Statistics Decision Tree

DATA

Look at Distribution
- Histogram
- Stem-and-Leaf

Describe Distribution-Moments
- Mean
- Standard Deviation/Variance
- Skewness
- Kurtosis

Determine Type of Distribution
- Normal
- Beta
- Gamma
- Exponential
- Log Normal
- General: Pearson Distributions

Test - Type of Distribution
- Normal Probability Plot
- Correlation Test for Normality
- Chi Square Test for Distribution

Compare Distribution to Limits
- Cp
- Cpk
- Variance from Target

DATA vs TIME

Look at Trend versus Time
- Trend Chart

Model Distribution vs Time
- Time Series Modeling
- Autocorrelation
- Partial Autocorrelation
- Moving Average
- EWMA
- AR
- MA
- ARIMA

Study Sources -Time Variation
- Gauge Capability
- Variance Components Analysis

Compare Trend to Limits
- Control Charts
- X-Bar
- R, S
- Individuals
- Moving R
- EWMA

DATA vs DATA

One Input Variable

Compare Variability
- F-ratio test (two levels)
- Bartlett's test (multiple levels)
- Cochran's test (multiple levels)

Compare Means
- Student's T Test (two levels)
- ANOVA (multiple levels)
- Nested ANOVA (multiple levels)

Compare Medians
- Mann-Whitney (two levels)
- Kruskal-Wallis (multiple levels)

Study Source of Variation
- Y vs X plot
- Correlation Coefficient
- Linear Regression

Compare Proportions
- Proportion Test
- Chi-Square Test

Multiple Input Variables

Compare Proportions
- Chi-Square Test

Screening Experiments
- Full Factorial
- Fractional Factorial

Analysis of Experiments
- ANOVA
- Multiple Linear Regression

Response Surface Modeling
- Box-Behnken Designs
- Central Composite Designs
- Multiple Linear Regression
- Stepwise Regression
- Contour Plots
- 3 D Mesh Plots

Model Response Distribution
- Monte Carlo Simulation
- Generation of System Moments

Optimization
- Optimization of Expected Value:
- Linear Programming
- Non Linear Programming
- Yield Surface Modeling™

Section 1: Data

DATA

Look at Distribution

- Histogram
- Stem-and-Leaf

Describe Distribution-Moments

- Mean
- Standard Deviation/Variance
- Skewness
- Kurtosis

Determine Type of Distribution

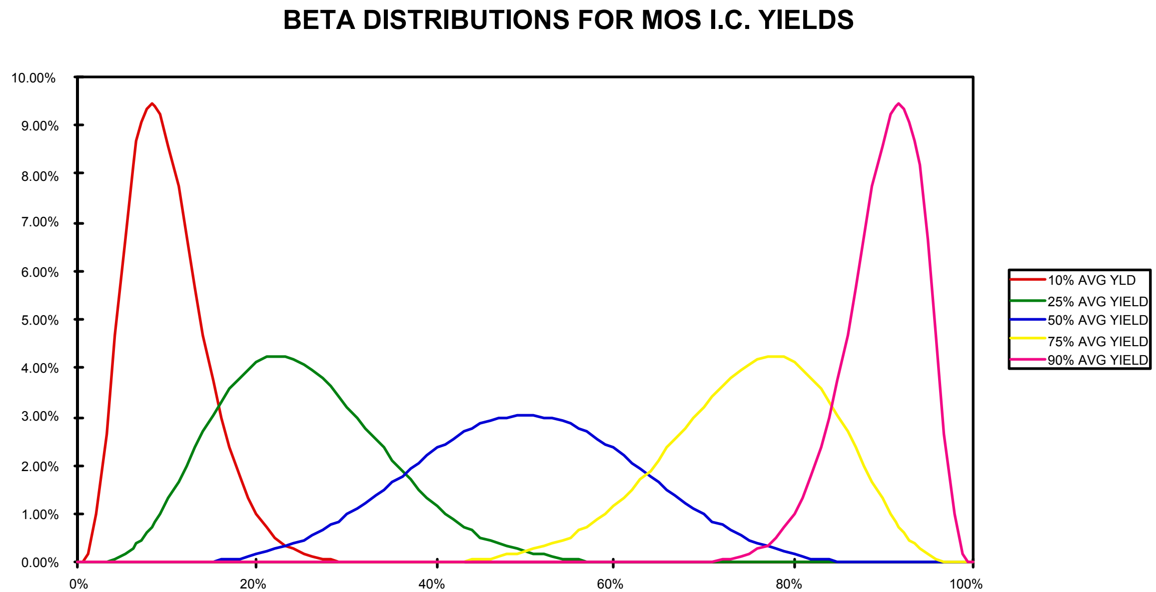
- Normal
- Beta
- Gamma
- Exponential
- Log Normal
- General: Pearson Distributions

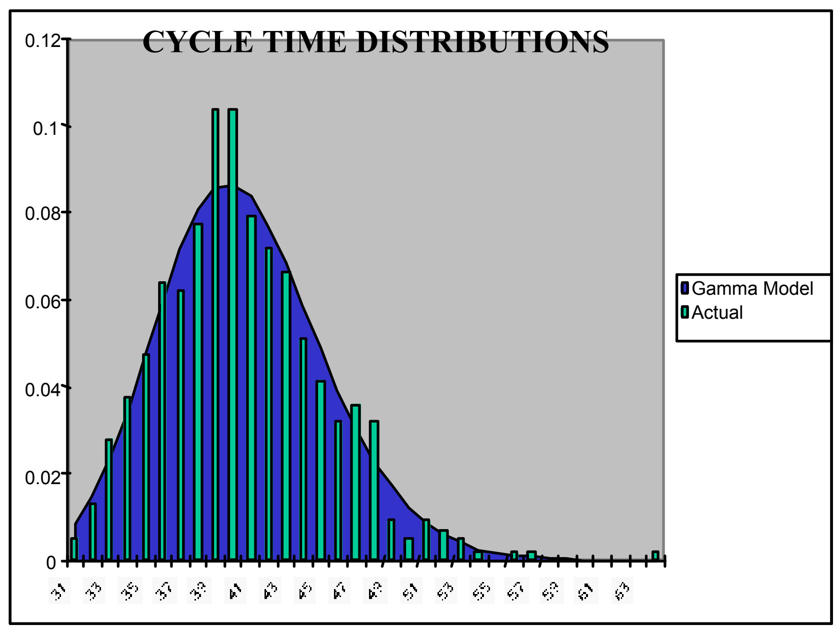
Test - Type of Distribution

- Normal Probability Plot
- Correlation Test for Normality
- Chi Square Test for Distribution

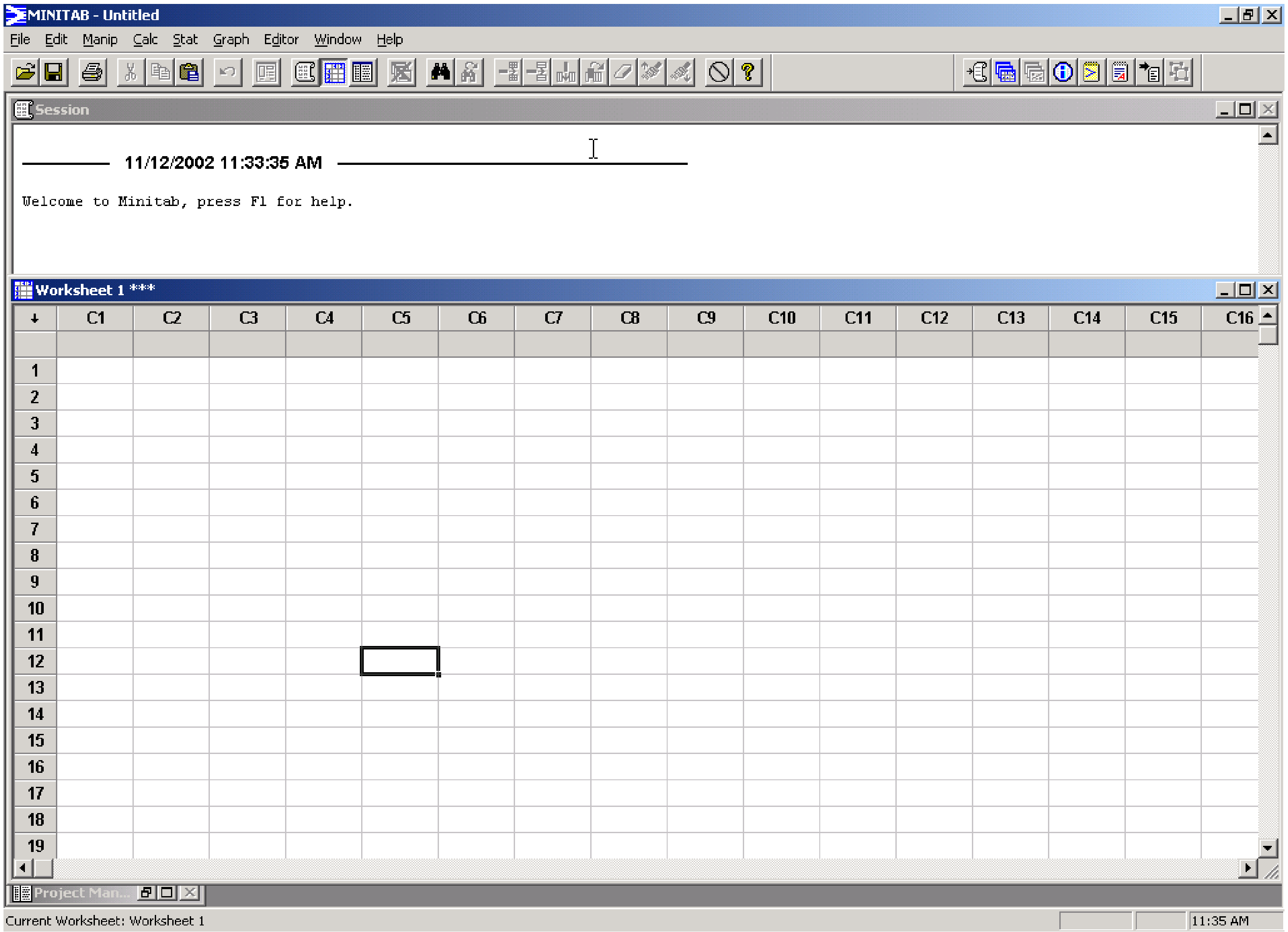
Compare Distribution to Limits

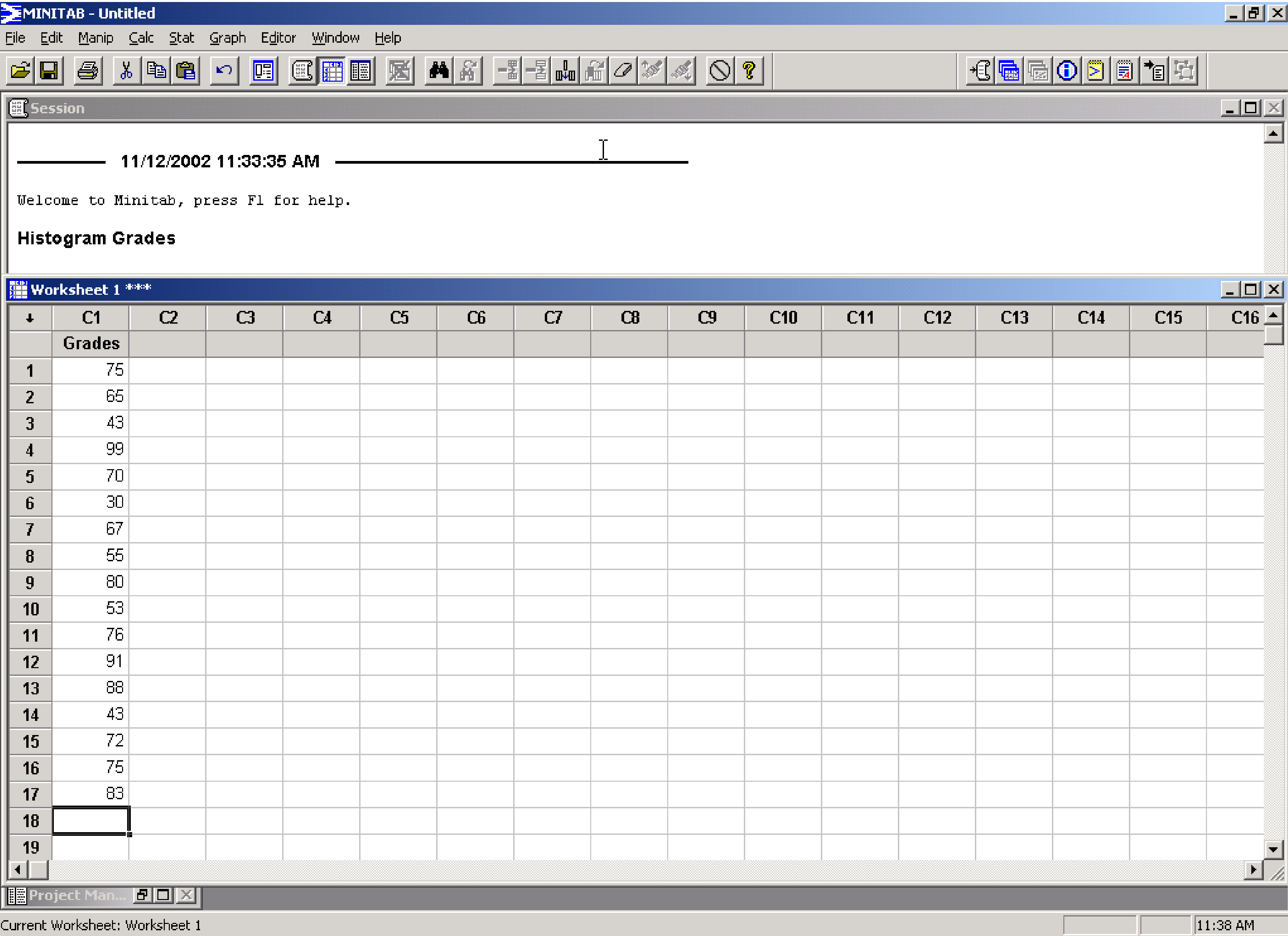
- Cp
- Cpk
- Variance from Target

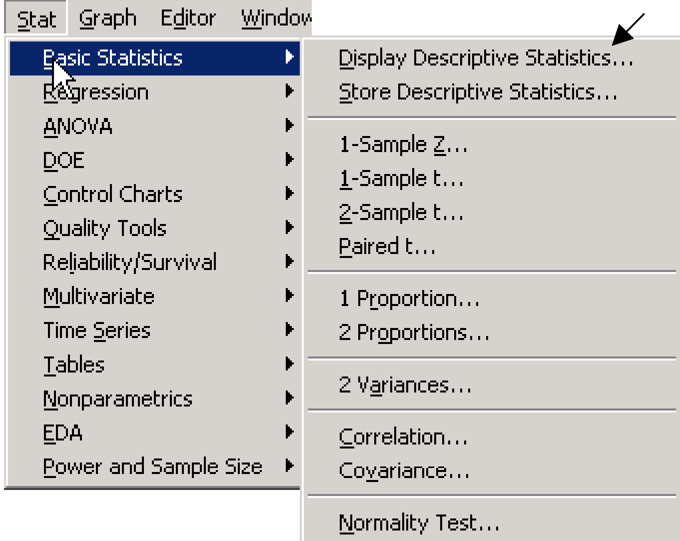


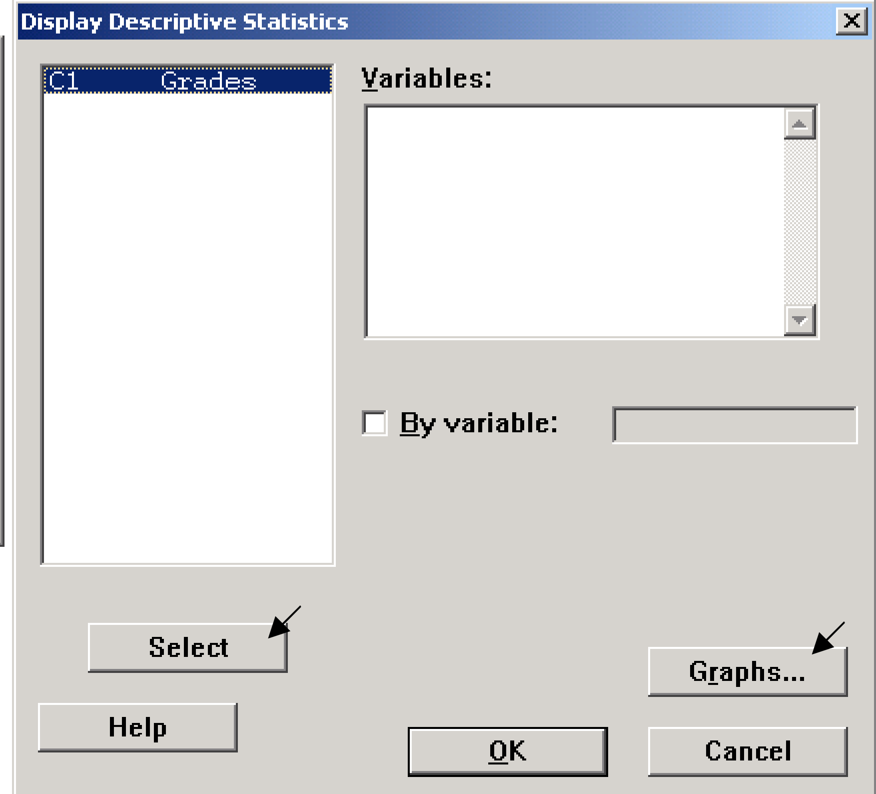


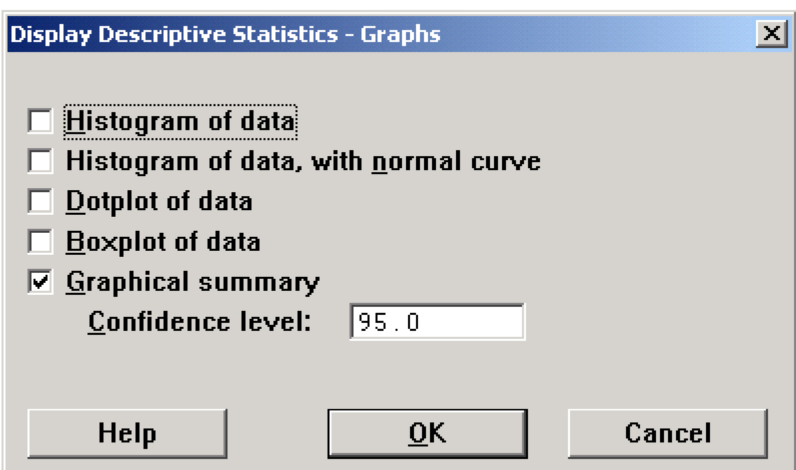
Minitab: Data Entry and Histograms

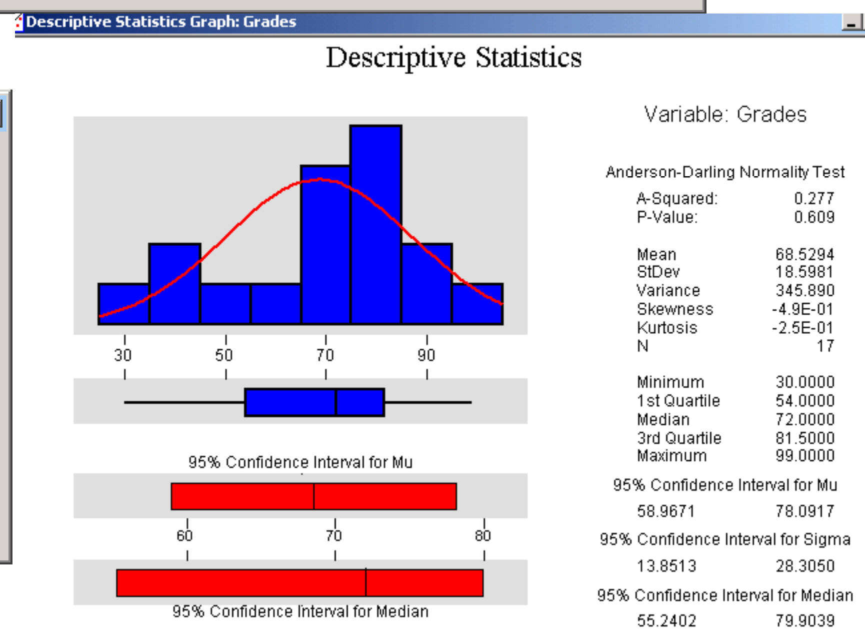



MINITAB - Untitled
File Edit Manip Calc Stat Graph Editor Window Help
Session
11/12/2002 11:33:35 AM
Welcome to Minitab, press F1 for help.
Histogram Grades
Worksheet 1 ***
C1 C2 C3 C4 C5 C6 C7 C8 C9 C10 C11 C12 C13 C14 C15 C16
Grades
1 75
2 65
3 43
4 99
5 70
6 30
7 67
8 55
9 80
10 53
11 76
12 91
13 88
14 43
15 72
16 75
17 83
18
19
Project Man...
Current Worksheet: Worksheet 1
11:38 AM

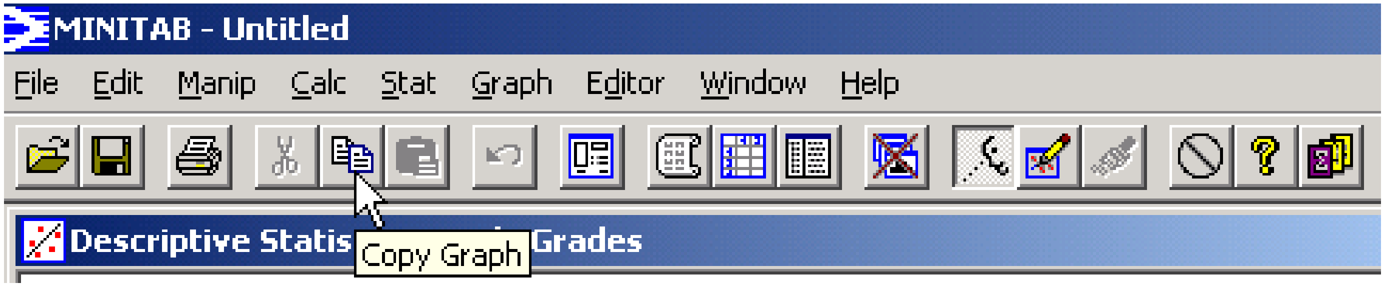

Stat Graph Editor Window
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Display Descriptive Statistics...
Store Descriptive Statistics...
1-Sample Z...
1-Sample t...
2-Sample t...
Paired t...
1 Proportion...
2 Proportions...
2 Variances...
Correlation...
Covariance...
Normality Test...


Display Descriptive Statistics
C1 Grades
Variables:
By variable:
Select
Help
Graphs...
OK
Cancel


Display Descriptive Statistics - Graphs
Histogram of data
Histogram of data, with normal curve
Dotplot of data
Boxplot of data
Graphical summary
Confidence level:
95.0
Help
OK
Cancel

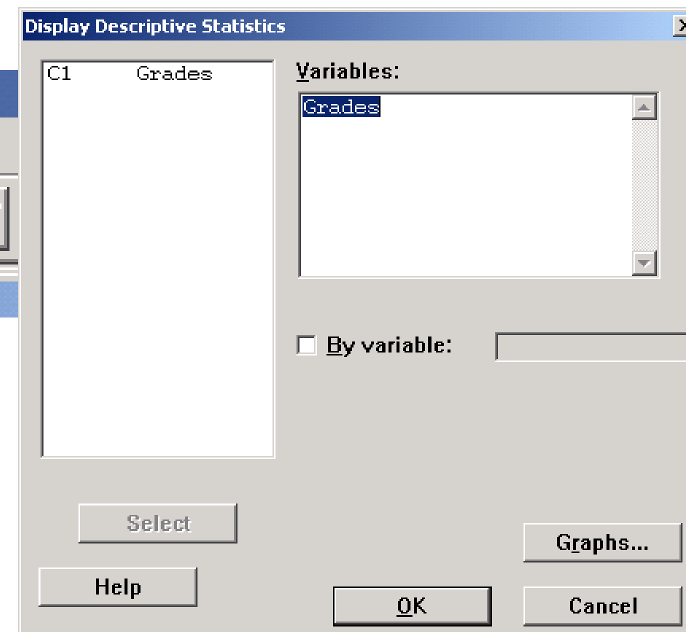

Descriptive Statistics Graph: Grades
Descriptive Statistics
Variable: Grades
Anderson-Darling Normality Test
A-Squared: 0.277
P-Value: 0.609
Mean 68.5294
StDev 18.5981
Variance 345.890
Skewness -4.9E-01
Kurtosis -2.5E-01
N 17
Minimum 30.0000
1st Quartile 54.0000
Median 72.0000
3rd Quartile 81.5000
Maximum 99.0000
95% Confidence Interval for Mu
58.9671 78.0917
95% Confidence Interval for Sigma
13.8513 28.3050
95% Confidence Interval for Median
55.2402 79.9039
30
50
70
90
95% Confidence Interval for Mu
60
70
80
95% Confidence Interval for Median

Some helpful toolbar options:



Copy Graph

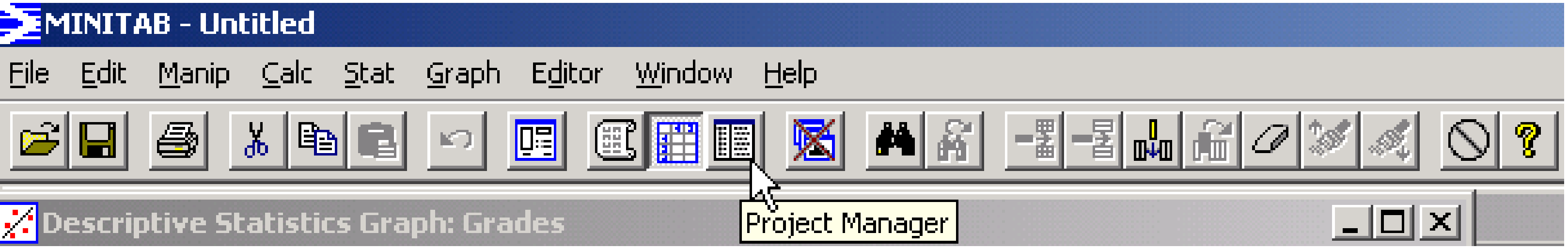
for pasting elsewhere



Edit Last Dialog

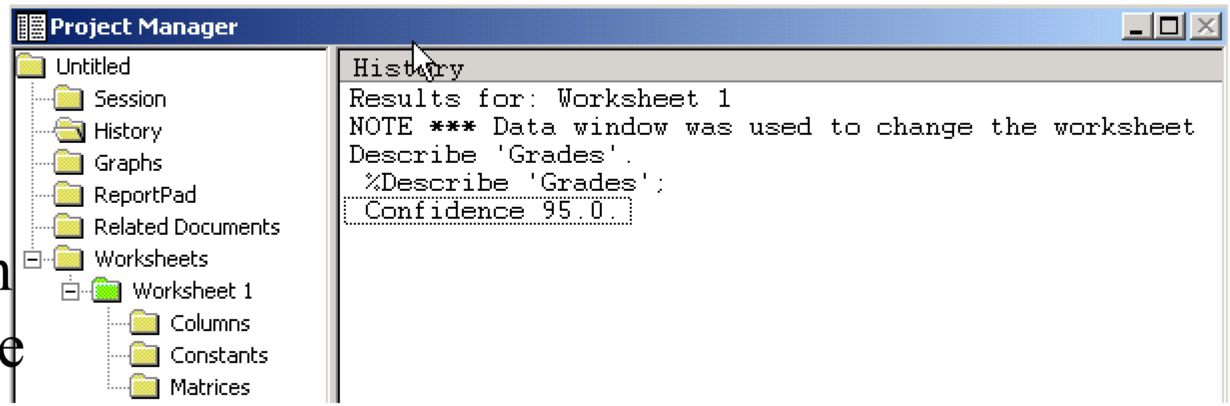
to make changes in analyses, graphs, transform equations,…

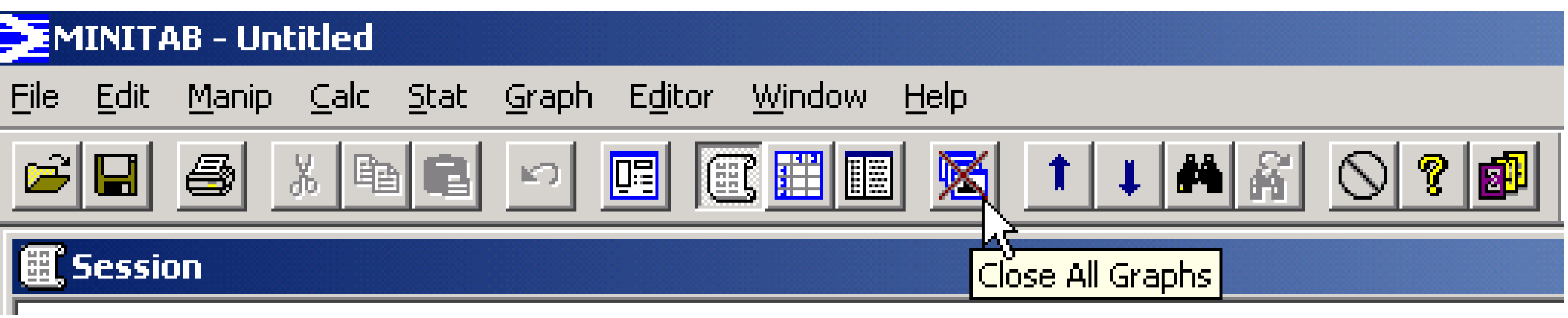
More helpful toolbar options:



Project Manager

to see all the files in the project in an organized structure

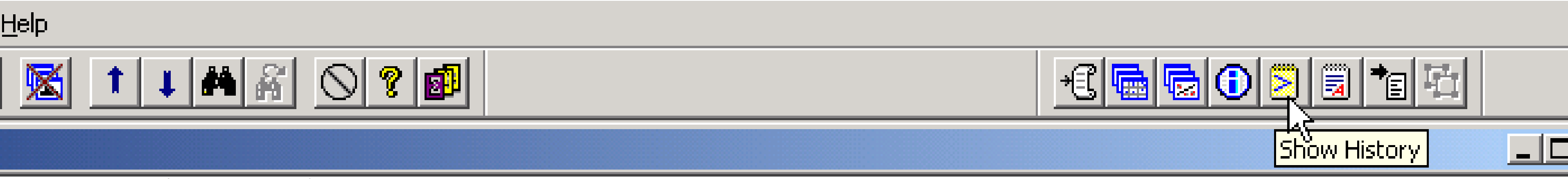




Close All Graphs

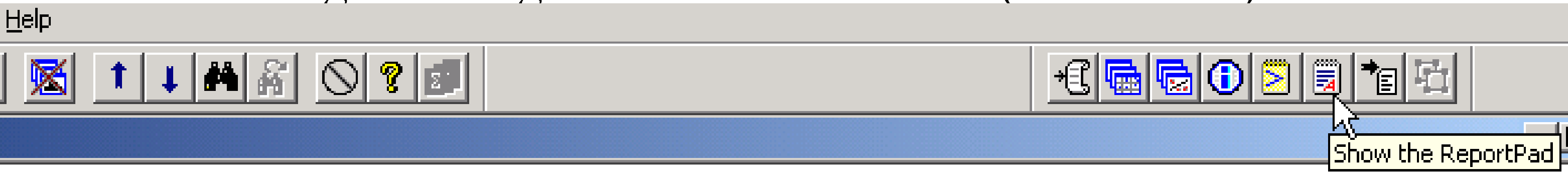
to clean up a bit

More helpful toolbar options:



Show History

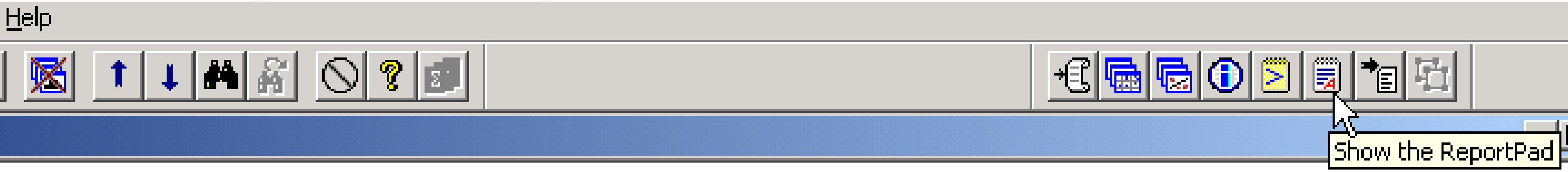
to review what you've done
OR
to use in generating an executable macro (".mtb" file)



Show the Report Pad

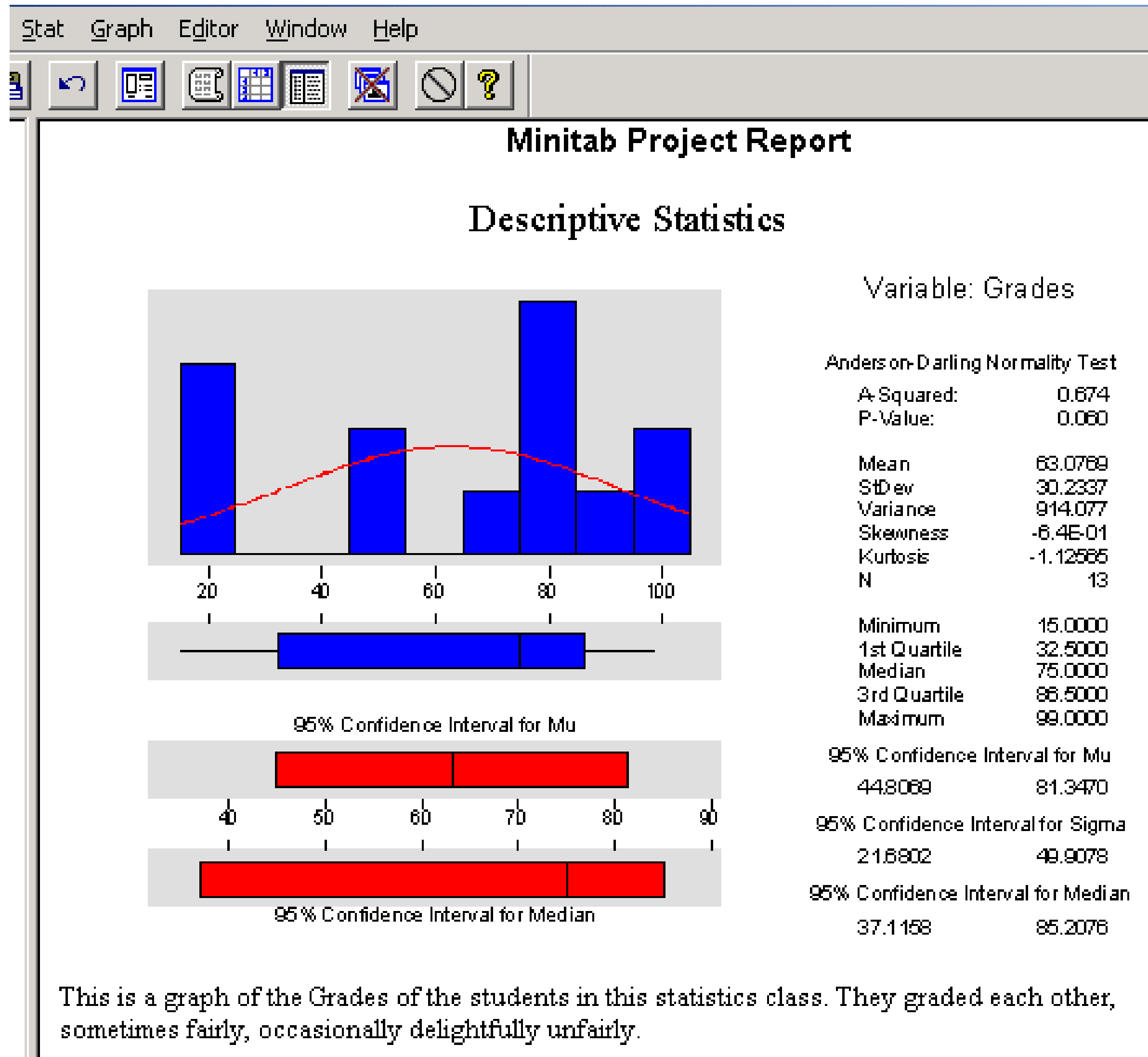
to bring up a convenient word processor within Minitab that allows you to combine graphs and text from Minitab with your own comments – to use in a report or paste into PowerPoint or …

More helpful toolbar options:



Show the Report Pad

to bring up a convenient word processor within Minitab that allows you to combine graphs and text from Minitab with your own comments – to use in a report or paste into PowerPoint or …



Distribution Parameters

- Mean
- Standard Deviation
- Variance
- Skewness
- Kurtosis

Measures of Location
(Central Tendency)

Measure	Symbol	Definition
Mean	$\overline{X}$	$\sum X \ / N$
Median	M	50% values $\geq$ M
		50% values $\leq$ M
Mode	Mo	Value or category with largest frequency

Note: $X_1, X_2, \ldots\ldots\ldots X_N$ are data points in sample of size N

Measures of Variability or Spread

Measure	Symbol	Definition
Range	R	$X_H - X_L$
Variance	S^2	$\frac{\sum (x_i - \bar{x})^2}{N - 1}$
Standard Deviation	S	$\sqrt{\frac{\sum (x_i - \bar{x})^2}{(N - 1)}}$

WHY USE $\frac{\sum (X - \bar{X})^2}{(n - 1)}$?

σ^2 = Population variance = $\frac{\sum (X - \mu)^2}{n}$

Desire to estimate this, using $\bar{x}$ instead of μ

(sample mean instead of population mean)

$$\sigma^2 = \sum \left[\frac{(X - \bar{x}) - (\mu - \bar{x})}{n}\right]^2 \}$$

$$\sigma^2 = \{ \sum \left[\frac{(X - \bar{x})}{n}^2 - \frac{2(X - \bar{x})(\mu - \bar{x})}{n} + \frac{(\mu - \bar{x})^2}{n}\right] \}$$

continued

Since $\sum (X - \bar{x}) = 0$, the middle term is zero:

$$\sigma^2 = \{ \sum \frac{[(X - \bar{x})^2}{n} + \frac{(\mu - \bar{x})^2]}{n} \}$$

The last term is the definition of standard error, or the variance of sample means about the population mean:

$$\text{Standard error} = \sum \frac{[(\mu - \bar{x})^2]}{n} = \frac{\sigma^2}{n}$$

continued

$$\sigma^2 = \sum \frac{(X - \bar{x})^2}{n} + \frac{\sigma^2}{n}$$

From the central limit theorem, standard error is equal to the population variance σ^2 / N; however, we don't know the population variance, and instead estimate it with s^2 / N

$$Est[\sigma^2] = s^2 = \sum \frac{(X - \bar{x})^2}{n} + \frac{s^2}{n}$$

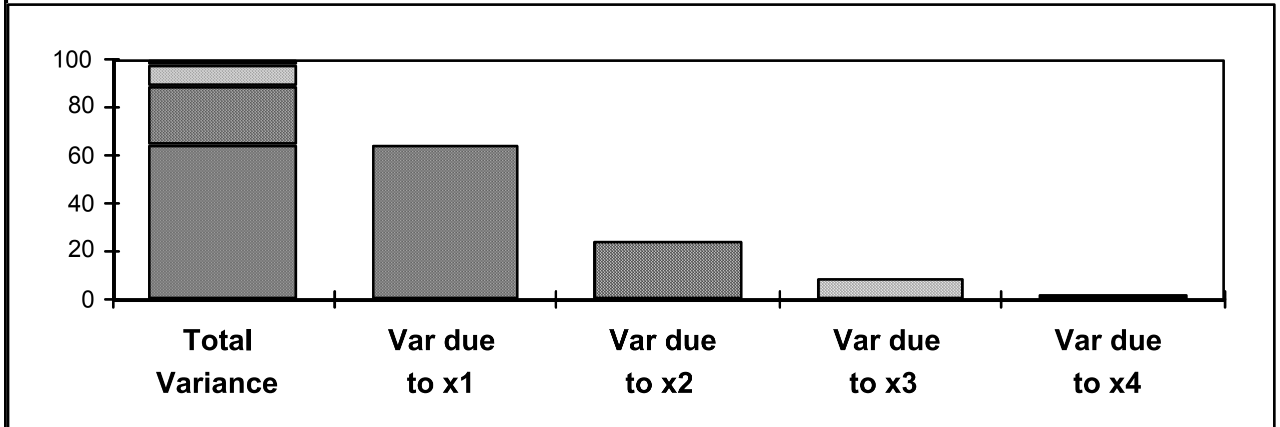
Multiply both sides by N, then combine s^2 terms:

$$s^2 (n - 1) = \sum (X - \bar{x})^2$$

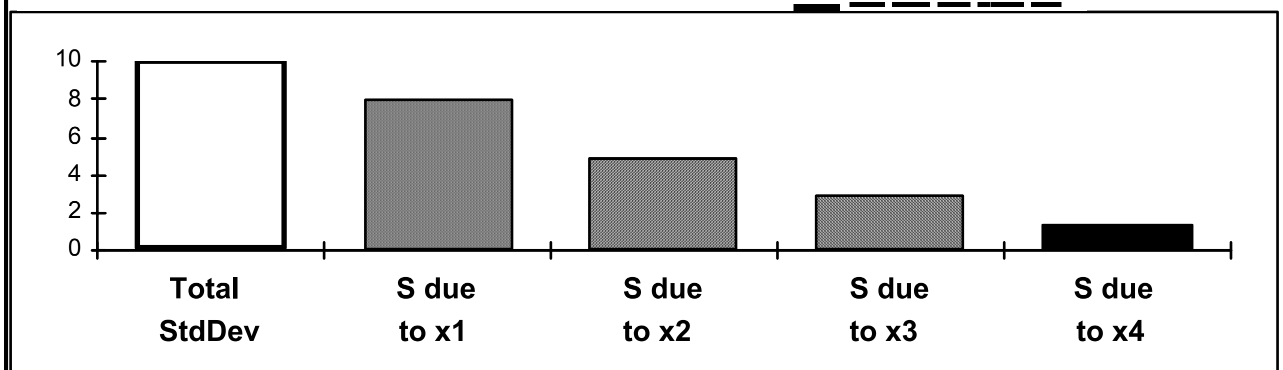
Divide both sides by (N - 1) to obtain sample variance

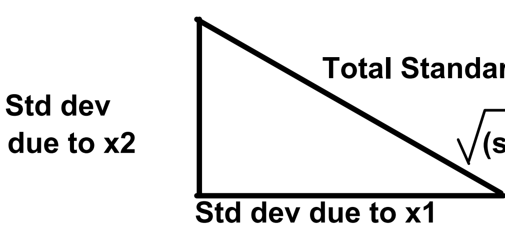
$$s^2 = \frac{\sum (X - \bar{x})^2}{(n - 1)}$$

Additivity Of Variances



Standard Deviations Combine





Total Standard deviation =

$$\sqrt{(\text{stddev due to x1})^2 + (\text{stddev due to x2})^2}$$

Moments of a Distribution

Kth Moment about T = $$\frac{\sum (x - T)^k}{n}$$

1st Moment about 0 = $$\frac{\sum (x - 0)^1}{n}$$

Mean

2nd Moment about mean = $$\frac{\sum (x - \bar{x})^2}{n}$$

Population VARIANCE

Skewness

The third moment about the mean is related to the asymmetry or skewness of a distribution:

$$\mu_3 = \frac{\sum (x - \bar{x})^3}{n}$$

A unimodal (i.e. a single peaked) distribution with $\mu_3 < 0$ is said to be skewed to the left — that is, it has a left "tail". If $\mu_3 > 0$, the distribution is skewed to the right. For symmetric distribution $\mu_3 = 0$

$$\text{Skewness} = \sqrt{\beta_1} = \frac{\mu_3}{s^3}$$

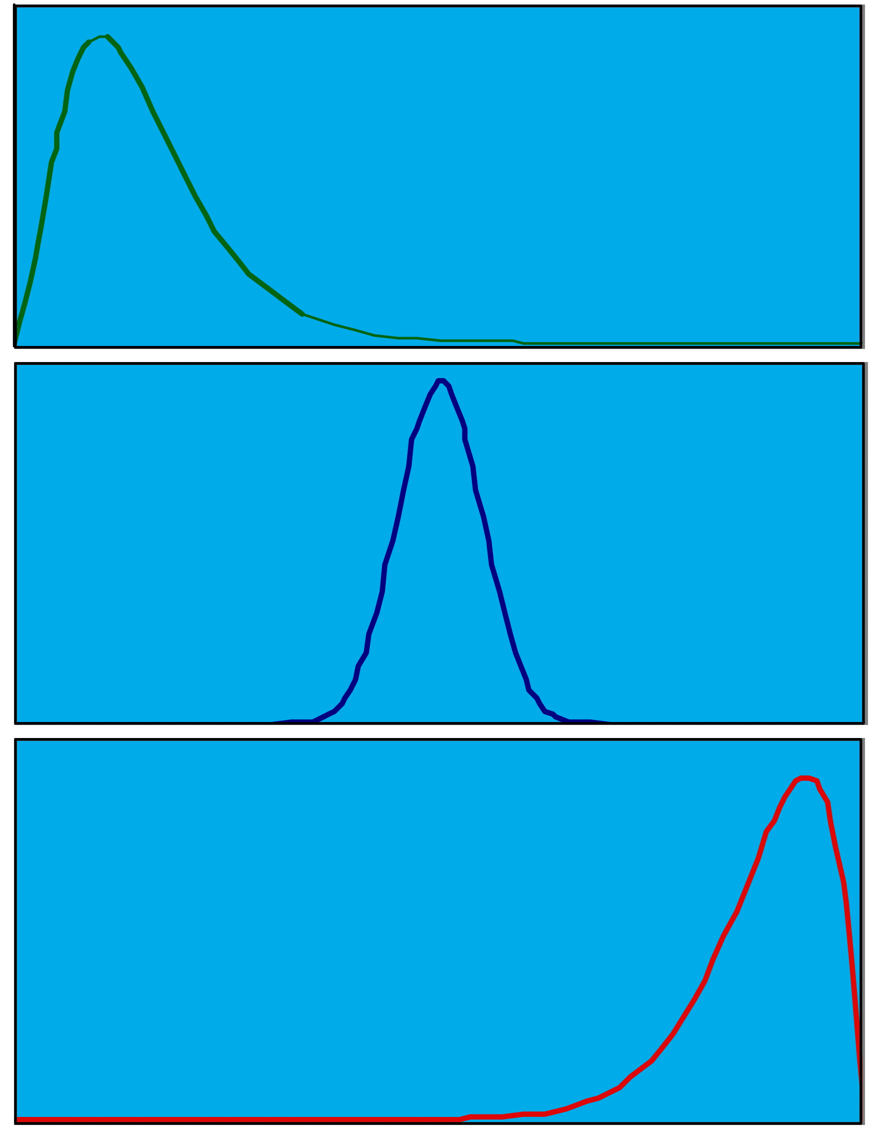
The quantity measures the skewness of the distribution relative to its degree of spread. This standardization allows us to compare the symmetry of two distributions whose scales of measurement differ.

SKEWNESS (CONT'D)

Positive Skew

Zero Skew

Negative Skew



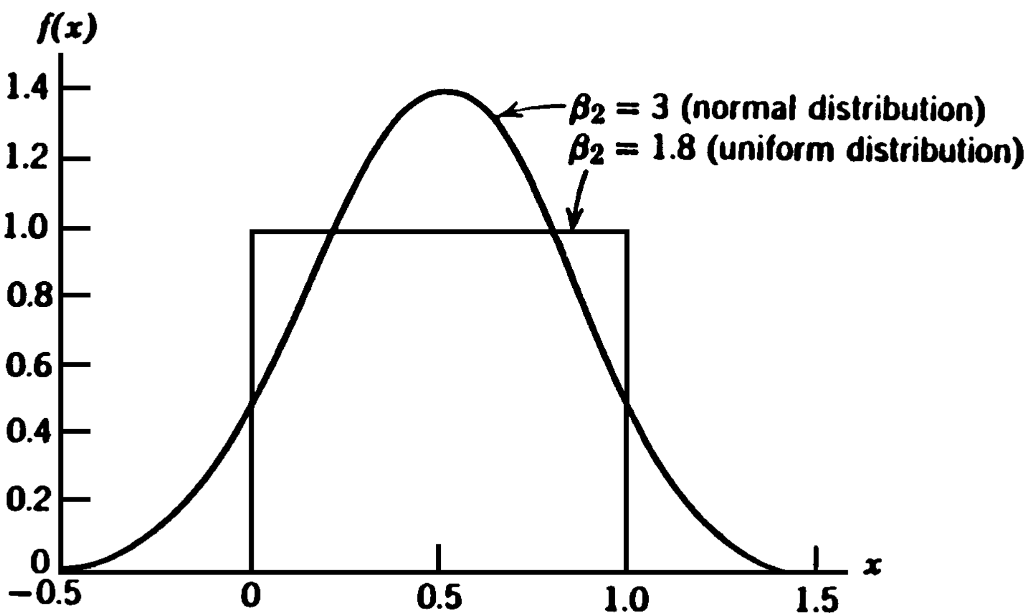
KURTOSIS

The fourth moment about the mean is related to the peakedness —also called kurtosis—of the distribution and is defined as

$$\mu_4 = \frac{\sum(x - \bar{x})^4}{n}$$

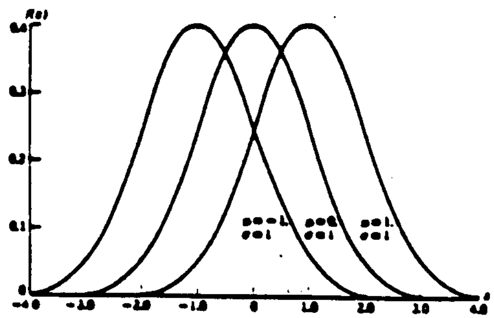
$$\text{Kurtosis} = \beta_2 = \frac{\mu_4}{s^4}$$

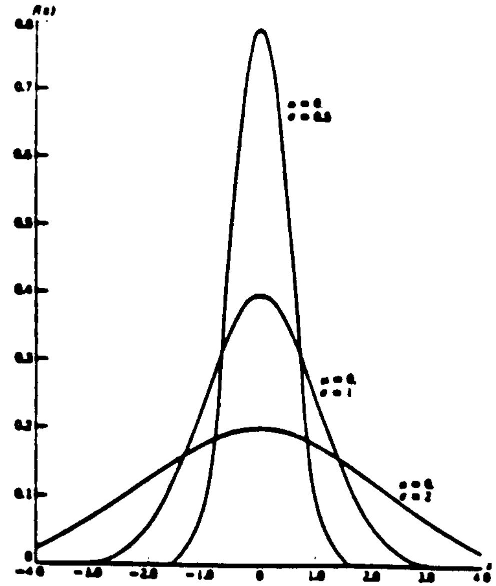
indicates how sharply the distribution reaches a peak. For example:



normal distribution. (β_2 = 3.0)
uniform distribution (β_2 = 1.8)

NORMAL DISTRIBUTION





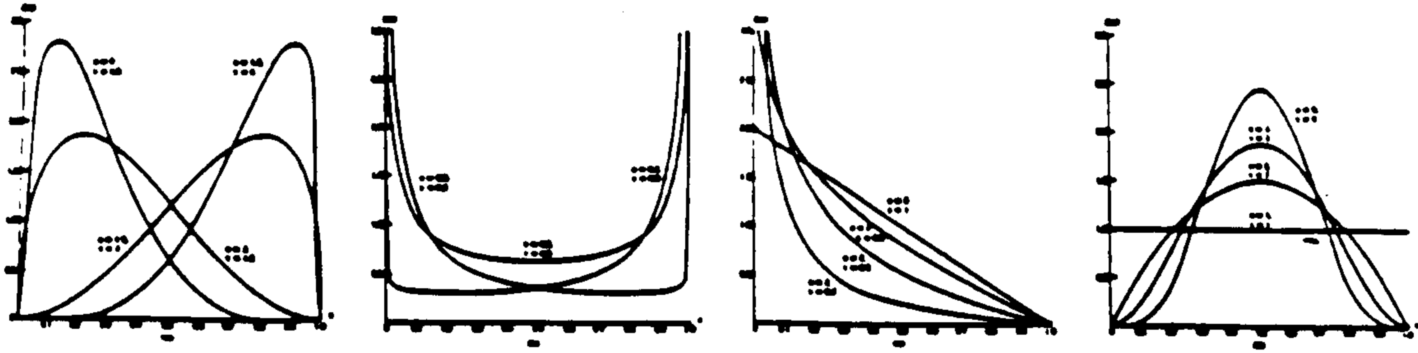
Parameters	Probability Density Function
$-\infty < \mu < \infty,$ $\sigma > 0$	$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{-(s-\mu)^{2}/2\sigma 2},$ $-\infty < x < \infty$

Expected Value	Variance	$\sqrt{\beta_1}$	β_2
μ	σ^2	0	3

Non-Normal Distributions

- Beta
- Gamma
- Exponential
- Log-normal
- Uniform

BETA DISTRIBUTION

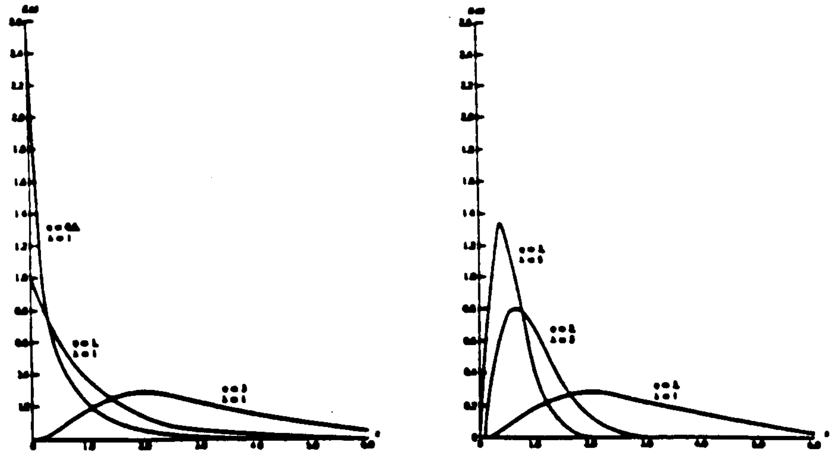


U J Symmetrical

Parameters	Probability Density Function
$\eta > 0, \gamma > 0$	$f(x) = \begin{cases} \dfrac{\Gamma(\eta+\gamma)}{\Gamma(\eta)\Gamma(\gamma)} x^{\gamma-1}(1-x)^{\eta-1}, & 0 \le x \le 1 \\ 0 & \text{elsewhere} \end{cases}$

Expected Value	Variance	β_1	β_2
$\dfrac{\gamma}{\eta+\gamma}$	$\dfrac{\eta\gamma}{(\eta+\gamma)^2(\eta+\gamma+1)}$	$*\sqrt{\beta_1} = \dfrac{2(n-\gamma)(\gamma+\eta+1)^{1/2}}{(n\gamma)^{1/2}(\gamma+\eta+2)}$	$\beta_2 = \dfrac{3(\eta+\gamma+1)[2(\gamma+\eta)^2+\eta\gamma(\eta+\gamma-6)]}{\eta\gamma(\eta+\gamma+2)(\eta+\gamma+3)}$

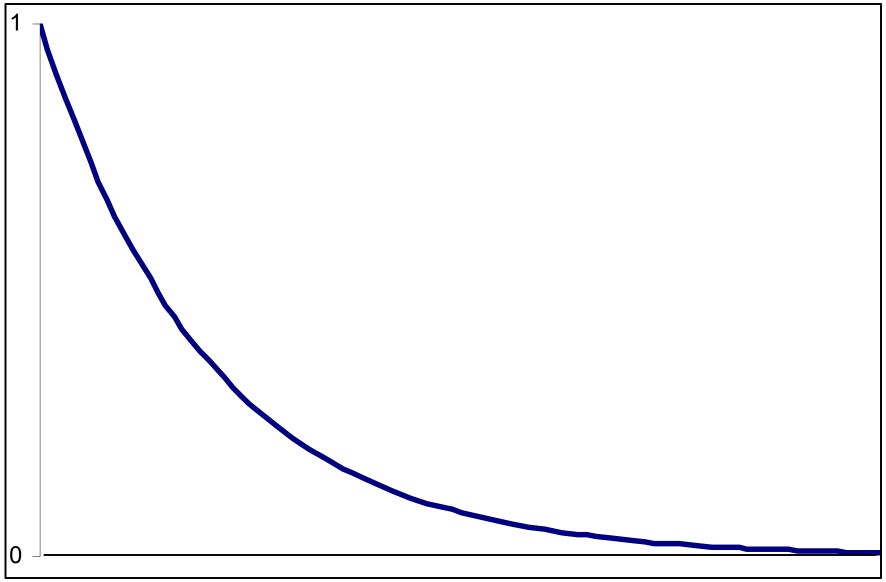
Gamma Distribution



Parameters	Probability Density Function
$\lambda > 0,$ $\eta > 0$	$f(\mathrm{x}) = \begin{cases} \dfrac{\lambda^{\eta}}{\Gamma(\eta)} \mathrm{x}^{\eta-1} e^{-\lambda \mathrm{x}}, \mathrm{x} \geq 0 \end{cases}$

Expected Value	Variance	β_1	β_2
$\dfrac{\eta}{\lambda}$	$\dfrac{\eta}{\lambda^2}$	$\dfrac{2}{\sqrt{\eta}}$	$\dfrac{3(\eta+2)}{\eta}$

Exponential Distribution

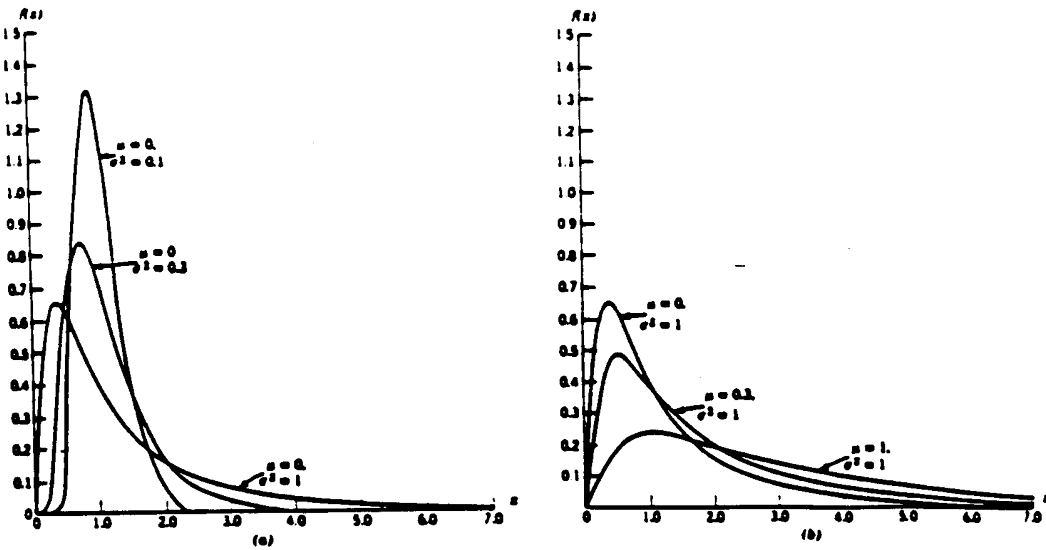


Parameters	Probability Density Function
$\lambda > 0$	$f(x) \begin{cases} = \lambda e^{-\lambda x}; x \geq 0 \\ = 0 \text{ elsewhere} \end{cases}$

Expected Value	Variance	β_1	β_2
$\frac{1}{\lambda}$	$\frac{1}{\lambda^2}$	2.0	9.0

$$\bar{x} \approx s$$

Log–Normal Distribution

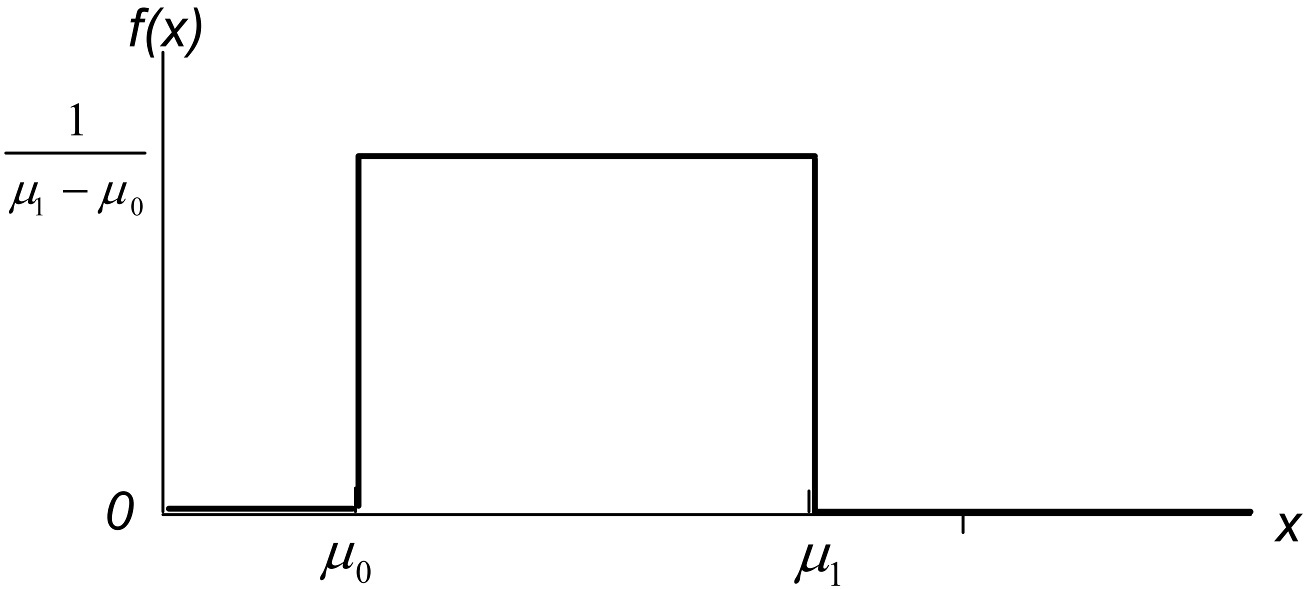


Parameters	Probability Density Function
$-\infty < \mu < \infty;$ $\sigma > 0$	$f(x) = \dfrac{1}{\sigma x\sqrt{2\pi}} \exp\left[-\dfrac{1}{2\sigma^2}(\log x - \mu)^2\right],$ $x \geq 0$

Expected Value	Variance	β_1	β_2
$e^{u+\frac{1}{2}\sigma^2}$	$e^{2\mu+\sigma^2}(e^{\sigma^2}-1)$	$(e^{\sigma^2}-1)^{\frac{1}{2}}(e^{\sigma^2}+2)$	

$\beta_2 = 3 + (w - 1)(w^3 + 3w^2 + 6w + 6);\ w = e^{\sigma^2}$

Uniform Distribution



Parameters	Probability Density Function
μ_0, μ_1, where $\mu_0 < \mu_1$	$f(x) = \begin{cases} \frac{1}{\mu_1 - \mu_0}, & \mu_0 \le x \le \mu_1 \\ 0 & \text{elsewhere} \end{cases}$

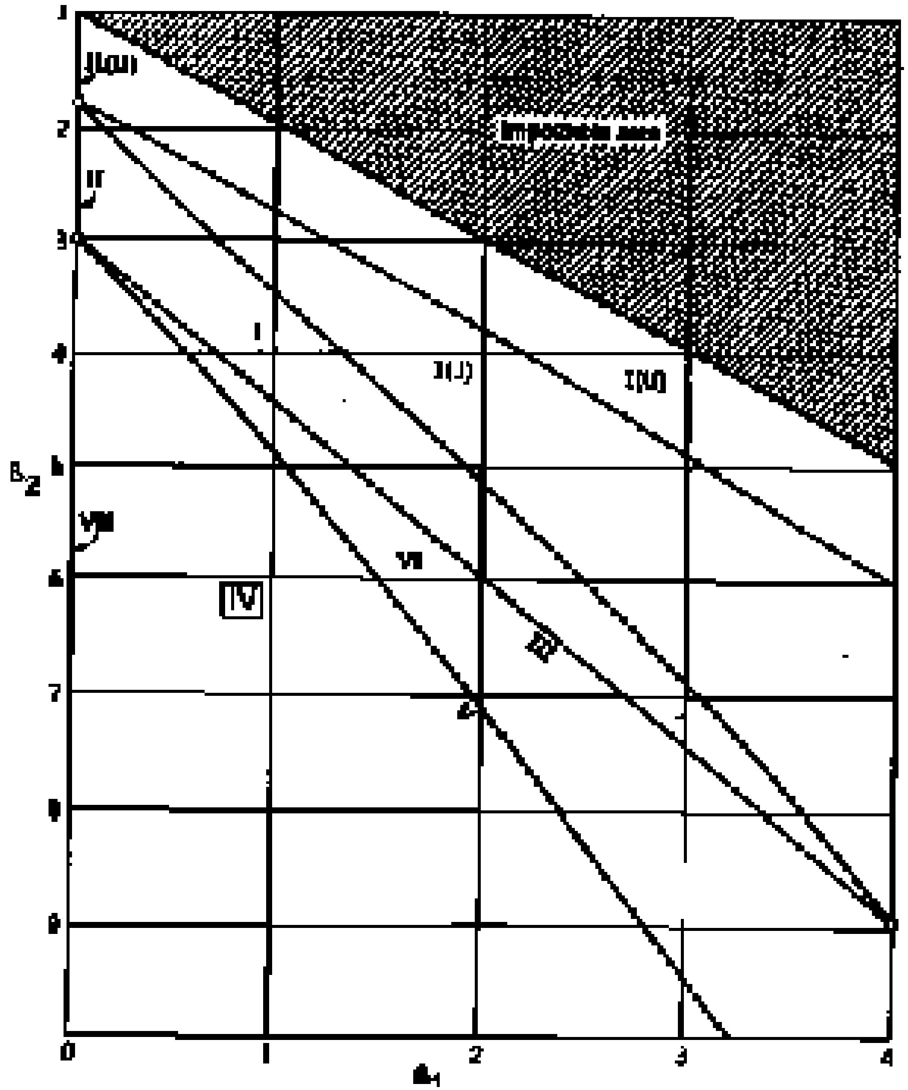
Mean	Variance	β_1	β_2
$\frac{\mu_0 + \mu_1}{2}$	$\frac{(\mu_1 - \mu_0)^2}{12}$	0	1.8

Pearson Distributions

$$\frac{df(x)}{dx} = \frac{(x - \phi_3)f(x)}{\phi_0 + \phi_1 x + \phi_2 x^2}$$

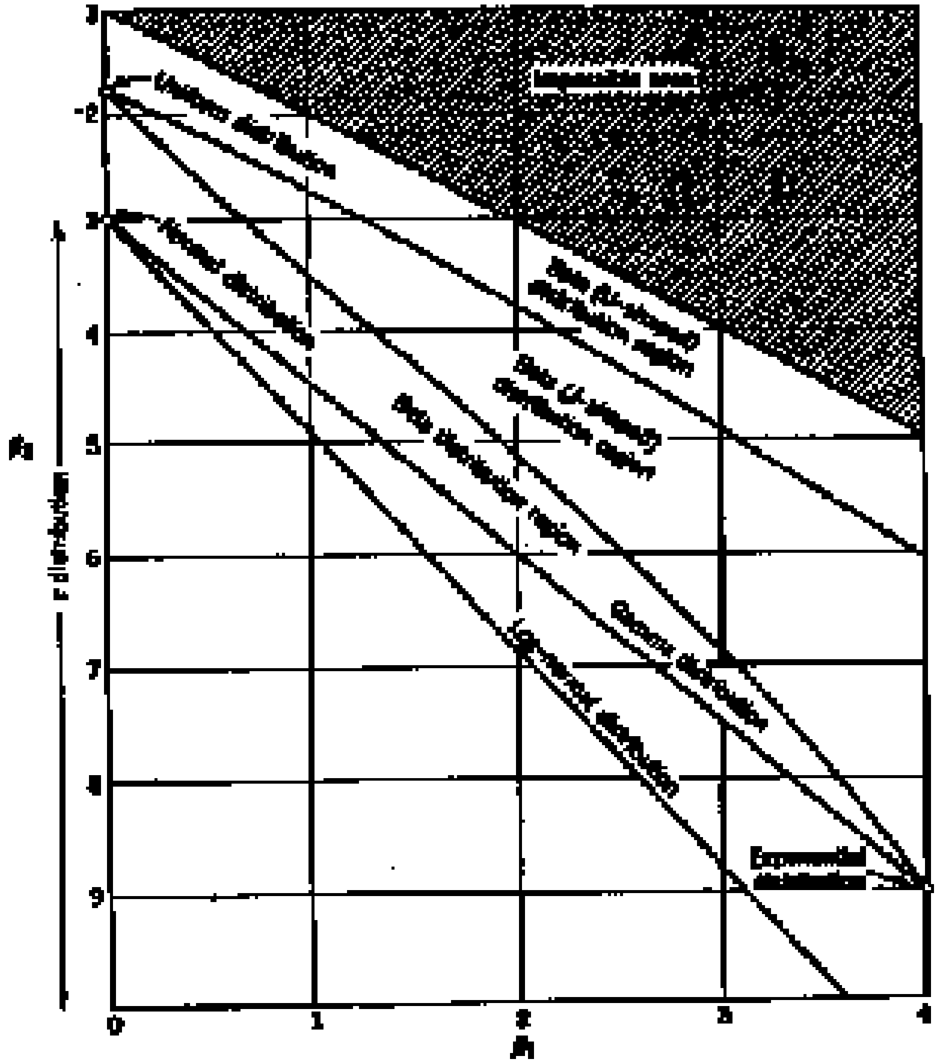
General equation for distributions.

PEARSON DISTRIBUTION

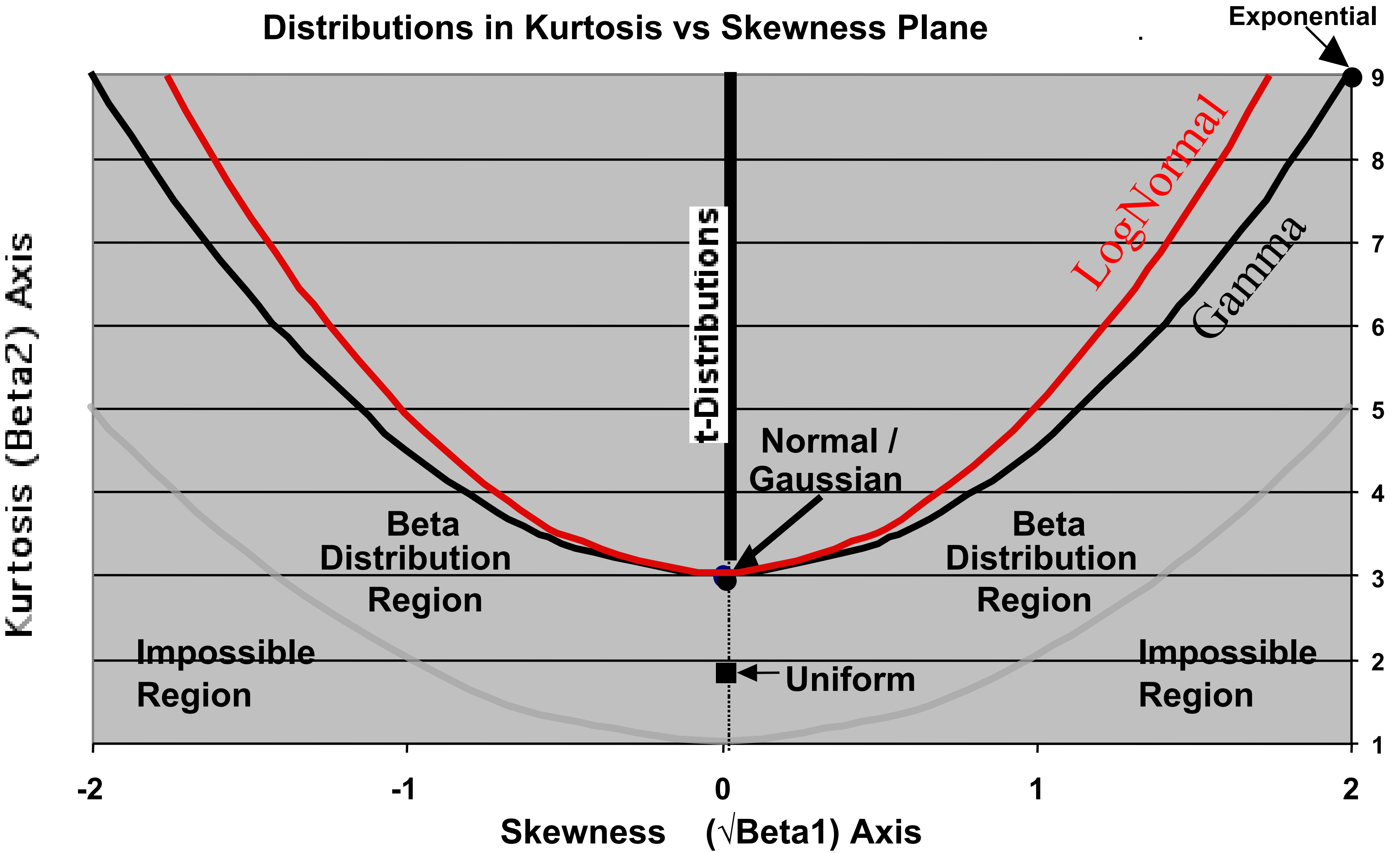


Region in (ß1,ß2) plane for various type Pearson distributions. Letters U and J denote U-shaped and J-shaped distributions. (from E.S. Pearson, Seminars, Princeton University, 1960.)

Regions of Distributions



Regions in (ß1,ß2) plane for various distributions. (From Professor E.S. Pearson, University College, London.)

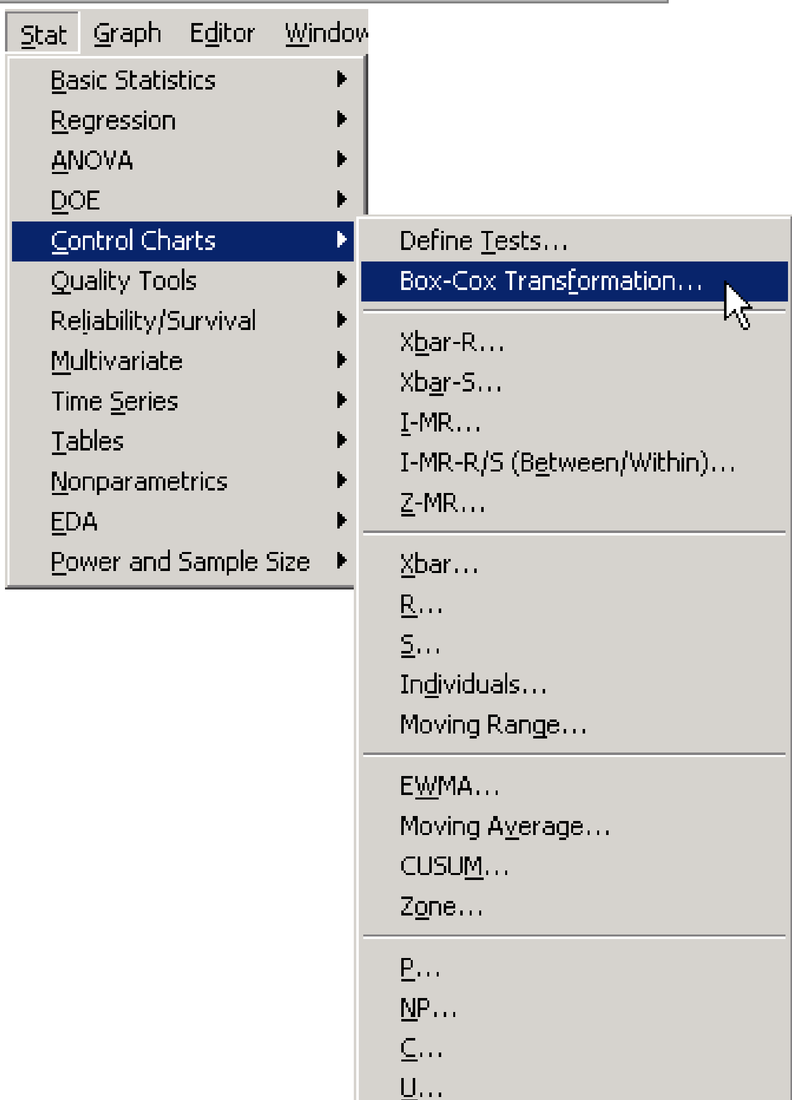


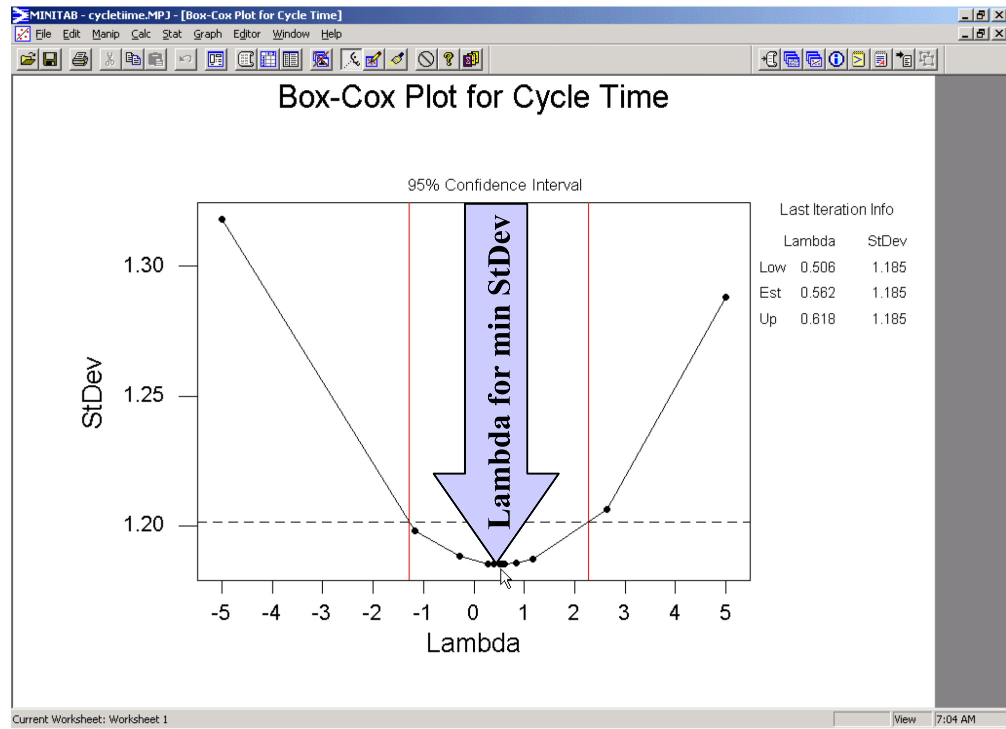
Distributions in Kurtosis vs Skewness Plane
Exponential
LogNormal
Gamma
t-Distributions
Normal / Gaussian
Beta Distribution Region
Beta Distribution Region
Impossible Region
Impossible Region
Uniform
Kurtosis (Beta2) Axis
Skewness (√Beta1) Axis
-2
-1
0
1
2
1
2
3
4
5
6
7
8
9

Box-Cox Transformation

Minitab's Box-Cox transformation can help determine how to transform data that doesn't follow a normal distribution to more nearly approximate normality.

File / Open "CYCLETIME.MPJ"





Depending on the lambda where the std dev reaches a minimum, The following transformations are suggested for the data:

Lambda	Transformation
-2.0	Inverse Square
-1.0	Inverse
-0.5	Inverse Square Root
0.0	Log (either Natural or Base 10)
0.5	Square Root
1.0	Untransformed
2.0	Square

Excel or other Files → Minitab

For Excel or other spreadsheet:
Data should be in the format of
different columns for different parameters,
different rows for different values of the same parameters.

Three primary ways to import data:

1) **Open Worksheet, select the type as "Excel" or another acceptable format.**

2) **Copy the data (including labels) to import from Excel or other table or worksheet.**

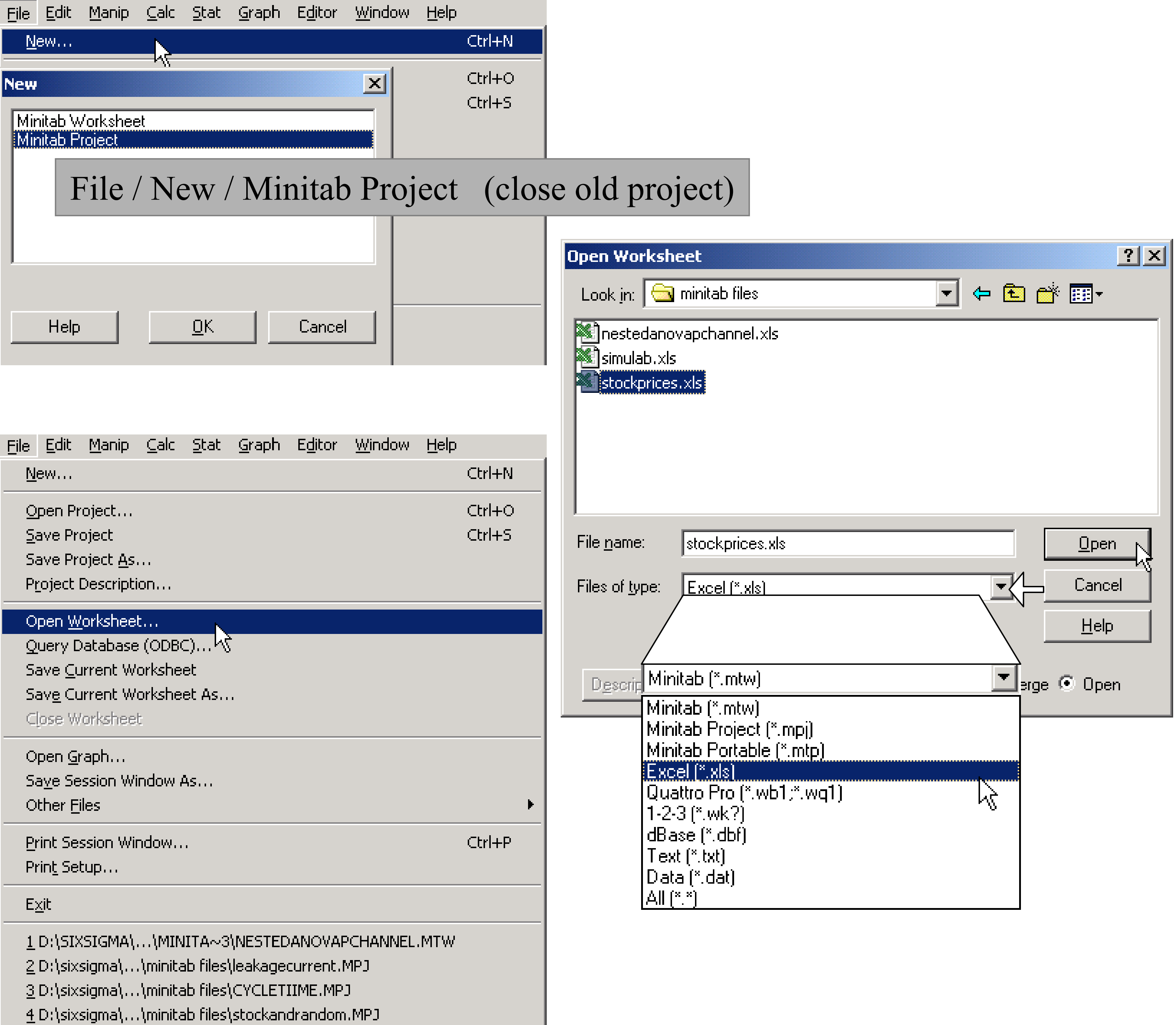
Go to the Minitab Worksheet page

If there are labels included, click on a label cell in the Minitab worksheet and paste.

If there are no labels, only data, click on a value cell below the label cell and paste.

Then, type in labels for each column.

3) **Go to the file menu on Minitab and choose the "import file" option.**
If the file format is consistent with one of the options Minitab supports,
select that format and import the file into Minitab.


File Edit Manip Calc Stat Graph Editor Window Help
New... Ctrl+N
Ctrl+O
Ctrl+S
New
Minitab Worksheet
Minitab Project
Help
OK
Cancel
File / New / Minitab Project (close old project)
Open Worksheet
Look in: minitab files
nestedanovapchannel.xls
simulab.xls
stockprices.xls
File name: stockprices.xls
Files of type: Excel (*.xls)
Open
Cancel
Help
Minitab (*.mtw)
Minitab (*.mtw)
Minitab Project (*.mpj)
Minitab Portable (*.mtp)
Excel (*.xls)
Quattro Pro (*.wb1;*.wq1)
1-2-3 (*.wk?)
dBase (*.dbf)
Text (*.txt)
Data (*.dat)
All (*.*)
Open
File Edit Manip Calc Stat Graph Editor Window Help
New... Ctrl+N
Open Project... Ctrl+O
Save Project Ctrl+S
Save Project As...
Project Description...
Open Worksheet...
Query Database (ODBC)...
Save Current Worksheet
Save Current Worksheet As...
Close Worksheet
Open Graph...
Save Session Window As...
Other Files
Print Session Window... Ctrl+P
Print Setup...
Exit
1 D:\SIXSIGMA\...\MINITA~3\NESTEDANOVAPCHANNEL.MTW
2 D:\sixsigma\...\minitab files\leakagecurrent.MPJ
3 D:\sixsigma\...\minitab files\CYCLETIIME.MPJ
4 D:\sixsigma\...\minitab files\stockandrandom.MPJ

Cleaning up data for analysis

*** When you extract data from databases, test systems.... - there often are "Outliers"**
*** Outliers are data points that don't fit with the predominant range of values for the rest of the data.**

1) Outliers shouldn't simply be tossed out without considering what they might be telling you.
<u>Sometimes, the outliers are the REAL story</u>
Example: A study with bacteria colonies being grown on petri dishes...one petri dish didn't fit in.
It turned out to be contaminated with mold. Result: the discovery of penicillin, our first antibiotic!

2) Outliers are sometimes individual values, and sometimes groups or clumps of values
- **Individual outliers are generally due to something being different - measurement problem, misprocessing...**
- **With some distributions (ex: lognormal), some values that look like outliers can actually be part of the distribution**
- **Clumps of outliers can be due to:**
    - *** Measurement problems, misprocessing**
    - *** "Pegging" on some sort of limit (testing limit, short, open...)**
    - *** Bimodal distribution (or trimodal, or...), which can be due to:**
        - **Different processes, tools, measurement systems ...**
        - **Nested variance: different lots, different wafers**
        - **Different populations (example: EU vs USA GSM)**
        - **U-shaped Beta distribution (peaks near 0% and 100%)**

Rules of Thumb:

a) Look at the histogram of the data. Where do the outliers lie? What does the distribution without the outliers look like?
b) What might the outliers be telling you? (see list 2 above)
c) If you are ready to discard the outliers, come up with a rational rule as to what is and what is not an outlier to discard.
- *** If the main distribution resembles a normal distribution, then may be able to use a "recursive filter" approach:**
    - **Obtain mean, std dev of entire data set with outliers, discard any points outside mean +/- 6 std dev; repeat as appropriate**
- *** For non-normal distributions, you may need to first use a transform (example: distribution of logarithms of values)**

CENTRAL LIMIT THEOREM

1) Standard Deviation of Means = $\frac{\text{Standard Deviation of Population}}{\text{SQRT (N)}}$

where N = Sample size for determining means

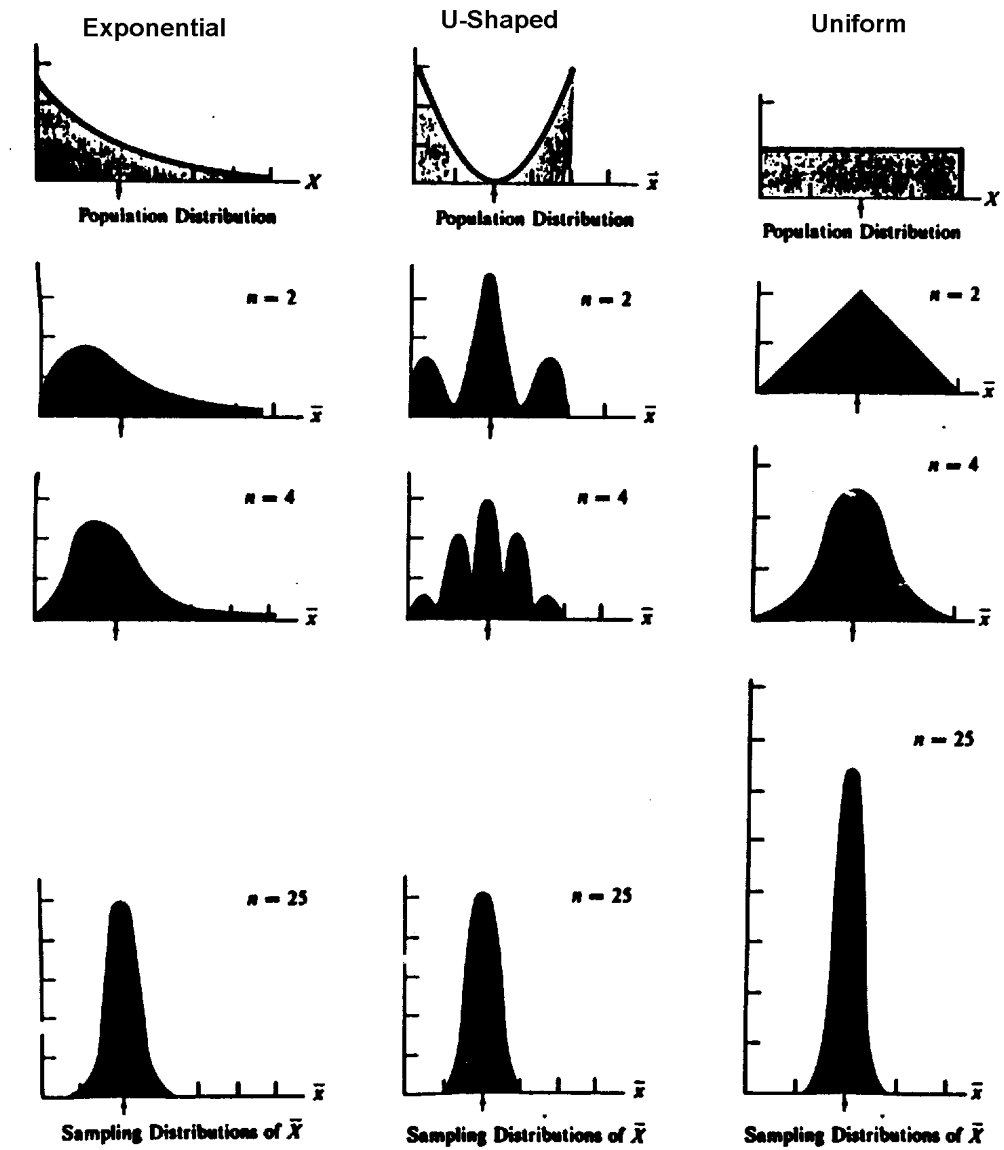
Note: used in ANOVA and T-Test Theory

2) Regardless of the shape of the distribution of the population,

the distribution of means will be more nearly normal.

Larger Sample Size → More nearly Normal Distribution of Means

Central Limit Theorem Effect on Non–Normal Distributions



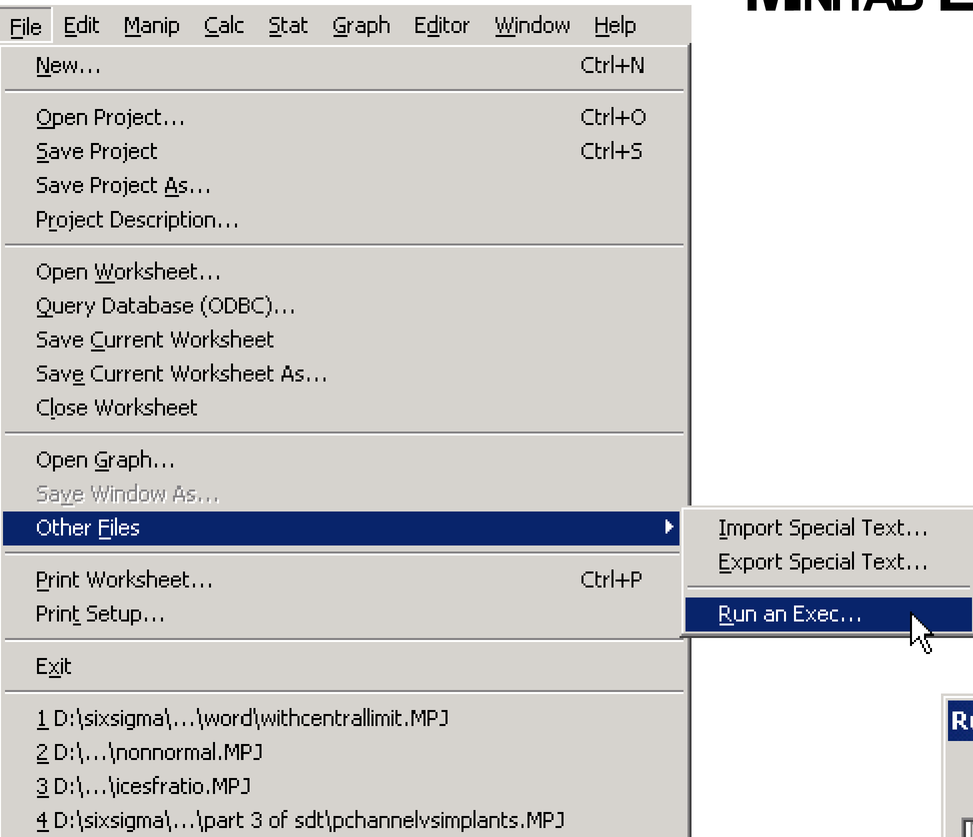
Minitab Execs / Macros

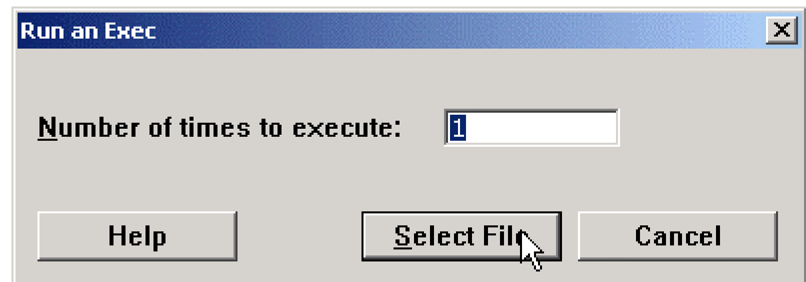
```
random 100 c1-c16;
normal 5 1.

Name c16 'normal'
name c17 'average 4'
name c18 'average 9'
name c19 'average 16'

let c17=(c1+c2+c3+c4)/4
let c18=(c1+c2+c3+c4+c5+c6+c7+c8+c9)/9
let c19=(c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c14+c15+c16)/16

% Describe C16,c17,c18,c19;
Confidence 95.0.
```

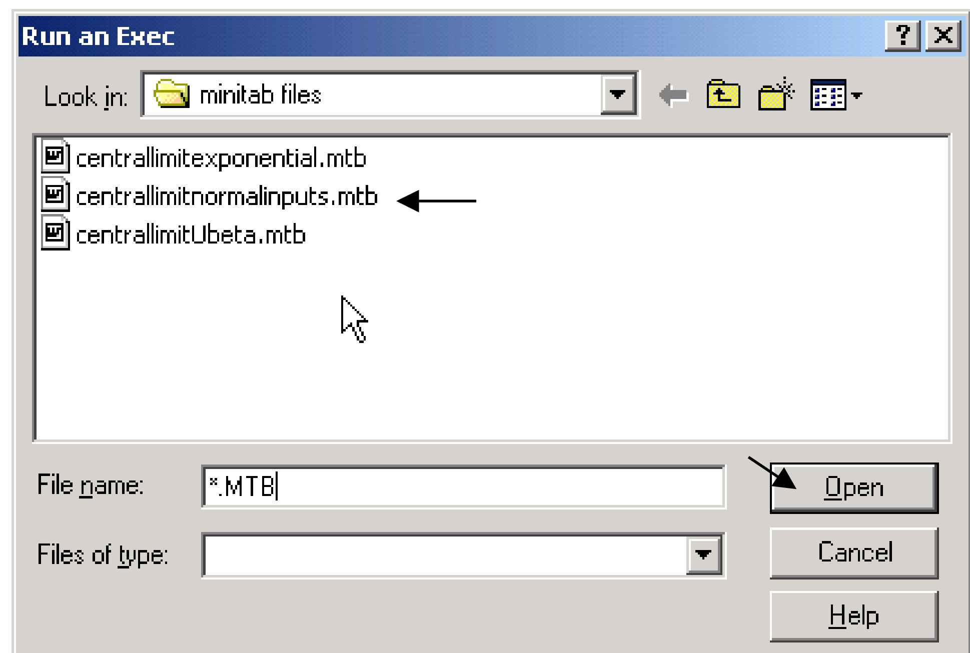
```
random 100 c1-c16;
exponential 5.

Name c16 'exponential'
name c17 'average 4'
name c18 'average 9'
name c19 'average 16'

let c17=(c1+c2+c3+c4)/4
let c18=(c1+c2+c3+c4+c5+c6+c7+c8+c9)/9
let c19=(c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c14+c15+c16)/16

% Describe C16,c17,c18,c19;
Confidence 95.0.
```

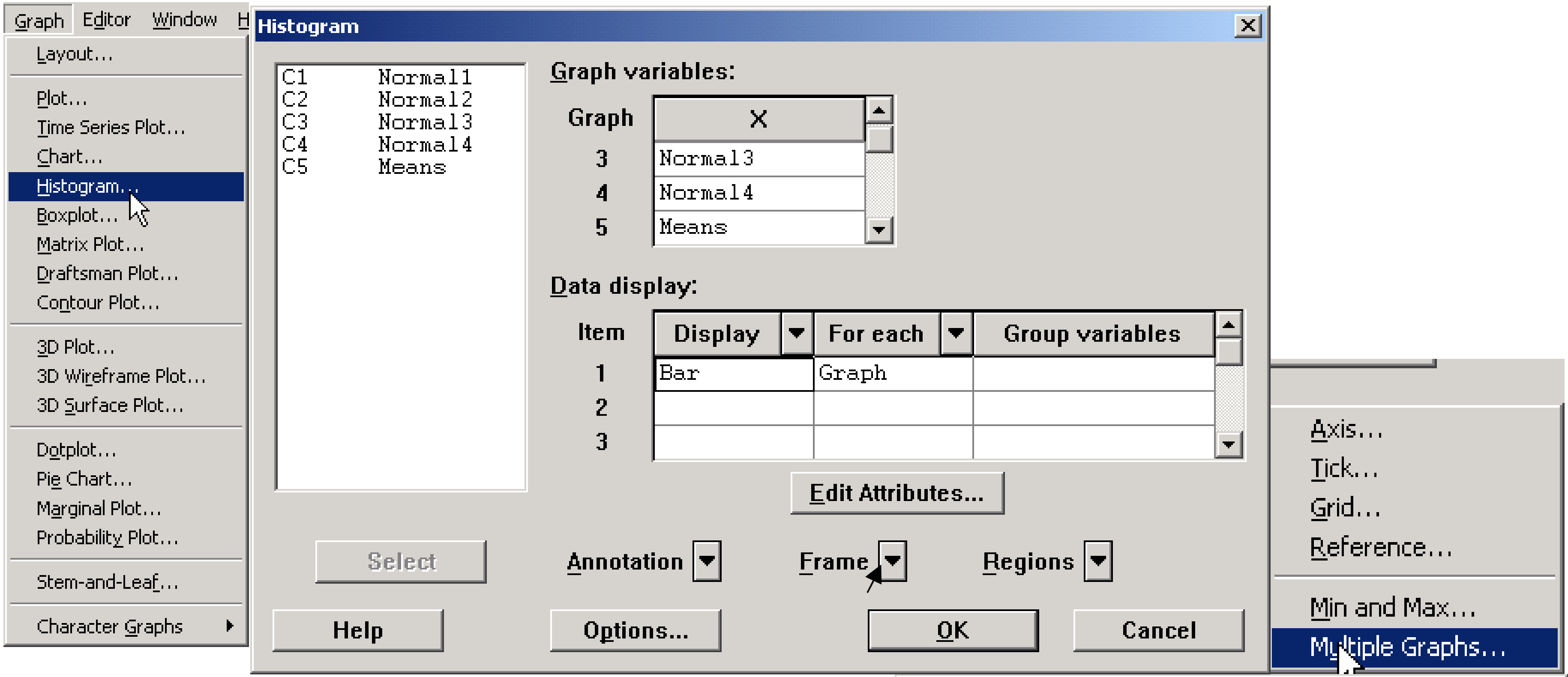
```
random 1000 c1-c16;
beta .5 .5.

Name c16 'beta U-shaped'
name c17 'average 4'
name c18 'average 9'
name c19 'average 16'
name c20 'average 2'

let c20=(c1+c2)/2
let c17=(c1+c2+c3+c4)/4
let c18=(c1+c2+c3+c4+c5+c6+c7+c8+c9)/9
Let c19=(c1+c2+c3+c4+c5+c6+c7+c8+c9+c10+c11+c12+c13+c14+c15+c16)/16

%Describe c16,c20,c17,c18,c19;
Confidence 95.0.
```

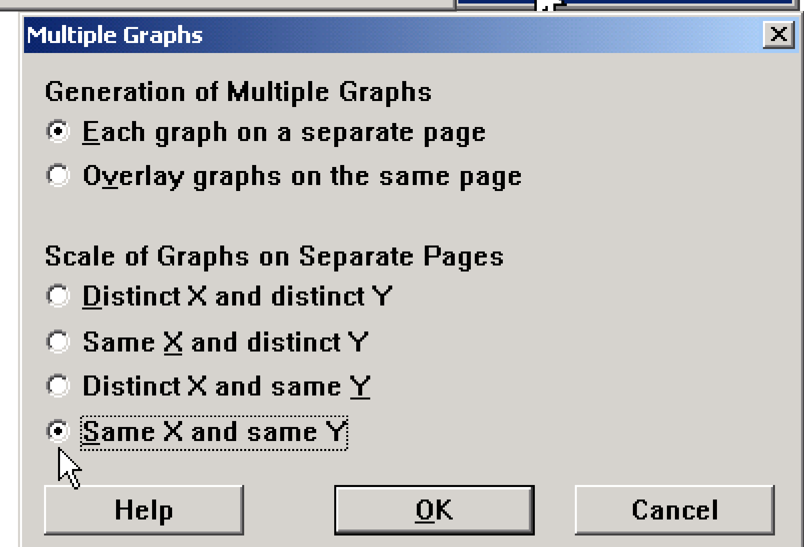
Minitab Example: Central Limit Theorem Normal Distribution



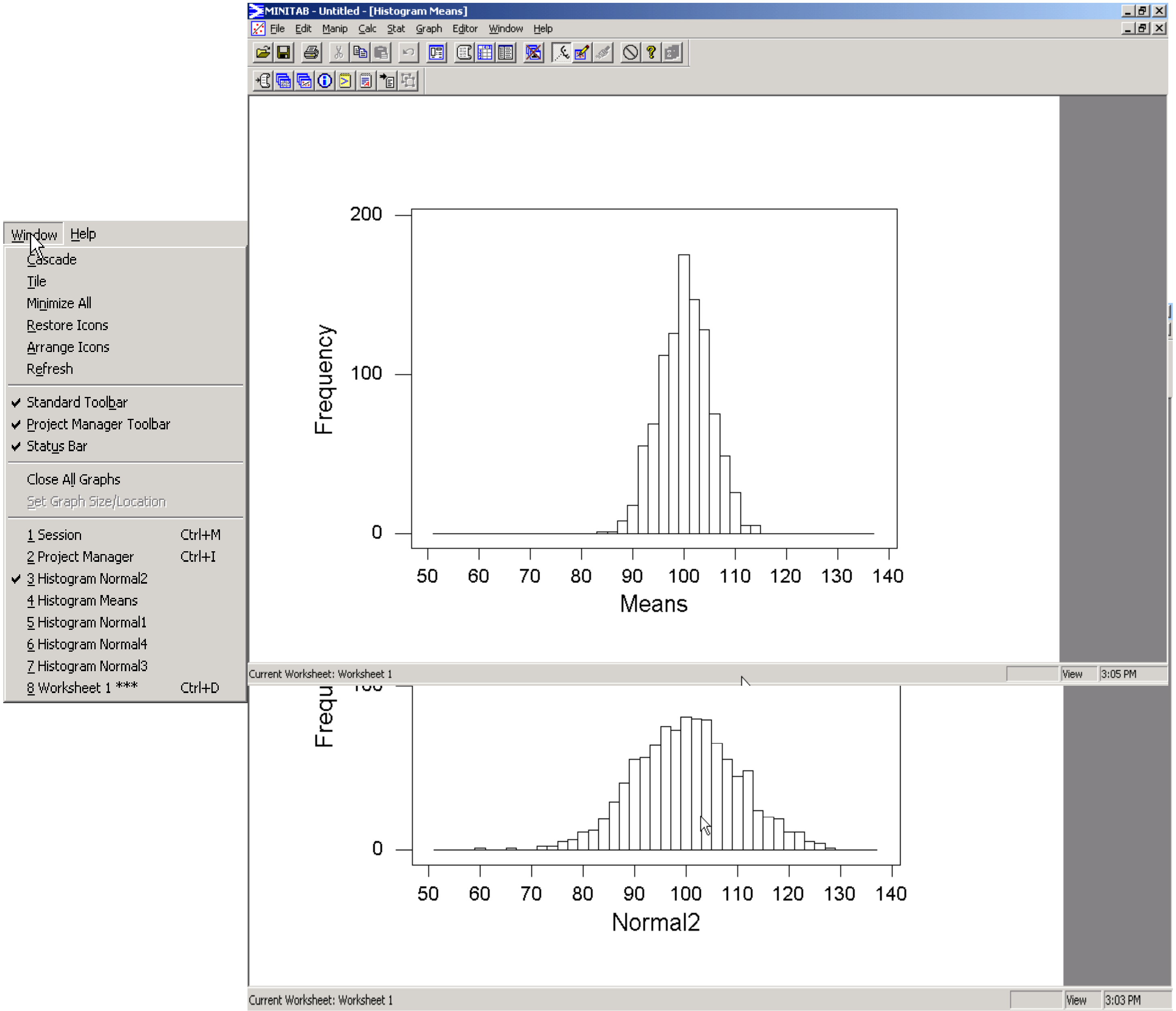






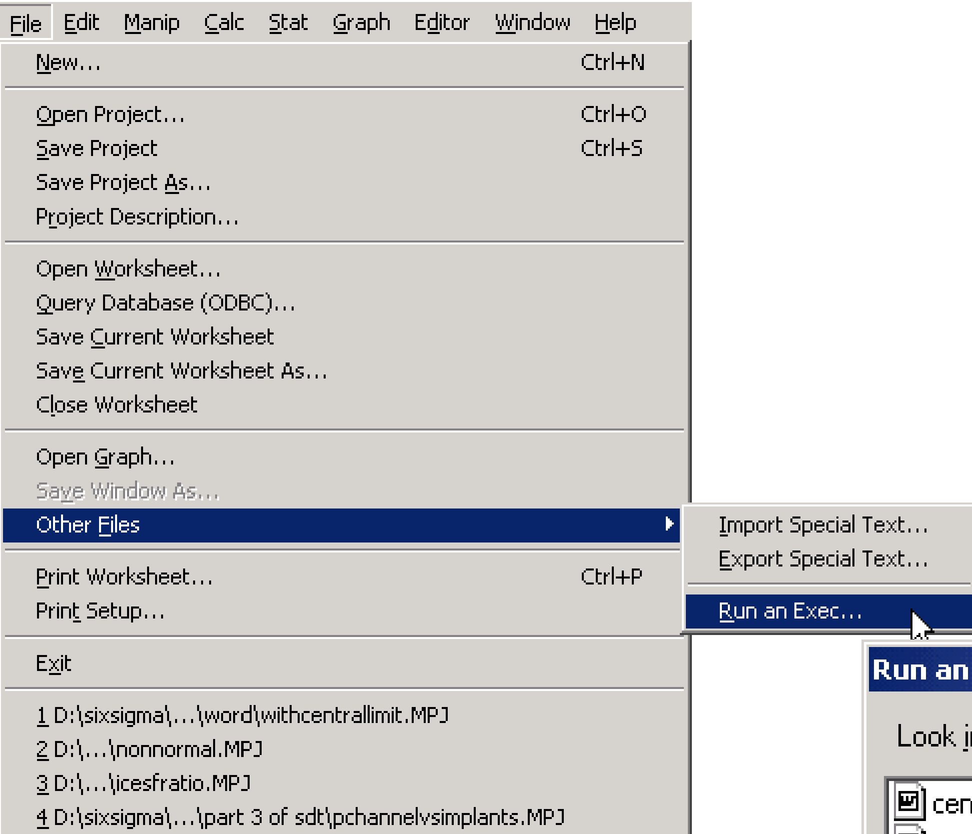


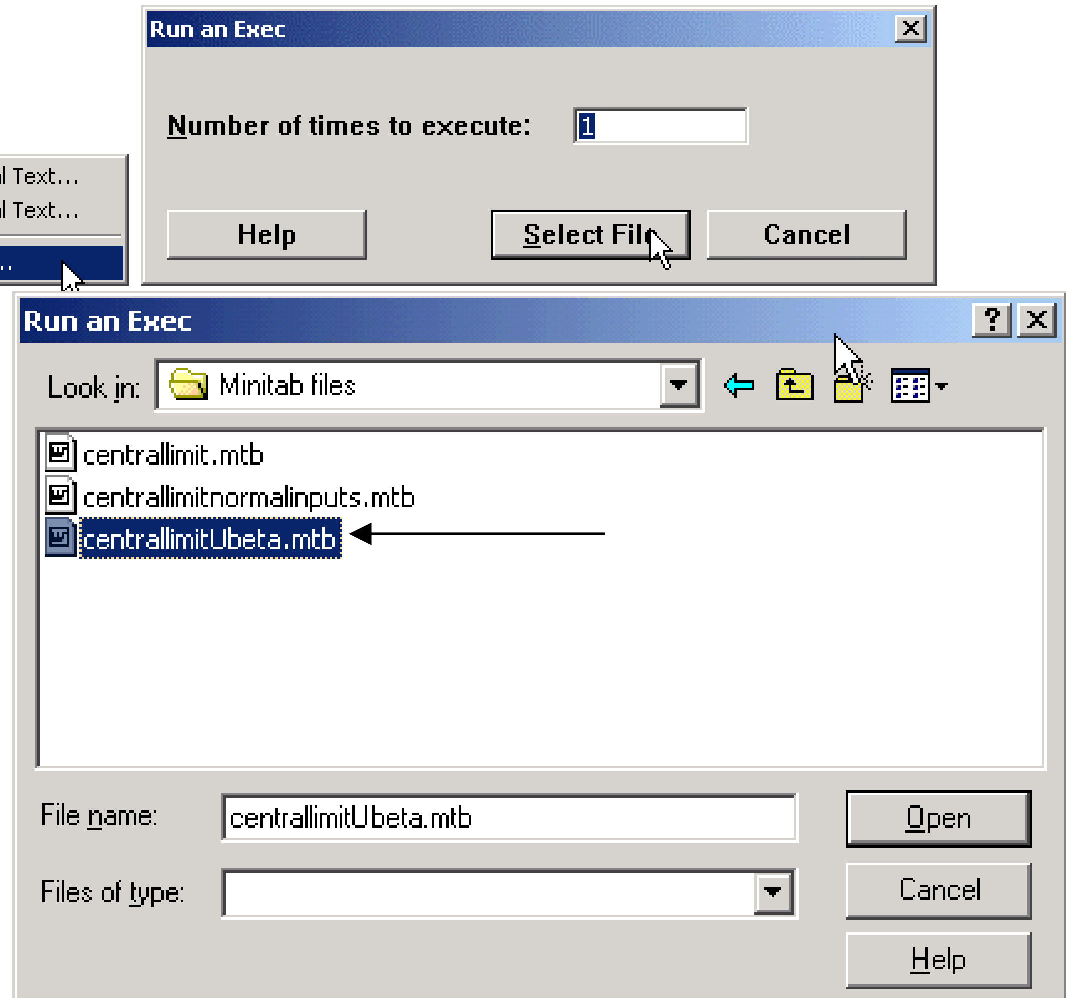
Select one of the random number columns and columns for averages of sample size 4, 9 and 16.

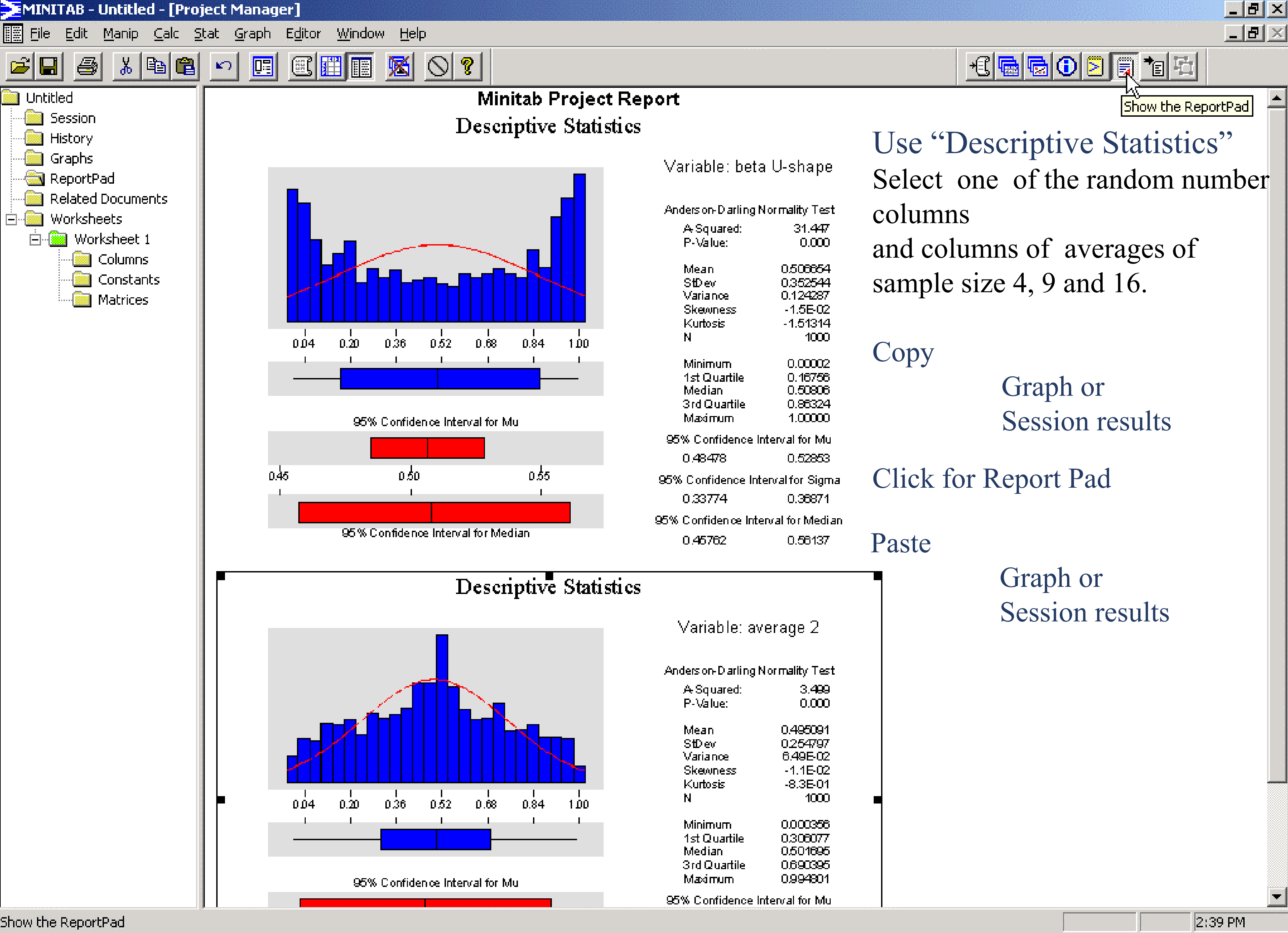

MINITAB - Untitled - [Histogram Means]
File Edit Manip Calc Stat Graph Editor Window Help
Window Help
Cascade
Tile
Minimize All
Restore Icons
Arrange Icons
Refresh
Standard Toolbar
Project Manager Toolbar
Status Bar
Close All Graphs
Set Graph Size/Location
1 Session Ctrl+M
2 Project Manager Ctrl+I
3 Histogram Normal2
4 Histogram Means
5 Histogram Normal1
6 Histogram Normal4
7 Histogram Normal3
8 Worksheet 1 *** Ctrl+D
Frequency
200
100
0
50 60 70 80 90 100 110 120 130 140
Means
Current Worksheet: Worksheet 1
View
3:05 PM
Frequ
0
50 60 70 80 90 100 110 120 130 140
Normal2
Current Worksheet: Worksheet 1
View
3:03 PM

MINITAB

U-Beta Demonstration







Use “Descriptive Statistics”

Select one of the random number columns
and columns of averages of sample size 4, 9 and 16.

Copy

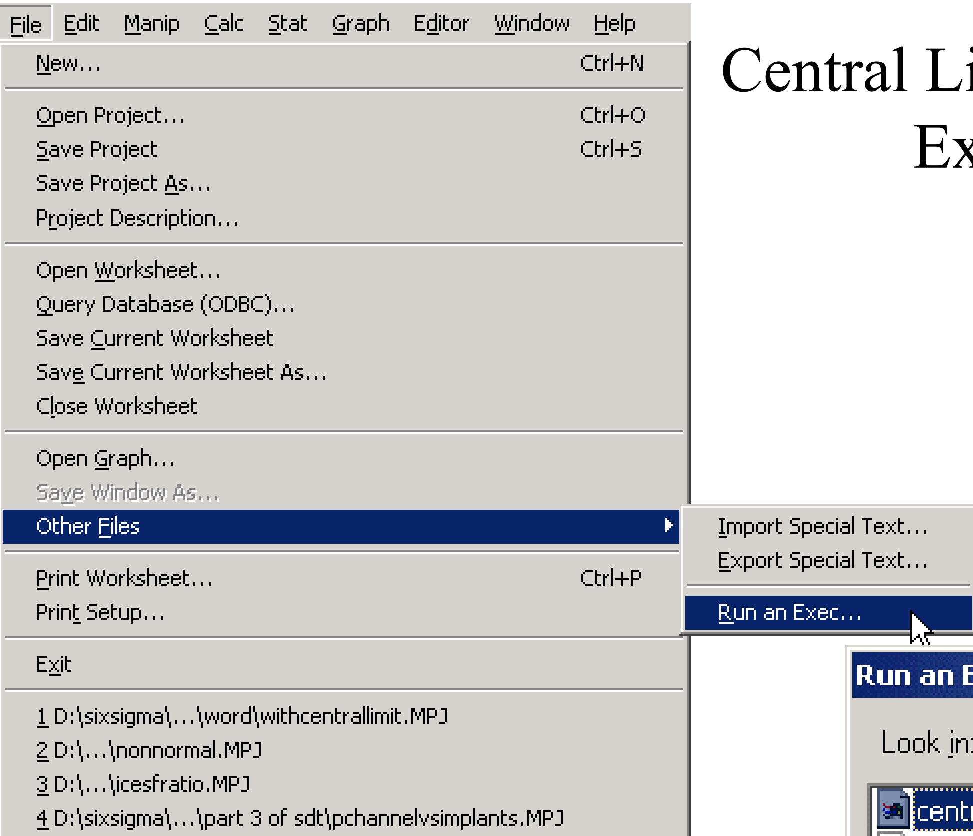
Graph or
Session results

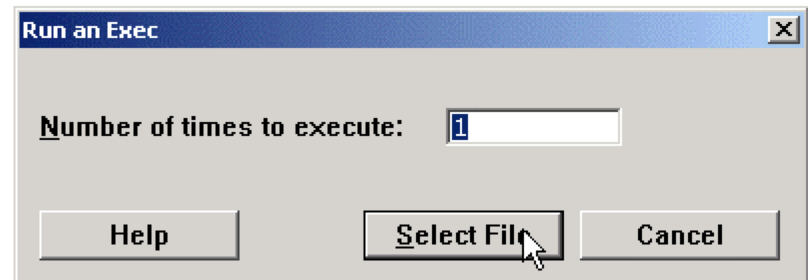
Click for Report Pad

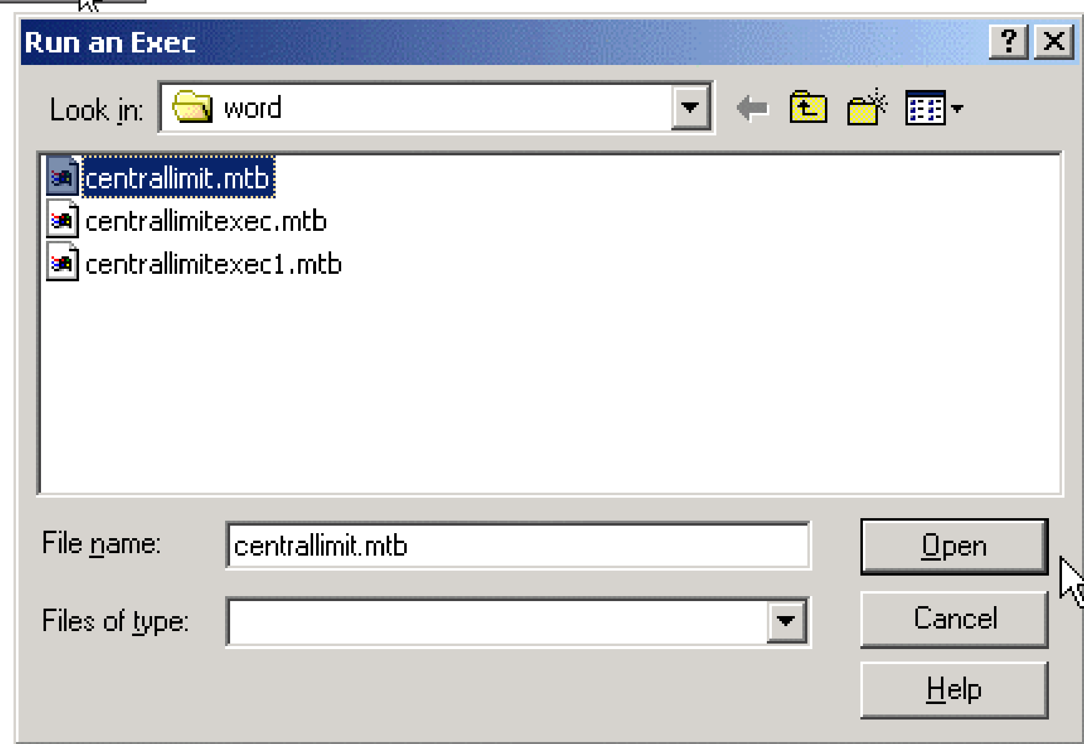
Paste

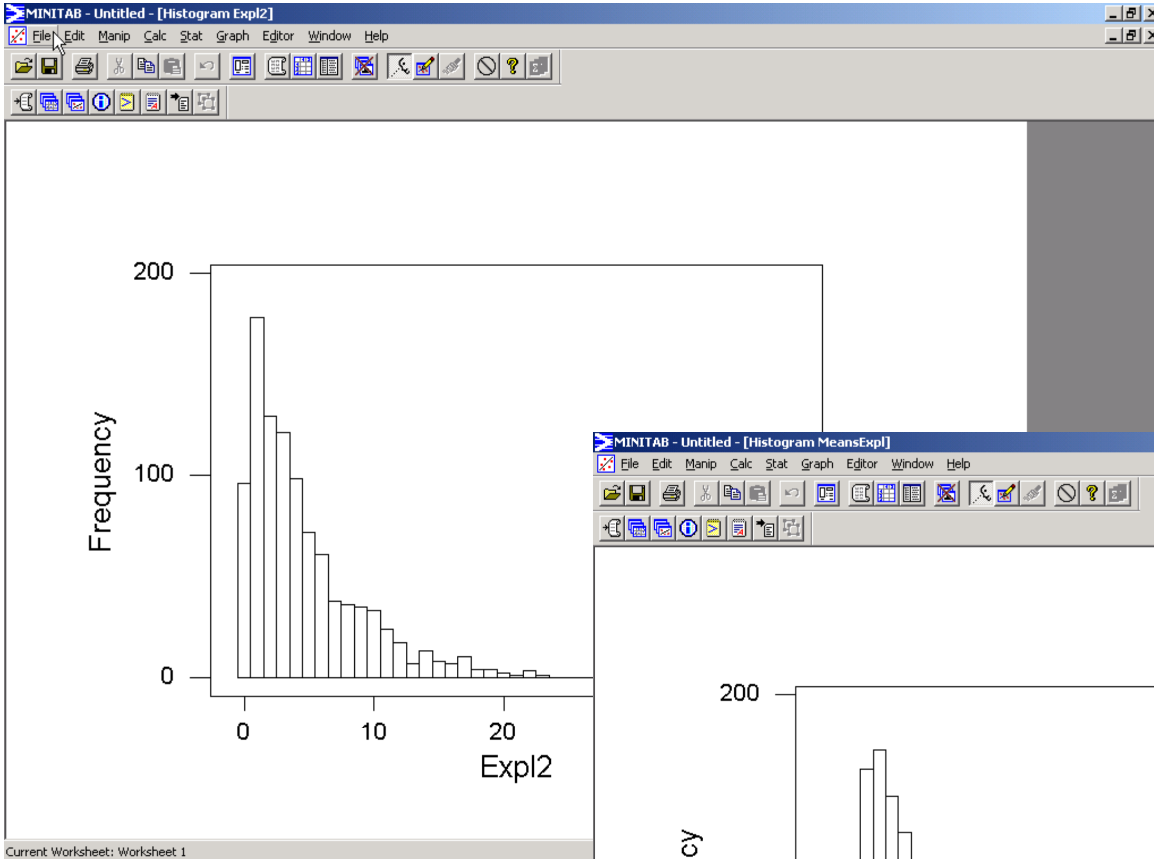
Graph or
Session results

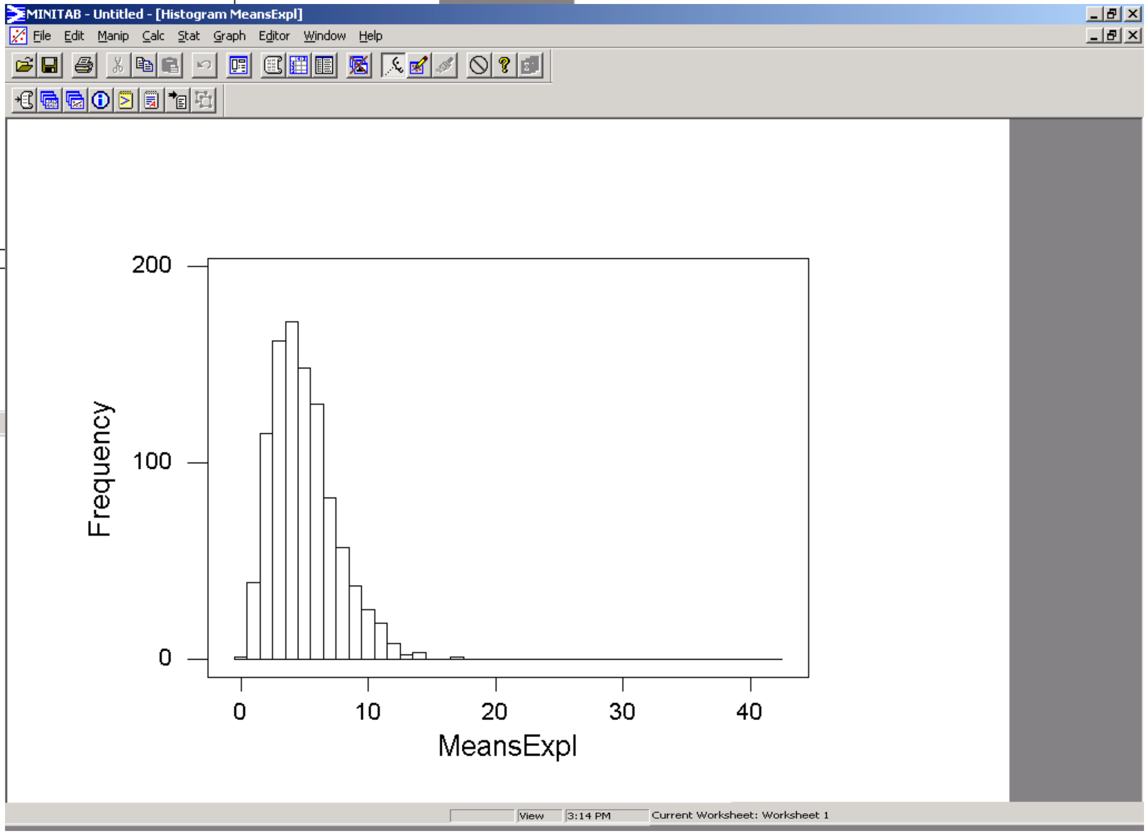
Central Limit Theorem Minitab Example: Exponential Distribution



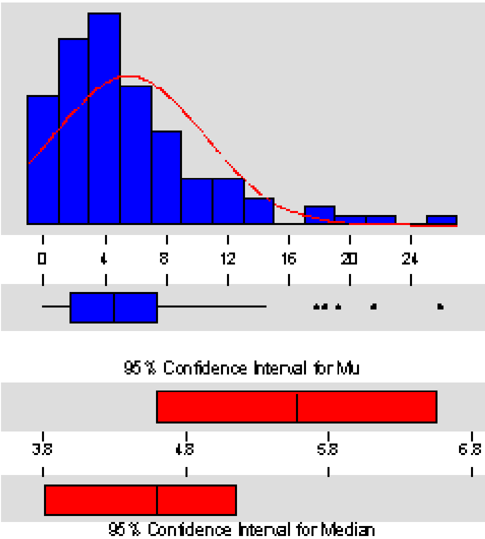





MINITAB - Untitled - [Histogram Expl2]
File Edit Manip Calc Stat Graph Editor Window Help
200
100
0
Frequency
0
10
20
Expl2
Current Worksheet: Worksheet 1


MINITAB - Untitled - [Histogram MeansExpl]
File Edit Manip Calc Stat Graph Editor Window Help
200
100
0
Frequency
0
10
20
30
40
MeansExpl
View
3:14 PM
Current Worksheet: Worksheet 1

Descriptive Statistics



Variable: exponential

Anderson-Darling Normality Test

A-Squared:	3.763
P-Value:	0.000
Mean	5.57415
StDev	4.93027
Variance	24.3075
Skewness	1.67354
Kurtosis	3.41393
N	100
Minimum	0.0103
1st Quartile	1.7638
Median	4.6021
3rd Quartile	7.4743
Maximum	25.8671

95% Confidence Interval for Mu

4.5959 6.5524

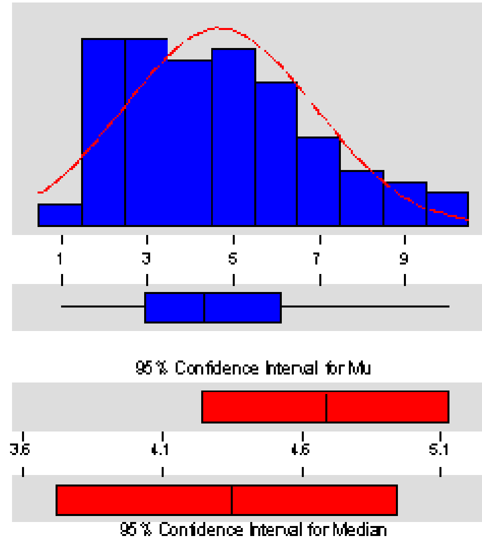
95% Confidence Interval for Sigma

4.3288 5.7274

95% Confidence Interval for Median

3.8220 5.1470

Descriptive Statistics



Variable: average 4

Anderson-Darling Normality Test

A-Squared:	1.240
P-Value:	0.003
Mean	4.68657
StDev	2.21823
Variance	4.92054
Skewness	0.574823
Kurtosis	-4.7E-01
N	100
Minimum	1.0261
1st Quartile	2.9878
Median	4.3501
3rd Quartile	6.0967
Maximum	10.0361

95% Confidence Interval for Mu

4.2464 5.1267

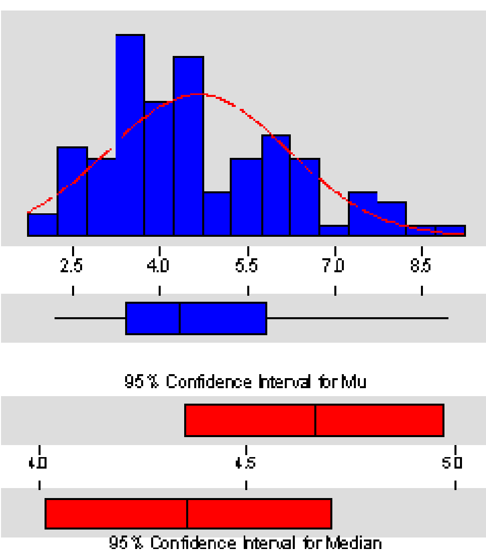
95% Confidence Interval for Sigma

1.9476 2.5769

95% Confidence Interval for Median

3.7276 4.9451

Descriptive Statistics



Variable: average 9

Anderson-Darling Normality Test

A-Squared:	1.415
P-Value:	0.001
Mean	4.66358
StDev	1.56564
Variance	2.45124
Skewness	0.658051
Kurtosis	-2.2E-01
N	100
Minimum	2.20521
1st Quartile	3.45675
Median	4.35676
3rd Quartile	5.83548
Maximum	8.96720

95% Confidence Interval for Mu

4.35292 4.97424

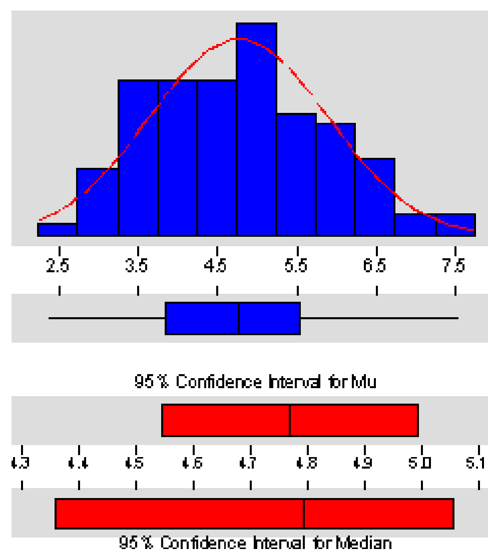
95% Confidence Interval for Sigma

1.37465 1.81877

95% Confidence Interval for Median

4.02079 4.70047

Descriptive Statistics



Variable: average 16

Anderson-Darling Normality Test

A-Squared:	0.275
P-Value:	0.653
Mean	4.77111
StDev	1.12540
Variance	1.26653
Skewness	0.183930
Kurtosis	-5.1E-01
N	100
Minimum	2.36864
1st Quartile	3.86102
Median	4.79291
3rd Quartile	5.55493
Maximum	7.54305

95% Confidence Interval for Mu

4.54781 4.99442

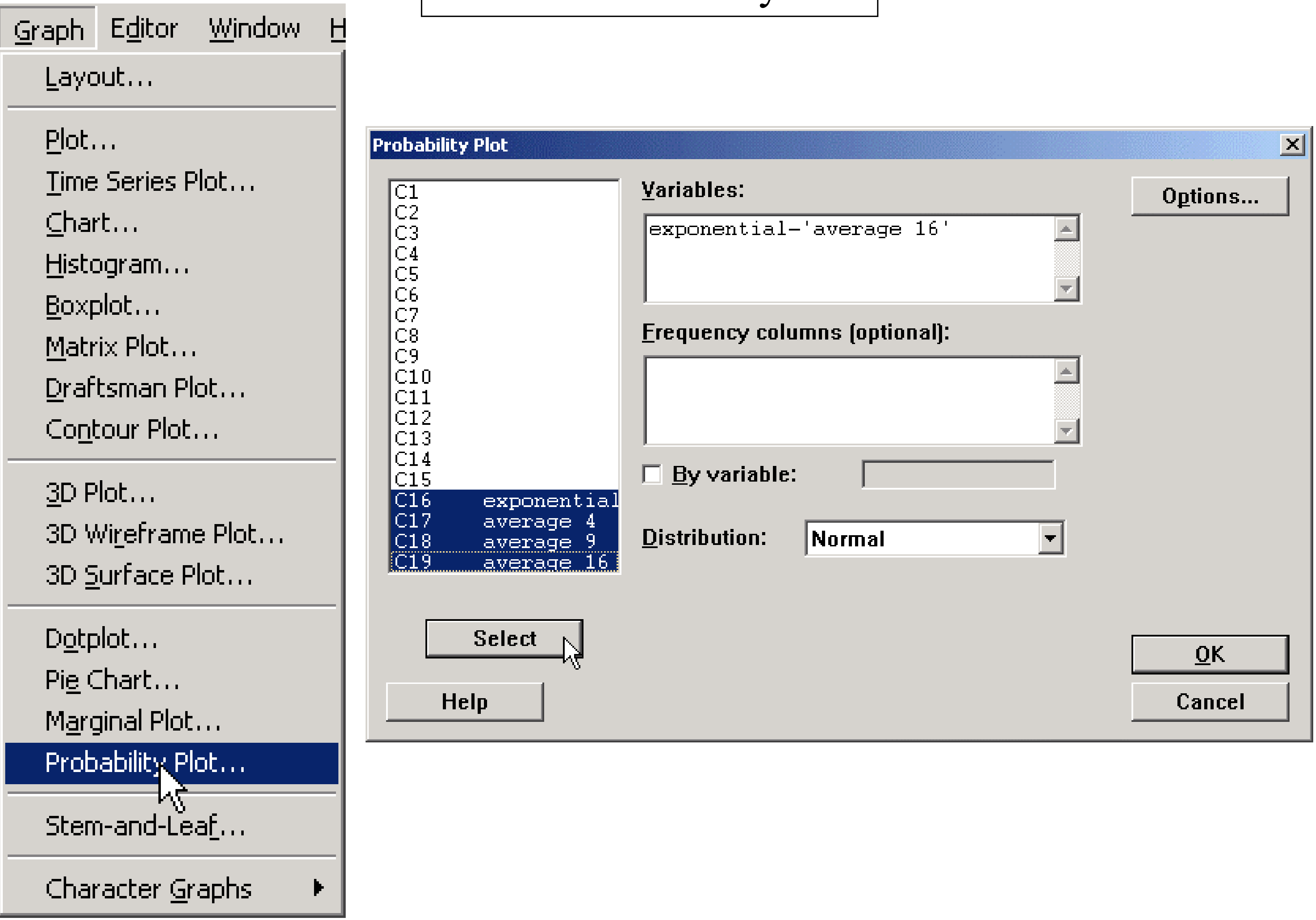
95% Confidence Interval for Sigma

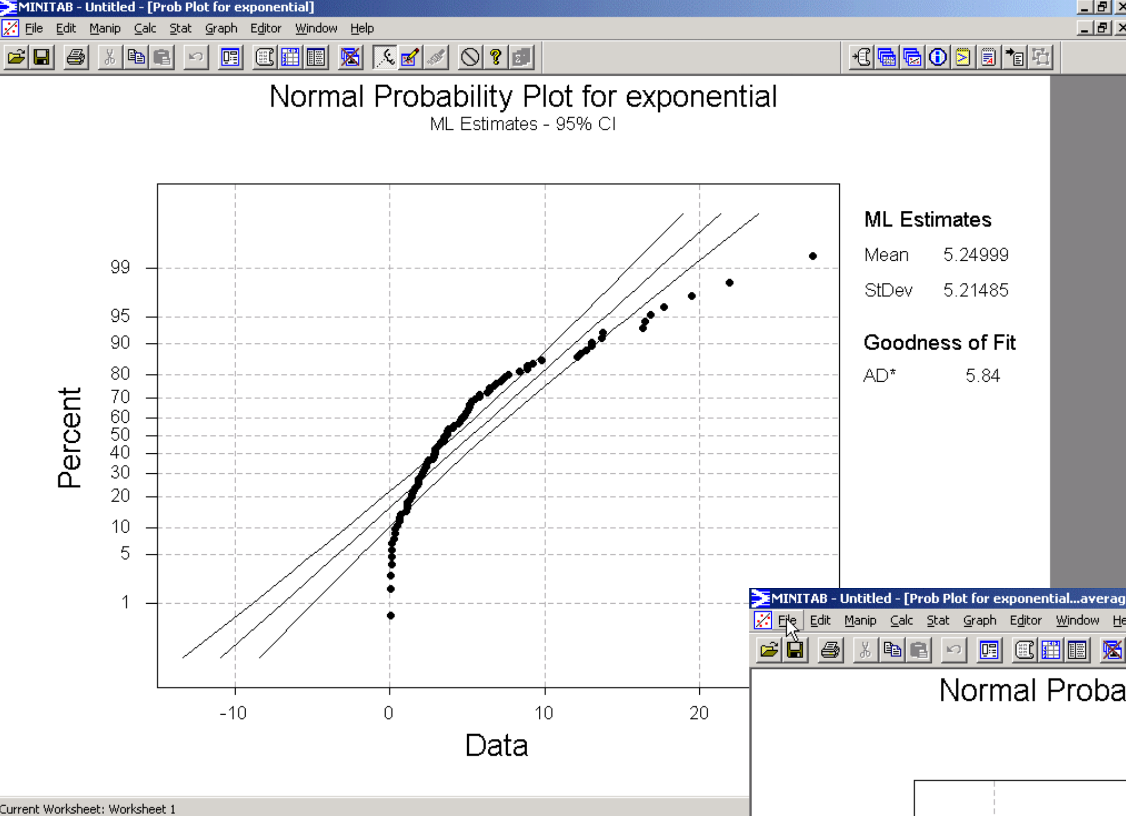
0.98811 1.30735

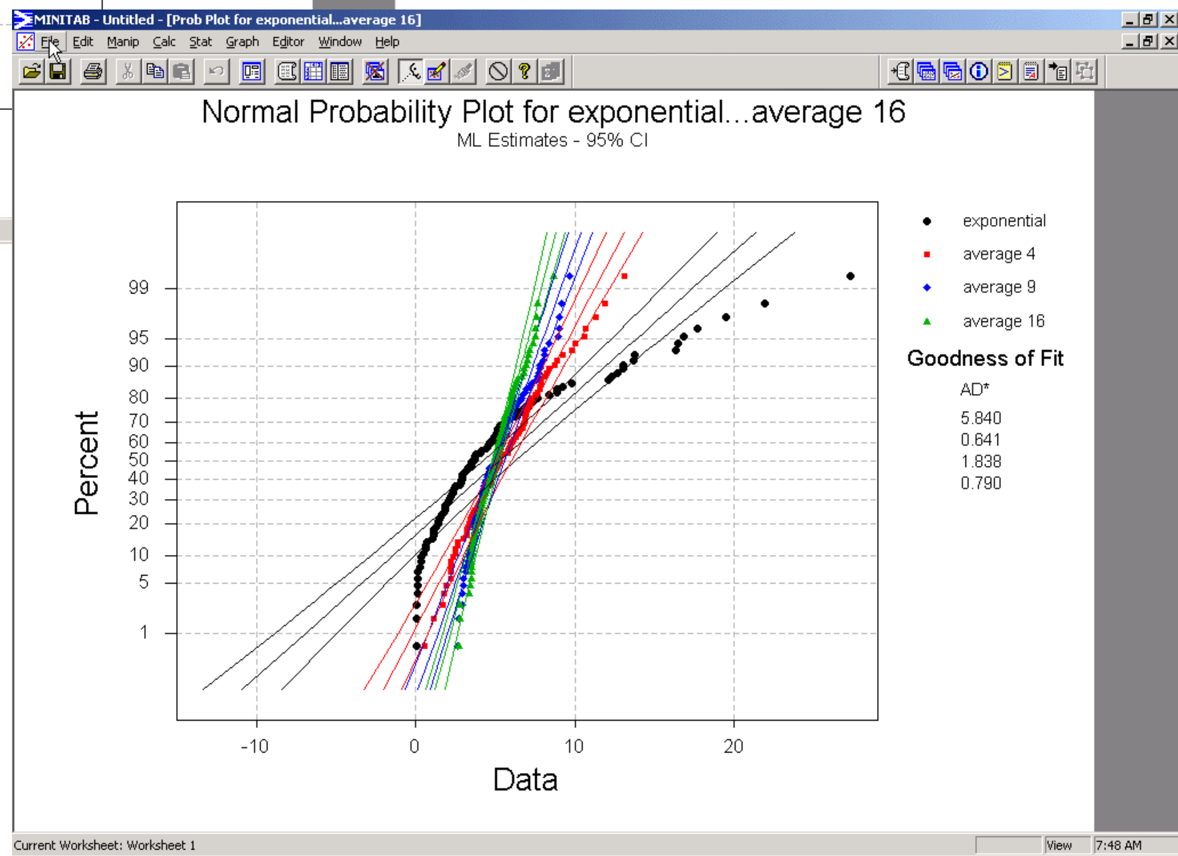
95% Confidence Interval for Median

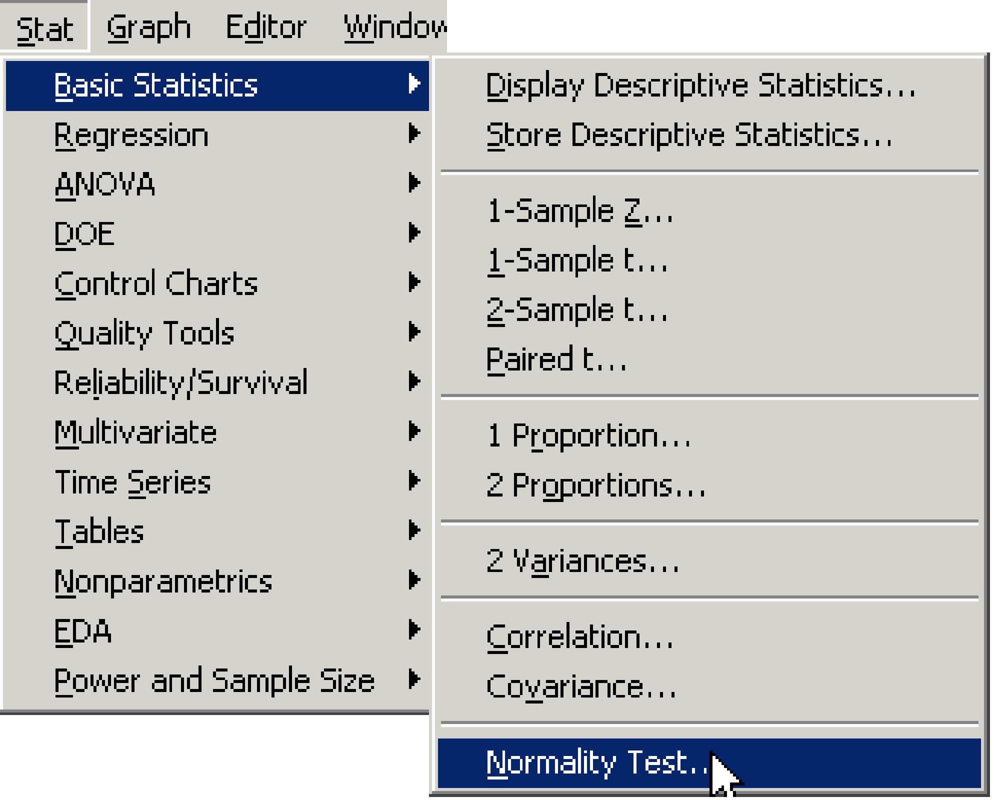
4.36164 5.05748

Normal Probability Plot



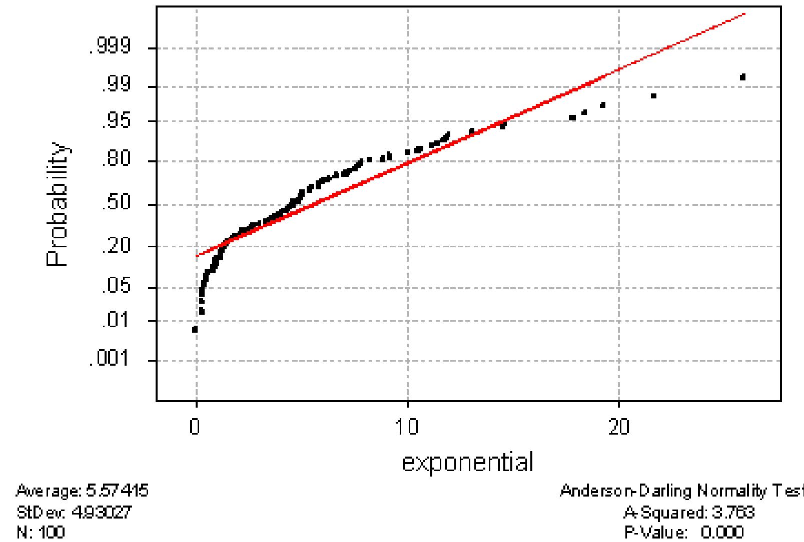

MINITAB - Untitled - [Prob Plot for exponential]
File Edit Manip Calc Stat Graph Editor Window Help
Normal Probability Plot for exponential
ML Estimates - 95% CI
ML Estimates
Mean 5.24999
StDev 5.21485
Goodness of Fit
AD* 5.84
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-10
0
10
20
Data
Current Worksheet: Worksheet 1


MINITAB - Untitled - [Prob Plot for exponential...average 16]
File Edit Manip Calc Stat Graph Editor Window Help
Normal Probability Plot for exponential...average 16
ML Estimates - 95% CI
exponential
average 4
average 9
average 16
Goodness of Fit
AD*
5.840
0.641
1.838
0.790
Percent
99
95
90
80
70
60
50
40
30
20
10
5
1
-10
0
10
20
Data
Current Worksheet: Worksheet 1
View
7:48 AM



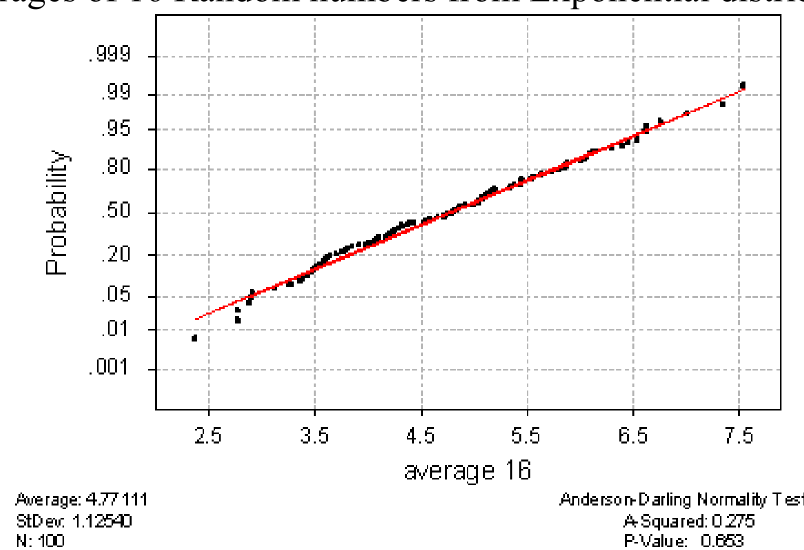
Normal Probability Plot

Random numbers from Exponential distribution

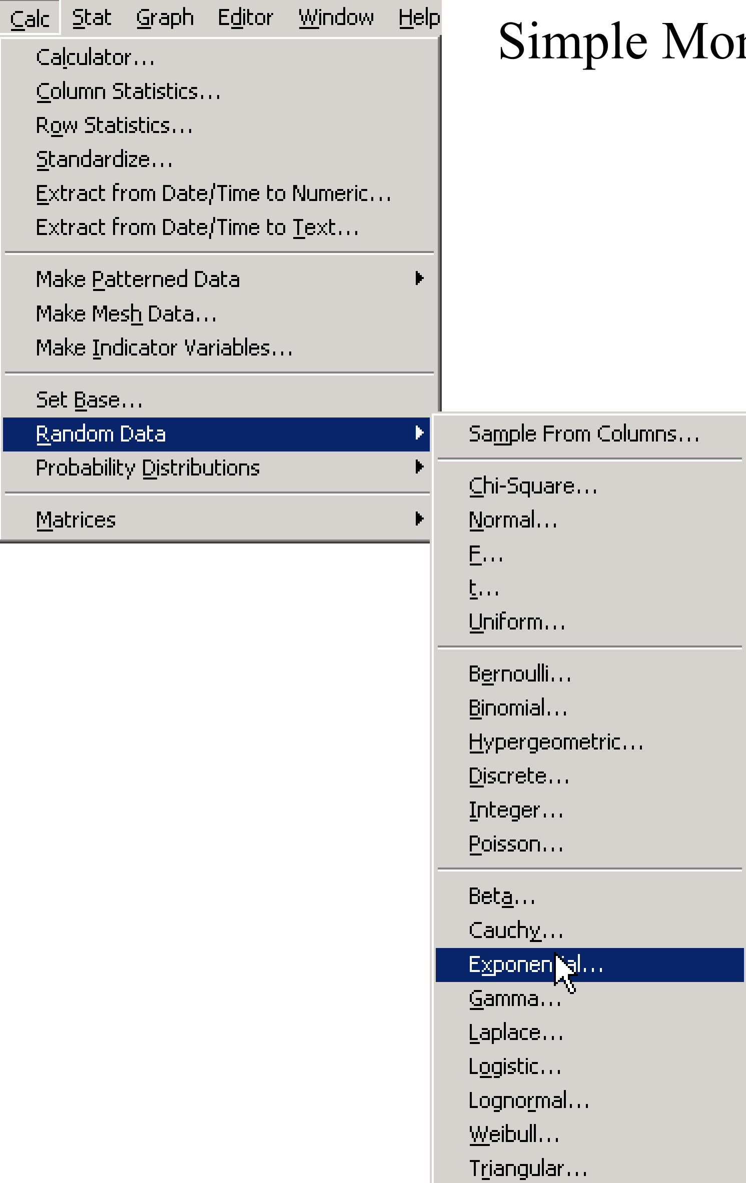


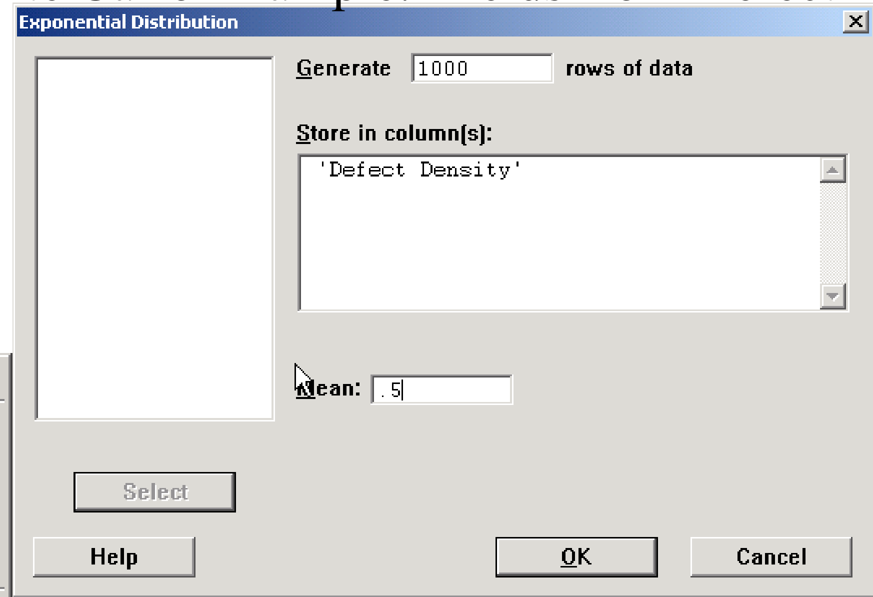
Normal Probability Plot

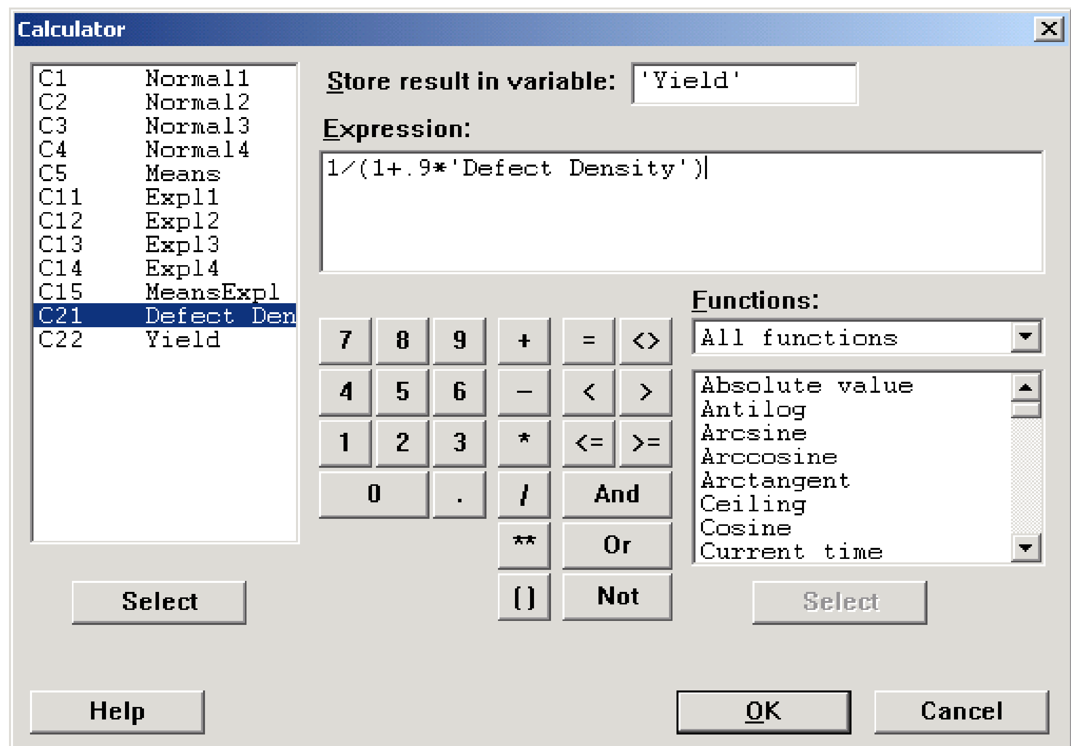
Averages of 16 Random numbers from Exponential distribution



Simple Monte Carlo Example: Yields from Defect Densities

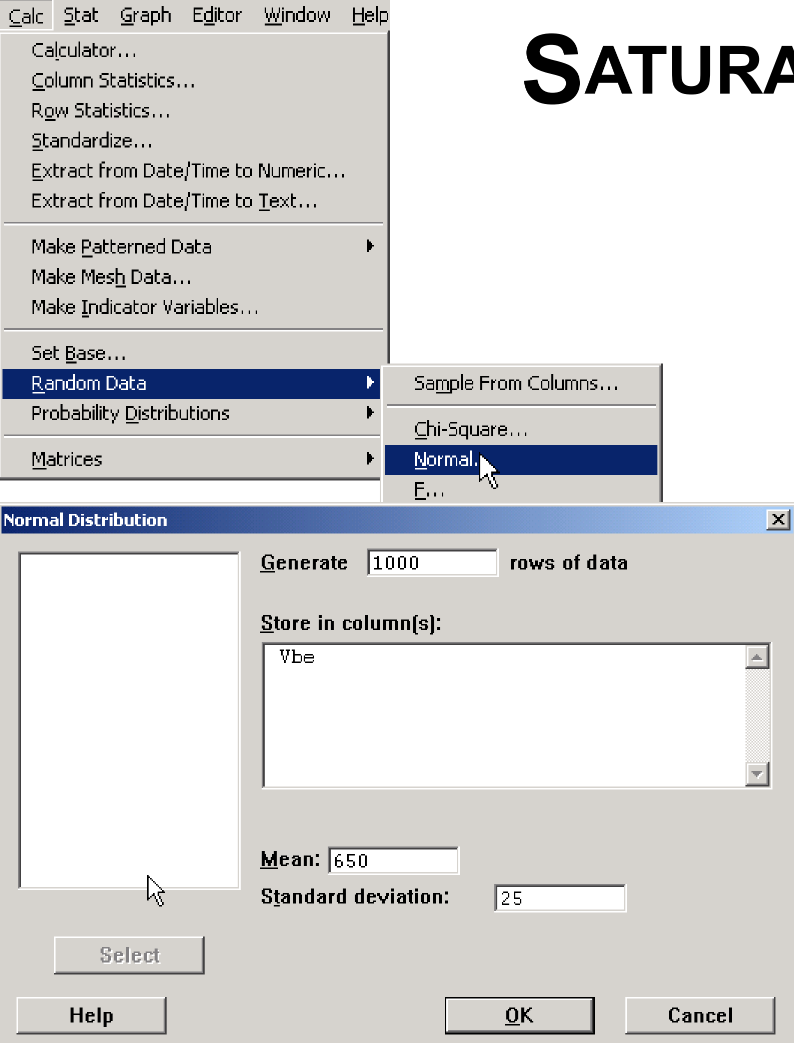


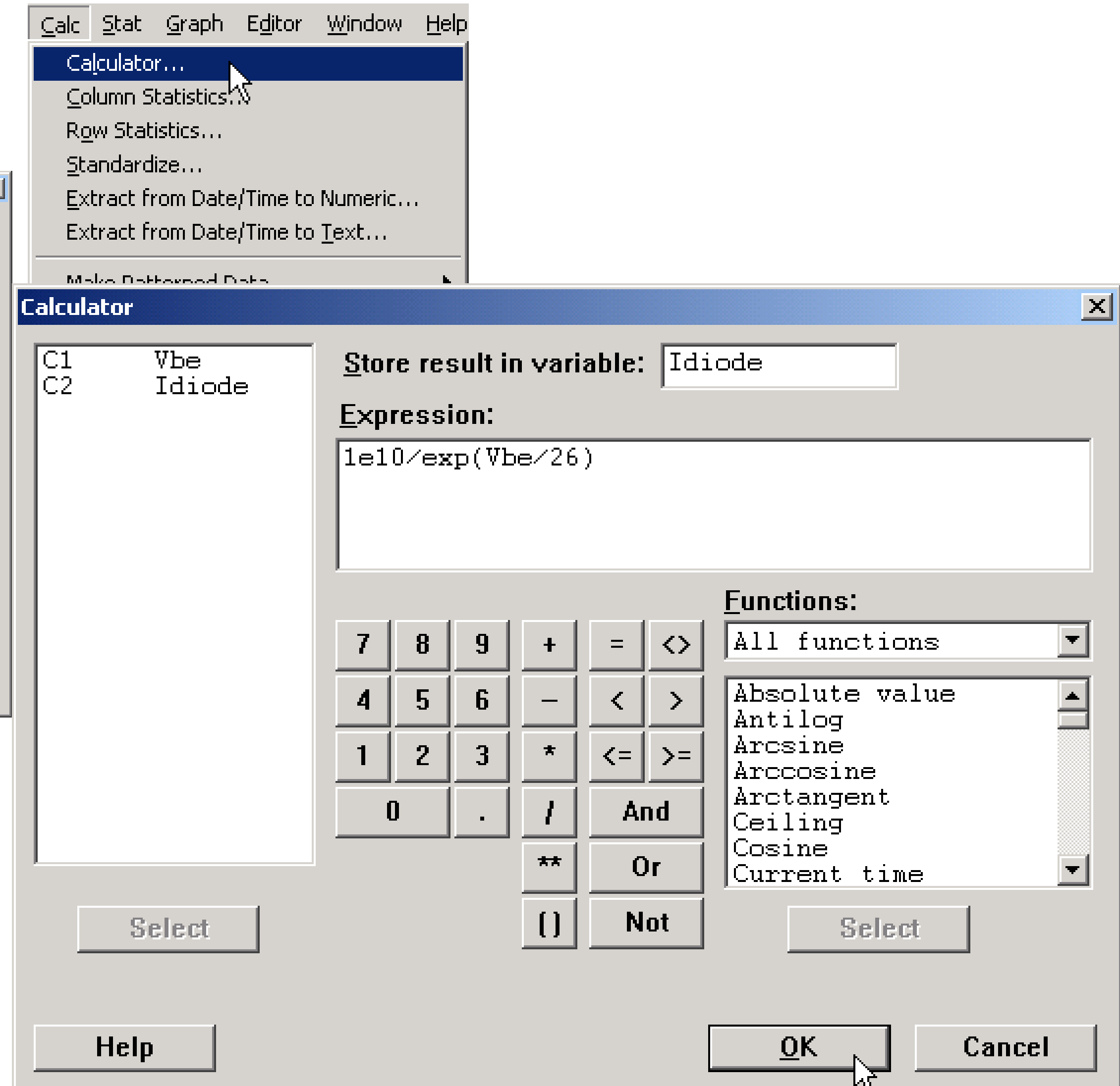




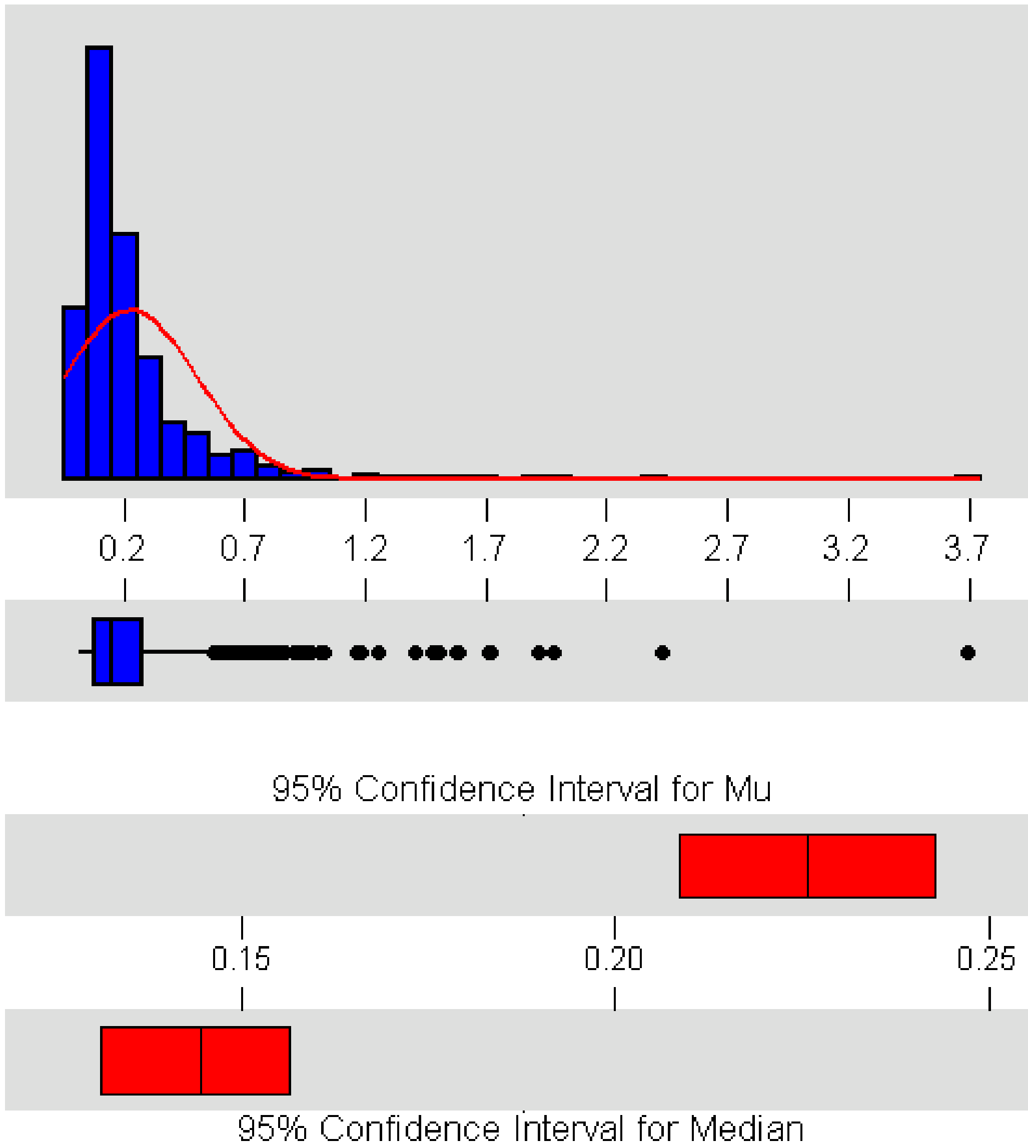
Saturation Current Simulation

$$I_{diode}=Isat\{e\hat{}(qV_{be}/kT)-1\}$$





Descriptive Statistics



Variable: Idiode

Anderson-Darling Normality Test

A-Squared:	88.000
P-Value:	0.000
Mean	0.225687
StDev	0.275615
Variance	7.60E-02
Skewness	4.50470
Kurtosis	35.6022
N	1000
Minimum	0.00901
1st Quartile	0.07095
Median	0.14434
3rd Quartile	0.27003
Maximum	3.69556

95% Confidence Interval for Mu

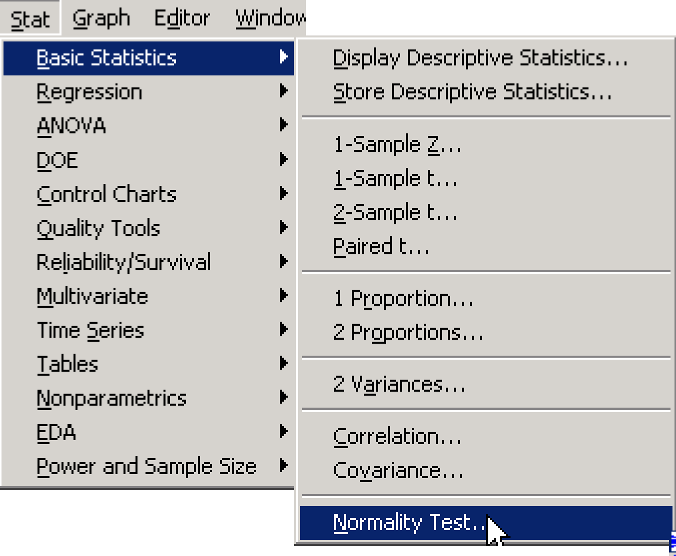
0.20858	0.24279

95% Confidence Interval for Sigma

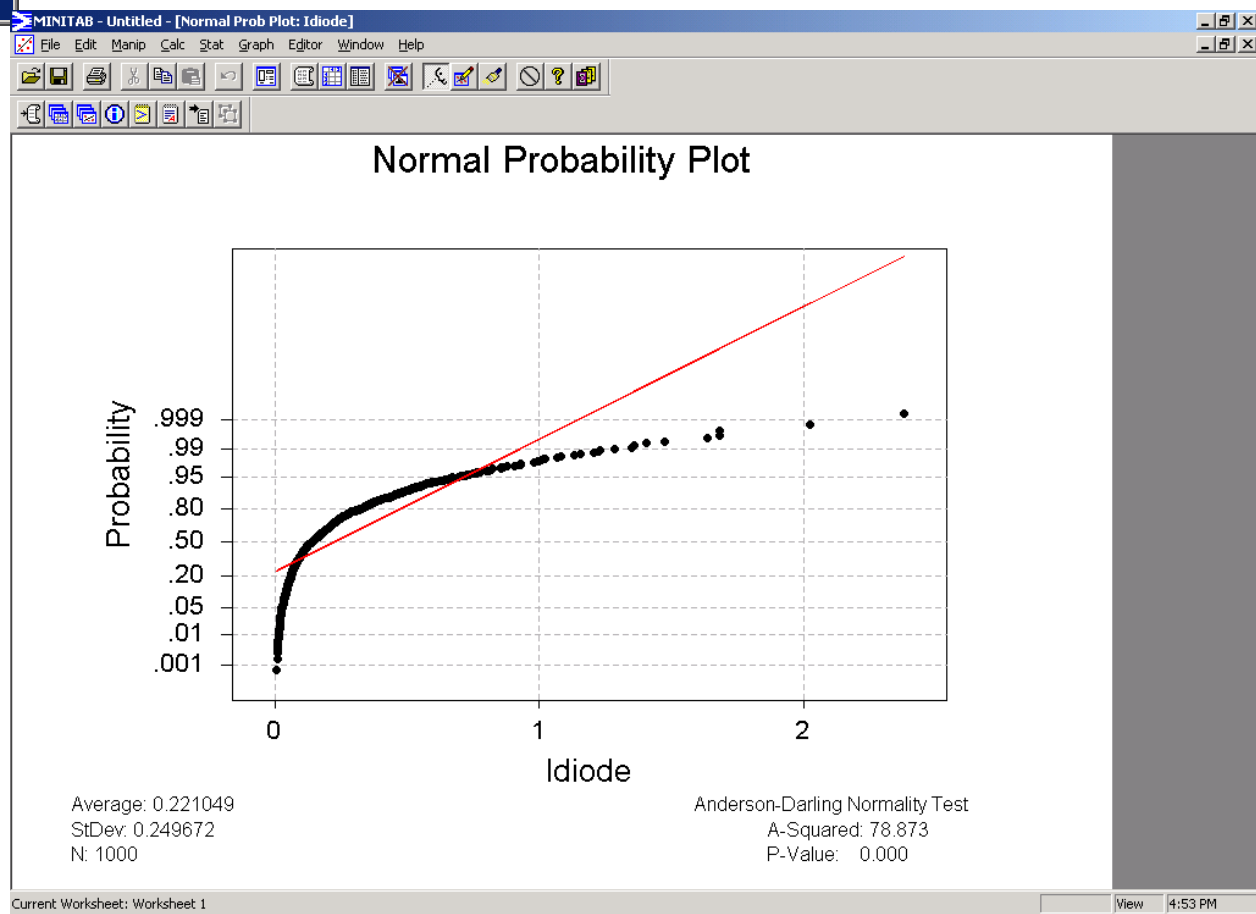
0.26404	0.28826

95% Confidence Interval for Median

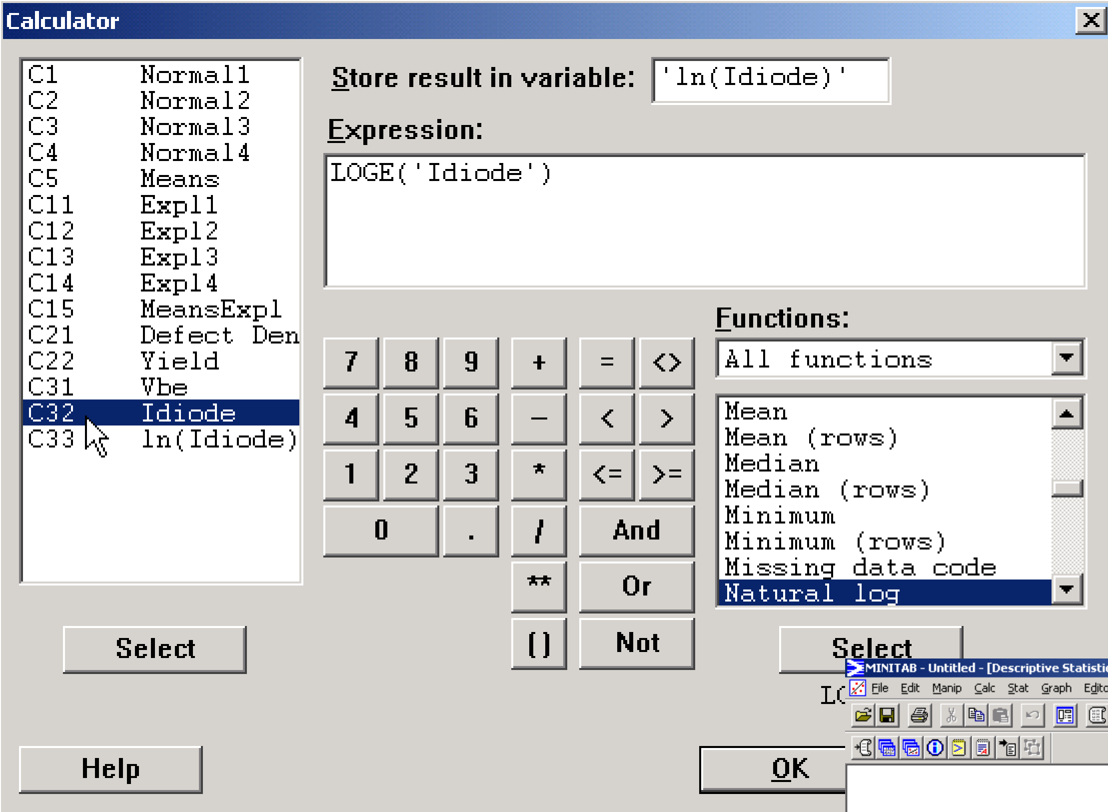
0.13117	0.15617

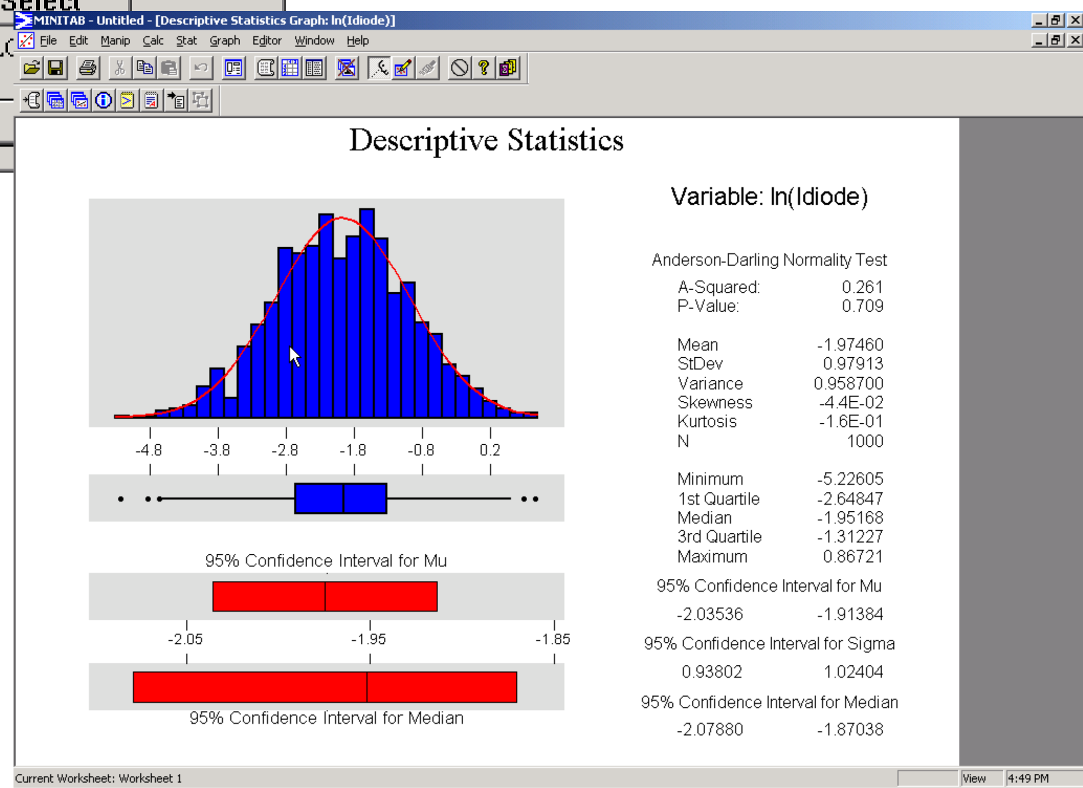


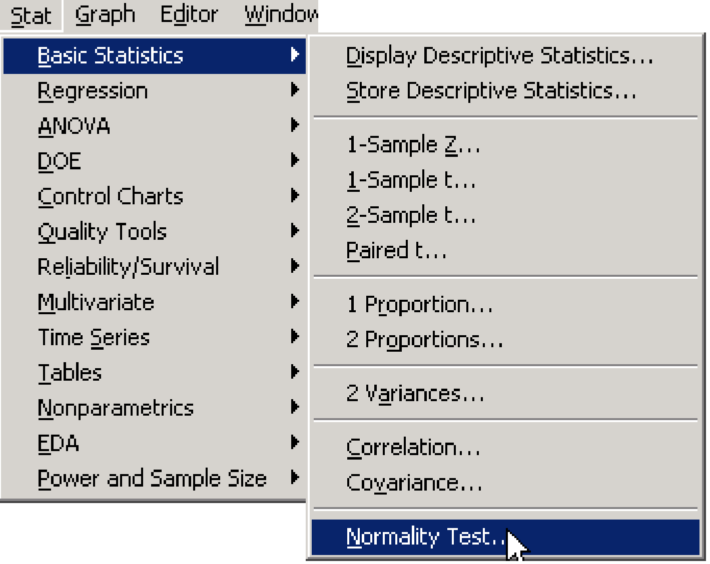
Stat Graph Editor Window
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Display Descriptive Statistics...
Store Descriptive Statistics...
1-Sample Z...
1-Sample t...
2-Sample t...
Paired t...
1 Proportion...
2 Proportions...
2 Variances...
Correlation...
Covariance...
Normality Test...

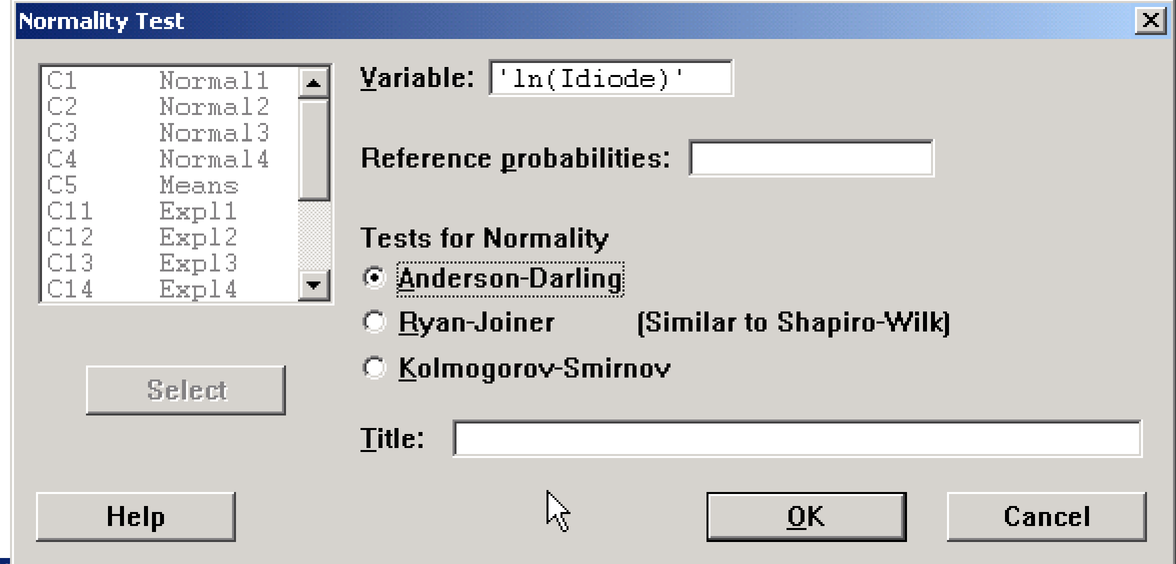


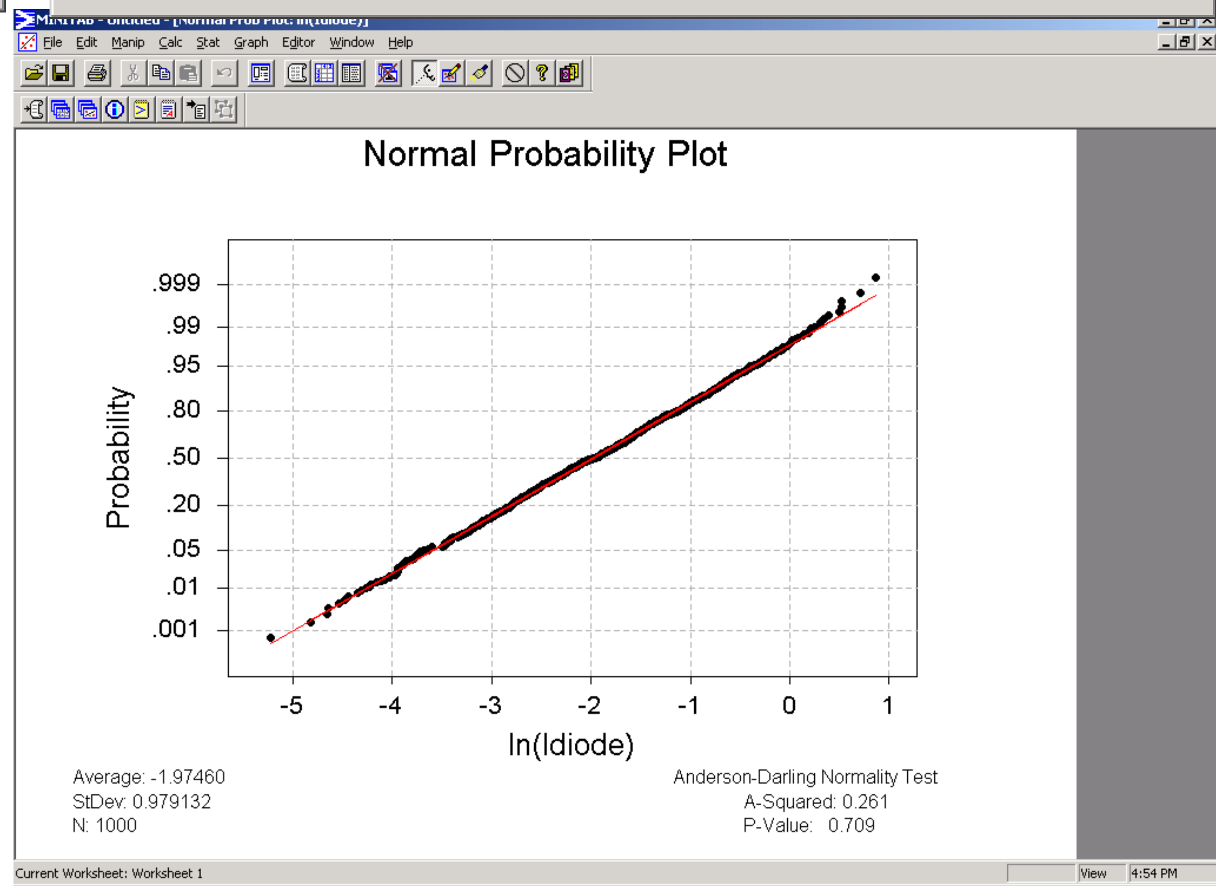
MINITAB - Untitled - [Normal Prob Plot: Idiode]
File Edit Manip Calc Stat Graph Editor Window Help
Normal Probability Plot
Probability
.999
.99
.95
.80
.50
.20
.05
.01
.001
0
1
2
Idiode
Average: 0.221049
StDev: 0.249672
N: 1000
Anderson-Darling Normality Test
A-Squared: 78.873
P-Value: 0.000
Current Worksheet: Worksheet 1
View
4:53 PM


Calculator
C1 Normal1
C2 Normal2
C3 Normal3
C4 Normal4
C5 Means
C11 Expl1
C12 Expl2
C13 Expl3
C14 Expl4
C15 MeansExpl
C21 Defect Den
C22 Yield
C31 Vbe
C32 Idiode
C33 ln(Idiode)
Store result in variable: 'ln(Idiode)'
Expression:
LOGE('Idiode')
Functions:
All functions
Mean
Mean (rows)
Median
Median (rows)
Minimum
Minimum (rows)
Missing data code
Natural log
7 8 9 + = <>
4 5 6 – < >
1 2 3 * <= >=
0 . / And
** Or
() Not
Select
Select
Help
OK


MINITAB - Untitled - [Descriptive Statistics Graph: ln(Idiode)]
File Edit Manip Calc Stat Graph Editor Window Help
Descriptive Statistics
Variable: ln(Idiode)
-4.8 -3.8 -2.8 -1.8 -0.8 0.2
95% Confidence Interval for Mu
-2.05 -1.95 -1.85
95% Confidence Interval for Median
Anderson-Darling Normality Test
A-Squared: 0.261
P-Value: 0.709
Mean -1.97460
StDev 0.97913
Variance 0.958700
Skewness -4.4E-02
Kurtosis -1.6E-01
N 1000
Minimum -5.22605
1st Quartile -2.64847
Median -1.95168
3rd Quartile -1.31227
Maximum 0.86721
95% Confidence Interval for Mu
-2.03536 -1.91384
95% Confidence Interval for Sigma
0.93802 1.02404
95% Confidence Interval for Median
-2.07880 -1.87038
Current Worksheet: Worksheet 1
View
4:49 PM


Stat
Graph
Editor
Window
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Display Descriptive Statistics...
Store Descriptive Statistics...
1-Sample Z...
1-Sample t...
2-Sample t...
Paired t...
1 Proportion...
2 Proportions...
2 Variances...
Correlation...
Covariance...
Normality Test...


Normality Test
C1 Normal1
C2 Normal2
C3 Normal3
C4 Normal4
C5 Means
C11 Expl1
C12 Expl2
C13 Expl3
C14 Expl4
Variable: 'ln(Idiode)'
Reference probabilities:
Tests for Normality
Anderson-Darling
Ryan-Joiner
(Similar to Shapiro-Wilk)
Kolmogorov-Smirnov
Select
Title:
Help
OK
Cancel


File Edit Manip Calc Stat Graph Editor Window Help
Normal Probability Plot
Probability
.999
.99
.95
.80
.50
.20
.05
.01
.001
-5
-4
-3
-2
-1
0
1
ln(Idiode)
Average: -1.97460
StDev: 0.979132
N: 1000
Anderson-Darling Normality Test
A-Squared: 0.261
P-Value: 0.709
Current Worksheet: Worksheet 1
View
4:54 PM

Risks

Reality

Your Conclusion	There is a difference	There is not a difference
There is a difference	(**Way to Go!)**	α **- Risk (Type I Error)** ← Producer's Risk
There is not a Difference	**ß-Risk (Type II Error)** ← Consumer's Risk	**(Way to Go!)**

Relation of Statistical Hypothesis Testing to U.S. Judicial System

		Defendant's True Status	
		Innocent	**Guilty**
Decision	Not Guilty	Correct Decision	Error
	Guilty	Error	Correct Decision

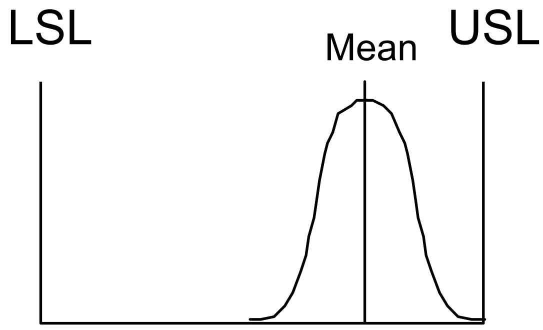
Alpha – Producer's risk

Beta – Consumer's risk
(risk to society)

Process Capability
Cp and Cpk

$$Cp = \frac{(USL - LSL)}{6\ \ s}$$

$$Cpk = \frac{[(NSL - MEAN)]}{3\ \ s}$$



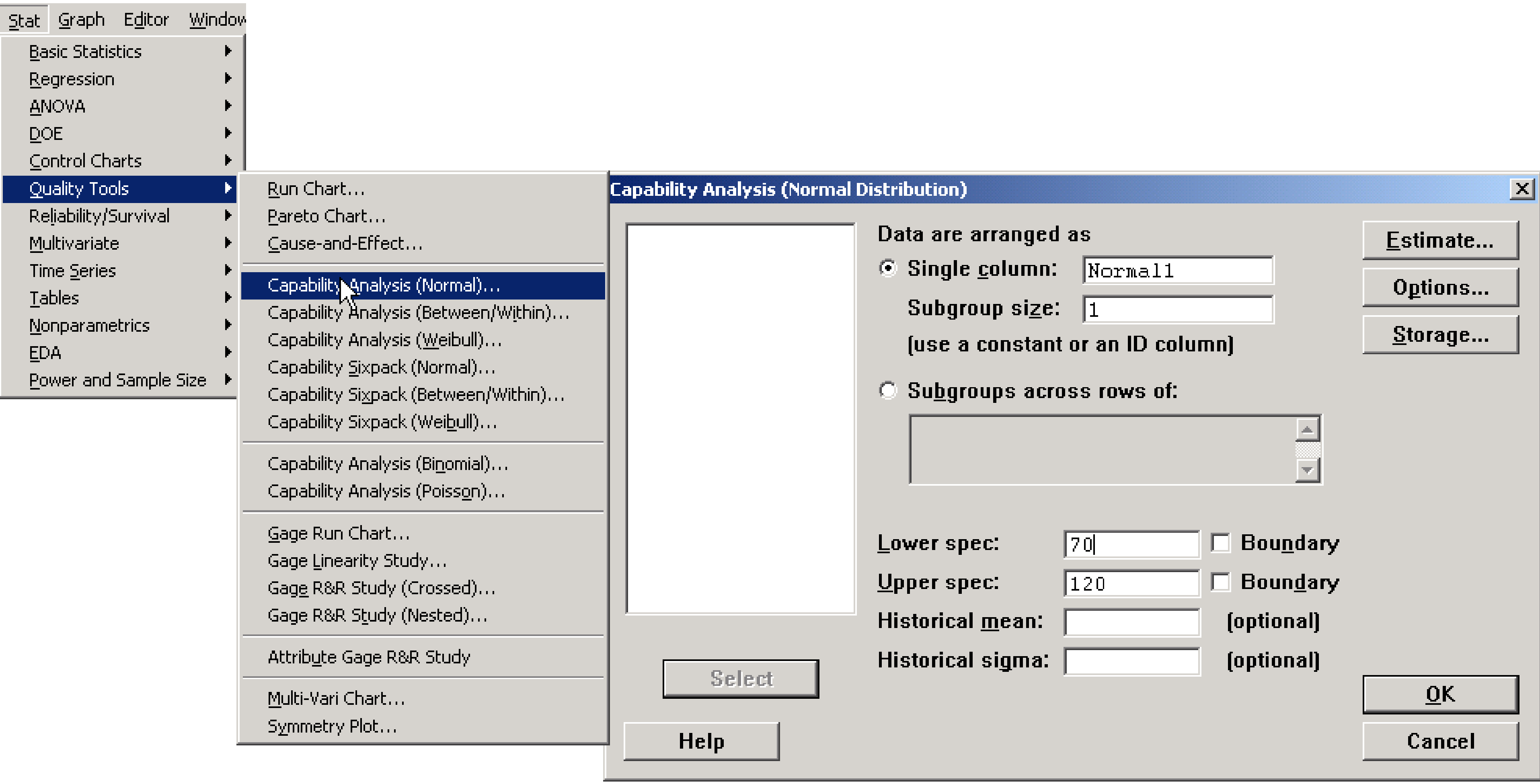
USL = UPPER SPEC LIMIT

LSL = LOWER SPEC LIMIT

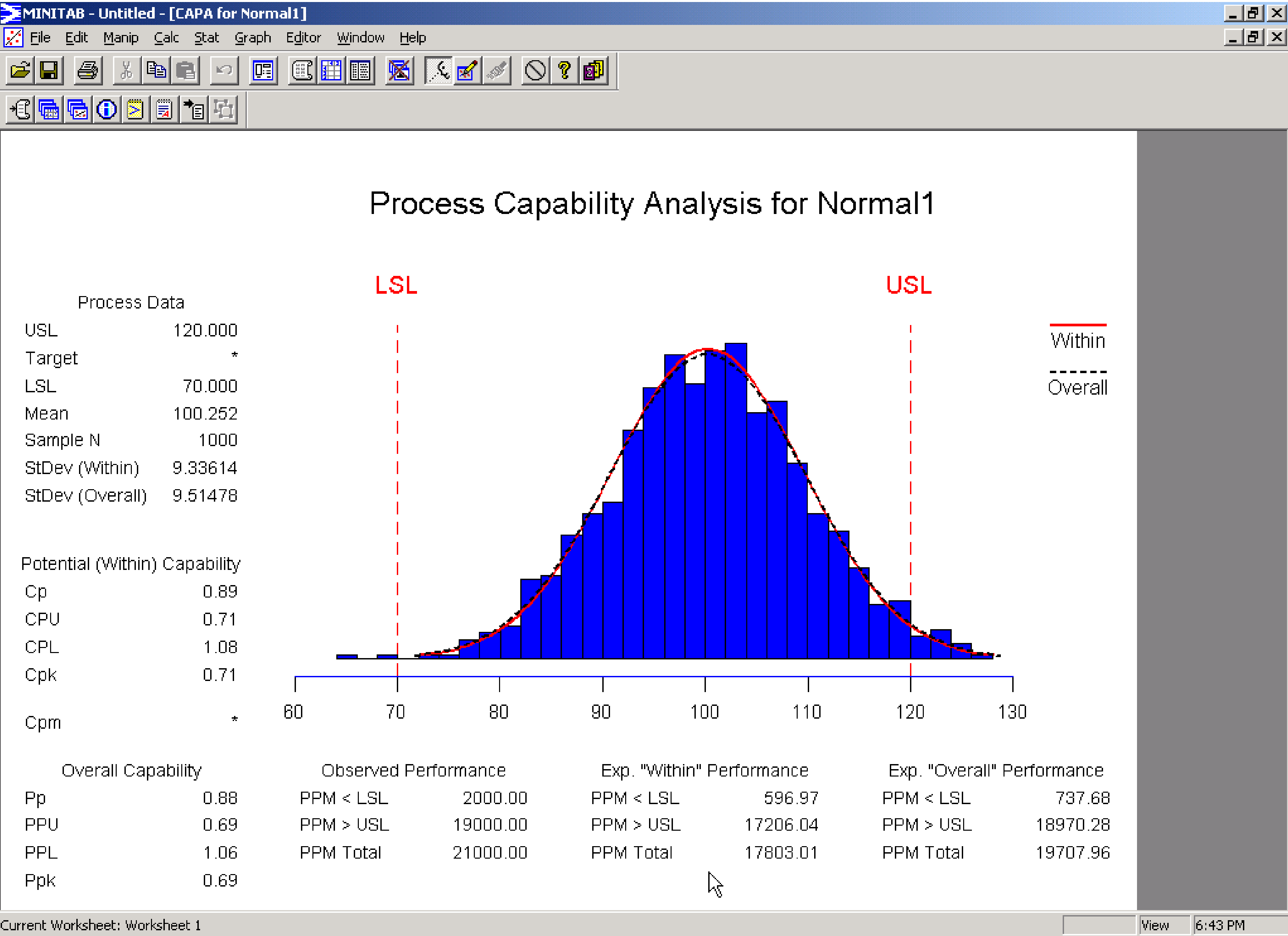
NSL = NEARER SPEC LIMIT (TO MEAN)

Z-Table and associated Cpk - Cumulative Normal Distribution Tail Probabilities

CPK	z	0.00	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.0	0.5000	0.4960	0.4920	0.4880	0.4840	0.4801	0.4761	0.4721	0.4681	0.4641
	0.1	0.4602	0.4562	0.4522	0.4483	0.4443	0.4404	0.4364	0.4325	0.4286	0.4247
	0.2	0.4207	0.4168	0.4129	0.4090	0.4052	0.4013	0.3974	0.3936	0.3897	0.3859
.1	0.3	0.3821	0.3783	0.3745	0.3707	0.3669	0.3632	0.3594	0.3557	0.3520	0.3483
	0.4	0.3446	0.3409	0.3372	0.3336	0.3300	0.3264	0.3228	0.3192	0.3156	0.3121
	0.5	0.3085	0.3050	0.3015	0.2981	0.2946	0.2912	0.2877	0.2843	0.2810	0.2776
.2	0.6	0.2743	0.2709	0.2676	0.2643	0.2611	0.2578	0.2546	0.2514	0.2483	0.2451
	0.7	0.2420	0.2389	0.2358	0.2327	0.2296	0.2266	0.2236	0.2206	0.2177	0.2148
	0.8	0.2119	0.2090	0.2061	0.2033	0.2005	0.1977	0.1949	0.1922	0.1894	0.1867
.3	0.9	0.1841	0.1814	0.1788	0.1762	0.1736	0.1711	0.1685	0.1660	0.1635	0.1611
	1.0	0.1587	0.1562	0.1539	0.1515	0.1492	0.1469	0.1446	0.1423	0.1401	0.1379
	1.1	0.1357	0.1335	0.1314	0.1292	0.1271	0.1251	0.1230	0.1210	0.1190	0.1170
.4	1.2	0.1151	0.1131	0.1112	0.1093	0.1075	0.1056	0.1038	0.1020	0.1003	0.0985
	1.3	0.0968	0.0951	0.0934	0.0918	0.0901	0.0885	0.0869	0.0853	0.0838	0.0823
	1.4	0.0808	0.0793	0.0778	0.0765	0.0749	0.0735	0.0721	0.0708	0.0694	0.0681
.5	1.5	0.0668	0.0655	0.0643	0.0630	0.0618	0.0606	0.0594	0.0582	0.0574	0.0559
	1.6	0.0548	0.0537	0.0526	0.0516	0.0505	0.0495	0.0485	0.0475	0.0465	0.0455
	1.7	0.0446	0.0436	0.0427	0.0418	0.0409	0.0401	0.0392	0.0384	0.0375	0.0367
.6	1.8	0.0359	0.0351	0.0344	0.0336	0.0329	0.0322	0.0314	0.0307	0.0301	0.0294
	1.9	0.0287	0.0281	0.0274	0.0268	0.0262	0.0256	0.0250	0.0244	0.0239	0.0233
	2.0	0.0228	0.0222	0.0217	0.0212	0.0207	0.0202	0.0197	0.0192	0.0188	0.0183
.7	2.1	0.0179	0.0174	0.0170	0.0166	0.0162	0.0158	0.0154	0.0150	0.0146	0.0143
	2.2	0.0139	0.0136	0.0132	0.0129	0.0125	0.0122	0.0119	0.0116	0.0113	0.0110
	2.3	0.0107	0.0104	0.0102	0.0099	0.0096	0.0094	0.0091	0.0089	0.0087	0.0084
.8	2.4	0.0082	0.0080	0.0078	0.0075	0.0073	0.0071	0.0069	0.0068	0.0066	0.0064
	2.5	0.0062	0.0060	0.0059	0.0057	0.0055	0.0054	0.0052	0.0051	0.0049	0.0048
	2.6	0.0047	0.0045	0.0044	0.0043	0.0041	0.0040	0.0039	0.0038	0.0037	0.0036
.9	2.7	0.0035	0.0034	0.0033	0.0032	0.0031	0.0030	0.0029	0.0028	0.0027	0.0026
	2.8	0.0026	0.0025	0.0024	0.0023	0.0023	0.0022	0.0021	0.0021	0.0020	0.0019
	2.9	0.0019	0.0018	0.0018	0.0017	0.0016	0.0016	0.0015	0.0015	0.0014	00014
1.0	3.0	0.0013	0.0013	0.0013	0.0012	0.0012	0.0011	0.0011	0.0011	0.0010	0.0010
	3.1	0.0010	0.0009	0.0009	0.0009	0.0008	0.0008	0.0008	0.0008	0.0007	0.0007
	3.2	0.0007	0.0007	0.0006	0.0006	0.0006	0.0006	0.0006	0.0005	0.0005	0.0005
1.1	3.3	0.0005	0.0005	0.0005	0.0004	0.0004	0.0004	0.0004	0.0004	0.0004	0.0003
	3.4	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0003	0.0002
	3.5	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002	0.0002
1.2	3.6	0.0002	0.0002	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	3.7	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
	3.8	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001	0.0001
1.3	3.9	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000	0.0000



Stat
Graph
Editor
Window
Basic Statistics
Regression
ANOVA
DOE
Control Charts
Quality Tools
Reliability/Survival
Multivariate
Time Series
Tables
Nonparametrics
EDA
Power and Sample Size
Run Chart...
Pareto Chart...
Cause-and-Effect...
Capability Analysis (Normal)...
Capability Analysis (Between/Within)...
Capability Analysis (Weibull)...
Capability Sixpack (Normal)...
Capability Sixpack (Between/Within)...
Capability Sixpack (Weibull)...
Capability Analysis (Binomial)...
Capability Analysis (Poisson)...
Gage Run Chart...
Gage Linearity Study...
Gage R&R Study (Crossed)...
Gage R&R Study (Nested)...
Attribute Gage R&R Study
Multi-Vari Chart...
Symmetry Plot...
Capability Analysis (Normal Distribution)
Data are arranged as
Single column:
Normal1
Subgroup size:
1
(use a constant or an ID column)
Subgroups across rows of:
Lower spec:
70
Boundary
Upper spec:
120
Boundary
Historical mean:
(optional)
Historical sigma:
(optional)
Estimate...
Options...
Storage...
Select
Help
OK
Cancel


MINITAB - Untitled - [CAPA for Normal1]
File Edit Manip Calc Stat Graph Editor Window Help
Process Capability Analysis for Normal1
LSL
USL
Within
Overall
Process Data
USL 120.000
Target *
LSL 70.000
Mean 100.252
Sample N 1000
StDev (Within) 9.33614
StDev (Overall) 9.51478
Potential (Within) Capability
Cp 0.89
CPU 0.71
CPL 1.08
Cpk 0.71
Cpm *
60 70 80 90 100 110 120 130
Overall Capability
Pp 0.88
PPU 0.69
PPL 1.06
Ppk 0.69
Observed Performance
PPM < LSL 2000.00
PPM > USL 19000.00
PPM Total 21000.00
Exp. "Within" Performance
PPM < LSL 596.97
PPM > USL 17206.04
PPM Total 17803.01
Exp. "Overall" Performance
PPM < LSL 737.68
PPM > USL 18970.28
PPM Total 19707.96
Current Worksheet: Worksheet 1
View
6:43 PM

TAGUCHI VARIANCE

$$S_T^{\,2} = S^2 + (\bar{x} - T)^2$$

Variance about mean — Variance of mean from target

NOTE:

$$\sigma^2 = S^2 = \frac{\sum (x - \bar{x})^2}{N}$$ Second moment about mean

$$S_T^{\,2} = \frac{\sum (x - T)^2}{N}$$ Second moment about target

- Can be used to separate the amount of the "problem" due to:

  - Deviation of mean from target (miscentered/mis-targeted) — $(\bar{x} - T)^2$

    from

    $+$

  - Deviation of parts from mean ("Process variability") — S^2

What is Quality?

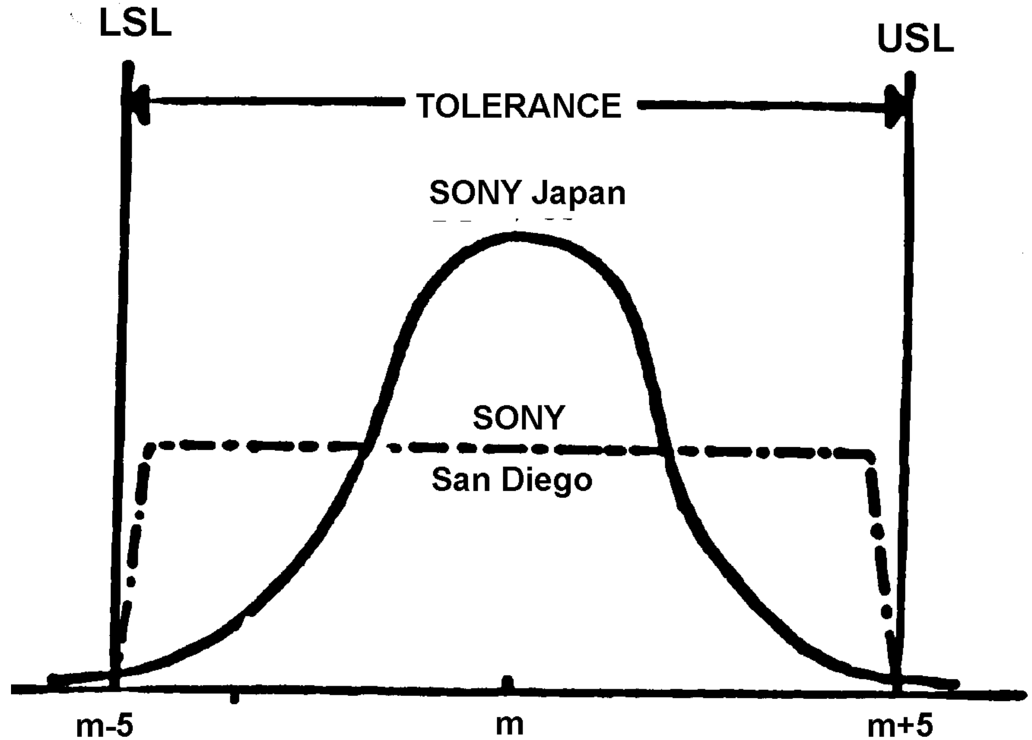
- Fitness for use — Juran
- Conformance to requirements — Crosby
- Meeting expectations of customers — Feigenbaum

- Minimization of the loss imparted to society by a product due to deviations of the product's functional characteristics from desired target values — Taguchi

- Conformance to target and reduction of variability — Luftig

CONFORMANCE TO SPECIFICATIONS VS. CONFORMANCE TO TARGET

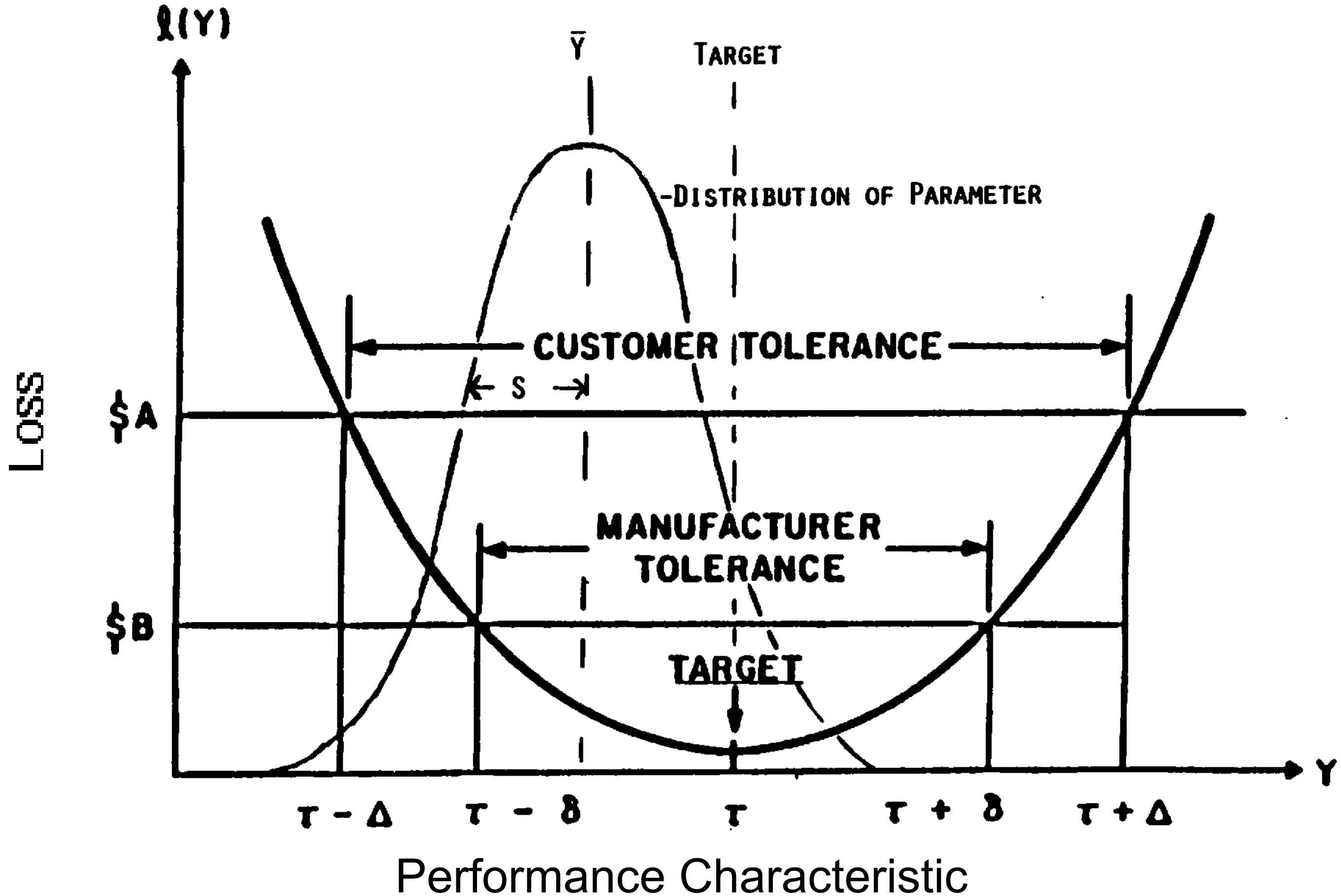
An example: Sony TV case



Distibuition of color concentration (from ASAHI, 4/17/79)

Taguchi Loss Function All Parts in Spec are Not of Equal Cost

$$\text{EXPECTED LOSS (\$)} = K\,[\,(\bar{Y} - T)^2 + S^2\,]$$



TAGUCHI'S QUALITY LOSS FUNCTION AND SPC

Vol limits: - 1.830 to - 1.683

Vol target: - 1.756

Vol mean: - 1.822

Vol Std Dev = .0097

$(\bar{x} - T)^2$	=	.0043	98%
S^2	=	.000093	2%
Total	=	.00445	100%
$(S_T$	=	.0667)	

98% of variance from ideal distribution is due to miscentering

$$Cp(T) = \frac{-1.683 - (-1.83)}{6 \times .0667} = 0.37$$

$$Cpk = \frac{-1.822 - (-1.83)}{3 \times .0097} = 0.28$$

$$Cp = \frac{-1.683 - (-1.83)}{6 \times .0097} = 2.53$$

MC10175 ECL VOL

Taguchi Loss Function -far from target!! BUT Cpk > 1.5 !!