Multivariate Capability Analysis Using Statgraphics

Presented by

Dr. Neil W. Polhemus

Multivariate Capability Analysis

- Used to demonstrate conformance of a process to requirements or specifications that involve more than one variable.
- Most important when the variables are not independent or when the requirements concern the joint behavior of the variables.
- This webinar will consider variable data only, although attribute data is sometimes considered.

Webinar Outline

- Follows Chapter 7 of my book: ***Process Capability Analysis: Estimating Quality*** (2017) published by CRC Press.
  - Visualizing Bivariate Data
  - Multivariate Normal Distribution
  - Multivariate Tests for Normality
  - Multivariate Capability Indices
  - Multivariate Statistical Tolerance Limits
  - Analysis of Non-normal Multivariate Data

Example

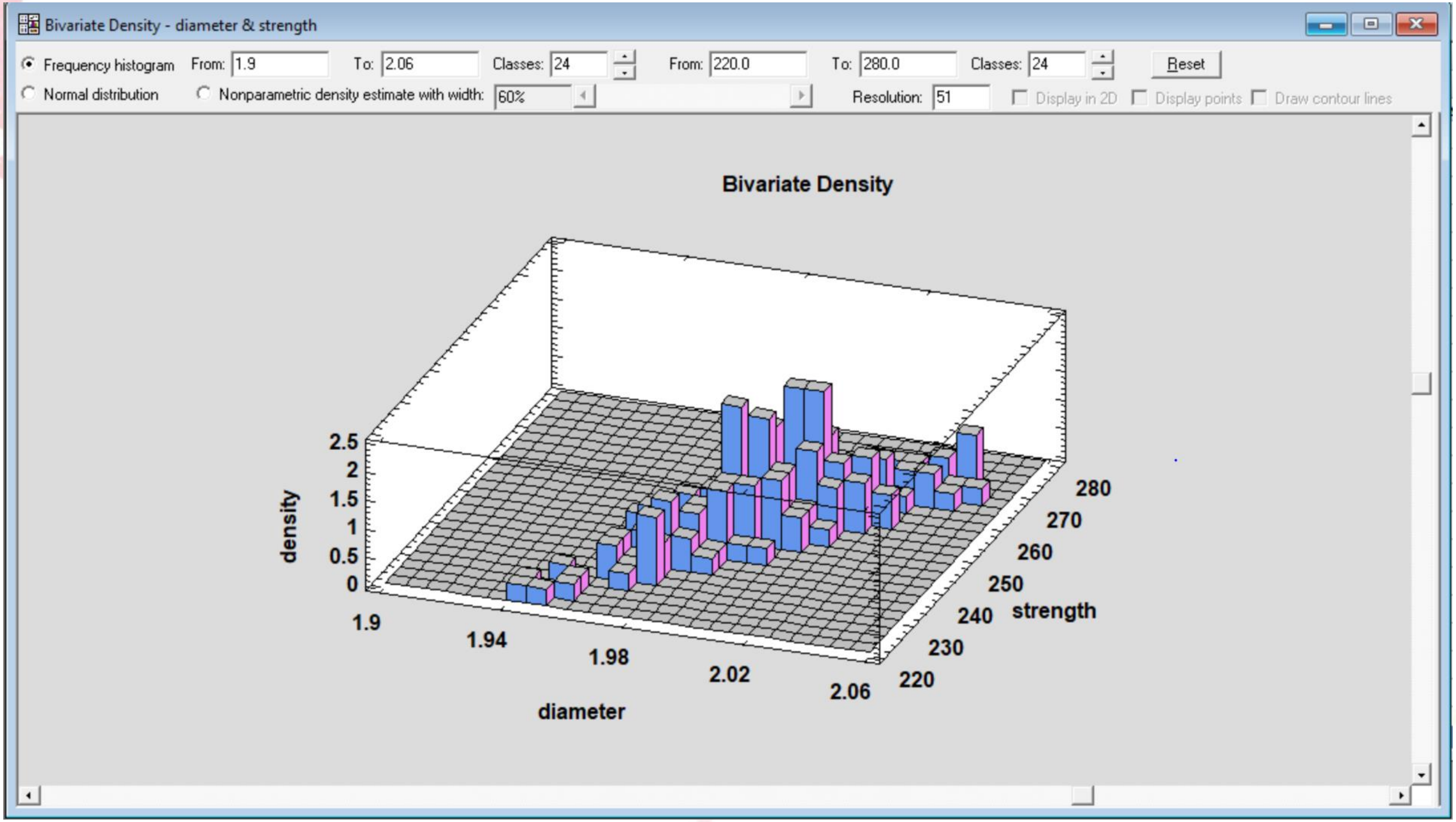
- Sample of n=200 medical devices.
- 2 variables:
  - $1.9 \leq$ **diameter** ≤ 2.1
  - **strength** ≥ 200

C:\Books\Capability Analysis\Dat...

	diameter	strength
	Numeric	Numeric
1	2.0185	249.864
2	2.02203	261.164
3	1.99604	248.786
4	1.99538	249.831
5	2.00224	253.328
6	1.99465	251.757
7	1.98519	241.059
8	2.01332	254.499
9	2.00017	254.337
10	2.00435	245.087
11	1.9896	240.632
12	1.98675	243.948

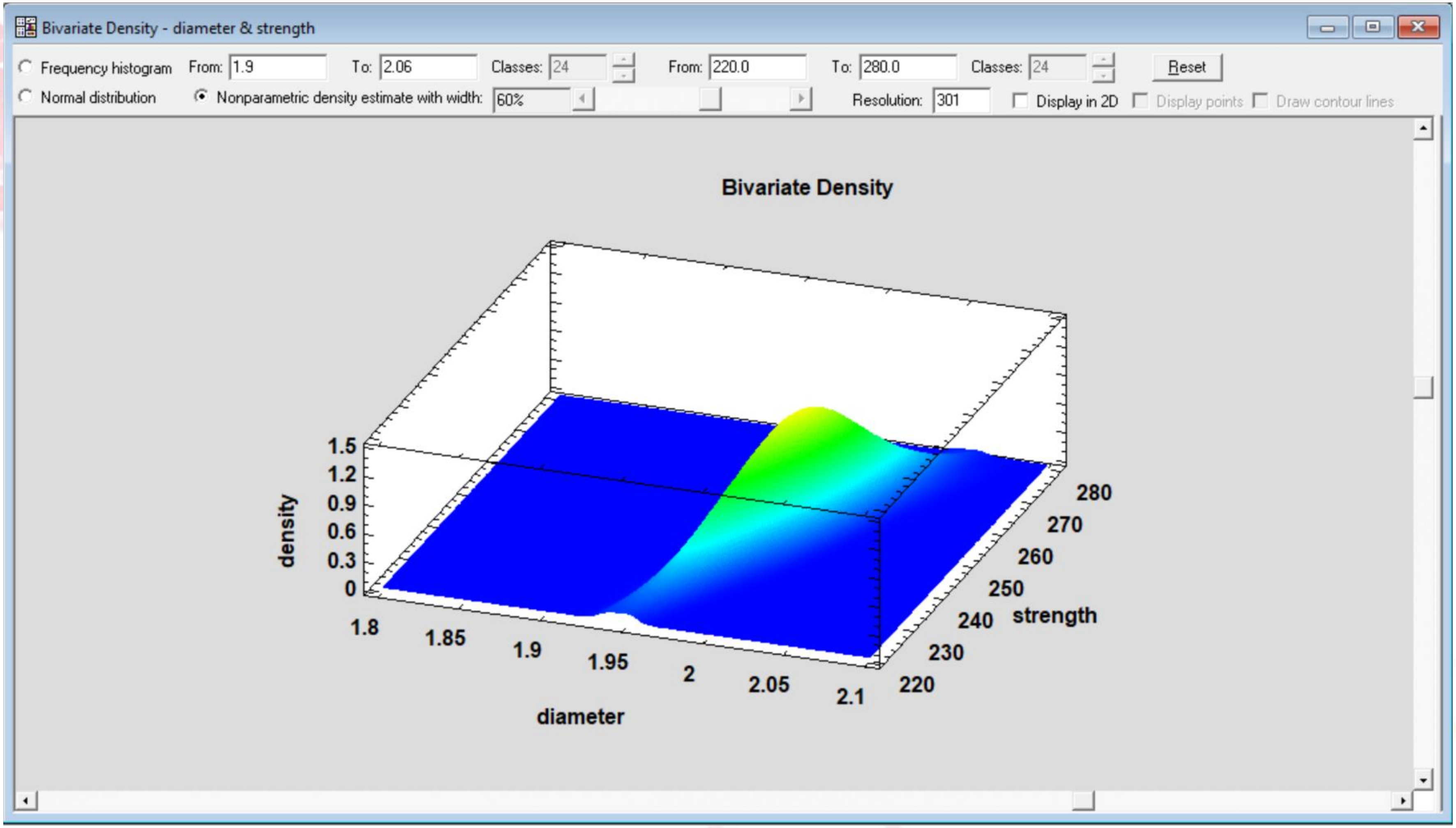
device:

Visualizing Bivariate Data

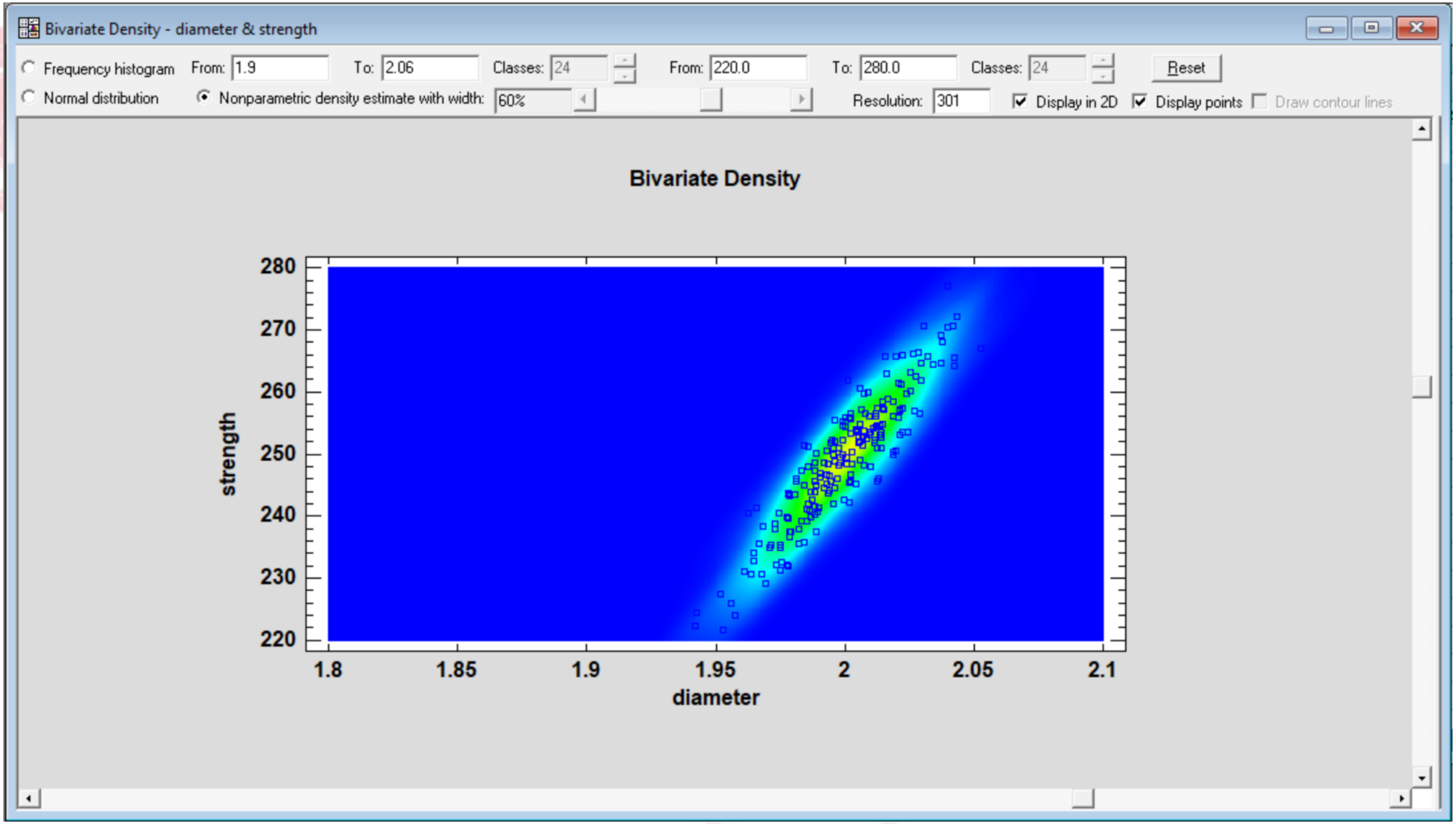


statgraphics® centurion
statgraphics.com

Bivariate Nonparametric Estimate



Bivariate Nonparametric Estimate



statgraphics.com

Multivariate Normal Distribution

$$f_X(X_1, X_2, \ldots, X_m) = \frac{1}{\sqrt{(2\pi)^m |\boldsymbol{\Sigma}|}} \exp\left(-\frac{1}{2}(\boldsymbol{X}-\boldsymbol{\mu})^T \boldsymbol{\Sigma}^{-1} (\boldsymbol{X}-\boldsymbol{\mu})\right)$$

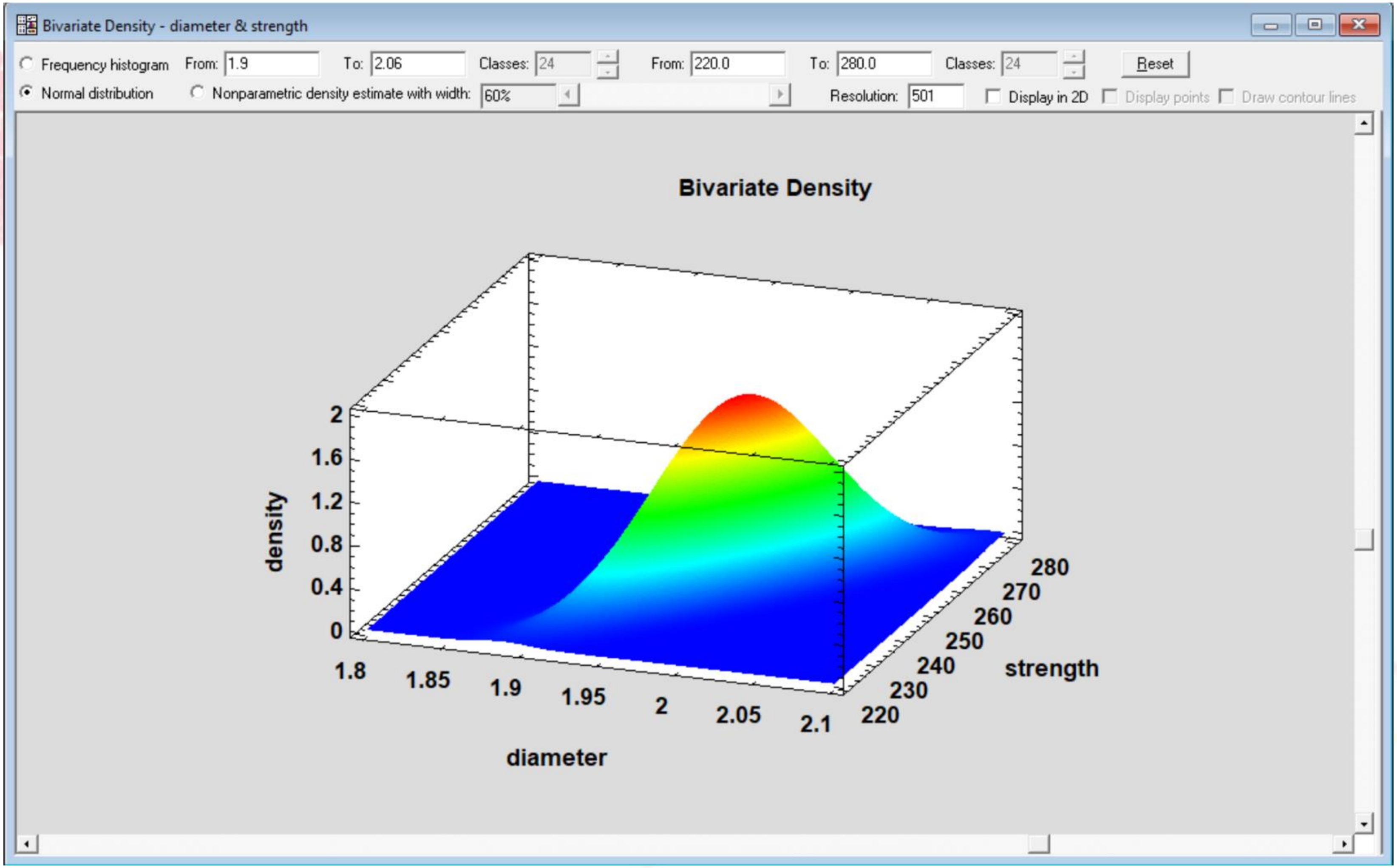
- Defined by:
  - a vector of *m* means μ
  - an *m* by *m* variance-covariance matrix Σ

- Important property: any linear combination of the *m* variables follows a univariate normal distribution

Bivariate Normal Distribution

- Defined by:
  - 2 means, μ_1 and μ_2
  - 2 standard deviations, σ_1 and σ_2
  - 1 correlation coefficient $-1 \leq \rho \leq 1$

Bivariate Normal Distribution

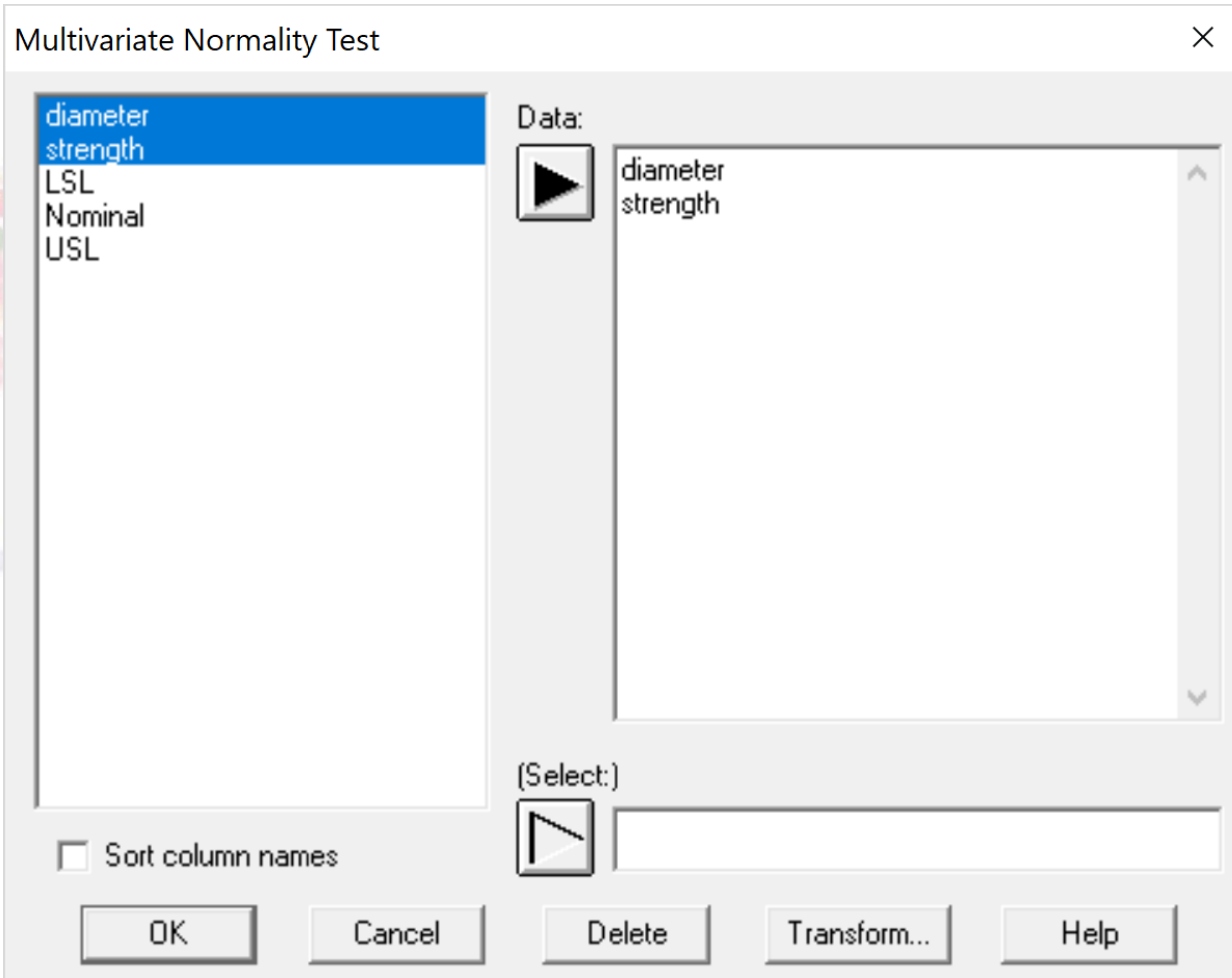


statgraphics centurion
statgraphics.com

Multivariate Tests for Normality

- Most widely used test is Roysten's test
- Roysten developed a test statistic called *H* that combines the Shapiro-Wilk *W* statistic for each of the *m* variables in the data set
- *H* is referred to a chi-square distribution with degrees of freedom that depend on the correlations amongst the variables

Data Input Dialog Box



statgraphics® centurion
statgraphics.com

Analysis Summary

Multivariate Normality Test

Data variables:
diameter
strength

	Mean	*Standard deviation*
diameter	1.99958	0.0208047
strength	249.3	10.4658

Sample Correlations

	diameter	strength
diameter	1.0	0.914186
strength	0.914186	1.0

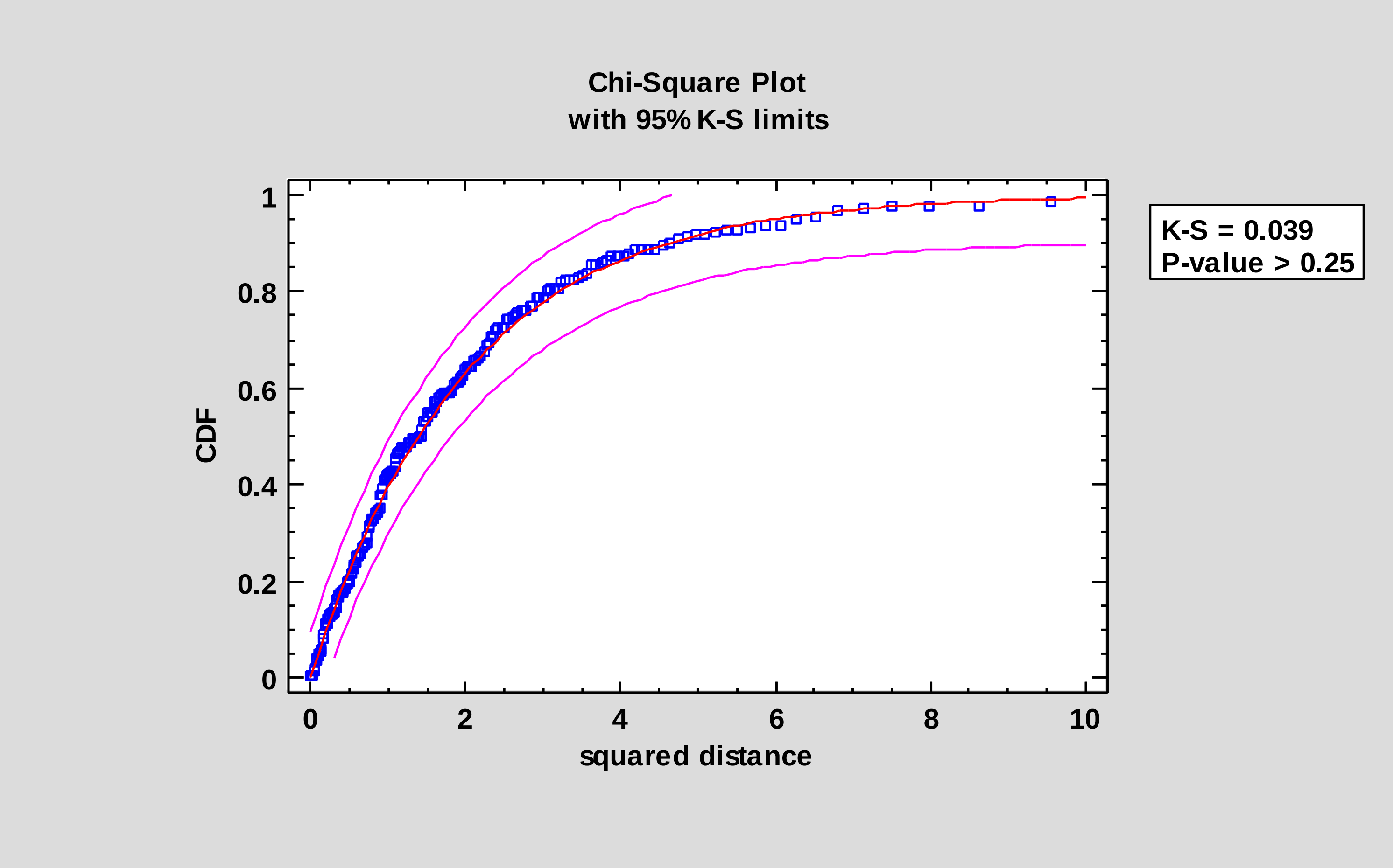
Number of observations = 200

Normality Tests

Test	*Statistic*	*P-Value*
Shapiro-Wilk W - diameter	0.997	0.9378
Shapiro-Wilk W - strength	0.993	0.5058
Royston's H	0.315	0.7178

The StatAdvisor
This procedure tests whether or not a multi-dimensional data sample may reasonably have come from a multivariate normal distribution. Based on Royston's H test, the hypothesis that the data come from a multivariate normal distribution cannot be rejected at the 5% significance level.

Chi-Square Plot



statgraphics® centurion

statgraphics.com

Multivariate Capability Indices

- In the case of univariate data, indices may be calculated that help summarize how well a process meets a set of specification limits.

- The most widely used index is C_{pk}, which measures how many standard deviations separate the process mean from the nearer specification limit.

$$C_{pk} = min\left(\frac{\mu - LSL}{3\sigma}, \frac{USL - \mu}{3\sigma}\right)$$

Multivariate Capability Indices

We may also write C_{pk} as

$$C_{pk} = \frac{Z_{min}}{3}$$

where

$$Z_{min} = min\left(\frac{\mu - LSL}{\sigma}, \frac{USL - \mu}{\sigma}\right)$$

Multivariate Capability Indices

Assuming that the variable of interest follows a univariate normal distribution, the probability of obtaining an observation more than Z_{min} standard deviations from the mean is approximately equal to

$$\theta = 1 - \Phi(Z_{min})$$

where $\Phi(Z)$ is the cumulative standard normal distribution evaluated at Z. If θ is specified, we can approximate C_{pk} using the inverse standard normal distribution

$$C_{pk} \approx \frac{1}{3}\Phi^{-1}(1-\theta)$$

statgraphics® centurion
statgraphics.com

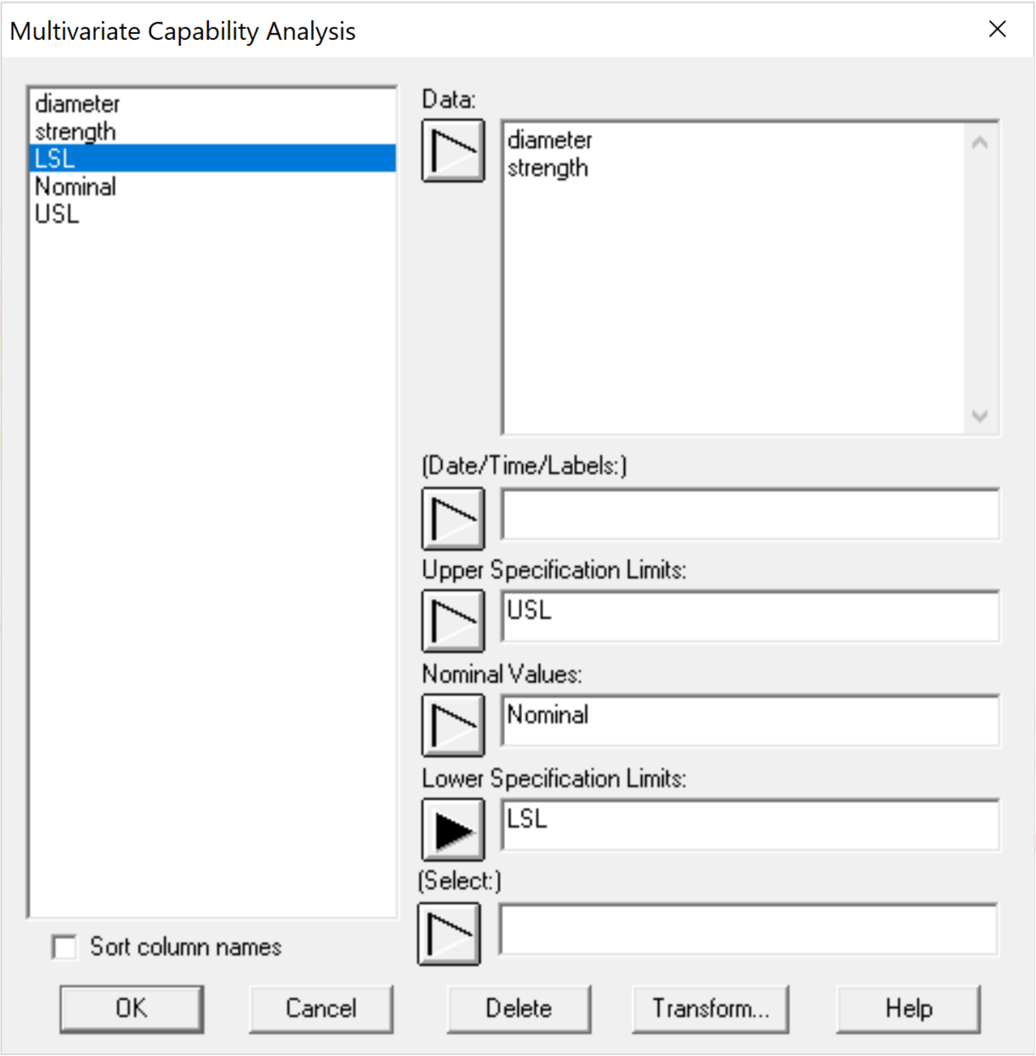
Multivariate C_{pk}

We will define a multivariate analog to C_{pk} according to

$$MC_{pk} = \frac{Z_{min}}{3}$$

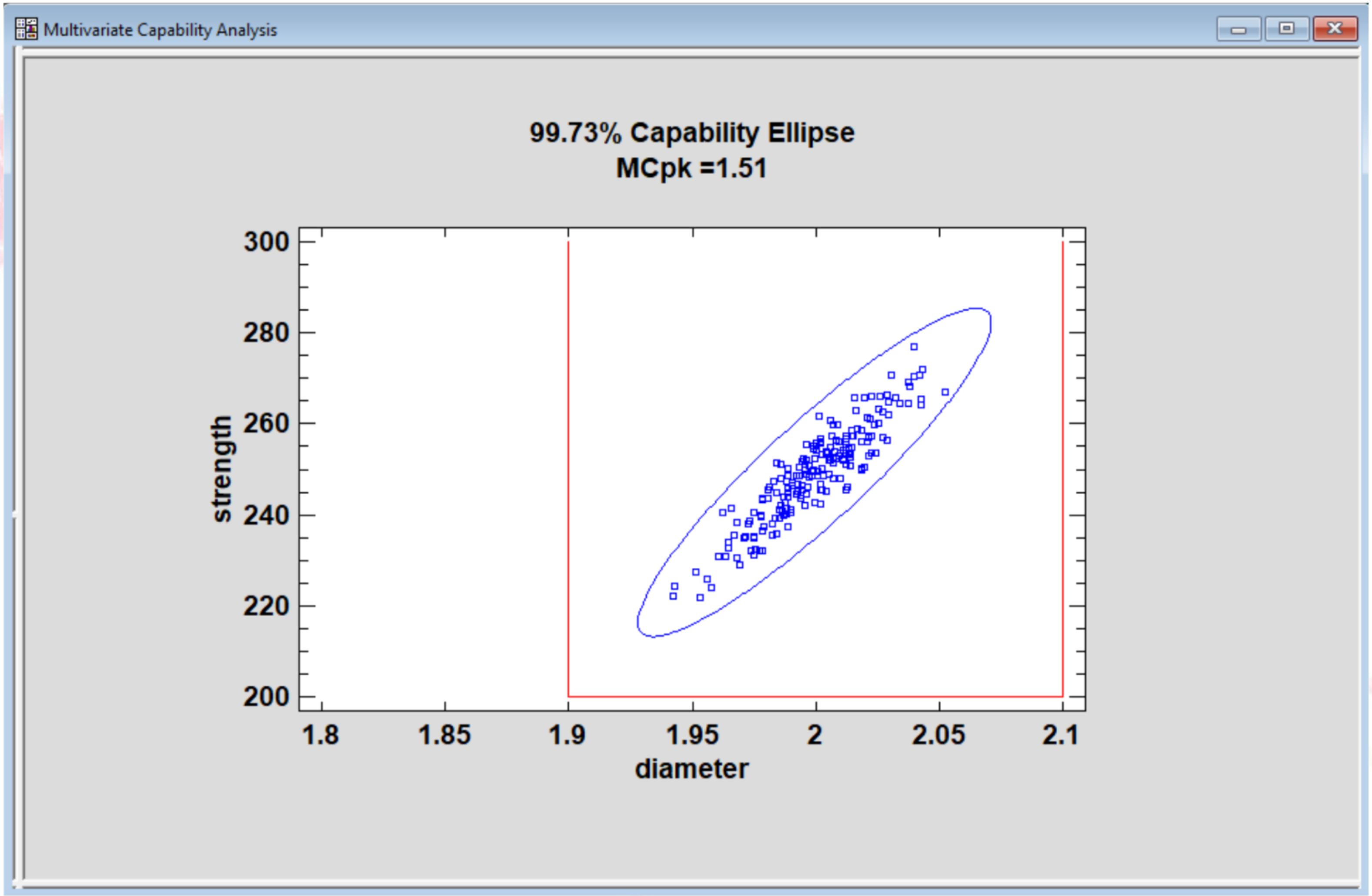
where Z_{min} is the value of a standard normal random variable which is exceeded with probability θ and θ is the probability that our vector of random variables is outside of the region defined by the specification limits when calculated from a multivariate normal distribution.

Data Input



statgraphics® centurion
statgraphics.com

Capability Ellipse



statgraphics® centurion

statgraphics.com

Analysis Summary

Multivariate Capability Analysis

Data variables:
diameter
strength

Number of complete cases: 200

	Sample	*Sample*			
Variable	*Mean*	*Std. Dev.*	*LSL*	*Nominal*	*USL*
diameter	1.99958	0.0208047	1.9	2.0	2.1
strength	249.3	10.4658	200.0		

	Observed	*Estimated*	*Estimated*
Variable	*Beyond Spec.*	*Beyond Spec.*	*DPM*
diameter	0.0%	0.000154463%	1.54463
strength	0.0%	0.000123616%	1.23616
Joint	0.0%	0.000227784%	2.27784

The StatAdvisor

This procedure determines the percentage of items beyond a set of multivariate specification limits. In this case, the estimated frequency of non-conformities with respect to at least one of the 2 variables equals 2.27784 per million.

statgraphics.com

Capability Indices

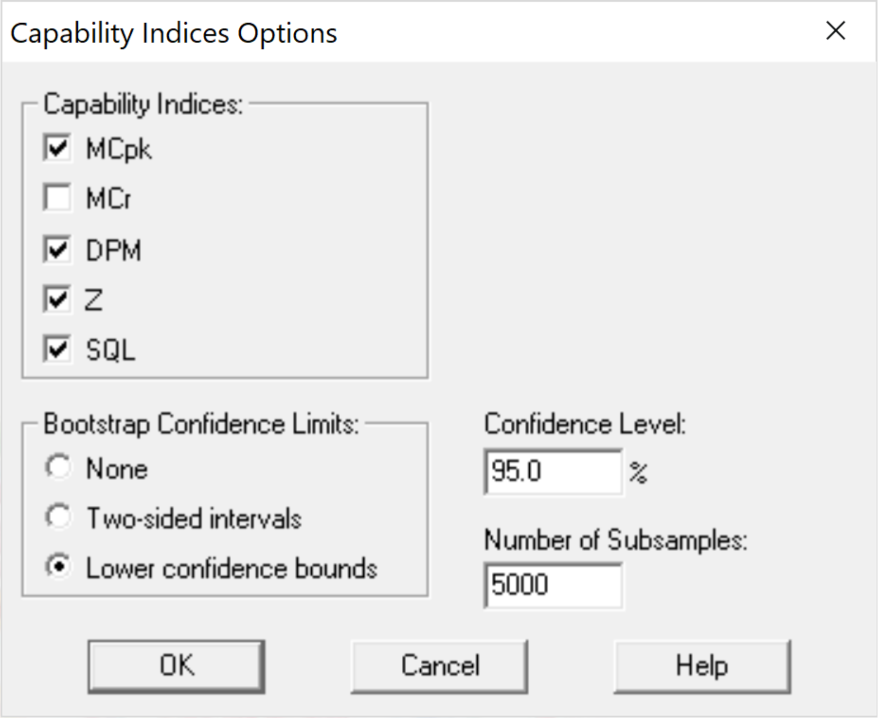
Capability Indices

Index	*Estimate*
MCpk	1.53
DPM	2.27784
Z	4.58428
SQL	6.08428

95.0% Confidence Bounds - Bootstrap Method (5000 subsamples)

	Lower Limit
MCpk	1.40289
*DPM	12.8435
Z	4.20868
SQL	5.70868

*Lower quality bound corresponds to upper confidence bound for this index.



Note: lower confidence bounds determined by bootstrapping.

statgraphics® centurion
statgraphics.com

Multivariate Statistical Tolerance Limits

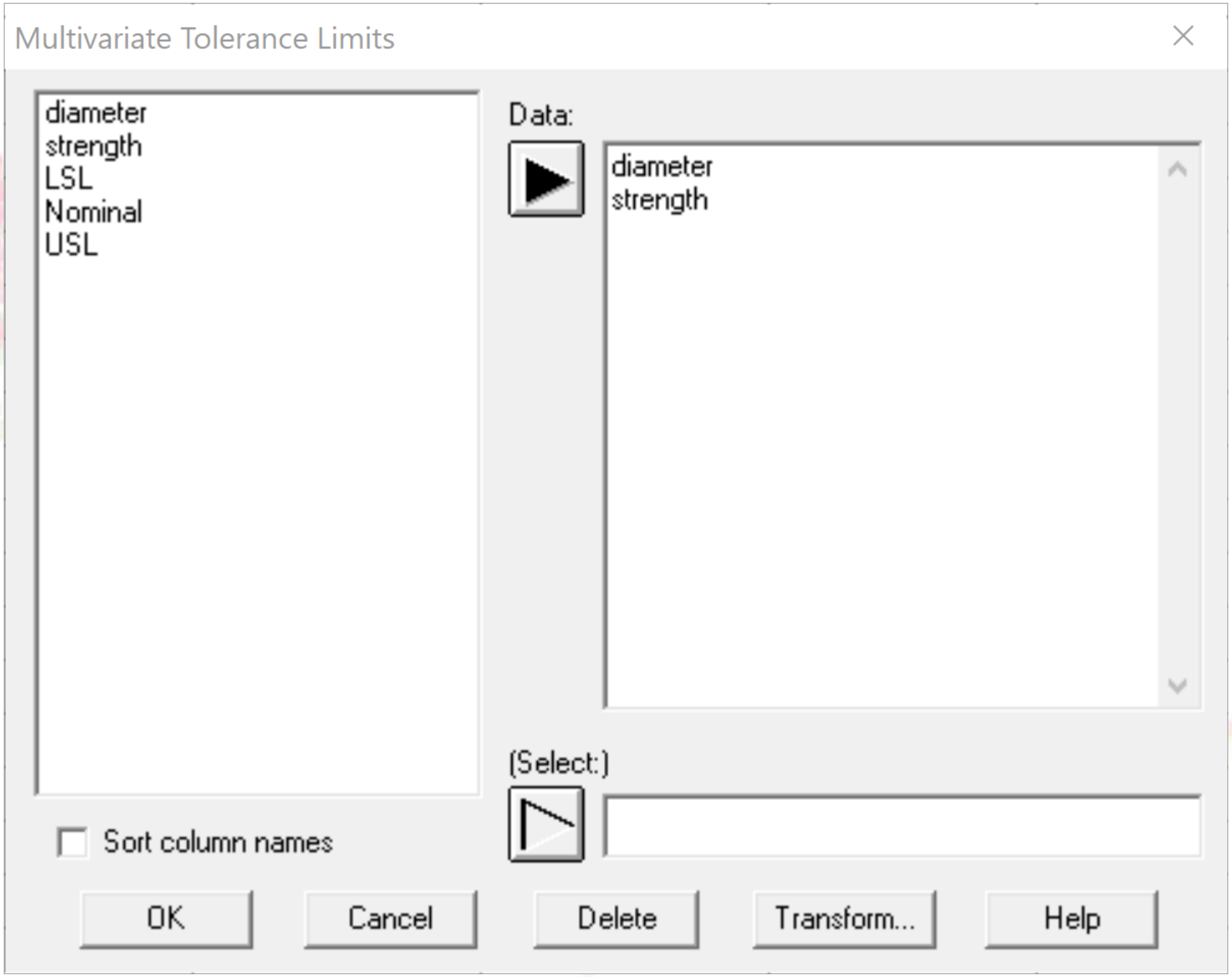
- Contain a specified percentage of a multivariate population at a specified level of confidence.
- For example, we may ask for an interval that contains at least 99% of all multivariate observations with 95% confidence.
- If entirely within the region defined by the specification limits, we know that at least 99% of all multivariate observations jointly satisfy those specifications.

Methods

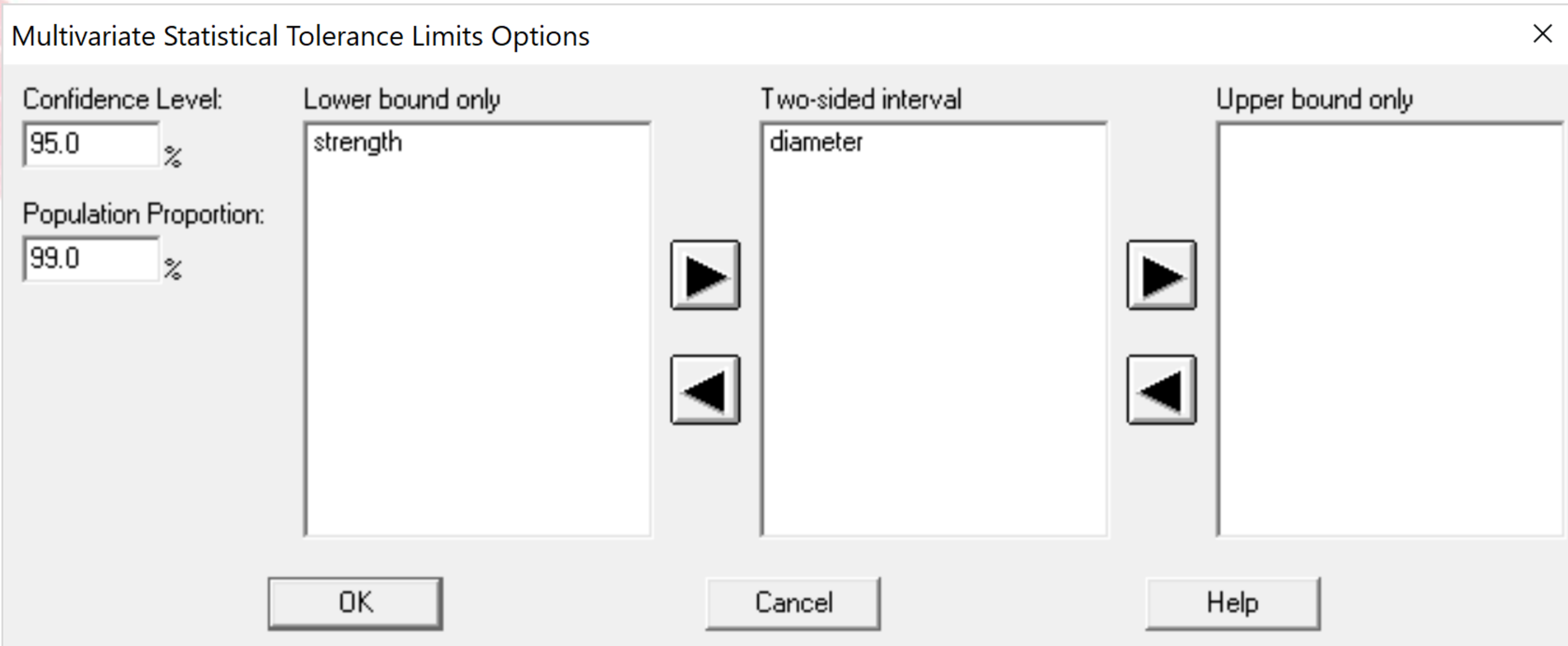
- Method 1: construct an elliptical tolerance region based on a multivariate normal distribution.

- Method 2: construct separate tolerance limits, adjusting each set of limits using a Bonferroni approach.

Data Input


Multivariate Tolerance Limits
diameter
strength
LSL
Nominal
USL
Data:
diameter
strength
(Select:)
Sort column names
OK
Cancel
Delete
Transform...
Help

Analysis Options



statgraphics centurion
statgraphics.com

Analysis Summary

Multivariate Tolerance Limits

Data variables:
    diameter
    strength

	Mean	*Standard deviation*
diameter	1.99958	0.0208047
strength	249.3	10.4658

Sample Correlations

	diameter	strength
diameter	1.0	0.914186
strength	0.914186	1.0

Number of observations = 200

95% Simultaneous Bonferroni Tolerance Limits for 99% of the Population

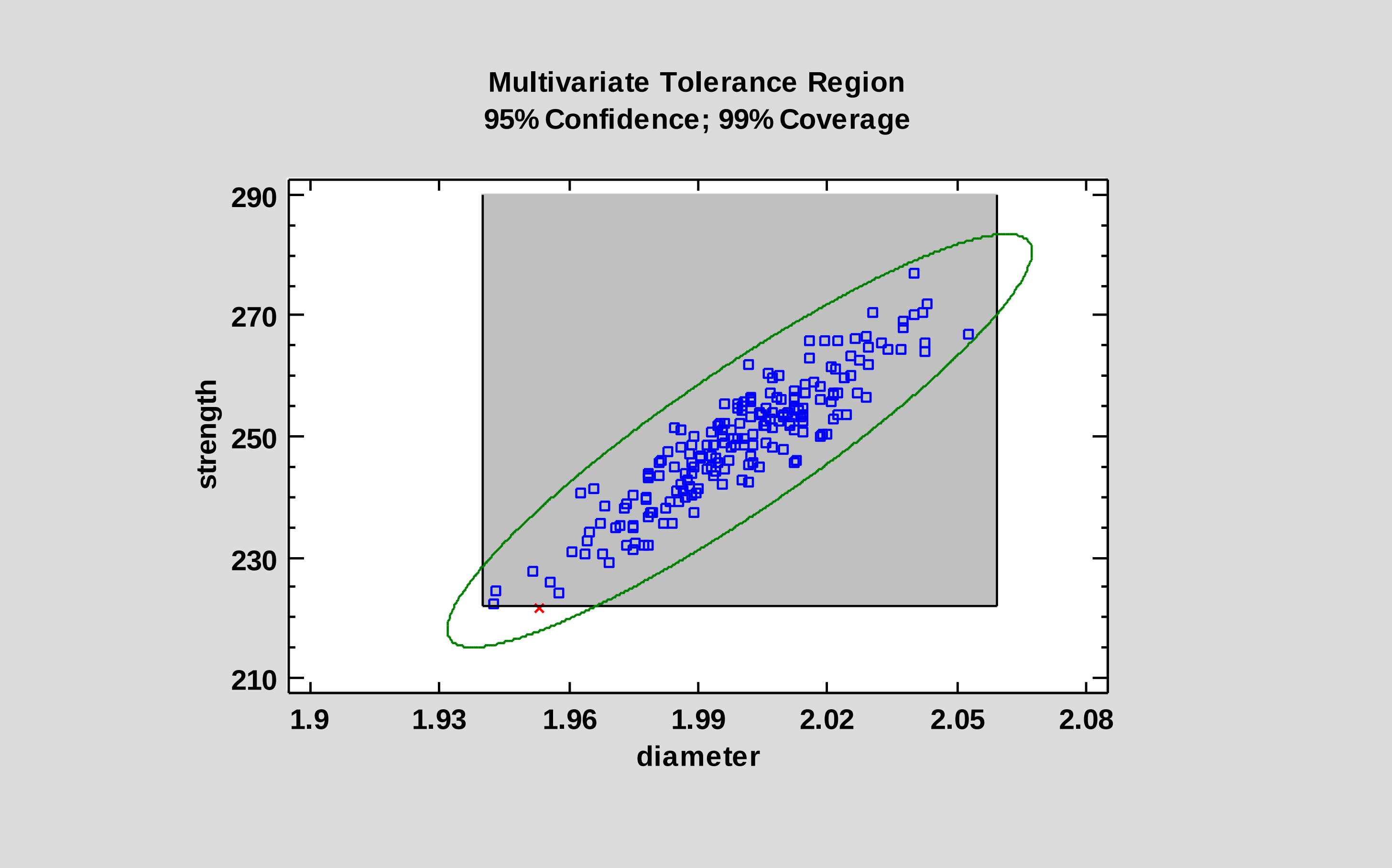
	Lower Limit	*Upper Limit*
diameter	1.93999	2.05917
strength	221.895	

Observations beyond Bonferroni limits: 1

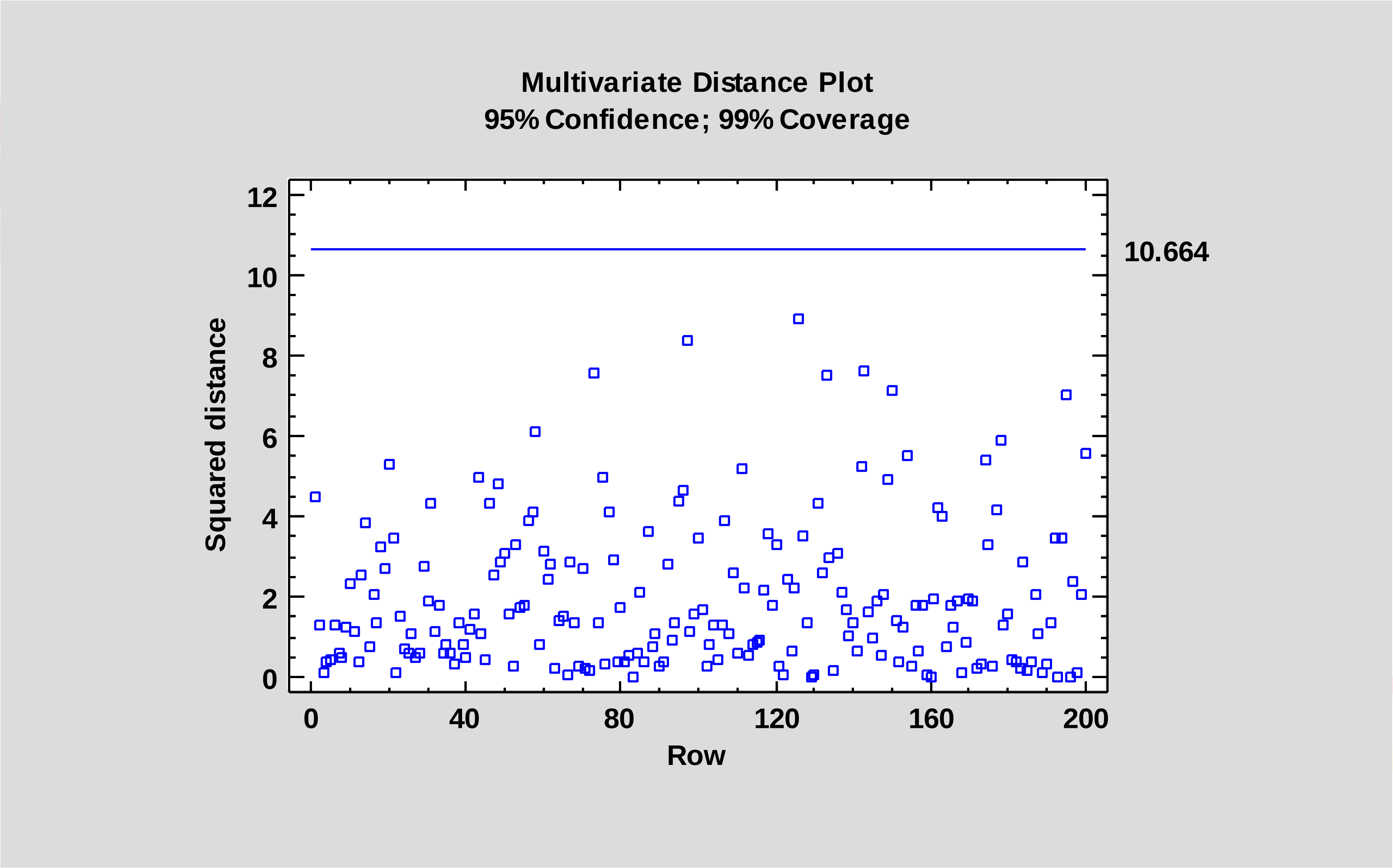
95% Elliptical Tolerance Region for 99% of the Population: Squared distance <= 10.664
Observations outside elliptical region: 0

The StatAdvisor
This pane displays multivariate tolerance limits for 2 variables. The limits contain 99% of the population from which the data come with 95% confidence, assuming that the data are a random sample from a multivariate normal distribution. The Bonferroni intervals provide simultaneous limits for each of the variables and may be somewhat conservative. The elliptical tolerance region is based on the distance of the observations from the centroid of the data in a multivariate space. Any multivariate observation with a squared generalized distance from the centroid greater than 10.664 is outside of the tolerance region.

Tolerance Region Plot



Distance Plot



statgraphics.com

Analysis of Multivariate Non-Normal Data

- Best approach is to transform one or more of the variables.
  - Apply the Box-Cox method to each variable separately
  - Perform a multivariate power transformation by finding powers to apply to each variable that maximize the joint profile likelihood

Multivariate Power Transformation

Multivariate Normality Test

Multivariate Normality Test

Data variables:

X1 (Random Numbers)

X2 (Random Numbers)

Power transformations: estimated simultaneously

Variable	*Power*
X1	-0.0802293
X2	0.0737869

	Mean	*Standard deviation*
X1	0.922674	0.014222
X2	1.16046	0.0335444

Sample Correlations

	X1	X2
X1	1.0	-0.880659
X2	-0.880659	1.0

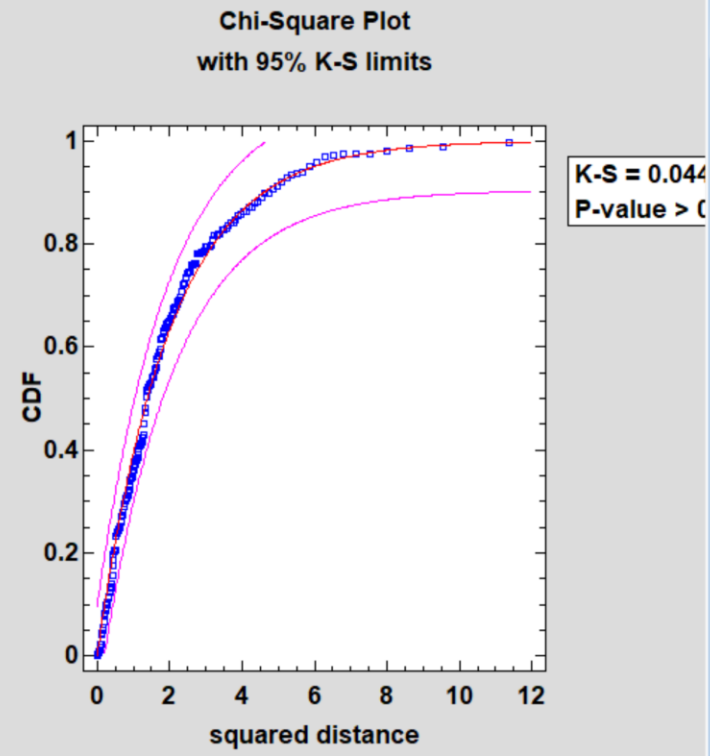
Number of observations = 200

Normality Tests

Test	*Statistic*	*P-Value*
Shapiro-Wilk W - X1	0.996	0.9192
Shapiro-Wilk W - X2	0.997	0.9764
Royston's H	0.009	0.9834

The StatAdvisor

This procedure tests whether or not a multi-dimensional data sample may reasonably have distribution. Based on Royston's H test, the hypothesis that the data come from a multivaria at the 5% significance level.



References

- StatFolios and data files are at:
  www.statgraphics.com/webinars

- Information on *Process Capability Analysis: Estimating Quality* may be found at:
  www.statgraphics.com/process-capability-analysis-book