

[HOME](#)

ENGINEERING STATISTICS HANDBOOK

[TOOLS & AIDS](#)[SEARCH](#)[BACK](#) [NEXT](#)

1. [Exploratory Data Analysis](#)
- 1.3. [EDA Techniques](#)
- 1.3.3. [Graphical Techniques: Alphabetic](#)

1.3.3.22. Probability Plot

*Purpose:
Check If
Data Follow
a Given
Distribution*

The probability plot ([Chambers 1983](#)) is a graphical technique for assessing whether or not a data set follows a given distribution such as the normal or Weibull.

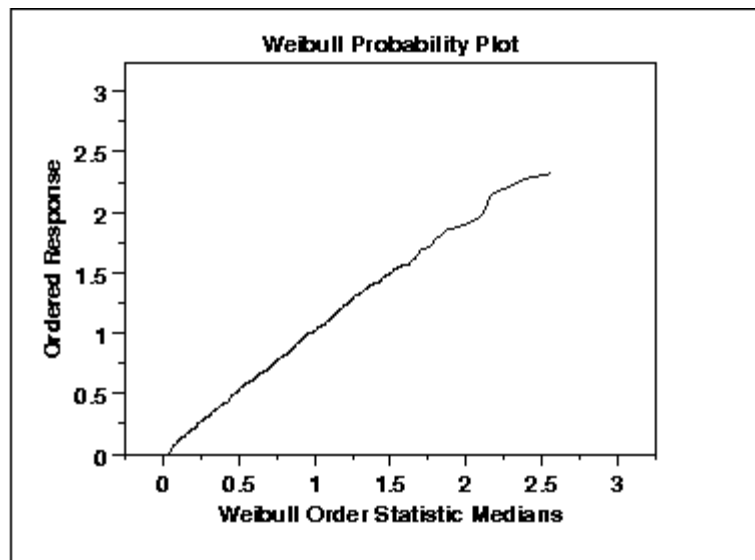
The data are plotted against a theoretical distribution in such a way that the points should form approximately a straight line. Departures from this straight line indicate departures from the specified distribution.

The correlation coefficient associated with the linear fit to the data in the probability plot is a measure of the goodness of the fit. Estimates of the [location and scale parameters](#) of the distribution are given by the intercept and slope. Probability plots can be generated for several competing distributions to see which provides the best fit, and the probability plot generating the highest correlation coefficient is the best choice since it generates the straightest probability plot.

For distributions with [shape parameters](#) (not counting location and scale parameters), the shape parameters must be known in order to generate the probability plot. For distributions with a single shape parameter, the [probability plot correlation coefficient](#) (PPCC) plot provides an excellent method for estimating the shape parameter.

We cover the special case of the [normal probability plot](#) separately due to its importance in many statistical applications.

Sample Plot



This data is a set of 500 [Weibull](#) random numbers with a shape parameter = 2, location parameter = 0, and scale parameter = 1. The Weibull probability plot indicates that the Weibull distribution does in fact fit these data well.

Definition:
Ordered
Response
Values
Versus Order
Statistic
Medians for
the Given
Distribution

The probability plot is formed by:

- Vertical axis: Ordered response values
- Horizontal axis: Order statistic medians for the given distribution

The order statistic medians are defined as:

$$N(i) = G(U(i))$$

where the $U(i)$ are the uniform order statistic medians (defined below) and G is the [percent point function](#) for the desired distribution. The percent point function is the inverse of the [cumulative distribution function](#) (probability that x is less than or equal to some value). That is, given a probability, we want the corresponding x of the cumulative distribution function.

The uniform order statistic medians are defined as:

$$m(i) = 1 - m(n) \text{ for } i = 1$$

$$m(i) = (i - 0.3175)/(n + 0.365) \text{ for } i = 2, 3, \dots, n-1$$

$$m(i) = 0.5^{**}(1/n) \text{ for } i = n$$

In addition, a straight line can be fit to the points and added as a reference line. The further the points vary from this line, the greater the indication of a departure from the specified distribution.

This definition implies that a probability plot can be easily generated for any distribution for which the percent point function can be computed.

One advantage of this method of computing probability plots is that the intercept and slope estimates of the fitted line are in fact estimates for the location and scale parameters of the distribution. Although this is not too important for the normal distribution (the location and scale are estimated by the mean and standard deviation, respectively), it can be useful for many other distributions.

Questions

The probability plot is used to answer the following questions:

- Does a given distribution, such as the Weibull, provide a good fit to my data?
- What distribution best fits my data?
- What are good estimates for the location and scale parameters of the chosen distribution?

Importance: Check distributional assumption

The discussion for the [normal probability plot](#) covers the use of probability plots for checking the fixed distribution assumption.

Some statistical models assume data have come from a population with a specific type of distribution. For example, in reliability applications, the Weibull, lognormal, and exponential are commonly used distributional models. Probability plots can be useful for checking this distributional assumption.

Related Techniques

[Histogram](#)
[Probability Plot Correlation Coefficient](#)

[\(PPCC\) Plot](#)
[Hazard Plot](#)
[Quantile-Quantile Plot](#)
[Anderson-Darling Goodness of Fit](#)
[Chi-Square Goodness of Fit](#)
[Kolmogorov-Smirnov Goodness of Fit](#)

Case Study The probability plot is demonstrated in the [airplane glass failure time](#) data case study.

Software Most general purpose statistical software programs support probability plots for at least a few common distributions. [Dataplot](#) supports probability plots for a large number of distributions.

NIST
SEMATECH

[HOME](#)

[TOOLS & AIDS](#)

[SEARCH](#)

[BACK](#) [NEXT](#)