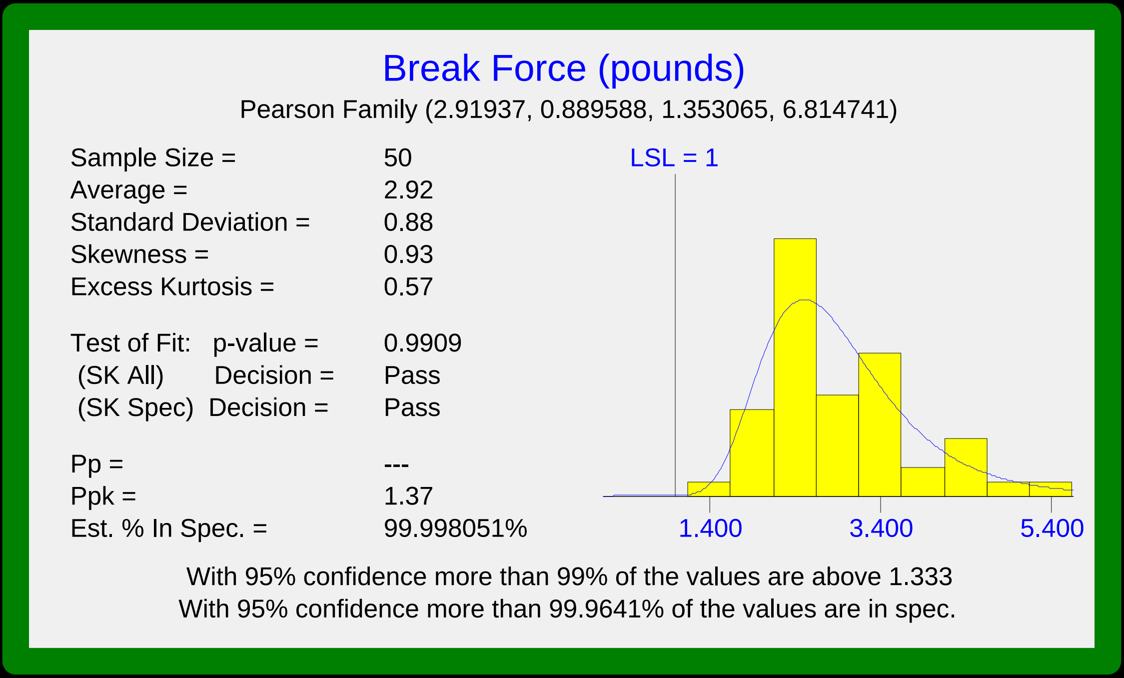
Distribution Analyzer



Break Force (pounds)

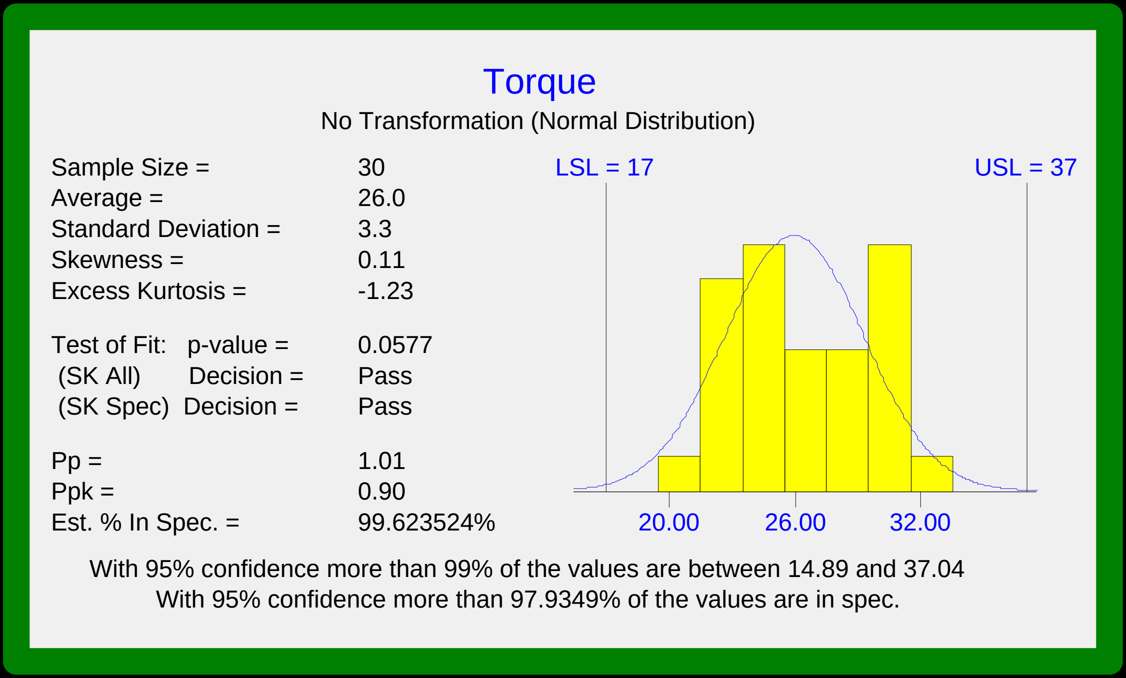
Pearson Family (2.91937, 0.889588, 1.353065, 6.814741)

Sample Size = 50
Average = 2.92
Standard Deviation = 0.88
Skewness = 0.93
Excess Kurtosis = 0.57

Test of Fit: p-value = 0.9909
(SK All) Decision = Pass
(SK Spec) Decision = Pass

Pp = ---
Ppk = 1.37
Est. % In Spec. = 99.998051%

With 95% confidence more than 99% of the values are above 1.333
With 95% confidence more than 99.9641% of the values are in spec.



Torque

No Transformation (Normal Distribution)

Sample Size = 30
Average = 26.0
Standard Deviation = 3.3
Skewness = 0.11
Excess Kurtosis = -1.23

Test of Fit: p-value = 0.0577
(SK All) Decision = Pass
(SK Spec) Decision = Pass

Pp = 1.01
Ppk = 0.90
Est. % In Spec. = 99.623524%

With 95% confidence more than 99% of the values are between 14.89 and 37.04
With 95% confidence more than 97.9349% of the values are in spec.

User's Guide

Distribution Analyzer

Version 1.2

Dr. Wayne A. Taylor

Taylor Enterprises, Inc.
5510 Fairmont Road, Suite A
Libertyville, IL 60048
(847) 367-1032
Fax: (847) 367-1037
E-mail: info@variation.com
Web: www.variation.com

This software package was created by:

Dr. Wayne A. Taylor

Dr. Taylor is the founder and chairman of Taylor Enterprises, Inc. where he is responsible for the development of the VarTran®, Change-Point Analyzer, Sampling Plan Analyzer and Distribution Analyzer software packages as well as providing consulting and training on the Six Sigma approach and tools including Robust Design, Tolerance Analysis, Design of Experiments, SPC, Acceptance Sampling, and Statistics.

Dr. Taylor retired from his position as Director of Quality Technologies at Baxter Healthcare Corporation where he was responsible for implementing Baxter's Six Sigma program. He had been with Baxter for 22 years.

Dr. Taylor is author of the books *Optimization and Variation Reduction in Quality* and *Guide to Acceptance Sampling*. His two courses *Successful Acceptance Sampling* and *Robust Tolerance Analysis* have been attended by thousands of engineers and scientists.

Dr. Taylor is a leading expert on acceptance sampling and process validation in the pharmaceutical, medical device and diagnostics industries. His articles on selecting statistically valid sampling plans have become standards in the industry. He is one of the authors of the Global Harmonization Task Force guideline on Process Validation.

The VarTran software and Dr. Taylor's course Robust Tolerance Analysis have rapidly become a key component of many companies DFSS (Design for Six Sigma) programs. Honeywell, the Six Sigma Academy and numerous other companies have adopted VarTran as an essential tool for designing high quality products.

Dr. Taylor received his Ph.D. in Statistics from Purdue University. He is a fellow of the American Society for Quality.

Table of Contents

1 Program Installation

This section describes the system resources required by Distribution Analyzer and how to install Distribution Analyzer on your computer.

1.1 System Requirements

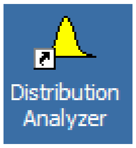
Distribution Analyzer runs on any 32-bit Windows operating system including Windows 95, 98, 2000, XP and Vista. It requires 4 MB of hard disk space to install.

1.2 Installing Distribution Analyzer

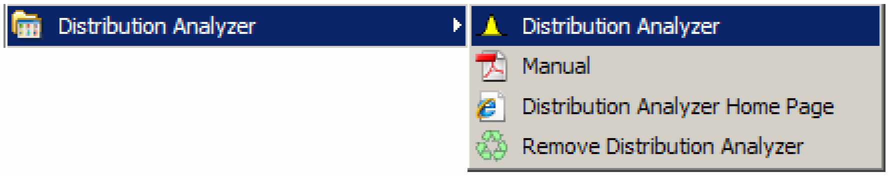
Distribution Analyzer is a shareware program. A full version of the software can be downloaded for a 30-day trial. If you want to continue using the software after trial period, you must register it.

To install Distribution Analyzer, download the file ***da12.exe*** from www.variation.com/da. Executing this file will run the installation program. The software will be automatically installed to the directory ***C:\Program Files\Taylor Enterprises\Distribution Analyzer***. Along with the software, the user manual and numerous example files are copied to this directory. Distribution Analyzer data files end with the extension "da".

A shortcut is added to the desktop for starting the program. It appears as to the right. Double clicking this shortcut will start the program.



In addition, four menu items are added to *All Programs* menu. Click on the Start button in the lower left corner of the screen and then click on the *All Programs* menu item. Find the Distribution Analyzer folder and click on it. This displays the following four menu items:

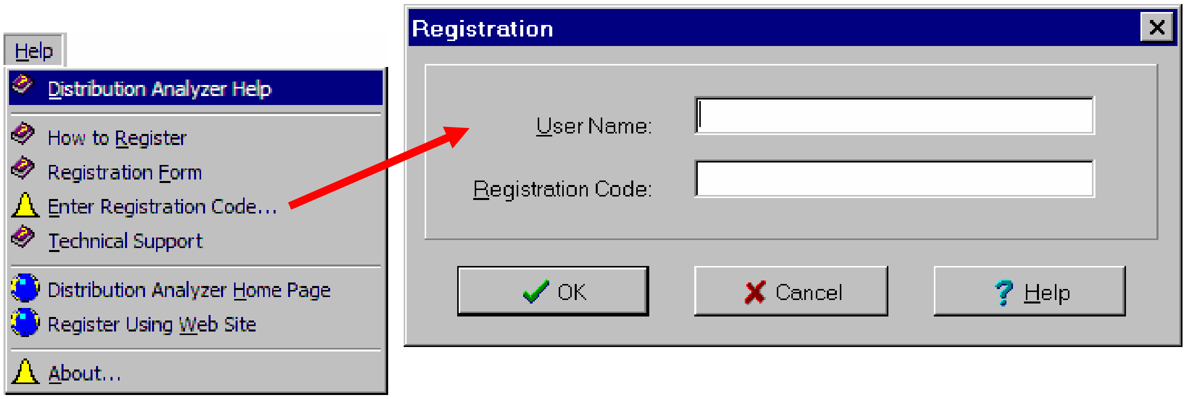


The first menu item starts the program. The second menu item displays program documentation in pdf format. You must have a copy of the free Adobe Acrobat Reader® to view this file. The third menu item displays the Distribution Analyzer home page. The final menu item uninstalls the program.

1.3 Registering the Software

Distribution Analyzer can be used free of charge for 30 days. After the 30 days trial period, you must register the software to continue to use it. To register the software, go to www.variation.com/da and click the *Register* button. Up to date pricing information will be provided along with further instructions. The registration process can be completed online. The *Help* menu of Distribution Analyzer also contains further details about registering the software.

Once the software is registered, you will receive a user name and registration code by email. These must be entered into the software. The user name and registration code can be manually entered into Distribution Analyzer by selecting the *Enter Registration Code* menu item on the *Help* menu. This displays the *Registration* dialog box shown below. The user name and registration code must both be entered exactly as provided (case-sensitive). Once done, click the *OK* button.



1.4 Uninstalling Distribution Analyzer

Distribution Analyzer can be uninstalled using the *Add or Remove Programs* icon on the *Control Panel*. Step-by-step instructions follow:

- Click the *Start* button.
- Select the *Control Panel* menu item.
- Double click the *Add or Remove Programs* icon.
- Select the program Distribution Analyzer and click the *Change/Remove* button.

This will remove the program, user manual and example files. It will not affect any files you created using the software. You can also uninstall it using the menu item described in Section 1.2.

2 Getting Started

Distribution Analyzer is used to test whether a set of data fits the normal distribution and, if not, to determine which distribution best fits the data. Associated with each distribution is a transformation that, when applied to the data, will convert data from that distribution to the normal distribution. Once the data is transformed to the normal distribution, Distribution Analyzer constructs confidence statements like the following:

- With 95% confidence more than 99% of the values are between 17.8 and 23.2 pounds (normal tolerance interval)
- With 95% confidence more than 99.34% of the values are within the specification limits (variables sampling plan)

Distribution Analyzer is specifically designed to aid in the execution of validation/verification/qualification studies designed to make a claim about the performance of a product or process. However, it can be used anytime one wants to test and fit distributions. It is also a valuable learning tool concerning distributions, their relationships and properties.

2.1 Uses and Capabilities

You will find many uses for Distribution Analyzer including:

1. **Validation/Verification/Qualification Studies:** Such studies commonly use normal tolerance intervals and variables sampling plans to make confidence statements like the ones above. Both these procedures make the assumption that the data fits the normal distribution. These two procedures are particularly sensitive to departures from normality so it is advised that a formal test for normality be applied to the data before applying either. Distribution Analyzer contains a specific test designed to detect departures from normality that invalid the use of these two procedures. This test allows one to proceed in many cases that traditional normality tests would fail. If the data is not normal, a transformation may be available that can be applied to the data to make it normal. Distribution Analyzer not only tests for normality and determines the best transformation; it constructs the normal tolerance interval and variables sampling plan confidence statements. Further it does additional analysis like checking for outliers, looking for time order effects and looking for differences between groups to aid in determining why a set of data fails normality.

2. **Test and Fit Distributions:** Distribution Analyzer has the ability to fit a wide range of distributions using both the method of moments and maximum likelihood methods. Both methods have been modified to ensure the distribution covers a specified range.

Distribution Analyzer has simplified the whole process of comparing and fitting distributions by characterizing all distributions in terms of their moments (average, standard deviation, skewness and kurtosis) rather than using a different set of Greek letters for each distribution.

3. **Learning Tool:** Distribution Analyzer can be used to learn about the many distributions and their relationships. Included is a skewness-kurtosis plot for understanding the range of shapes each distribution can fit and the relationships between the different distributions. For each distribution you can view density plots, calculate probabilities and explore the effect of changing the parameters. Finally, you can generate random values for any of the distributions as well as for dice experiments to experience handling different types of data.

Distribution Analyzer has numerous capabilities, many not found elsewhere:

1. **Robust and Specific Tests for Normality:** While Distribution Analyzer contains the traditional Anderson-Darling and Shapiro-Wilks tests, it contains two equally powerful tests designed to overcome shortcomings of these two tests. The first is the Skewness-Kurtosis All test that, like the previous two tests, tests for all departures from normality. However, this test is not adversely affected by ties in the data, as are the other two tests. The second is the Skewness-Kurtosis Specific test. This test is designed to detect those departures from normality that invalidate the use of a normal tolerance interval or variables sampling plan. Specifically, they are designed to detect tails that are heavier than the normal distribution. It does not reject when the tails are equal to or less than the normal distribution so that the confidence statements remain valid, although potentially conservative.

2. **Improved Methods of Fitting Distributions:** In determining which distribution best fits your data, Distribution Analyzer uses both the method of moments and maximum likelihood methods. Both methods have been modified to ensure the distribution covers a specified range. This avoids the issues of spec limits being outside the bounds of the fitted distribution making transformation of the spec limits impossible. A wide range of distributions are covered:

   - Beta Distribution
   - Exponential Distribution
   - Extreme Value Distributions (Smallest Extreme Value, Largest Extreme Value, Weibull, Fréchet)
   - Gamma Distribution
   - Johnson Family of Distributions
   - Logistic and Loglogistic Distributions
   - Lognormal Distribution
   - Normal Distribution
   - Pearson Family of Distribution (Includes Inverse Beta and Inverse Gamma)
   - Uniform

Not only are the above distributions covered, but the negative of these distributions are included to facilitate the fitting of data with negative skewness. Further, if there are physical bounds (like zero), these bounds can be used to pre-transform the data to the unbounded case. This effectively expands the above list of distributions to include the Log-Beta, Log-Pearson and much more. Distribution Analyzer automates the whole process by letting you click a button "Fit Best Distribution". However, you also have the ability to completely control the selection process including the distribution fit, the method of fitting and whether to pre-transform.

3. **Integrated Supporting Analysis:** Data can fail a normality test for a variety of other reasons including a shift over time, a mixture of different groups and the presence of outliers. Whenever testing that a distribution fits the data, the data is automatically checked for outliers, shifts over time and differences between groups and the user notified if anything of importance is detected.

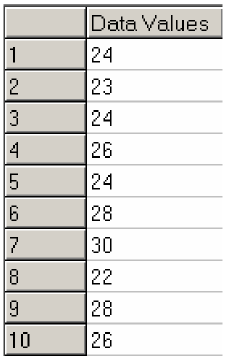
4. **Detailed Information about Each Distribution:** For each distribution you can view density plots, calculate probabilities and explore the effect of changing the parameters. You can view the distribution on a skewness-kurtosis plot for understanding the range of shapes each distribution can fit and its relationships with other distributions.

5. **Generating Data:** Random numbers from any of the distributions can be generated. You can also use Distribution Analyzer to perform simulated dice experiments to illustrate the types of physical phenomena that create different distributions.

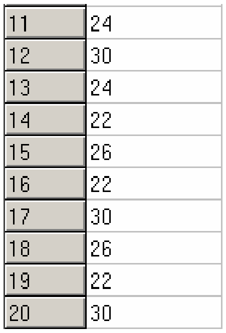
The remainder of Chapter 2 highlights the process of testing for normality and transforming data to allow one to quickly get started. The remaining chapters then provide more complete details.

2.2 Testing for Normality

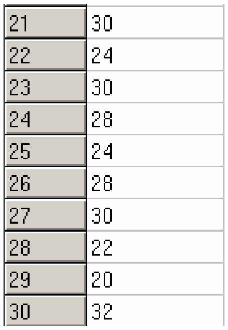
Take as an example the torque values shown below. There is a lower spec limit of 17 pounds and upper spec limit of 37 pounds:



	Data Values
1	24
2	23
3	24
4	26
5	24
6	28
7	30
8	22
9	28
10	26

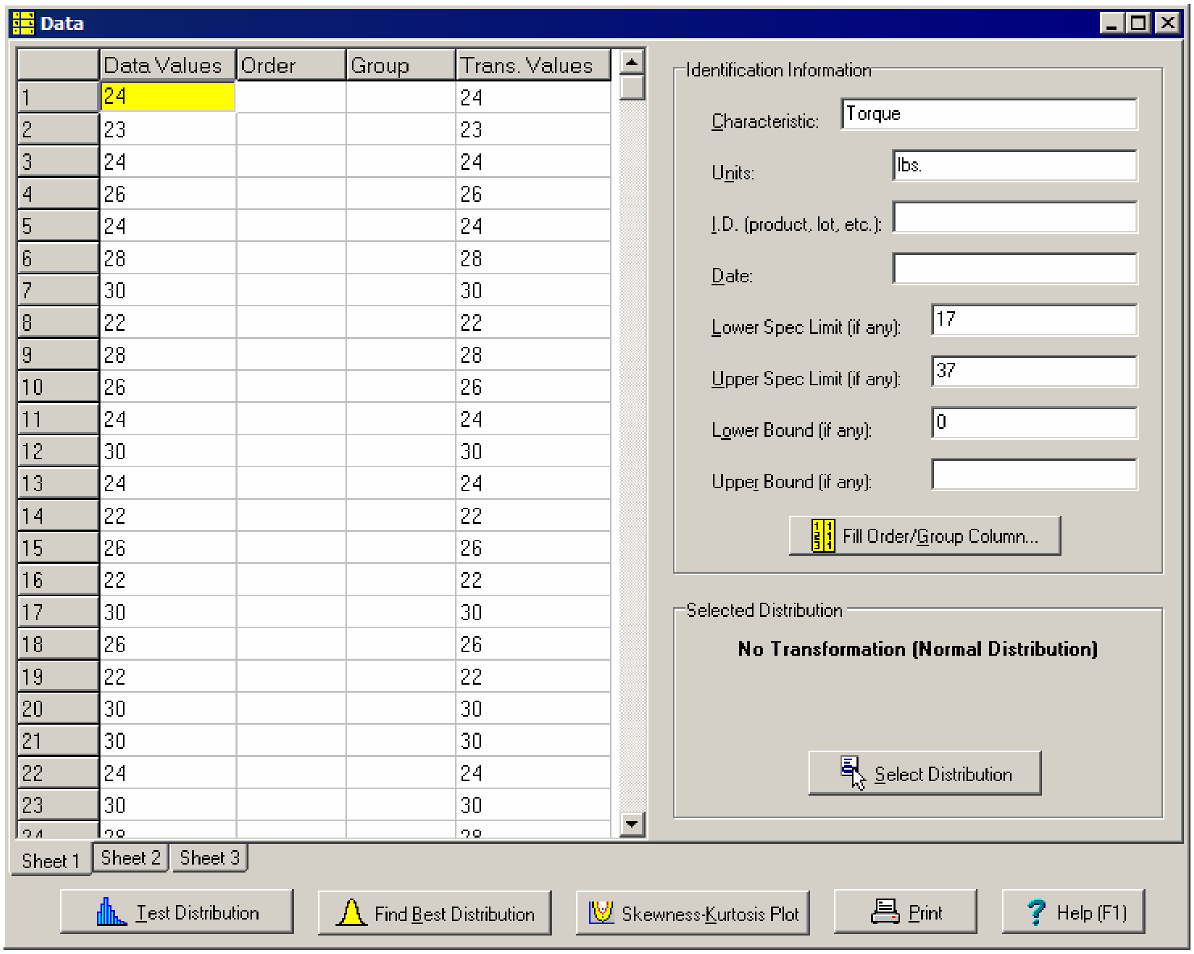


11	24
12	30
13	24
14	22
15	26
16	22
17	30
18	26
19	22
20	30



21	30
22	24
23	30
24	28
25	24
26	28
27	30
28	22
29	20
30	32

Start by entering the data in the first column of the *Data* window. The column labeled "Trans. Value" automatically displays the original values because no transformation has been selected.



	Data Values	Order	Group	Trans. Values
1	24			24
2	23			23
3	24			24
4	26			26
5	24			24
6	28			28
7	30			30
8	22			22
9	28			28
10	26			26
11	24			24
12	30			30
13	24			24
14	22			22
15	26			26
16	22			22
17	30			30
18	26			26
19	22			22
20	30			30
21	30			30
22	24			24
23	30			30

Identification Information

Characteristic: Torque
Units: lbs.
I.D. (product, lot, etc.):
Date:
Lower Spec Limit (if any): 17
Upper Spec Limit (if any): 37
Lower Bound (if any): 0
Upper Bound (if any):

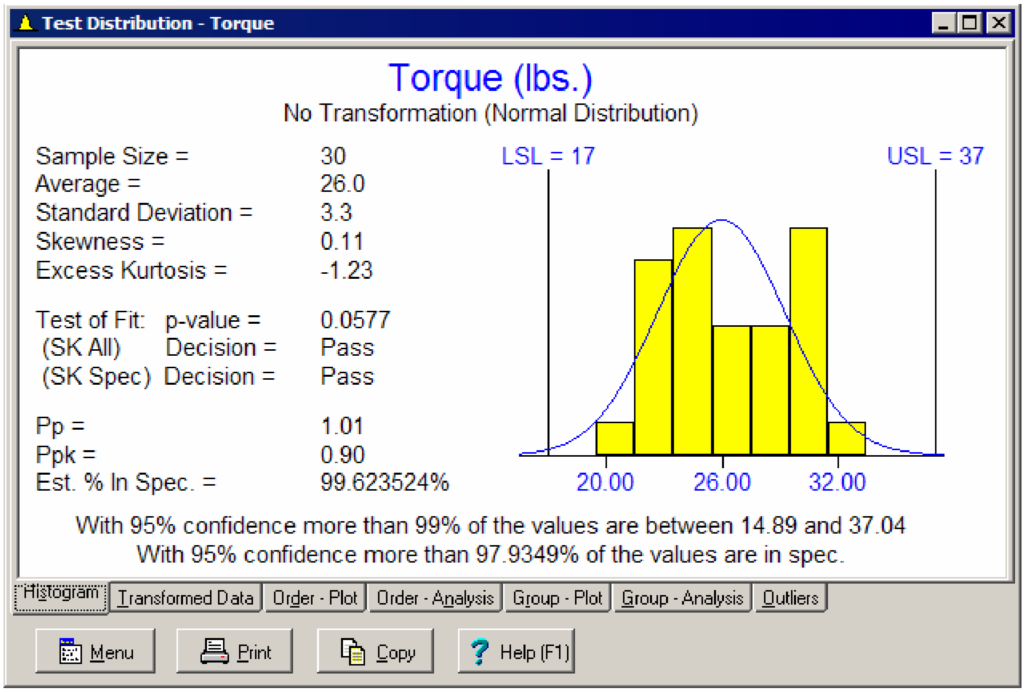
Selected Distribution

No Transformation (Normal Distribution)

The Characteristic, Units, I.D. and Date fields are optional. In this case “Torque” was entered in the *Characteristic* edit box and “lbs.” was entered in the *Units* edit box. Also enter the upper and lower spec limits.

Finally enter any physical bounds. In this case negative values are impossible, so there is a physical lower bound of zero. For yield data, reported as a percent, there is a lower bound of zero percent and upper bound of 100 percent.

Click the *Test Distribution* button to test whether the data fits the normal distribution and, if so, to perform additional analysis. The *Test Distribution* window appears as shown below:



Torque (lbs.)
No Transformation (Normal Distribution)

Sample Size = 30
Average = 26.0
Standard Deviation = 3.3
Skewness = 0.11
Excess Kurtosis = -1.23

Test of Fit: p-value = 0.0577
(SK All) Decision = Pass
(SK Spec) Decision = Pass

Pp = 1.01
Ppk = 0.90
Est. % In Spec. = 99.623524%

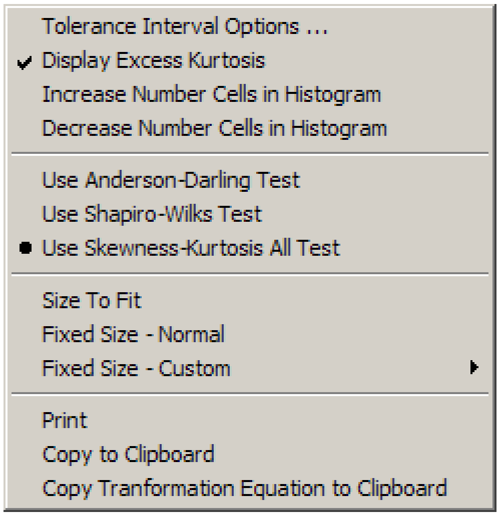
With 95% confidence more than 99% of the values are between 14.89 and 37.04
With 95% confidence more than 97.9349% of the values are in spec.

A histogram of the data is shown along with estimates of the average, standard deviation, skewness and excess kurtosis. As no transformation was selected, the best fit normal curve is shown in blue.

Also shown are the results of two tests for normality. The first is a general test for all departures from normality called the Skewness-Kurtosis All test (SK All). This test has a p-value of 0.0577. The p-value is the probability that the data or one more extreme than it would have been generated if the data came from the normal distribution. The smaller the p-value, the more evidence there is that the data does not come from the normal distribution. The rule is that one rejects the normal distribution if the p-value is 0.05 or below. This corresponds to data with a 1 in 20 chance or below of being generated by the normal distribution. If the normality test fails, one can state: “With 95% confidence the data is not from the normal distribution.” If one passes the normality test, one can state: “No significant departure from normality was detected.” The p-value of 0.0577 passes, although barely.

The second test for normality is the Skewness-Kurtosis Specific test (SK Spec). This test is designed to only reject for those departures from the normality that invalidate the confidence statements associated with normal tolerance intervals and variables sampling plans. Passing this test indicates it is OK to use these two procedures even if the other normality test fails. There is no associated p-value for this test. Just a pass/fail result is displayed. However, the interpretation of this decision is the same as before and the same confidence statements can be made. This test also passes. As a result, it is OK to proceed with normal tolerance intervals and variables sampling plans.

Distribution Analyzer also can perform the Anderson-Darling and Shapiro-Wilks tests. To see the results of these tests, click the *Menu* button (or right mouse click) and the popup menu to the right will appear. Then select the desired test. The default general test for all departures from normality is the Skewness-Kurtosis All test. This can be changed to either of the other two tests using the *Options* menu item on the *Analysis* menu. The Skewness-Kurtosis Specific test is always performed.



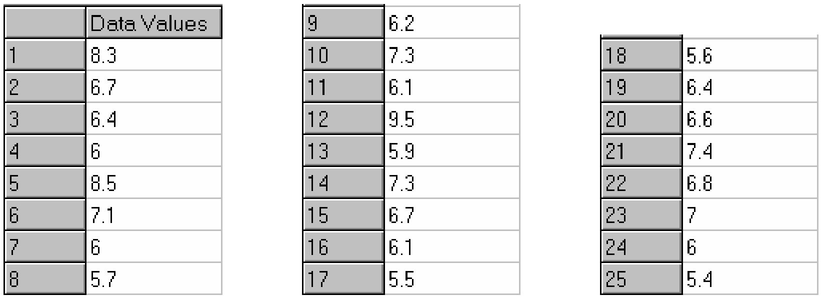
P_p, and P_{pk} are also shown. These are capability indexes for Statistical Process Control (SPC). They directly relate to the defect rate. Next is the estimated percentage of values inside the specifications. This estimate is made using the average and standard deviation of the data to calculate the distance to each of the spec limits. A table of the normal distribution is then used to obtain the corresponding estimate. This estimate depends on the data fitting the normal distribution.

At the bottom are the normal tolerance interval, "With 95% confidence more than 99% of the values are between 14.89 and 37.04.", and the variables sampling plan, "With 95% confidence more than 97.9349% of the values are in spec." These statements also depend on the data fitting the normal distribution. The confidence level and percent in the interval can be adjusted using the *Tolerance Interval Options* menu item on the above popup menu. The normal tolerance interval and variables sampling plan are closely related. The normal tolerance interval gives an interval containing a specified percentage of values. The variables sampling plan gives the percentage of values in a specified interval (namely the specifications). The variables sampling plan generally is more useful for validation/verification/qualification studies.

By default, the estimated percentage in spec and the two confidence statements are only displayed if one of the two normality tests passes. This can be changed using the *Options* menu item on the *Analysis* menu.

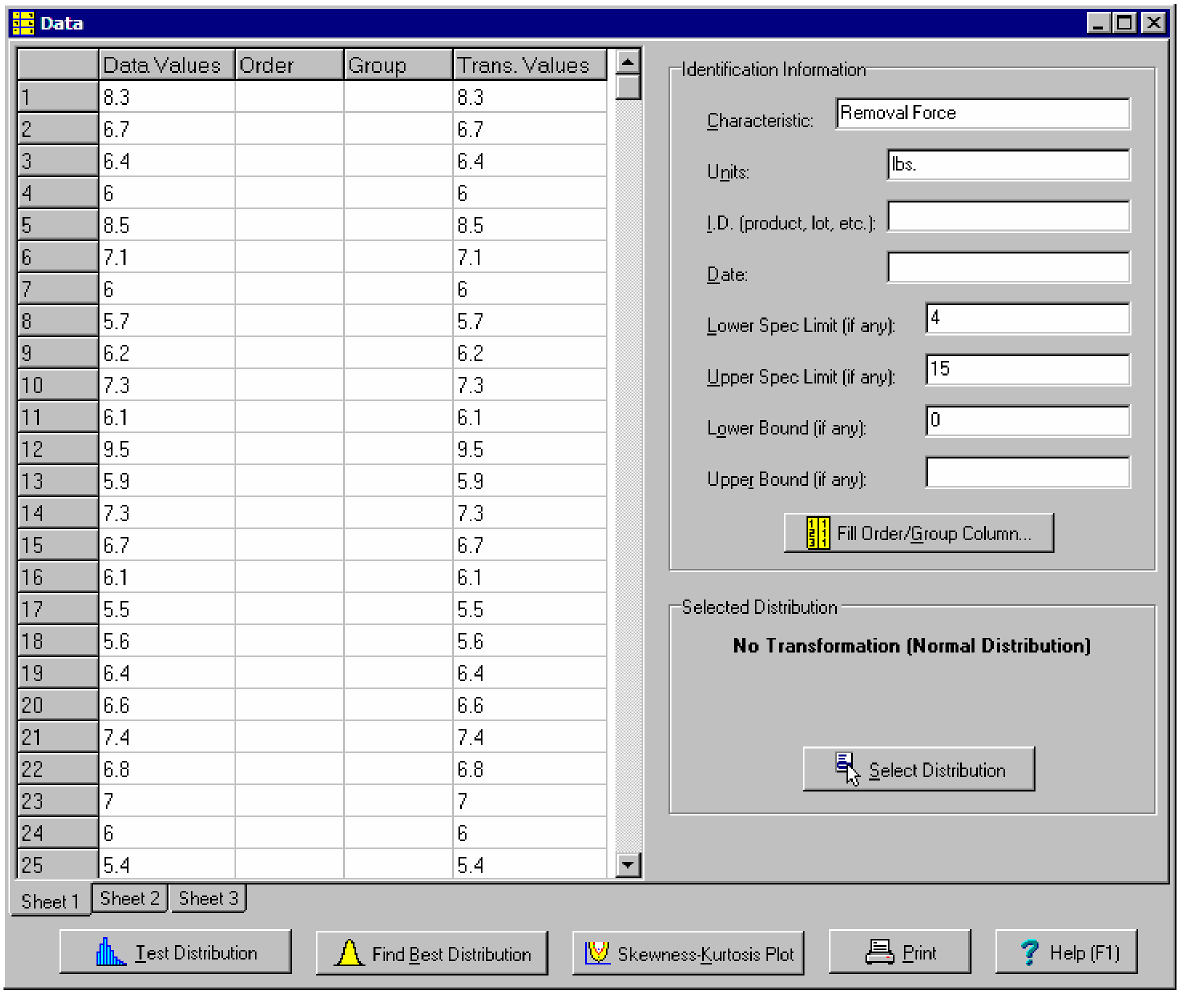
2.3 Transforming Data

Take as a second example the removal force values shown below. There is a lower spec limit of 4 pounds and upper spec limit of 15 pounds.



	Data Values
1	8.3
2	6.7
3	6.4
4	6
5	8.5
6	7.1
7	6
8	5.7
9	6.2
10	7.3
11	6.1
12	9.5
13	5.9
14	7.3
15	6.7
16	6.1
17	5.5
18	5.6
19	6.4
20	6.6
21	7.4
22	6.8
23	7
24	6
25	5.4

Start by entering the data in the first column of the *Data* window. The column labeled "Trans. Value" automatically displays the original values because no transformation has been selected.



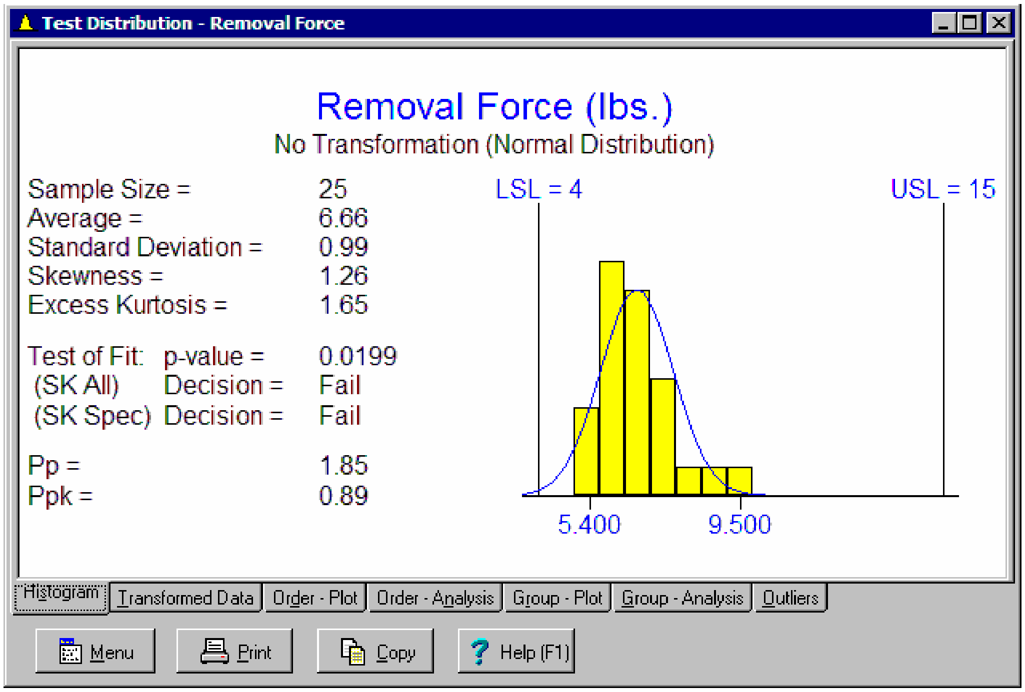
	Data Values	Order	Group	Trans. Values
1	8.3			8.3
2	6.7			6.7
3	6.4			6.4
4	6			6
5	8.5			8.5
6	7.1			7.1
7	6			6
8	5.7			5.7
9	6.2			6.2
10	7.3			7.3
11	6.1			6.1
12	9.5			9.5
13	5.9			5.9
14	7.3			7.3
15	6.7			6.7
16	6.1			6.1
17	5.5			5.5
18	5.6			5.6
19	6.4			6.4
20	6.6			6.6
21	7.4			7.4
22	6.8			6.8
23	7			7
24	6			6
25	5.4			5.4

Characteristic: Removal Force
Units: lbs.
I.D. (product, lot, etc.):
Date:
Lower Spec Limit (if any): 4
Upper Spec Limit (if any): 15
Lower Bound (if any): 0
Upper Bound (if any):

No Transformation (Normal Distribution)

"Removal Force" was entered in the *Characteristic* edit box and "lbs." was entered in the *Units* edit box. The upper and lower specification limits were also entered. Since negative values are impossible, a lower bound of zero was entered.

Click the *Test Distribution* button to test whether the data fits the normal distribution and, if so, to perform additional analysis. The *Test Distribution* window appears as shown below:



Removal Force (lbs.)

No Transformation (Normal Distribution)

Sample Size = 25
Average = 6.66
Standard Deviation = 0.99
Skewness = 1.26
Excess Kurtosis = 1.65

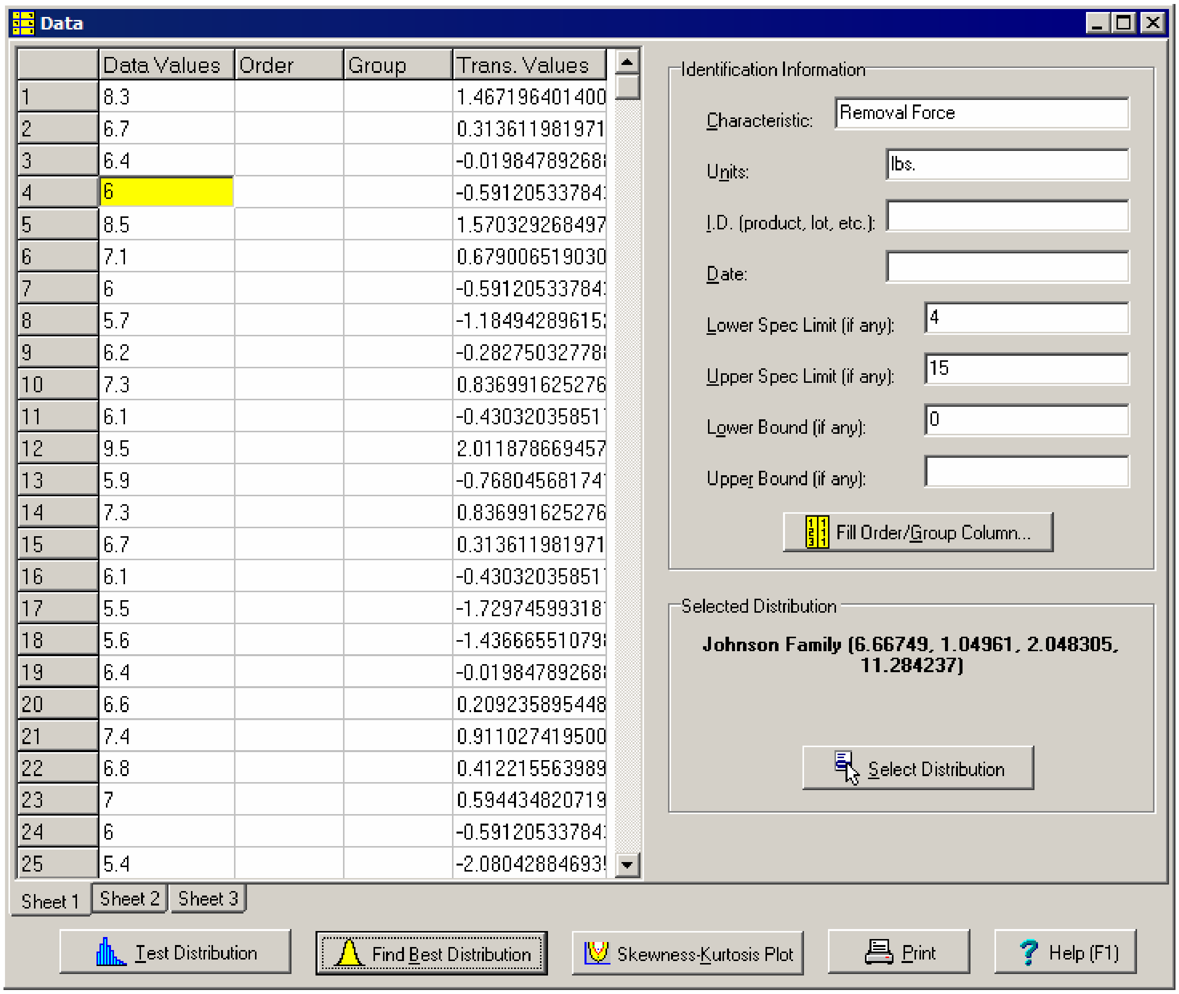
Test of Fit: p-value = 0.0199
(SK All) Decision = Fail
(SK Spec) Decision = Fail

Pp = 1.85
Ppk = 0.89

What is different this time is that both normality tests fail. As a result neither the estimated percentage in spec nor the two confidence statements are displayed. This requires that the data be transformed. Transforming the data means to apply a function like the log to the values so that the transformed values fit the normal distribution. The same transformation is applied to the specifications limits. The percentage in spec and the two confidence statements can then be correctly calculated using the transformed values.

Identifying the best transformation is done by identifying the distribution that best fits the data. For every distribution there is a transformation that will make data from that distribution fit the normal distribution.

Click the *Find Best Distribution* button to determine which distribution best fits the data. Distribution Analyzer searches through dozens of distributions to determine which best fits your data. Once done, it displays the selected distribution in the *Data* window as shown below. In this case, the best fit distribution is the Johnson Distribution (6.66749, 1.04961, 2.048305, 11.284237). The values given after the name of the distribution are the parameters of the distribution represented as the average, standard deviation, skewness and kurtosis. The transformed values are displayed in the "*Trans. Values*" column.



	Data Values	Order	Group	Trans. Values
1	8.3			1.467196401400
2	6.7			0.313611981971
3	6.4			-0.01984789268
4	6			-0.59120533784
5	8.5			1.570329268497
6	7.1			0.679006519030
7	6			-0.59120533784
8	5.7			-1.18494289615
9	6.2			-0.28275032778
10	7.3			0.836991625276
11	6.1			-0.43032035851
12	9.5			2.011878669457
13	5.9			-0.76804568174
14	7.3			0.836991625276
15	6.7			0.313611981971
16	6.1			-0.43032035851
17	5.5			-1.72974599318
18	5.6			-1.43666551079
19	6.4			-0.01984789268
20	6.6			0.209235895448
21	7.4			0.911027419500
22	6.8			0.412215563989
23	7			0.594434820719
24	6			-0.59120533784
25	5.4			-2.08042884693

Identification Information

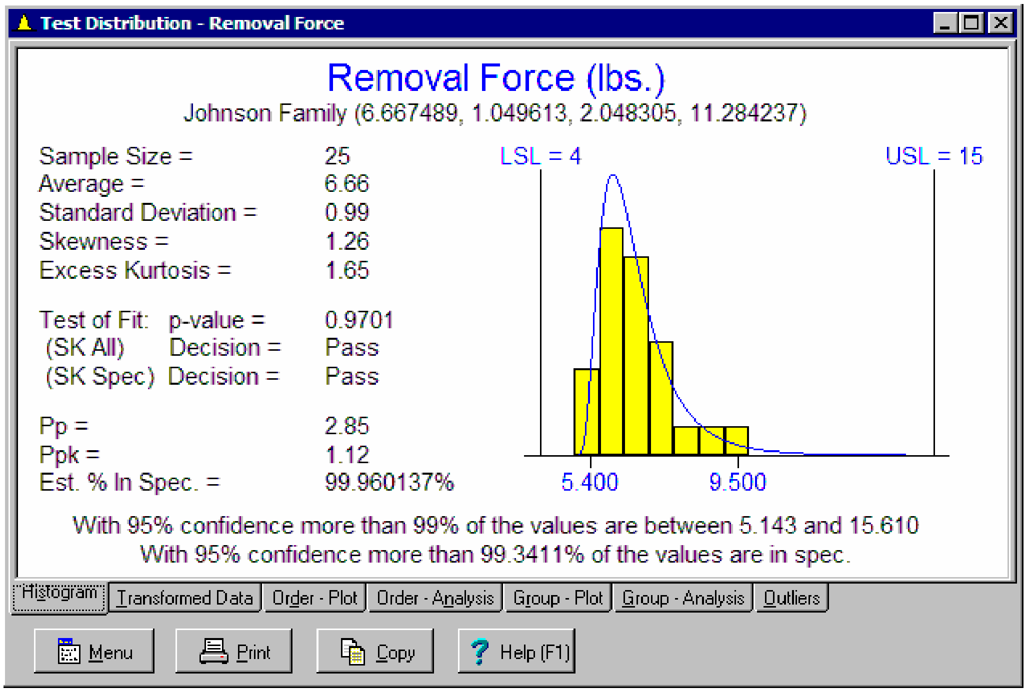
Characteristic: Removal Force
Units: lbs.
I.D. (product, lot, etc.):
Date:
Lower Spec Limit (if any): 4
Upper Spec Limit (if any): 15
Lower Bound (if any): 0
Upper Bound (if any):

Selected Distribution

Johnson Family (6.66749, 1.04961, 2.048305, 11.284237)

Transforming data has the potential for being abused. There may be several distributions that fit the data. The decision of which one to use should be based solely on which one results in transformed values best fitting the normal distribution. Distribution Analyzer returns the distribution maximizing the p-value of the Skewness-Kurtosis All test for normality.

Next click the *Test Distribution* button again to see if the Johnson distribution fits the data. The results are shown below. This time both tests pass. The Johnson distribution fits the data and can be used to analyze and transform the data. The Johnson distribution fit to the data is shown in blue. As a result, further analysis is performed and shown at the bottom. The resulting normal tolerance interval is: "With 95% confidence 99% of the values are between 5.143 and 15.610." The resulting variables sampling plan gives: "With 95% confidence statement more than 99.3411% of the values are in spec."



Removal Force (lbs.)

Johnson Family (6.667489, 1.049613, 2.048305, 11.284237)

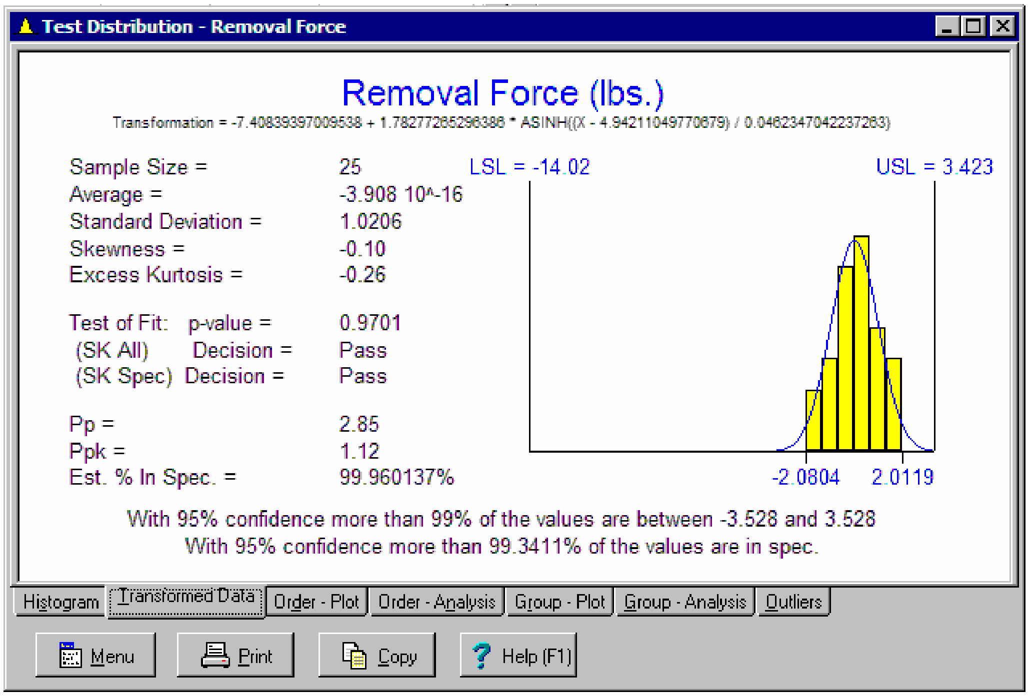
Sample Size = 25
Average = 6.66
Standard Deviation = 0.99
Skewness = 1.26
Excess Kurtosis = 1.65

Test of Fit: p-value = 0.9701
(SK All) Decision = Pass
(SK Spec) Decision = Pass

Pp = 2.85
Ppk = 1.12
Est. % In Spec. = 99.960137%

With 95% confidence more than 99% of the values are between 5.143 and 15.610
With 95% confidence more than 99.3411% of the values are in spec.

To further understand the transformation and how it works, the second tab of the *Test Distribution* window shows the transformation:



Removal Force (lbs.)

Transformation = -7.40839397009538 + 1.78277265296386 * ASINH((X - 4.94211049770679) / 0.0462347042237263)

Sample Size = 25
Average = -3.908 10^-16
Standard Deviation = 1.0206
Skewness = -0.10
Excess Kurtosis = -0.26

Test of Fit: p-value = 0.9701
(SK All) Decision = Pass
(SK Spec) Decision = Pass

Pp = 2.85
Ppk = 1.12
Est. % In Spec. = 99.960137%

With 95% confidence more than 99% of the values are between -3.528 and 3.528
With 95% confidence more than 99.3411% of the values are in spec.

The transformation equation, in EXCEL format, is shown at the top:

= -7.40839397009538 + 1.78277265296386 * ASINH((X - 4.94211049770679) / 0.0462347042237263)

A histogram of the transformed values is shown along with the best fit normal curve. The transformed values themselves are shown in the last column of the *Data* window. The same transformation is applied to the spec limits as well. Distribution Analyzer uses special routines for fitting the different distributions to ensure that the spec limits are safely within the range of the distribution and can be transformed. This allows Distribution Analyzer to find transformations for nearly every set of data.

The P_{pk}, P_p, percent in spec, normal tolerance interval and confidence statement relative to the spec limits are all calculated using the transformed values. These same results were displayed on the previous page except the tolerance interval (-3.528, 3.528) is transformed back into the original units of measure (5.143, 15.610) by applying the inverse of the transformation equation. From the above analysis, it is seen more values are predicted to exceed the upper spec limit than the lower.

Distribution Analyzer has numerous other capabilities including:

- Test for shifts over time if the order the data points were collected is indicated in the *Order* column of the *Data* window.
- Test for differences between groups like cavities, nozzles and operators if the groups are indicated in the *Group* column of the *Data* window.
- Identify potential outliers.
- Generate random values.
- Understand relationships between distributions using a skewness-kurtosis plot.

These capabilities are more fully described in the next chapter.

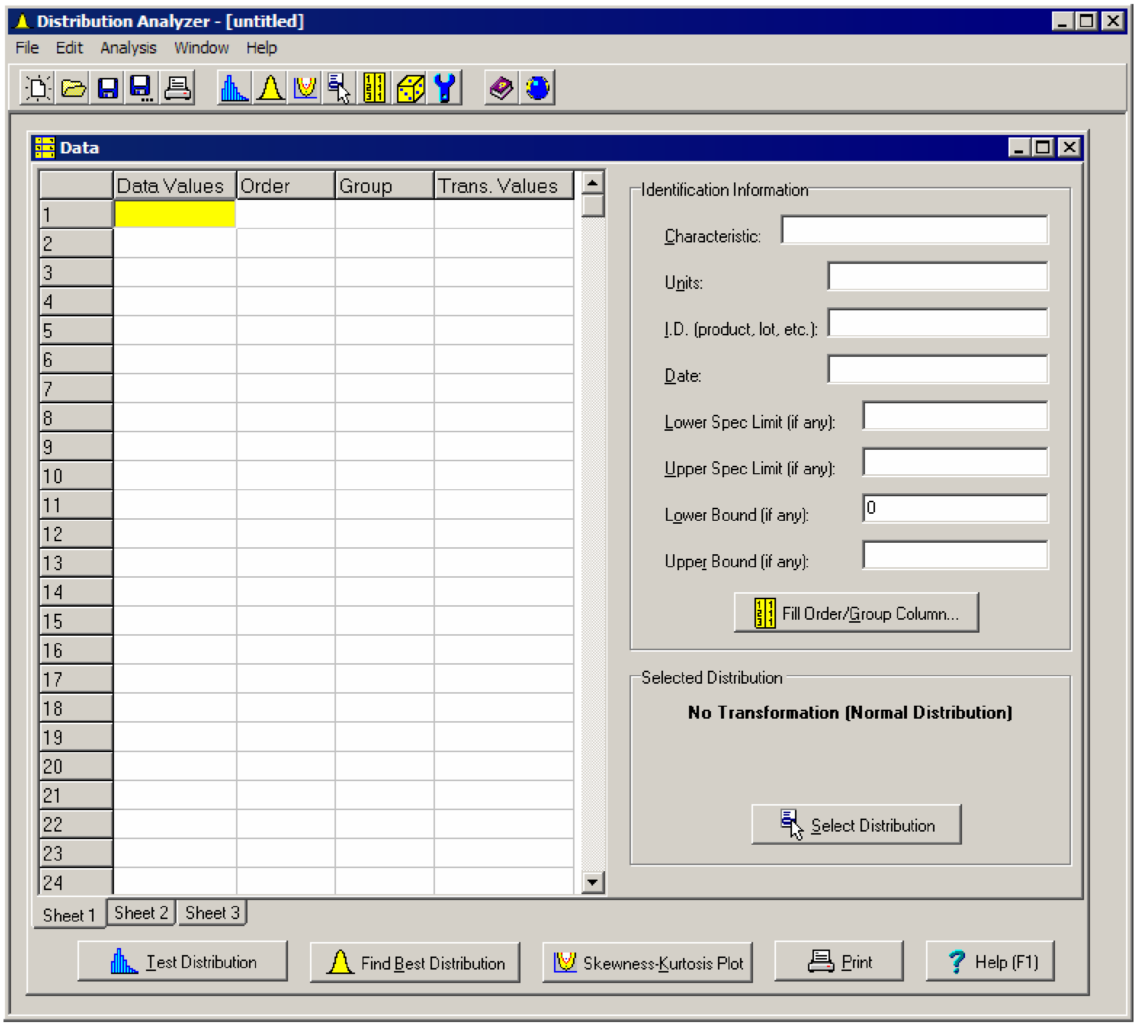
3

Program Details

This section describes more completely how to enter information into Distribution Analyzer and how to use the software to perform different analysis.

3.1 Main Window

When the program is first started, the main window shown below is displayed. The main window consists of from top to bottom: the caption bar, the menu bar, the toolbar and the interior containing the *Data* window.



Lower Bound (if any): 0

No Transformation (Normal Distribution)

Caption Bar: The caption bar contains the program name and the name of any associated file. When the program is not maximized to cover the entire screen, dragging the caption bar moves the window. Double clicking on the caption bar maximizes/restores the window. On the left of the caption bar is the system menu button. Clicking this button displays the system menu containing items to move, size, and close the window. The program can also be closed by double clicking on this button. On the right of the caption bar are the minimize button –, the maximize/restore button □, and the close button ×.

Menu Bar: The menu bar provides a list of drop-down menus containing menu items. These menu items serve as the primary means of telling the program what to do. Menu items exist for adding/editing data, performing analysis, printing the results and much more. Chapter 4 gives a complete description of all the menus.

Toolbar: The toolbar contains buttons serving as shortcuts for the most commonly used menu items. To see what a particular button does, hold the mouse cursor over the top of the button. A description of the button will appear.

Interior: The interior initially contains the *Data* child window. Later we will encounter other child windows that can also be displayed in the interior. If a child window extends outside the interior of the main window, scroll bars are provided for shifting the child windows up/down and left/right.

A sizing border surrounds the window except when the main window is maximized. Dragging the border causes the window to be resized.

3.2 Data Window

The *Data* window is a permanent child window for entering the data to be analyzed. Each set of data should be entered on a separate tab sheet. Each tab sheet is analyzed independently of the others. The columns for entering the data are:

***Data Values* Column:** Enter the values to be analyzed. Generally a minimum of 15 values must be entered but this can be changed to as few as 8 using the *Advanced Options* dialog box.

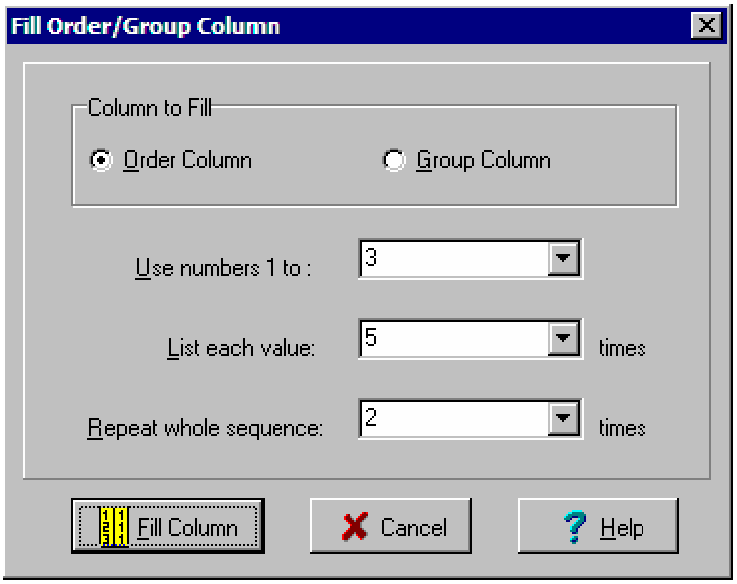
***Order* Column:** Optionally, enter the time order the values were collected in. This may be either numbers or labels. The data does not have to appear in time order as the values entered in the *Order* column are sorted into numeric or alphanumeric order. If several values are taken at the same time, use the same number or label for all values. The *Order* column can be automatically filled by clicking the *Fill Order/Group Column* button to display the *Fill Order/Group Column* dialog box. If the order is specified, additional analysis is performed to see if the values shifted over time and the results are displayed in the *Order - Analysis* tab of the *Test Distribution* window.

***Group* Column:** Optionally, enter the group (cavity, operator, line, lot, etc.) associated with each value. This may be either numbers or labels. The *Group* column can be automatically filled by clicking the *Fill Order/Group Column* button to display the *Fill Order/Group Column* dialog box. If the groups are specified, additional analysis is performed to see if the groups are different and the results are displayed in the *Group - Analysis* tab of the *Test Distribution* window.

Transformed Values Column: You cannot type into this column. Instead it is used to display the transformed values. If no distribution/transformation is selected, the values are identical to the first column. Distributions/transformations are selected by clicking the *Find Best Distribution* or *Select Distribution* buttons.

Either type the data in directly or paste it from the clipboard using the *Paste* menu item. The *Edit* menu has further menu items for selecting, copying, cutting and clearing cells as well as for adding, moving, deleting and renaming tab sheets.

Click the *Fill Order/Group Column* button to display the *Fill Order/Group Column* dialog box. This dialog box is a time saving feature designed to save having to type values into the *Order* or *Group* columns of the currently selected tab in the *Data* window. It can only be used when the data is in a patterned order.


Column to Fill
● Order Column
○ Group Column
Use numbers 1 to : 3
List each value: 5 times
Repeat whole sequence: 2 times

To specific the values to enter into the selected column, specify:

***Column to Fill* Radio Group Box:** Select whether to fill *Order* column or *Group* column.

***Use Numbers 1 to* Combo Box:** Enter the number of subgroups when filling the *Order* column or the number of groups when filling the *Group* column. It must be an integer of 2 or more.

***List Each Value ? Times* Combo Box:** Repeat each value the specified number of times before going to the next number. It must be an integer of 1 or more.

***Repeat Whole Sequence ? Times* Combo Box:** Repeat the whole sequence specified by the previous 2 controls the specified number of times. It must be an integer of 1 or more.

Click the *Fill Column* button to generate the values for the selected column in the currently selected tab of the *Data* window. If the tab already contains data, you are prompted to make sure it is OK to overwrite the data. To exit without generating any values, click the *Cancel* button or press the *Esc* key.

In the example above, the following values are generated in the *Order* column:

1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3, 1, 1, 1, 1, 1, 2, 2, 2, 2, 2, 3, 3, 3, 3, 3

Each tab sheet also contains the following controls for entering additional information:

***Characteristic* Edit Box:** Optionally, use to enter the name of the characteristics. If no characteristic is entered, the name of the sheet is used.

***Units* Edit Box:** Optionally, use to enter the units for the characteristic.

***I.D.* Edit Box:** Optionally, use to further differentiate data sets with product name, lot number, lane, cavity etc.

***Date* Edit Box:** Optionally, used to enter date or any other further information.

***Lower Spec Limit* Edit Box:** Used to enter the lower spec limit, if one exists.

***Upper Spec Limit* Edit Box:** Used to enter the upper spec limit, if one exists.

***Lower Bound* Edit Box:** Used to enter the lower physical bound, if one exists. A physical bound represents a barrier below which it is impossible for values to occur. Frequently zero is the lower physical bound when negative values are impossible. This bound can be used to pre-transform the data.

***Upper Bound* Edit Box:** Used to enter the upper physical bound, if one exists. This bound can be used to pre-transform the data.

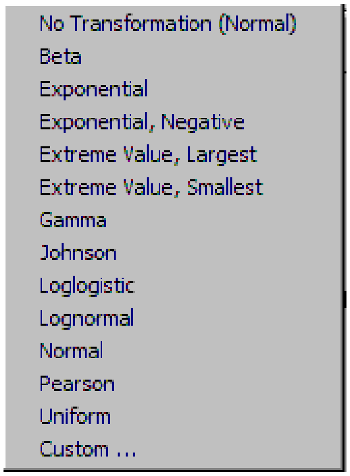
If there are more than 3 sets of data, the *Edit* menu can be used to add additional sheets. It also contains menu items for deleting, renaming and reordering the sheets.

Once the data is entered, it is time to start the analysis using the following buttons:

***Test Distribution* Button:** Displays the *Test Distribution* window where the data is analyzed to see if the selected distribution fits the data (Section 3.3). The selected distribution is displayed in the *Selected Distribution* group box. It is initially set to "No Transformation (Normal Distribution)", in which case the data is tested to see if it fits the normal distribution. Additional analysis are also performed including: estimates of moments, capability indexes, tolerance interval, confidence statement relative to spec limits, test if order effect, test if groups are different and check for outliers.

***Find Best Distribution* Button:** Used to determine which of the available distributions best fits the data with or without pre-transforming the data (Section 3.4). The result is displayed in the *Selected Distribution* group box. The distribution producing the highest p-value for the Skewness-Kurtosis All Test for normality when applied to the transformed data is selected, excluding those not including the spec limits within the range of the distribution.

***Select Distribution* Button:** Displays popup menu shown to the right for selecting a specific distribution (Section 3.4). If your data commonly fits a distribution other than the normal distribution, then select it using this menu. When a distribution is selected, it is fit to the data using both method of moments and maximum likelihood. The best fit is returned. If you don't know what distribution to use, click the *Find Best Distribution* button instead. The following distributions and family of distributions can be fit: Beta, Exponential, Negative Exponential, Largest Extreme Value, Smallest Extreme Value, Gamma, Johnson, Loglogistic, Lognormal, Normal, Pearson and Uniform. The *Custom* menu item displays the *Select Distribution to Fit Data* dialog box. This dialog gives you complete control over the process including the distribution to fit, the method used to fit the data and whether to pre-transform the data or not. The difference between selecting the *No Transformation* menu item and the *Normal* menu item is that the *No Transformation* menu item does not transform the data (Y=X) and the *Normal* menu item transforms the data using the transformation Y = (X-Average) / Standard Deviation.



***Skewness-Kurtosis Plot* Button:** Displays the *Skewness-Kurtosis Plot* window for exploring the shapes and relationships of the different distributions (Section 3.5). The skewness and kurtosis of any entered data is also displayed on the plot as an aid in identifying which distributions might fit the data.

Finally, Section 3.6 describes how to use Distribution Analyzer to generate random set of data from the different distributions.

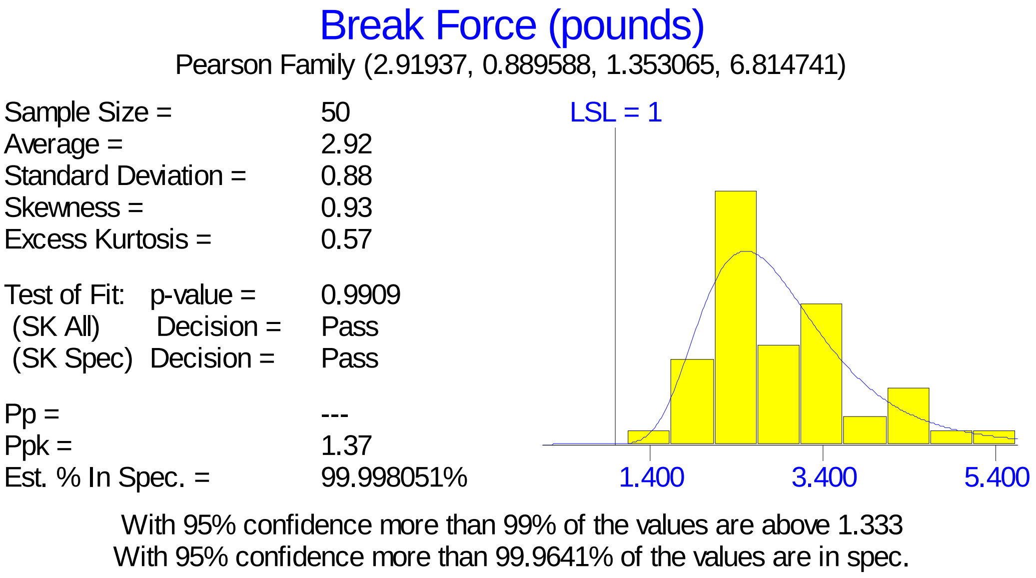
3.3 Test Distribution Window

Clicking the *Test Distribution* button on the *Data* window analyzes the data on the current sheet and displays the results in a *Test Distribution* window. If a distribution is selected in the *Selected Distribution* group box, the associated transformation is first applied to the data. The results appear on the 7 tabs shown below:

Tab 1: Histogram
Tab 2: Transformed Data
Tab 3: Order - Plot
Tab 4: Order - Analysis
Tab 5: Group - Plot
Tab 6: Group - Analysis
Tab 7: Outliers

Each tab has options that can be set. By performing a right mouse click over the graphic or clicking the *Menu* button, a popup menu will appear to set these options. Click the *Print* button to print the current tab and click the *Copy* button to copy the current tab to the clipboard.

Tab 1: Histogram



Break Force (pounds)
Pearson Family (2.91937, 0.889588, 1.353065, 6.814741)

```
Sample Size =              50
Average =                  2.92
Standard Deviation =       0.88
Skewness =                 0.93
Excess Kurtosis =          0.57

Test of Fit:  p-value =    0.9909
 (SK All)      Decision =  Pass
 (SK Spec)   Decision =    Pass

Pp =                       ---
Ppk =                      1.37
Est. % In Spec. =          99.998051%
```

With 95% confidence more than 99% of the values are above 1.333
With 95% confidence more than 99.9641% of the values are in spec.

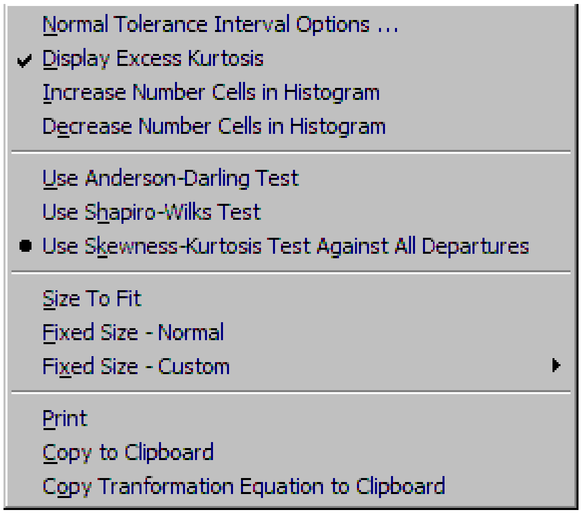
Tab 1 of the *Test Distribution* window displays a histogram of the original data along with the density of the selected distribution. In the graphic above, the Pearson distribution was selected. The parameters in parenthesis are the parameters of the distribution (average, standard deviation, skewness and kurtosis). The normal distribution is shown if no distribution was selected or fit to the data. Also shown are:

Moments: The sample size, average, standard deviation, skewness and kurtosis (or excess kurtosis) of the data.

Test of Fit: Two normality tests are performed. When no transformation is selected, these tests are applied directly to data values and test for normality. When a transformation/distribution is selected, these tests are applied to the transformed values and test for whether the selected distribution fits. The first test is a general test for all departures from normality. The Skewness-Kurtosis All test (SK All) is the default test but the Anderson-Darling test (AD) and Shapiro-Wilks test (SW) are also available. They can be selected using the *Analysis Options* dialog box or the popup menu described below. The p-value and decision are given. The test passes if the p-value > 0.05. The second test is the Skewness-Kurtosis Specific test (SK Spec). This test is designed to only reject for those departures from normality that invalid the tolerance interval and confidence statement relative to the spec limits. Passing this test indicates the statements are valid. Only the pass/fail decision is given. If the normality test fails, one can state: “With 95% confidence the data is not from the normal distribution.” If it passes, one can state: “No significant departure from normality was detected.”

Capability Indexes: P_p, P_{pk} and the estimated defect rate are shown. The estimated defect rate may be in terms of percent in spec, percent out of spec or reliability, depending on the option selected in the *Analysis Options* dialog box and/or the *Tolerance Interval Options* dialog box. The estimated defect rate assumes the selected distribution fits the data. By default, it is only shown if the selected distribution fits the data. This can be altered using the *Analysis Options* dialog box and/or the *Tolerance Interval Options* dialog box. When a transformation/distribution is selected, the capability indexes and defect rate are calculated using the transformed values and spec limits.

Tolerance Interval and Confidence Statement Relative To Spec Limits: These two statements assume the selected distribution fits the data. By default, they are only shown if the selected distribution fits the data. This can be altered along with the confidence level, etc. using the *Analysis Options* dialog box and/or the *Tolerance Interval Options* dialog box. If the general test for normality passes, the statements are accurate. If the general test fails but the Skewness-Kurtosis Specific test passes, the statements provide conservative bounds. When a transformation/distribution is selected, these statements are calculated using the transformed values and spec limits. The normal tolerance interval is then transformed back to the original units.



By performing a right mouse click over the graphic or clicking the *Menu* button, the following popup menu will appear. The menu items are as follows:

Tolerance Interval Options.: Displays the *Tolerance Interval Options* dialog box for altering the confidence level, percentage in interval, type (upper, lower, 2-sided), units to use (% in spec, % out of spec, reliability) and whether to display tolerance interval only it the distribution fits. This dialog box is described in more detail later in this section.

Display Excess Kurtosis: If checked, the excess kurtosis is display. Otherwise, the kurtosis is displayed.

Increase Number Cells in Histogram: Selecting this menu item increases the number of cells in the histogram meaning fewer values fall in each cell. This menu item can be selected multiple times. This menu item is grayed out (not available) when the histogram is such that every unique value is contained in its own cell.

Decrease Number Cells in Histogram: Selecting this menu item decreases the number of cells in the histogram meaning more values fall in each cell. This menu item can be selected multiple times.

Use Anderson-Darling Test: Selects Anderson-Darling test (AD) as general test for normality.

Use Shapiro-Wilks Test: Selects Shapiro-Wilks test (SW) as general test for normality.

Use Skewness-Kurtosis All Test: Selects Skewness-Kurtosis All test (SK All) as general test for normality.

Size To Fit: Sizes plot to fit window. Plot will shrink and expand to fit window when window is resized.

Fixed Size - Normal: Sizes the plot so it is easy to read. If the plot is too large to fit the window, scroll bars are added.

Fixed Size - Custom: Can specify the size of the plot. If the plot is too large to fit the window, scroll bars are added.

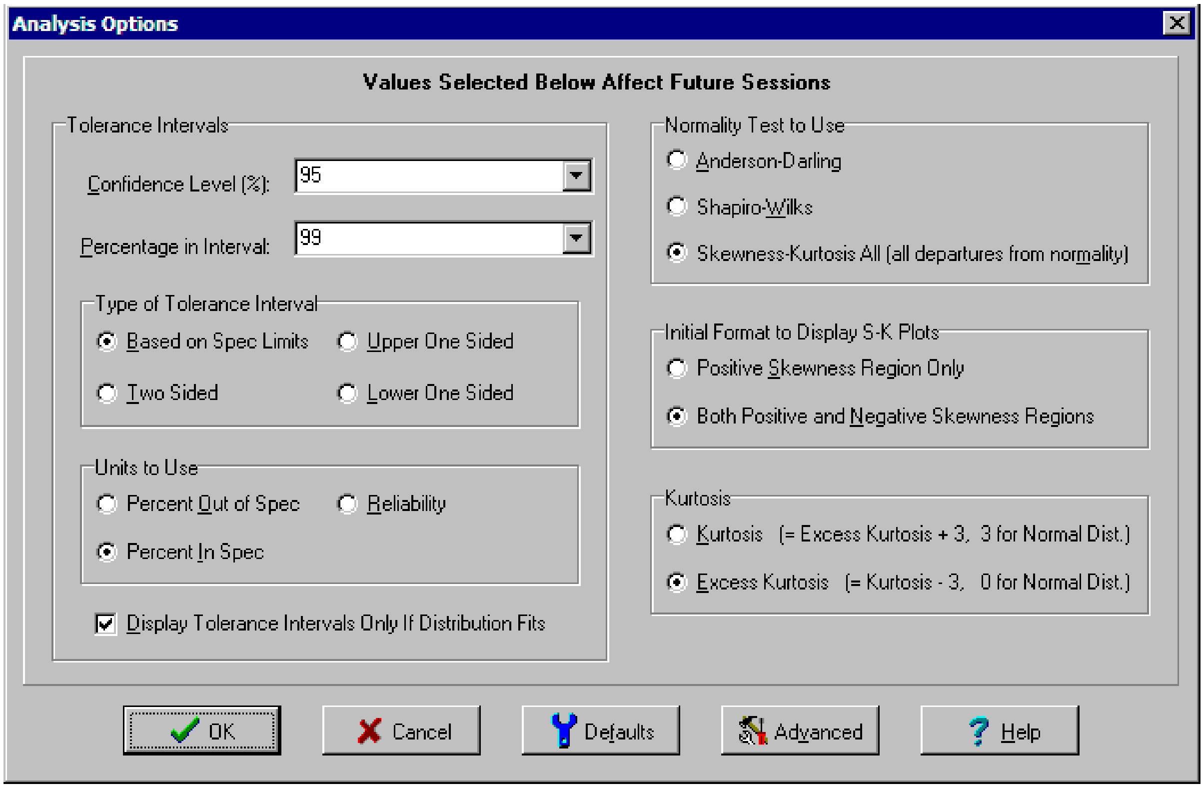
Print: Prints the plot.

Copy to Clipboard: Copies the plot to the clipboard in Windows Meta file (Picture) format.

Copy Transformation Equation to Clipboard: Copies the equation for the transformation to the clipboard in EXCEL format.

Setting Tolerance Intervals Options

Default options for the T*est Distribution* and *Skewness-Kurtosis Plot* windows are set using *Analysis Options* dialog box displayed using the *Analysis* menu. Changing these options will not affect existing windows but will affect all new windows and future sessions.



Analysis Options

Values Selected Below Affect Future Sessions

Tolerance Intervals

Confidence Level (%): 95

Percentage in Interval: 99

Type of Tolerance Interval

● Based on Spec Limits ○ Upper One Sided
○ Two Sided ○ Lower One Sided

Units to Use

○ Percent Out of Spec ○ Reliability
● Percent In Spec

☑ Display Tolerance Intervals Only If Distribution Fits

Normality Test to Use

○ Anderson-Darling
○ Shapiro-Wilks
● Skewness-Kurtosis All (all departures from normality)

Initial Format to Display S-K Plots

○ Positive Skewness Region Only
● Both Positive and Negative Skewness Regions

Kurtosis

○ Kurtosis (= Excess Kurtosis + 3, 3 for Normal Dist.)
● Excess Kurtosis (= Kurtosis - 3, 0 for Normal Dist.)

The following options can be set:

***Confidence Level* Combo Box:** Confidence level as a percentage to use for constructing tolerance intervals and confidence statements. Must be a real number greater than equal to 50.0 but below 100.0. Recommend 95.0.

***Percentage in Interval* Combo Box:** For tolerance intervals, percentage of values in the interval. It must be a real number greater than 0.0 but below 100.0. Initially set to 99.0.

***Type of Tolerance Interval* Radio Group:** For tolerance intervals, can select whether to display an upper, lower or two sided tolerance interval. A fourth option is provided to use the type of tolerance interval matching the type of specifications. A two sided tolerance is used for two sided specification and so on. If no specs are provided, a two sided tolerance interval is displayed.

***Units to Use* Radio Group:** Can select whether to report estimated defect rates and confidence statements relative to the spec limits in terms of the percent in spec, percent out of spec or the reliability.

***Display Tolerance Intervals Only If Distribution Fits* Check Box:** The estimated defect rate, tolerance interval and confidence statement relative to the spec limits are only valid if the distribution used adequately fits the data. Checking this box will only display them when one of the normality tests passes when applied to the transformed values. It is recommended this box be checked.

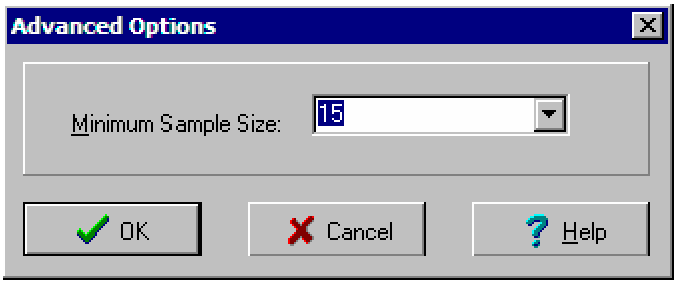
***Normality Test to Use* Radio Group:** It is recommended that the Skewness-Kurtosis All Test (SK All) be used to test whether the distribution fits. However, two alternative tests are also provided: Anderson-Darling and Shapiro-Wilks.

***Initial Format to Display S-K Plot* Radio Group:** In *Skewness-Kurtosis Plot* window, select whether to initially display either the positive and negative skewness regions or just the positive skewness region.

***Kurtosis* Radio Group:** Specify whether to report kurtosis or excess kurtosis.

When done, click the *OK* button. If any errors are found, an error message will be displayed and the errors must be corrected before the dialog box can be closed. To set everything back to the original defaults, click the *Defaults* button. You can also click the *Advanced* button to display and change additional advanced options. This displays the *Advanced Options* dialog box. To exit without updating the options, click the *Cancel* button or press the *Esc* key.

The *Advanced Options* dialog box is used to set advanced options that should not generally be changed by the user.



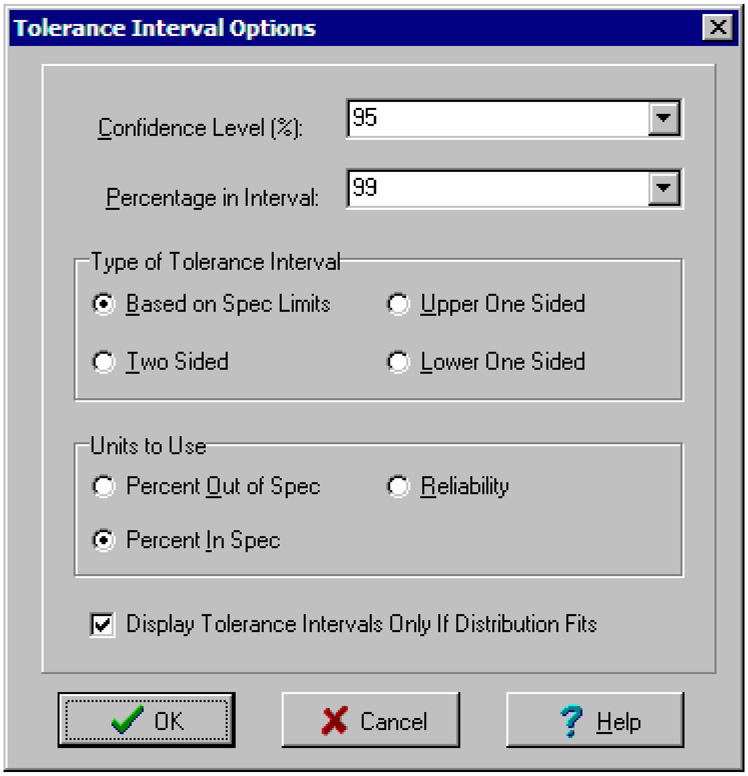
Minimum Sample Size: 15

The available options are:

***Minimum Sample Size* Combo Box:** It is recommended that at least 15 samples be used when fitting and testing distributions but the program allows the user to adjust this policy. The program will not perform an analysis if the sample size is less than the specified minimum sample size. The minimum sample size must be an integer of at least 8.

Once the options have been specified, click the *OK* button or press the *Enter* key to save the results and close the dialog box. If any errors are found, an error message will be displayed and the errors must be corrected before the dialog box can be closed. Clicking the *Cancel* button or pressing the *Esc* key instead restores the options to their former values and closes the dialog box.

To modify the options used by the current T*est Distribution* window, use the *Tolerance Interval Options* dialog box shown below and the menu items on the popup menu. Changing these options will only affect the current *Test Distribution* window.



Tolerance Interval Options

Confidence Level (%): 95

Percentage in Interval: 99

Type of Tolerance Interval

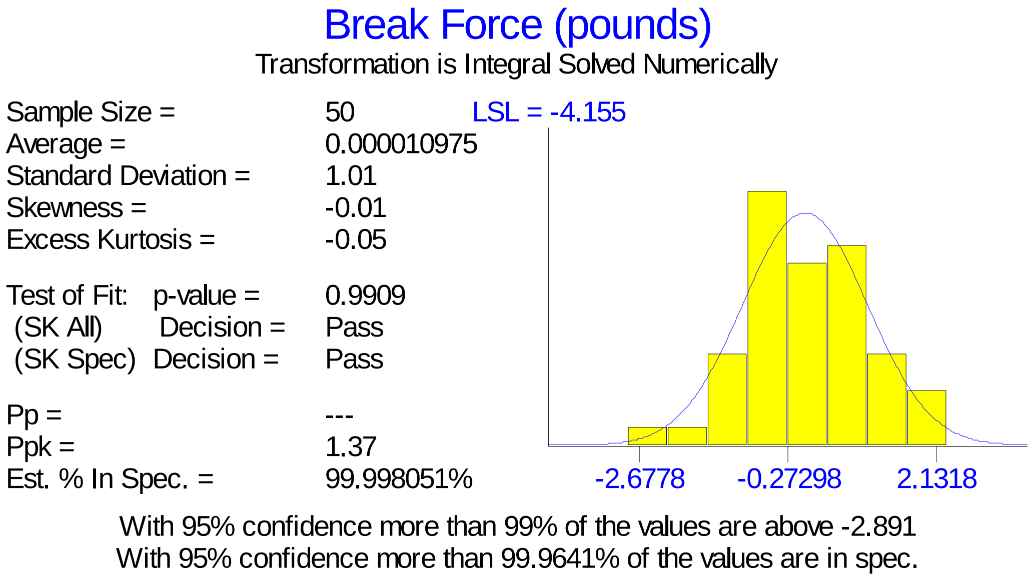
● Based on Spec Limits
○ Upper One Sided
○ Two Sided
○ Lower One Sided

Units to Use

○ Percent Out of Spec
○ Reliability
● Percent In Spec

☑ Display Tolerance Intervals Only If Distribution Fits

Tab 2: Transformed Data



Break Force (pounds)

Transformation is Integral Solved Numerically

```
Sample Size =             50
Average =                 0.000010975
Standard Deviation =      1.01
Skewness =                -0.01
Excess Kurtosis =         -0.05

Test of Fit:  p-value =   0.9909
 (SK All)      Decision = Pass
 (SK Spec)  Decision =    Pass

Pp =                      ---
Ppk =                     1.37
Est. % In Spec. =         99.998051%
```

With 95% confidence more than 99% of the values are above -2.891

With 95% confidence more than 99.9641% of the values are in spec.

Tab 2 of the *Test Distribution* window displays a histogram of the transformed values along with the density of the normal distribution. The equation used to transfer the data is shown at the top. Also shown are:

Moments: The sample size, average, standard deviation, skewness and kurtosis (or excess kurtosis) of the transformed values.

Test of Fit: Identical to Tab 1.

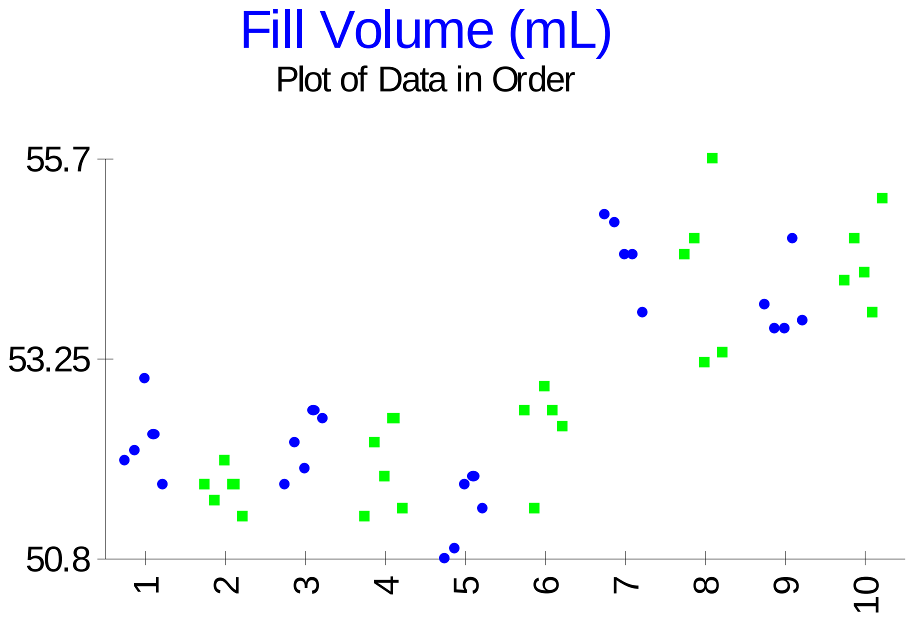
Capability Indexes: Identical to Tab 1.

Tolerance Interval: For transformed values. The inverse of the transformation equation is applied to this interval to obtain the tolerance interval in Tab 1.

Confidence Statement Relative To Spec Limits: Identical to Tab 1.

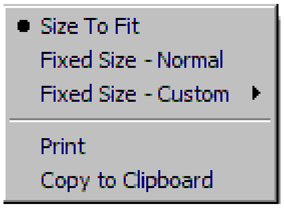
By performing a right mouse click over the graphic or clicking the *Menu* button, the same popup menu as on tab 1 will appear.

Tab 3: Order - Plot



Tab 3 of the *Test Distribution* window displays a plot of the values in time order like the one shown above. In order for this plot to be displayed, the *Order* column must be filled out in the *Data* window. An analysis for whether shifts occurred over time is displayed on Tab 4: Order - Analysis.

By performing a right mouse click over the graphic or clicking the *Menu* button, the following popup menu will appear. The menu items are as follows:



Size To Fit: Sizes plot to fit window. Plot will shrink and expand to fit window when window is resized.

Fixed Size - Normal: Sizes plot so easy to read. If the plot is too large to fit the window, scroll bars are added.

Fixed Size - Custom: Can specify the size of the plot. If the plot is too large to fit the window, scroll bars are added.

Print: Prints the plot.

Copy to Clipboard: Copies the plot to the clipboard in Windows Meta file (Picture) format.

Tab 4: Order Analysis

Fill Volume (mL)

Table of Significant Shifts

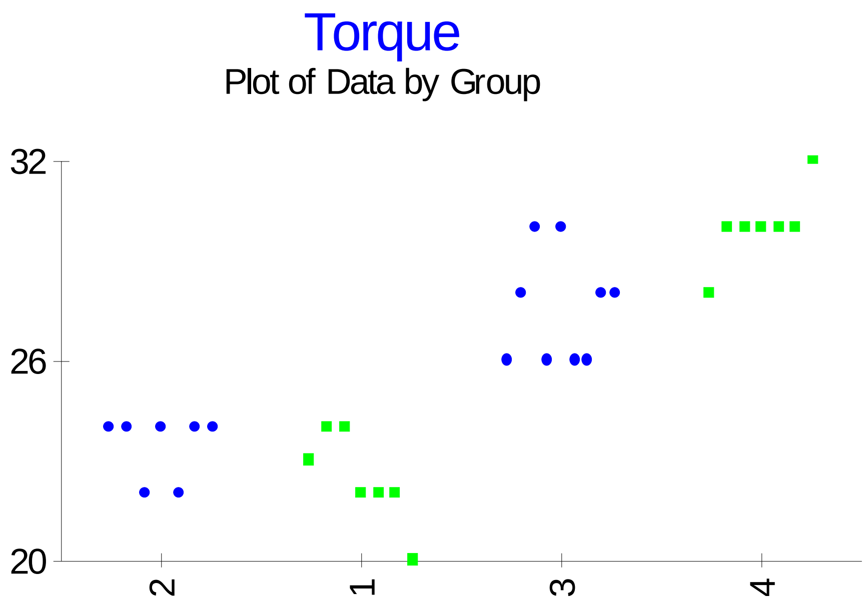
Subgroup	Confidence Level
7	98.3%

Tab 4 of the *Test Distribution* window displays the results of a change-point analysis to determine if there are shifts in the data over time. In order for this analysis to be performed, the *Order* column must be filled out in the *Data* window. A plot of the data in time order is displayed on Tab 3: Order - Plot. If there are significant shifts, this tab will be boldrd in the *Test Distribution* window.

A change-point analysis can detect multiple shifts. If it detects shifts, the shifts are listed in a table. For each shift, the estimated first point or subgroup following the change is listed along with the confidence level representing the confidence that the shift occurred. Only shifts detected with 95% confidence or better are listed.

By performing a right mouse click over the graphic or clicking the *Menu* button, the same popup menu as on Tab 3 will appear.

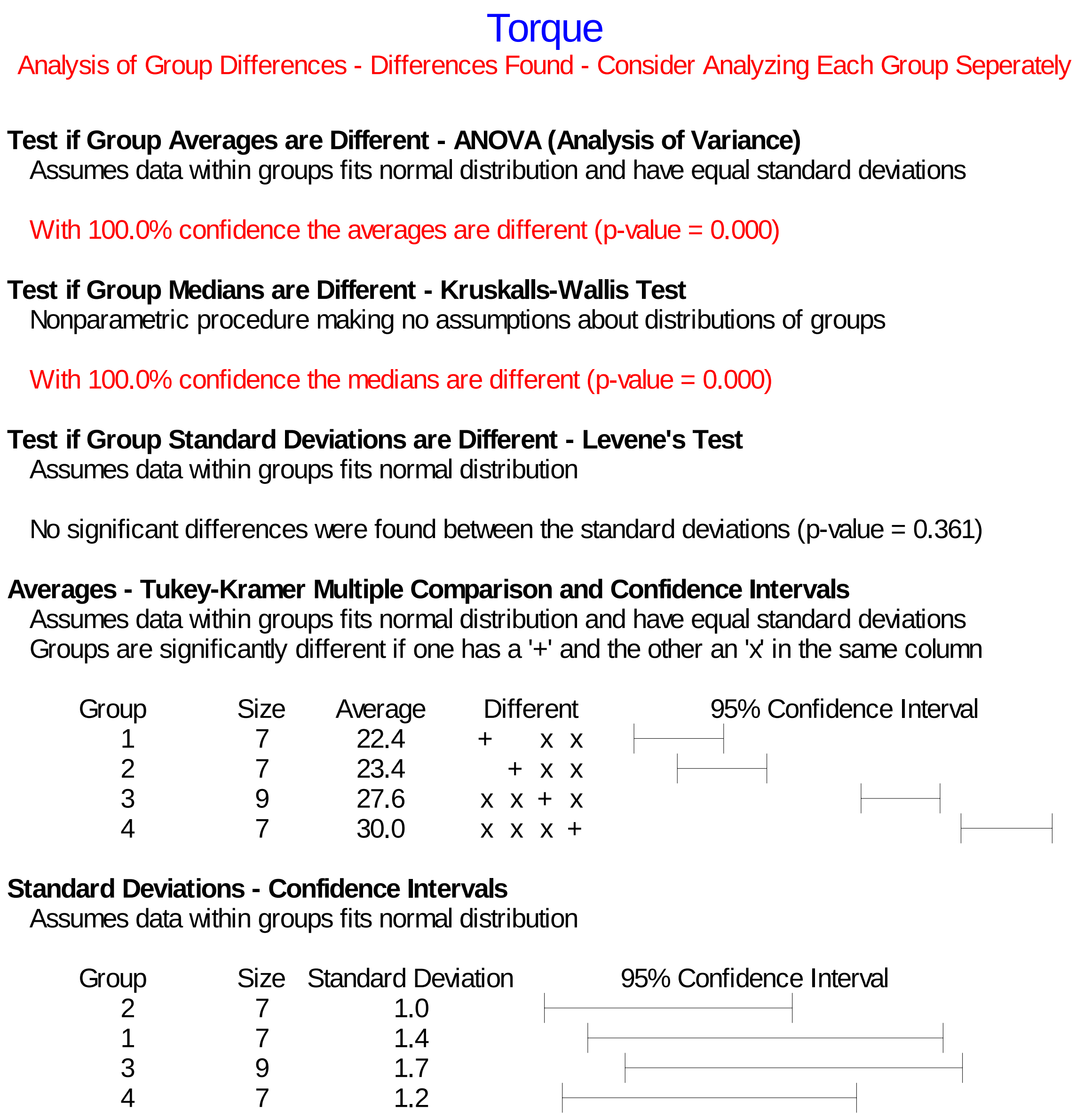
Tab 5: Group Plot



Tab 5 of the *Test Distribution* window displays a plot of the values by group like the one shown above. In order for this plot to be displayed, the *Group* column must be filled out in the *Data* window. An analysis for whether differences exist between the groups is displayed on Tab 6: Group - Analysis.

By performing a right mouse click over the graphic or clicking the *Menu* button, the same popup menu as on Tab 3 will appear.

Tab 6: Group - Analysis



```
                                   Torque
 Analysis of Group Differences - Differences Found - Consider Analyzing Each Group Seperately

Test if Group Averages are Different - ANOVA (Analysis of Variance)
   Assumes data within groups fits normal distribution and have equal standard deviations

   With 100.0% confidence the averages are different (p-value = 0.000)

Test if Group Medians are Different - Kruskalls-Wallis Test
   Nonparametric procedure making no assumptions about distributions of groups

   With 100.0% confidence the medians are different (p-value = 0.000)

Test if Group Standard Deviations are Different - Levene's Test
   Assumes data within groups fits normal distribution

   No significant differences were found between the standard deviations (p-value = 0.361)

Averages - Tukey-Kramer Multiple Comparison and Confidence Intervals
   Assumes data within groups fits normal distribution and have equal standard deviations
   Groups are significantly different if one has a '+' and the other an 'x' in the same column

         Group      Size   Average    Different          95% Confidence Interval
           1          7      22.4      +     x  x
           2          7      23.4         +  x  x
           3          9      27.6      x  x  +  x
           4          7      30.0      x  x  x  +

Standard Deviations - Confidence Intervals
   Assumes data within groups fits normal distribution

         Group      Size   Standard Deviation       95% Confidence Interval
           2          7           1.0
           1          7           1.4
           3          9           1.7
           4          7           1.2
```

Tab 6 of the *Test Distribution* window displays the results of several analyses to detect differences between groups. In order for these analyses to be performed, the *Group* column must be filled out in the *Data* window. A plot of the data by group is displayed on Tab 5: Group - Plot. If significant differences between the groups are detected, this tab will be bolded in the *Test Distribution* window. Any significant differences are highlighted in red.

The following analyses are performed:

Group Averages Different: The first analysis performed is an Analysis of Variance (ANOVA) to see if the averages of the groups are different. This analysis assumes the data within the groups fits the normal distribution and that the standard deviations are equal. The p-value and confidence level are reported. A significant difference is reported if the p-value is less than or equal to 0.05 and the results are highlighted in red.

Group Medians Different: The second analysis performed is the Kruskall-Wallis test to see if the medians of the groups are different. This is a nonparametric test that makes no assumptions about the distributions of the groups. It is an alternative to the above ANOVA. The p-value and confidence level are reported. A significant difference is reported if the p-value is less than or equal to 0.05 and the results are highlighted in red.

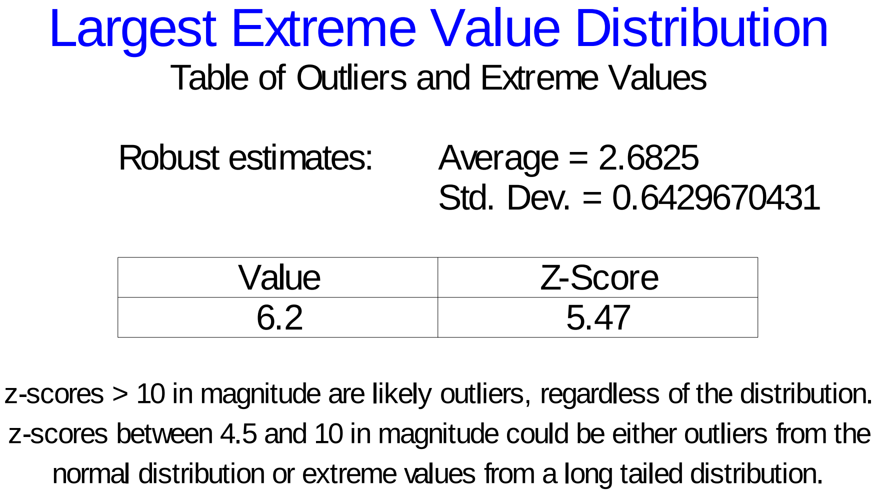
Group Standard Deviations Different: The third analysis performed is the Levene's test to see if the standard deviations of the groups are different. This analysis assumes the data within the groups fits the normal distribution. The p-value and confidence level are reported. A significant difference is reported if the p-value is less than or equal to 0.05 and the results are highlighted in red.

Comparison of Averages: The average of each group is displayed along with 95% confidence intervals. This analysis assumes the data within the groups fits the normal distribution and that the standard deviations are equal. If the ANOVA or Kruskall-Wallis test is significant, this section can be used to further compare the group averages. A Tukey-Kramer Multiple Comparison is also performed to determine which groups are significantly different. For each group, there is a corresponding column with a "+" sign in it. Any other groups with "x" signs in the same column are significantly different. Above, group 1 is significantly different from groups 3 and 4 but is not significantly different than group 2.

Comparison of Standard Deviations: The standard deviation of each group is displayed along with 95% confidence intervals. These confidence intervals assume the data within the groups fits the normal distribution. If Levene's test is significant, this section can be used to further compare the group standard deviations.

By performing a right mouse click over the graphic or clicking the *Menu* button, the same popup menu as on Tab 3 will appear.

Tab 7: Outliers



Largest Extreme Value Distribution

Table of Outliers and Extreme Values

Robust estimates: Average = 2.6825
Std. Dev. = 0.6429670431

Value	Z-Score
6.2	5.47

z-scores > 10 in magnitude are likely outliers, regardless of the distribution.
z-scores between 4.5 and 10 in magnitude could be either outliers from the normal distribution or extreme values from a long tailed distribution.

Tab 7 of the *Test Distribution* window displays any potential outliers in the data. These points should be reviewed to see if they are in error and replaced if this is the case. They might also represent extreme values for a long tailed distribution so transforming the data is another option. If they cannot be eliminated and a distribution cannot be fit to the data, the use of a nonparametric tolerance interval or attribute sampling plan may be required.

For each point a z-score is calculated. The z-score is:

$$\frac{\text{Value} - \text{Average}}{\text{Standard Deviation}}$$

A z-score of 10 means the value is 10 standard deviations above the average. A point is marked as likely being an outlier if it is more than 10 standard deviations from the average (z-score greater than 10 or less than -10). Robust estimates of the average and standard deviation are used for this calculation called 20% trimmed estimates. This is so that an outlier does distinguish itself by inflating the estimate of the standard deviation. Points that are from 4.5 to 10 standard deviations from the average are marked as either outliers relative to the normal distribution or extreme values from a long tailed distribution.

The outlier shown above is from a set of data generated using the largest extreme value distribution. Even though the one point is flagged as a potential outlier, it is in reality just part of a long tail.

By performing a right mouse click over the graphic or clicking the *Menu* button, the same popup menu as on Tab 3 will appear.

3.4 Transforming Data

If the data does not fit the normal distribution; consider transforming it, among other things. Before deciding to transform the data, consider the following items:

1. **Did a shift occur in the middle of the data?** This would indicate the process in unstable. Generally the cause of the shift should be identified and eliminated.
2. **Are there multiple sources that are different?** For example, there might be consistent differences between the different cavities of an injection molding process. If a significant difference is detected between the cavities, consider testing each cavity separately.
3. **Is data truncated**? For example, the supplier might 100% inspect the components before shipping them. If product is 100% inspected and the 100% inspection removes significant numbers of out of spec units, then the data must be handled as attribute data rather than variables data. An attribute sampling plan can be used instead to demonstrate any claims about units being in spec.
4. **Is there poor measurement resolution?** This will be evidenced by frequent ties in the data. The Anderson-Darling and Shapiro-Wilks tests should not be used in this case. Make sure the Skewness-Kurtosis based tests are used.
5. **Are there outliers?** If it can be demonstrated outlier values are measurement related, they can be eliminated or replaced. One way of doing this when testing is nondestructive is to retest it multiple times and demonstrate the result is consistently different than the first test result. In which case, the result can be replaced by the average of the retests. Outliers are easily confused with long tails of nonnormal distributions, so don't be afraid to try transforming the data in this case, once reviewing the potential outliers.
6. **Is there too much data?** Nothing is truly normal. With enough data very small departures from normality can be detected that are of no practical concern. If there are over several hundred data points, it is better to use the estimates of the skewness and kurtosis to judge it the data is sufficiently normal.

If none of these issues are identified and it appears the underlying reason the data failed the normality tests is that it comes from some distribution other than the normal distribution, then a transformation is appropriate.

The recommended method of determining the transformation is to click the *Find Best Distribution* button in the *Data* window. This determines which of the available distributions best fits the data with or without pre-transforming the data. It also uses all available methods of fitting the distribution (method of moments and maximum likelihood). The distribution producing the highest p-value for the Skewness-Kurtosis All test for normality when applied to the transformed data is selected, excluding those not including the spec limits within the range of the distribution. The result is displayed in the *Selected Distribution* group box of the *Data* window.

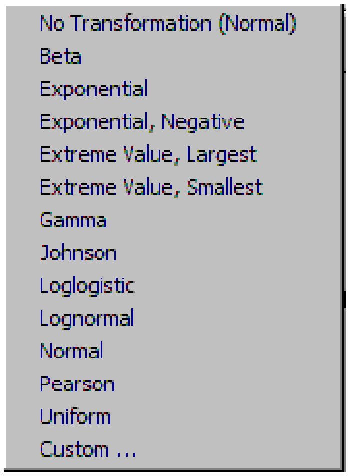
Pre-transforming the data is often helpful when the data has physical bounds. Pre-transforming data takes data that is bounded and converts it to unbounded data. Bounds, if they exist, can be specified along with the data in the *Data* window. Physical bounds and pre-transforming data is defined in more detail on the Glossary. A simple example is to take positive data, bounded below by zero, and apply the log pre-transformation to the data. Including pre-transformations essentially doubles the number of distributions fit by Distribution Analyzer. It means distributions like the log-gamma and log-Johnson can be fit to the data.

Distribution Analyzer fits a distribution to the data using both of the commonly used approaches: method of moments and maximum likelihood. However, it modifies both approaches to ensure that the distribution fit includes the spec limits within the range of the distribution. Distributions like the normal distribution are unbounded and never cause a problem. Distributions like the lognormal and gamma are bounded below and it is possible that the spec limit falls below the lower bound of the distribution. When this is the case, the transformation associated with the distribution cannot be used to transform the spec limit so no further analysis can be performed.

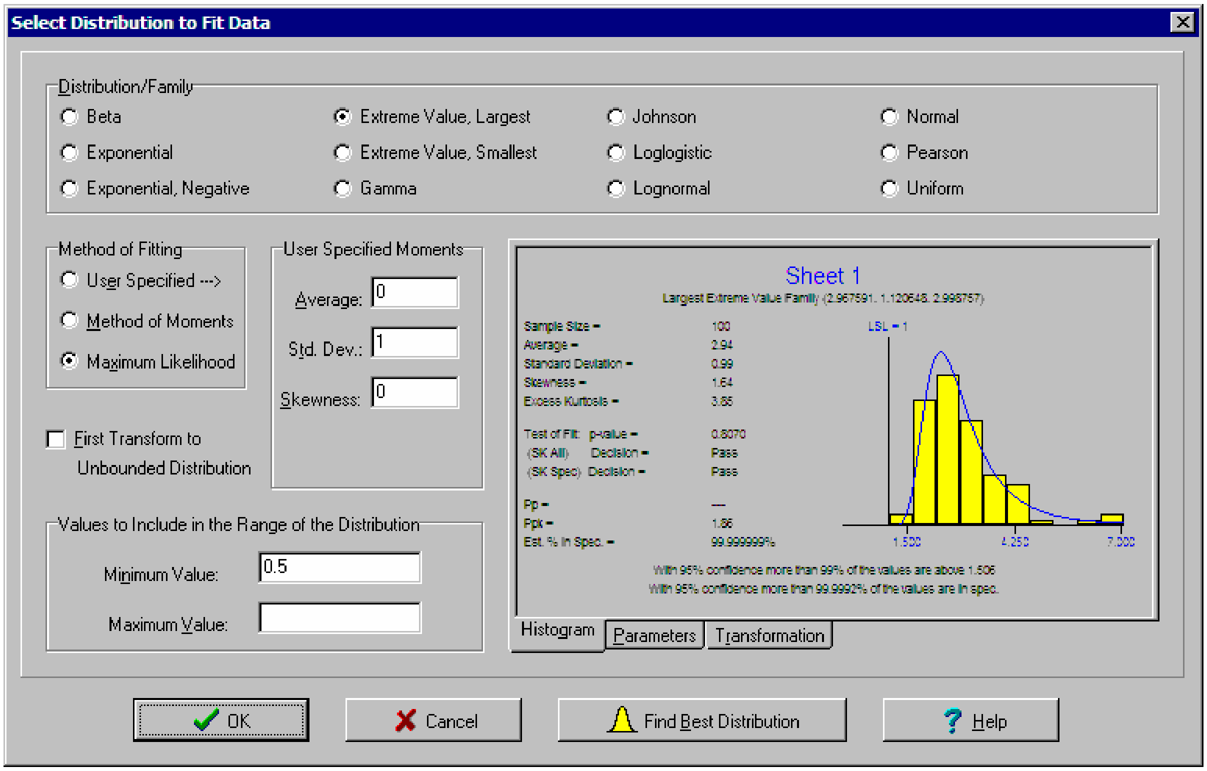
The method of moments approach does not even guarantee all the data will be within the range of the distribution. Distribution Analyzer reduces the number of moments it fits to ensure the data and spec limits are within the range of the distribution. For example, fitting the lognormal distribution generally involves matching the average, standard deviation and skewness of the distribution to that of the data. If this results in a lower bound that does not include the spec limits and data, then a lognormal is fit to the data matching the lower bound, average and standard deviation. A lower bound of one standard deviation below the lower spec limit is used.

The maximum likelihood method guarantees all the data is within the range of the distribution but does not guarantee the spec limits are. This method involves finding the values for the parameters that maximizes the log-likelihood function. This optimization is constrained to ensure the selected values include the spec limits.

If one knows the distribution most likely to fit the data up front, then it is best to start with the known distribution, rather than starting with the normal distribution. To fit a specific distribution to the data, click the *Select Distribution* button in the *Data* window. This displays the popup menu shown to the right for selecting a specific distribution. When a distribution is selected, it is fit to the data using both method of moments and maximum likelihood. The best fit is returned. The difference between selecting the *No Transformation* menu item and the *Normal* menu item is that the *No Transformation* menu item does not transform the data (Y=X) and the *Normal* menu item transforms the data using the transformation Y = (X-Average) / Standard Deviation.



The last *Custom* menu item displays the *Select Distribution to Fit Data* dialog box shown below. This dialog gives you complete control over the process including the distribution to fit, the method used to fit the data and whether to pre-transform the data or not. The user is encouraged to use the *Find Best Distribution* button in the *Data* window instead. Selecting transformations has the potential of being abused. The distribution that best fits the data is generally the one that should be used.



Select Distribution to Fit Data

Distribution/Family

○ Beta
○ Exponential
○ Exponential, Negative
● Extreme Value, Largest
○ Extreme Value, Smallest
○ Gamma
○ Johnson
○ Loglogistic
○ Lognormal
○ Normal
○ Pearson
○ Uniform

Method of Fitting

○ User Specified -->
○ Method of Moments
● Maximum Likelihood

User Specified Moments

Average: 0
Std. Dev.: 1
Skewness: 0

☐ First Transform to Unbounded Distribution

Values to Include in the Range of the Distribution

Minimum Value: 0.5
Maximum Value:

Sheet 1

Largest Extreme Value Family (2.967591, 1.120648, 2.998757)

Sample Size = 100
Average = 2.94
Standard Deviation = 0.99
Skewness = 1.64
Excess Kurtosis = 3.85

Test of Fit: p-value = 0.8070
(SK All) Decision = Pass
(SK Spec) Decision = Pass

Pp = ---
Ppk = 1.86
Est. % In Spec. = 99.999999%

With 95% confidence more than 99% of the values are above 1.506
With 95% confidence more than 99.9992% of the values are in spec.

The following options can be set:

***Distribution/Family* Radio Group:** Select the distribution or family of distributions to use from the list. The following are available: beta, exponential, negative exponential, largest extreme value, smallest extreme value, gamma, Johnson, loglogistic, lognormal, normal, Pearson and uniform.

***Method of Fitting* Radio Group:** Used to select method of fitting data to selected distribution. There are three options: (1) enter user specified moments, (2) method of moments and (3) maximum likelihood method. The user specified moments approach uses the moments entered in the *User Specified Moments* group to determine the parameters of the distribution and completely ignores the data. The other two approaches use the data to estimate the parameters of the distribution.

***User Specified Moments* Group:** Used to enter up to four moments depending on the distribution. All distributions require the first two moments (average and standard deviation). Many also require the skewness. A few also require the kurtosis. These moments are only used when the *User Specified* radio button is selected in the *Method of Fitting* radio group.

***Average* Edit Box:** First moment required for all distributions. It must be a real number.

***Standard Deviation* Edit Box:** Second moment required for all distributions. It must be a real number greater than zero.

***Skewness* Edit Box:** Third moment. It is only displayed for certain distributions. Can be any real number except for the largest extreme value (> -5.6051382), smallest extreme value (< 5.6051382) and loglogistic (> -4.284783 and < 4.284783).

***Kurtosis/Excess Kurtosis* Edit Box:** Fourth moment. It is only displayed for certain distributions. Enter either the kurtosis or excess kurtosis depending on which is selected in the *Analysis Options* dialog box. The kurtosis must be a real number greater than 1.0 and satisfying Kurtosis > Skewness * Skewness + 1. The excess kurtosis must be a real number greater than -2.0 and satisfying Excess Kurtosis > Skewness * Skewness - 2.

***First Transform to Unbounded Distribution* Check Box:** Checking this box pre-transforms the data using the Lower and/or Upper bounds of the currently selected tab in the *Data* window. If no bounds are provided, this box is grayed out.

***Minimum Value* Edit Box:** Specifying a minimum value forces the distribution to include this value in the range of the fitted distribution. It may be left blank. The value has no affect on unbounded distributions but does effect bounded distributions. This value is by default set to one standard deviation below the lower spec limit, if one exists, to ensure the lower spec limit is within the range of the distribution fit to the data. This ensures the lower spec limit can be transformed along with the data.

***Maximum Value* Edit Box:** Specifying a maximum value forces the distribution to include this value in the range of the fitted distribution. It may be left blank. The value has no affect on unbounded distributions but does affect bounded distributions. This value is by default set to one standard deviation above the upper spec limit, if one exists, to ensure the upper spec limit is within the range of the distribution fit to the data. This ensures the upper spec limit can be transformed along with the data.

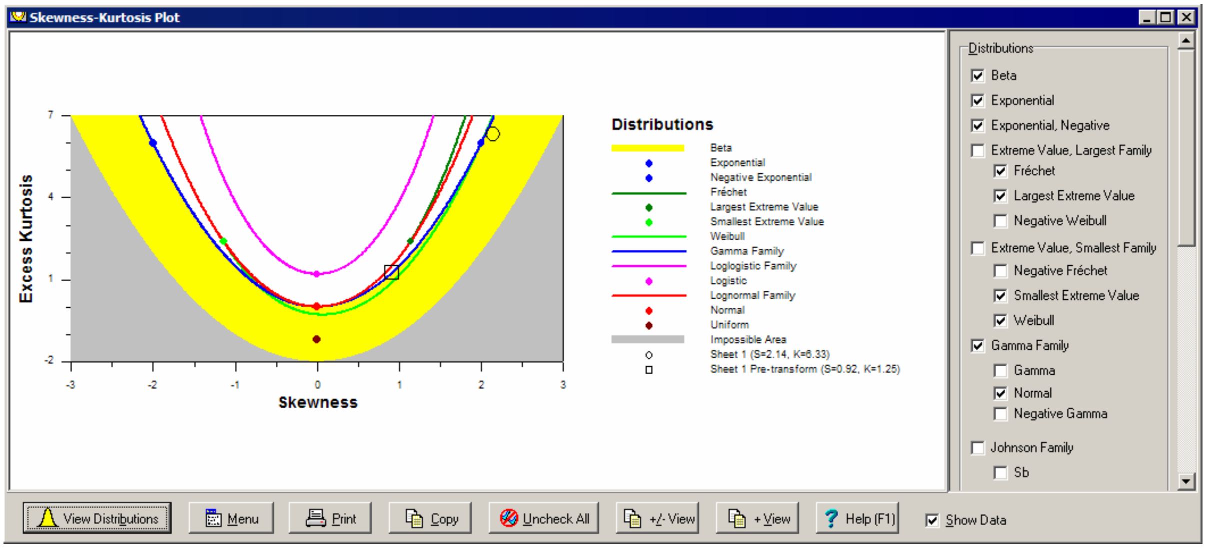
A picture of the distribution fit to the data is displayed if the selected approach can fit the data. Otherwise, an error message is displayed. The parameters of the distribution are given on the second tab. The corresponding transformation in EXCEL format is displayed on the third tab. Right clicking the mouse while the cursor is over one of these tabs displays a popup menu for printing the graphic or copying it to the clipboard.

Click the *Find Best Distribution* button to determine which of the available distributions best fits the data with or without pre-transforming the data. The distribution producing the highest p-value for the Skewness-Kurtosis All Test for normality when applied to the transformed data is selected. This button attempts to return a distribution that covers the specified range but, in certain cases, the method of moments approach may fit a distribution not covering the range.

When done, click the *OK* button. If any errors are found, an error message will be displayed and the errors must be corrected before the dialog box can be closed. To exit without changing the distribution, click the *Cancel* button or press the Esc *key*.

3.5 Skewness-Kurtosis Plot Window

The *Skewness-Kurtosis Plot* window is a child window that displays a skewness-kurtosis plot. This plot is useful for exploring the shapes and relationships of the different distributions. This window is displayed by selecting the *View Skewness-Kurtosis Plot* menu item on the *Analysis* menu, by clicking on the *View Skewness-Kurtosis Plot* button on the toolbar, or by clicking the *Skewness-Kurtosis Plot* button in the *Data* window.



Distributions
- ☑ Beta
- ☑ Exponential
- ☑ Exponential, Negative
- ☐ Extreme Value, Largest Family
  - ☑ Fréchet
  - ☑ Largest Extreme Value
  - ☐ Negative Weibull
- ☐ Extreme Value, Smallest Family
  - ☐ Negative Fréchet
  - ☑ Smallest Extreme Value
  - ☑ Weibull
- ☑ Gamma Family
  - ☐ Gamma
  - ☑ Normal
  - ☐ Negative Gamma
- ☐ Johnson Family
  - ☐ Sb

☑ Show Data

Use the *Distributions* panel at the right of the window to select which distributions and family of distribution to display. The skewness and kurtosis of any data entered in the *Data* window can also be displayed on the plot.

A skewness-kurtosis plot indicates the range of skewness and kurtosis values a distribution can fit. Two-parameter distributions like the normal distribution are represented by a single point. Three parameters distributions like the lognormal distribution are represented by a curve. Four parameter distributions like the beta distribution are represented by a shaded region. At the bottom of the plot is a gray shaded region called the impossible region. No distributions can fall into this region.

The skewness-kurtosis plot can be used in many ways:

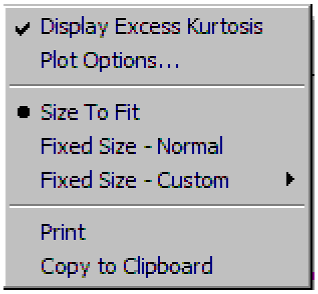
- Locate the point on the plot that corresponds to a set of data and see which distributions are nearby and might fit the data.
- See which distributions are close to each other. For example, the exponential distribution is at the point where the gamma and Weibull distributions intersect and is a special case of both distributions. Another example is that the normal distribution is on the curve of the lognormal distribution. The lognormal distribution limits to the normal distribution as the skewness goes to zero.

- See the relationships between distributions. For example, the lognormal distribution's curve is above the gamma distribution's curve. This means that for the same skewness, the lognormal distribution has a higher kurtosis (heavier tails) than the gamma distribution.

There are a number of buttons at the bottom on the window to alter, print and copy the skewness-kurtosis plot:

***View Distributions* Button:** Displays the *Select/View Distribution* dialog box to find further information about the distributions including viewing the density function, calculating probabilities and determining bounds. This dialog box is described more fully later in the section.

***Menu* Button:** Displays popup menu below for customizing the plot. The first menu item selects whether to display the kurtosis or excess kurtosis on the left axis. The *Plot Options* menu item displays the *Plot Options* dialog box for altering the style, labels and axis of the skewness-kurtosis plot. These items can also be altered by double clicking on the plot itself. The *Plot Options* dialog box is described more fully later in the section.



***Print* Button:** Prints the skewness-kurtosis plot. Same as clicking the *Print* button on the toolbar or selecting the *Print* menu item on the *File* menu.

***Copy* Button:** Copies the skewness-kurtosis plot to the clipboard.

***Uncheck All* Button:** Removes all checks from distributions in the *Distributions* panel in order to start over selecting distributions.

+/- *View* Button: Displays both the positive and negative skewness regions of the plot by setting the scale for the X-axis from -3 to 3 in the *Plot Options* dialog box.

+ *View* Button: Displays only the positive skewness region of the plot by setting the scale for the X-axis from 0 to 3 in the *Plot Options* dialog box.

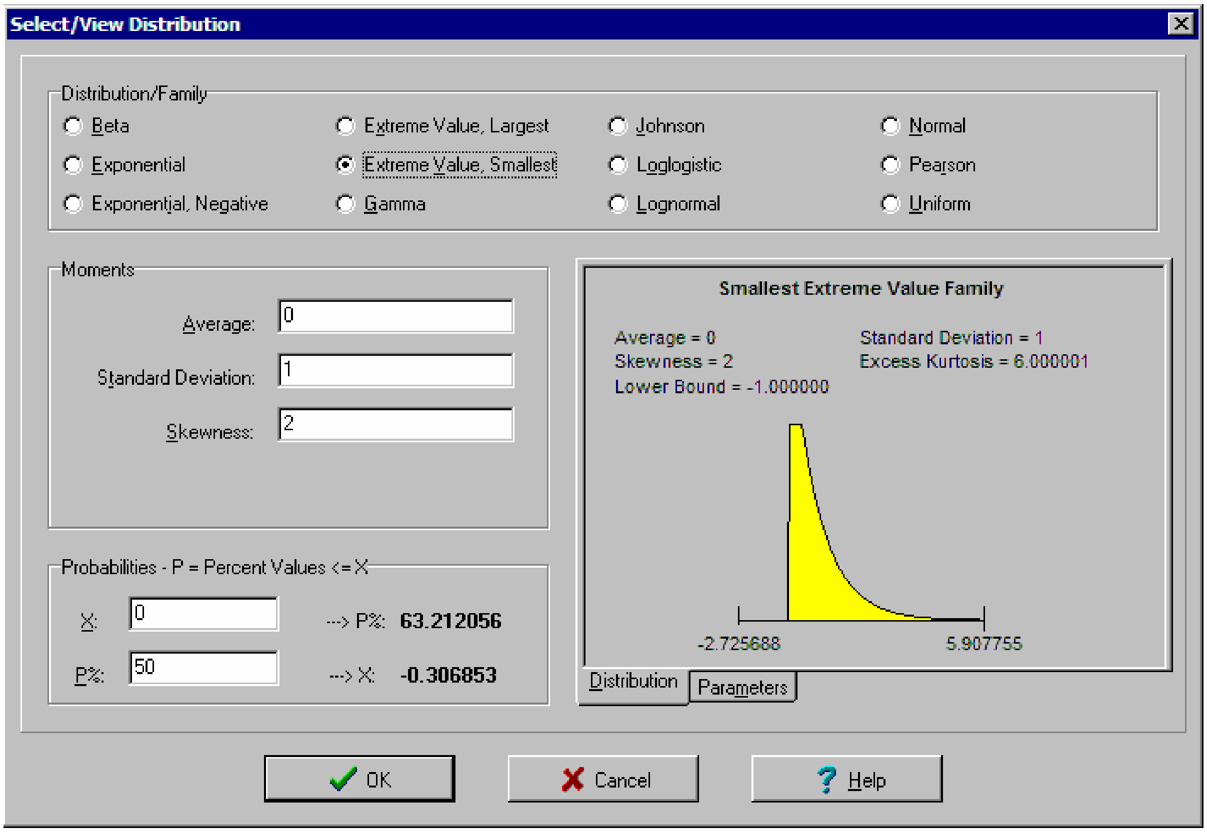
***Help* Button:** Displays help on using this window.

***Show Data* Check Box:** When checked, points are displayed on the skewness-kurtosis plot representing data found on the tabs in the *Data* window. If lower or upper physical bounds are also specified, then points are also displayed representing the pre-transformed data.

To close the *Skewness-Kurtosis Plot* window, click the X in the upper right corner.

Select/View Distribution Dialog Box

The *Select/View Distribution* dialog box is used to specify a distribution and its parameters. It is used by the *Skewness-Kurtosis Plot* window to view different distributions to better understand their shapes. It is also used by the *Generate Random Values* dialog box to select a distribution to use when generating random numbers. This dialog box is displayed by clicking the *View Distributions* button on the *Skewness-Kurtosis Plot* window and by clicking the *Select Distribution* button on the *Generate Random Values* dialog box.



Select/View Distribution

Distribution/Family
○ Beta
○ Exponential
○ Exponential, Negative
○ Extreme Value, Largest
● Extreme Value, Smallest
○ Gamma
○ Johnson
○ Loglogistic
○ Lognormal
○ Normal
○ Pearson
○ Uniform

Moments
Average: 0
Standard Deviation: 1
Skewness: 2

Probabilities - P = Percent Values <= X
X: 0 --> P%: 63.212056
P%: 50 --> X: -0.306853

Smallest Extreme Value Family
Average = 0
Standard Deviation = 1
Skewness = 2
Excess Kurtosis = 6.000001
Lower Bound = -1.000000

The following options can be set:

***Distribution/Family* Radio Group:** Select the distribution or family of distributions to use from the list. The following are available: beta, exponential, negative exponential, largest extreme value, smallest extreme value, gamma, Johnson, loglogistic, lognormal, normal, Pearson and uniform.

***Moments* Group:** Used to enter up to four moments depending on the distribution. All distributions require the first two moments (average and standard deviation). Many also require the skewness. A few require both the skewness and kurtosis.

***Average* Edit Box:** First moment required for all distributions. It must be a real number.

Standard Deviation **Edit Box:** Second moment required for all distributions. It must be a real number greater than zero.

Skewness **Edit Box:** Third moment. It is only displayed for certain distributions. Can be any real number except for the largest extreme value (> -5.6051382), smallest extreme value (< 5.6051382) and loglogistic (> -4.284783 and < 4.284783).

Kurtosis/Excess Kurtosis **Edit Box:** Fourth moment. It is only displayed for certain distributions. Enter either the kurtosis or excess kurtosis depending on which is selected in the *Analysis Options* dialog box. The kurtosis must be a real number greater than 1.0 and satisfying Kurtosis > Skewness * Skewness + 1. The excess kurtosis must be a real number greater than -2.0 and satisfying Excess Kurtosis > Skewness * Skewness - 2. Otherwise, the specified value falls into the impossible region.

X **Edit Box:** Used to explore the distribution. Enter a value for X and the percent of values less than X will be displayed (distribution function). For the distribution shown above, 63.2% of values fall below 0.

P **Edit Box:** Used to explore the distribution. Enter a percentage for P and the value of X that has P% of values below it will be displayed (inverse of distribution function). For the above example, 50% of the values fall below -0.31.

A picture of the selected distribution is displayed if the moments are valid. Otherwise, an error message is displayed. The parameters of the distribution are also given on a second tab. Right clicking the mouse while the cursor is over one of these tabs displays a popup menu used for printing the graphic or copying it to the clipboard.

When done, click the *OK* button. If any errors are found, an error message will be displayed and the errors must be corrected before the dialog box can be closed. To exit without changing the distribution, click the *Cancel* button or press the *Esc* key.

Plot Options Dialog Box

Selecting the *Plot Options* menu item from the popup menu associated with the *Menu* button displays the *Plot Options* dialog box shown on the next page. There are seven tabs displayed for changing different features of the plot. The tabs are summarized in the table below. Most of these tabs can be displayed automatically by double clicking on the appropriate part of the plot. For example, to display the *X-Axis* tab, double click on the X-Axis.

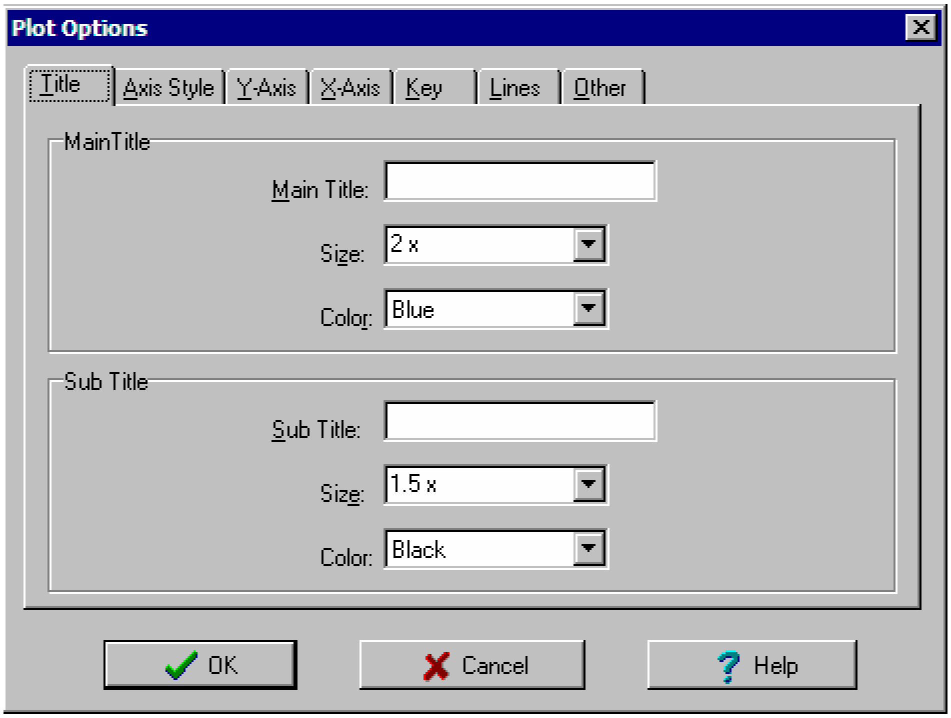
Title	Changes title text and style.
Axis Style	Changes the axis style, tic mark style and the plot's aspect ratio.
Y-Axis	Changes y-axis label, scale, style and orientation.
X-Axis	Changes x-axis label, scale and style.
Key	Used to specify whether to display the key and for setting style options.

Lines	Changes the labels, values and styles associated with the lines.
Other	Changes the resolution of the plot and whether to automatically update axis scales.

When you are done, click the *OK* button to close the dialog box and update the plot. An alternative is to press the *Enter* key. Clicking the *Cancel* button instead closes the dialog box without saving any changes. An alternative is to press the *Esc* key. The *Help* button provides help concerning the *Plot Options* dialog box.

Plot Options Dialog Box – *Title* Tab

The *Title* tab is used to change the text, color and size of the title and subtitle at the top of the plot. Initially there are no titles.



Main Title:

Size: 2 x

Color: Blue

Sub Title:

Size: 1.5 x

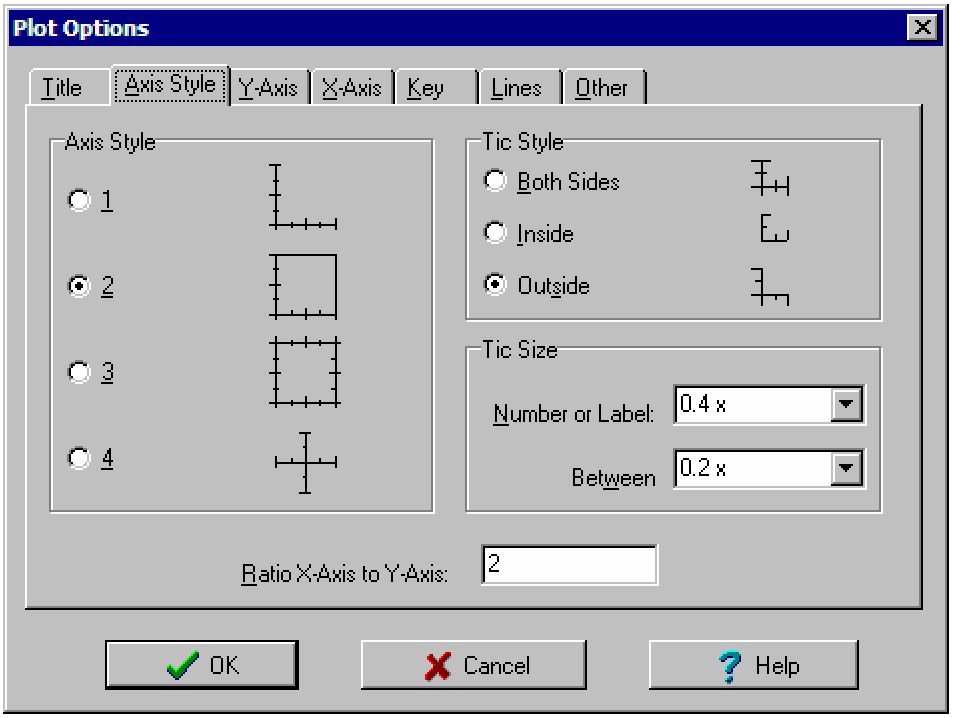
Color: Black

The ***Main Title* edit box** is used to change the text for the main title. It may be left blank if no title is desired. The ***Main Title Size* combo box** is used to select the size for displaying the main title. Normal size is the same size as the system font. The default is twice the normal size. You can select one of the values in the list or type your own value. Entering a value of 1.3 results in the title being 1.3 times normal size. The ***Main Title Color* list box** is used to change the color of the main title.

The ***Sub Title* edit box**, ***Sub Title Size* combo box** and **Sub Title Color list box** are used to enter an optional subtitle. By default it is blank. They work the same as the main title controls.

Plot Options Dialog Box – *Axis Style* Tab

The *Axis Style* tab is used to change the axis style, tic mark style, tic mark size and the plot's aspect ratio.



Plot Options

Axis Style

○ 1

● 2

○ 3

○ 4

Tic Style

○ Both Sides

○ Inside

● Outside

Tic Size

Number or Label: 0.4 x

Between: 0.2 x

Ratio X-Axis to Y-Axis: 2

The ***Axis Style* radio buttons** are used to select one of four different axis styles. The ***Tic Style* radio buttons** are used to select from three different tic styles. Tic marks can appear next to the values on an axis as well as in between values. The ***Tic Size Number or Label* combo box** is used to select the size for displaying tics marks appearing next to values. The ***Tic Size Between* combo box** is used to select the size for displaying tics marks appearing between values. The default is to make the tic marks appearing next to values 0.4 times normal size and the tic marks appearing between values smaller at 0.2 times normal size.

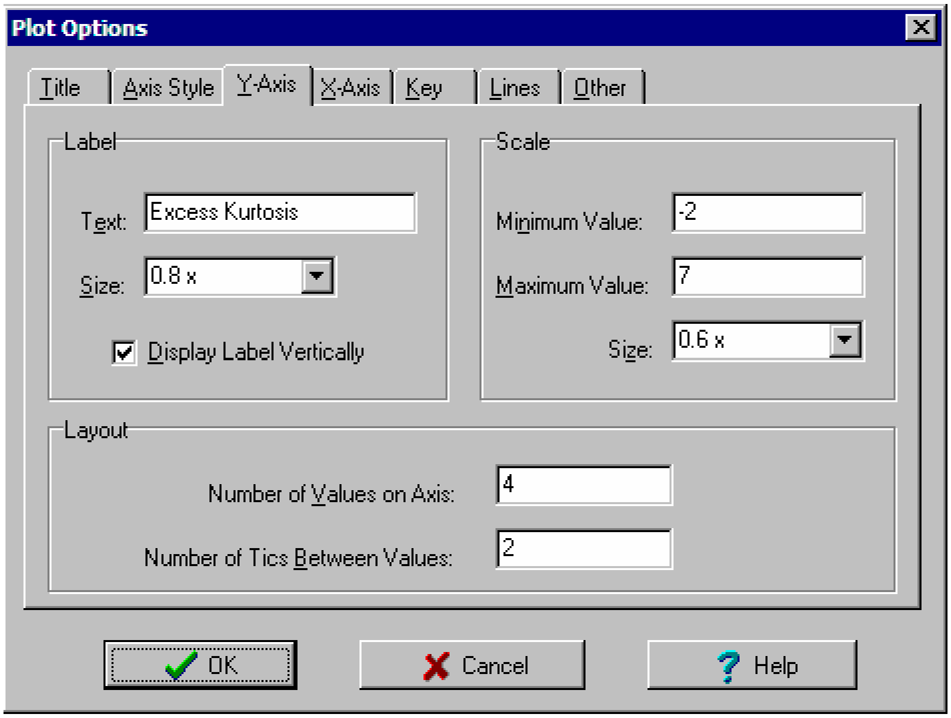
The ***Ratio X-Axis to Y-Axis* edit box** is used to change the plot's aspect ratio. A value of 1 results in a square plotting region. A value of 2 makes the x-axis twice as long as the y-axis. A value of 0.5 makes the x-axis half as long as the y-axis. The y-axis size remains fixed.

Plot Options Dialog Box – *Y-Axis* Tab

The *Y-Axis* tab is used to change the y-axis label, scale, tic marks and orientation. Double clicking on the y-axis displays this tab.

The ***Text* edit box** is used to change the axis label. By default this is the phrase “Kurtosis” or “Excess Kurtosis”. The ***Label Size* combo box** is used to select the size for displaying the axis label. The default is 1.5 times normal size. The ***Display Label Vertically* check box** allows the label to be displayed either vertically or horizontally.

The ***Minimum Value*** and ***Maximum Value* edit boxes** are used to change the scale of the y-axis. The minimum value must be less than the maximum value. The ***Value Size* combo box** is used to select the size for displaying axis values. The default is normal size.

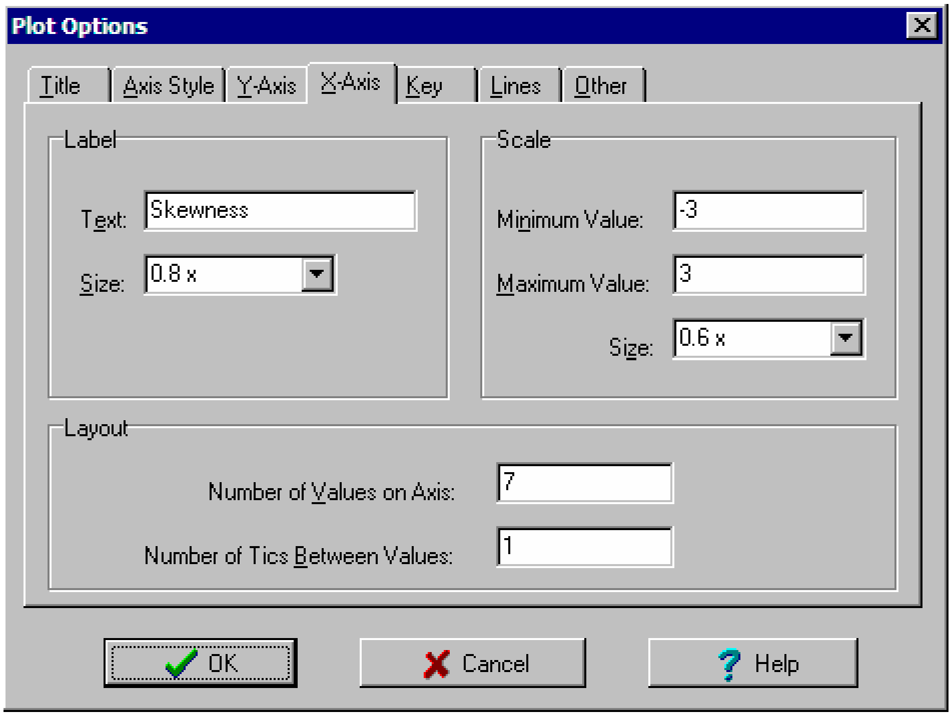


Text: Excess Kurtosis
Size: 0.8 x
☑ Display Label Vertically
Minimum Value: -2
Maximum Value: 7
Size: 0.6 x
Number of Values on Axis: 4
Number of Tics Between Values: 2

The ***Number of Values on Axis* edit box** is used to change the number of values displayed along the y-axis. The minimum is two. The default is 3. The ***Number of Tics Between Values* edit box** is used to change the number of tic marks displayed between each value on the y-axis. A value of zero can be entered. The default is 4.

Plot Options Dialog Box – *X-Axis* Tab

The *X-Axis* tab is used to change the x-axis label, scale and tic marks. Double clicking on the x-axis displays this tab. It functions identical to the *Y-Axis* tab except it affects the bottom or x-axis.



Label
Text: Skewness
Size: 0.8 x

Scale
Minimum Value: -3
Maximum Value: 3
Size: 0.6 x

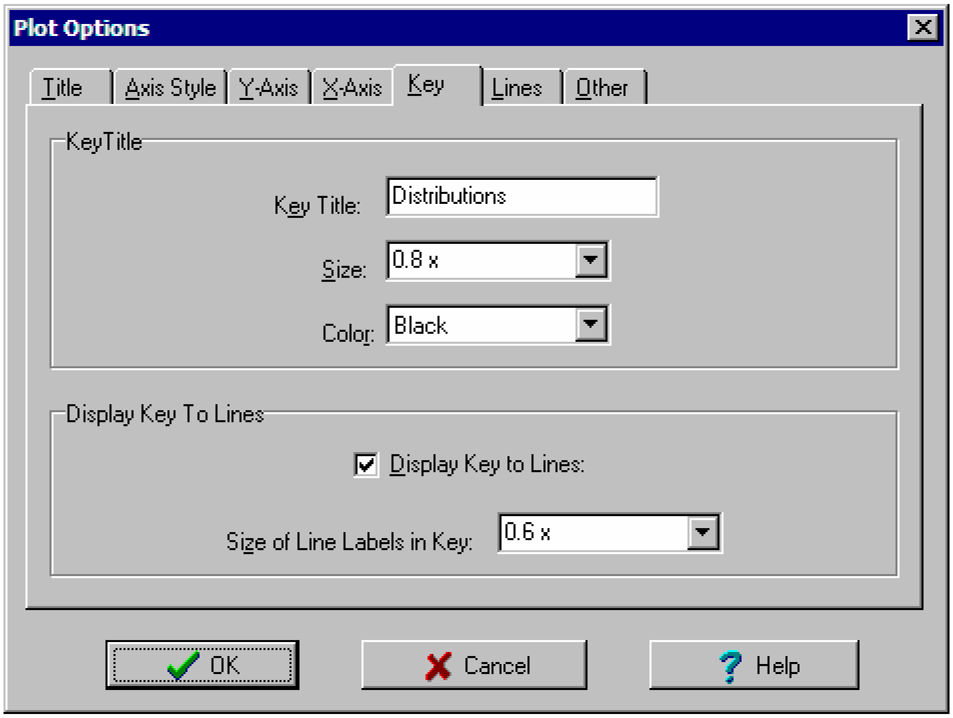
Layout
Number of Values on Axis: 7
Number of Tics Between Values: 1

Plot Options Dialog Box – *Key* Tab

The *Key* tab is used to change the key title and style. On an interaction plot, the key describes how the lines correspond to the different distributions. Double clicking on the key title displays this tab.

The ***Key Title* edit box** is used to change the text for the key title. It may be left blank if no title is desired. The ***Key Title Size* combo box** is used to select the size for displaying the key title. Normal size is the same size as the system font and is the default size. You can select one of the values in the list or type your own value. Entering a value of 1.3 results in the title being 1.3 times normal size. The ***Key Title Color* list box** is used to change the color of the key title.

The ***Display Key To Line* check box** is used to specify whether to display the key to the lines. Generally the key is desired but can be removed for purposes of printing and copying to the clipboard. The ***Size of Line Labels in Key* combo box** is used to select the size for displaying the line labels in the key. The default is normal size.



KeyTitle

Key Title: Distributions

Size: 0.8 x

Color: Black

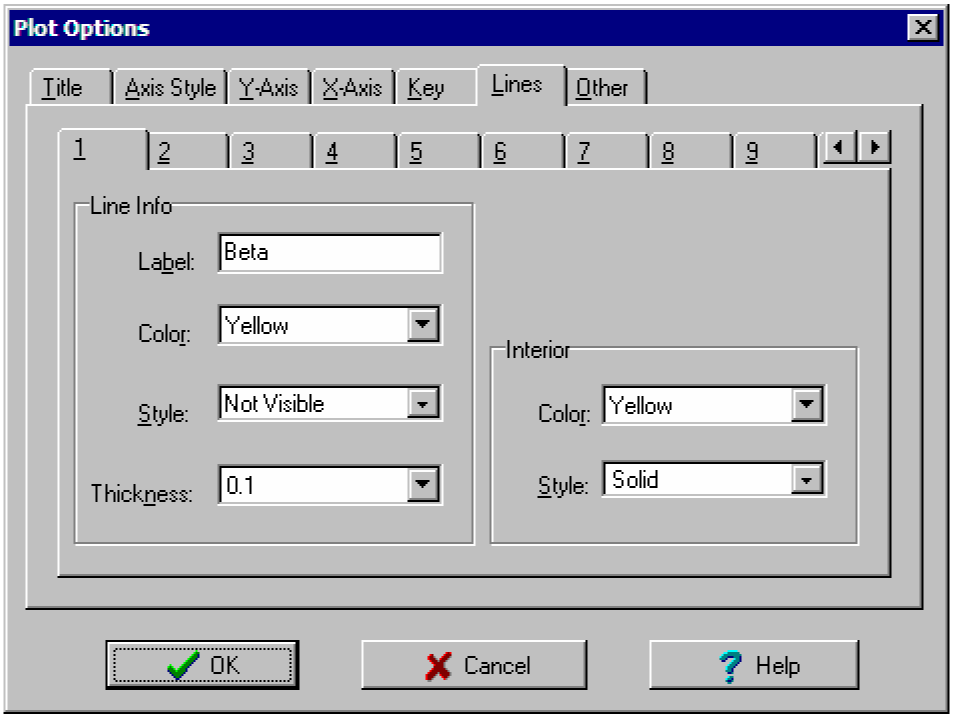
Display Key To Lines

☑ Display Key to Lines:

Size of Line Labels in Key: 0.6 x

Plot Options Dialog Box – *Lines* Tab

The *Lines* tab is used to change the label and style of a point, line or region representing a distribution. A short cut is to double click on the key for the lines. Double clicking on a line in either the plot or key displays this tab.



Line Info

Label: Beta

Color: Yellow

Style: Not Visible

Thickness: 0.1

Interior

Color: Yellow

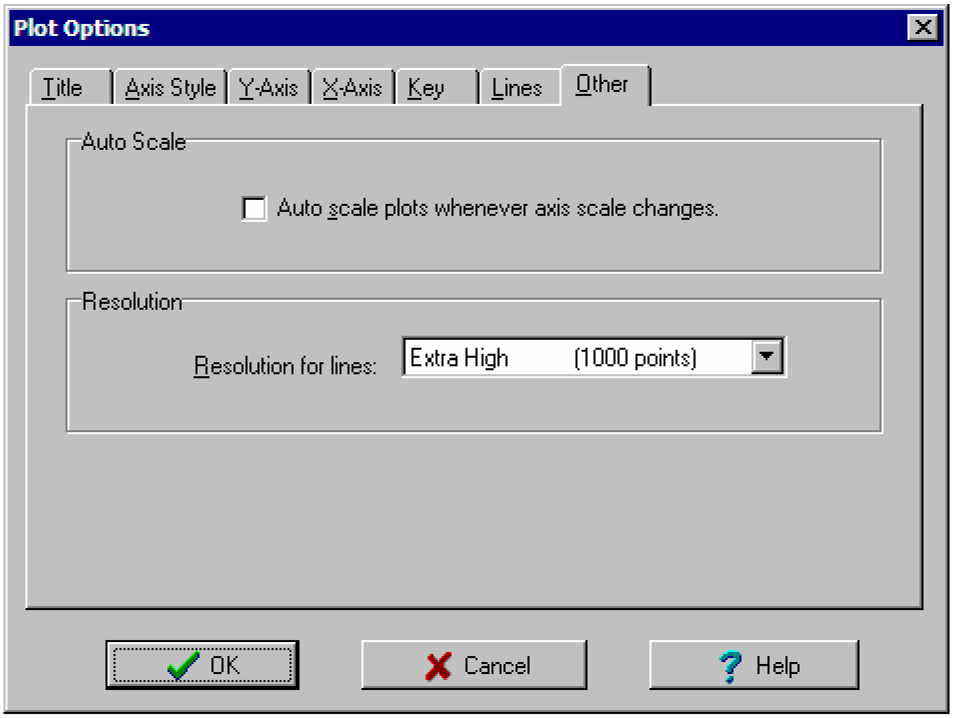
Style: Solid

The ***Line* tabs** are used to select which line to change. The values displayed in the other controls are for the line whose tab is selected.

The ***Label* edit box** is used to change the label shown next to the line in the key for lines. The ***Color* list box** is used to change the color of the line. The ***Style* list box** is used to select the line style. If a line style other than a solid line is selected, the line thickness is set to hairline. The ***Thickness* combo box** is used to change the line thickness of solid lines. Only solid lines can have a thickness other than hairline. If a thickness other than hairline is selected, the line style will be set to solid. The ***Interior Color* list box** is used to change the color of the region. The ***Interior Style* list box** is used to select the region style.

Plot Options Dialog Box – *Other* Tab

The *Other* tab is used to change whether automatic scaling is performed and the plot's resolution.



☐ Auto scale plots whenever axis scale changes.

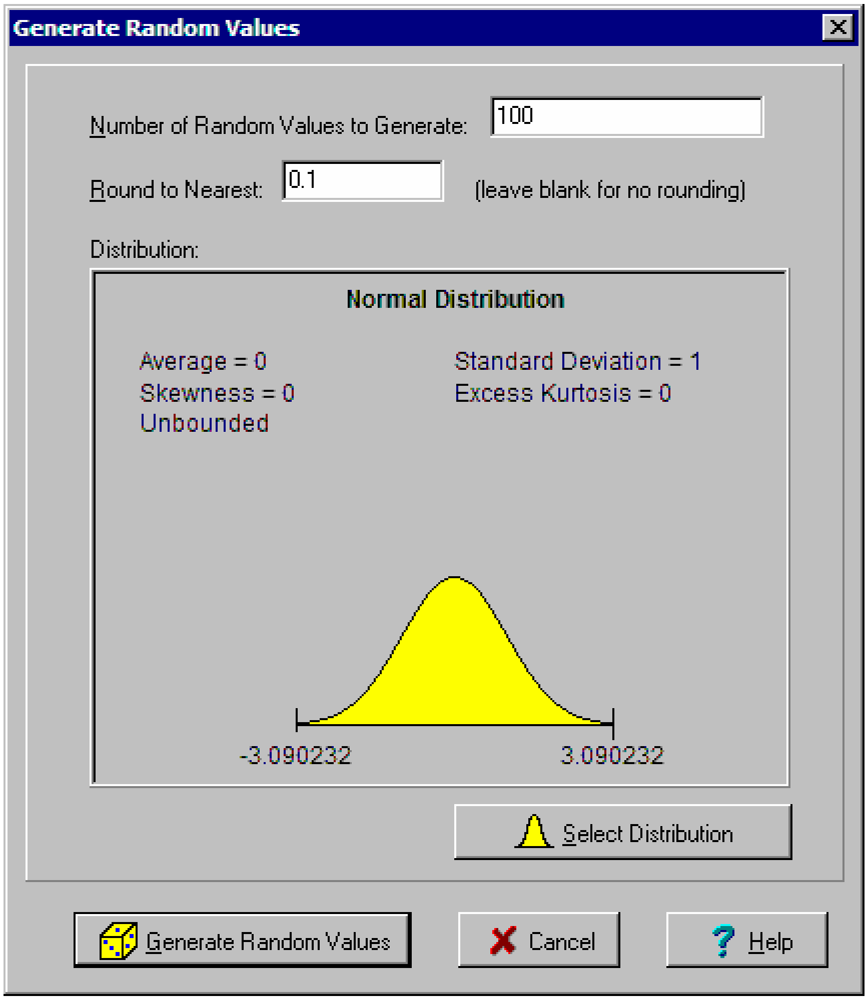
Resolution for lines: Extra High (1000 points)

The ***Automatic Scaling* check box** is used to select whether automatic scaling is active. Changing the x-axis scale can cause the curves to shift so that they are no longer visible on the plot. Automatic scaling automatically changes the plot's scales whenever such a change occurs so that the curves remain visible. Uncheck this box if you want to override the default scaling. It is recommended auto-scaling not be used.

The ***Resolution* list box** is used to select the plot's resolution. The resolution is the number of points used to construct the curve. The default is low resolution which speeds up the drawing of the plots. Higher resolutions can be used to produce smoother plots.

3.6 Generating Random Values

Random values from any of the distributions can be generated using the *Generate Random Values* dialog box. This dialog box is displayed by selecting the *Generate Random Values* menu item on the *Analysis* menu or by clicking on the *Generate Random Values* button on the toolbar.



Generate Random Values

Number of Random Values to Generate: 100

Round to Nearest: 0.1 (leave blank for no rounding)

Distribution:

Normal Distribution

Average = 0
Standard Deviation = 1
Skewness = 0
Excess Kurtosis = 0
Unbounded

This dialog box is used to create a set of data by generating random values that are saved in the currently selected tab in the *Data* window. You can select the number of values, the precision and the distribution.

***Number of Random Values to Generate* Edit Box:** Enter the number of values to generate. It must be an integer from 1 to 9,999. At least 15 are required to perform an analysis using the generated data (may be changed in the *Advanced Options* dialog box).

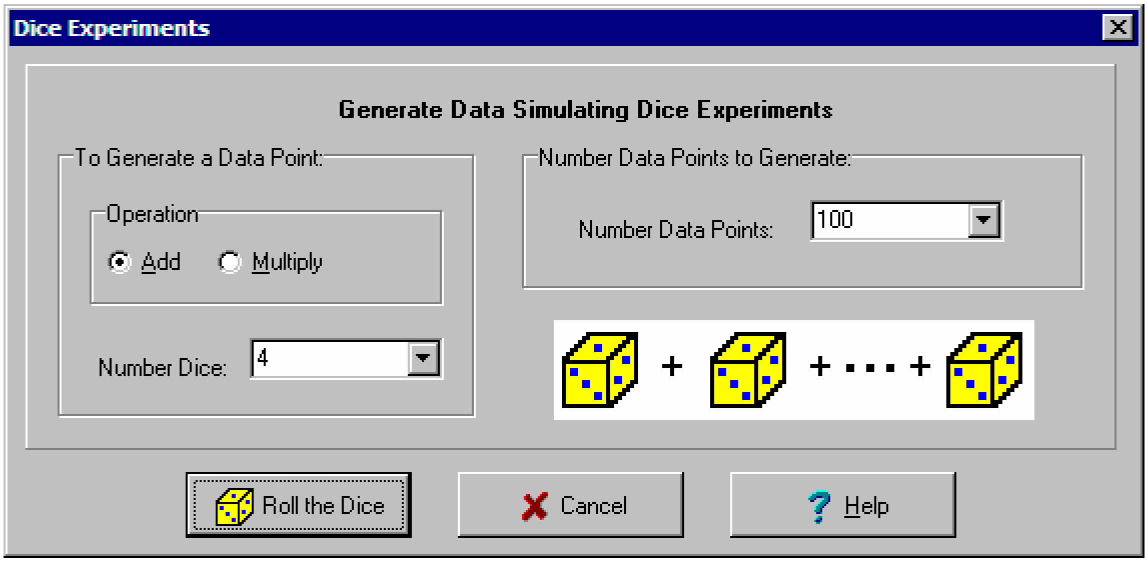
***Precision* Edit Box:** Specifies the precision values are rounded to. For example, if a precision of 0.2 is specified, values will be a multiple of 0.2. This will result in the values like -0.2, 0.0, 0.2, 0.4, and 0.6. It must be a positive number or blank.

***Select Distribution* Button:** Click this button to select a different distribution or to change the parameters of the distribution. It displays the *Select/View Distributio*n dialog box described in the previous section.

The currently selected distribution is displayed in the graphic box. Right clicking the mouse while the cursor is over the graphic displays a popup menu for printing the graphic or copying it to the clipboard.

Click the *Generate Random Values* button to generate the data and save it in the currently selected tab of the *Data* window. If the tab already contains data, you are prompted to make sure it is OK to overwrite the data. To exit without generating any data, click the *Cancel* button or press the *Esc* key.

A second way to generate random values is to perform dice experiments. The *Dice Experiments* dialog box is displayed by selecting the *Dice Experiments* menu item on the *Analysis* menu.



Dice Experiments

Generate Data Simulating Dice Experiments

To Generate a Data Point:

Operation: ● Add ○ Multiply

Number Dice: 4

Number Data Points to Generate:

Number Data Points: 100

This dialog box is used to create a set of data points mimicking the rolling of dice that are saved in the currently selected tab in the *Data* window. This is intended to aid in demonstrating how adding dice tends to the normal distribution and multiplying dice tends to the lognormal distribution.

Operation* Radio Group Box*: Select whether to add or multiply dice.

***Number Dice* Edit Box:** Specifies the number of dice to roll and either add or multiply together in order to create a single data point. It must be an integer from 1 to 20.

***Number Data Points* Edit Combo Box:** Enter the number of data points to generate. It must be an integer from 1 to 10000.

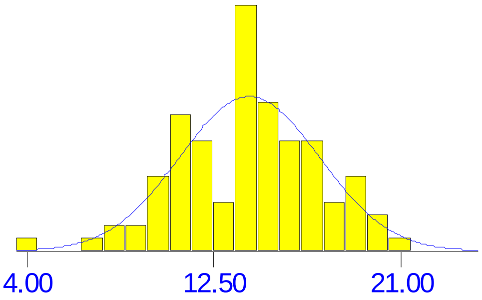
Click the *Roll the Dice* button to generate the data and save it in the currently selected tab of the *Data* window. If the tab already contains data, you are prompted to make sure it is OK to overwrite the data. To exit without generating any data, click the *Cancel* button or press the *Esc* key.

Dice experiments can be used to demonstrate the central limit theorem. The central limit theorem states that as items are added and subtracted together, under certain restrictions, the result will tend to the normal distribution. The normal distribution is the distribution of addition and subtraction. To see the central limit theory in practice, go to the *Dice Experiments* dialog box and specify 4 dice be added together as shown above. The normal distribution fits the resulting data as shown below.

Add 4 Dice

No Transformation (Normal Distribution)

Sample Size =	100
Average =	14.1
Standard Deviation =	3.2
Skewness =	-0.24
Excess Kurtosis =	0.14
Test of Fit: p-value =	0.5200
(SK All) Decision =	Pass
(SK Spec) Decision =	No specs



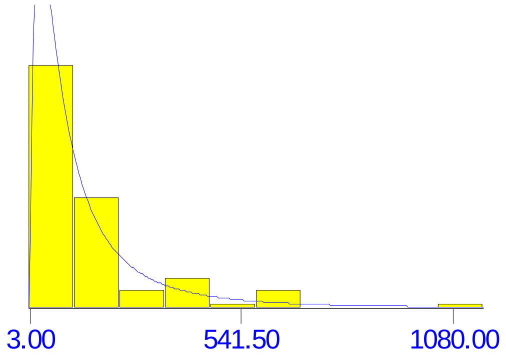
With 95% confidence more than 99% of the values are between 4.75 and 23.45

Similarly, the lognormal distribution is the distribution of multiplication and division. The central limit theorem states that as positive items are multiplied and divided, under certain restrictions, the result will tend to the lognormal distribution. To see the central limit theory in practice, go to the Dice Experiments dialog box and specify 4 dice be multiplied together. The lognormal distribution fits the resulting data as shown below.

Multiply 4 Dice

Lognormal Family (151.719, 227.061, 7.596354)

Sample Size =	100
Average =	146.3
Standard Deviation =	174.7
Skewness =	2.63
Excess Kurtosis =	8.85
Test of Fit: p-value =	0.8405
(SK All) Decision =	Pass
(SK Spec) Decision =	No specs



With 95% confidence more than 99% of the values are between 1.03 and 2053.11

4

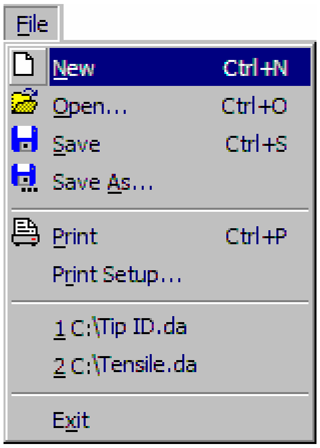
Menus and Toolbar

This chapter covers the menus and toolbar. Many of the menu items have already been covered elsewhere. However, a few have not yet been described. This chapter explains those remaining menu items.

The menu bar contains the menus listed below.

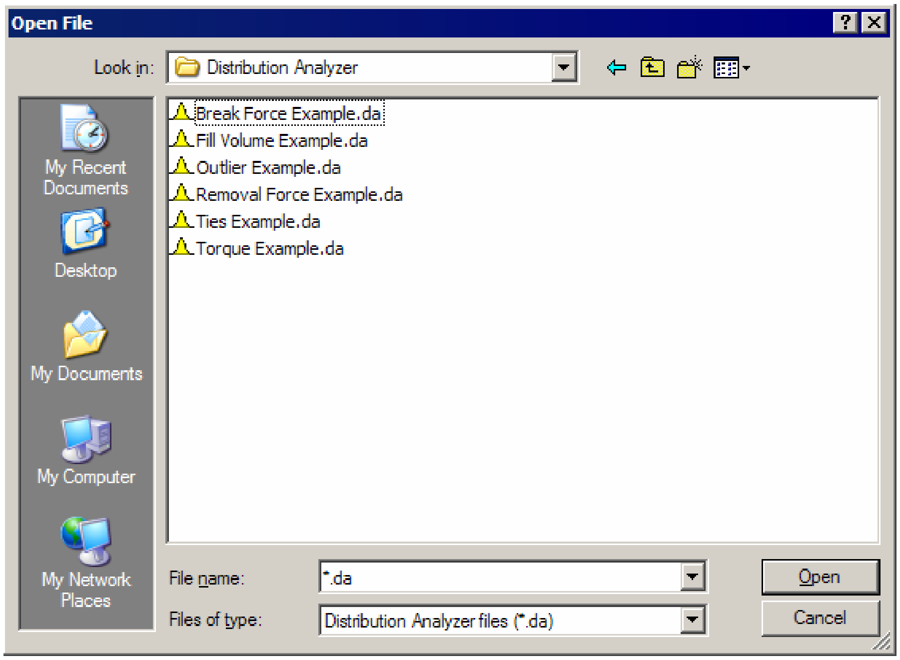
File **menu:**	Used to start a new session, save a session, open a previously saved session, setup the printer, print and exit the program.
Edit **menu:**	Used to help edit the spreadsheet in the *Data* window. It includes menu items for copying, cutting, pasting and clearing cells and for adding, deleting, moving and renaming sheets.
Analysis **menu:**	Used to perform different analyses including testing whether the selected distribution fits the data, selecting specific distributions and selecting the best distribution. Also contains items to learn more about the different distributions including generating example sets of data and skewness-kurtosis plots.
Window **menu:**	Used to rearrange and select child windows displayed in interior of main window.
Help **menu:**	Used to obtain help, register the software, obtain technical support and to link to our web site.

4.1 File Menu



The *File* menu is the first menu on the menu bar. Used to start a new session, save a session, open a previously saved session, setup the printer, print and exit the program. The first group of five buttons on the toolbar contains shortcuts for those menu items with icons. The available menu items are:

New	Begins a new session. Shortcuts are Ctrl+N and the button on the toolbar. First asks whether to save the current session.
Open...	Restores a previously saved session. An *Open File* dialog box will be opened for selecting the file to open. Distribution Analyzer files have the extension "da". Shortcuts are Ctrl+O and the button on the toolbar. You will first be asked whether to save the current session.

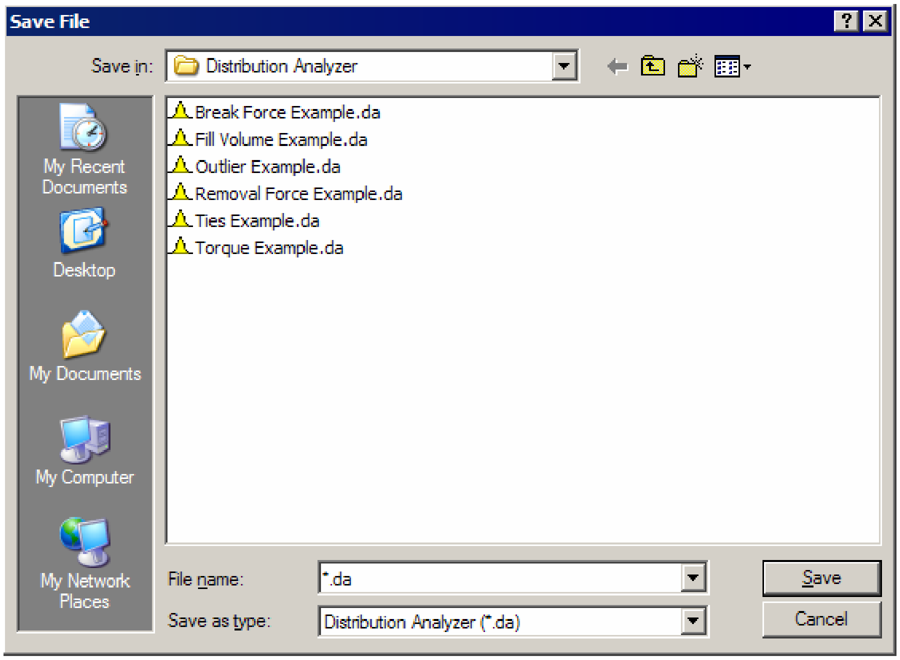


Look in: Distribution Analyzer

Break Force Example.da
Fill Volume Example.da
Outlier Example.da
Removal Force Example.da
Ties Example.da
Torque Example.da

File name: *.da
Files of type: Distribution Analyzer files (*.da)

Save	Saves changes to a session back into the file which was originally opened to start the session. If no file is associated with the session, the *Save As* menu item is automatically invoked instead. The file associated with a session is displayed on the caption bar of the main window. Shortcuts are Ctrl+S and the button on the toolbar.
Save As...	Saves the session in a new file. A *Save File* dialog box will be opened for selecting the file to save the session in. An extension of "da" will automatically be added to indicate it is a Distribution Analyzer file. Shortcut is the button on the toolbar.



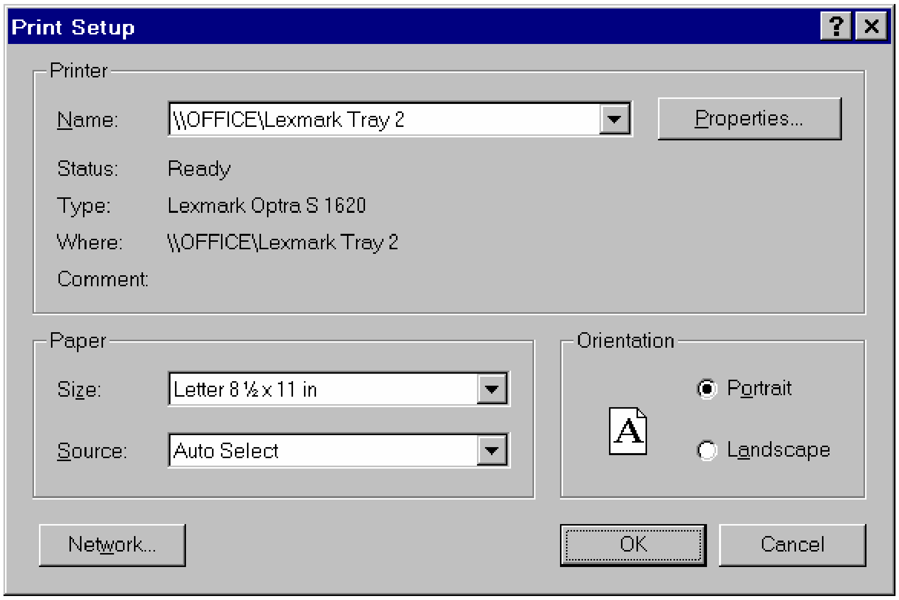
Save in: Distribution Analyzer

Break Force Example.da
Fill Volume Example.da
Outlier Example.da
Removal Force Example.da
Ties Example.da
Torque Example.da

File name: *.da
Save as type: Distribution Analyzer (*.da)

Print Prints the contents of the active child window. Shortcuts are Ctrl+P and the button on the toolbar.

Print Setup… Displays the *Print Setup* dialog box for selecting the printer to use and for changing printer options such as portrait versus landscape, paper size and paper source.



Name: \\OFFICE\Lexmark Tray 2
Status: Ready
Type: Lexmark Optra S 1620
Where: \\OFFICE\Lexmark Tray 2
Comment:

Size: Letter 8 ½ x 11 in
Source: Auto Select

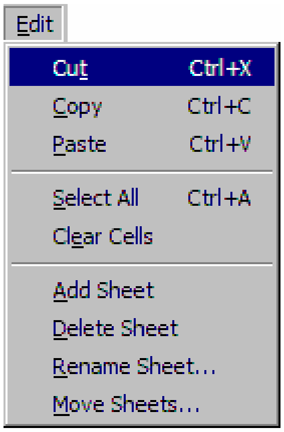
● Portrait
○ Landscape

The *Name* list box is used to select which printer to print to. Clicking on the *Properties* button displays a dialog box for modifying additional printer options.

Exit	Exits the program. First asks whether to save the current session. Same as double clicking the system menu button (icon in top-left corner) or clicking the close button (x in top-right corner). A shortcut is to press Alt-F4.

In addition, the most recently saved sessions are displayed between the Print Setup and Exit menu items. Selecting one of them will reopen that session.

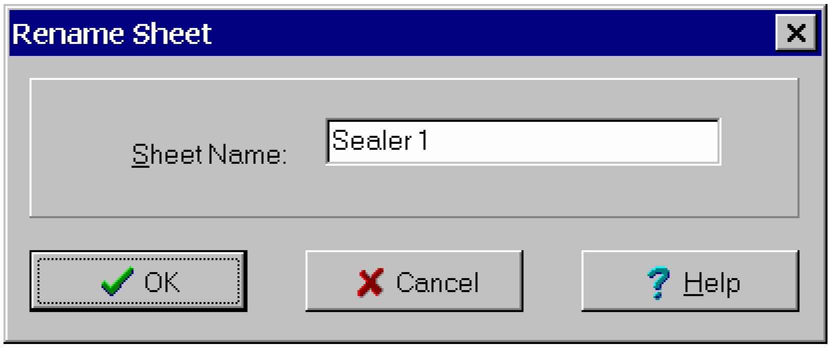
4.2 Edit Menu



The *Edit* menu is the second menu on the menu bar. It is used to help edit the spreadsheet in the *Data* window. It includes menu items for copying, cutting, pasting and clearing cells and for adding, deleting, moving and renaming sheets. The available menu items are:

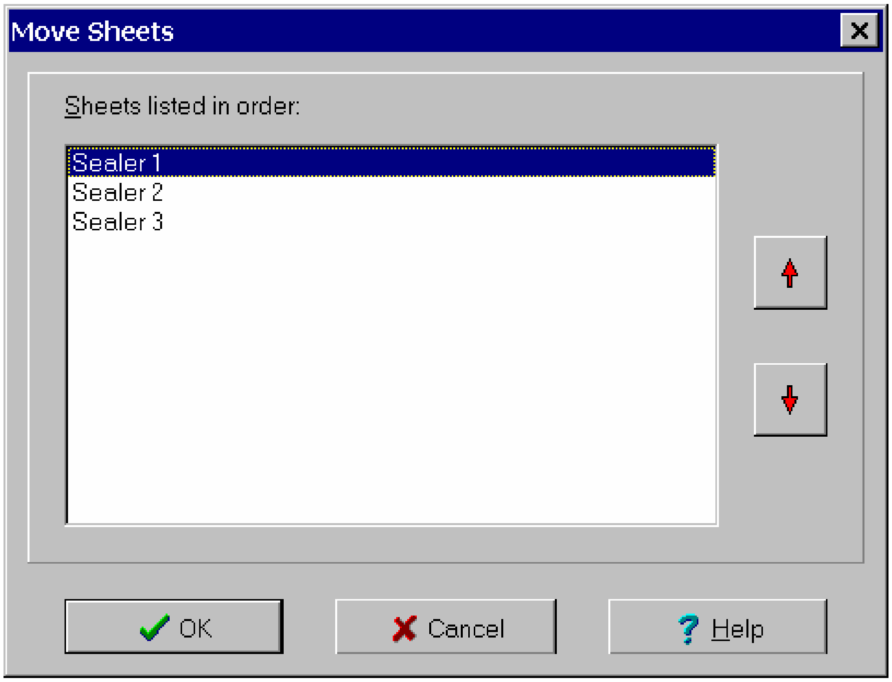
Cut	Cuts currently selected cells from the spreadsheet. They are placed in the clipboard so that they can be pasted into this or some other program. Selected cells are cleared (made blank). Shortcut is Ctrl+X.
Copy	Copies currently selected cells of spreadsheet. They are placed in the clipboard so that they can be pasted into this or some other program. Selected cells are left unchanged. Shortcut is Ctrl+C.
Paste	Pastes cells from the clipboard into the spreadsheet. All cells in the clipboard will be pasted with the selected cell being the top-left pasted cell. Shortcut is Ctrl+V.
Select All	All cells containing data are selected. This is typically used in conjunction with the *Copy* menu item to copy the entire data sheet into the clipboard. Shortcut is Ctrl+A.
Clear Cells	Clears (makes blank) the currently selected cells of the spreadsheet. Prompts first to confirm.

Add Sheet	Inserts a new data sheet behind the current sheets.
Delete Sheet	Deletes the current data sheet. Prompts first to confirm.
Rename Sheet	Displays the *Rename Sheet* dialog box for entering a new name for the current sheet. Shortcut is to double click on the current name on the tab.



Sheet Name: Sealer 1

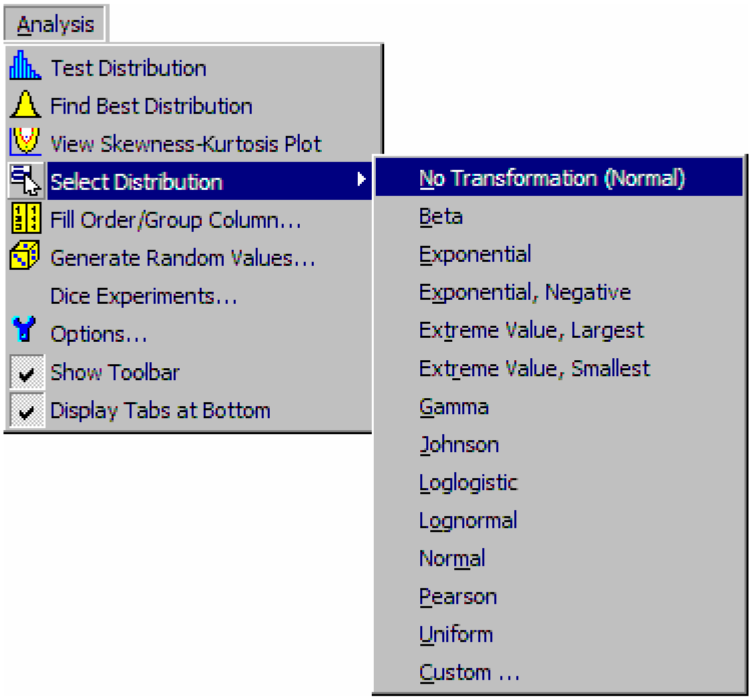
Move Sheets	Displays the *Move Sheets* dialog box for reordering the sheets. Drag and drop the sheets into a new order or select a sheet and use the up and down buttons to move the sheet in the list.



Sheets listed in order:

Sealer 1
Sealer 2
Sealer 3

4.3 Analysis Menu



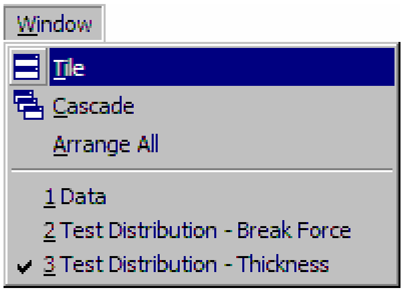
The *Analysis* menu is the third menu on the menu bar. It is used to perform different analyses including testing whether the selected distribution fits the data, selecting specific distributions and selecting the best distribution. Also contains items to learn more about the different distributions including generating example sets of data and skewness-kurtosis plots. The second group of buttons on the toolbar are shortcuts for the menu items from the *Analysis* menu. The available menu items are:

Test Distribution	Tests whether the selected distribution displayed in the *Data* window fits the corresponding data. The results are displayed in the *Test Distribution* window. Shortcuts are to click the *Test Distribution* button on the toolbar or in the *Data* window.
Find Best Distribution	Determines the best distribution to fit to the data. It fits all the available distributions using both the method of moments and maximum likelihood methods, with and without pre-transforming the data. It returns the distribution giving the highest p-value per the Skewness-Kurtosis All test. The result is displayed in the *Selected Distribution* box of the *Data* window. Shortcuts are to click the *Find Best Distribution* button on the toolbar or in the *Data* window.

View Skewness-Kurtosis Plot	Displays a skewness-kurtosis plot in a *Skewness-Kurtosis Plot* window summarizing the shapes each distribution can take and illustrating relationships between the different distributions. This is primarily a learning tool. However, if data exists, its location is shown on the plot. Shortcuts are to click the *View Skewness-Kurtosis Plot* button on the toolbar or in the *Data* window.
Select Distribution	Displays the *Distribution* submenu for selecting a specific distribution to fit to the data. When a distribution is selected, it is fit to the data using both method of moments and maximum likelihood. The best fit is returned. The following distributions and family of distributions can be fit: Beta, Exponential, Negative Exponential, Largest Extreme Value, Smallest Extreme Value, Gamma, Johnson, Loglogistic, Lognormal, Normal, Pearson and Uniform. In addition there is a *Custom* menu item that displays the *Select Distribution to Fit* dialog box allowing complete control in fitting the data including the method of fitting the data, whether pre-transformation are performed and more. Shortcuts are to click the *Select Distribution* button on the toolbar or to click the *Select Distribution* button in the *Data* window. The difference between selecting the *No Transformatio*n menu item and the *Normal* menu item is that the *No Transformation* menu item does not transform the data (Y=X) and the *Normal* menu item transforms the data using the transformation Y = (X-Average) / Standard Deviation.
Fill Order/Group Column	Displays the *Fill Order/Group Column* dialog box for generating patterned data for the *Group* and *Order* columns in the *Data* window. This saves time relative to typing the information into these columns.
Generate Random Values	Displays the *Generate Random Values* dialog box for generating example sets of data from any of the available distributions. This is a learning tool. A shortcut is to click the *Generate Random Values* button on the toolbar.
Dice Experiments	Displays the *Dice Experiments* dialog box for generating example sets of data based on experiments involving rolling dice. This is a learning tool.

Options	Displays the *Analysis Options* dialog box for specifying default options for calculating tolerance intervals, confidence statements, and testing for normality. A shortcut is to click the *Options* button on the toolbar.
Show Toolbar	By default, the toolbar is displayed. This menu item can be used to remove the toolbar.
Display Tabs at Bottom	By default, the tabs in the *Data* and *Test Distribution* windows are displayed at the bottom. This menu item can be used to move the tabs to the top of the window.

4.4 Window Menu

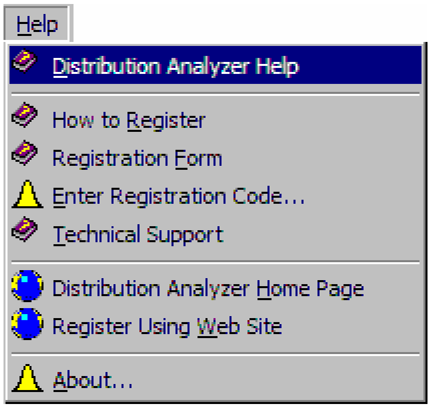


The *Window* menu is the fourth menu on the menu bar. It is used to rearrange the child windows or to select a particular child window. A list of opened child windows is appended to the bottom of this menu. These can be used to select a child window. Selecting a child window brings it to the top. Clicking on a child window has the same affect but the menu items can be used when the child window is buried under other windows.

Tile	Displays all the child windows in a tile pattern. The windows are reduced in size so that none of the windows overlap. A shortcut is to press shift+F4.
Cascade	Stacks the child windows slightly offset from each other so that the caption of each window is visible. A shortcut is to press shift+F5.
Arrange Icons	Places the icons of all child windows that are minimized along the bottom of the parent window.

At the bottom of the *Window* menu, a list of currently opened child windows is displayed. Selecting one of these menu items will activate that child window (display it on top of the others).

4.5 Help Menu



The *Help* menu is the last menu on the menu bar. It is used to obtain help, to register the software, to obtain technical support and to link to our web site. The last group of buttons on the toolbar provides short cuts for some of the more commonly used menu items on the *Help* menu. The available menu items are:

Distribution Analyzer Help	Displays table of contents or index of the help system. A short cut is to press the *F1* key.
How to Register	Displays instructions for registering the software. Registration and payment is required to continue using the software after the 30-day evaluation period.
Registration Form	Displays a registration form that can be printed out and used to register your copy of Distribution Analyzer by fax or mail.
Enter Registration Code	If the registration information is in the clipboard, selecting this menu item will automatically complete the registration process. If not, the *Registration* dialog box will be displayed for entering the user name and registration code (See Section 1.3).
Technical Support	Displays instructions for obtaining technical support.
Distribution Analyzer Home Page	If you have an internet connection and a web browser, displays the home page for Distribution Analyzer containing downloads of the most recent version and up to date information.

Register Using Web Site	If you have an internet connection and a web browser, displays the web page for registering your copy of Distribution Analyzer over our secure web server using a credit card.
About	Displays copyright and registration information in the *About* dialog box.

Distribution Analyzer

Version 1.0

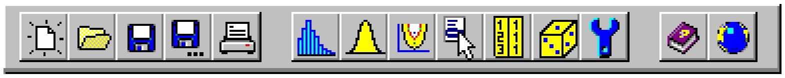
Registered to: Wayne Taylor

Registration Date: April 25, 2006

Copyright © 2006-2007 Taylor Enterprises, Inc.
All Rights Reserved

www.variation.com
info@variation.com
(847) 367-1032

4.6 Toolbar



The toolbar buttons provide short cuts for the more frequently used menu items. Holding the mouse cursor over a toolbar button displays a popup description of the button. The menu items associated with each button is given in the table below.

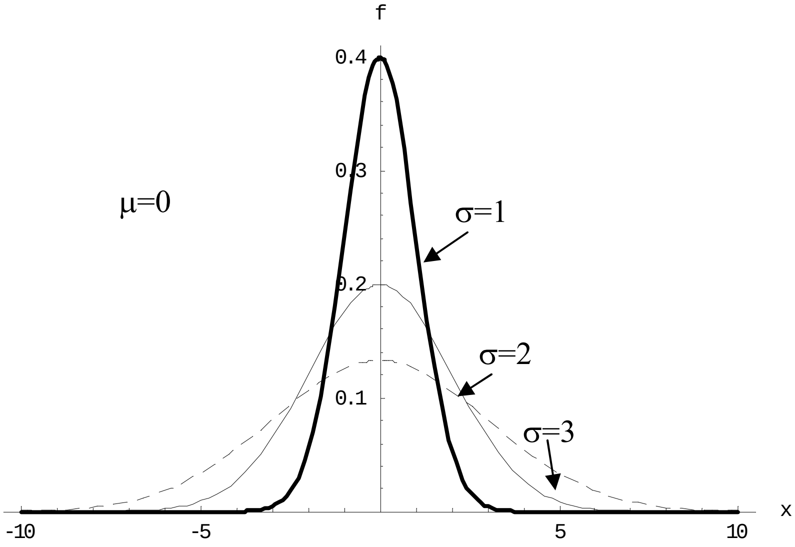
Control	Menu Item	Menu
🗋	***New***	***File***
📂	***Open…***	***File***
💾	***Save***	***File***
💾	***Save As…***	***File***
🖨	***Print***	***File***
📊	***Test Distribution***	***Analysis***
▲	***Find Best Distribution***	***Analysis***
⋃	***View Skewness-Kurtosis Plot***	***Analysis***
☰	***Select Distribution***	***Analysis***
⁞	***Fill Order/Group Column***	***Analysis***
🎲	***Generate Random Values***	***Analysis***
🔧	***Options***	***Analysis***
📕	***Distribution Analyzer Help***	***Help***
🌐	***Distribution Analyzer Home Page***	***Help***

5 Distributions

This chapter provides descriptions of the distributions and family of distributions included in Distribution Analyzer.

5.1 What is a Distribution?

A distribution is a model for how a set of data is suppose to behave. If describes this behavior in terms of probabilities. From this model it can be determined the probability a value is below a certain value or the probability a value is within a certain range. This model is described in terms of the density function of the distribution. Density functions for the normal distribution are shown below. They are essentially smoothed out histograms (where the width of the bars go to zero).



The mathematical form of the density function is:

$$f(x|\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

Parameters:	Location:	μ	$-\infty < \mu < \infty$
	Scale:	σ	$\sigma > 0$

Probabilities are obtained from the density function as areas under the curve. The probability that a data point is between 1 and 2 is the area under the density function between 1 and 2. For all density functions, the total area under the curve is one.

The parameters of a distribution are variables included in the density function so that the distribution can be adapted to a variety of situations. Of greatest importance is the number of parameters as shown below:

2 Parameters: The two parameters determine the average and standard deviation of the distribution. Such distributions are represented as a point on a skewness-kurtosis plot as they have fixed values of the skewness and kurtosis. Examples are the exponential, normal and uniform distributions.

3 Parameters: The three parameters determine the average, standard deviation and skewness of the distribution. Such distributions are represented as a curve on a skewness-kurtosis plot as the kurtosis depends of the skewness. Examples are the gamma and lognormal distributions.

4 Parameters: The four parameters determine the average, standard deviation, skewness and kurtosis of the distribution. Such distributions are represented as a region on a skewness-kurtosis plot as they can take on a variety of skewness and kurtosis values. Examples are the beta, Johnson and Pearson distributions.

Different books and articles will sometimes parameterize the same distribution differently. One set of parameters can always be calculated from the other. Further, sometimes different numbers of parameters are used so there are 2 and 3 parameter versions of the lognormal distribution. This greatly complicates comparing and using distributions. For this reason, Distribution Analyzer represents all the distributions in terms of their moments (average, standard deviation, skewness and kurtosis). Further, the distributions are broadened to always include at least two parameters so that they cover all possible averages and standard deviations.

Certain distributions like the normal distribution and logistic distribution are unbounded. Values generated from these distributions range from -infinity to infinity. Distributions like the lognormal distribution and gamma distribution have lower bounds. They range from this lower bound to infinity. Distributions like the negative of the lognormal distribution and negative of the gamma distribution have upper bounds. They range from -infinity to this upper bound. Finally distributions like the beta distribution have both upper and lower bounds. They range from the lower bound to the upper bound.

A family of distributions is several distributions combined so that they cover a well define region in a skewness-kurtosis plot. For example, the lognormal family of distributions includes the lognormal, negative lognormal and normal distributions. This allows the family to fit all possible average, standard deviation and skewness values. It appears as a curve in the skewness-kurtosis plot. The lognormal, negative lognormal and normal distributions are distinct distributions because they have different density and distribution functions.

The negative of a distribution is the distribution of -X (negative of the values) when X varies according to the distribution. For example, if the value X follows the lognormal distribution, then -X follows the negative lognormal distribution.

The moments of -X are:

Average of (-X) = - (Average of X)

Standard Deviation of (-X) = Standard Deviation of X

Skewness of (-X) = - (Skewness of X)

Kurtosis of (-X) = Kurtosis of X

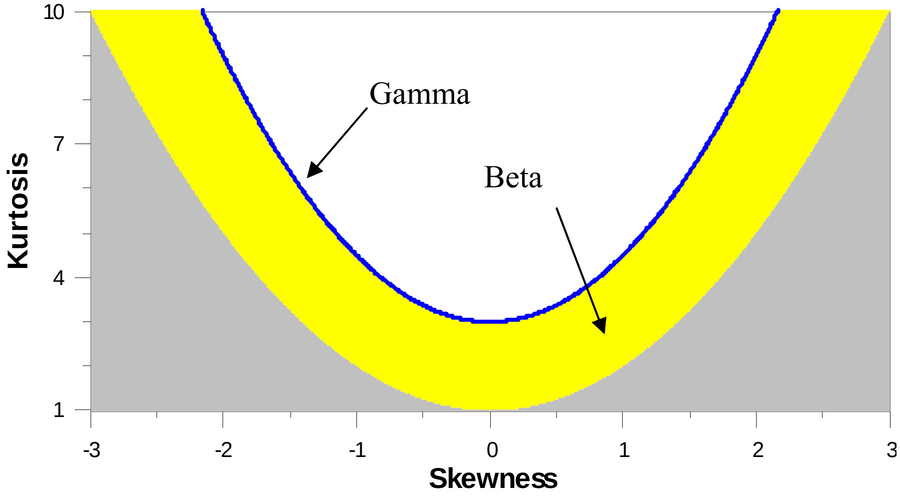
Most importantly, the skewness changes signs. There are many distributions like the lognormal distribution that only fit positive skewness data. Combining the lognormal (skewness > 0) with the negative of the lognormal (skewness < 0) and the normal distribution (skewness = 0) creates a family of distributions that coves all possible combinations of the average, standard deviation and skewness.

The remainder of this chapter provides specific details concerning the distributions used by Distribution Analyzer.

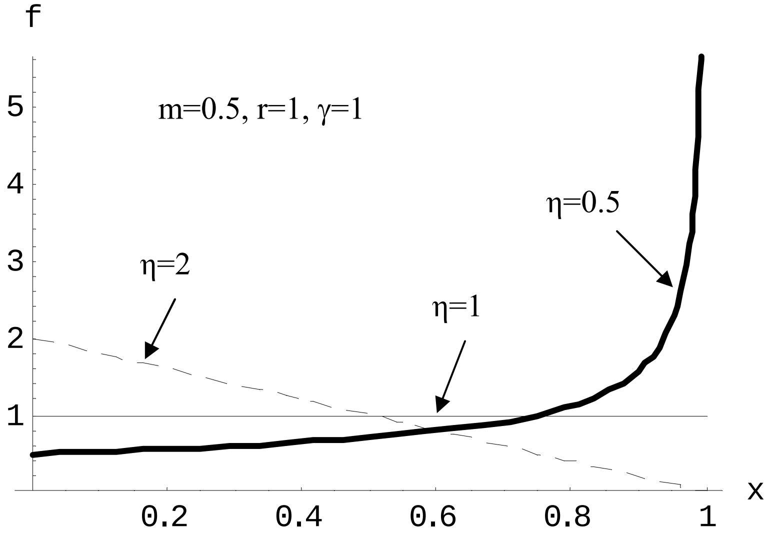
5.2 Beta Distribution

Shape: The beta distribution is a 4-parameter distribution that is represented by a region between the gamma curve and the impossible region (gray area) on a skewness-kurtosis plot as shown below. For a specified skewness, it covers the following range of kurtosis values:

$$\text{Skewness}^2 + 1 < \text{Kurtosis} < 3\left(1 + \frac{\text{Skewness}^2}{2}\right)$$



Density Function: A plot of the density function of the beta distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f\left(x \middle| m, r, \gamma, \eta\right) = \begin{cases} 0 & x < m - \frac{r}{2} \text{ or } x > m + \frac{r}{2} \\ \frac{\Gamma(\gamma + \eta)}{r\Gamma(\gamma)\Gamma(\eta)} \left(\frac{x - m + \frac{r}{2}}{r} \right)^{\gamma - 1} \left(\frac{m + \frac{r}{2} - x}{r} \right)^{\eta - 1} & m - \frac{r}{2} \le x \le m + \frac{r}{2} \end{cases}$$

Parameters:	Location:	m	$-\infty < m < \infty$
	Scale:	r	$r > 0$
	Shape:	γ	$\gamma > 0$
		η	$\eta > 0$

Bounds: Bounded below by m-r/2 and above by m+r/2.

Moments: The moments of the beta distribution can be calculated from the parameters as shown below:

Mean: $m + r\left(\frac{\gamma}{\gamma + \eta} - \frac{1}{2} \right)$

Standard Deviation: $r\sqrt{\frac{\eta\gamma}{(\eta + \gamma)^2(\eta + \gamma + 1)}}$

Skewness: $\frac{2(\eta - \gamma)\sqrt{\gamma + \eta + 1}}{\sqrt{\eta\gamma}(\eta + \gamma + 2)}$ which can range from -∞ to ∞.

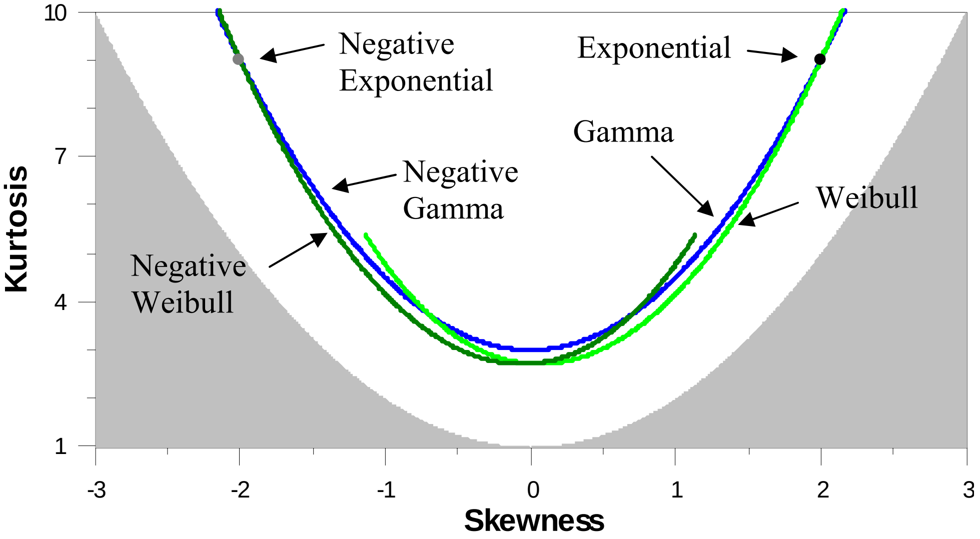
Kurtosis: $\frac{3(\eta + \gamma + 1)\left[2(\eta + \gamma)^2 + \gamma\eta(\eta + \gamma - 6)\right]}{\eta\gamma(\eta + \gamma + 2)(\eta + \gamma + 3)}$

Properties:

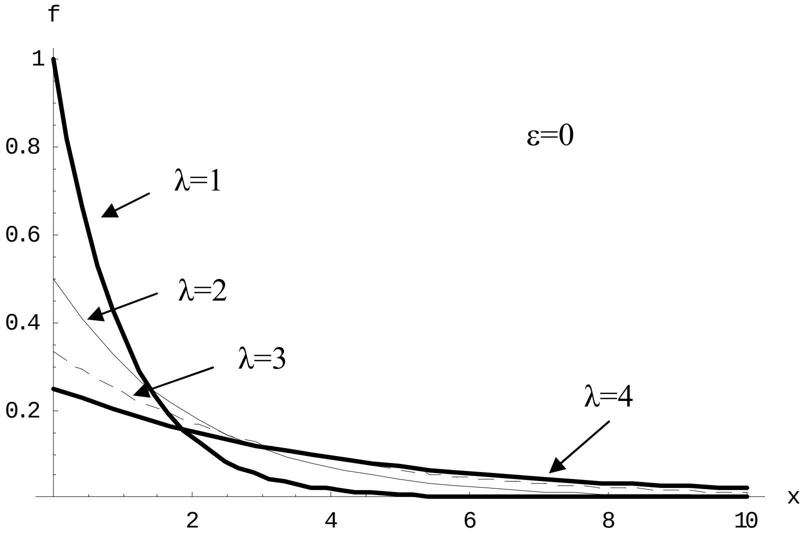
- The uniform distribution is a special case.

5.3 Exponential Distribution

Shape: The exponential distribution is a 2-parameter distribution and covers any specified average and standard deviation. It is represented by a single point with a skewness of 2 and kurtosis of 9 (excess kurtosis of 6) on a skewness-kurtosis plot as shown below. It is at the intersection of the gamma and Weibull distributions.



Density Function: The density function of the exponential distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\lambda) = \begin{cases} 0 & x \le \varepsilon \\ \frac{1}{\lambda} e^{-\frac{(x-\varepsilon)}{\lambda}} & x > \varepsilon \end{cases}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	λ	$\lambda > 0$

Bounds: Bounded below by ε.

Moments: The moments of the exponential distribution can be calculated from the parameters as shown below:

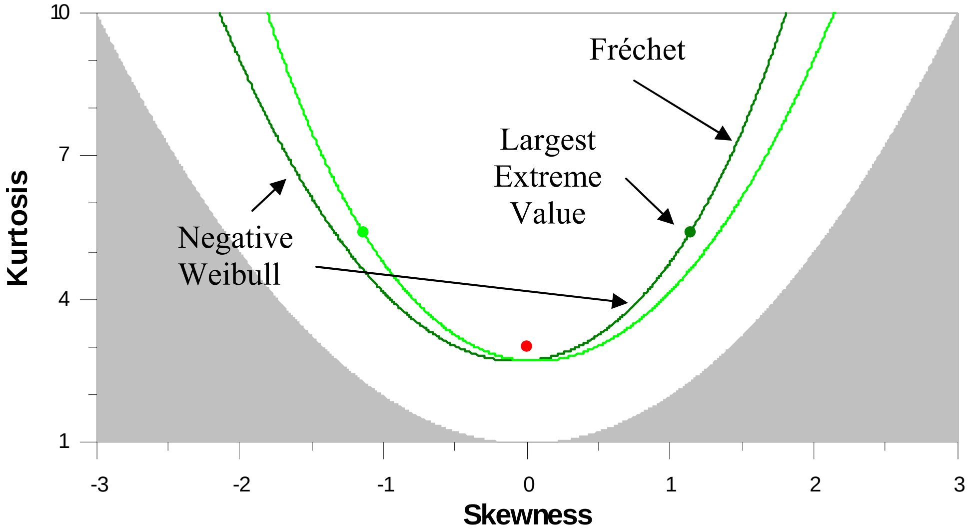
Mean:	$\varepsilon + \lambda$
Standard Deviation:	λ
Skewness:	2
Kurtosis:	9

Properties:

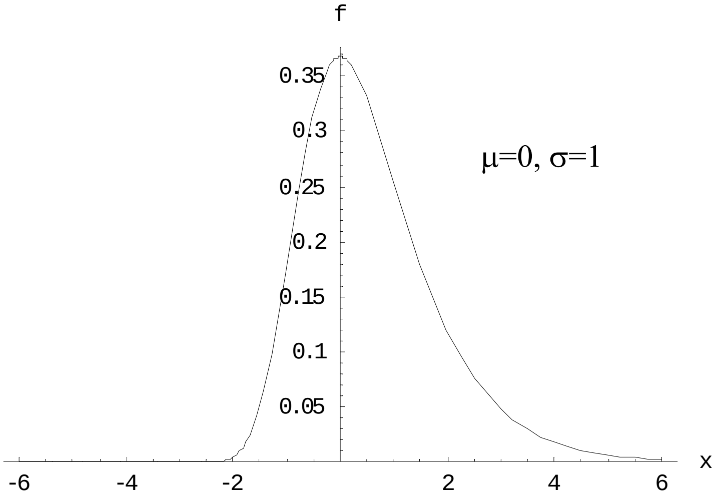
- The exponential distribution is a special case of the both the gamma and Weibull distributions falling at the intersection of these two curves on the skewness-kurtosis plot.

5.4 Extreme Value, Largest Family (Fréchet)

Shape: The largest extreme value family of distributions is made up of three distributions: Fréchet, negative Weibull and largest extreme value. It covers any specified average, standard deviation and any skewness below 5.6051382. Together they form a 3-parameter family of distributions that is represented by a curve on a skewness-kurtosis plot as shown below. The Fréchet distribution covers the portion of the curve with skewness above 1.139547. The negative Weibull distribution covers the portion of the curve with skewness below 1.139547. The largest extreme value distribution handles the remaining case of skewness equal to 1.139547.



Density Function - Largest Extreme Value Distribution: The density function of the largest extreme value distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\mu,\sigma)=\frac{1}{\sigma}e^{\left[-\frac{1}{\sigma}(x-\mu)-e^{\left[-\frac{1}{\sigma}(x-\mu)\right]}\right]} \quad \text{for } -\infty < x < \infty$$

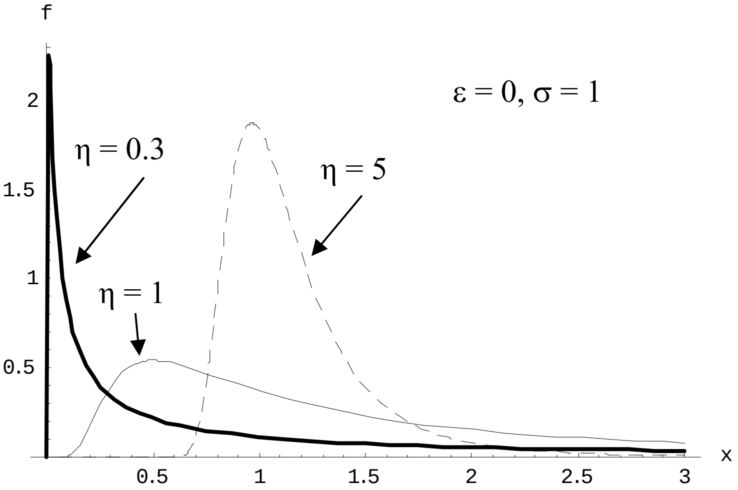
Parameters:	Location:	μ	$-\infty < \mu < \infty$	mode
	Scale:	σ	$\sigma > 0$	

Bounds: Unbounded

Moments - Largest Extreme Value Distribution: The moments of the largest extreme value distribution can be calculated from the parameters as shown below:

Mean:	$\mu + \text{Euler Gamma } \sigma = \mu + 0.57721566490153286061\,\sigma$
Standard Deviation:	$\frac{\pi}{\sqrt{6}}\sigma = 1.2825498301618640955\,\sigma$
Skewness:	$\frac{12\sqrt{6}\ \text{Zeta}[3]}{\pi^3} = 1.1395470994046486575$
Kurtosis:	5.4

Density Function - Fréchet: The density function of the Fréchet distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\sigma,\eta)=\begin{cases}\dfrac{\eta}{\sigma}\left(\dfrac{x-\varepsilon}{\sigma}\right)^{-\eta-1}e^{-\left(\frac{x-\varepsilon}{\sigma}\right)^{-\eta}} & x>\varepsilon\\ 0 & x\leq\varepsilon\end{cases}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	σ	$\sigma > 0$
	Shape:	η	$\eta > 0$

Bounds: Bounded below by ε.

Moments - Fréchet: The moments of the Fréchet distribution can be calculated from the parameters as shown below:

Mean: $$\varepsilon+\sigma\Gamma\left(1-\frac{1}{\eta}\right)$$

Standard Deviation: $$\sigma\sqrt{\Gamma\left(1-\frac{2}{\eta}\right)-\left(\Gamma\left(1-\frac{1}{\eta}\right)\right)^2}$$

Skewness: $$\frac{\Gamma\left(1-\frac{3}{\eta}\right)-3\Gamma\left(1-\frac{2}{\eta}\right)\Gamma\left(1-\frac{1}{\eta}\right)+2\left(\Gamma\left(1-\frac{1}{\eta}\right)\right)^3}{\left(\Gamma\left(1-\frac{2}{\eta}\right)-\left(\Gamma\left(1-\frac{1}{\eta}\right)\right)^2\right)^{3/2}}$$ for $\eta>3$

Kurtosis: $$\frac{\Gamma\left(1-\frac{4}{\eta}\right)-4\Gamma\left(1-\frac{3}{\eta}\right)\Gamma\left(1-\frac{1}{\eta}\right)+6\Gamma\left(1-\frac{2}{\eta}\right)\left(\Gamma\left(1-\frac{1}{\eta}\right)\right)^2-3\left(\Gamma\left(1-\frac{1}{\eta}\right)\right)^4}{\left(\Gamma\left(1-\frac{2}{\eta}\right)-\left(\Gamma\left(1-\frac{1}{\eta}\right)\right)^2\right)^2}$$ for $\eta>4$

Properties:

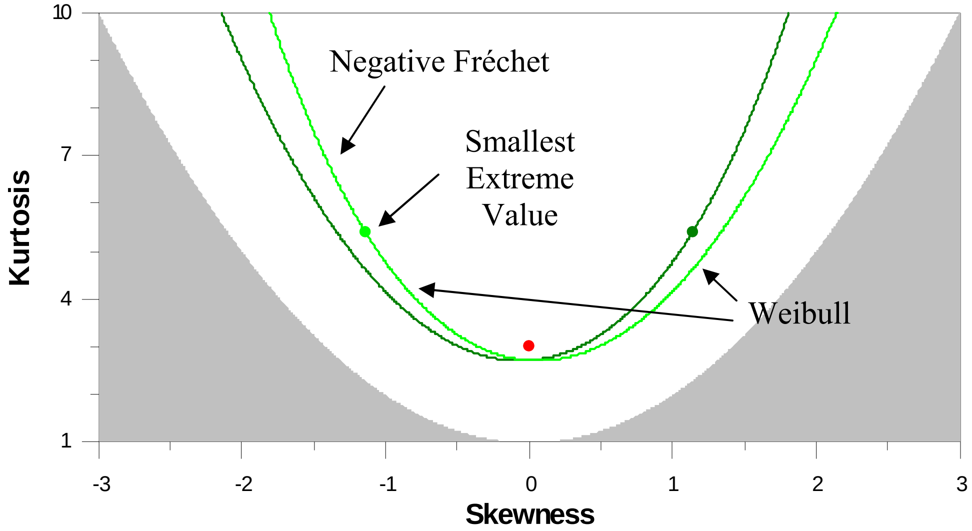
- The largest extreme value family of distributions are the negatives of the smallest extreme value family of distributions. In particular, the largest extreme distribution is the negative of the smallest extreme value distribution.

- The negative exponential distribution is a special case of the negative Weibull distribution and thus falls on the negative Weibull curve in the skewness-kurtosis plot.

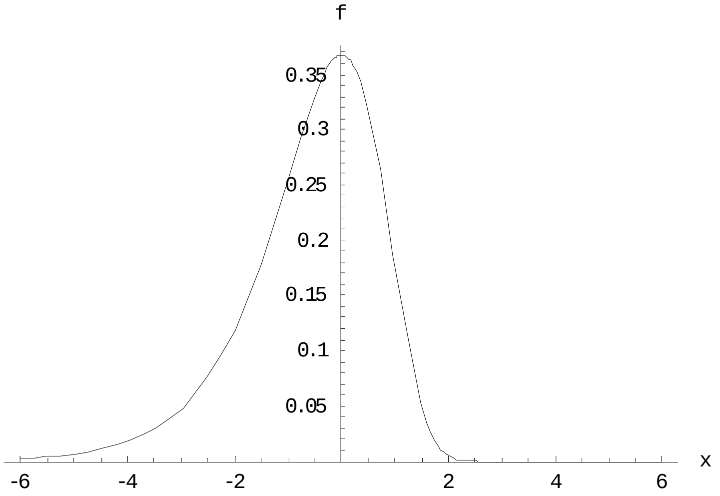
- The negative Weibull and largest extreme value distributions are the distributions of maximums. Under certain restrictions, the maximum of distributions without upper bounds tends to the largest extreme value distribution and the maximum of distributions bounded above tends to the negative Weibull.

5.5 Extreme Value, Smallest Family (Weibull)

Shape: The smallest extreme value family of distributions is made up of three distributions: Weibull, negative Fréchet and smallest extreme value. It covers any specified average, standard deviation and any skewness above -5.6051382. Together they form a 3-parameter family of distributions that is represented by a curve on a skewness-kurtosis plot as shown below. The Weibull distribution covers the portion of the curve with skewness above -1.139547. The negative Fréchet distribution covers the portion of the curve with skewness below -1.139547. The smallest extreme value distribution handles the remaining case of skewness equal to -1.139547.



Density Function - Smallest Extreme Value Distribution: The density function of the smallest extreme value distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\mu,\sigma) = \frac{1}{\sigma} e^{\left[\frac{1}{\sigma}(x-\mu) - e^{\left[\frac{1}{\sigma}(x-\mu)\right]}\right]} \text{ for } -\infty < x < \infty$$

Parameters:	Location:	μ	$-\infty < \mu < \infty$	mode
	Scale:	σ	$\sigma > 0$	

Bounds: Unbounded

Moments - Smallest Extreme Value Distribution: The moments of the smallest extreme value distribution can be calculated from the parameters as shown below:

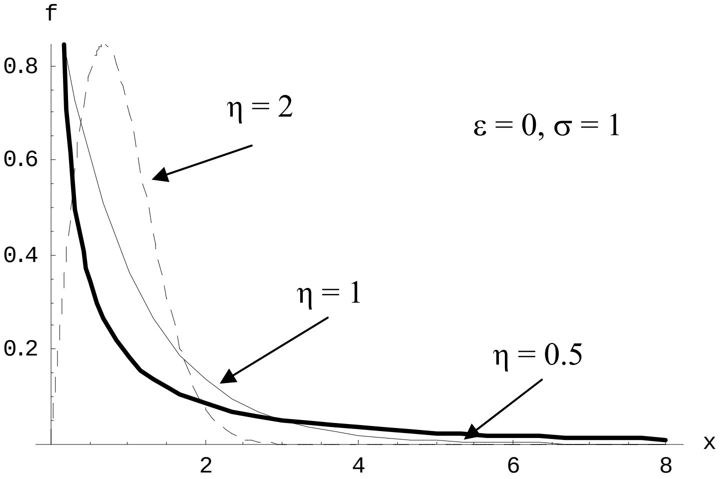
Mean: $\mu - \text{Euler Gamma } \sigma = \mu - 0.57721566490153286061\,\sigma$

Standard Deviation: $\frac{\pi}{\sqrt{6}}\sigma = 1.2825498301618640955\,\sigma$

Skewness: $-\frac{12\sqrt{6}\ \text{Zeta}[3]}{\pi^3} = -1.1395470994046486575$

Kurtosis: 5.4

Density Function - Weibull: The density function of the Weibull distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\sigma,\eta)=\begin{cases} 0 & x \le \varepsilon \\ \dfrac{\eta}{\sigma}\left(\dfrac{x-\varepsilon}{\sigma}\right)^{\eta-1} e^{-\left(\frac{x-\varepsilon}{\sigma}\right)^{\eta}} & x > \varepsilon \end{cases}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	σ	$\sigma > 0$
	Shape:	η	$\eta > 0$

Bounds: Bounded below by ε.

Moments – Weibull: The moments of the Weibull distribution can be calculated from the parameters as shown below:

Mean:

$$\varepsilon+\sigma\Gamma\left(1+\frac{1}{\eta}\right)$$

Standard Deviation:

$$\sigma\sqrt{\Gamma\left(1+\frac{2}{\eta}\right)-\left(\Gamma\left(1+\frac{1}{\eta}\right)\right)^2}$$

Skewness:

$$\frac{\Gamma\left(1+\frac{3}{\eta}\right)-3\Gamma\left(1+\frac{2}{\eta}\right)\Gamma\left(1+\frac{1}{\eta}\right)+2\left(\Gamma\left(1+\frac{1}{\eta}\right)\right)^3}{\left(\Gamma\left(1+\frac{2}{\eta}\right)-\left(\Gamma\left(1+\frac{1}{\eta}\right)\right)^2\right)^{3/2}}$$

Kurtosis:

$$\frac{\Gamma\left(1+\frac{4}{\eta}\right)-4\Gamma\left(1+\frac{3}{\eta}\right)\Gamma\left(1+\frac{1}{\eta}\right)+6\Gamma\left(1+\frac{2}{\eta}\right)\left(\Gamma\left(1+\frac{1}{\eta}\right)\right)^2-3\left(\Gamma\left(1+\frac{1}{\eta}\right)\right)^4}{\left(\Gamma\left(1+\frac{2}{\eta}\right)-\left(\Gamma\left(1+\frac{1}{\eta}\right)\right)^2\right)^2}$$

Properties:

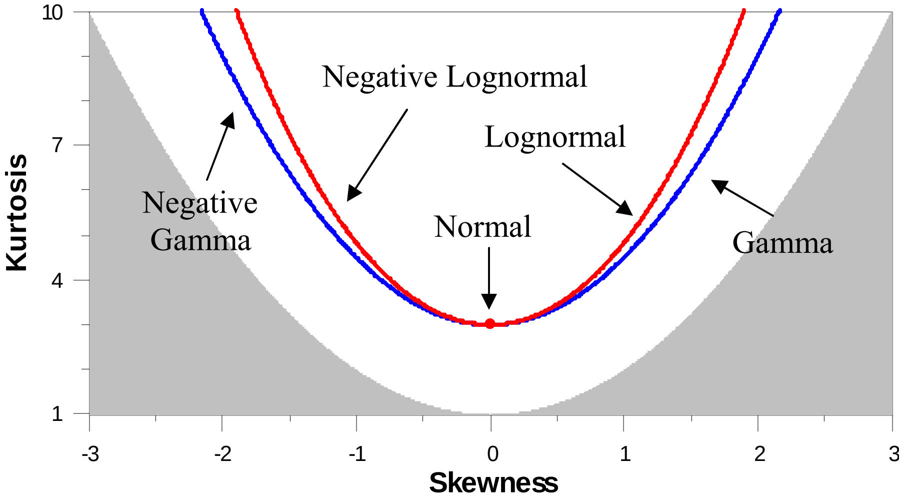
- The smallest extreme value family of distributions are the negatives of the largest extreme value family of distributions. In particular, the smallest extreme distribution is the negative of the largest extreme value distribution.

- The exponential distribution is a special case of the Weibull distribution and thus falls on the Weibull curve in the skewness-kurtosis plot.

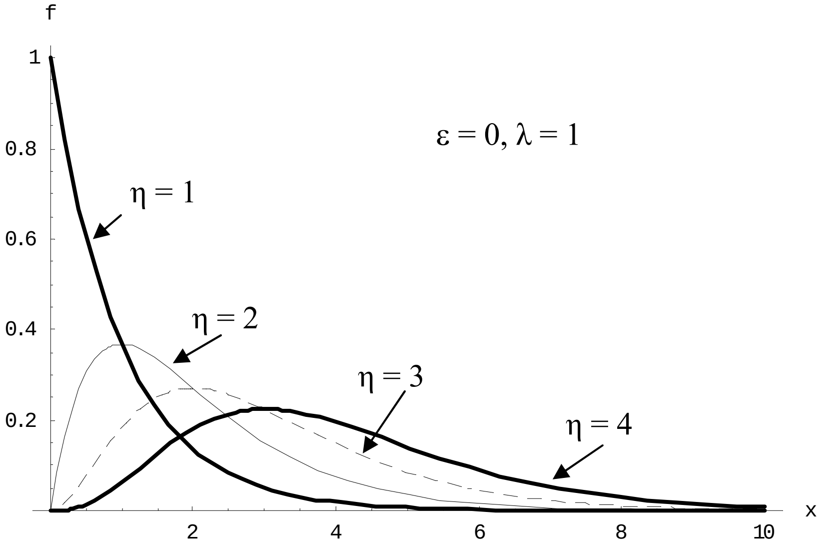
- The Weibull and smallest extreme value distributions are the distributions of minimums. Under certain restrictions, the minimum of distributions without lower bounds tends to the smallest extreme value distribution and the minimum of distributions bounded below tends to the Weibull.

5.6 Gamma Family of Distributions

Shape: The gamma family of distributions is made up of three distributions: gamma, negative gamma and normal. It covers any specified average, standard deviation and skewness. Together they form a 3-parameter family of distributions that is represented by a curve on a skewness-kurtosis plot as shown below. The gamma distribution covers the positive skewness portion of the curve. The negative gamma distribution covers the negative skewness portion of the curve. The normal distribution handles the remaining case of zero skewness. The gamma curve falls below the lognormal curve.



Density Function: The density function of the gamma distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\lambda,\eta)=\begin{cases}0 & x\le\varepsilon\\ \dfrac{1}{\lambda^{\eta}\Gamma(\eta)}(x-\varepsilon)^{\eta-1}e^{-\frac{x-\varepsilon}{\lambda}} & x>\varepsilon\end{cases}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$	
	Scale:	λ	$\lambda > 0$	instantaneous failure rate
	Shape:	η	$\eta > 0$	number of failures

Bounds: Bounded below by ε.

Moments: The moments of the gamma distribution can be calculated from the parameters as shown below:

Mean: $\varepsilon + \lambda\eta$

Standard Deviation: $\lambda\sqrt{\eta}$

Skewness: $\dfrac{2}{\sqrt{\eta}}$ which is always positive

Kurtosis: $3+\dfrac{6}{\eta}$

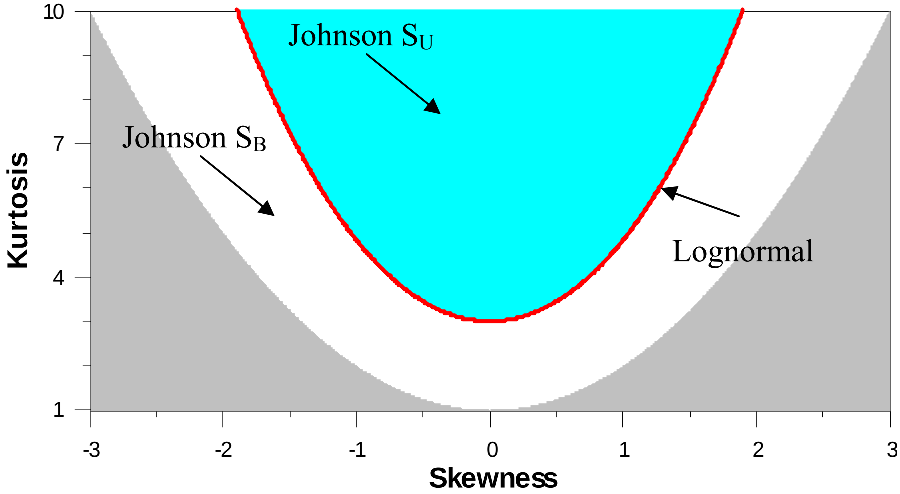
Properties:

- As the skewness goes to zero, both the gamma and negative gamma distributions limit to the normal distribution. This means that in some cases the gamma and normal distributions can be difficult to distinguish between. As a result, some sets of data may fit both the gamma and normal distributions.

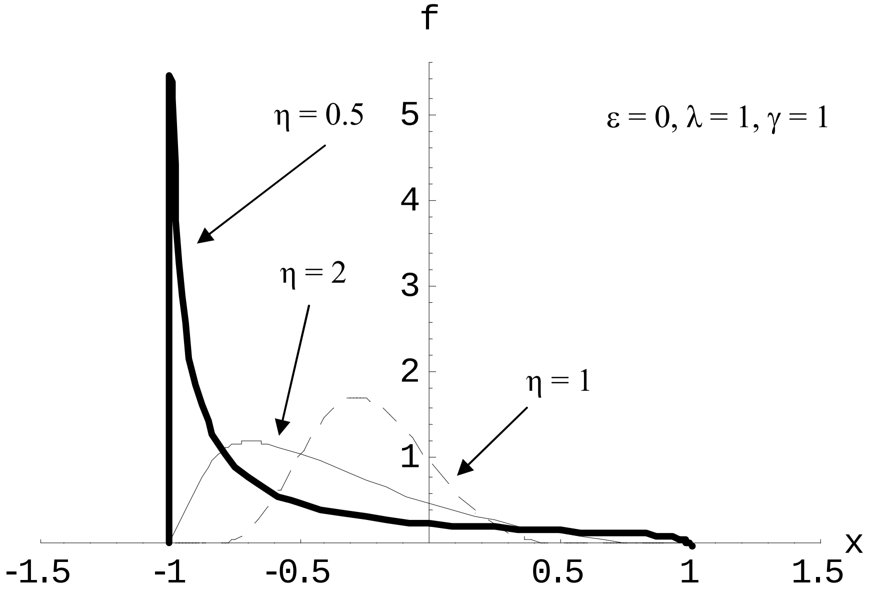
- The exponential distribution is a special case of the gamma distribution and thus falls on the gamma curve in the skewness-kurtosis plot.

5.7 Johnson Family

Shape: The Johnson family of distributions is made up of three distributions: Johnson S_U, Johnson S_B and lognormal. It covers any specified average, standard deviation, skewness and kurtosis. Together they form a 4-parameter family of distributions that covers the entire skewness-kurtosis region other than the impossible region. The Johnson S_U distribution covers the area above the lognormal curve and the Johnson S_B covers the area below the normal curve.



Density Function - S_B: The density function of the Johnson S_B distribution is shown below:



The equation, parameters and bounds of the density function are:

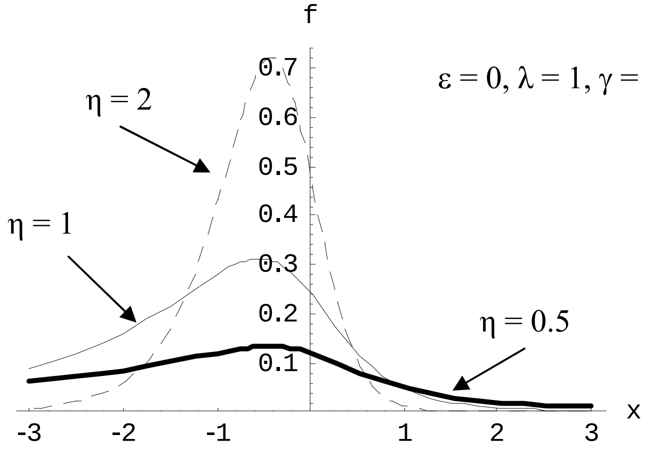
$$f(x|\eta,\gamma,r,m)=\begin{cases} 0 & x \le m-\frac{r}{2},\ x \ge m+\frac{r}{2} \\ \dfrac{\eta\, r}{\sqrt{2\pi}\left(x-m+\frac{r}{2}\right)\left(m+\frac{r}{2}-x\right)} e^{-\frac{1}{2}\left(\gamma+\eta \ln\left(\frac{x-m+\frac{r}{2}}{m+\frac{r}{2}-x}\right)\right)^2} & m-\frac{r}{2} < x < m+\frac{r}{2} \end{cases}$$

Parameters:	Location:	m	$-\infty < m < \infty$
	Scale:	r	$r > 0$
	Shape:	γ	$-\infty < \gamma < \infty$
	Shape:	δ	$\eta > 0$

Bounds: Lower bound = $m - r/2$
Upper bound = $m + r/2$

Moments - S_B: The moments of the Johnson S_B distribution do not have a simple expression.

Density Function - S_U: The density function of the Johnson S_U distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\eta,\gamma,\lambda,\varepsilon)=\frac{\eta}{\sqrt{(x-\varepsilon)^2+\lambda^2}\sqrt{2\pi}}\,e^{-\frac{1}{2}\left(\gamma+\eta\,\operatorname{asinh}\left(\frac{x-\varepsilon}{\lambda}\right)\right)^2}\quad \text{for } -\infty < x < \infty$$

Parameters:	Location:	ε	$-\infty<\varepsilon<\infty$
	Scale:	λ	$\lambda>0$
	Shape:	γ	$-\infty<\gamma<\infty$
		η	$\eta>0$

Bounds: Unbounded

Moments - S_U: The moments of the Johnson S_U distribution can be calculated from the parameters as shown below:

Mean:

$$\lambda\left(-e^{\frac{1}{2\eta^2}}\sinh\left(\frac{\gamma}{\eta}\right)\right)+\varepsilon \text{ where } \sinh=\frac{e^x-e^{-x}}{2}$$

Standard Deviation:

$$\lambda\sqrt{\frac{e^{\frac{1}{\eta^2}}-1}{2}\left(e^{\frac{1}{\eta^2}}\cosh\left(2\frac{\gamma}{\eta}\right)+1\right)}$$

Skewness:

$$-\frac{e^{\frac{1}{2\eta^2}}\left(e^{\frac{1}{\eta^2}}-1\right)^{1/2}\left(3\sinh\left(\frac{\gamma}{\eta}\right)+e^{\frac{1}{\eta^2}}\left(e^{\frac{1}{\eta^2}}+2\right)\sinh\left(3\frac{\gamma}{\eta}\right)\right)}{\sqrt{2}\left(1+e^{\frac{1}{\eta^2}}\cosh\left(\frac{2\gamma}{\eta}\right)\right)^{3/2}}$$

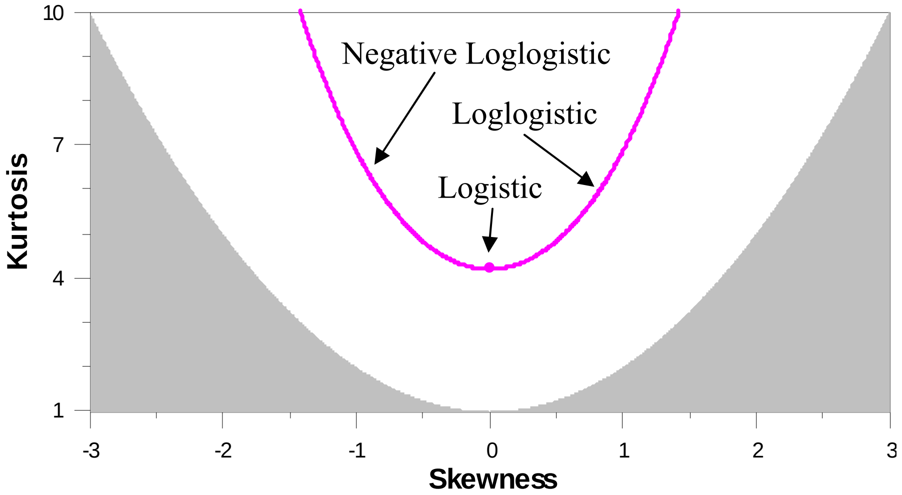
Kurtosis:

$$\frac{\left(\begin{array}{l}3+6e^{\frac{1}{\eta^2}}+4e^{\frac{2}{\eta^2}}\left(e^{\frac{1}{\eta^2}}+2\right)\cosh\left(\frac{2\gamma}{\eta}\right)\\ \qquad +e^{\frac{2}{\eta^2}}\left(-3+3e^{\frac{2}{\eta^2}}+2e^{\frac{3}{\eta^2}}+e^{\frac{4}{\eta^2}}\right)\cosh\left(\frac{4\gamma}{\eta}\right)\end{array}\right)}{2\left(1+e^{\frac{1}{\eta^2}}\cosh\left(\frac{2\gamma}{\eta}\right)\right)^2}$$

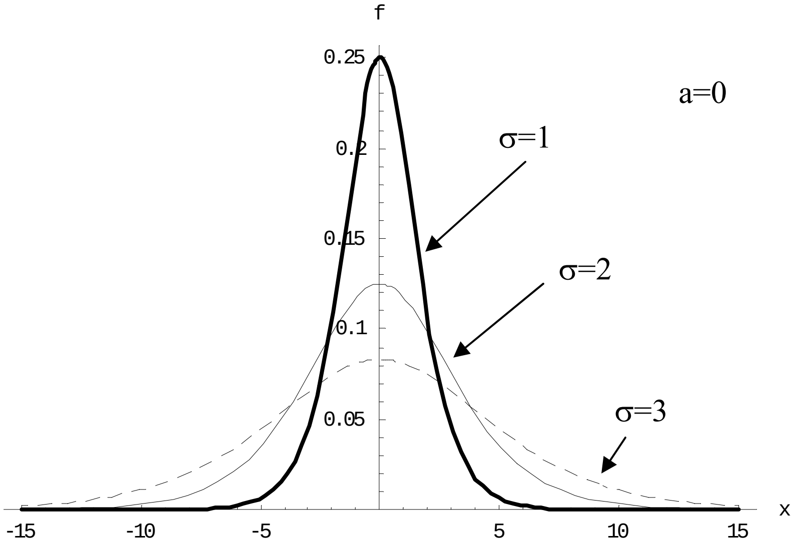
where $\cosh=\dfrac{e^x+e^{-x}}{2}$.

5.8 Loglogistic Family

Shape: The loglogistic family of distributions is made up of three distributions: loglogistic, negative loglogistic and logistic. It covers any specified average, standard deviation and any skewness in the range of -4.2847830295411833030 to 4.2847830295411833030. Together they form a 3-parameter family of distributions that is represented by a curve on a skewness-kurtosis plot as shown below. The loglogistic distribution covers the positive skewness portion of the curve. The negative loglogistic distribution covers the negative skewness portion of the curve. The logistic distribution handles the remaining case of zero skewness.



Density Function - Logistic: The density function of the logistic distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|a,b) = \frac{e^{\frac{x-a}{b}}}{b\left(1+e^{\frac{x-a}{b}}\right)^2} \quad \text{for } -\infty < x < \infty$$

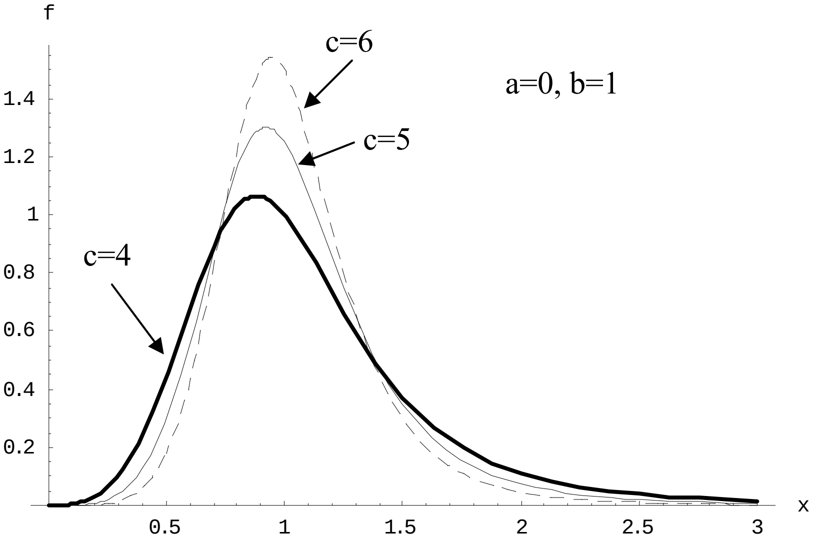
Parameters:	Location:	a	$-\infty < a < \infty$
	Scale:	b	$b > 0$

Bounds: Unbounded

Moments - Logistic: The moments of the logistic distribution can be calculated from the parameters as shown below:

Mean:	a
Standard Deviation:	$\frac{\pi b}{\sqrt{3}}$
Skewness:	0
Kurtosis:	4.2

Density Function - Loglogistic: The density function of the loglogistic distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f\left(x\middle|a,b\right)=\frac{e^{\frac{x-a}{b}}}{b\left(1+e^{\frac{x-a}{b}}\right)^{2}}\quad \text{for } -\infty < x < \infty$$

Parameters:	Location:	a	$-\infty < a < \infty$
	Scale:	b	$b > 0$
	Shape:	c	$c > 4$

Bounds: Bounded below by a.

Moments - Loglogistic: The moments of the loglogistic distribution can be calculated from the parameters as shown below:

Mean:

$$a+b\frac{\pi}{c}\operatorname{Csc}\left(\frac{\pi}{c}\right)\qquad (\operatorname{Csc}(x) = 1/\sin(x))$$

Standard Deviation:

$$b\sqrt{\frac{2\pi}{c}\operatorname{Csc}\left(\frac{2\pi}{c}\right)-\left(\frac{\pi}{c}\operatorname{Csc}\left(\frac{\pi}{c}\right)\right)^{2}}\quad \text{for } c>2.$$

Skewness:

$$\frac{2\pi^{2}\operatorname{Csc}\left(\frac{\pi}{c}\right)^{3}-6c\pi\operatorname{Csc}\left(\frac{\pi}{c}\right)\operatorname{Csc}\left(\frac{2\pi}{c}\right)+3c^{2}\operatorname{Csc}\left(\frac{3\pi}{c}\right)}{\sqrt{\pi}\left(-\pi\operatorname{Csc}\left(\frac{\pi}{c}\right)^{2}+2c\operatorname{Csc}\left(\frac{2\pi}{c}\right)\right)^{3/2}}$$

Kurtosis:

$$\frac{-3\pi^{3}\operatorname{Csc}\left(\frac{\pi}{c}\right)^{4}-12\pi c^{2}\operatorname{Csc}\left(\frac{\pi}{c}\right)\operatorname{Csc}\left(\frac{3\pi}{c}\right)+4c^{3}\operatorname{Csc}\left(\frac{4\pi}{c}\right)+6c\pi^{2}\operatorname{Csc}\left(\frac{\pi}{c}\right)^{3}\operatorname{Sec}\left(\frac{\pi}{c}\right)}{\pi\left(-\pi\operatorname{Csc}\left(\frac{\pi}{c}\right)^{2}+2c\operatorname{Csc}\left(\frac{2\pi}{c}\right)\right)^{2}}$$

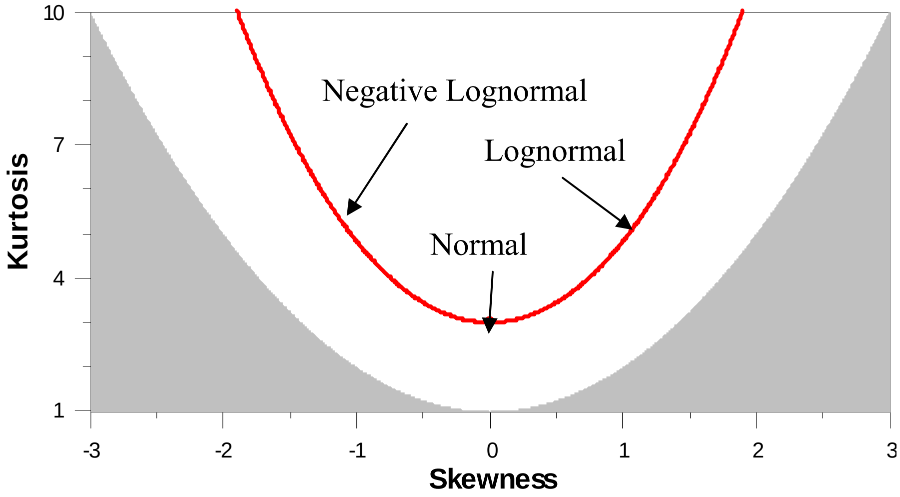
(Sec(x) = 1/cos(x))

Properties:

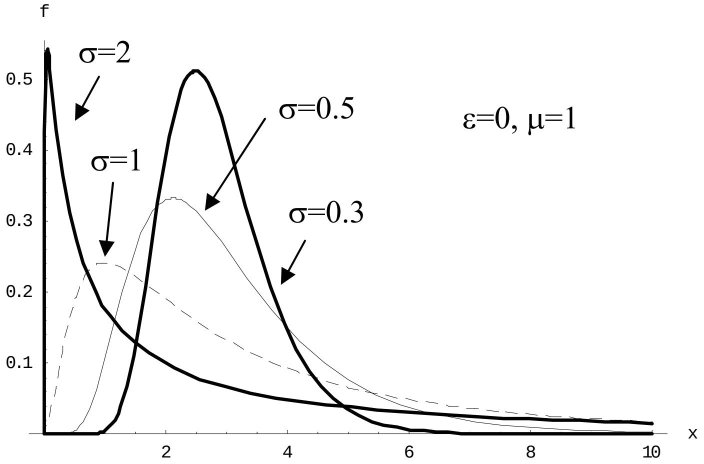
- Commonly used in growth models

5.9 Lognormal Family

Shape: The lognormal family of distributions is made up of three distributions: lognormal, negative lognormal and normal. It covers any specified average, standard deviation and skewness. Together they form a 3-parameter family of distributions that is represented by a curve on a skewness-kurtosis plot as shown below. The lognormal distribution covers the positive skewness portion of the curve. The negative lognormal distribution covers the negative skewness portion of the curve. The normal distribution handles the remaining case of zero skewness.



Density Function: The density function of the lognormal distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\mu,\sigma)=\begin{cases} 0 & x \le \varepsilon \\ \dfrac{1}{(x-\varepsilon)\sigma\sqrt{2\pi}}e^{-\frac{(\ln(x-\varepsilon)-\mu)^2}{2\sigma^2}} & x > \varepsilon \end{cases}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	μ	$-\infty < \mu < \infty$
	Shape:	σ	$\sigma > 0$

Bounds: Bounded below by ε.

Moments: The moments of the lognormal distribution can be calculated from the parameters as shown below:

Mean: $\varepsilon + e^{\left(\mu+\frac{\sigma^2}{2}\right)}$

Standard Deviation: $\sqrt{e^{\left(2\mu+\sigma^2\right)}\left(e^{\sigma^2}-1\right)}$

Skewness: $\sqrt{e^{\sigma^2}-1}\left(e^{\sigma^2}+2\right)$

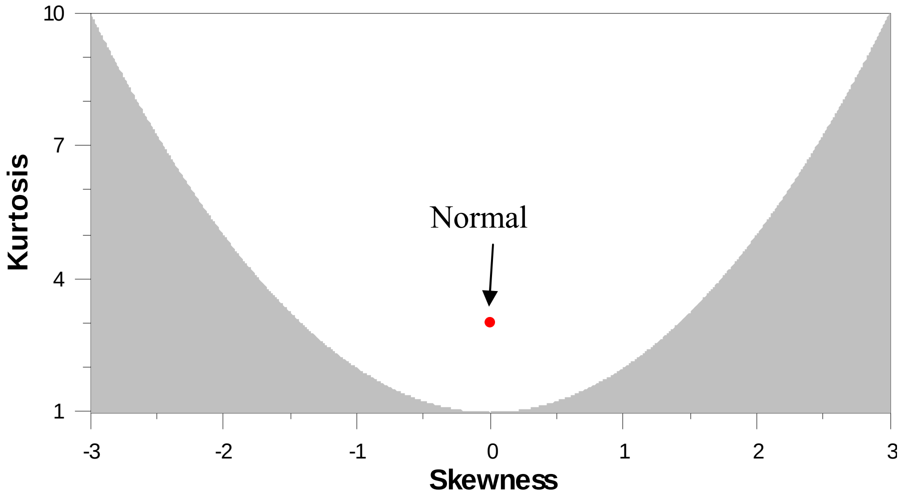
Kurtosis: $3+\left(e^{\sigma^2}-1\right)\left(e^{3\sigma^2}+3e^{2\sigma^2}+6e^{\sigma^2}+6\right)$

Properties:

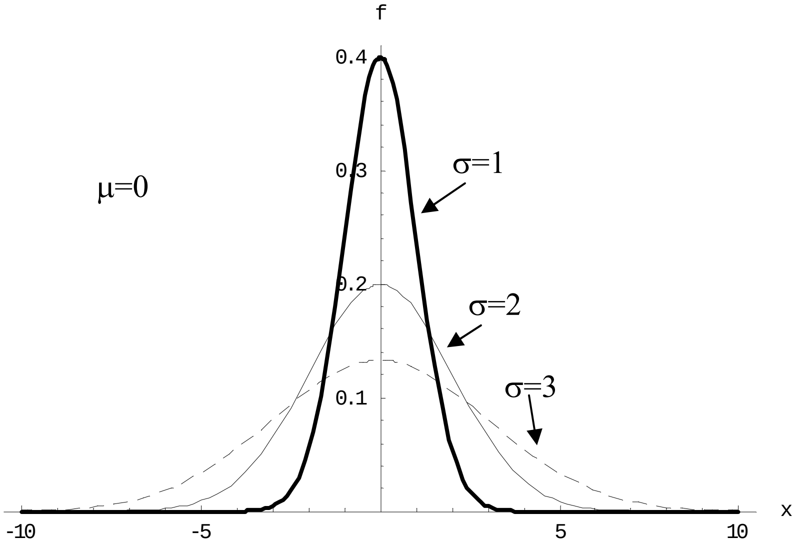
- As the skewness goes to zero, both the lognormal and negative lognormal distributions limit to the normal distribution. This means that in some cases the lognormal and normal distributions can be difficult to distinguish between. As a result, some sets of data may fit both the lognormal and normal distributions.
- The lognormal distribution is the distribution of multiplication and division. The central limit theorem states that as positive items are multiplied and divided, under certain restrictions, the result will tend to the lognormal distribution.

5.10 Normal Distribution

Shape: The normal distribution is a 2-parameter distribution and covers any specified average and standard deviation. It is represented by a single point with a skewness of zero and kurtosis of three (excess kurtosis of zero) on a skewness-kurtosis plot as shown below:



Density Function: The density function of the normal distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\mu,\sigma)=\begin{cases} 0 & x \le \varepsilon \\ \dfrac{1}{(x-\varepsilon)\sigma\sqrt{2\pi}}e^{-\frac{(\ln(x-\varepsilon)-\mu)^2}{2\sigma^2}} & x > \varepsilon \end{cases}$$

Parameters: Location: μ $-\infty < \mu < \infty$

Scale: σ $\sigma > 0$

Bounds: Unbounded

Moments: The moments of the normal distribution can be calculated from the parameters as shown below:

Mean: μ

Standard Deviation: σ

Skewness: 0

Kurtosis: 3

Properties:

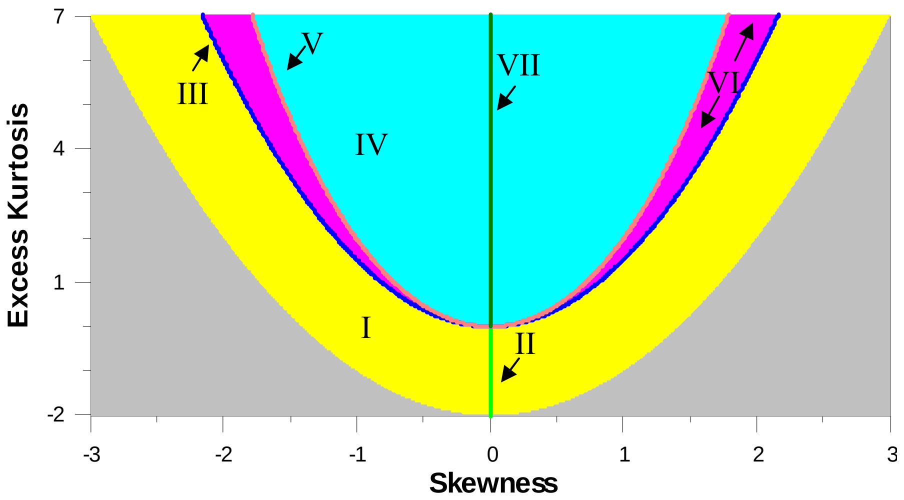
- The normal distribution is the distribution of addition and subtraction. The central limit theorem states that as items are added and subtracted together, under certain restrictions, the result will tend to the normal distribution.

5.11 Pearson Family

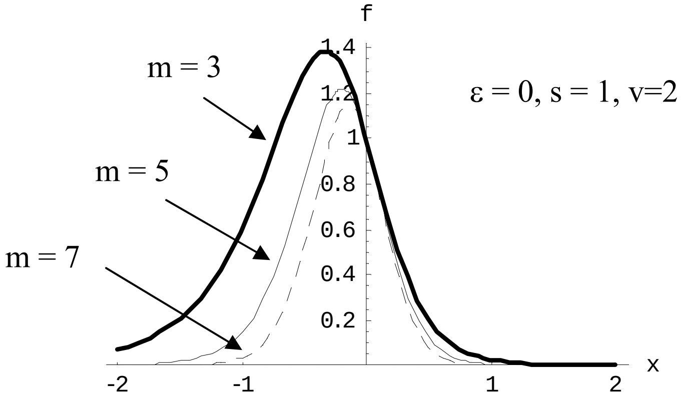
Shape: The Pearson family of distributions is made up of seven distributions: Type I-VII. It covers any specified average, standard deviation, skewness and kurtosis. Together they form a 4-parameter family of distributions that covers the entire skewness-kurtosis region other than the impossible region. The seven types are described below.

- Type I: Beta Distribution
- Type II: Special case of beta distribution that is symmetrical
- Type III: Gamma Distribution
- Type IV: Region above Type V
- Type V: 3 parameter distribution represented by curve
- Type VI: Region between Gamma and Type V
- Type VII: Special case of Type IV that is symmetrical

The special cases can be ignored (II and VII). Type I and III are alias for distribution already covered. That leaves Type IV, V and VI as new distributions.



Density Function – Type IV: The density function of the Type IV Pearson distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon, s, v, m) = C\left(1+\left(\frac{x-\varepsilon}{s}\right)^2\right)^{-m} e^{-v\tan^{-1}\left(\frac{x-\varepsilon}{s}\right)} \quad \text{for } -\infty < x < \infty$$

$$\text{where } C = \frac{1}{s\,F[2m-2, v]} \quad \text{where } F(r,v) = e^{-\frac{1}{2}v\pi}\int_0^{\pi} e^{v\,\phi}\sin(\phi)^r\,d\phi$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	s	$s > 0$
	Shape:	v	$-\infty < v < \infty$
	Shape:	m	$m > 2.5$

Bounds: Unbounded

Moments – Type IV: The moments of the Type IV Pearson distribution can be calculated from the parameters as shown below:

Skewness: $$-\frac{2v}{m-2}\sqrt{\frac{2m-3}{v^2+(2m-2)^2}}$$

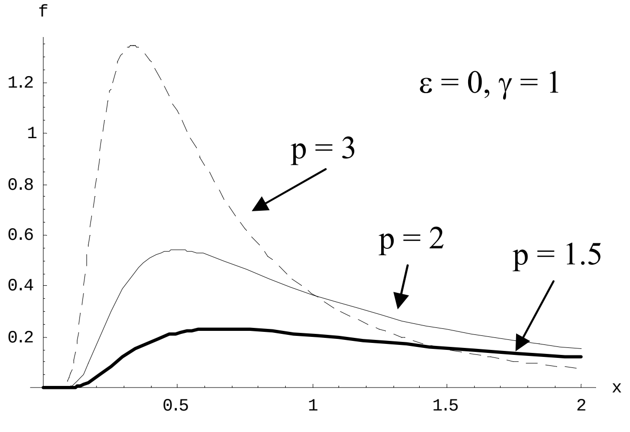
Kurtosis: $$\frac{(3m-6)\text{Skewness}^2+6m-9}{2m-5}$$

Standard Deviation: $$\frac{4s}{\sqrt{16(r-1)-\text{Skewness}^2(r-2)^2}}$$

$$\text{where } r = \frac{6(\text{Kurtosis - Skewness}^2\text{ - 1})}{2\text{Kurtosis - 3Skewness}^2\text{ - 6}}$$

Mean: $$\varepsilon - \frac{v\,s}{r} = \varepsilon + \text{Standard Deviation} \times \frac{\text{Skewness}(r\text{ - }2)}{4}$$

Density Function – Type V: The density function of the Type V Pearson distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|\varepsilon,\gamma,p)=\begin{bmatrix} 0 & x\le\varepsilon \\ \dfrac{1}{\gamma^{1-p}\,\Gamma(p-1)}(x-\varepsilon)^{-p}e^{-\frac{\gamma}{x-\varepsilon}} & x>\varepsilon \end{bmatrix}$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	γ	$\gamma > 0$
	Shape:	p	$p > 1$

Bounds: Bounded below by ε.

Moments – Type V: The moments of the Type V Pearson distribution can be calculated from the parameters as shown below:

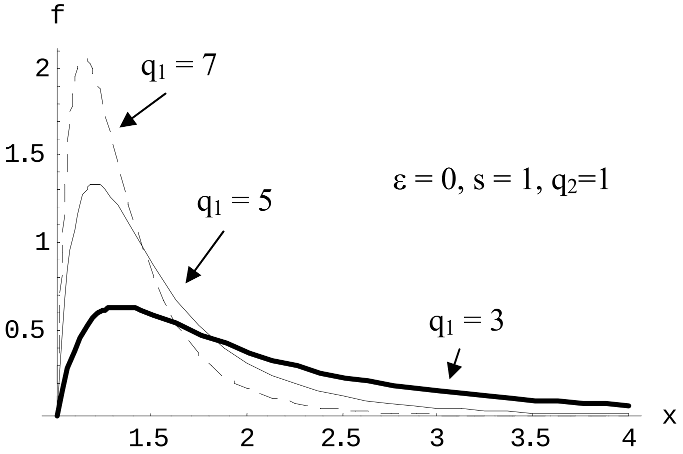
Mean: $\varepsilon+\dfrac{\gamma}{p-2}$ assuming $p > 2$

Standard Deviation: $\dfrac{\gamma}{(p-2)\sqrt{(p-3)}}$ assuming $p > 3$

Skewness: $\dfrac{4\sqrt{p-3}}{p-4}$ assuming $p > 4$, which is always > 0

Kurtosis: $\dfrac{3(p+4)(p-3)}{(p-4)(p-5)}$ assuming $p > 5$

Density Function – Type VI: The density function of the Type VI Pearson distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f\left(x \middle| \varepsilon, s, q_2, q_1\right) = \text{Constant}\left(\frac{x-\varepsilon}{s}-1\right)^{q_2}\left(\frac{x-\varepsilon}{s}\right)^{-q_1} \quad \text{for } x > s+\varepsilon \text{ if } s > 0$$

$$\text{for } x < s+\varepsilon \text{ if } s < 0$$

Parameters:	Location:	ε	$-\infty < \varepsilon < \infty$
	Scale:	s	$s \neq 0$
	Shape:	q_2	$-\infty < q_2 < q_1 < \infty$
	Shape:	q_1	

Bounds: If $s > 0$ lower bound $= s + \varepsilon$
If $s < 0$, upper bound $= s + \varepsilon$

Moments – Type VI: The moments of the Type VI Pearson distribution can be calculated from the parameters as shown below:

Mean: $$\frac{s(q_1-1)}{q_1-q_2-2}+\varepsilon$$

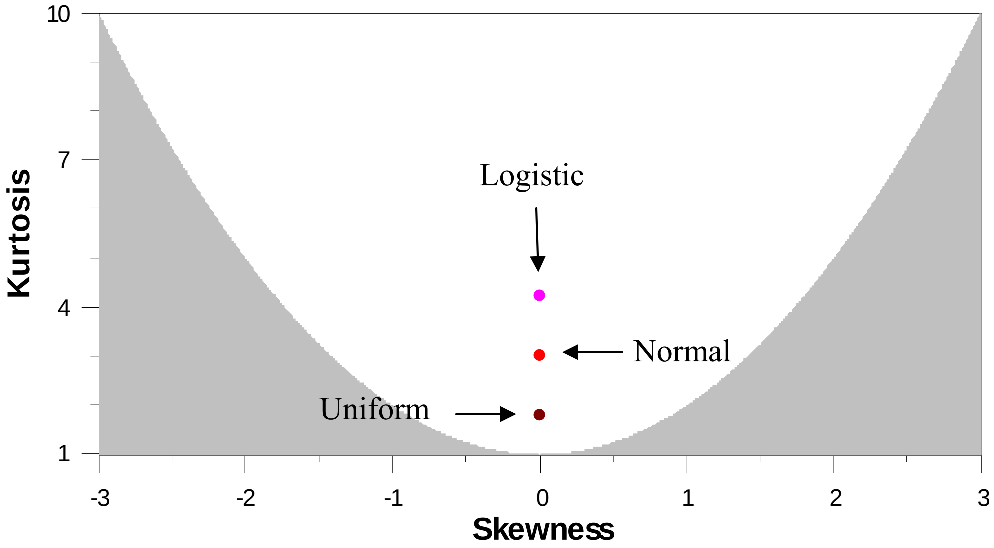
Standard Deviation: $$\sqrt{\frac{s^2(q_1-1)(q_2+1)}{(q_1-q_2-3)(q_1-q_2-2)^2}}$$

Skewness: $$\text{Sign}[s]\left|\frac{\dfrac{2(q_1+q_2)}{(q_1-q_2-4)}}{\sqrt{\dfrac{(q_1-1)(q_2+1)}{(q_1-q_2-3)}}}\right|$$

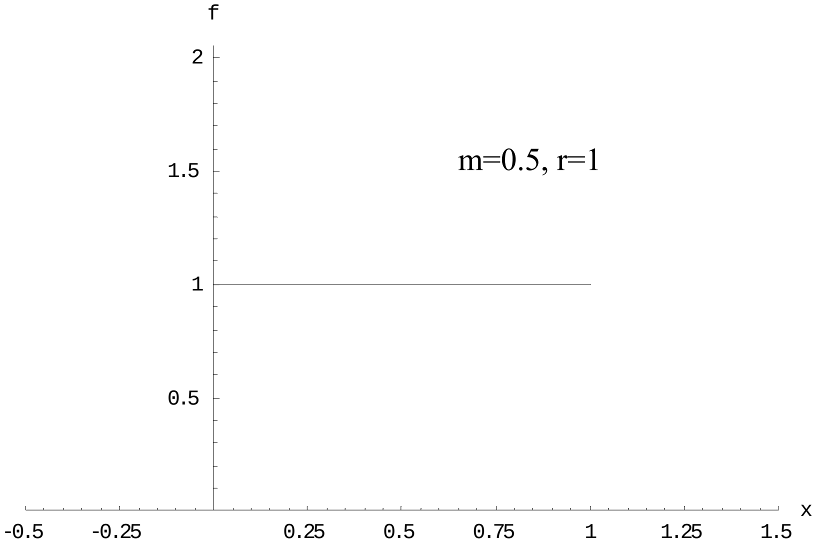
Kurtosis: $$\frac{3(q_1-q_2-3)(4-5q_1+3q_1^2+5q_2-2q_1q_2+q_1^2q_2+3q_2^2-q_1q_2^2)}{(q_1-1)(q_2+1)(q_1-q_2-4)(q_1-q_2-5)}$$

5.12 Uniform Distribution

Shape: The uniform distribution is a 2-parameter distribution and covers any specified average and standard deviation. It is represented by a single point with a skewness of zero and kurtosis of 1.8 (excess kurtosis of -1.2) on a skewness-kurtosis plot as shown below:



Density Function: The density function of the uniform distribution is shown below:



The equation, parameters and bounds of the density function are:

$$f(x|m,r)=\begin{cases}\frac{1}{r} & m-\frac{r}{2}\le x\le m+\frac{r}{2}\\ 0 & \text{otherwise}\end{cases}$$

Parameters:	Location:	m	$-\infty < m < \infty$
	Scale:	r	$r > 0$.

Bounds: Bounded below by m-r/2 and above by m+r/2

Moments: The moments of the uniform distribution can be calculated from the parameters as shown below:

Mean:	m
Standard Deviation:	$\frac{r}{\sqrt{12}}$
Skewness:	0
Kurtosis:	1.8

Properties:

- The uniform distribution is a special case of the beta distribution.

Glossary

Alpha Level

See p-value.

Analysis of Variance (ANOVA)

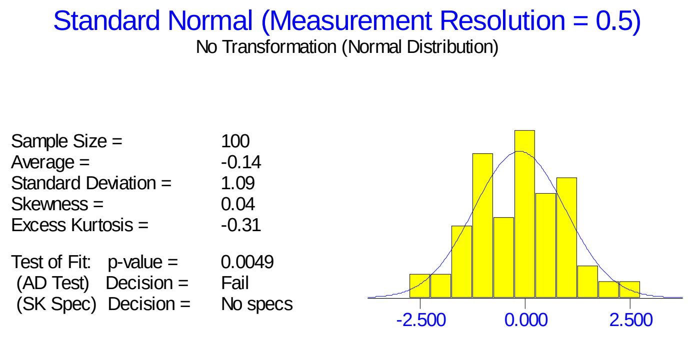
An Analysis of Variance (ANOVA) is performed to test if the averages of the groups are different. This analysis assumes the data within the groups fits the normal distribution and that the standard deviations are equal. The p-value and confidence level are reported. A significant difference is reported if the p-value is less than or equal to 0.05.

Anderson-Darling Normality Test

The Anderson-Darling test for normality is one of three general normality tests designed to detect all departures from normality. While it is sometimes touted as the most powerful test, no one test is best against all alternatives and the other 2 tests are of comparable power. The p-values given by Distribution Analyzer for this test may differ slightly from those given in other software packages as they have been corrected to be accurate to 3 significant digits.

The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05. Failing the normality test allows you to state with 95% confidence the data does not fit the normal distribution. Passing the normality test only allows you to state no significant departure from normality was found.

The Anderson-Darling test, while having excellent theoretical properties, has a serious flaw when applied to real world data. The Anderson-Darling test is severely affected by ties in the data due to poor precision. When a significant number of ties exist, the Anderson-Darling will frequently reject the data as non-normal, regardless of how well the data fits the normal distribution. Below is an example of data generated from the normal distribution but rounded to the nearest 0.5 to create ties. A tie is when identical values occurs more than once in the data set:



Both the Shapiro-Wilks test (p-value = 0.1311) and Skewness-Kurtosis All test (p-value = 0.9930) pass this set of data. The Shapiro-Wilks test is also affected by ties, but not nearly as bad as the Anderson-Darling test. The Skewness-Kurtosis All test is not affected by ties and thus the default test.

Average

The first moment of a distribution. The average is the measure of location indicating the center of the distribution. The formula used for estimating the average from a set of data is:

$$\frac{x_1 + x_2 + \cdots + x_n}{n} = \frac{\sum_{i=1}^{n} x_i}{n}$$

where n is the sample size and x_i represents the data points.

Beta Distribution

See Section 5.2.

Bi-modal Data

Histograms can appear to have multiple peaks (modes). Such data is called bi-modal or multi-modal. As a result, it is likely to fail a normality test.

Multi-model data generally indicates the data consists of a mixture of product. Shifts over time and differences between cavities or nozzles can both produce multi-modal data.

Bounds, Distribution

Certain distributions like the normal distribution and logistic distribution are unbounded. Values generated from these distributions range from -infinity to infinity.

Distributions like the lognormal distribution and gamma distribution have lower bounds. They range from this lower bound to infinity.

Distributions like the negative of the lognormal distribution and negative of the gamma distribution have upper bounds. They range from -infinity to this upper bound.

Finally distributions like the beta distribution have both upper and lower bounds. They range from the lower to the upper bound.

Bounds, Physical

Physical bounds for the data occur when the data is restricted to certain values. For example, the radius is bounded below by zero as only positive values can result. Yield results are bounded below by 0% and above by 100%. Speed, if the object can move in two directions is unbounded. Most real data is bounded. Bounded data can be more difficult to fit a distribution to. To aid in fitting a distribution, it is often helpful to pre-transform the

	data. Physical bounds, if they exist, can be specified along with the data in the *Data* window.
Capability Index	Capability indexes, like P_p and P_{pk}, are measures of how well the data fits within the specification limit(s) associated with Statistical Process Control (SPC). The higher the values, the better the fit. They are used by variables sampling plans to make pass/fail decisions.
Change-Point Analysis	A change-point analysis is performed to detect if the process shifted over time. It can detect multiple shifts. If it detects shifts, the shifts are listed in a table. For each shift, the estimated first point or subgroup following the change is listed along with the confidence level representing the confidence that the shift occurred. Only shifts detected with 95% confidence or better are listed.
Child Window	A sizable window displayed in the interior of the main or parent window. Distribution Analyzer has a permanent *Data* window and allows multiple *Test Distribution* and *Skewness-Kurtosis Plot* windows, which are all child windows.
Clipboard	A common resource used by all Window programs for sharing information. Copying and cutting information from a program places this information into the clipboard. Pasting retrieves information from the clipboard. Information can be placed into the clipboard in many formats. Distribution Analyzer places data in the clipboard in tab delimited format so that a wide range of programs can access this data like EXCEL. Plots and tables from the *Test Distribution* window are place into the clipboard in Windows Meta file (picture) format. They can be pasted into Word. In order to edit them in Word, they must first be converted. To do this, right mouse click over the graphic and select the *Edit Picture* menu item.
Confidence Level	Confidence levels are associated with normal tolerance intervals, variables sampling plans, tests of fit, tests of group differences and tests if there was a shift over time (order). The confidence statement can be thought of as representing the probability the statement or conclusion is correct. By default, confidence levels of 95% are used. You can adjust the confidence level used for normal tolerance intervals and variables sampling plans using the *Analysis Options* dialog box or *Tolerance Interval Options* dialog box.

For tests of fit (normality tests), the confidence level is based on the significant level, alpha-level or p-value. The p-value is the probability that the data or one more extreme than it would have been generated had the data come from the selected distribution. A p-value of 0.05 would indicate that the chance of the observed data is low, 1 in 20, due to variation alone. This is good evidence that the data is not from the selected distribution. A p-value of 0.5 would indicate that there is a 50-50 chance of the something as extreme as the observed data assuming the selected distribution. This is consistent with the selected distribution. The smaller the p-value, the greater the evidence that the data did not come from the selected distribution.

For tests of fit and other tests, the confidence level is calculated from the p-value as 100*(1 - p-value). Therefore:

Confidence Level	p-value
99%	0.01
95%	0.05
90%	0.1
80%	0.2
50%	0.5

Confidence Statement Relative to Spec Limits

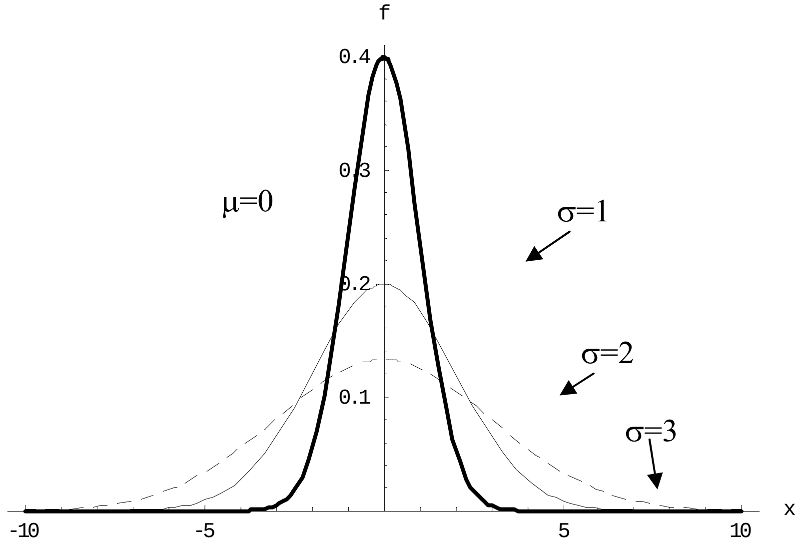
A confidence statement relative to spec limits is statement like: "With 95% confidence more than 99.87% are in spec." Such a statement results from passing a variables sampling plan. Based on the P_p and P_{pk} values, a variables sampling plan is identified that would pass and the resulting confidence statement constructed. Confidence statements relative to spec limits are displayed in Tabs 1 and 2 of the *Test Distribution* window.

Confidence statements relative to spec limits assume the underlying data fits the normal distribution. Before using one, you should pass a normality test. By default, the confidence statement relative to the spec limits is only displayed if one of the normality tests passes. For data that fits some other distribution, the data is first transformed along with the spec limits. P_p, P_{pk} and the confidence statement relative to the spec limits are then calculated using the transformed values.

An alternate approach is to use normal tolerance intervals.

Density Function

A distribution is commonly described in terms of the density function. Density functions for the normal distribution are shown below. They are essentially smoothed out histograms (where the width of the bars go to zero).



The mathematical form of the density function is:

$$f\left(x\middle|\mu,\sigma\right)=\frac{1}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad \text{for } -\infty < x < \infty$$

Parameters:	Location:	μ	$-\infty < \mu < \infty$
	Scale:	σ	$\sigma > 0$

Probabilities are obtained from the density function as areas under the curve. The probability that a data point is between 1 and 2 is the area under the density function between 1 and 2. For all density functions, the total area under the curve is one.

Distribution

A distribution is a model for how a set of data is suppose to behave. If describes this behavior in terms of probabilities. From this model it can be determined the probability a value is below a certain value or the probability a value is within a certain range. This model is described in terms of the either the density function or the distribution function of the distribution. These two forms are interchangeable as one can be determined based on the other.

Different physical phenomena produce different distributions. When rolling a 6-sided die, one gets 1 to 6 with equal probability. If one tosses a coin 10 times and counts the number of heads, one gets a number from 0 to 10 following a distribution called the binomial distribution. Potentially every physical could produce a unique distribution. Fortunately, many data sets fit the normal distribution due to it being the limiting distribution of addition and subtraction. Another common distribution is the lognormal distributions due to it being the limiting distribution of multiplication and division. For minimum and maximums there are the smallest extreme value family of distributions (including the Weibull distribution) and largest extreme value family of distributions. However, there are many other distributions that might fit your data. The distributions and family of distributions included in Distribution Analyzer are:

Beta Distribution
Exponential Distribution
Negative Exponential Distribution
Gamma Family of Distributions
Johnson Family of Distributions
Largest Extreme Value Family of Distributions
Loglogistic Family of Distributions
Lognormal Family of Distributions
Normal Distribution
Pearson Family of Distributions
Smallest Extreme Value Family of Distributions
Uniform Distribution

For every distribution there is a transformation that makes data from that distribution fit the normal distribution. Identifying the distribution that fits the data is identical to identifying the transformation that makes the data fit the normal distribution.

Distribution Function

The distribution function F(x|parameters) is the probability of a value less than x. It can be calculated from the density function f(y|parameters) as follows:

$$F(x|\text{parameters}) = \int_{-\infty}^{x} f(y|\text{parameters})\,dy$$

The equation is based on integrals. In plain English the distribution function F(x|parameters) is the area under the density function f(y|parameters) for y ranging from -infinity to x.

Excess Kurtosis

The excess kurtosis = kurtosis - 3. This results in the normal distribution having an excess kurtosis of zero. An excess kurtosis above 0 indicates the tails are heavier than the normal distribution. An excess kurtosis below 0 indicates the tails are lighter than the normal distribution. An excess kurtosis value of 1 and above or -1 and below represents a sizable departure from normality. The formula used for estimating the excess kurtosis from a set of data is:

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^4}{S^4} - \frac{3(n-1)^2}{(n-2)(n-3)}$$

where n is the sample size, x_i represents the data points, $\overline{X}$ is the average and S is the standard deviation.

Exponential Distribution	See Section 5.3.
Extreme Value, Largest Family	See Section 5.4.
Extreme Value, Smallest Family	See Section 5.5.
Family of Distributions	A family of distributions is several distributions combined so that they cover a well define region in a skewness-kurtosis plot. For example, the lognormal family of distributions includes the lognormal, negative lognormal and normal distributions. This allows the family to fit all possible average, standard deviation and skewness values. It appears as a curve in the skewness-kurtosis plot. The lognormal, negative lognormal and normal distributions are distinct distributions because they have different density and distribution functions. There are two families that allow one to fit all possible average, standard deviation, skewness and kurtosis values (excluding the impossible region): Johnson, combining 3 distributions, and Pearson, combining 7 distributions.
Fréchet Distribution	See Section 5.4.
Gamma Distribution	See Section 5.6.
Groups	Groups are categories that the data can be divided into. Examples are: operator producing or testing the unit, line the unit was manufactured on and the cavity that the unit was made in. When the values fall into to such groups there is the possibility that the groups are different. If each group fits the normal distribution, but if some groups are shifted left or right relative to other groups, the resulting histogram can appear to have multiple peaks (modes). Such data is called bi-modal or multi-modal. As a result, it is likely to fail a normality test. Such multi-modal data may pass the Skewness-Kurtosis Specific test and you may be able to proceed with a variables sampling plan or normal tolerance interval. However, it may sometimes be necessary to divide the data into the groups and test each group separately. For example, in validating an injection molding process, the die used may have 4 cavities for forming 4 parts at a time. If the cavities are different, it may be necessary to validate each cavity separately. Before the data is split into subgroups,

there should be statistical evidence the groups are different. The group each value is from should be entered in the *Group* column of the *Data* window. Then a comparison of the groups is automatically performed and displayed in Tab 6 of the *Test Distribution* window.

The difference between groups and ordered data is that groups are categories that are unordered whereas ordered data consists of subgroups that are ordered relative to time. Both can produce multi-modal data.

Impossible Region

The impossible region in the skewness-kurtosis plot represents values for the skewness and kurtosis that cannot occur. For all distributions, kurtosis ≥ $\text{skewness}^2 + 1$. This implies kurtosis ≥ 1. The border of the impossible region, where kurtosis = $\text{skewness}^2 + 1$, corresponds to the 2-point distribution. The 2-point distribution takes on 2 values, x and y, where x occurs with probability p and y occurs with probability 1-p.

Johnson Family of Distributions

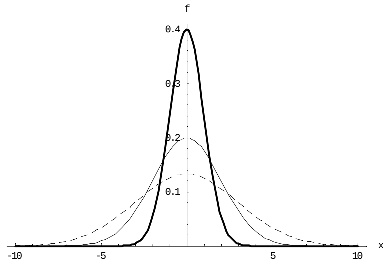
See Section 5.7.

Kruskall-Wallis Test

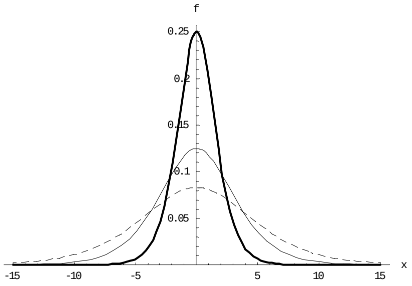
The Kruskall-Wallis test checks to see if the medians of the groups are different. This is a nonparametric test that makes no assumptions about the distributions of the groups. It is an alternative to the an ANOVA.

Kurtosis

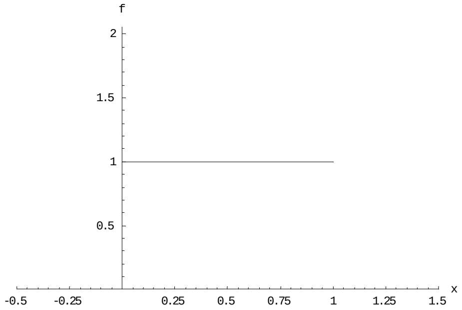
The fourth moment of a distribution and the second shape parameter. The kurtosis is a measure of the heavy the tails are. The normal distribution, shown below, has a kurtosis of 3:



A kurtosis greater than 3 means the tails are heavier than the normal distribution. In order to have more units in the extreme tails also means there must also be more units near the middle of the distribution. Such distributions appear to have a very high peak in the middle with wide plateaus for tail. An example is the logistic distribution with a kurtosis of 4.2 shown below:



A kurtosis less than 3 means the tails are lighter than the normal distribution like the Uniform distribution with a kurtosis of 1.8 shown below:



A kurtosis value of 4 and above or 2 and below represents a sizable departure from normality. The formula used for estimating the kurtosis from a set of data is:

$$\frac{n(n+1)}{(n-1)(n-2)(n-3)} \quad \frac{\sum_{i=1}^{n}\left(x_i - \overline{X}\right)^4}{S^4} - \frac{3(3n-5)}{(n-2)(n-3)}$$

where n is the sample size, x_i represents the data points, $\overline{X}$ is the average and S is the standard deviation.

The excess kurtosis = kurtosis - 3. This results in the normal distribution having an excess kurtosis of zero.

Largest Extreme Value Family of Distributions	See Section 5.4.
Levene's Test	Levene's test checks to see if the standard deviations of the groups are different. This analysis assumes the data within the groups fits the normal distribution.
Logistic Distribution	See Section 5.8.
Loglogistic Family of Distributions	See Section 5.8.

Lognormal Family of Distributions

See Section 5.9.

Maximum Likelihood Method

An approach to fitting a distribution to a set of data that involves maximizing the likelihood function. The likelihood function is based on the density function of the distribution. If f(x|parameters) is the density function, then the likelihood function is:

$$f\left(x_1 \middle| \text{parameters}\right) \times f\left(x_2 \middle| \text{parameters}\right) \times \cdots \times f\left(x_n \middle| \text{parameters}\right)$$

where n is the sample size and x_i represents the data points. The likelihood function is akin to finding the values of the parameters making the resulting data most likely (probable).

Using the maximum likelihood method always results in the data being within the range of the distribution. However, it is possible the spec limits are outside the range of the distribution and cannot be transformed.

This approach has been further adapted so that the user can specify a range of values that must be in the range of the selected distribution. This range can be specified using the *Select Distribution to Fit Data* dialog box. When the *Find Best Distribution* button is clicked in the *Data* window, the required range is automatically specified as at least 1 standard deviation beyond any spec limits. This assures the spec limits are also within the range of the distribution and can be transformed.

An alternative approach for fitting data is the method of moments method.

Menu Bar

See Chapter 4.

Method of Moments

An approach to fitting a distribution to a set of data that involves matching the moments of the data to the distribution. The number of moments that are matched depends on the number of parameters of the distribution. For the normal distribution, with 2 parameters, the average and standard deviation are matched. For the Johnson family of distributions, with 4 parameters, the average, standard deviation, skewness and kurtosis are all matched.

The method of moments approach has the disadvantage that the resulting estimates can be inconsistent with the data set. This can occur for bounded distributions. It is possible that the distribution fit to the data has data values below the lower bound or above the upper bound of the distribution. When this occurs, Distribution

Analyzer reduces the number of moments that are matched and instead matches the bounds.

For example, the Beta distribution has 4 parameters and both upper and lower bounds. Initially Distribution Analyzer attempts to match all four moments. If a value occurs below the lower bound, the program will match the lower bound, average, standard deviation and skewness. If a value occurs above the upper bound, the program will match the upper bound, average, standard deviation and skewness. If values occur both below the lower bound above the upper bound, the program will match the lower bound, upper bound, average and standard deviation. As a result, the data will fall within the range of the fitted distribution. This approach has not been implemented for the Johnson distribution so it is still possible that data be outside the range of the fitted distribution.

This approach has been further adapted so that the user can specify a range of values that must be in the range of the selected distribution. This range can be specified using the *Select Distribution to Fit Data* dialog box. When the *Find Best Distribution* button is clicked in the *Data* window, the required range is automatically specified as at least 1 standard deviation beyond any spec limits. This assures the spec limits are also within the range of the distribution and can be transformed.

An alternative approach for fitting data is the maximum likelihood method.

Moments of a Distribution

The first four moments of a distribution are:

- Average - measures the location of the distribution.
- Standard Deviation - measures the width of the distribution. Referred to as the scale parameter.
- Skewness - measures the symmetry of the distribution. Referred to as the first shape parameter.
- Kurtosis - measure of the heaviness of the tails and peakedness of the center of a distribution. Referred to as the second shape parameter.

Different distributions with the same moments will have similar density functions and behave nearly identical.

Multi-Modal Data

Data that has multiple peaks in the histogram. Data with 2 peaks is often referred to as bimodal (See Bi-Modal).

Negative Exponential Distribution

See Section 5.3.

Negative of a Distribution

The negative of a distribution is the distribution of -X (negative of the values) when X varies according to the distribution. For example, if the value X follows the lognormal distribution, then -X follows the negative lognormal distribution. For symmetrical distributions like the normal distribution, the negative of the distribution is the same as the original distribution.

The moments of -X are:

Average of (-X) = - (Average of X)
Standard Deviation of (-X) = Standard Deviation of X
Skewness of (-X) = - (Skewness of X)
Kurtosis of (-X) = Kurtosis of X

The average and skewness change sign while the standard deviation and kurtosis are not affected.

Normal Distribution

See Section 5.10.

Normal Tolerance Interval

A normal tolerance interval is a statistical procedure for constructing an interval like: "With 95% confidence, 99% of the values fall between 1.32 and 1.43." Such an interval is called a 2-sided tolerance interval. One-sided intervals can also be constructed like: "With 95% confidence, 99% of the values fall below 1.45" (upper tolerance interval) or "With 95% confidence, 99% of the values fall above 1.30" (lower tolerance interval). Normal tolerance intervals for the data are displayed in Tabs 1 and 2 of the *Test Distribution* window.

To construct a normal tolerance interval you must specify: confidence level, percent in interval and whether to use a 2-sided, upper or lower tolerance interval. These options can be set using the *Analysis Options* and *Tolerance Interval Options* dialog boxes.

Normal tolerance intervals assume the underlying data fits the normal distribution. Before using one, you should pass a normality test. By default, the normal tolerance interval is only displayed if one of the normality tests passes. For data that fits some other distribution, the data is first transformed and the transformed values are used to construct a normal tolerance interval. This interval is then transformed back to the original units of measures.

It is commonly desired to make a confidence statement like: "With 95% confidence, 99% of the values are in spec." One way of accomplishing this goal is to construct a normal tolerance interval. If this interval falls inside the specs, then the same confidence statement can be made relative to the spec limits. This is a valid approach. A similar, but slightly more powerful approach is to use a variables sampling plan.

Normality Test

See Test of Fit.

Ordered Data

Ordered data is when data is collected over time. It may be that 3 samples are collected at 10 points in time. The groups of 3 samples are referred to as subgroups. It may be that the underlying data fits the normal distribution, but if a shift occurs in the middle of collecting the data, the resulting histogram can appear to have multiple peaks (modes). Such data is called bi-modal or multi-modal. As a result, it is likely to fail a normality test.

Such multi-modal data may pass the Skewness-Kurtosis Specific test and you may be able to proceed with a variables sampling plan or normal tolerance interval. To determine if a shift occurred, the subgroup each value is from should be entered in the *Order* column of the *Data* window. Then an analysis for shifts over time is automatically performed and displayed in Tab 4 of the *Test Distribution* window.

The difference between groups and ordered data is that groups are categories that are unordered whereas ordered data consists of subgroups that are ordered relative to time. Both can produce multi-modal data.

Outlier

A true outlier is a point that is not from the same distribution as the other values, but instead something happened to it (typically an error) to make it different than the other values. For example, consider a filling operation where bags are filled with a solution. Bags are consistently in the 50 to 55 mL range. However, occasionally a half filled bag is found (15-35 mL). An investigation into the cause identified these bags were being removed from the filling nozzle before the cycle was completed. As a result, some the solution missed going into the bag. An outlier can be the result of either a manufacturing error or a measurement error.

It can be difficult to distinguish between an outlier relative to the normal distribution or an extreme value out in the tail of a long tailed distribution. For a data point to be considered an outlier relative to the normal distribution, it must generally be at least 4.5 standard deviations from the average. In Tab 7: Outliers of the *Test Distribution* window values are flagged as definite outliers if they are 10 or more standard deviations from the average. Between 4.5 and 10 standard deviations they are flagged as either outliers from the normal distributions or extreme values from a long tailed distribution.

p-value

For significance test including tests of fit there is a hypothesized condition (called null hypothesis or H0) that one is testing to see if it is true. For a test of fit the hypothesized condition is that the selected distribution generated the data. For a test that the means are equal, the hypothesized condition is equal means. The p-value is then probability that the data or one more extreme than it would have been generated under the hypothesized condition. A p-value of 0.05 would indicate that the chance of the observed data is low, 1 in 20, due to variation alone. This is good evidence that the data was not generated under the hypothesized condition. The hypothesized condition is rejected if the p-value is 0.05 or below. This provides 95% confidence the hypothesized condition is not true, i.e., the data does not fit the selected distribution or the means are not the equal.

The smaller the p-value, the greater the evidence that the data did not come from the selected distribution. For tests of fit and other tests, the confidence level is calculated from the p-value as 100*(1 - p-value). Therefore:

Confidence Level	p-value
99%	0.01
95%	0.05
90%	0.1

The p-value is also known as alpha level or significance level.

Parameters of a Distribution

The parameters of a distribution are variables included in the density function so that the distribution can be adapted to a variety of situations. Of greatest importance is the number of parameters as shown below:

2 Parameters: The two parameters determine the average and standard deviation of the distribution. Such distributions are represented as a point on a skewness-kurtosis plot as they have fixed values of the skewness and kurtosis. Examples are the exponential, normal and uniform distributions.

3 Parameters: The three parameters determine the average, standard deviation and skewness of the distribution. Such distributions are represented as a curve on a skewness-kurtosis plot as the kurtosis depends of the skewness. Examples are the gamma and lognormal distributions.

4 Parameters: The four parameters determine the average, standard deviation, skewness and kurtosis of the distribution. Such distributions are represented as a region on a skewness-kurtosis plot as they can take on a variety of skewness and kurtosis values. Examples are the beta, Johnson and Pearson distributions.

Different books and articles will sometimes parameterize the same distribution differently and even give them different names. One set of parameters can always be calculated from the other. Further, sometimes different numbers of parameters are used so there are 2 and 3 parameter versions of the lognormal distribution. This greatly complicates comparing and using distributions. For this reason, Distribution Analyzer re-parameterizes all the distributions in terms of the average, standard deviation, skewness and kurtosis as needed for use in the *Select Distribution to Fit Data* and *Select/View Distribution* dialog boxes. Further, the distributions are expanded to always include the average and standard deviation as parameters. There are tabs in both dialog boxes that display the traditional parameters as well.

Pearson Family of Distributions

See Section 5.11.

P_p

Capability index that compares process variation to the width of the spec limits:

$$P_p = \frac{(\text{Upper Spec Limit} - \text{Lower Spec Limit})}{\text{6 Standard Deviations}}$$

The numerator is the width of the spec limits. The denominator is 6 standard deviations, which can be thought of as the width of the process. For the normal distribution 99.7% of values fall within ±3 standard deviations of the average or into an interval 6 standard deviations wide. A P_p value of 1 means the process variation fills the spec limits. A P_p of 2 means the specs are twice as wide as the process. The larger P_p is, the better the capability.

P_p is similar to the capability index C_p. They are both calculated using the formula above. However, they use different methods for estimating the standard deviation. P_p uses the total standard deviation. C_p requires that the units be collected in subgroups over time and uses the within subgroup standard deviation, ignoring the effects of shifts between subgroups. P_p estimates actual performance, while C_p estimates the capability the process could achieve if made stable over time.

The acceptance criteria for a variable sampling plan can be stated in terms of the capability indexes P_p and P_{pk}.

Note: P_p does not depend on the average. Further P_p can only be calculated when there are both upper and lower spec limits.

P_{pk} Capability index that measure the relative distance to the nearest spec limit:

$$P_{pk} = \text{Minimum}\left(\frac{\text{Upper Spec Limit} - \text{Average}}{\text{3 Standard Deviations}}, \frac{\text{Average - Lower Spec Limit}}{\text{3 Standard Deviations}}\right)$$

The numerator is the distance to the nearest spec limit. For a one-sided spec limit (lower spec only or upper spec only) only use the portion of the formula for that spec limit. The denominator is 3 standard deviations, which can be thought of as the half the width of the process. For the normal distribution 99.7% of values fall within ±3 standard deviations of the average or into an interval 6 standard deviations wide. A P_{pk} value of 1 means the distance between the average and the nearest spec limit is 3 standard deviations and thus the process fills this interval and touches the spec limit. A P_{pk} value of 2 means the distance between the average and the nearest spec limit is 6 standard deviations and thus the process fills only half this interval. This leaves a safety margin. The larger P_{pk} is, the better the capability.

P_{pk} is similar to the capability index C_{pk}. They are both calculated using the formula above. However, they use different methods for estimating the standard deviation. P_{pk} uses the total standard deviation. C_{pk} requires that the units be collected in subgroups over time and uses the within subgroup standard deviation, ignoring the effects of shifts between subgroups. P_{pk} estimates actual performance, while C_{pk} estimates the capability the process could achieve if made stable over time.

The acceptance criteria for a variable sampling plan can be stated in terms of the capability indexes P_p and P_{pk}.

Note: P_{pk} depends on both the average and standard deviation. Further P_{pk} can be calculated when there is any combination of spec limits: lower only, upper only and both upper and lower spec limits.

Precision

The precision of the data is the difference between reportable values. For example, if the data consists of values like 2.1, 2.2, and 2.3, the precision is 0.1.

Pre-Transforming Data

To aid in fitting a distribution, it is often helpful to pre-transform data that has physical bounds. Pre-transforming data takes data that is bounded and converts it to unbounded data. Bounds, if they exist, can be specified along with the data in the *Data* window.

When a lower bound (LB) is specified, the data can be pre-transformed using the transformation:

$$\mathrm{Ln}(\mathrm{X} - \mathrm{LB})$$

This transforms the range from LB to infinity to -infinity to infinity. A value equal to LB cannot be transformed by this equation. Such a value is replaced with LB + Precision/4 before performing the pre-transform. Take as an example the values below reported to the nearest 0.1 units (precision) and the lower bound is zero. Then the values correspond to:

Value	Interval	Midpoint
0	[0, 0.05]	0.025
0.1	[0.05, 0.15]	0.1
0.2	[0.15, 0.25]	0.2
0.3	[0.25, 0.35]	0.3

This is why 0.025 = precision/4 is used in place of zero when performing the pre-transform.

When an upper bound (UB) is specified, the data can be pre-transformed using the transformation:

$$-\mathrm{Ln}(\mathrm{UB} - \mathrm{X})$$

This transforms the range from -infinity to UB to -infinity to infinity. A value equal to UB cannot be transformed by this equation. Such a value is replaced with UB - Precision/4 before performing the pre-transform.

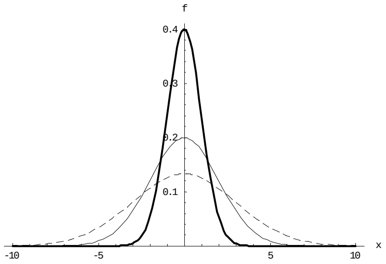
When both a lower bound (LB) and upper bound (UB) are specified, the data can be pre-transformed using the transformation:

$$\mathrm{Ln}\left(\frac{\mathrm{X} - \mathrm{LB}}{\mathrm{UB} - \mathrm{X}}\right)$$

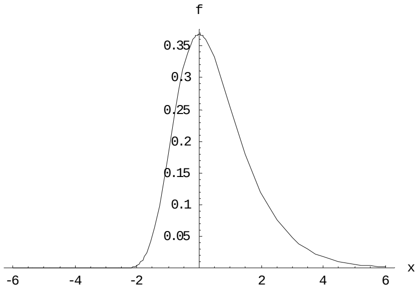
This transforms the range from LB to UB to -infinity to infinity.

A value equal to LB cannot be transformed by this equation. Such a value is replaced with LB + Precision/4 before performing the pre-transform. A value also equal to UB cannot be transformed by this equation. Such a value is replaced with UB - Precision/4 before performing the pre-transform.

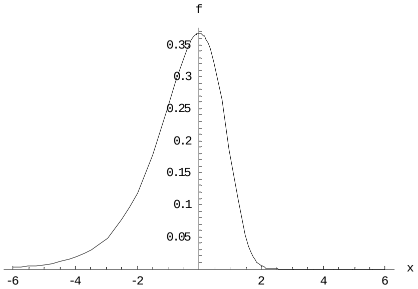
Range, Distribution	See Bounds, Distribution.
Reliability	When used in the content of a variable sampling plan or normal tolerance interval, reliability refers to the percent in spec. 99% reliable means 99% are in spec or, equivalently, 1% is out of spec. This use of the word reliability is different than reliability in terms of mean time to failure of a device or equipment used for an extended period of time.
Shape Parameters	The skewness and kurtosis are commonly referred to as the shape parameters. Skewness measures symmetry and kurtosis measures the heaviness of the tails. This is in contrast to the average (location parameter) and standard deviation (scale parameter).
Shapiro-Wilks Normality Test	The Shapiro-Wilks test for normality is one of three general normality tests designed to detect all departures from normality. It is comparable in power to the other two tests. The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05. Failing the normality test allows you to state with 95% confidence the data does not fit the normal distribution. Passing the normality test only allows you to state no significant departure from normality was found. The Shapiro-Wilks test is not as affected by ties as the Anderson-Darling test, but is still affected. The Skewness-Kurtosis All test is not affected by ties and thus the default test.
Significance Level	See p-value.
Skewness	The third moment of a distribution and the first shape parameter. The skewness is measure of the symmetry of the distribution. A skewness of zero means the distribution is symmetrical like the normal distribution shown below:



A positive skewness means the upper tail is longer than the lower tail like the Largest Extreme Value distribution with a skewness of 1.14 shown below:



A negative skewness means the lower tail is longer than the upper tail like the Smallest Extreme Value distribution with a skewness of -1.14 shown below:



A skewness value of 1 and above or -1 and below represents a sizable departure from normality. The formula used for estimating the skewness from a set of data is:

$$\frac{n}{(n-1)(n-2)} \quad \frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^3}{S^3}$$

where n is the sample size, x_i represents the data points, $\overline{X}$ is the average and S is the standard deviation.

Skewness-Kurtosis All Normality Test (All Departures From Normality)

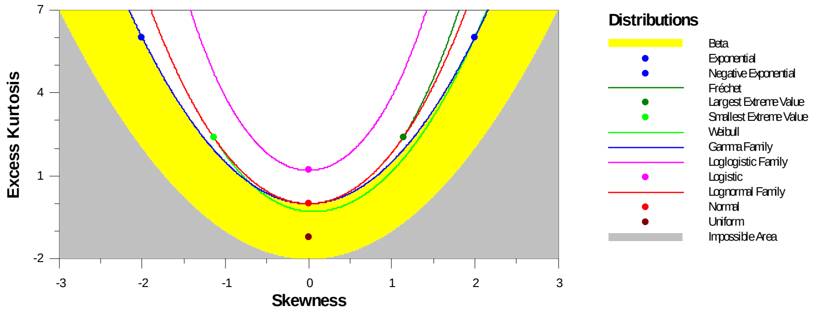
The Skewness-Kurtosis All test for normality is one of three general normality tests designed to detect all departures from normality. It is comparable in power to the other two tests. The normal distribution has a skewness of zero and kurtosis of three. The test is based on the difference between the data's skewness and zero and the data's kurtosis and three.

The test rejects the hypothesis of normality when the p-value is less than or equal to 0.05. Failing the normality test allows you to state with 95% confidence the data does not fit the normal distribution. Passing the normality test only allows you to state no significant departure from normality was found.

It is the default test because it is not affected by ties like both the Anderson-Darling and Shapiro-Wilks tests.

Skewness-Kurtosis Plot

A skewness-kurtosis plot indicates the range of skewness and kurtosis values a distribution can fit. An example is shown below:



Two-parameter distributions like the normal distribution are represented by a single point. Three parameters distributions like the lognormal distribution are represented by a curve. Four parameter distributions like the beta distribution are represented by a shaded region. At the bottom of the plot is a gray shaded region called the impossible region. No distributions can fall into this region.

Skewness-Kurtosis Specific Test (Heavy Tails Towards Spec)

The Skewness-Kurtosis Specific test for normality is not a general test for normality designed to detect all departures for normality like the other tests. This test is designed to only reject specific departures from normality that invalidate the confidence statements associated with variables sampling plans and normal tolerance intervals. It is designed to answer the question: "Is it OK to use a variables sampling plan or normal tolerance interval?" Passing this test is sufficient justification to use a variables sampling plan or normal tolerance interval, even if the other three tests fail.

The confidence statements are valid so long as the tails of the distribution are no larger than the tails of the normal distribution. This means the normal distribution bounds the tails of the distribution rather than exactly fit it. This test rejects when the tails are longer than the normal distribution, which occurs when there is a positive kurtosis or skewness in the direction of a spec limit. Depending on the type of specification involved, the Skewness-Kurtosis Specific test rejects when:

Lower Spec Limit Only: Rejects if negative skewness or a positive kurtosis.

Upper Spec Limit Only: Rejects if positive skewness or a positive kurtosis.

Two-Sided Spec Limit Only: Rejects if either positive or negative skewness or a positive kurtosis.

The Skewness-Kurtosis test does not give a p-value but instead just indicates pass/fail. If you fail, you can state with 95% confidence the data is not from the normal distribution as before.

The risk of proceeding if you pass the Skewness-Kurtosis Specific test is that of a false rejection. The tails may be less than the normal distribution. The confidence statements associated with the variables sampling plan or normal tolerance interval may overestimate the defect rate or range of the distribution resulting in failing a set of data that deserves to pass. If the Skewness-Kurtosis Specific test passes, but the other normality test fails, consider going ahead and transforming the data to avoid a false rejection. This will result in more accurate statements. However, it the Skewness-Kurtosis Specific test passes and the confidence statements meet the acceptance criteria, then the data clearly passes making a transformation unnecessary.

Smallest Extreme Value Family of Distributions

See Section 5.5.

Standard Deviation

The second moment of a distribution. The standard deviation is the measure of scale indicating the width of the distribution. The formula used for estimating the standard deviation from a set of data is:

$$\sqrt{\frac{\sum_{i=1}^{n}\left(x_i-\overline{X}\right)^2}{n-1}}$$

where n is the sample size, x_i represents the data points and $\overline{X}$ is the average.

Test of Fit

To test whether a distribution fits a set of data, Distribution Analyzer uses the transformation to normality associated with that distribution and transforms the data. It then applies a normality test to see if the transformed values fit the normal distribution. Distribution analyzer includes three normality tests:

Anderson-Darling test (AD)
Shapiro-Wilks test (SW)
Skewness-Kurtosis All test (SK All)

The above three tests are designed to detect all departures form

	normality. They answer the question: "Does the data fit the normal distribution?" A fourth specialized test is: Skewness-Kurtosis Specific test (SK Spec) This test is designed to only reject for specific departures that invalidate the confidence statements associated with variables sampling plans and normal tolerance intervals. It is designed to answer the question: "Is it OK to use a variables sampling plan or normal tolerance interval?" Passing this test is sufficient justification to use a variables sampling plan or normal tolerance interval, even if the other three tests fail.
Tolerance Interval	See Normal Tolerance Interval.
Toolbar	See Section 4.6.
Transformation	A transformation is a function applied to the data points before analyzing them. The most common transformation is the log transformation. If the original data is from the lognormal distribution, taking the log of the values will cause them to fit the normal distribution. For every distribution, there is a transformation that makes data from that distribution fit the normal distribution. Identifying the distribution that fits the data is identical to identifying the transformation that makes the data fit the normal distribution.
Tukey-Kramer Multiple Comparison	A Tukey-Kramer Multiple Comparison is used to determine which groups are significantly different.
Uniform Distribution	See Section 5.12.
Variables Sampling Plan	A variables sampling plan is a statistical procedure for making a pass/fail decision. When the sampling plan passes, there is an associated confidence statement relative to the spec limits that can be made like: "With 95% confidence, 99% of the values are in spec." Variables sampling plans make the pass/fail decision based on the capability indexes P_p and P_{pk}. One has two options for using variables sampling plans. First, you can use tables of variables sampling plans to determine the acceptance criteria for P_p and P_pk. Distribution Analyzer then calculates and displays these capability indexes, allowing a pass/fail decision to be made. Second, you can state an

acceptance criteria like "With 95% confidence more than 99% of values must be in spec." Then use Distribution Analyzer to construct the confidence statement relative to the spec limits to see if the study passes.

Variables sampling plans assume the underlying data fits the normal distribution. Before using one, you should pass a normality test. By default, the confidence statement relative to the spec limits is only displayed if one of the normality tests passes. For data that fits some other distribution, the data is first transformed along with the spec limits. P_p, P_{pk} and the confidence statement relative to the spec limits are then calculated using the transformed values.

Weibull Distribution See Section 5.5.

z-score The z-score associated with a value is:

$$\frac{\text{Value} - \text{Average}}{\text{Standard Deviation}}$$

A z-score of 10 means the value is 10 standard deviations from the average. A point is marked as likely being an outlier if it is more than 10 standard deviations from the average (z-score greater than 10 or less than -10).

References

D'Agostino, Ralph B. and Stephens, Michael A. (1986). *Goodness of Fit Techniques*. Marcel Dekker, Inc., New York, NY.

Elderton, William Palin (1953), *Frequency Curves and Correlation*. Harren Press, Washington, D.C.

Johnson, Norman L. et. al. (1994). *Continuous Univariate Distributions Volumes 1 & 2 Second Edition*, John Wiley & Sons, New York, NY.

Rose, Colin and Smith, Murray D. (2001). *Mathematical Statistics with Mathematica*. Spring-Verlang, New York, NY.

Shapiro, Samuel S. (1990). *How to Test Normality and Other Distributional Assumptions*. American Society for Quality, Milwaukee, WI.

Taylor, Wayne A. (2007). *Guide to Normality Testing and Transformations.* Taylor Enterprises, Inc., Libertyville, IL.

Thode Jr., Henry C. (2002). *Testing for Normality.* Marcel Dekker Inc, New York, NY.

Index

Taylor Enterprises, Inc

5510 Fairmont Road, Suite A, Libertyville, IL 60048