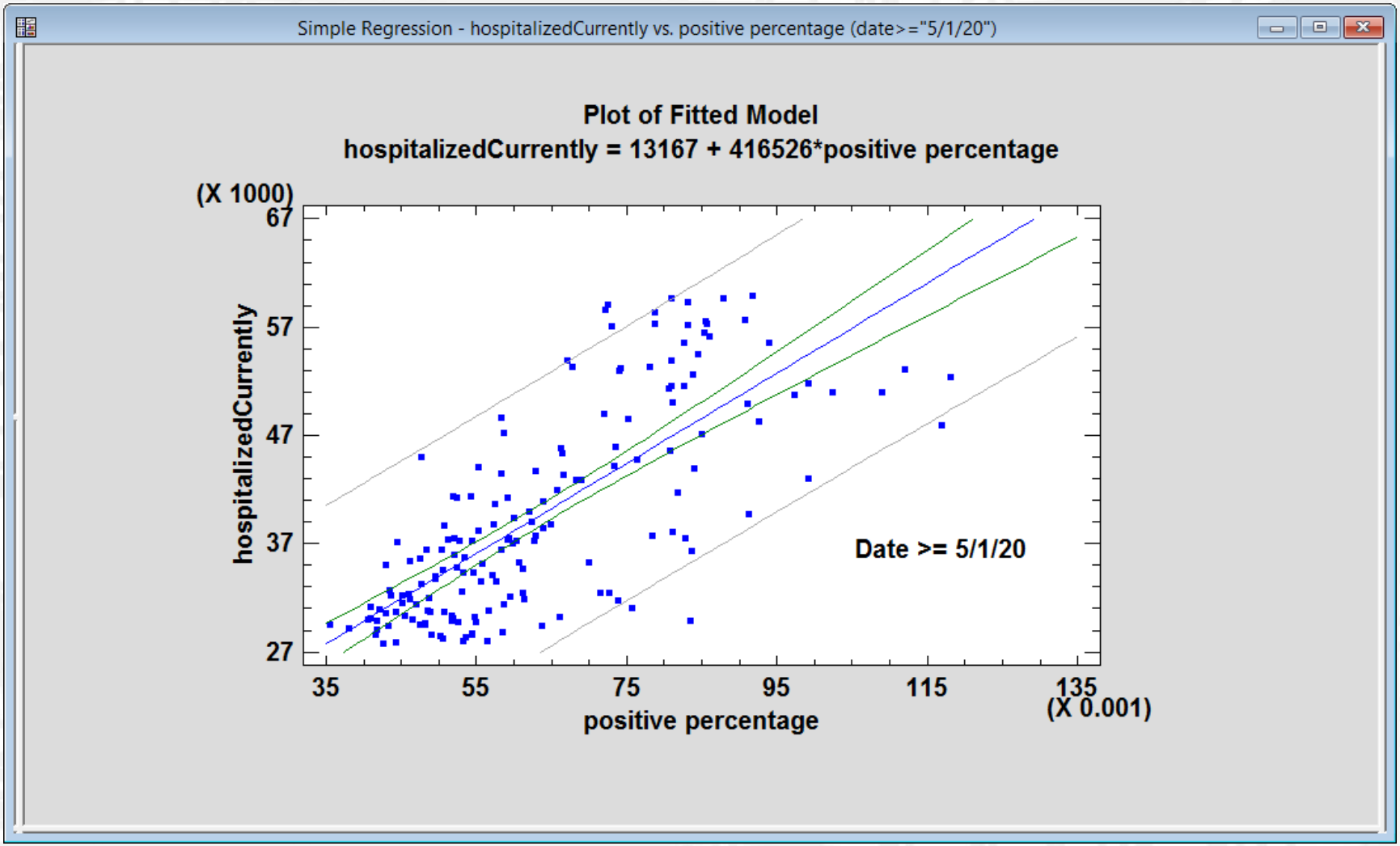
Backforecasting: yes
Estimated white noise variance = 1.06146E6 with 213 degrees of freedom
Estimated white noise standard deviation = 1030.27
Number of iterations: 6



Revised Tabbed Dialog Boxes



Regression



STATGRAPHICS.COM

Dashboard

Dashboard

Page: 1 | Next Page | Prev Page | First Page | Last Page | Help | Hide splitters

Simple Regression - hospitalizedCurrently vs

Dependent variable: hospitalizedCurrently
Independent variable: positive percentage
Selection variable: date>="5/1/20"
Linear model: Y = a + b*X
Number of observations: 171

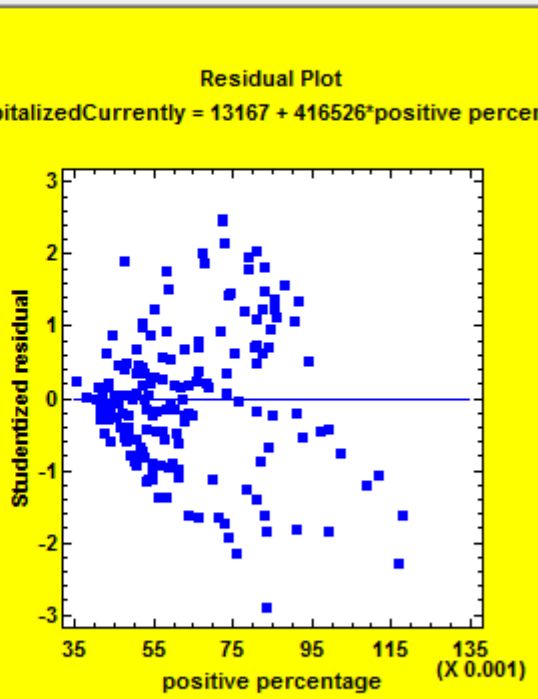
Coefficients

Parameter	Least Squares Estimate	Standard Er
Intercept	13167.0	1878.26
Slope	416526.	28389.1

Analysis of Variance

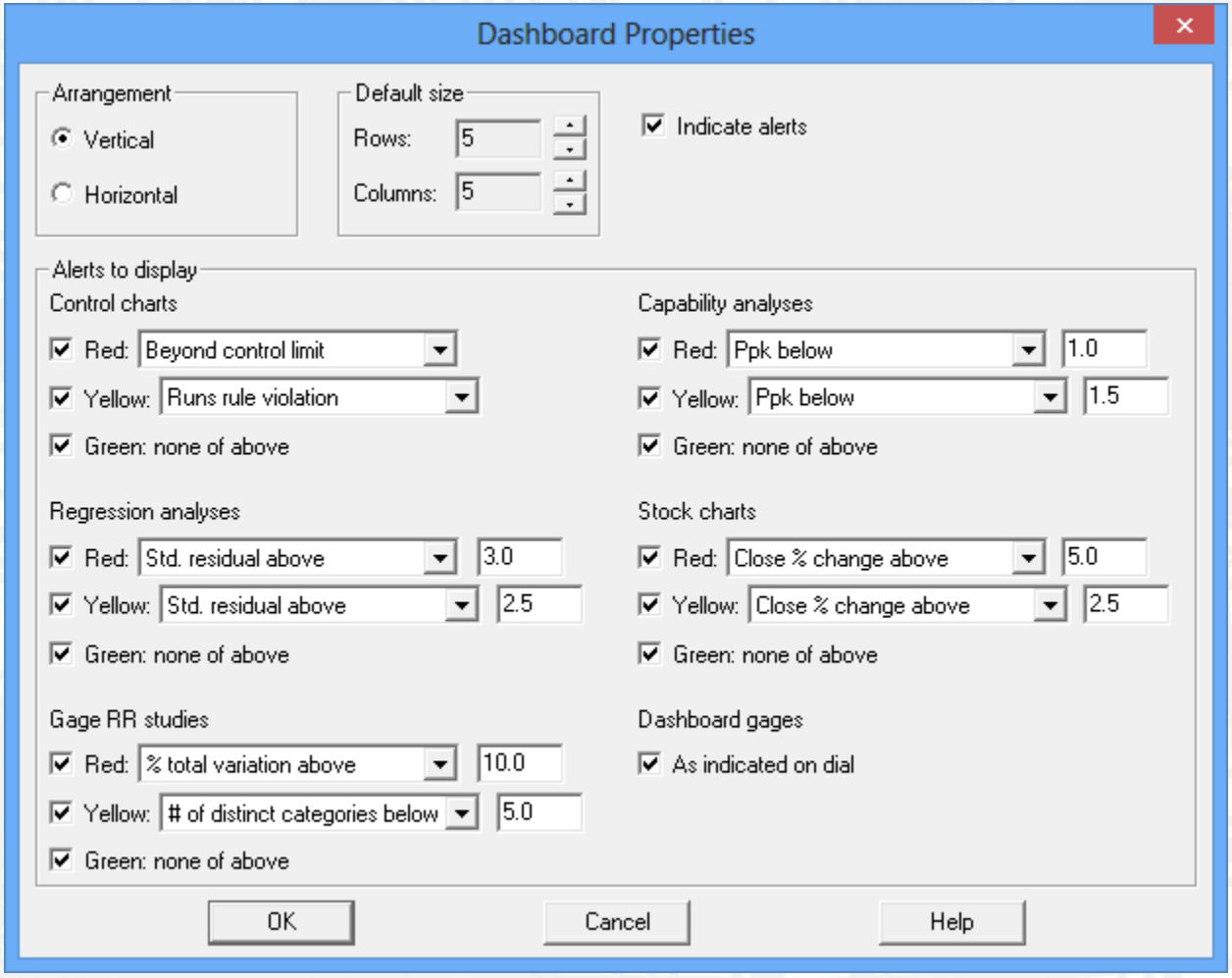
Source	Sum of Squares	Df
Model	8.8027E9	1
Residual	6.9107E9	169
Total (Corr.)	1.57134E10	170

Correlation Coefficient = 0.748467
R-squared = 56.0203 percent
R-squared (adjusted for d.f.) = 55.7601 percent
Standard Error of Est. = 6394.66
Mean absolute error = 4925.77
Durbin-Watson statistic = 0.508004 (P=0.0000)



STATGRAPHICS.COM

Dashboard Properties



Distribution Fitting

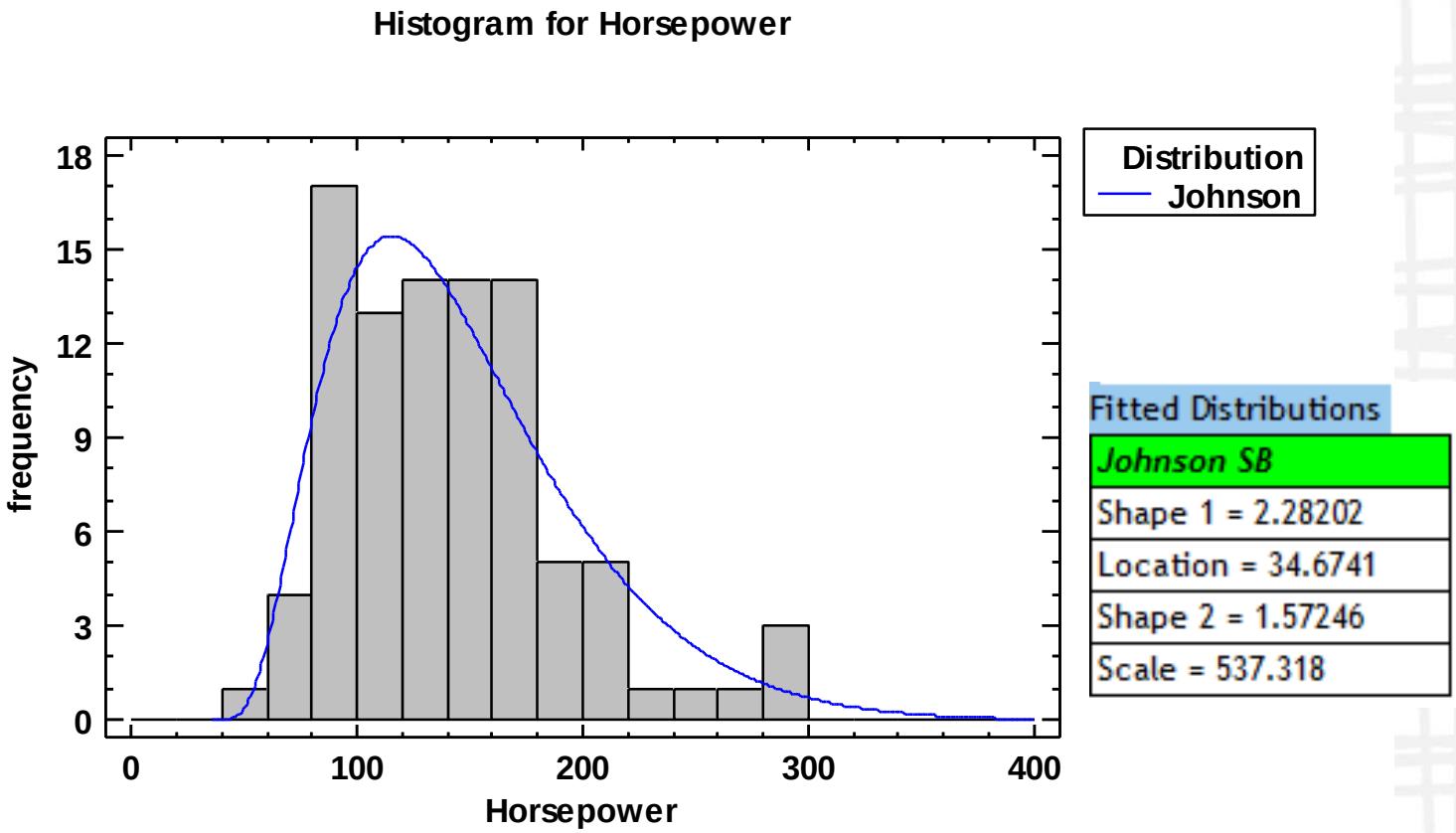
Added several distributions:

- Johnson family
- Zero-inflated Poisson
- Zero-inflated negative binomial
- Mixture of univariate normal distributions
- Mixture of bivariate normal distributions

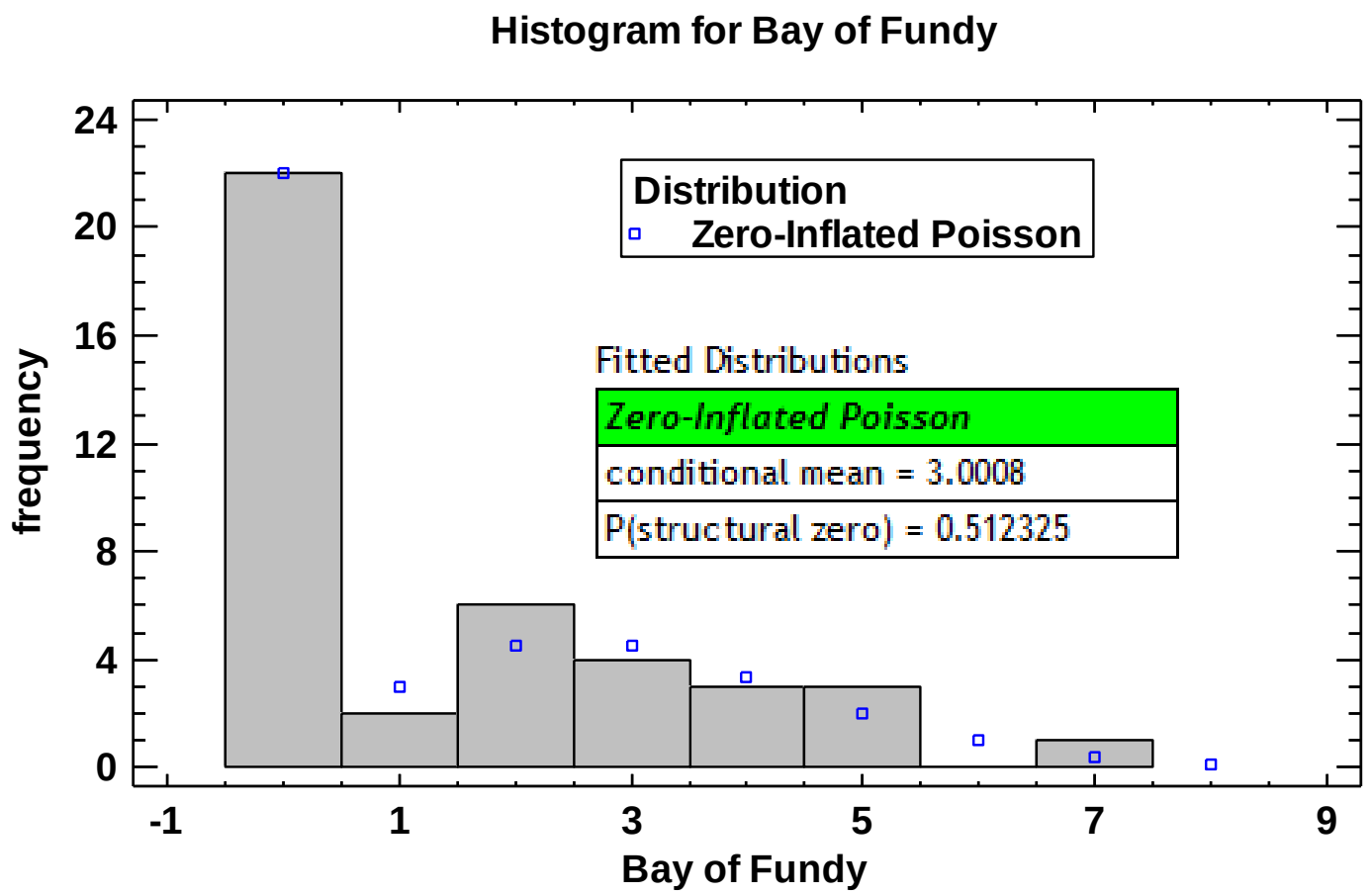
Added ability to deal with data having arbitrary censoring:

- Fits distributions with any combination of left, right, and interval censored data

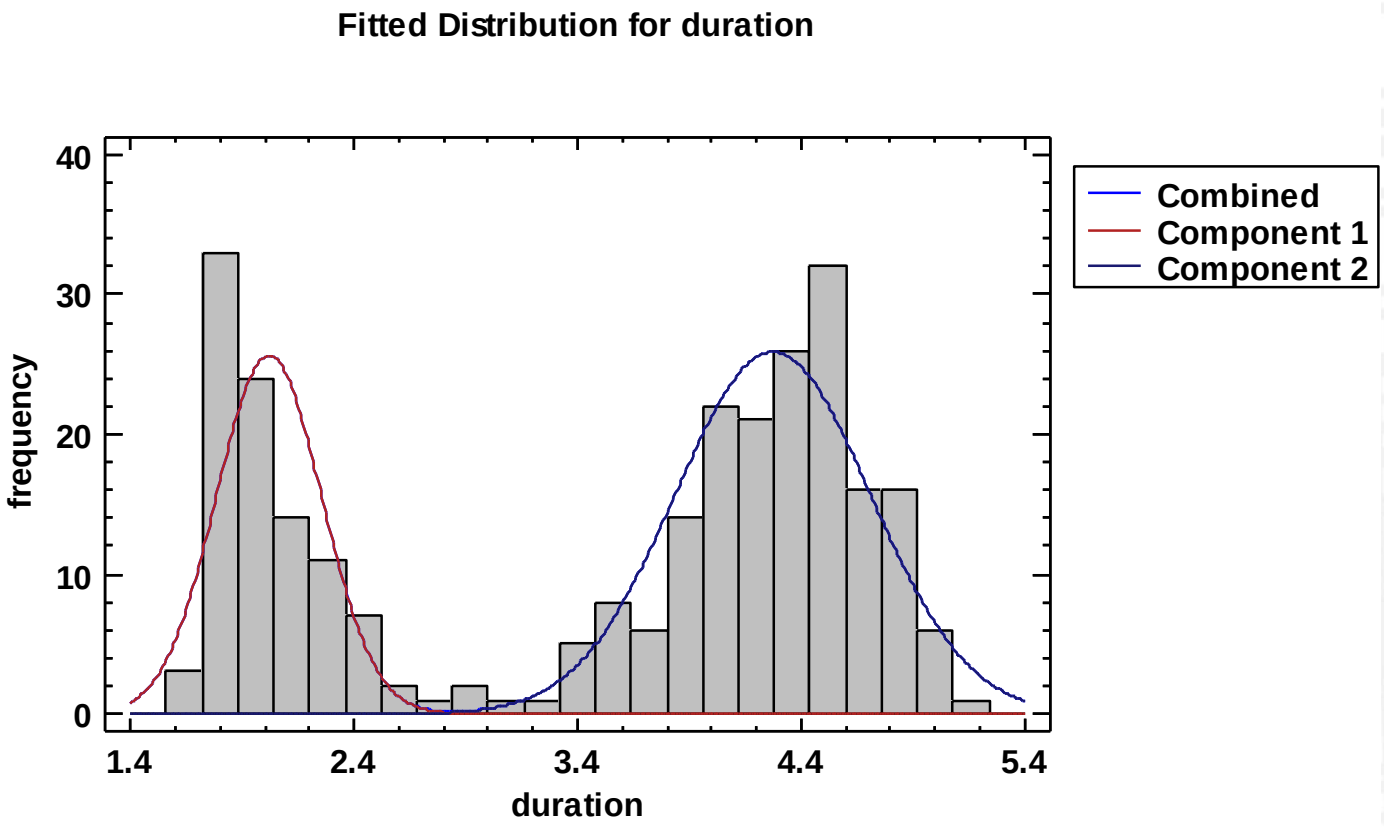
Johnson Distributions



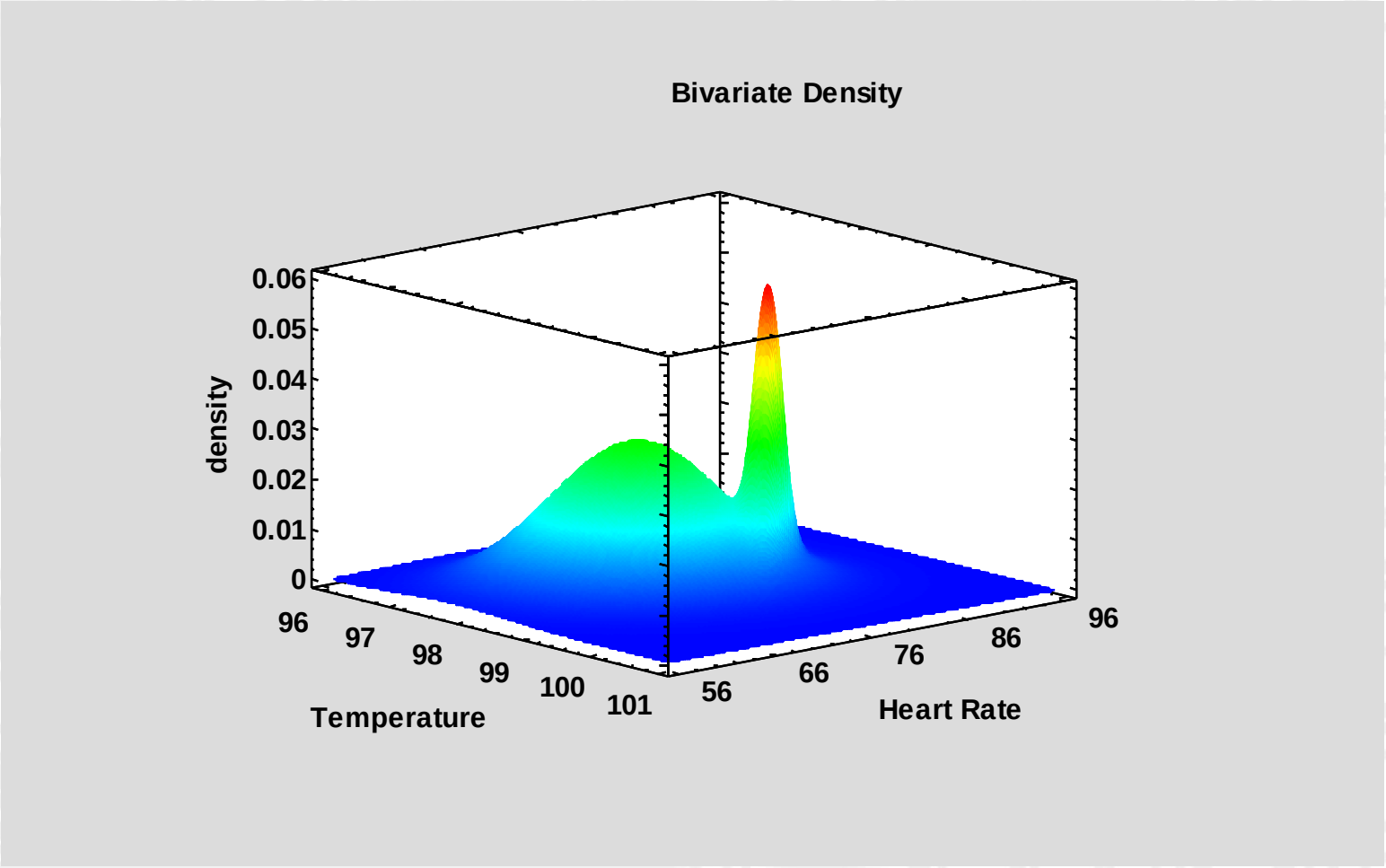
Zero-Inflated Distributions



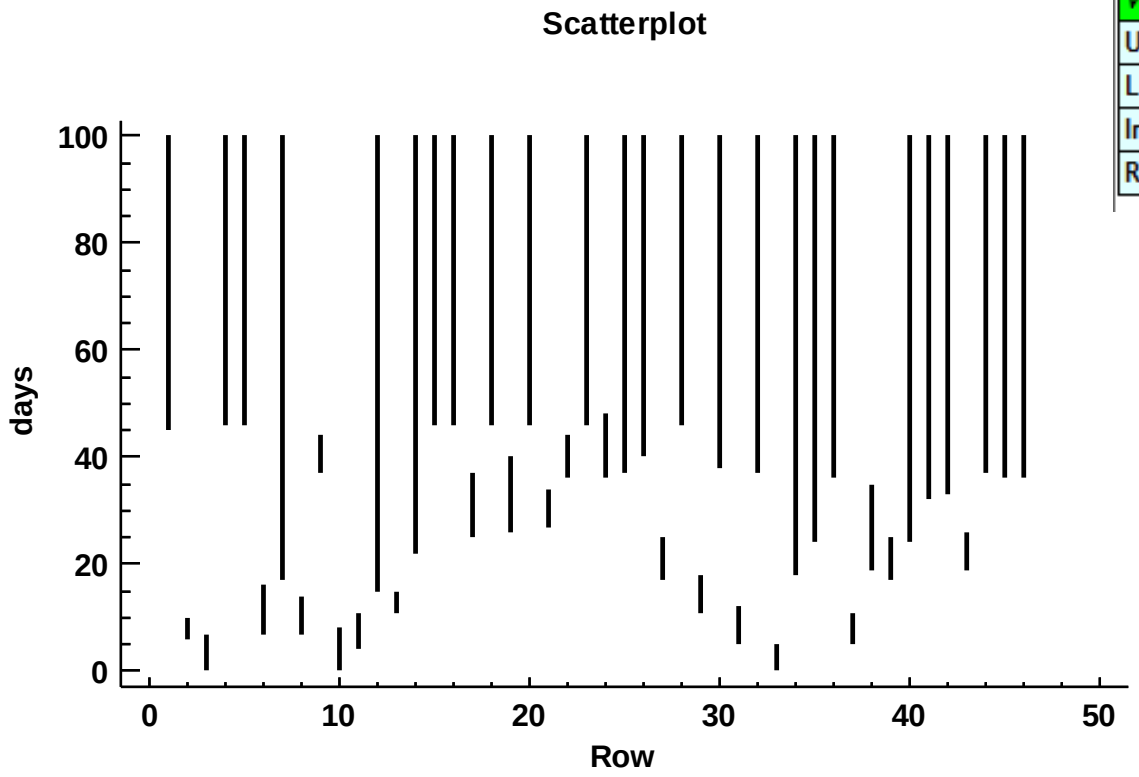
Univariate Mixture Distribution



Bivariate Mixture Distribution



Arbitrarily Censored Data



Observations

Value	*Frequency*
Uncensored	0
Left-censored	3
Interval-censored	18
Right-censored	25

STATGRAPHICS.COM

Regression and ANOVA

New procedures:

- Quantile regression
- Zero-inflated Poisson and negative binomial regression
- Piecewise linear regression
- Stability studies

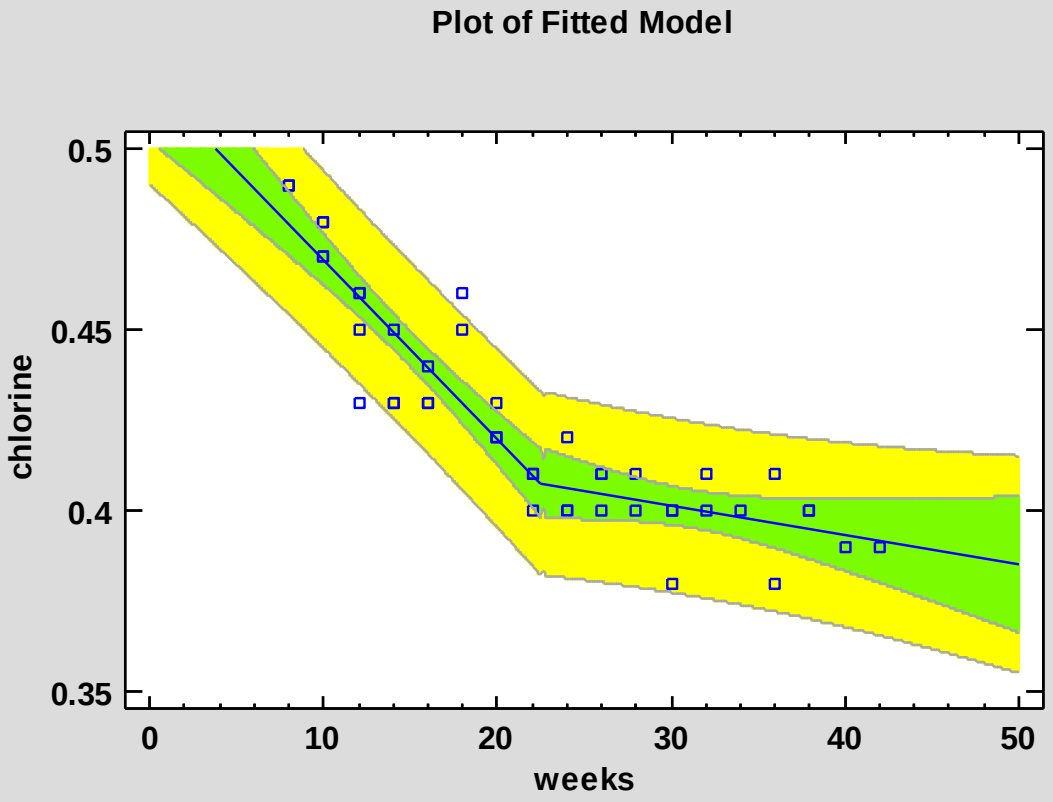
New features in GLM:

- Easier entry of interactions and second-order terms
- Stepwise variable selection with both quantitative and categorical factors

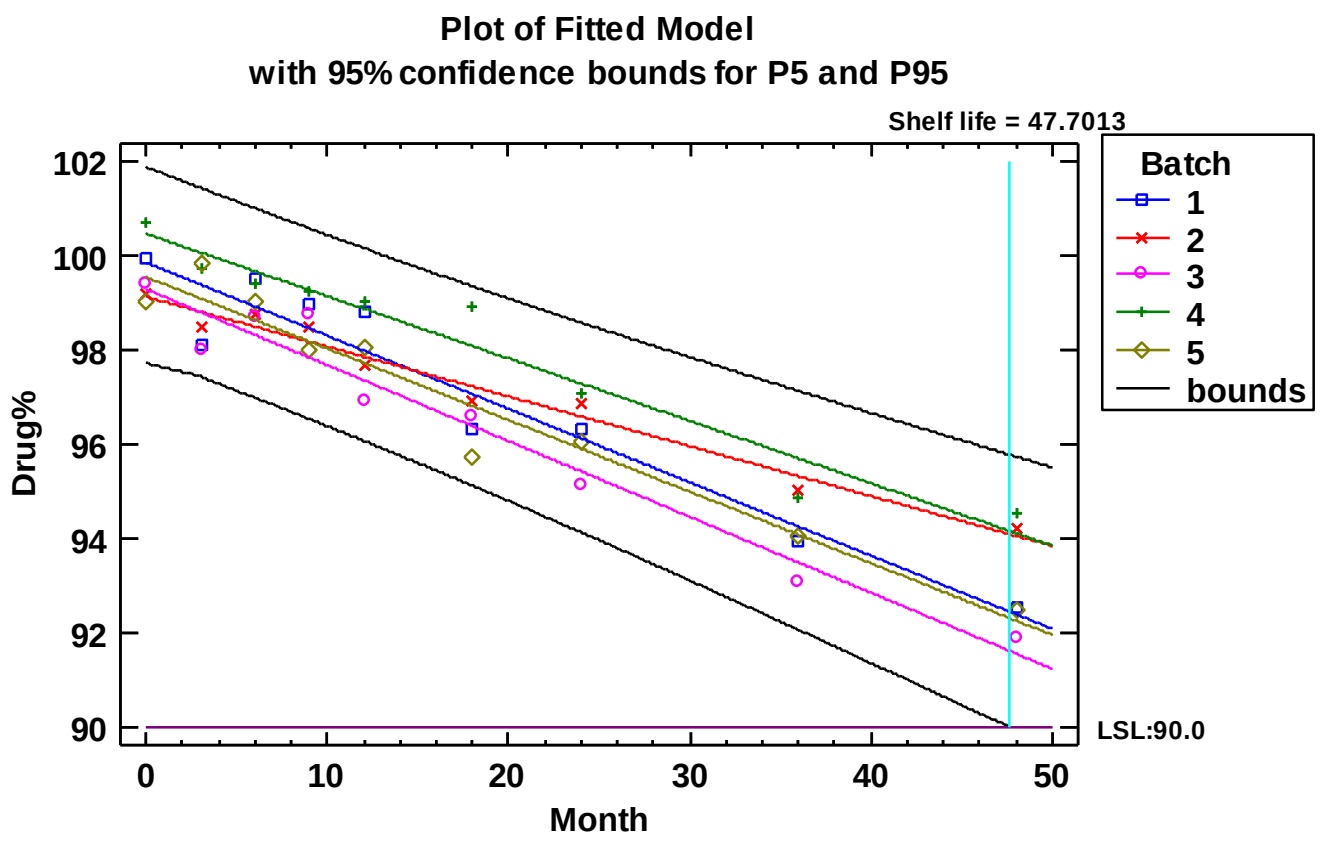
New gage study procedure based on GLM:

- Handles unbalanced data
- Allows for additional factors

Piecewise Linear Regression

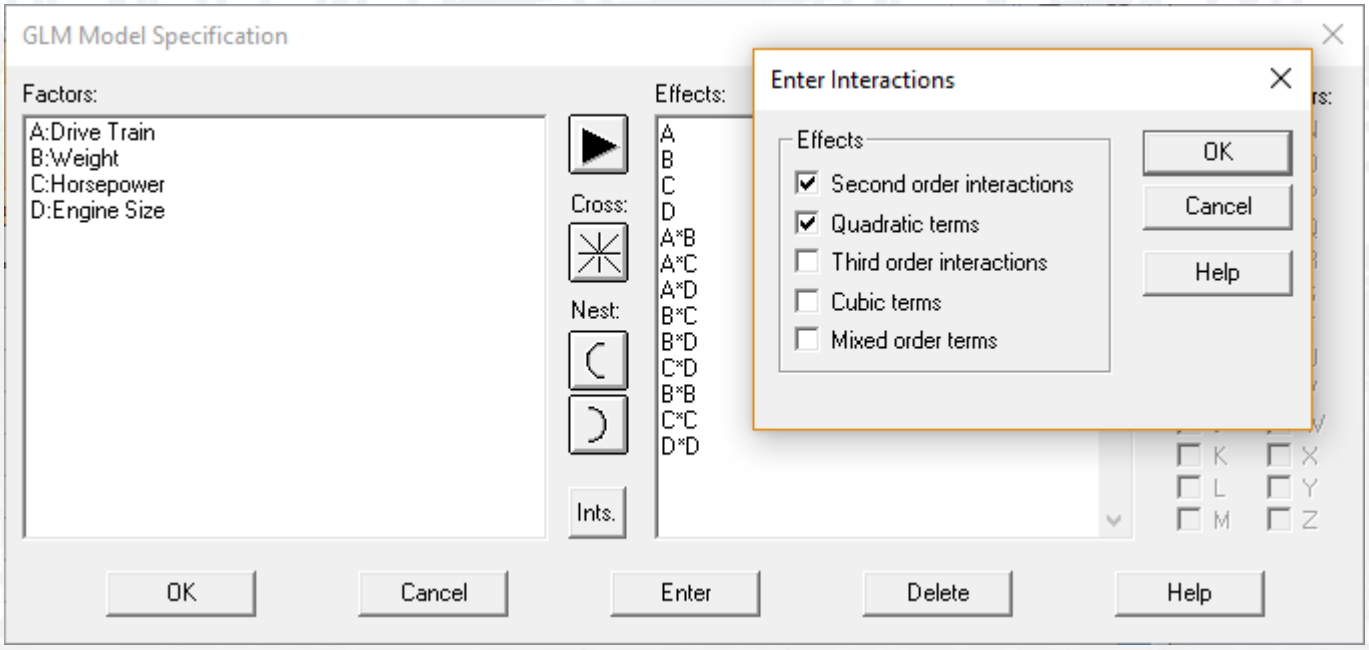


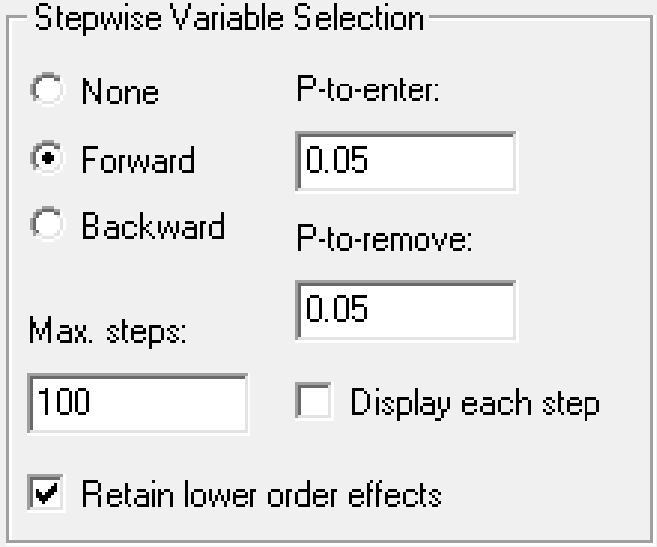
Stability Study



STATGRAPHICS.COM

GLM Changes





STATGRAPHICS.COM

Gage Study (GLM)

Gage R&R - GLM Method - Coating Thickness

Operators: Operator

Parts: Part

Measurements: Coating Thickness

Additional random factors:

Coating

3 operators 10 parts 4 trials

Gage Repeatability and Reproducibility Report

Measurement Unit	*Estimated Variance*	*Percent Contribution*	*Percent of R&R*
Total R&R	24.0274	10.56%	
Repeatability	6.34826	2.79%	26.42%
Reproducibility	17.6791	7.77%	73.58%
Operators	11.4351	5.0235%	47.5919%
Coating	6.24406	2.7431%	25.9872%
Part-to-Part	203.604	89.4446%	
Parts	203.604	89.4446%	
Total Variation	227.632	100%	

Number of distinct categories (ndc): 4

Interface to Python

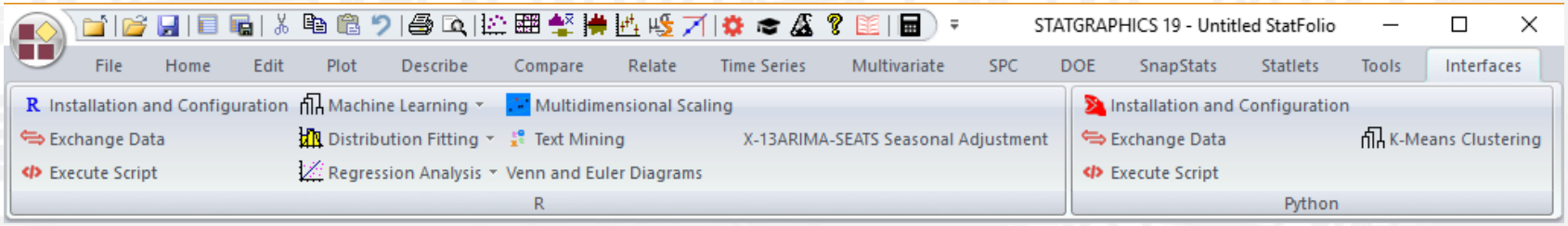
Tools:

- Installation and Configuration
- Data Exchange
- Execution of Scripts

Libraries:

- Access to Procedures in Scikit-Learn Library

Python Interface



C:\Users\neil.STATPOINT\AppData\Local\Programs\Python\Python37\python.exe

```
Python 3.7.0 (v3.7.0:1bf9cc5093, Jun 27 2018, 04:59:51) [MSC v.1914 64 bit (AMD64)] on win32
Type "help", "copyright", "credits" or "license" for more information.
>>> import pandas
>>> import numpy
>>> data = pandas.read_csv(r'C:\Users\NEIL~1.STA\AppData\Local\Temp\statgraphics_data.csv')
>>> data=data.replace(-32768,numpy.NaN)
>>>
>>> data.info
<bound method DataFrame.info of         date  states  positive  ...    positiveIncrease_1  totalTestResultsIncrease
   day
0    20200304      14       118  ...                   NaN                     NaN   3/4/20
1    20200305      24       176  ...                  58.0                   263.0   3/5/20
2    20200306      36       223  ...                  47.0                   665.0   3/6/20
3    20200307      51       341  ...                 118.0                   356.0   3/7/20
4    20200308      51       417  ...                  76.0                   602.0   3/8/20
5    20200309      51       584  ...                 167.0                  1199.0   3/9/20
6    20200310      51       778  ...                 194.0                   634.0  3/10/20
7    20200311      51      1054  ...                 276.0                  2539.0  3/11/20
8    20200312      51      1315  ...                 261.0                  2232.0  3/12/20
9    20200313      51      1922  ...                 607.0                  6179.0  3/13/20
10   20200314      51      2450  ...                 528.0                  4017.0  3/14/20
11   20200315      51      3173  ...                 723.0                  6172.0  3/15/20
12   20200316      56      4019  ...                 846.0                 14399.0  3/16/20
13   20200317      56      5722  ...                1703.0                 13203.0  3/17/20
14   20200318      56      7730  ...                2008.0                 20629.0  3/18/20
15   20200319      56     11719  ...                3989.0                 26917.0  3/19/20
16   20200320      56     17033  ...                5314.0                 34308.0  3/20/20
17   20200321      56     23197  ...                6164.0                 43926.0  3/21/20
18   20200322      56     31879  ...                8682.0                 46236.0  3/22/20
19   20200323      56     42152  ...               10273.0                 54131.0  3/23/20
```

Machine Learning

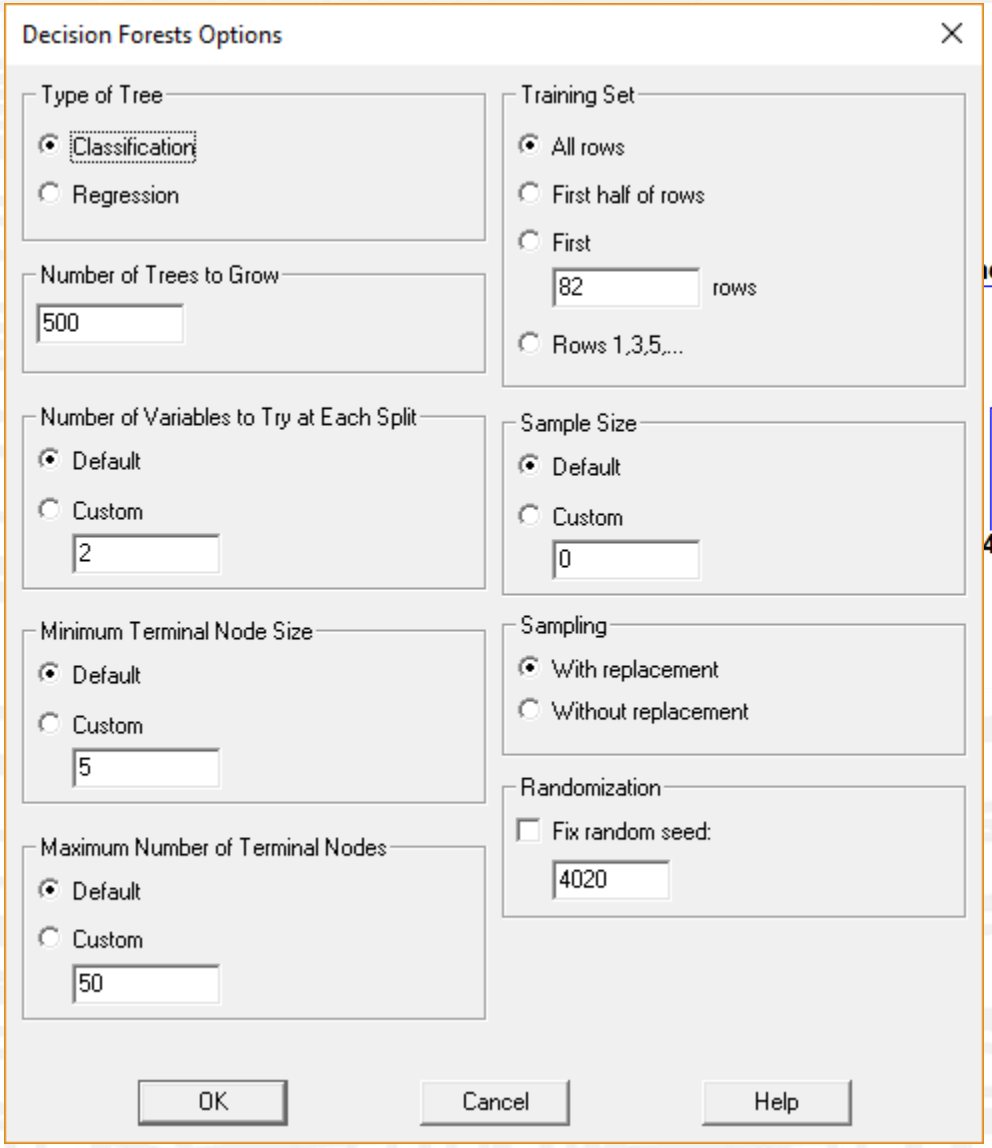
R:

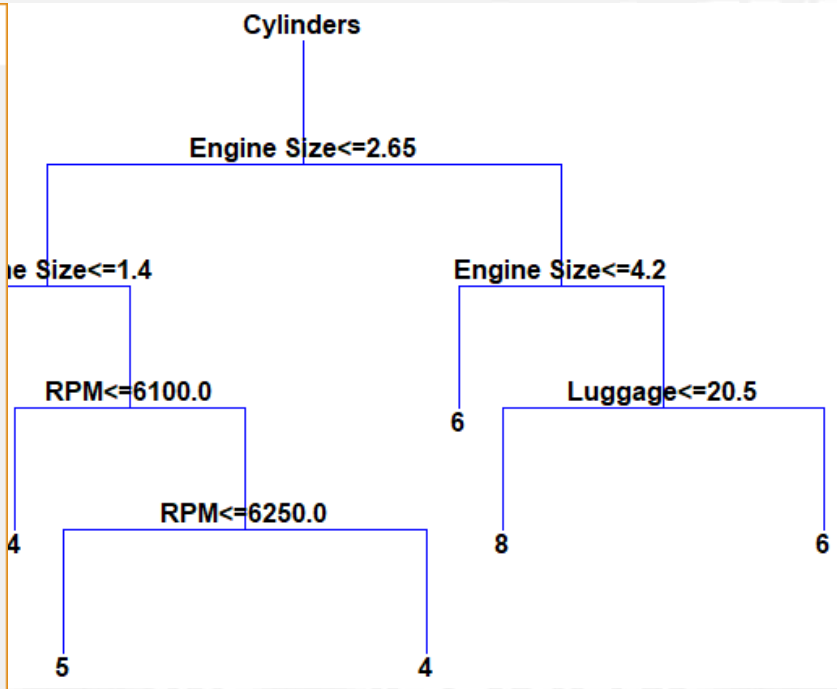
- Classification and regression trees (V18)
- Decision forests

Python:

- K-means clustering
- Support vector machines (V19.2)

Decision Forests





STATGRAPHICS.COM

K-Means Clustering

K-Means Clustering Options

Number of clusters: 3

Standardize variables

Verbose output

Algorithm
- Automatic
- Classical EM-style
- Elkan

Estimation
Number of runs: 10
Maximum iterations: 300
Relative tolerance: 0.0001

Initial seeds
- Smart selection
- Random selection
- From input dialog

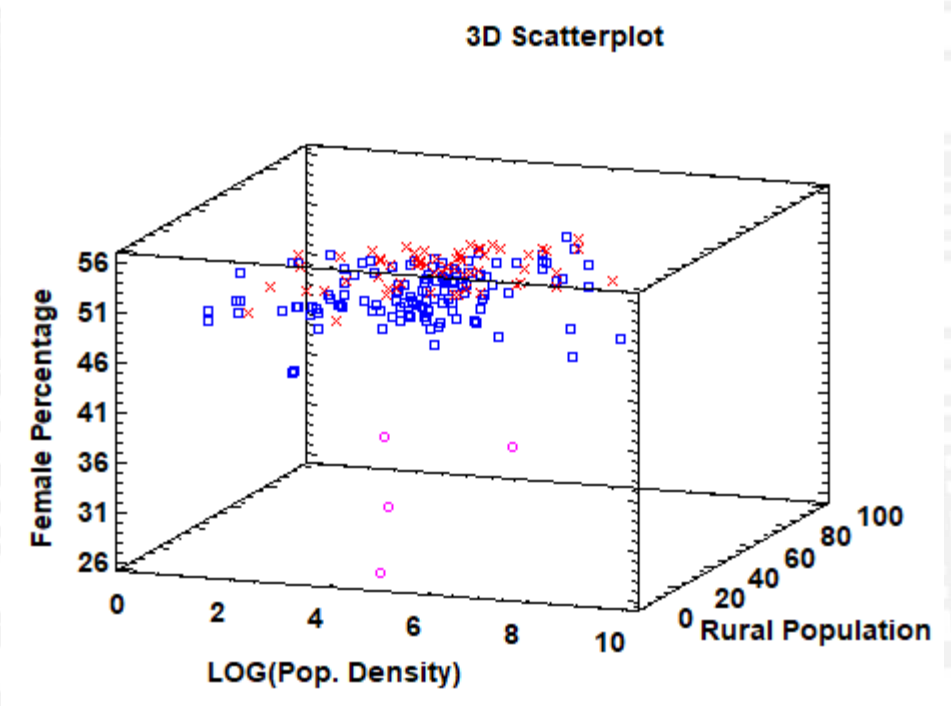
Precompute distances
- Automatic
- Always
- Never

Center the data

Randomization
Fix random seed: 30109

Missing value treatment
- Exclude incomplete cases
- Assign incomplete cases to nearest cluster centroid
- Assign incomplete cases to cluster of nearest neighbor
- Replace missing values with column means
- Replace missing values with column medians
- Replace missing values with most frequent value

OK Cancel Help



STATGRAPHICS.COM

Other Enhancements

Data visualization:

- Venn and Euler diagrams
- Waterfall plots
- Additional line on barcharts

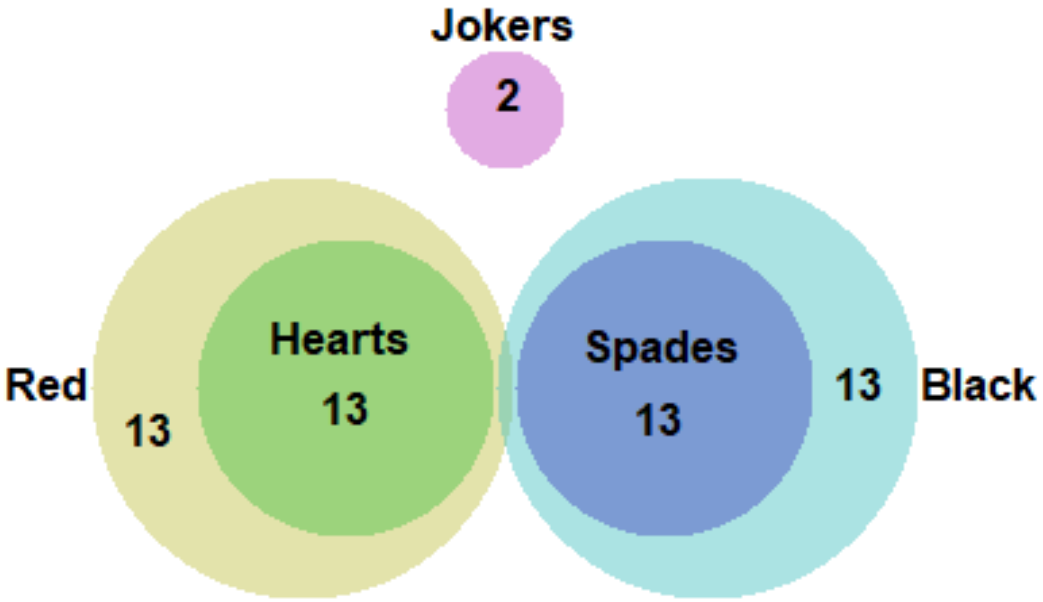
Statistical tests:

- Equivalence tests for standard deviations
- Wald-Wolfowitz test for comparing 2 samples
- Modified Levene's test for homogeneity of variances

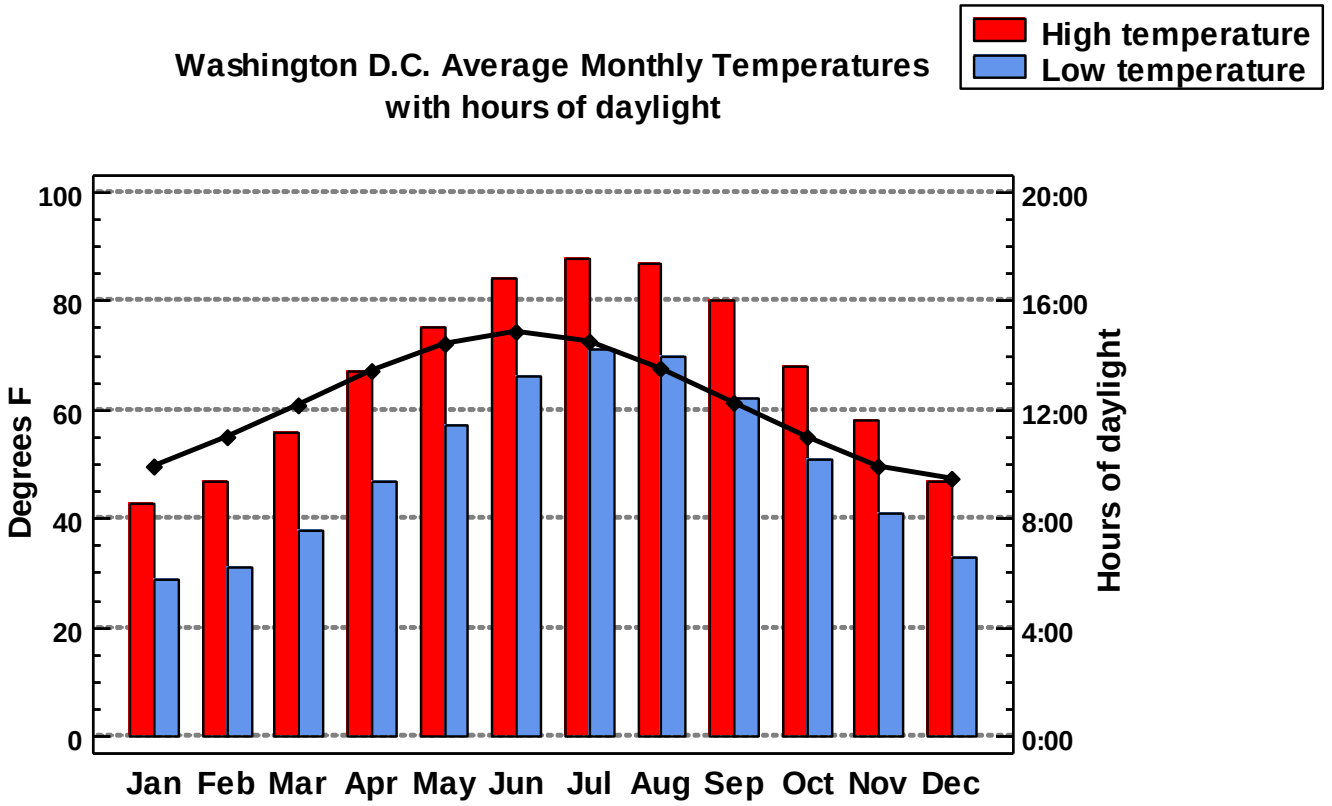
Design of experiments:

- Alias optimal designs
- Augmentation of existing designs with optimal additional runs

Venn and Euler Diagrams

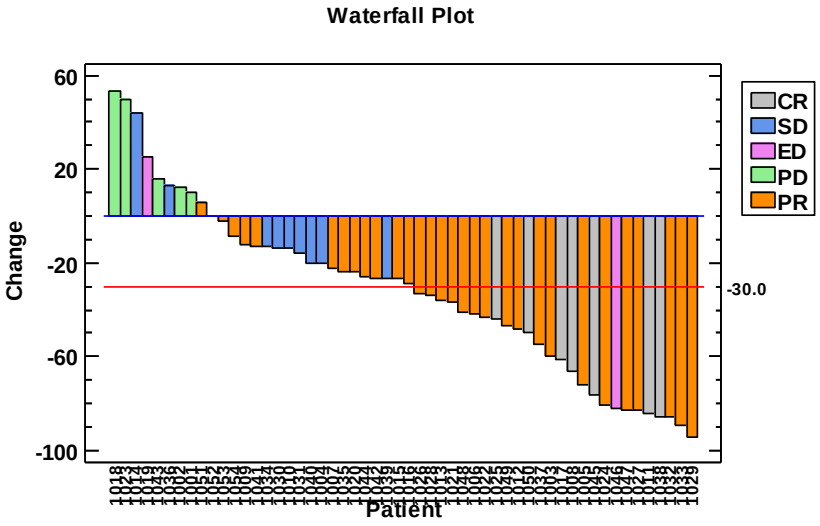


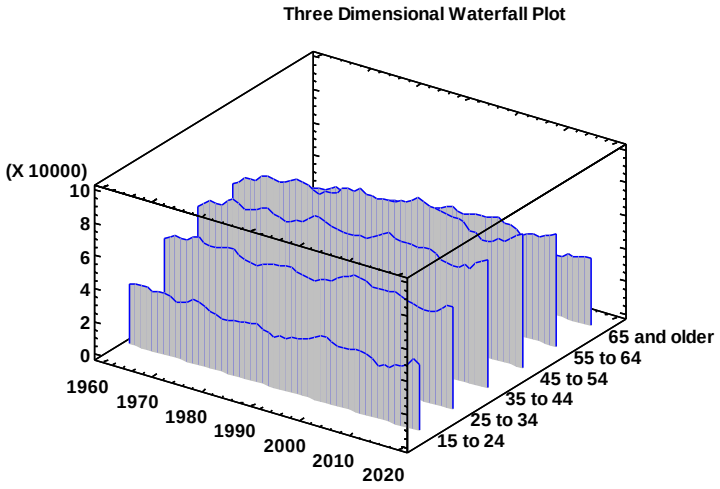
Lines on Barcharts

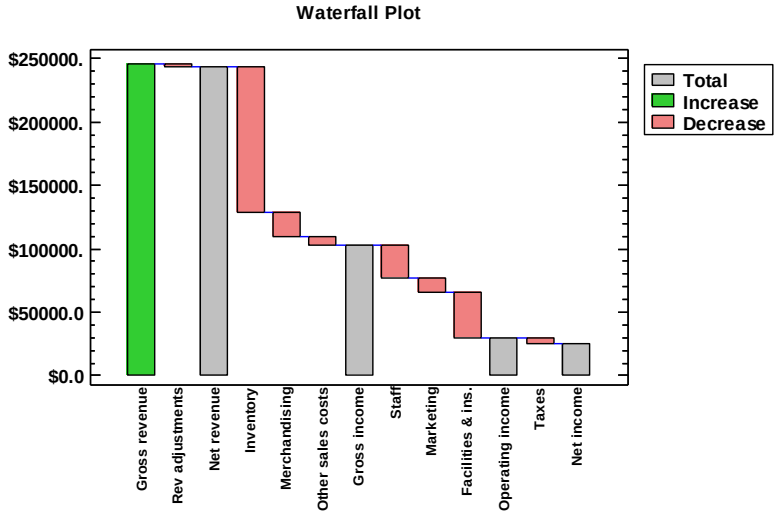


STATGRAPHICS.COM

Waterfall Plots

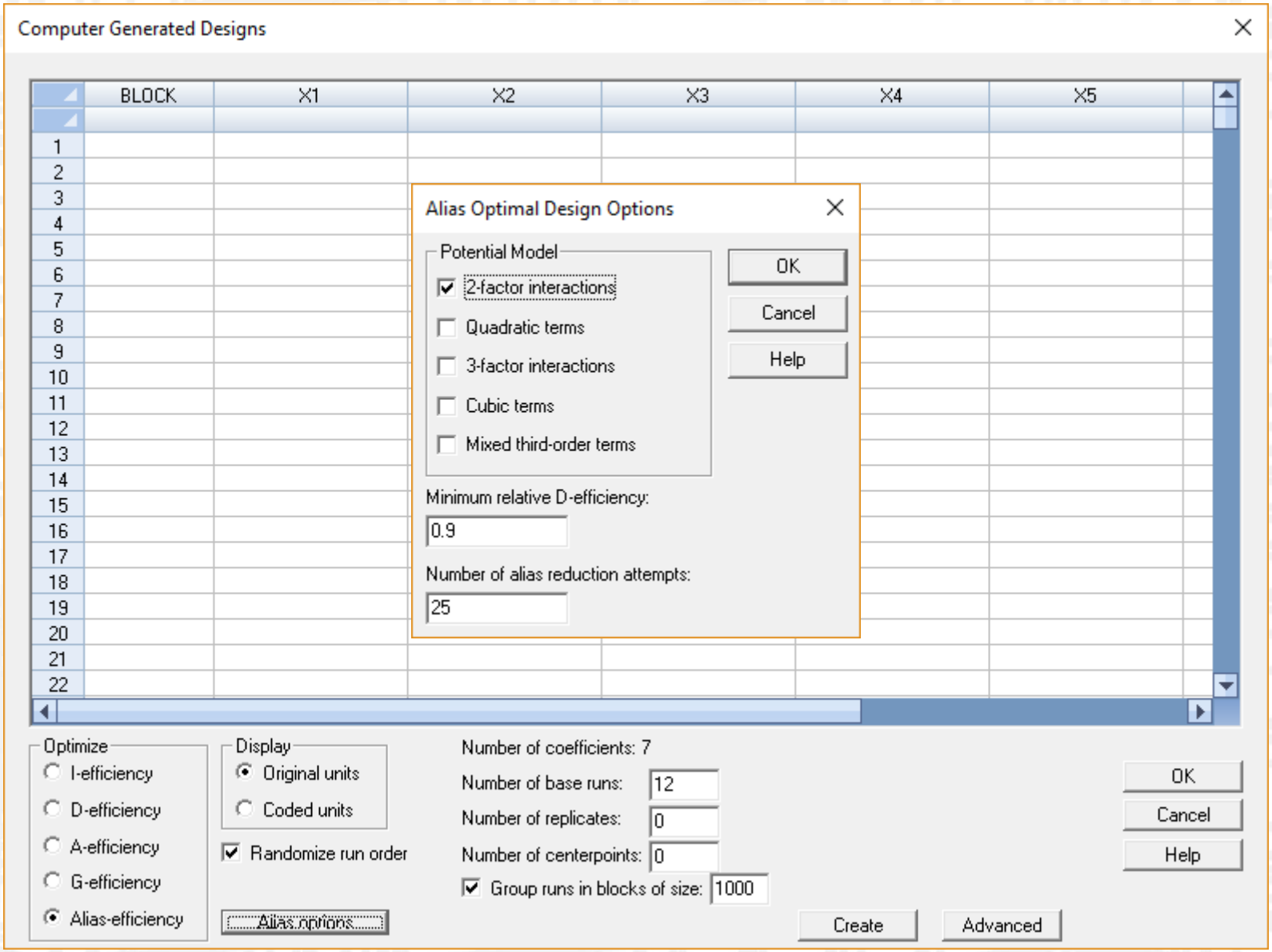






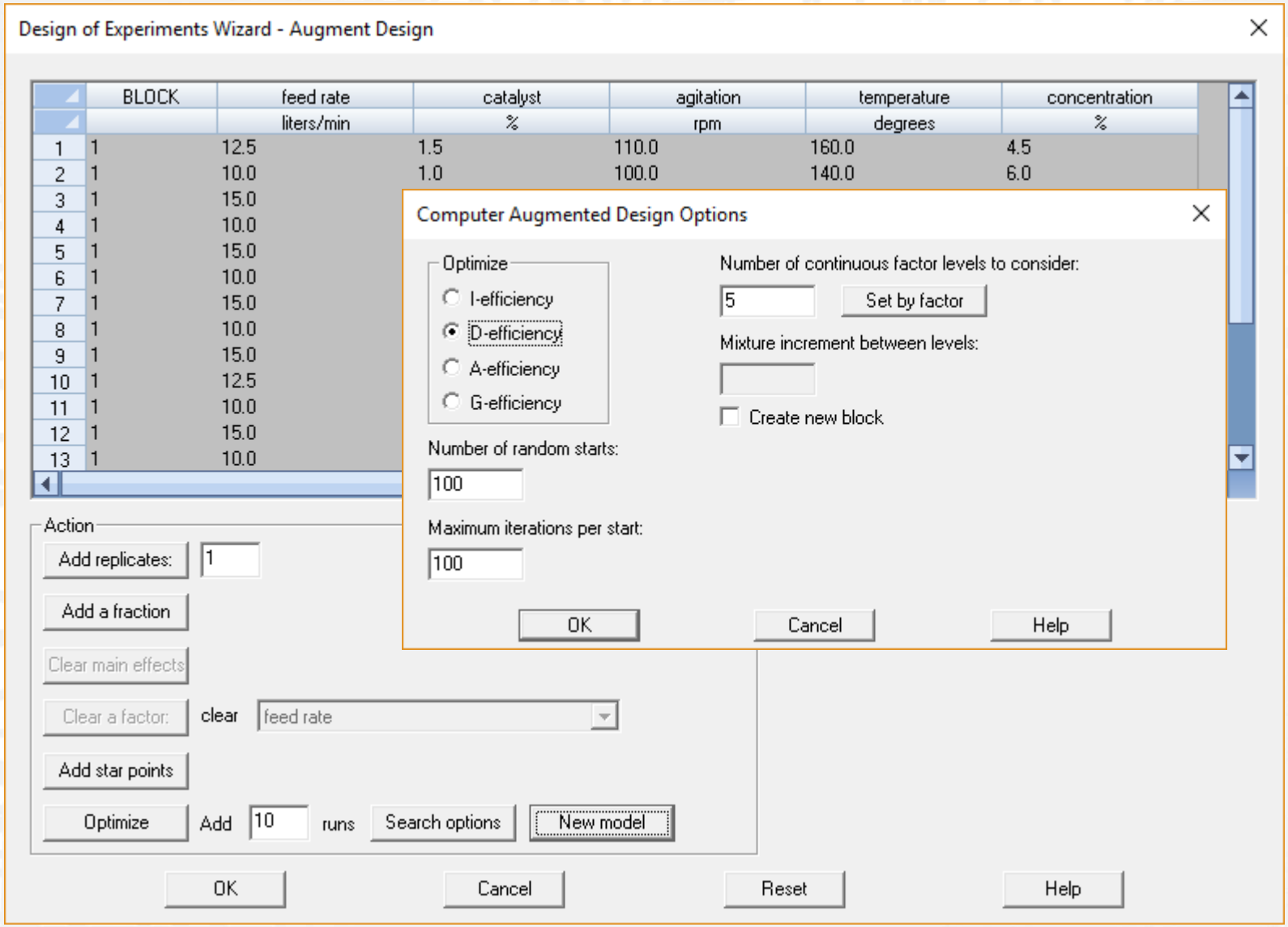
STATGRAPHICS.COM

DOE – Alias Optimal Designs



STATGRAPHICS.COM

DOE – Augment Design



STATGRAPHICS.COM

More Videos

Videos are available to learn more about each of the new features.

You'll find them at:

www.statgraphics.com/instructional_videos

Also check our website for upcoming webinars.

References

- The Covid Tracking Project – covidtracking.com