RELIABILITY PLOTTING

V7, August 11, 2014, by John Zorich (johnzorich@yahoo.com; www.johnzorich.com)

Reliability Plotting (see Tobias & Trindade, *Applied Reliability*, chapter 6) is a method for determining the shape of a distribution and using that knowledge to estimate what percentage of that population is "in specification" (that percentage is commonly referred to as the "% Reliability"). The distribution's shape is identified by means of a "Probability Plot" of the measurement ("variables") data derived from a random or representative sample taken from the population.

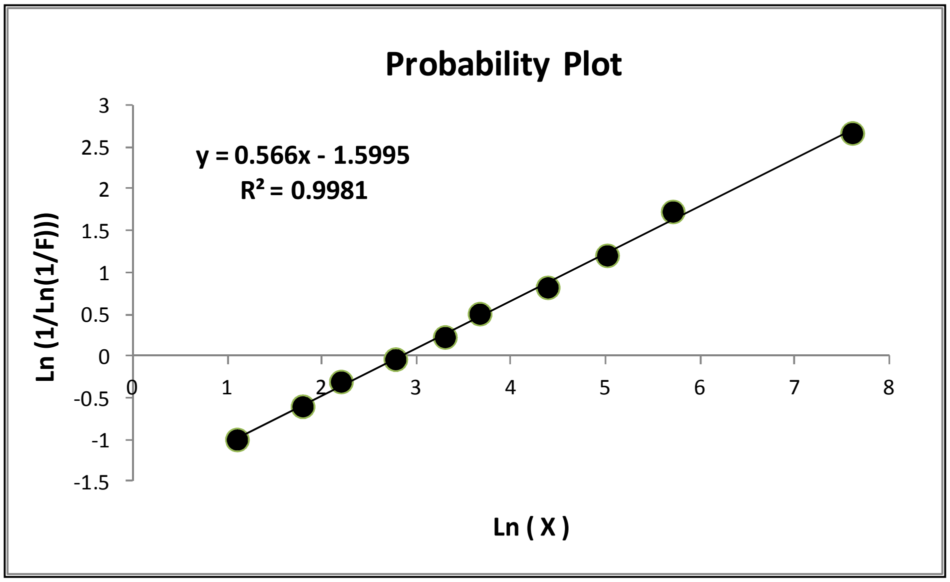
"Reliability Plots" are cumulative distribution probability plots that have had their Y and/or X-axis values transformed in such a way as to linearize the plotted points. Examples of commonly used transformations for the X-axis include square-root, natural log, or inverse; examples of commonly used transformations for the Y-axis include Ln(1/(1–F)), Ln(Ln(1/(1–F))), Ln(F/ (1–F), or Ln(1/(Ln(1/F))), where "F" is represents the %Cumulative expressed as a the "median rank" --- see formula below).

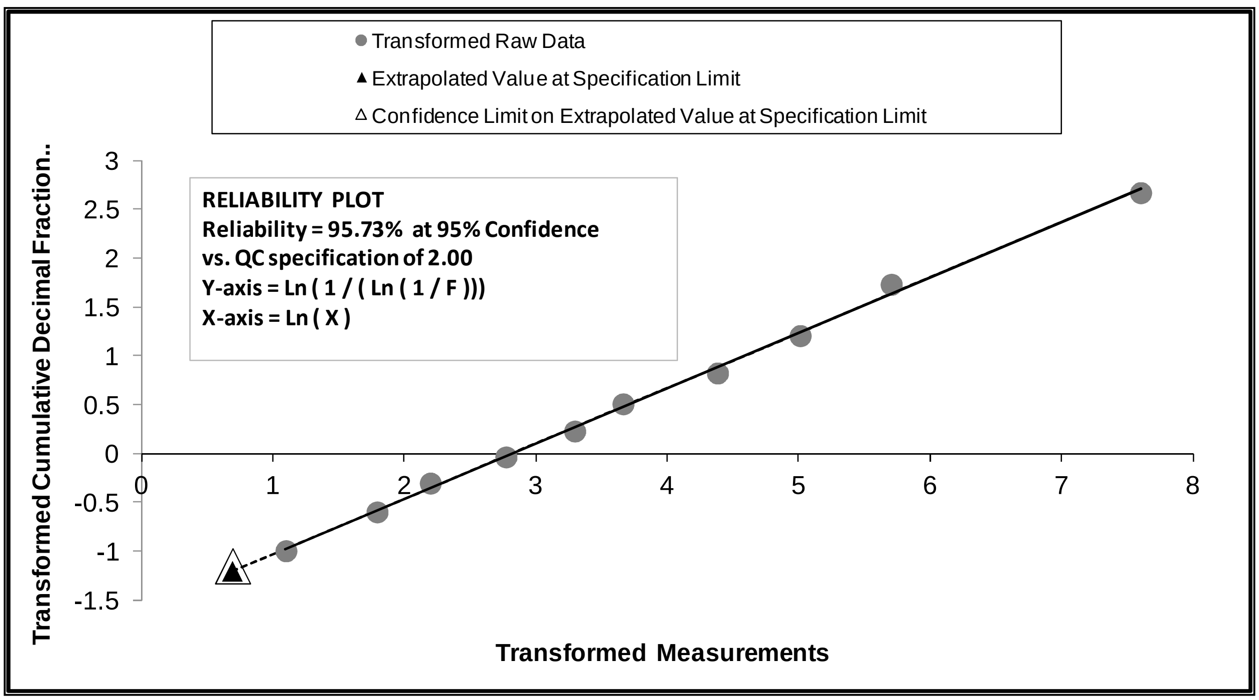
After such linearization, the methods of "linear regression" are employed to determine the transformed %Cumulative (on the Y-axis) that corresponds to the extrapolated transformed X-axis value for the QC specification. The upper 1-sided confidence limit is then calculated for that transformed Y-axis value; next, that confidence limit is reverse-transformed, yielding a value in units of untransformed %Cumulative; finally, that %Cumulative is subtracted from 100%, to yield the %Reliability at whatever confidence level was used to calculate the confidence limit.

For example, the distribution of the following N=10 data set is non-normal (based upon its "normal probability plot" being extremely curved).

3, 6, 9, 16, 27, 39, 80, 150, 300, 2000

Even though on a "**Log**-Normal Probability Plot" (i.e., where the X-axis is in Log units) that data produces a line with a high correlation coefficient (0.985), the plot has a definite curve to it (and therefore cannot be accurately extrapolated by linear regression). However, if the %Cumulative is transformed using Ln(1/(Ln(1/F))), then a probability plot of Ln(X) is very straight (see 1st chart below); and, on a Reliability Plot, can be extrapolated accurately to a QC specification of 2.00 (see 2nd chart below).


Probability Plot
y = 0.566x - 1.5995
R² = 0.9981
Ln (1/Ln(1/F))
Ln (X)
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
0
1
2
3
4
5
6
7
8


Transformed Raw Data
Extrapolated Value at Specification Limit
Confidence Limit on Extrapolated Value at Specification Limit
RELIABILITY PLOT
Reliability = 95.73% at 95% Confidence
vs. QC specification of 2.00
Y-axis = Ln (1 / (Ln (1 / F)))
X-axis = Ln (X)
Transformed Cumulative Decimal Fraction..
Transformed Measurements
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
0
1
2
3
4
5
6
7
8

That entire process is summarized in slightly more detail, in the following 6 sequential steps, for a one-sided lower specification limit:

1. The sample data values are charted onto a "cumulative probability plot", with the X-axis values being the raw data points arranged in numerical order with the lowest values on the left-hand side of the axis, and with the Y-axis values being the corresponding cumulative percentages (expressed as "F" values, i.e., the "median ranks" --- see formula below). The resulting plot typically looks like an "S" shaped curve, with the Y-axis decimal values ranging from near 0.00 to near 1.00 (which is the equivalent of a range from near 0% to near 100%).

2. "Transformations" are chosen for the Y-axis values and/or the X-axis values in attempts to straighten-out that S-shaped curve. See list of transformations, below; these transformations have been in use since reliability plotting was first invented, in the mid-20th century. In most (but not all) cases, the best straight line is the one that generates the highest Correlation Coefficient, as determined by the classic method of Linear Regression.

3. After the combination of X-and-Y axis transformations is found that provides the best straight line, that line is extrapolated to the point on the X-axis that corresponds to the QC or Verification/Validation pass/fail criterion (a.k.a., "spec limit").

4. The Y-axis value that corresponds to that extrapolated X-axis value then has its 1-sided confidence limit calculated (typically, a 95% confidence limit is chosen), using the classic method that is appropriate for the mean of a point on a Linear Regression line.

5. That confidence limit is then back-transformed. For example, if the Y-axis had been transformed by taking the Log(base 10) of every % cumulative value, then the back-transformation is achieved by taking the anti-log of the confidence limit (that is, using the confidence limit as a power on a base of 10). The resulting value represents the decimal % of the population that is out-of-specification.

6. That %-out-of-specification value is then subtracted from 100%, to yield the desired result, namely the % Reliability. It can then be claimed that, at the confidence level used in step 4 above, the % of the population that meets specification is no worse than the calculated % Reliability.

FORMULAS:

F = "median rank" = BETAINV (0.5 , rank , SampleSize – Rank + 1)
where "betainv" is the Excel function, and "rank" = 1 for the smallest number in the sample, "rank" = 2, for the next smallest, and so on to where "rank" = N for the largest value in the sample.

The generalized linear regression equation is... **Yei** = a + b • Xi. The following is the formula for calculating the 1-sided confidence limit on the plotted Y-value at a single point on a linear regression line (i.e., the transformed Y-value for the hollow triangle on the chart above).

= Ysl +/− t x See x [(1 / N)+((Xsl − Xavg)^2) / (Sum((Xi − Xavg)^2))]^0.5

where...

Ysl = Y-axis transformed “F” value corresponding to the Specification Limit
(i.e., transformed Y-value for solid triangle on the chart above)

+/− = use “+” if out-of-specification is below the spec limit; otherwise use “–”

t = one-side, t-Table value at alpha = 1 – Confidence, and df = N – 2
using Excel, t = TINV (2 * (1 – Confidence) , (N – 2))

N = number of X,Y points plotted on the chart ("N" is <u>not</u> the sample size; it is the number of dots on
the chart, not counting the solid and hollow triangles)

Xsl = transformed Specification Limit

Xavg = average of the transformed X values of the plotted X,Y points
(this is not the same as the average of the transformed raw data)
(do not include the specification limit in this average)

Xi = each of the transformed X values of the plotted X,Y points

"**Sum**" here means to add up each (Xi – Xavg)^2, from i = 1 thru i = N

See = Std Error of Estimate = [(Sum ((**Yei** – Yi) ^2)) / (N – 2)] ^0.5
Yi = transformed plotted “F” value corresponding to a plotted **Xi**

Lists for commonly used X-axis Transformations to linearity:

In the formulas below, A, B, D, and E are constants (negative or positive, whole numbers or fractional) chosen to help linearize the reliability plot.

= 1 / X

= SQRT (X)

= ASINH (SQRT (X))

= SQRT (X + A)

= ((X ^ B) – 1) / B   << this is called the Box-Cox Transformation

= LN (X + D)

= 1 / (X + E)

The following transformations can be used only with X values between 0 and 1:

= LN (X / (1 – X)   << this is called the Logit Transformation

= ASIN (SQRT (X))

= 0.5 * LN ((1 + X) / (1 – X)   << this is called the Fisher Transformation

Lists for commonly used Y-axis Transformations to linearity:

In the formulas below, “F” = the calculated Median Rank and
“C” = the user-chosen “shape parameter” constant
** = formula has been standardized by setting other shape parameters to a value of 1.000

= NORMSINV (F)   << this is the “Normal” (Z-table) transformation

= NORMSINV [1 − (1 − F) ^ (1 / C)]   << Power Normal

= EXP (NORMSINV (F))   << Three Parameter LogNormal **

= EXP (NORMSINV [1 − (1 − F) ^ (1 / C)]) << Power LogNormal **

= LN (LN (1 / (1 − F)))   << Smallest Extreme Value

= LN (1 / (1 − F)) ^ (1 / C)   << Weibull **

= LN (F / (1 − F))   << Logistic

= LN (1 / (LN (1 / F)))   << Largest Extreme Value

= TAN (PI() * (F − 0.5))   << Cauchy distribution

---

end of article

