PPLOT[1]

PPLOT produces probability plots of one or more sequence or time series variables. The variables can be standardized, differenced, and/or transformed before plotting. Expected normal values or deviations from expected normal values can be plotted. PPLOT can be used to investigate whether the data are from a specified distribution: normal, lognormal, logistic, exponential, Weibull, gamma, beta, uniform, Pareto, Laplace, half normal, chi-square and Student's *t*.

Notation

The following notation is used throughout this chapter unless otherwise stated:

$\overline{X}$	Sample mean
S	Sample standard deviation
$\overline{LX}$	Sample mean for $\ln(x_i)$
LS	Sample standard deviation for $\ln(x_i)$
x_i	Value of the *i*th observation
$x_{(i)}$	The *i*th smallest observation
R_i	Corresponding rank for x_i
n	Sample size
$fr_{dist}(x_i)$	Fractional rank of x_i for the specified distribution function
$a_{dist}(x_i)$	Score for the specified distribution function
α	Location parameter
β	Scale parameter
γ	Shape parameter
ν	Degrees of freedom

[1] This procedure was introduced in SPSS 7.0 and replaces the NPPLOT procedure of earlier releases.

Fractional Ranks

Based on the rank R_i for the observation x_i, the fractional rank $fr_{dist}(x_i)$ is computed and used to estimate the expected cumulative distribution function of X. One of four methods can be selected to calculate the fractional rank $fr_{dist}(x_i)$:

$$fr_{dist}(x_i) = \begin{cases} (R_i - \frac{3}{8}) / (n + \frac{1}{4}) & \text{Blom} \\ (R_i - \frac{1}{2}) / n & \text{Rankit} \\ (R_i - \frac{1}{3}) / (n + \frac{1}{3}) & \text{Tukey} \\ R_i / (n+1) & \text{Van der Waerden} \end{cases}$$

Scores

The score of the specified distribution for case i is defined as

$$a_{dist}(x_i) = F_{dist}{}^{-1}(fr_{dist}(x_i)) \qquad i = 1, \ldots, n$$

where $F_{dist}{}^{-1}$ is the inverse cumulative specified distribution function.

P-P Plot

For a P-P plot, the fractional rank and the cumulative specified distribution function F_{dist} are plotted:

$$\left(fr_{dist}(x_i), F_{dist}(x_i) \right) \qquad i = 1, \ldots, n$$

Q-Q Plot

For a Q-Q plot, the observations and the score for the specified distribution function are plotted.

$$\left(x_i, a_{dist}(x_i)\right) \qquad i = 1, \ldots, n$$

Distributions

The distributions and their parameters are listed below. Parameters may be either specified by users or estimated from the data. Any parameter values specified by the user should satisfy the conditions indicated.

Beta(β_1, β_2)	β_1 (> 0) and β_2 (> 0) are scale parameters.
Chi-square(ν)	ν (> 0) is the degrees of freedom.
Exponential(β)	β (> 0) is a scale parameter.
Gamma(γ, β)	γ (> 0) is a shape parameter and β (> 0) is the scale parameter.
Half Normal(β)	β (> 0) is a scale parameter and the location parameter is 0.
Laplace(α,β)	α is the location parameter and β (> 0) is the scale parameter.
Logistic(α,β)	α is the location parameter and β (> 0) is the scale parameter.
Lognormal(β,γ)	β (> 0) is a scale parameter and γ (> 0) is a shape parameter.
Normal(α, β)	α is the location parameter and β (> 0) is the scale parameter.
Pareto(β,b);	β (> 0) is scale parameter and b (> 0) is an index of inequality.

Student's $t(\nu)$	$\nu\,(>0)$ is the degrees of freedom specified by the user.	
Uniform(a,b)	a is a lower bound and b is an upper bound.	
Weibull(β,γ)	$\beta\,(>0)$ is a scale parameter and $\gamma\,(>0)$ is a shape parameter.	

Estimates of the Parameters

The estimates for parameters for each distribution are defined below.

Beta($\beta_1,\ \beta_2$)	$\hat{\beta}_1 = \overline{X}\left\{\frac{\overline{X}(1-\overline{X})}{S^2}-1\right\}$	scale parameter
	$\hat{\beta}_2 = (1-\overline{X})\left\{\frac{\overline{X}\left(1-\overline{X}\right)}{S^2}-1\right\}$	scale parameter
Chi-square(ν)	ν is the degrees of freedom specified by the user.	
Exponential(β)	$\hat{\beta}=\frac{1}{\overline{X}}$	scale parameter
Gamma($\gamma,\ \beta$)	$\hat{\gamma}=\frac{\overline{X}^2}{S^2}$	shape parameter
	$\hat{\beta}=\frac{\overline{X}}{S^2}$	scale parameter
Half Normal(β)	$\hat{\beta}=\sqrt{x_1^2+\ldots+x_n^2}$	scale parameter
Laplace(α,β)	$\hat{\alpha}=\overline{X}$	location parameter
	$\hat{\beta}=\sqrt{\frac{S^2}{2}}$	scale parameter
Logistic(α,β)	$\hat{\alpha}=\overline{X}$	location parameter
	$\hat{\beta}=\sqrt{3}\left(\frac{S}{\pi}\right),\ \pi=3.1415927$	scale parameter

Lognormal	$\hat{\beta} = \exp(L\bar{X})$	scale parameter
	$\hat{\gamma} = LS$	shape parameter
Normal(α, β)	$\hat{\alpha} = \bar{X}$	location parameter
	$\hat{\beta} = S$	scale parameter
Pareto(β,b);	$\hat{\beta} = \min\{x_1, \ldots, x_n\}$	scale parameter
	$\hat{b} = \frac{1}{\overline{L}\overline{X} - \ln(\hat{\beta})}$	index of inequality
Student's $t(v)$	v is the degrees of freedom specified by the user.	
Uniform(a,b)	$\hat{a} = \min\{x_1, \ldots, x_n\}$	lower bound
	$\hat{b} = \max\{x_1, \ldots, x_n\}$	upper bound
Weibull(β,γ)	$\hat{\beta} = \frac{\sum_{i=1}^{n} U_i Y_i - n\bar{U}\bar{Y}}{\sum_{i=1}^{n} (U_i - \bar{U})^2}$	scale parameter
	$\hat{\gamma} = \exp(-((\bar{Y} - \hat{\beta}\bar{U}) / \hat{\beta}))$	shape parameter

where $Y_i = \ln(-\ln(1 - fr_{dist}(x_i)))$ and $U_i = \ln(x_i)$

References

Kotz, S., and Johnson, N. L., eds. 1988. *Encyclopedia of statistical sciences.* John Wiley & Sons, Inc.: New York.