

## 1. EDA

1. EDA Introduction
  1. What is EDA?
  2. How Does EDA differ from Classical Data Analysis?
    1. Model
    2. Focus
    3. Techniques
    4. Rigor
    5. Data Treatment
    6. Assumptions
  3. How Does EDA Differ from Summary Analysis?
  4. What are the EDA Goals?
  5. The Role of Graphics
  6. An EDA/Graphics Example
  7. General Problem Categories
2. EDA Assumptions
  1. Underlying Assumptions
  2. Importance
  3. Techniques for Testing Assumptions
  4. Interpretation of 4-Plot
  5. Consequences
    1. Consequences of Non-Randomness
    2. Consequences of Non-Fixed Location Parameter
    3. Consequences of Non-Fixed Variation Parameter
    4. Consequences Related to Distributional Assumptions

3. EDA Techniques
  1. Introduction
  2. Analysis Questions
  3. Graphical Techniques: Alphabetic

1. Autocorrelation Plot

The command to generate an autocorrelation plot is  
AUTOCORRELATION PLOT Y

The appearance of the autocorrelation plot can be controlled by appropriate settings of the LINE, CHARACTER, and SPIKE commands.  
Dataplot draws the following curves on the autocorrelation plot:  
The autocorrelations.  
A reference line at zero.  
A reference line at the upper 95% confidence limit.  
A reference line at the lower 95% confidence limit.  
A reference line at the upper 99% confidence limit.  
A reference line at the lower 99% confidence limit.

For example, to draw the autocorrelations as spikes, the zero reference line as a solid line, the 95% lines as dashed lines, and the 99% lines as dotted lines, enter the command

```
LINE BLANK SOLID DASH DASH DOT DOT
CHARACTER BLANK ALL
SPIKE ON OFF OFF OFF OFF OFF
SPIKE BASE 0
```

By default, the confidence bands are fixed width.  
This is appropriate for testing for white noise (i.e., randomness).  
For Box-Jenkins modeling, variable-width confidence bands are more appropriate.  
Enter the following command for variable-width confidence bands:

```
SET AUTOCORRELATION BAND BOX-JENKINS
To restore fixed-width confidence bands, enter
SET AUTOCORRELATION BAND WHITE-NOISE
```

1. Autocorrelation Plot: Random Data
2. Autocorrelation Plot: Moderate Autocorrelation
3. Autocorrelation Plot: Strong Autocorrelation and Autoregressive Model
4. Autocorrelation Plot: Sinusoidal Model

2. Bihistogram

BIHISTOGRAM Y1 Y2

As with the standard histogram, the class width, the lower class limit, and the upper class limit can be controlled with the commands  
CLASS WIDTH <value>  
CLASS LOWER <value>  
CLASS UPPER <value>

In addition, relative bihistograms, cumulative bihistograms, and relative cumulative bihistograms can be generated with the commands  
RELATIVE BIHISTOGRAM Y1 Y2  
CUMULATIVE BIHISTOGRAM Y1 Y2  
RELATIVE CUMULATIVE BIHISTOGRAM Y1 Y2

### 3. Block Plot

#### 4. Bootstrap Plot

BOOTSTRAP <STAT> PLOT Y  
where <STAT> is one of the following:

```
MEAN
MIDMEAN
MIDRANGE
MEDIAN
TRIMMED MEAN
WINSORIZED MEAN
GEOMETRIC MEAN
HARMONIC MEAN
SUM
PRODUCT
MINIMUM
MAXIMUM
STANDARD DEVIATION
VARIANCE
STANDARD DEVIATION OF MEAN
VARIANCE OF MEAN
RELATIVE STANDARD DEVIATION
RELATIVE VARIANCE
AVERAGE ABSOLUTE DEVIATION
MEDIAN ABSOLUTE DEVIATION
LOWER QUANTILE
LOWER HINGE
UPPER QUANTILE
UPPER HINGE
FIRST DECILE
SECOND DECILE
THIRD DECILE
FOURTH DECILE
FIFTH DECILE
SIXTH DECILE
SEVENTH DECILE
EIGHTH DECILE
NINTH DECILE
PERCENTILE
SKEWNESS
KURTOSIS
AUTOCORRELATION
AUTOCOVARIANCE
SINE FREQUENCY
COSINE FREQUENCY
TAGUCHI SN0
TAGUCHI SN+
TAGUCHI SN-
TAGUCHI SN00
```

The BOOTSTRAP PLOT command is almost always followed by a histogram or some other distributional plot.

Dataplot automatically stores the following internal parameters after a BOOTSTRAP PLOT command:

```
BMEAN - mean of the plotted bootstrap values
BSD - standard deviation of the plotted bootstrap values
B001 - the 0.1 percentile of the plotted bootstrap values
```

B005 - the 0.5 percentile of the plotted bootstrap values  
 B01 - the 1.0 percentile of the plotted bootstrap values  
 B025 - the 2.5 percentile of the plotted bootstrap values  
 B05 - the 5.0 percentile of the plotted bootstrap values  
 B10 - the 10 percentile of the plotted bootstrap values  
 B20 - the 20 percentile of the plotted bootstrap values  
 B80 - the 80 percentile of the plotted bootstrap values  
 B90 - the 90 percentile of the plotted bootstrap values  
 B95 - the 95 percentile of the plotted bootstrap values  
 B975 - the 97.5 percentile of the plotted bootstrap values  
 B99 - the 99 percentile of the plotted bootstrap values  
 B995 - the 99.5 percentile of the plotted bootstrap values  
 B999 - the 99.9 percentile of the plotted bootstrap values

These internal parameters are useful for generating confidence intervals and can be printed (PRINT BMEAN) or used as any user-defined parameter could (e.g., LET UCL = B95).

To specify the number of bootstrap subsamples to use, enter the command  
 BOOTSTRAP SAMPLE <N>  
 where <N> is the number of samples you want.  
 The default is 500 (it may be 100 in older implementations).  
 Dataplot can also generate bootstrap estimates for statistics that are not directly supported.  
 The following example shows a bootstrap calculation for the mean of 500 normal random numbers.  
 Although we can do this directly in Dataplot, this demonstrates the steps necessary for an unsupported statistic.  
 The subsamples are generated with a loop.  
 The BOOTSTRAP INDEX and BOOTSTRAP SAMPLE commands generate a single subsample which is stored in Y2.  
 The desired statistic is then calculated for Y2 and the result stored in an array.  
 After the loop, the array XMEAN contains the 100 mean values.

```

LET Y = NORMAL RANDOM NUMBERS FOR I = 1 1 500
LET N = SIZE Y
LOOP FOR K = 1 1 500
  LET IND = BOOTSTRAP INDEX FOR I = 1 1 N
  LET Y2 = BOOTSTRAP SAMPLE Y IND
  LET A = MEAN Y2
  LET XMEAN(K) = A
END OF LOOP
HISTOGRAM XMEAN
  
```

- Box-Cox Linearity Plot**  
 BOX-COX LINEARITY PLOT Y X  
 where Y and X are the response variables.
- Box-Cox Normality Plot**  
 BOX-COX NORMALITY PLOT Y  
 where Y is the response variable.
- Box Plot**  
 BOX PLOT Y X  
 The BOX PLOT command is usually preceded by the commands  
 CHARACTER BOX PLOT  
 LINE BOX PLOT  
 These commands set the default line and character settings for the box plot.  
 You can use the CHARACTER and LINE commands to choose your own line and character settings if you prefer.  
 To show the outliers as circles, enter the command  
 FENCES ON
- Complex Demodulation Amplitude Plot**  
 CHI-SQUARE TEST Y A  
 where Y is the response variable and A is the value being tested.  
 Return to the Chi-Square Test for the Standard Deviation Page  
 Dataplot Command for Complex Demodulation Amplitude Plot  
 The Dataplot command for a complex demodulation amplitude plot is  
 COMPLEX DEMODULATION AMPLITUDE PLOT Y  
 where Y is the response variable.
- Complex Demodulation Phase Plot**  
 DEMODULATION FREQUENCY <VALUE>  
 COMPLEX DEMODULATION PHASE PLOT Y  
 where Y is the response variable.  
 The DEMODULATION FREQUENCY is used to specify the desired frequency for the COMPLEX DEMODULATION PLOT.  
 The value of the demodulation frequency is usually obtained from a spectral plot.  
 Return to the Complex Demodulation Phase Plot Page
- Contour Plot**
  - DOE Contour Plot**  
 CONTOUR PLOT Z X Y Z0  
 The variables X and Y define the grid, the Z variable is the response variable, and Z0 defines the desired contour levels.  
 Currently,  
 Dataplot only supports contour plots over regular grids.  
 Dataplot does provide 2D interpolation capabilities to form regular grids from irregular data.  
 Dataplot also does not support labels for the contour lines or solid fills between contour lines.
- DOE Scatter Plot**  
 The Dataplot command for generating a dex scatter plot is  
 DEX SCATTER PLOT Y X1 X2 X3 X4 X5  
 where Y is the response variable and X1, X2, X3, X4, and X5 are the factor variables.  
 The number of factor variables can vary, and is at least one.  
 The DEX SCATTER PLOT is typically preceded by the commands  
 CHARACTER X BLANK  
 LINE BLANK SOLID  
 However, you can set the plot character and line settings to whatever seems appropriate.  
 It is often desirable to provide alphabetic labels for the factors.  
 For example, if there are 2 factors, time and temperature, the following commands could be used to define alphabetic labels:
 

```

XLIMITS 1 2
XTIC OFFSET 0.5 0.5
MAJOR XTIC MARK NUMBER 2
MINOR XTIC MARK NUMBER 0
XTIC MARK LABEL FORMAT ALPHA
XTIC MARK LABEL CONTENT TIME TEMPERATURE
      
```
- DOE Mean Plot**  
 The Dataplot command to generate a dex mean interaction effects plot is  
 DEX MEAN INTERACTION EFFECTS PLOT Y X1 X2 X3 X4 X5  
 where Y is the response variable and X1, X2, X3, X4, and X5 are the factor variables.  
 The number of factor variables can vary, and is at least one.  
 Dataplot supports the following additional plots for other location statistics  
 DEX MEDIAN INTERACTION EFFECTS PLOT Y X1 X2 X3 X4 X5  
 DEX MIDMEAN INTERACTION EFFECTS PLOT Y X1 X2 X3 X4 X5  
 DEX TRIMMED MEAN INTERACTION EFFECTS PLOT Y X1 X2 X3 X4 X5  
 DEX WINSORIZED MEAN INTERACTION EFFECTS PLOT Y X1 X2 X3 X4 X5  
 If you want the raw data plotted rather than a statistic, enter  
 DEX INTERACTION EFFECTS PLOT Y X1 X2 X3 X4 X5  
 The LINE and CHARACTER commands can be used to control the appearance of the plot.  
 For example, a typical sequence of commands might be  
 LINE SOLID SOLID  
 CHARACTER CIRCLE BLANK  
 CHARACTER FILL ON  
 This draws the connecting line between the levels of a factor and the overall mean reference line as solid lines.  
 In addition, the level means are drawn with a solid fill circle.  
 This command is a variant of the SCATTER PLOT MATRIX command.  
 There are a number of options to control the appearance of these plots.  
 In Dataplot, you can enter HELP SCATTER PLOT MATRIX for details.
- DOE Standard Deviation Plot**  
 The Dataplot command to generate a dex standard deviation plot is  
 DEX STANDARD DEVIATION PLOT Y X1 X2 X3 X4 X5  
 where Y is the response variable and X1, X2, X3, X4, and X5 are the factor variables.

The number of factor variables can vary, and is at least one.  
 Dataplot supports the following additional plots for other scale statistics.

```
DEX VARIANCE PLOT Y X1 X2 X3 X4 X5
DEX MEDIAN ABSOLUTE VALUE PLOT Y X1 X2 X3 X4 X5
DEX AVERAGE ABSOLUTE VALUE PLOT Y X1 X2 X3 X4 X5
DEX RANGE VALUE PLOT Y X1 X2 X3 X4 X5
DEX MIDRANGE VALUE PLOT Y X1 X2 X3 X4 X5
DEX MINIMUM PLOT Y X1 X2 X3 X4 X5
DEX MAXIMUM PLOT Y X1 X2 X3 X4 X5
```

The LINE and CHARACTER commands can be used to control the appearance of the plot. For example, a typical sequence of commands might be

```
LINE SOLID SOLID
CHARACTER CIRCLE BLANK
CHARACTER FILL ON
```

This draws the connecting line between the levels of a factor and the overall mean reference line as solid lines.

In addition, the level means are drawn with a solid fill circle.

It is often desirable to provide alphabetic labels for the factors.

For example, if there are 2 factors, time and temperature, the following commands could be used to define alphabetic labels:

```
XLIMITS 1 2
XTIC OFFSET 0.5 0.5
MAJOR XTIC MARK NUMBER 2
MINOR XTIC MARK NUMBER 0
XTIC MARK LABEL FORMAT ALPHA
XTIC MARK LABEL CONTENT TIME TEMPERATURE
```

#### 14. Histogram

HISTOGRAM Y

where Y is the response variable.

The different variants of the histogram can be generated with the commands

```
RELATIVE HISTOGRAM Y
CUMULATIVE HISTOGRAM Y
RELATIVE CUMULATIVE HISTOGRAM Y
```

The class width, the start of the first class, and the end of the last class can be specified with the commands

```
CLASS WIDTH <value>
CLASS LOWER <value>
CLASS UPPER <value>
```

By default,

Dataplot uses a class width of  $0.3 \times SD$  where SD is the standard deviation of the data.

The lower class limit is the sample mean minus 6 times the sample standard deviation.

Similarly, the upper class limit is the sample mean plus 6 times the sample standard deviation.

By default,

Dataplot uses the probability normalization for relative histograms.

If you want the relative counts to sum to one instead, enter the command

```
SET RELATIVE HISTOGRAM PERCENT
```

To reset the probability interpretation, enter

```
SET RELATIVE HISTOGRAM AREA
```

1. Histogram Interpretation: Normal
2. Histogram Interpretation: Symmetric, Non-Normal, Short-Tailed
3. Histogram Interpretation: Symmetric, Non-Normal, Long-Tailed
4. Histogram Interpretation: Symmetric and Bimodal
5. Histogram Interpretation: Bimodal Mixture of 2 Normals
6. Histogram Interpretation: Skewed (Non-Normal) Right
7. Histogram Interpretation: Skewed (Non-Symmetric) Left
8. Histogram Interpretation: Symmetric with Outlier

#### 15. Lag Plot

LAG PLOT Y

The appearance of the lag plot can be controlled with appropriate settings for the LINE and CHARACTER commands.

Typical settings for these commands would be

```
LINE BLANK
CHARACTER X
```

To generate a linear fit of the points on the lag plot when an autoregressive fit is suggested, enter the following commands

```
LAG PLOT Y
LINEAR FIT YPLOT XPLOT
```

The variables YPLOT and XPLOT are internal variables that store the coordinates of the most recent plot.

Return to the Lag Plot Page

1. Lag Plot: Random Data
2. Lag Plot: Moderate Autocorrelation
3. Lag Plot: Strong Autocorrelation and Auto regressiveModel
4. Lag Plot: Sinusoidal Models and Outliers

#### 16. Linear Correlation Plot

The Dataplot command to generate a linear correlation plot is

```
LINEAR CORRELATION PLOT Y X TAG
```

where Y is the response variable, X is the independent variable, and TAG is the group id variable.

The appearance of the plot can be controlled with appropriate settings for the LINE and CHARACTER commands.

Typical settings would be

```
CHARACTER X BLANK
LINE BLANK SOLID
```

#### 17. Linear Intercept Plot

The Dataplot command to generate a linear intercept plot is

```
LINEAR INTERCEPT PLOT Y X TAG
```

where Y is the response variable, X is the independent variable, and TAG is the group id variable.

The appearance of the plot can be controlled with appropriate settings for the LINE and CHARACTER commands.

Typical settings would be

```
CHARACTER X BLANK
LINE BLANK SOLID
```

#### 18. Linear Slope Plot

The Dataplot command to generate a linear slope plot is

```
LINEAR SLOPE PLOT Y X TAG
```

where Y is the response variable, X is the independent variable, and TAG is the group id variable.

The appearance of the plot can be controlled with appropriate settings for the LINE and CHARACTER commands.

Typical settings would be

```
CHARACTER X BLANK
LINE BLANK SOLID
```

#### 19. Linear Residual Standard Deviation Plot

The Dataplot command to generate a linear residual standard deviation plot is

```
LINEAR RESSD PLOT Y X TAG
```

where Y is the response variable, X is the independent variable, and TAG is the group id variable.

The appearance of the plot can be controlled with appropriate settings for the LINE and CHARACTER commands.

Typical settings would be

```
CHARACTER X BLANK
LINE BLANK SOLID
```

#### 20. Mean Plot

The Dataplot command to generate a mean plot is

```
MEAN PLOT Y X
```

where Y is a response variable and X is a group id variable.

Dataplot supports this command for a number of other common location statistics.

For example,

```
MEDIAN PLOT Y X and MID-RANGE PLOT Y X
```

compute the median and mid-range instead of the mean for each group.

#### 21. Normal Probability Plot

The Dataplot command to generate a normal probability plot is

```
NORMAL PROBABILITY PLOT Y
```

where Y is the response variable.

If your data are already grouped (i.e., Y contains counts for the groups identified by X), the Dataplot command is

```
NORMAL PROBABILITY PLOT Y X
```

Dataplot returns the following internal parameters when it generates a probability plot.

PPCC the correlation coefficient of the fitted line on the probability plot.

This is a measure of how well the straight line fits the probability plot.

PPA0 the intercept term for the fitted line on the probability plot.

This is an estimate of the location parameter.  
PPA1 the slope term for the fitted line on the probability plot.  
This is an estimate of the scale parameter.  
SDPPA0 the standard deviation of the intercept term for the fitted line on the probability plot.  
SDPPA1 the standard deviation of the slope term for the fitted line on the probability plot.  
PPRESSD the residual standard deviation of the fitted line on the probability plot.  
This is a measure of the adequacy of the fitted line.  
PPRESDF the residual degrees of freedom of the fitted line on the probability plot.

- Normal Probability Plot: Normally Distributed Data
- Normal Probability Plot: Data Have Short Tails
- Normal Probability Plot: Data Have Long Tails
- Normal Probability Plot: Data are Skewed Right

22. Probability Plot  
The Dataplot command for a probability plot is  
`<dist> PROBABILITY PLOT Y`  
where `<dist>` is the name of the specific distribution.  
Dataplot currently supports probability plots for over 70 distributions.  
For example,  
`NORMAL PROBABILITY PLOT Y`  
`EXPONENTIAL PROBABILITY PLOT Y`  
`DOUBLE EXPONENTIAL PROBABILITY PLOT Y`  
`CAUCHY PROBABILITY PLOT Y`  
For some distributions, you may need to specify one or more shape parameters.  
For example, to specify the shape parameter for the gamma distribution, you might enter the following commands:  
`LET GAMMA = 2`  
`GAMMA PROBABILITY PLOT Y`  
Enter the command `LIST DISTRIBUTIONS` to see a list of distributions for which Dataplot supports probability plots (and to see what parameters need to be specified).  
Dataplot returns the following internal parameters when it generates a probability plot.  
PPCC the correlation coefficient of the fitted line on the probability plot.  
This is a measure of how well the straight line fits the probability plot.  
PPA0 the intercept term for the fitted line on the probability plot.  
This is an estimate of the location parameter.  
PPA1 the slope term for the fitted line on the probability plot.  
This is an estimate of the scale parameter.  
SDPPA0 the standard deviation of the intercept term for the fitted line on the probability plot.  
SDPPA1 the standard deviation of the slope term for the fitted line on the probability plot.  
PPRESSD the residual standard deviation of the fitted line on the probability plot.  
This is a measure of the adequacy of the fitted line.  
PPRESDF the residual degrees of freedom of the fitted line on the probability plot.

- Probability Plot Correlation Coefficient Plot
- Quantile-Quantile Plot
- Run-Sequence Plot
- Scatter Plot

The Dataplot command to generate a scatter plot is  
`PLOT Y X`  
The appearance of the plot can be controlled by appropriate settings of the `CHARACTER` and `LINE` commands and their various attribute-setting commands.

- Scatter Plot: No Relationship
- Scatter Plot: Strong Linear (positive correlation) Relationship
- Scatter Plot: Strong Linear (negative correlation) Relationship
- Scatter Plot: Exact Linear (positive correlation) Relationship
- Scatter Plot: Quadratic Relationship
- Scatter Plot: Exponential Relationship
- Scatter Plot: Sinusoidal Relationship (damped)
- Scatter Plot: Variation of Y Does Not Depend on X (homoscedastic)
- Scatter Plot: Variation of Y Does Depend on X (heteroscedastic)
- Scatter Plot: Outlier
- Scatterplot Matrix

The Dataplot command to generate a scatterplot matrix is  
`SCATTER PLOT MATRIX X1 X2 ... XK`  
The appearance of the plot can be controlled by appropriate settings of the `CHARACTER` and `LINE` commands and their various attribute-setting commands.

In addition,  
Dataplot provides a number of `SET` commands to control the appearance of the scatterplot matrix.  
The most common commands are:  
`SET MATRIX PLOT LOWER DIAGONAL <ON/OFF>`  
This command controls whether or not the plots below the diagonal are plotted.  
`SET MATRIX PLOT TAG <ON/OFF>`  
If `ON`, the last variable on the `SCATTER PLOT MATRIX` command is not plotted directly.  
Instead, it is used as a group-id variable.  
You can use the `CHARACTER` and `LINE` commands to set the plot attributes for each group.  
`SET MATRIX PLOT FRAME <DEFAULT/USER/CONNECTED>`  
If `DEFAULT`, the plot frames are connected (that is, it does a `FRAME CORNER COORDINATES 0 0 100 100`).  
The axis tic marks and labels are controlled automatically.  
If `CONNECTED`, then it is similar to `DEFAULT` except the current value of `FRAME CORNER COORDINATES` is used.  
This is useful for putting a small gap between the plots (e.g., enter `FRAME CORNER COORDINATES 3 3 97 97` before generating the scatterplot matrix).

If `USER`,  
Dataplot does not connect the plot frames.  
The tic marks and labels are as the user set them.  
`SET MATRIX PLOT FIT <NONE/LOWESS/LINEAR/QUADRATIC>` This controls whether a lowess fit, a linear fit, a quadratic fit line, or no fit is superimposed on the plot points.  
If `lowess`,  
a rather high value of the lowess fraction is recommended (e.g.,  
`LOWESS FRACTION 0.6`).  
In Dataplot, enter `HELP SCATTER PLOT MATRIX` for additional options for this plot.

12. Conditioning Plot  
`CONDITION PLOT Y X COND`  
Y is the response variable,  
X is the independent variable,  
and `COND` is the conditioning variable.  
Dataplot expects `COND` to contain a discrete number of distinct values.  
Dataplot provides a number of commands for creating a discrete variable from a continuous variable.  
For example,  
suppose `X2` is a continuous variable that we want to split into 4 regions.  
We could enter the following sequence of commands to create a discrete variable from `X2`.  
`LET COND = X2`  
`LET COND = 1 SUBSET X2 = 0 TO 99.99`  
`LET COND = 2 SUBSET X2 = 100 TO 199.99`  
`LET COND = 3 SUBSET X2 = 200 TO 299.99`  
`LET COND = 4 SUBSET X2 = 300 TO 400`  
The `SUBSET` feature can be used as above to create whatever ranges we want.

A simpler,  
more automatic way is to use the `CODE` command in Dataplot.  
For example,  
`LET COND = CODE4 X2`  
splits the data into quartiles and assigns a value of 1 to 4 to `COND` based on what quartile the corresponding value of `X2` is in.  
The appearance of the plot can be controlled by appropriate settings of the `CHARACTER` and `LINE` commands and their various attribute-setting commands.

In addition,  
Dataplot provides a number of `SET` commands to control the appearance of the conditioning plot.  
In Dataplot,  
enter `HELP CONDITION PLOT` for details.

27. Spectral Plot

- Spectral Plot: Random Data
- Spectral Plot: Strong Autocorrelation and Autoregressive Model
- Spectral Plot: Sinusoidal Model

The Dataplot command to generate a spectral plot is  
SPECTRAL PLOT Y

28. Standard Deviation Plot  
The Dataplot command to generate a standard deviation plot is  
STANDARD DEVIATION PLOT Y X  
where Y is a response variable and X is a group id variable.  
Dataplot supports this command for a number of other common scale statistics.

For example,  
AAD PLOT Y X and MAD PLOT Y X compute the average absolute deviation and median absolute deviation,  
respectively,  
instead of the standard deviation for each group.

29. Star Plot  
STAR PLOT X1 TO XP FOR I = 10 1 10  
where there are p response variables called X1,  
X2,  
... ,  
XP.  
Note that this syntax prints one star,  
specifically the tenth row of the X1,  
X2,  
... ,  
XP variables.  
Typically,  
multiple star plots will be displayed on the same page.

For example,  
to plot the first 25 rows on the same page,  
enter the following sequence of commands  
MULTI PLOT CORNER COORDINATES 0 0 100 100  
MULTI PLOT 5 5  
LOOP FOR K = 1 1 25  
STAR PLOT X1 TO XP FOR I = K 1 K  
END OF LOOP

30. Weibull Plot  
WEIBULL PLOT Y  
WEIBULL PLOT Y X  
where Y is the response variable containing failure times and X is an optional censoring variable.  
A value of 1 indicates the item failed by the failure mode of interest while a value of 0 indicates that the item failed by a failure mode that is not  
of interest.  
The appearance of the plot can be controlled with appropriate settings for the LINE and CHARACTER commands.

For example,  
to draw the raw data with the "X" character and the 2 reference lines as dashed lines,  
enter the commands  
LINE BLANK DASH DASH  
CHARACTER X BLANK BLANK  
WEIBULL PLOT Y X  
Dataplot saves the following internal parameters after the Weibull plot.  
ETA - the estimated characteristic life  
BETA - the estimated shape parameter  
SDETA - the estimated standard deviation of ETA  
SDBETA - the estimated standard deviation of BETA  
BPT1 - the estimated 0.1% point of failure times  
BPT5 - the estimated 0.5% point of failure times  
B1 - the estimated 1% point of failure times  
B5 - the estimated 5% point of failure times  
B10 - the estimated 10% point of failure times  
B20 - the estimated 20% point of failure times  
B50 - the estimated 50% point of failure times  
B80 - the estimated 80% point of failure times  
B90 - the estimated 90% point of failure times  
B95 - the estimated 95% point of failure times  
B99 - the estimated 99% point of failure times  
B995 - the estimated 99.5% point of failure times  
B999 - the estimated 99.9% point of failure times

31. Youden Plot  
1. DOE Youden Plot  
The Dataplot command to generate a Youden plot is  
YOUDEN PLOT Y1 Y2 LAB  
where Y1 and Y2 are the response variables and LAB is a laboratory (or run number) identifier.  
The LINE and CHARACTER commands can be used to control the appearance of the Youden plot.

For example,  
if there are 5 labs,  
a typical sequence would be  
LINE BLANK ALL  
CHARACTER 1 2 3 4 5  
YOUDEN PLOT Y X LAB

32. 4-Plot  
4-PLOT Y  
where Y is the response variable.

33. 6-Plot  
FIT Y X  
6-PLOT Y X  
where Y is the response variable and X is the independent variable.

4. Graphical Techniques: By Problem Category

5. Quantitative Techniques

1. Measures of Location  
Various measures of location can be computed in Dataplot as follows:  
LET A = MEAN Y  
LET A = MEDIAN Y  
LET A = MIDMEAN Y  
LET P1 = 10  
LET P2 = 10  
LET A = TRIMMED MEAN Y  
LET P1 = 10  
LET P2 = 10  
LET A = WINSORIZED Y  
In the above,  
P1 and P2 are used to set the percentage of values that are trimmed or Winsorized.  
Use P1 to set the percentage for the lower tail and P2 the percentage for the upper tail.

2. Confidence Limits for the Mean  
The following commands can be used in Dataplot to generate a confidence interval for the mean or to generate a one sample t-test,  
respectively.  
CONFIDENCE LIMITS Y  
T TEST Y U0  
where Y is the response variable and U0 is a parameter or scalar value that defines the hypothesized value.

3. Two-Sample t-Test for Equal Means  
1. Data Used for Two-Sample t-Test

4. One-Factor ANOVA  
The Dataplot command for a one way analysis of variance is  
ANOVA Y X  
where Y is a response variable and X is a group identifier variable.  
Dataplot is currently limited to the balanced case (i.e.,  
each level has the same number of observations) and it does not compute interaction effect estimates.

5. Multi-factor Analysis of Variance  
ANOVA Y X1

```

ANOVA Y X1 X2
ANOVA Y X1 X2 X3
ANOVA Y X1 X2 X3 X4
ANOVA Y X1 X2 X3 X4 X5
where Y is the response variable and X1,
X2,
X3,
X4,
and X5 are factor variables.
Daplot allows up to 10 factor variables.
Daplot is currently limited to the balanced case (i.e.,
each level has the same number of observations) and it does not compute interaction effect estimates.

```

6. Measures of Scale

The various scale measures can be computed in Daplot as follows:

```

LET A = VARIANCE Y
LET A = STANDARD DEVIATION Y
LET A = AVERAGE ABSOLUTE DEVIATION Y
LET A = MEDIAN ABSOLUTE DEVIATION Y
LET A = RANGE Y
LET A1 = LOWER QUARTILE Y
LET A2 = UPPER QUARTILE Y
LET IQRANGE = A2 - A1

```

7. Bartlett's Test

```

BARTLETT TEST Y X

```

where Y is the response variable and X is the group id variable.  
The above computes the standard form of Bartlett's test.

To compute the Dixon-Massey form of Bartlett's test,  
the Daplot command is one of the following (these are synonyms,  
not distinct commands)

```

DIXON BARTLETT TEST Y X
DIXON MASSEY BARTLETT TEST Y X
DM BARTLETT TEST Y X

```

8. Chi-Square Test for the Standard Deviation

1. Data Used for Chi-Square Test for the Standard Deviation

9. F-Test for Equality of Two Standard Deviations

10. Levene Test for Equality of Variances

The Daplot command for the Levene test is

```

LEVENE TEST Y X

```

where Y is the response variable and X is the group id variable.

11. Measures of Skewness and Kurtosis

Daplot Commands for skewness and kurtosis are

```

LET A = SKEWNESS Y
LET A = KURTOSIS Y

```

where Y is the response variable.

Daplot can also generate plots of the skewness and kurtosis for grouped data or one-factor data with the following commands:

```

SKEWNESS PLOT Y X
KURTOSIS PLOT Y X

```

where Y is the response variable and X is the group id variable.

12. Autocorrelation

To generate the lag 1 autocorrelation value in Daplot, enter

```

LET A = AUTOCORRELATION Y

```

where Y is the response variable.

In Daplot, the easiest way to generate the autocorrelations for lags greater than 1 is:

```

AUTOCORRELATION PLOT Y
LET AC = YPLOT
LET LAG = XPLOT
RETAIN AC LAG SUBSET TAGPLOT = 1

```

The AUTOCORRELATION PLOT command generates an autocorrelation plot for lags 0 to N/4.  
It also generates 95% and 99% confidence limits for the autocorrelations.  
Daplot stores the plot coordinates in the internal variables XPLOT,  
YPLOT,  
and TAGPLOT.

The 2 LET commands and the RETAIN command are used to extract the numerical values of the autocorrelations.  
The variable LAG identifies the lag while the corresponding row of AC contains the autocorrelation value.

13. Runs Test for Detecting Non-randomness

The Daplot command to generate a run sequence plot is

```

RUN SEQUENCE PLOT Y

```

Equivalently, you can enter

```

PLOT Y

```

The appearance of the plot can be controlled with appropriate settings of the LINE,  
CHARACTER,  
SPIKE,  
and BAR commands and their associated attribute-setting commands.

Return to the Run Sequence Plot Page

Daplot Command for the Runs Test

The Daplot command for a runs test is

```

RUNS TEST Y

```

where Y is a response variable.

14. Anderson-Darling Test

```

ANDERSON DARLING NORMAL TEST Y
ANDERSON DARLING LOGNORMAL TEST Y
ANDERSON DARLING EXPONENTIAL TEST Y
ANDERSON DARLING WEIBULL TEST Y
ANDERSON DARLING EXTREME VALUE TYPE I TEST Y

```

where Y is the response variable.

15. Chi-Square Goodness-of-Fit Test

```

<dist> CHI-SQUARE GOODNESS OF FIT TEST Y
<dist> CHI-SQUARE GOODNESS OF FIT TEST Y X
<dist> CHI-SQUARE GOODNESS OF FIT TEST Y XL XU

```

where <dist> is one of 70+ built-in distributions.

Daplot supports the chi-square goodness-of-fit test for all distributions that support the cumulative distribution function.  
To see a list of supported distributions,  
enter the command LIST DISTRIBUTIONS.

Some specific examples are

```

NORMAL CHI-SQUARE GOODNESS OF FIT TEST Y
LOGISTIC CHI-SQUARE GOODNESS OF FIT TEST Y
DOUBLE EXPONENTIAL CHI-SQUARE GOODNESS OF FIT TEST Y

```

You can specify the location and scale parameters (for any of the supported distributions) by entering

```

LET CHSLOC = value
LET CHSSCAL = value

```

You may need to enter the values for 1 or more shape parameters for distributions that require them.

For example,  
to specify the shape parameter gamma for the gamma distribution,  
enter the commands

```

LET GAMMA = value
GAMMA CHI-SQUARE GOODNESS OF FIT TEST Y

```

Daplot also allows you to control the class width,  
the lower limit (i.e.,  
start of the first bin),  
and the upper limit (i.e.,  
the end value for the last bin).

These commands are

```

CLASS WIDTH value
CLASS LOWER value
CLASS UPPER value

```

In most cases, the default Daplot class intervals will be adequate.  
If your data are already binned, you can enter the commands

NORMAL CHI-SQUARE GOODNESS OF FIT TEST Y X  
 NORMAL CHI-SQUARE GOODNESS OF FIT TEST Y XL XU  
 In both commands above,  
 Y is the frequency variable.  
 If one X variable is given,  
 Dataplot assumes that it is the bin mid point and that bins have equal width.  
 If two X variables are given,  
 Dataplot assumes that these are the bin end points and that the bin widths are not necessarily of equal width.  
 Unequal bin widths are typically used to combine classes with small frequencies since the chi-square approximation for the test may not be accurate if there are frequency classes with less than five observations.

16. Kolmogorov-Smirnov Goodness-of-Fit Test  
 <dist> KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST Y  
 where <dist> is one of 60+ built-in distributions.  
 The K-S goodness of fit test is supported for all Dataplot internal continuous distributions that support the CDF (cumulative distribution function).  
 The command LIST DISTRIBUTIONS shows the currently supported distributions in Dataplot.  
 Some specific examples are

NORMAL KOLM-SMIR GOODNESS OF FIT Y  
 LOGISTIC KOLM-SMIR GOODNESS OF FIT Y  
 DOUBLE EXPONENTIAL KOLM-SMIR GOODNESS OF FIT Y  
 You can specify the location and scale parameters by entering  
 LET KSLC = value  
 LET KSSCALE = value

You may need to enter the values for 1 or more shape parameters for distributions that require them.  
 For example,  
 to specify the shape parameter gamma for the gamma distribution,  
 enter the commands

LET GAMMA = value  
 GAMMA KOLMOGOROV-SMIRNOV GOODNESS OF FIT TEST Y  
 Be aware that you should not use the same data to estimate these distributional parameters as you use to calculate the K-S test as the critical values of the K-S test assume the distribution is fully specified.  
 The empirical cdf function can be plotted with the following command  
 EMPIRICAL CDF PLOT Y

17. Grubbs' Test for Outliers  
 GRUBBS <MINIMUM/MAXIMUM> TEST Y  
 where Y is the response variable.  
 Dataplot identifies one outlier at a time.  
 The MINIMUM or MAXIMUM keyword is optional.  
 If omitted,  
 the most extreme value will be checked (regardless of whether it is in the minimum or maximum direction).

18. Yates Analysis  
 1. Defining Models and Prediction Equations  
 2. Important Factors  
 The Dataplot command for a Yates analysis is  
 YATES Y  
 where Y is a response variable in Yates order.

6. Probability Distributions
  1. What is a Probability Distribution
  2. Related Distributions
  3. Families of Distributions
  4. Location and Scale Parameters
  5. Estimating the Parameters of a Distribution
    1. Method of Moments
    2. Maximum Likelihood
    3. Least Squares
    4. PPCC and Probability Plots
  6. Gallery of Distributions
    1. Normal Distribution

Dataplot can compute the probability functions for the power normal distribution with the following commands.  
 cdf  
 LET Y = PNRCDF(X,P,SD,MU)  
 pdf  
 LET Y = PNRPDF(X,P,SD,MU)  
 ppf  
 LET Y = PNRPPF(X,P,SD,MU)  
 hazard  
 LET Y = PNRHAZ(X,P,SD,MU)  
 cumulative hazard  
 LET Y = PNRCHAZ(X,P,SD,MU)  
 survival  
 LET Y = 1 - PNRCDF(X,P,SD,MU)  
 inverse survival  
 LET Y = PNRPPF(1-X,P,SD,MU)  
 probability plot  
 P = value  
 LET SD = value (defaults to 1)  
 POWER NORMAL PROBABILITY PLOT Y  
 ppcc plot  
 POWER NORMAL PPCC PLOT Y  
 In the above,

X can be a number,  
 a parameter,  
 or a variable.  
 SD and MU are the scale and location parameters,  
 respectively,  
 and they are optional (a location of 0 and scale of 1 are used if they are omitted).  
 If given,  
 SD and MU can be a number,  
 a parameter,  
 or a variable.  
 However,  
 they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
 PLOT PNRPDF(X,10,1) FOR X = -5 0.01 5

2. Uniform Distribution  
 cdf  
 LET Y = UNICDF(X,A,B)  
 pdf  
 LET Y = UNIPDF(X,A,B)  
 ppf  
 LET Y = UNIPPF(X,A,B)  
 hazard  
 LET Y = UNHAZ(X,A,B)  
 cumulative hazard  
 LET Y = UNICHAZ(X,A,B)  
 survival  
 LET Y = 1 - UNICDF(X,A,B)  
 inverse survival  
 LET Y = UNIPPF(1-X,A,B)  
 random numbers  
 LET Y = UNIFORM RANDOM NUMBERS FOR I = 1 1 1000  
 probability plot  
 UNIFORM PROBABILITY PLOT Y  
 parameter estimation  
 The method of moment estimators can be computed with the commands  
 LET YMEAN = MEAN Y  
 LET YSD = STANDARD DEVIATION Y  
 LET A = YMEAN - SQRT(3)\*YSD  
 LET B = YMEAN + SQRT(3)\*YSD  
 The maximum likelihood estimators can be computed with the commands

```

        LET YRANGE = RANGE Y
        LET YMIDRANG = MID-RANGE Y
        LET A = YMIDRANG - 0.5*YRANGE
        LET B = YMIDRANG + 0.5*YRANGE
    In the above,
    X can be a number,
    a parameter,
    or a variable.
A and B are the lower and upper limits of the uniform distribution and they are optional (A is 0 and B is 1 if they are omitted).
The location parameter is A and the scale parameter is (B - A).
If given,
    A and B can be a number,
    a parameter,
    or a variable.
However,
    they are typically either a number or a parameter.
    These functions can be used in the Dataplot PLOT and FIT commands as well. For example,
        PLOT UNIPDF(X) FOR X = 0 0.1 1

```

### 3. Cauchy Distribution

Dataplot can compute the probability functions for the Cauchy distribution with the following commands.

```

cdf
LET Y = CAUCDF(X,A,B)
pdf
LET Y = CAUPDF(X,A,B)
ppf
LET Y = CAUPPF(X,A,B)
hazard
LET Y = CAUHAZ(X,A,B)
cumulative hazard
LET Y = CAUHAZ(X,A,B)
survival
LET Y = 1 - CAUCDF(X,A,B)
inverse survival
LET Y = CAUPPF(1-X,A,B)
random numbers
LET Y = CAUCHY RANDOM NUMBERS FOR I = 1 1 1000
probability plot
CAUCHY PROBABILITY PLOT Y
where X can be a number,
a parameter,
or a variable.
A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).
If given,
    A and B can be a number,
    a parameter,
    or a variable.
However,
    they are typically either a number or a parameter.
    These functions can be used in the Dataplot PLOT and FIT commands as well. For example,
        PLOT CAUPDF(X) FOR X = -5 0.01 5

```

### 4. t Distribution

```

cdf
LET Y = TCDF(X,NU,A,B)
pdf
LET Y = TPDF(X,NU,A,B)
ppf
LET Y = TPPF(X,NU,A,B)
random numbers
LET NU = value
LET Y = T RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET NU = value
T PROBABILITY PLOT Y
ppcc plot
LET NU = value
T PPCC PLOT Y
In the above,
X can be a number,
a parameter,
or a variable.
NU is the shape parameter (= number of degrees of freedom).
NU can be a number,
a parameter,
or a variable.
However,
    it is typically either a number or a parameter.
A and B are the location and scale parameters,
respectively,
and they are optional (a location of 0 and scale of 1 are used if they are omitted).
If given,
    A and B can be a number,
    a parameter,
    or a variable.
However,
    they are typically either a number or a parameter.
    These functions can be used in the Dataplot PLOT and FIT commands as well. For example,
        PLOT TPDF(X) FOR X = -4 0.01 4

```

### 5. F Distribution

```

cdf
LET Y = FCDF(X,NU1,NU2,A,B)
pdf
LET Y = FPDF(X,NU1,NU2,A,B)
ppf
LET Y = FPPF(X,NU1,NU2,A,B)
random numbers
LET NU1 = value
LET NU2 = value
LET Y = F RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET NU1 = value
LET NU2 = value
F PROBABILITY PLOT Y
where X can be a number,
a parameter,
or a variable.
NU1 and NU2 are the shape parameters (= number of degrees of freedom).
NU1 and NU2 can be a number,
a parameter,
or a variable.
However,
    they are typically either a number or a parameter.
A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).
If given,
    A and B can be a number,
    a parameter,
    or a variable.
However,
    they are typically either a number or a parameter.
    These functions can be used in the Dataplot PLOT and FIT commands as well. For example,
        PLOT FPDF(X,10,10) FOR X = 0 0.01 5
    Return to the F Distribution Page
    Dataplot Command for F Test for Equality of Two Standard Deviations

```



The Dataplot command for the F test for the equality of two standard deviations is  
 F TEST Y1 Y2  
 where Y1 is the data for sample one and Y2 is the data for sample two.

#### 6. Chi-Square Distribution

Dataplot can compute the probability functions for the chi-square distribution with the following commands.

```

cdf
LET Y = CHSCDF(X,NU,NU2,A,B)
pdf
LET Y = CHSPDF(X,NU,A,B)
ppf
LET Y = CHSPPF(X,NU,A,B)
random numbers
LET NU = value
LET Y = CHI-SQUARE RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET NU = value
CHI-SQUARE PROBABILITY PLOT Y
ppcc plot
LET NU = value
CHI-SQUARE PPCC PLOT Y
where X can be a number,

```

a parameter,  
 or a variable.  
 NU is the shape parameter (number of degrees of freedom).  
 NU can be a number,  
 a parameter,  
 or a variable.

However,  
 it is typically either a number or a parameter.

A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,

A and B can be a number,  
 a parameter,  
 or a variable.

However,  
 they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
 PLOT CHSPDF(X,5) FOR X = 0 0.01 5

#### 7. Exponential Distribution

```

cdf
LET Y = EXPCDF(X,A,B)
pdf
LET Y = EXPPDF(X,A,B)
ppf
LET Y = EXPPPF(X,A,B)
hazard
LET Y = EXPHAZ(X,A,B)
cumulative hazard
LET Y = EXPCHAZ(X,A,B)
survival
LET Y = 1 - EXPCDF(X,A,B)
inverse survival
LET Y = EXPPPF(1-X,A,B)
random numbers
LET Y = EXPONENTIAL RANDOM NUMBERS FOR I = 1 1 1000
probability plot
EXPONENTIAL PROBABILITY PLOT Y
parameter estimation
If your data are not censored, enter the commands
SET CENSORING TYPE NONE
EXPONENTIAL MLE Y
If your data have type 1 censoring at fixed time t0, enter the commands
LET TEND = censoring time
SET CENSORING TYPE 1
EXPONENTIAL MLE Y X
If your data have type 2 censoring, enter the commands
SET CENSORING TYPE 2
EXPONENTIAL MLE Y X
Y is the response variable and X is the censoring variable where a value of 1 indicates a failure time and a value of 0 indica

```

tes a censoring time.

In addition to the point estimates,  
 confidence intervals for the parameters are generated.

In the above,

X can be a number,  
 a parameter,  
 or a variable.

A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,

A and B can be a number,  
 a parameter,  
 or a variable.

However,  
 they are typically either a number or parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
 PLOT EXPPDF(X) FOR X = 0 0.01 4

#### 8. Weibull Distribution

```

cdf
LET Y = WEICDF(X,GAMMA,A,B)
pdf
LET Y = WEIPDF(X,GAMMA,A,B)
ppf
LET Y = WEIPPF(X,GAMMA,A,B)
hazard
LET Y = WEIHAZ(X,GAMMA,A,B)
cumulative hazard
LET Y = WEICHAZ(X,GAMMA,A,B)
survival
LET Y = 1 - WEICDF(X,GAMMA,A,B)
inverse survival
LET Y = WEIPPF(1-X,GAMMA,A,B)
random numbers
LET GAMMA = value
LET Y = WEIBULL RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET GAMMA = value
WEIBULL PROBABILITY PLOT Y
ppcc plot
LET GAMMA = value
WEIBULL PPCC PLOT Y
parameter estimation
If your data are not censored, enter the commands
SET CENSORING TYPE NONE
WEIBULL MLE Y
If your data have type 1 censoring at fixed time t0, enter the commands
LET TEND = censoring time
SET CENSORING TYPE 1
WEIBULL MLE Y X
If your data have type 2 censoring, enter the commands
SET CENSORING TYPE 2
WEIBULL MLE Y X
Y is the response variable and X is the censoring variable where a value of 1 indicates a failure time and a value of 0 indica

```

tes a censoring time.

In addition to the point estimates,  
confidence intervals for the parameters are generated.

In the above,  
X can be a number,  
a parameter,  
or a variable.  
GAMMA is the shape parameter and is required.  
It can be a number,  
a parameter,  
or a variable.

It is typically a number or a parameter.  
A and B are the location and scale parameters,  
respectively,  
and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,  
A and B can be a number,  
a parameter,  
or a variable.

However,  
they are typically either a number or a parameter.  
These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
PLOT WEIPDF(X,2) FOR X = 0.01 0.01 5

#### 9. Lognormal Distribution

Dataplot can compute the probability functions for the lognormal distribution with the following commands.

```
cdf
LET Y = LGNCDF(X,SD,A,B)
pdf
LET Y = LGNPDF(X,SD,A,B)
ppf
LET Y = LGNPPF(X,SD,A,B)
hazard
LET Y = LGNHAZ(X,SD,A,B)
cumulative hazard
LET Y = LGNCHAZ(X,SD,A,B)
survival
LET Y = 1 - LGNCDF(X,SD,A,B)
inverse survival
LET Y = LGNPPF(1-X,SD,A,B)
random numbers
LET SD = value
LET Y = LOGNORMAL RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET SD = value
LOGNORMAL PROBABILITY PLOT Y
ppcc plot
LET SD = value
LOGNORMAL PPCC PLOT Y
parameter estimation
LOGNORMAL MLE Y
```

This returns point estimates for the shape and scale parameters. It does not handle censored data and it does not generate confidence intervals for the parameters.  
where X can be a number,

a parameter,  
or a variable.  
SD is the shape parameter and is optional.  
It can be a number,  
a parameter,  
or a variable.

It is typically a number or a parameter.  
A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,  
A and B can be a number,  
a parameter,  
or a variable.

However,  
they are typically either a number or a parameter.  
These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
PLOT LGNPDF(X,5) FOR X = 0.01 0.01 5

#### 10. Fatigue Life Distribution

```
cdf
LET Y = FLCDF(X,GAMMA,A,B)
pdf
LET Y = FLPDF(X,GAMMA,A,B)
ppf
LET Y = FLPPF(X,GAMMA,A,B)
hazard
LET Y = FLHAZ(X,GAMMA,A,B)
cumulative hazard
LET Y = FLCHAZ(X,GAMMA,A,B)
survival
LET Y = 1 - FLCDF(X,GAMMA,A,B)
inverse survival
LET Y = FLPPF(1-X,GAMMA,A,B)
random numbers
LET GAMMA = value
LET Y = FATIGUE LIFE RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET GAMMA = value
FATIGUE LIFE PROBABILITY PLOT Y
ppcc plot
LET GAMMA = value
FATIGUE LIFE PPCC PLOT Y
where X can be a number,
```

a parameter,  
or a variable.  
FLMA is the shape parameter and is required.  
It can be a number,  
a parameter,  
or a variable.

It is typically a number or a parameter.  
A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,  
A and B can be a number,  
a parameter,  
or a variable.

However,  
they are typically either a number or a parameter.  
These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
PLOT FLPDF(X,2) FOR X = 0.01 0.01 10

#### 11. Gamma Distribution

```
cdf
LET Y = GAMCDF(X,GAMMA,A,B)
pdf
LET Y = GAMPDF(X,GAMMA,A,B)
ppf
LET Y = GAMPF(X,GAMMA,A,B)
hazard
LET Y = GAMHAZ(X,GAMMA,A,B)
cumulative hazard
LET Y = GAMCHAZ(X,GAMMA,A,B)
survival
LET Y = 1 - GAMCDF(X,GAMMA,A,B)
```

```

inverse survival
LET Y = GAMPPF(1-X,GAMMA,A,B)
random numbers
LET GAMMA = value
LET Y = Gamma RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET GAMMA = value
Gamma PROBABILITY PLOT Y
ppcc plot
LET GAMMA = value
Gamma PPCC PLOT Y
maximum likelihood
GAMMA MLE Y

```

This returns a point estimate for the full-sample case. It does not provide confidence intervals for the parameters and it does not handle censored data.

where X can be a number,  
a parameter,  
or a variable.

GAMMA is the shape parameter and is required.  
It can be a number,  
a parameter,  
or a variable.

It is typically a number or a parameter.  
A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).  
If given,  
A and B can be a number,  
a parameter,  
or a variable.

However,  
they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
PLOT GAMPDF(X,2) FOR X = 0.01 0.01 10

12. Double Exponential Distribution  
Dataplot can compute the probability functions for the double exponential distribution with the following commands.

```

cdf
LET Y = DEXCDF(X,A,B)
pdf
LET Y = DEXPDF(X,A,B)
ppf
LET Y = DEXPPF(X,A,B)
hazard
LET Y = DEXHAZ(X,A,B)/(1 - DEXCDF(X,A,B))
cumulative hazard
LET Y = -LOG(1 - DEXCHAZ(X,A,B))
survival
LET Y = 1 - DEXCDF(X,A,B)
inverse survival
LET Y = DEXPPF(1-X,A,B)
random numbers
LET Y = DOUBLE EXPONENTIAL RANDOM NUMBERS FOR I = 1 1 1000
probability plot
DOUBLE EXPONENTIAL PROBABILITY PLOT Y
maximum likelihood
LET MU = MEDIAN Y
LET BETA = MEDIAN ABSOLUTE DEVIATION Y

```

where X can be a number,  
a parameter,  
or a variable.

A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).  
If given,  
A and B can be a number,  
a parameter,  
or a variable.

However,  
they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
PLOT DEXPDF(X) FOR X = -5 0.01 5

13. Power Normal Distribution  
14. Power Lognormal Distribution  
Dataplot can compute the probability functions for the power lognormal distribution with the following commands.

```

cdf
LET Y = PLNCDF(X,P,SD,MU)
pdf
LET Y = PLNPDF(X,P,SD,MU)
ppf
LET Y = PLNPPF(X,P,SD,MU)
hazard
LET Y = PLNHAZ(X,P,SD,MU)
cumulative hazard
LET Y = PLNCHAZ(X,P,SD,MU)
survival
LET Y = 1 - PLNCDF(X,P,SD,MU)
inverse survival
LET Y = PLNPPF(1-X,P,SD,MU)
probability plot
LET P = value
LET SD = value (defaults to 1)
POWER LOGNORMAL PROBABILITY PLOT Y
ppcc plot
LET SD = value POWER LOGNORMAL PPCC PLOT Y

```

In the above,  
X can be a number,  
a parameter,  
or a variable.

SD and MU are the scale and location parameters,  
respectively,  
and they are optional (a location of 0 and scale of 1 are used if they are omitted).  
If given,  
SD and MU can be a number,  
a parameter,  
or a variable.

However,  
they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example, the command  
PLOT PLNPDF(X,5,1) FOR X = 0.01 0.01 5

15. Tukey-Lambda (Gumbel) Distribution  
cdf
LET Y = LAMCDF(X,LAMBDA,A,B)
pdf
LET Y = LAMPDF(X,LAMBDA,A,B)
ppf
LET Y = LAMPPF(X,LAMBDA,A,B)
random numbers
LET LAMBDA = value
LET Y = TUKEY-LAMBDA RANDOM NUMBERS FOR I = 1 1 1000
probability plot
LET LAMBDA = value
TUKEY-LAMBDA PROBABILITY PLOT Y
ppcc plot
TUKEY-LAMBDA PPCC PLOT Y

In the above,  
X can be a number,

a parameter,  
or a variable.  
LAMBDA is the shape parameter and is required.

It can be a number,  
a parameter,  
or a variable.

It is typically a number or a parameter.

A and B are the location and scale parameters,  
respectively,  
and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,

A and B can be a number,  
a parameter,  
or a variable.

However,

they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,  
PLOT LAMPDF(X,0.14) FOR X = -5 0.01 5

#### 16. Extreme Value Type I Distribution

To specify the form of the Gumbel distribution based on the smallest value, enter the command

SET MINMAX 1

To specify the form of the Gumbel distribution based on the largest value, enter the command

SET MINMAX 2

One of these commands must be entered before using the commands below.

Dataplot can compute the probability functions for the extreme value type I distribution with the following commands.

cdf

LET Y = EVICDF(X,A,B)

pdf

LET Y = EVIPDF(X,A,B)

ppf

LET Y = EVIPPF(X,A,B)

hazard

LET Y = EV1HAZ(X,A,B)

cumulative hazard

LET Y = EV1CHAZ(X,A,B)

survival

LET Y = 1 - EVICDF(X,A,B)

inverse survival

LET Y = EVIPPF(1-X,A,B)

random numbers

LET Y = EXTREME VALUE TYPE 1 RANDOM NUMBERS FOR I = 1 1 1000

probability plot

EXTREME VALUE TYPE 1 PROBABILITY PLOT Y

maximum likelihood

EV1 MLE Y

This returns a point estimate for the full sample case. It does not provide confidence intervals for the parameters and it does not handle censored data.

In the above,

X can be a number,  
a parameter,  
or a variable.

A and B are the location and scale parameters and they are optional (a location of 0 and scale of 1 are used if they are omitted).

If given,

A and B can be a number,  
a parameter,  
or a variable.

However,

they are typically either a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well. For example,

SET MINMAX 1

PLOT EVIPDF(X) FOR X = -4 0.01 4

#### 17. Beta Distribution

#### 18. Binomial Distribution

Dataplot can compute the probability functions for the binomial distribution with the following commands.

cdf

LET Y = BINCDF(X,P,N)

pdf

LET Y = BINPDF(X,P,N)

ppf

LET Y = BINPPF(F,P,N)

random numbers

LET N = value

LET P = value

LET Y = BINOMIAL RANDOM NUMBERS FOR I = 1 1 1000

probability plot

LET N = value

LET P = value

BINOMIAL PROBABILITY PLOT Y

where X can be a number,

a parameter,  
or a variable.

P and N are the shape parameters and are required.

They can be a number,

a parameter,  
or a variable.

They are typically a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well.

For example,

PLOT BINPDF(X,0.5,100) FOR X = 0 1 100

#### 19. Poisson Distribution

Dataplot can compute the probability functions for the Poisson distribution with the following commands.

cdf

LET Y = POICDF(X,LAMBDA)

pdf

LET Y = POIPDF(X,LAMBDA)

ppf

LET Y = POIPPF(X,LAMBDA)

random numbers

LET LAMBDA = value

LET Y = POISSON RANDOM NUMBERS FOR I = 1 1 1000

probability plot

LET LAMBDA = value

POISSON PROBABILITY PLOT Y

ppcc plot

POISSON PPCC PLOT Y

where X can be a number,

a parameter,  
or a variable.

LAMBDA is the shape parameter and is required.

It can be a number,

a parameter,  
or a variable.

It is typically a number or a parameter.

These functions can be used in the Dataplot PLOT and FIT commands as well.

For example,

PLOT POIPDF(X,15) FOR X = 0 1 50

#### 7. Tables for Probability Distributions

1. Cumulative Distribution Function of the Standard Normal Distribution
2. Upper Critical Values of the Student's-t Distribution
3. Upper Critical Values of the F Distribution
4. Critical Values of the Chi-Square Distribution

5. Critical Values of the t\* Distribution  
6. Critical Values of the Normal PPCC Distribution

4. EDA Case Studies

- Case Studies Introduction
- Case Studies
  - Normal Random Numbers
    - Background and Data

Dataplot Commands to generate 1,000 normal random numbers with a location of 50 and a scale of 20 are

```
LET LOC = 50
LET SCALE = 20
LET Y = NORM RAND NUMBERS FOR I = 1 1 1000
LET Y = LOC + SCALE*Y
```

Programs that automatically generate random numbers are typically controlled by a seed, which is usually an integer value.

The importance of the seed is that it allows the random numbers to be replicated.

That is,

giving the program the same seed should generate the same sequence of random numbers.

If the ability to replicate the set of random numbers is not important, you can give any valid value for the seed.

In Dataplot,

the seed is an odd integer with a minimum (and default) value of 305.

Seeds less than 305 generate the same sequence as 305 and even numbers generate the same sequence as the preceding odd number.

To change the seed value to 401 in Dataplot, enter the command:

```
SEED 401
```
    - Graphical Output and Interpretation
    - Quantitative Output and Interpretation

The Dataplot command to generate a table of summary statistics is

```
SUMMARY Y
```

where Y is the response variable.
    - Work This Example Yourself
  - Uniform Random Numbers
    - Background and Data

Dataplot Commands to generate 1,000 uniform random numbers in the interval (-100,100) are

```
LET A = -100
LET B = 100
LET Y = UNIFORM RANDOM NUMBERS FOR I = 1 1 1000
LET Y = A + (B-A)*Y
```

A similar technique can be used for any package that can generate standard uniform random numbers.

Simply multiply by the scale value (equals upper limit minus lower limit) and add the location value.

Programs that automatically generate random numbers are typically controlled by a seed, which is usually an integer value.

The importance of the seed is that it allows the random numbers to be replicated.

That is,

giving the program the same seed should generate the same sequence of random numbers.

If the ability to replicate the set of random numbers is not important, you can give any valid value for the seed.

In Dataplot,

the seed is an odd integer with a minimum (and default) value of 305.

Seeds less than 305 generate the same sequence as 305 and even numbers generate the same sequence as the preceding odd number.

To change the seed value to 401 in Dataplot, enter the command:

```
SEED 401
```
    - Graphical Output and Interpretation
    - Quantitative Output and Interpretation
    - Work This Example Yourself
  - Random Walk
    - Background and Data

Dataplot Commands for the Generation of Random Walk Numbers To generate a random walk with 1,000 points requires the following

```
LET Y = UNIFORM RANDOM NUMBERS FOR I = 1 1 1000
LET Y2 = Y - 0.5
LET RW = CUMULATIVE SUM Y2
```
    - Test Underlying Assumptions
    - Develop A Better Model
    - Validate New Model
    - Work This Example Yourself
  - Josephson Junction Cryothermometry
    - Background and Data
    - Graphical Output and Interpretation
    - Quantitative Output and Interpretation
    - Work This Example Yourself
  - Beam Deflections
    - Background and Data
    - Test Underlying Assumptions
    - Develop a Better Model
    - Validate New Model
    - Work This Example Yourself
  - Filter Transmittance
    - Background and Data
    - Graphical Output and Interpretation
    - Quantitative Output and Interpretation
    - Work This Example Yourself
  - Standard Resistor
    - Background and Data
    - Graphical Output and Interpretation
    - Quantitative Output and Interpretation
    - Work This Example Yourself
  - Heat Flow Meter 1
    - Background and Data
    - Graphical Output and Interpretation
    - Quantitative Output and Interpretation
    - Work This Example Yourself
  - Fatigue Life of Aluminum Alloy Specimens
    - Background and Data
    - Graphical Output and Interpretation
  - Ceramic Strength
    - Background and Data
    - Analysis of the Response Variable
    - Analysis of the Batch Effect
    - Analysis of the Lab Effect
    - Analysis of Primary Factors
    - Work This Example Yourself

---

3. References For Chapter 1: EDA

---

2. Measurement Process Characterization

- Characterization
  - What are the issues for characterization?
    - Purpose
    - Reference base
    - Bias and Accuracy
    - Variability
  - What is a check standard?
    - Assumptions
    - Data collection
    - Analysis
- Statistical control of a measurement process
  - What are the issues in controlling the measurement process?
  - How are bias and variability controlled?

1. Shewhart control chart
  1. EWMA control chart
2. Data collection
3. Monitoring bias and long-term variability
4. Remedial actions
3. How is short-term variability controlled?
  1. Control chart for standard deviations
  2. Data collection
  3. Monitoring short-term precision
  4. Remedial actions
3. Calibration
  1. Issues in calibration
    1. Reference base
    2. Reference standards
  2. What is artifact (single-point) calibration?
  3. What are calibration designs?
    1. Elimination of special types of bias
      1. Left-right (constant instrument) bias
      2. Bias caused by instrument drift
    2. Solutions to calibration designs
      1. General matrix solutions to calibration designs
    3. Uncertainties of calibrated values
      1. Type A evaluations for calibration designs
      2. Repeatability and level-2 standard deviations
      3. Combination of repeatability and level-2 standard deviations
      4. Calculation of standard deviations for 1,1,1,1 design
      5. Type B uncertainty
      6. Expanded uncertainties
  4. Catalog of calibration designs
    1. Mass weights
      1. Design for 1,1,1
      2. Design for 1,1,1,1
      3. Design for 1,1,1,1,1
      4. Design for 1,1,1,1,1,1
      5. Design for 2,1,1,1
      6. Design for 2,2,1,1,1
      7. Design for 2,2,2,1,1
      8. Design for 5,2,2,1,1,1
      9. Design for 5,2,2,1,1,1,1
      10. Design for 5,3,2,1,1,1
      11. Design for 5,3,2,1,1,1,1
      12. Design for 5,3,2,2,1,1,1
      13. Design for 5,4,4,3,2,2,1,1
      14. Design for 5,5,2,2,1,1,1,1
      15. Design for 5,5,3,2,1,1,1,1
      16. Design for 1,1,1,1,1,1,1,1 weights
      17. Design for 3,2,1,1,1,1 weights
      18. Design for 10 and 20 pound weights
    2. Drift-elimination designs for gage blocks
      1. Doiron 3-6 Design
      2. Doiron 3-9 Design
      3. Doiron 4-8 Design
      4. Doiron 4-12 Design
      5. Doiron 5-10 Design
      6. Doiron 6-12 Design
      7. Doiron 7-14 Design
      8. Doiron 8-16 Design
      9. Doiron 9-18 Design
      10. Doiron 10-20 Design
      11. Doiron 11-22 Design
    3. Designs for electrical quantities
      1. Left-right balanced design for 3 standard cells
      2. Left-right balanced design for 4 standard cells
      3. Left-right balanced design for 5 standard cells
      4. Left-right balanced design for 6 standard cells
      5. Left-right balanced design for 4 references and 4 test items
      6. Design for 8 references and 8 test items
      7. Design for 4 reference zeners and 2 test zeners
      8. Design for 4 reference zeners and 3 test zeners
      9. Design for 3 references and 1 test resistor
      10. Design for 4 references and 1 test resistor
    4. Roundness measurements
      1. Single trace roundness design
      2. Multiple trace roundness designs
    5. Designs for angle blocks
      1. Design for 4 angle blocks
      2. Design for 5 angle blocks
      3. Design for 6 angle blocks
    6. Thermometers in a bath
    7. Humidity standards
      1. Drift-elimination design for 2 reference weights and 3 cylinders
  5. Control of artifact calibration
    1. Control of precision
      1. Example of control chart for precision
    2. Control of bias and long-term variability
      1. Example of Shewhart control chart for mass calibrations
      2. Example of EWMA control chart for mass calibrations
  6. Instrument calibration over a regime
    1. Models for instrument calibration
    2. Data collection
    3. Assumptions for instrument calibration
    4. What can go wrong with the calibration procedure
      1. Example of day-to-day changes in calibration
    5. Data analysis and model validation
      1. Data on load cell #32066
    6. Calibration of future measurements
    7. Uncertainties of calibrated values
      1. Uncertainty for quadratic calibration using propagation of error
      2. Uncertainty for linear calibration using check standards
      3. Comparison of check standard analysis and propagation of error
  7. Instrument control for linear calibration
    1. Control chart for a linear calibration line
4. Gauge R & R studies
  1. What are the important issues?
  2. Design considerations
  3. Data collection for time-related sources of variability
    1. Simple design
    2. 2-level nested design
    3. 3-level nested design
  4. Analysis of variability
    1. Analysis of repeatability
    2. Analysis of reproducibility
    3. Analysis of stability
      1. Example of calculations
  5. Analysis of bias
    1. Resolution
    2. Linearity of the gauge
    3. Drift
    4. Differences among gauges
    5. Geometry/configuration differences
    6. Remedial actions and strategies
  6. Quantifying uncertainties from a gauge study

5. Uncertainty analysis
  1. Issues
  2. Approach
    1. Steps
  3. Type A evaluations
    1. Type A evaluations of random components
      1. Type A evaluations of time-dependent effects
      2. Measurement configuration within the laboratory
    2. Material inhomogeneity
      1. Data collection and analysis
    3. Type A evaluations of bias
      1. Inconsistent bias
      2. Consistent bias
      3. Bias with sparse data
  4. Type B evaluations
    1. Standard deviations from assumed distributions
  5. Propagation of error considerations
    1. Formulas for functions of one variable
    2. Formulas for functions of two variables
    3. Propagation of error for many variables
  6. Uncertainty budgets and sensitivity coefficients
    1. Sensitivity coefficients for measurements on the test item
    2. Sensitivity coefficients for measurements on a check standard
    3. Sensitivity coefficients for measurements from a 2-level design
    4. Sensitivity coefficients for measurements from a 3-level design
    5. Example of uncertainty budget
  7. Standard and expanded uncertainties
    1. Degrees of freedom
  8. Treatment of uncorrected bias
    1. Computation of revised uncertainty
6. Case studies
  1. Gauge study of resistivity probes
    1. Background and data
      1. Database of resistivity measurements
    2. Analysis and interpretation
    3. Repeatability standard deviations
    4. Effects of days and long-term stability
    5. Differences among 5 probes
    6. Run gauge study example using Dataplot
    7. Dataplot macros
  2. Check standard for resistivity measurements
    1. Background and data
      1. Database for resistivity check standard
    2. Analysis and interpretation
      1. Repeatability and level-2 standard deviations
    3. Control chart for probe precision
    4. Control chart for bias and long-term variability
    5. Run check standard example yourself
    6. Dataplot macros
  3. Evaluation of type A uncertainty
    1. Background and data
      1. Database of resistivity measurements
      2. Measurements on wiring configurations
    2. Analysis and interpretation
      1. Difference between 2 wiring configurations
    3. Run the type A uncertainty analysis using Dataplot
    4. Dataplot macros
  4. Evaluation of type B uncertainty and propagation of error
7. References

=====

3. Production Process Characterization
  1. Introduction to Production Process Characterization
    1. What is PPC?
    2. What are PPC Studies Used For?
  3. Terminology/Concepts
    1. Distribution (Location, Spread and Shape)
    2. Process Variability
      1. Controlled/Uncontrolled Variation
    3. Propagating Error
    4. Populations and Sampling
    5. Process Models
    6. Experiments and Experimental Design
  4. PPC Steps
2. Assumptions / Prerequisites
  1. General Assumptions
  2. Continuous Linear Model
  3. Analysis of Variance Models (ANOVA)
    1. One-Way ANOVA
      1. One-Way Value-Splitting
    2. Two-Way Crossed ANOVA
      1. Two-way Crossed Value-Splitting Example
    3. Two-Way Nested ANOVA
      1. Two-Way Nested Value-Splitting Example
  4. Discrete Models
3. Data Collection for PPC
  1. Define Goals
  2. Process Modeling
  3. Define Sampling Plan
    1. Identifying Parameters, Ranges and Resolution
    2. Choosing a Sampling Scheme
    3. Selecting Sample Sizes
    4. Data Storage and Retrieval
    5. Assign Roles and Responsibilities
4. Data Analysis for PPC
  1. First Steps
  2. Exploring Relationships
    1. Response Correlations
    2. Exploring Main Effects
    3. Exploring First Order Interactions
  3. Building Models
    1. Fitting Polynomial Models
    2. Fitting Physical Models
  4. Analyzing Variance Structure
  5. Assessing Process Stability
  6. Assessing Process Capability
  7. Checking Assumptions
5. Case Studies
  1. Furnace Case Study
    1. Background and Data
    2. Initial Analysis of Response Variable
    3. Identify Sources of Variation
    4. Analysis of Variance
    5. Final Conclusions
    6. Work This Example Yourself
  2. Machine Screw Case Study
    1. Background and Data
    2. Box Plots by Factors
    3. Analysis of Variance

4. Throughput
5. Final Conclusions
6. Work This Example Yourself

## 6. References

### 4. Process Modeling - Detailed Table of Contents

1. Introduction to Process Modeling
  1. What is process modeling?
  2. What terminology do statisticians use to describe process models?
  3. What are process models used for?
    1. Estimation
    2. Prediction
    3. Calibration
    4. Optimization
  4. What are some of the different statistical methods for model building?
    1. Linear Least Squares Regression
    2. Nonlinear Least Squares Regression
    3. Weighted Least Squares Regression
    4. LOESS (aka LOWESS)

The following example shows how to use Dataplot to obtain least squares estimates for data generated from a Weibull distribution.

```
. Generate some Weibull data
SET MINMAX MIN
LET GAMMA = 5
LET Y = WEIBULL RAND NUMB FOR I = 1 1 1000
. Bin the data
SET RELATIVE HISTOGRAM AREA
RELATIVE HISTOGRAM Y
LET ZY = YPLOT
LET ZX = XPLOT
RETAIN ZY ZX SUBSET YPLOT > 0
. Specify some starting values
LET SHAPE = 3
LET LOC = MINIMUM Y
LET SCALE = 1
. Now perform the least squares fit
FIT ZY = WEIPDF(ZX,SHAPE,LOC,SCALE)
```

The RELATIVE HISTOGRAM generates a relative histogram.

The command SET RELATIVE HISTOGRAM specifies that the relative histogram is created so that the area under the histogram is 1 (i.e., the integral is 1) rather than the sum of the bars equaling 1.

This effectively makes the relative histogram an estimator of the underlying density function.

Dataplot saves the coordinates of the histogram in the internal variables XPLOT and YPLOT.

The SUBSET command eliminates zero frequency classes.

The FIT command then performs the least squares fit.

The same general approach can be used to compute least squares estimates for any distribution for which Dataplot has a pdf function.

The primary difficulty with the least squares fitting is that it can be quite sensitive to starting values.

For distributions with no shape parameters,

the probability plot can be used to determine starting values for the location and scale parameters.

For distributions with a single shape parameter,

the ppcc plot can be used to determine a starting value for the shape parameter and the probability plot used to determine starting values for the location and scale parameters.

The approach above can be used in any statistical software package that provides non-linear least squares fitting and a method for defining the probability density function (either built-in or user definable).

### 2. Underlying Assumptions for Process Modeling

1. What are the typical underlying assumptions in process modeling?
  1. The process is a statistical process.
  2. The means of the random errors are zero.
  3. The random errors have a constant standard deviation.
  4. The random errors follow a normal distribution.
  5. The data are randomly sampled from the process.
  6. The explanatory variables are observed without error.

### 3. Data Collection for Process Modeling

1. What is design of experiments (DOE)?
2. Why is experimental design important for process modeling?
3. What are some general design principles for process modeling?
4. I've heard some people refer to "optimal" designs, shouldn't I use those?
5. How can I tell if a particular experimental design is good for my application?

### 4. Data Analysis for Process Modeling

1. What are the basic steps for developing an effective process model?
2. How do I select a function to describe my process?
  1. Incorporating Scientific Knowledge into Function Selection
  2. Using the Data to Select an Appropriate Function
  3. Using Methods that Do Not Require Function Specification
3. How are estimates of the unknown parameters obtained?
  1. Least Squares
  2. Weighted Least Squares
4. How can I tell if a model fits my data?
  1. How can I assess the sufficiency of the functional part of the model?
  2. How can I detect non-constant variation across the data?
  3. How can I tell if there was drift in the measurement process?
  4. How can I assess whether the random errors are independent from one to the next?
  5. How can I test whether or not the random errors are distributed normally?
  6. How can I test whether any significant terms are missing or misspecified in the functional part of the model?
  7. How can I test whether all of the terms in the functional part of the model are necessary?
5. If my current model does not fit the data well, how can I improve it?
  1. Updating the Function Based on Residual Plots
  2. Accounting for Non-Constant Variation Across the Data
  3. Accounting for Errors with a Non-Normal Distribution

### 5. Use and Interpretation of Process Models

1. What types of predictions can I make using the model?
  1. How do I estimate the average response for a particular set of predictor variable values?
  2. How can I predict the value and estimate the uncertainty of a single response?
2. How can I use my process model for calibration?
  1. Single-Use Calibration Intervals
3. How can I optimize my process using the process model?

### 6. Case Studies in Process Modeling

1. Load Cell Calibration
  1. Background & Data
  2. Selection of Initial Model
  3. Model Fitting - Initial Model
  4. Graphical Residual Analysis - Initial Model
  5. Interpretation of Numerical Output - Initial Model
  6. Model Refinement
  7. Model Fitting - Model #2
  8. Graphical Residual Analysis - Model #2
  9. Interpretation of Numerical Output - Model #2
  10. Use of the Model for Calibration
  11. Work This Example Yourself
2. Alaska Pipeline
  1. Background and Data
  2. Check for Batch Effect
  3. Initial Linear Fit
  4. Transformations to Improve Fit and Equalize Variances
  5. Weighting to Improve Fit
  6. Compare the Fits
  7. Work This Example Yourself
3. Ultrasonic Reference Block Study



1. Background and Data
  2. Initial Non-Linear Fit
  3. Transformations to Improve Fit
  4. Weighting to Improve Fit
  5. Compare the Fits
  6. Work This Example Yourself
  4. Thermal Expansion of Copper Case Study
    1. Background and Data
    2. Rational Function Models
    3. Initial Plot of Data
    4. Quadratic/Quadratic Rational Function Model
    5. Cubic/Cubic Rational Function Model
    6. Work This Example Yourself
  7. References For Chapter 4: Process Modeling
  8. Some Useful Functions for Process Modeling
    1. Univariate Functions
      1. Polynomial Functions
        1. Straight Line
        2. Quadratic Polynomial
        3. Cubic Polynomial
      2. Rational Functions
        1. Constant / Linear Rational Function
        2. Linear / Linear Rational Function
        3. Linear / Quadratic Rational Function
        4. Quadratic / Linear Rational Function
        5. Quadratic / Quadratic Rational Function
        6. Cubic / Linear Rational Function
        7. Cubic / Quadratic Rational Function
        8. Linear / Cubic Rational Function
        9. Quadratic / Cubic Rational Function
        10. Cubic / Cubic Rational Function
        11. Determining m and n for Rational Function Models
- 
5. Process Improvement
    1. Introduction
      1. What is experimental design?
      2. What are the uses of DOE?
      3. What are the steps of DOE?
    2. Assumptions
      1. Is the measurement system capable?
      2. Is the process stable?
      3. Is there a simple model?
      4. Are the model residuals well-behaved?
    3. Choosing an experimental design
      1. What are the objectives?
      2. How do you select and scale the process variables?
      3. How do you select an experimental design?
        1. Completely randomized designs
        2. Randomized block designs
          1. Latin square and related designs
          2. Graeco-Latin square designs
          3. Hyper-Graeco-Latin square designs
        3. Full factorial designs
          1. Two-level full factorial designs
          2. Full factorial example
          3. Blocking of full factorial designs
        4. Fractional factorial designs
          1. A 2<sup>3</sup>-1 design (half of a 2<sup>3</sup>)
          2. Constructing the 2<sup>3</sup>-1 half-fraction design
          3. Confounding (also called aliasing)
          4. Fractional factorial design specifications and design resolution
          5. Use of fractional factorial designs
          6. Screening designs
          7. Summary tables of useful fractional factorial designs
        5. Plackett-Burman designs
        6. Response surface designs
          1. Central Composite Designs (CCD)
          2. Box-Behnken designs
          3. Comparisons of response surface designs
          4. Blocking a response surface design
        7. Adding centerpoints
        8. Improving fractional factorial design resolution
          1. Mirror-Image foldover designs
          2. Alternate foldover designs
        9. Three-level full factorial designs
        10. Three-level, mixed-level and fractional factorial designs
    4. Analysis of DOE data
      1. What are the steps in a DOE analysis?
      2. How to "look" at DOE data
      3. How to model DOE data
      4. How to test and revise DOE models
      5. How to interpret DOE results
      6. How to confirm DOE results (confirmatory runs)
      7. Examples of DOE's
        1. Full factorial example
        2. Fractional factorial example
        3. Response surface model example
    5. Advanced topics
      1. What if classical designs don't work?
      2. What is a computer-aided design?
        1. D-Optimal designs
        2. Repairing a design
      3. How do you optimize a process?
        1. Single response case
          1. Single response: Path of steepest ascent
          2. Single response: Confidence region for search path
          3. Single response: Choosing the step length
          4. Single response: Optimization when there is adequate quadratic fit
          5. Single response: Effect of sampling error on optimal solution
          6. Single response: Optimization subject to experimental region constraints
        2. Multiple response case
          1. Multiple responses: Path of steepest ascent
          2. Multiple responses: The desirability approach
          3. Multiple responses: The mathematical programming approach
      4. What is a mixture design?
        1. Mixture screening designs
        2. Simplex-lattice designs
        3. Simplex-centroid designs
        4. Constrained mixture designs
        5. Treating mixture and process variables together
      5. How can I account for nested variation (restricted randomization)?
      6. What are Taguchi designs?
      7. What are John's 3/4 fractional factorial designs?
      8. What are small composite designs?
      9. An EDA approach to experimental design
        1. Ordered data plot
        2. DOE scatter plot
        3. DOE mean plot

4. Interaction effects matrix plot  
 5. Block plot  
     BLOCK PLOT Y X1 X2 X3 etc. XP  
     where  
     Y is the response variable,  
     X1, X2, X3, etc. are the one or more nuisance (= secondary) factors, and  
     XP is the primary factor of interest.  
     The following commands typically precede the block plot.  
     CHARACTER 1 2  
     LINE BLANK BLANK  
     These commands set the plot character for the primary factor.  
 Although 1 and 2 are useful indicators,  
 the choice of plot character is at the discretion of the user.

6. DOE Youden plot
7. |Effects| plot
  1. Statistical significance
  2. Engineering significance
  3. Numerical significance
  4. Pattern significance
8. Half-normal probability plot
9. Cumulative residual standard deviation plot
  1. Motivation: What is a Model?
  2. Motivation: How do we Construct a Goodness-of-fit Metric for a Model?
  3. Motivation: How do we Construct a Good Model?
  4. Motivation: How do we Know When to Stop Adding Terms?
  5. Motivation: What is the Form of the Model?
  6. Motivation: What are the Advantages of the LinearCombinatoric Model?
  7. Motivation: How do we use the Model to Generate Predicted Values?
  8. Motivation: How do we Use the Model Beyond the Data Domain?
  9. Motivation: What is the Best Confirmation Point for Interpolation?
  10. Motivation: How do we Use the Model for Interpolation?
  11. Motivation: How do we Use the Model for Extrapolation?
10. DOE contour plot

The Dataplot command for generating a linear dex contour plot is  
 DEX CONTOUR PLOT Y X1 X2 Y0  
 The variables X1 and X2 are the two factor variables,  
 Y is the response variable,  
 and Y0 defines the desired contour levels.  
 Dataplot does not have a built-in quadratic dex contour plot.

However,  
 the macro DEXCONTQ.DP will generate a quadratic dex contour plot.  
 Enter LIST DEXCONTQ.DP for more information.

1. How to Interpret: Axes
2. How to Interpret: Contour Curves
3. How to Interpret: Optimal Response Value
4. How to Interpret: Best Corner
5. How to Interpret: Steepest Ascent/Descent
6. How to Interpret: Optimal Curve
7. How to Interpret: Optimal Setting

6. Case Studies
  1. Eddy Current Probe Sensitivity Case Study
    1. Background and Data
    2. Initial Plots/Main Effects
    3. Interaction Effects
    4. Main and Interaction Effects: Block Plots
    5. Estimate Main and Interaction Effects
    6. Modeling and Prediction Equations
    7. Intermediate Conclusions
    8. Important Factors and Parsimonious Prediction
    9. Validate the Fitted Model
    10. Using the Fitted Model
    11. Conclusions and Next Step
    12. Work This Example Yourself
  2. Sonoluminescent Light Intensity Case Study
    1. Background and Data
    2. Initial Plots/Main Effects
    3. Interaction Effects
    4. Main and Interaction Effects: Block Plots
    5. Important Factors: Youden Plot
    6. Important Factors: |Effects| Plot
    7. Important Factors: Half-Normal Probability Plot
    8. Cumulative Residual Standard Deviation Plot
    9. Next Step: DOE Contour Plot
    10. Summary of Conclusions
    11. Work This Example Yourself

#### 7. A Glossary of DOE Terminology

#### 8. References

#### 6. Process or Product Monitoring and Control

1. Introduction
  1. How did Statistical Quality Control Begin?
  2. What are Process Control Techniques?
  3. What is Process Control?
  4. What to do if the process is "Out of Control"?
  5. What to do if "In Control" but Unacceptable?
  6. What is Process Capability?
2. Test Product for Acceptability: Lot Acceptance Sampling
  1. What is Acceptance Sampling?
  2. What kinds of Lot Acceptance Sampling Plans (LASPs) are there?
  3. How do you Choose a Single Sampling Plan?
    1. Choosing a Sampling Plan: MIL Standard 105D
    2. Choosing a Sampling Plan with a given OC Curve
  4. What is Double Sampling?
  5. What is Multiple Sampling?
  6. What is a Sequential Sampling Plan?
  7. What is Skip Lot Sampling?
3. Univariate and Multivariate Control Charts
  1. What are Control Charts?
  2. What are Variables Control Charts?
    1. Shewhart X-bar and R and S Control Charts
    2. Individuals Control Charts
    3. Cusum Control Charts
      1. Cusum Average Run Length
    4. EWMA Control Charts
  3. What are Attributes Control Charts?
    1. Counts Control Charts
    2. Proportions Control Charts
  4. What are Multivariate Control Charts?
    1. Hotelling Control Charts
    2. Principal Components Control Charts
    3. Multivariate EWMA Charts
4. Introduction to Time Series Analysis
  1. Definitions, Applications and Techniques
  2. What are Moving Average or Smoothing Techniques?
    1. Single Moving Average
    2. Centered Moving Average

3. What is Exponential Smoothing?
  1. Single Exponential Smoothing
  2. Forecasting with Single Exponential Smoothing
  3. Double Exponential Smoothing
  4. Forecasting with Double Exponential Smoothing(LASP)
  5. Triple Exponential Smoothing
  6. Example of Triple Exponential Smoothing
  7. Exponential Smoothing Summary
4. Univariate Time Series Models
  1. Sample Data Sets
    1. Data Set of Monthly CO2 Concentrations
    2. Data Set of Southern Oscillations
  2. Stationarity
  3. Seasonality
    1. Seasonal Subseries Plot
 

```
LET PERIOD = <value>
LET START = <value>
SEASONAL SUBSERIES PLOT Y
```

The value of PERIOD defines the length of the seasonal period (e.g., 12 for monthly data) and START identifies which group the series starts with (e.g., if you have monthly data that starts in March, set START to 3).

The appearance of the plot can be controlled by appropriate settings of the CHARACTER and LINE commands and their various attribute-setting commands.

  4. Common Approaches to Univariate Time Series
  5. Box-Jenkins Models
  6. Box-Jenkins Model Identification
    1. Model Identification for Southern Oscillations Data
    2. Model Identification for the CO2 Concentrations Data
    3. Partial Autocorrelation Plot
 

The command to generate a partial autocorrelation plot is

```
PARTIAL AUTOCORRELATION PLOT Y
```

The appearance of the partial autocorrelation plot can be controlled by appropriate settings of the LINE, CHARACTER, and SPIKE commands.

Dataplot draws the following curves on the autocorrelation plot:

The autocorrelations.

A reference line at zero.

A reference line at the upper 95% confidence limit.

A reference line at the lower 95% confidence limit.

A reference line at the upper 99% confidence limit.

A reference line at the lower 99% confidence limit.

For example, to draw the partial autocorrelations as spikes, the zero reference line as a solid line, the 95% lines as dashed lines, and the 99% line as dotted lines, enter the command

```
LINE BLANK SOLID DASH DASH DOT DOT
CHARACTER BLANK ALL
SPIKE ON OFF OFF OFF OFF OFF
SPIKE BASE 0
```

  7. Box-Jenkins Model Estimation
  8. Box-Jenkins Model Diagnostics
    1. Box-Ljung Test
  9. Example of Univariate Box-Jenkins Analysis
  10. Box-Jenkins Analysis on Seasonal Data
5. Multivariate Time Series Models
  1. Example of Multivariate Time Series Analysis
5. Tutorials
  1. What do we mean by "Normal" data?
  2. What do we do when data are "Non-normal"?
  3. Elements of Matrix Algebra
    1. Numerical Examples
    2. Determinant and Eigenstructure
  4. Elements of Multivariate Analysis
    1. Mean Vector and Covariance Matrix
    2. The Multivariate Normal Distribution
    3. Hotelling's T squared
      1. T2 Chart for Subgroup Averages -- Phase I
      2. T2 Chart for Subgroup Averages -- Phase II
      3. Chart for Individual Observations -- Phase I
      4. Chart for Individual Observations -- Phase II
      5. Charts for Controlling Multivariate Variability
      6. Constructing Multivariate Charts
  5. Principal Components
    1. Properties of Principal Components
    2. Numerical Example
6. Case Studies in Process Monitoring
  1. Lithography Process
    1. Background and Data
    2. Graphical Representation of the Data
    3. Subgroup Analysis
    4. Shewhart Control Chart
    5. Work This Example Yourself
  2. Aerosol Particle Size
    1. Background and Data
    2. Model Identification
    3. Model Estimation
    4. Model Validation
    5. Work This Example Yourself
7. References

---

7. Product and Process Comparisons
  1. Introduction
    1. What is the scope?
    2. What assumptions are typically made?
    3. What are statistical tests?
      1. Critical values and p values
    4. What are confidence intervals?
    5. What is the relationship between a test and a confidence interval?
    6. What are outliers in the data?
    7. What are trends in sequential process or product data?
  2. Comparisons based on data from one process
    1. Do the observations come from a particular distribution?
      1. Chi-square goodness-of-fit test
      2. Kolmogorov-Smirnov test
      3. Anderson-Darling and Shapiro-Wilk tests
    2. Are the data consistent with the assumed process mean?
      1. Confidence interval approach
      2. Sample sizes required
    3. Are the data consistent with a nominal standard deviation?
      1. Confidence interval approach
      2. Sample sizes required
    4. Does the proportion of defectives meet requirements?
      1. Confidence intervals
      2. Sample sizes required
    5. Does the defect density meet requirements?
    6. What intervals contain a fixed percentage of the population values?
      1. Approximate intervals that contain most of the population values
      2. Percentiles
      3. Tolerance intervals for a normal distribution

4. Tolerance intervals based on the largest and smallest observations

The Dataplot command for tolerance intervals is

TOLERANCE Y

where Y is the response variable. Both normal and nonparametric tolerance intervals are printed.

3. Comparisons based on data from two processes

1. Do two processes have the same mean?
  1. Analysis of paired observations
  2. Confidence intervals for differences between means
2. Do two processes have the same standard deviation?
3. How can we determine whether two processes produce the same proportion of defectives?
4. Assuming the observations are failure times, are the failure rates (or Mean Times To Failure) for two distributions the same?
5. Do two arbitrary processes have the same central tendency?

4. Comparisons based on data from more than two processes

1. How can we compare several populations with unknown distributions (the Kruskal-Wallis test)?

KRUSKAL WALLIS TEST Y X

where Y is the response variable and X is the group identifier variable.
2. Assuming the observations are normal, do the processes have the same variance?
3. Are the means equal?
  1. 1-Way ANOVA overview
  2. The 1-way ANOVA model and assumptions
  3. The ANOVA table and tests of hypotheses about means
  4. 1-Way ANOVA calculations
  5. Confidence intervals for the difference of treatment means
  6. Assessing the response from any factor combination
  7. The two-way ANOVA
  8. Models and calculations for the two-way ANOVA
4. What are variance components?
5. How can we compare the results of classifying according to several categories?
6. Do all the processes have the same proportion of defects?
7. How can we make multiple comparisons?
  1. Tukey's method
  2. Scheffe's method
  3. Bonferroni's method
  4. Comparing multiple proportions: The Marascuillo procedure

5. References

=====

8. Assessing Product Reliability

1. Introduction

1. Why is the assessment and control of product reliability important?
  1. Quality versus reliability
  2. Competitive driving factors
  3. Safety and health considerations
2. What are the basic terms and models used for reliability evaluation?
  1. Repairable systems, non-repairable populations and lifetime distribution models
  2. Reliability or survival function
  3. Failure (or hazard) rate
  4. "Bathtub" curve
  5. Repair rate or ROCOF
3. What are some common difficulties with reliability data and how are they overcome?
  1. Censoring
  2. Lack of failures
4. What is "physical acceleration" and how do we model it?
5. What are some common acceleration models?
  1. Arrhenius
  2. Eyring
  3. Other models
6. What are the basic lifetime distribution models used for non-repairable populations?
  1. Exponential
  2. Weibull
  3. Extreme value distributions
  4. Lognormal
  5. Gamma
  6. Fatigue life (Birnbaum-Saunders)
  7. Proportional hazards model
7. What are some basic repair rate models used for repairable systems?
  1. Homogeneous Poisson Process (HPP)
  2. Non-Homogeneous Poisson Process (NHPP) - power law
  3. Exponential law
8. How can you evaluate reliability from the "bottom-up" (component failure mode to system failure rate)?
  1. Competing risk model
  2. Series model
  3. Parallel or redundant model
  4. R out of N model
  5. Standby model
  6. Complex systems
9. How can you model reliability growth?
  1. NHPP power law
  2. Duane plots

Dataplot Command for Duane Plot

DUANE PLOT Y

where Y is a response variable containing failure times.

  3. NHPP exponential law
10. How can Bayesian methodology be used for reliability evaluation?

2. Assumptions/Prerequisites

1. How do you choose an appropriate life distribution model?
  1. Based on failure mode
  2. Extreme value argument
  3. Multiplicative degradation argument
  4. Fatigue life (Birnbaum-Saunders) model
  5. Empirical model fitting - distribution free (Kaplan-Meier) approach
2. How do you plot reliability data?
  1. Probability plotting
  2. Hazard and cum hazard plotting

EXPONENTIAL HAZARD PLOT Y X

NORMAL HAZARD PLOT Y X

LOGNORMAL HAZARD PLOT Y X

WEIBULL HAZARD PLOT Y X

where Y is a response variable containing failure times and X is a censoring variable (0 means failure time, 1 means censoring time).

  3. Trend and growth plotting (Duane plots)
3. How can you test reliability model assumptions?
  1. Visual tests
  2. Goodness of fit tests
  3. Likelihood ratio tests
  4. Trend tests
4. How do you choose an appropriate physical acceleration model?
5. What models and assumptions are typically made when Bayesian methods are used for reliability evaluation?

3. Reliability Data Collection

1. How do you plan a reliability assessment test?
  1. Exponential life distribution (or HPP model) tests
  2. Lognormal or Weibull tests
  3. Reliability growth (Duane model)
  4. Accelerated life tests
  5. Bayesian gamma prior model

#### 4. Reliability Data Analysis

##### 1. How do you estimate life distribution parameters from censored data?

###### 1. Graphical estimation

###### 2. Maximum likelihood estimation

Dataplot performs maximum likelihood estimation for a few specific distributions as documented in the table below.

Unless specified otherwise, censored data are not supported and only point estimates are generated (i.e., no confidence intervals for the parameters).

For censored data, create an id variable that is equal to 1 for a failure time and equal to 0 for a censoring time.

Type I censoring is censoring at a fixed time  $t_0$ .

Type II censoring is censoring after a pre-determined number of units have failed.

Normal

NORMAL MAXIMUM LIKELIHOOD Y

Exponential

EXPONENTIAL MAXIMUM LIKELIHOOD Y

Confidence intervals are generated for the parameters and both type I and type II censoring are supported.

For type I censoring, enter the following commands

SET CENSORING TYPE 1

LET TEND = censoring time

EXPONENTIAL MAXIMUM LIKELIHOOD Y ID

For type II censoring, enter the following commands

SET CENSORING TYPE 2

EXPONENTIAL MAXIMUM LIKELIHOOD Y ID

Weibull

WEIBULL MAXIMUM LIKELIHOOD Y

Confidence intervals are generated for the parameters and both type I and type II censoring are supported.

For type I censoring, enter the following commands

SET CENSORING TYPE 1

LET TEND = censoring time

WEIBULL MAXIMUM LIKELIHOOD Y ID

For type II censoring, enter the following commands

SET CENSORING TYPE 2

WEIBULL MAXIMUM LIKELIHOOD Y ID

Lognormal

LOGNORMAL MAXIMUM LIKELIHOOD Y

Double

Exponential

DOUBLE EXPONENTIAL MAXIMUM LIKELIHOOD Y

Pareto

PARETO MAXIMUM LIKELIHOOD Y

Gamma

GAMMA MAXIMUM LIKELIHOOD Y

Inverse

Gaussian

INVERSE GAUSSIAN MAXIMUM LIKELIHOOD Y

Gumbel

GUMBEL MAXIMUM LIKELIHOOD Y

Binomial

BINOMIAL MAXIMUM LIKELIHOOD Y

Poisson

POISSON MAXIMUM LIKELIHOOD Y

##### 3. A Weibull maximum likelihood estimation example

##### 2. How do you fit an acceleration model?

###### 1. Graphical estimation

###### 2. Maximum likelihood

###### 3. Fitting models using degradation data instead of failures

##### 3. How do you project reliability at use conditions?

##### 4. How do you compare reliability between two or more populations?

##### 5. How do you fit system repair rate models?

###### 1. Constant repair rate (HPP/exponential) model

###### 2. Power law (Duane) model

###### 3. Exponential law model

##### 6. How do you estimate reliability using the Bayesian gamma prior model?

##### 7. References For Chapter 8: Assessing Product Reliability