CHAPTER 8 Probability Library Functions

The following is a list of available probability functions. These functions can operate on numbers, parameters, variables, and arithmetic expressions.

There are currently 39 supported distributions with 4 possible functions (the probability density function, the cumulative distribution function, the percent point function, and the sparsity function). All distributions support the cumulative distribution function, 33 support the probability density function, 38 support the percent point function, and 7 support the sparsity function. These functions are often referred to as the cdf, pdf, ppf, and sf functions respectively. Of the 39 distributions, 33 are continuous distributions and 6 are discrete. All of the currently supported distributions are univariate except for the bivariate normal distribution. There is one miscellaneous function for finding the non-centrality parameter of the non-central chi-square distribution.

A probability density function, f(x), for a continuous distribution is one that satisfies the following 3 properties:

3. It is non-negative for all real x.

4. The integral of the function from minus infinity to positive infinity is zero.

5. The probability that X is between 2 points a and b (a < b) is equal to the definite integral of f(x) from a to b.

For a discrete distribution, property 2 becomes the sum over all x is equal to 1. Property 3 becomes f(x) equals the probability at a given x. The probability density is called the probability mass function for discrete distributions. However, we will use the term probability density for both continuous and discrete distributions.

For both discrete and continuous distributions, the cumulative distribution function for a given value x is the probability from minus infinity to x, i.e., F(x) =). For continuous distributions, this is the integral of the density function from minus infinity to x. For discrete distributions, this is the sum over all X from minus infinity to x.

The percent point function is the inverse of the cumulative distribution function. That is, given a probability and its cumulative distribution function, it returns the x from the corresponding cumulative distribution function. The sparsity function, which is rarely used in practice, is the derivative of the percent point function. The input value is between 0 and 1 for both of these functions.

Additional probability functions are frequently added to DATAPLOT. You can enter the DATAPLOT command LIST DISTRIBU to see a list of currently supported distributions. You can then enter the DATAPLOT command HELP PROBABILITY FUNCTIONS to find the names of new functions not documented in this chapter.

Beta distribution

BETCDF(ALPHA,BETA)	Compute the beta cumulative distribution function.
BETPDF(ALPHA,BETA)	Compute the beta probability density function.
BETPPF(P,ALPHA,BETA)	Compute the beta percent point function.

Bivariate normal distribution

BVNCDF(X1,X2,RHO)	Compute the bivariate normal cumulative distribution function.
BVNPDF(X1,X2,RHO)	Compute the bivariate normal probability density function.

Binomial distribution

BINCDF(X,P,N)	Compute the binomial cumulative distribution function.
BINPDF(X,P,N)	Compute the binomial probability density function.
BINPPF(X,P,N)	Compute the binomial percent point function.

Cauchy distribution

CAUCDF(X)	Compute the Cauchy cumulative distribution function.
CAUPDF(X)	Compute the Cauchy probability density function.
CAUPPF(P)	Compute the Cauchy percent point function.
CAUSF(P)	Compute the Cauchy sparsity function.

Chi-square distribution

CHSCDF(X,NU)	Compute the chi-squared cumulative distribution function.
CHSPDF(X,NU)	Compute the chi-squared probability density function.
CHSPPF(P,NU)	Compute the chi-squared percent point function.

Discrete uniform distribution

DISCDF(X,N)	Compute the discrete uniform cumulative distribution function.
DISPDF(X,N)	Compute the discrete uniform probability density function.
DISPPF(P,N)	Compute the discrete uniform percent point function.

Doubly non-central F distribution

DNFCDF(X,V1,V2,L1,L2)	Compute the doubly non-central F cumulative distribution function.
DNFPPF(P,V1,V2,L1,L2)	Compute the doubly non-central F percent point function.

Doubly non-central t distribution

DNTCDF(X,V,L1,L2)	Compute the doubly non-central t cumulative distribution function.
DNTPPF(P,V,L1,L2)	Compute the doubly non-central t percent point function.

Double exponential (or Laplace) distribution

DEXCDF(X)	Compute the double exponential cumulative distribution function.
DEXPDF(X)	Compute the double exponential probability density function.
DEXPPF(P)	Compute the double exponential percent point function.
DEXSF(P)	Compute the double exponential sparsity function.

Extreme Value Type I (or Frechet) distribution

EV1CDF(X)	Compute the extreme value type I cumulative distribution function.
EV1PDF(X)	Compute the extreme value type I probability density function.
EV1PPF(P)	Compute the extreme value type I percent point function.

Extreme Value Type II (or Gumbel) distribution

EV2CDF(X,GAMMA)	Compute the extreme value type II cumulative distribution function.
EV2PDF(X,GAMMA)	Compute the extreme value type II probability density function.
EV2PPF(P,GAMMA)	Compute the extreme value type II percent point function.

Exponential distribution

EXPCDF(X)	Compute the exponential cumulative distribution function.
EXPPDF(X)	Compute the exponential probability density function.
EXPPPF(P)	Compute the exponential percent point function.
EXPSF(P)	Compute the exponential sparsity function.

F distribution

FCDF(X,NU1,NU2)	Compute the F cumulative distribution function.
FPDF(X,NU1,NU2)	Compute the F probability density function.

FPPF(P,NU1,NU2)	Compute the F percent point function.

Fatigue Life (or Birnbaum-Saunders) distribution

FLCDF(X,GAMMA)	Compute the fatigue life cumulative distribution function.
FLPDF(X,GAMMA)	Compute the fatigue life probability density function.
FLPPF(X,GAMMA)	Compute the fatigue life percent point function.

Gamma distribution

GAMCDF(X,GAMMA)	Compute the gamma cumulative distribution function.
GAMPDF(X,GAMMA)	Compute the gamma cumulative distribution function.
GAMPPF(P,GAMMA)	Compute the gamma percent point function.

Generalized Pareto distribution

GEPCDF(X,GAMMA)	Compute the generalized Pareto cumulative distribution function.
GEPPDF(X,GAMMA)	Compute the generalized Pareto probability density function.
GEPCDF(X,GAMMA)	Compute the generalized Pareto percent point function.

Geometric distribution

GEOCDF(X,P)	Compute the geometric cumulative distribution function.
GEOPDF(X,P)	Compute the geometric probability density function.
GEOPPF(X,P)	Compute the geometric percent point function.

Half-Normal distribution

HFNCDF(X)	Compute the half-normal cumulative distribution function.
HFNPDF(X)	Compute the half-normal probability density function.
HFNPPF(P)	Compute the half-normal percent point function.

Hypergeometric distribution

HYPCDF(X,K,N,M)	Compute the hypergeometric cumulative distribution function.
HYPPDF(X,K,N,M)	Compute the hypergeometric probability density function.
HYPPPF(X,K,N,M)	Compute the hypergeometric percent point function.

Inverse Gaussian distribution

IGCDF(X,GAMMA)	Compute the inverse Gaussian cumulative distribution function.
IGCDF(X,GAMMA)	Compute the inverse Gaussian probability density function.
IGPPF(X,GAMMA)	Compute the inverse Gaussian percent point function.

Logormal distribution

LGNCDF(X)	Compute the lognormal cumulative distribution function.
LGNPDF(X)	Compute the lognormal probability density function.
LGNPPF(P)	Compute the lognormal percent point function.

Logistic distribution

LOGCDF(X)	Compute the logistic cumulative distribution function.
LOGPDF(X)	Compute the logistic probability density function.
LOGPPF(P)	Compute the logistic percent point function.
LOGSF(P)	Compute the logistic sparsity function.

Negative Binomial distribution

NBCDF(X,P,N)	Compute the negative binomial cumulative distribution function.
NBPDF(X,P,N)	Compute the negative binomial probability density function.
NBPPF(X,P,N)	Compute the negative binomial percent point function.

Non-central beta distribution

NCBCDF(X,A,B,LAM)	Compute the non-central beta cumulative distribution function.
NCBPPF(P,A,B,LAM)	Compute the non-central beta percent point function.

Non-central chi-square distribution

NCCCDF(X,V,LAMBDA)	Compute the non-central chi-square cumulative distribution function.
NCCNCP(X,V,LAMBDA)	Compute the non-central chi-square non-centrality parameter.
NCCPPF(P,V,LAMBDA)	Compute the non-central chi-square percent point function.

Non-central F distribution

NCFCDF(X,V1,V2,L1,L2)	Compute the non-central F cumulative distribution function.
NCFPPF(P,V1,V2,L1,L2)	Compute the non-central F percent point function.

Non-central t distribution

NCTCDF(X,V,DELTA)	Compute the non-central t cumulative distribution function.
NCTPPF(P,V,DELTA)	Compute the non-central t percent point function.

Normal distribution

NORCDF(X)	Compute the normal cumulative distribution function.
NORPDF(X)	Compute the normal probability density function.
NORPPF(P)	Compute the normal percent point function.
NORSF(P)	Compute the normal sparsity function.

Pareto distribution

PARCDF(X,GAMMA)	Compute the Pareto cumulative distribution function.
PARPDF(X,GAMMA)	Compute the Pareto probability density function.
PARPPF(X,GAMMA)	Compute the Pareto percent point function.

Poisson distribution

POICDF(X,P,N)	Compute the Poisson cumulative distribution function.
POIPDF(X,P,N)	Compute the Poisson probability density function.
POIPPF(X,P,N)	Compute the Poisson percent point function.

Reciprocal Inverse Gaussian distribution

RIGCDF(X,GAMMA)	Compute the reciprocal inverse Gaussian cumulative distribution function.
RIGCDF(X,GAMMA)	Compute the reciprocal inverse Gaussian probability density function.
RIGPPF(X,GAMMA)	Compute the reciprocal inverse Gaussian percent point function.

Semi-circular distribution

SEMCDF(X)	Compute the semi-circular cumulative distribution function.
SEMPDF(X)	Compute the semi-circular probability density function.
SEMPPF(P)	Compute the semi-circular percent point function.

t distribution

TCDF(X,NU)	Compute the t cumulative distribution function.
TPDF(X,NU)	Compute the t probability density function.
TPPF(P,NU)	Compute the t percent point function.

Triangular distribution

TRICDF(X,C	Compute the triangular cumulative distribution function.
TRIPDF(X,C)	Compute the triangular probability density function.
TRIPPF(P,C)	Compute the triangular percent point function.

Tukey-Lambda distribution

LAMCDF(X,LAMBDA)	Compute the Tukey-Lambda cumulative distribution function.
LAMPDF(X,LAMBDA)	Compute the Tukey-Lambda probability density function.
LAMPPF(P,LAMBDA)	Compute the Tukey-Lambda percent point function.
LAMSF(P,LAMBDA)	Compute the Tukey-Lambda sparsity function.

Uniform distribution

UNICDF(X)	Compute the uniform cumulative distribution function.
UNIPDF(X)	Compute the uniform probability density function.
UNIPPF(P)	Compute the uniform percent point function.
UNISF(P)	Compute the uniform sparsity function.

Von Mises distribution

VONCDF(X,B	Compute the Von Mises cumulative distribution function.
VONPDF(X,B)	Compute the Von Mises probability density function.
VONPPF(P,B)	Compute the Von Mises percent point function.

Wald distribution

WALCDF(X,GAMMA)	Compute the Wald cumulative distribution function.
WALPDF(X,GAMMA)	Compute the Wald probability density function.
WALPPF(X,GAMMA)	Compute the Wald percent point function.

Weibull distribution

WEICDF(X,GAMMA)	Compute the Weibull cumulative distribution function.
WEIPDF(X,GAMMA)	Compute the Weibull probability density function.
WEIPPF(X,GAMMA)	Compute the Weibull percent point function.

General considerations

1. The following are excellent sources for information on a large range of probability distributions.

   "Discrete Univariate Distributions," Johnson and Kotz, Houghton Mifflin, 1969.

   "Continuous Univariate Distributions - Volumes 1 and 2," Johnson and Kotz, Houghton Mifflin, 1970.

   "Statistical Distributions", Second Edition, Evans, Hastings, and Peacock, John Wiley and Sons, 1993.

   The Evans, Hastings, and Peacock book provides quick summaries of a large range of distributions. The Johnson and Kotz volumes provide detailed descriptions of a large range of distributions. New editions, with substantial updates, have recently been published (from John Wiley and Sons) for the Johnson and Kotz texts.

2. The extreme value distributions (the Weibull, the extreme value type I, the extreme value type II, and the generalized Pareto) can be based on either the minimum or the maximum order statistic. This is specified with the SET MINMAX command. This command is required before using any of the probability functions for these distributions. There is no default value for the MINMAX parameter. Additional details are provided in the documentation for the individual commands.

3. Many of these distributions have a standard and a general form of the distribution. DATAPLOT calculates the function for the standard form of the distribution. The plot of the function for a member of the general form of the distribution will have the same shape as the standard form, but with different location and scale parameters. The Cauchy, double exponential, extreme value type I, extreme value type II, exponential, fatigue life, gamma, generalized Pareto, half-normal, inverse Gaussian, lognormal, logistic, normal, Pareto, reciprocal inverse Gaussian, triangular, uniform, Von Mises, Wald, and Weibull distributions have a standard and general form. Fortunately, the general form of the distribution is straightforward to compute from the standard form.

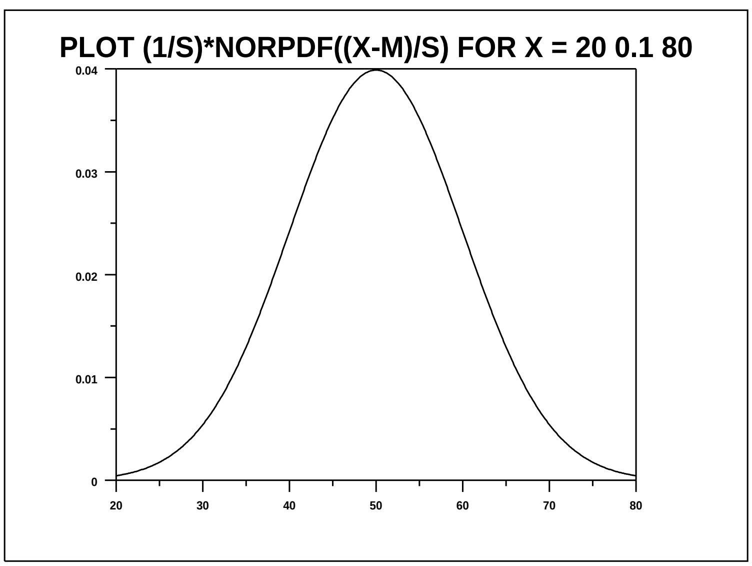
For example, to calculate a probability density, a cumulative distribution, a percent point, a sparsity function value, and to plot the probability density function on the original scale for the normal distribution with mean 50 and standard deviation 10, do the following:

```
. Specify the location and scale parameters
. (mean and standard deviation for normal distribution)
LET M = 50
LET S = 10
. Define a given x value and a significance level
LET X = 62
LET P = 0.95
LET A = NORCDF((X-M)/S)
LET B = (1/S)*NORPDF((X-M)/S)
LET C = S*NORPPF(P) + M
LET D = S*NORSF(P)
PRINT “CDF VALUE (X = ^X) = ^A”
PRINT “PDF VALUE (X = ^X) = ^B”
PRINT “PPF VALUE (ALPHA = ^P) = ^C”
PRINT “SPARSITY FUNCTION VALUE (ALPHA = ^P) = ^D”
```

The following output is generated.

```
CDF VALUE (X = 62) = 0.194187
PDF VALUE (X = 62) = 0.019419
PPF VALUE (ALPHA = 0.95) = 66.44854
SPARSITY FUNCTION VALUE (ALPHA = 0.95) = 96.95968
```

```
TITLE AUTOMATIC
PLOT (1/S)*NORPDF((X-M)/S) FOR X = 20 0.1 80
```



The technique is the same for the other distributions that have a standard and a general form. The M and S represent the location and scale parameters and are not necessarily the mean and standard deviation. The documentation for the individual commands gives the functional form of the general distribution, but it does not discuss the transformation from the standard form to the general form since it is the same for all distributions. Simply use the given scale and location parameter with the appropriate probability distribution as in the above example.

4. Random numbers can be generated from more than 30 of the distributions listed in this chapter. The random numbers will be for the standard form of the distribution. To generate random numbers for the general form, multiply by the scale factor and add the location parameter. For example, to generate random numbers from a normal distribution with mean 50 and standard deviation 10, enter the following commands:

```
LET Y = NORMAL RANDOM NUMBERS FOR I = 1 1 100
LET Y = 10*Y + 50
```

Random number generation is covered in chapter 5 of this manual.

In addition, probability plots can be generated for each of the 38 distributions that have a percent point function available. For most of the distributions with a family parameter, a probability plot correlation coefficient plot can be generated to determine the best value of the family parameter. The PROBABILITY PLOT and PPCC PLOT commands are documented in chapter 2 of the DATAPLOT Reference Manual - Volume I.

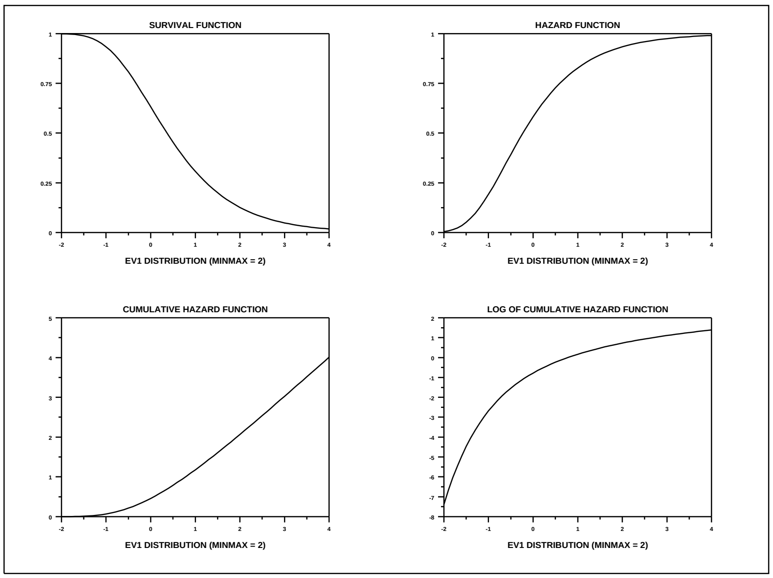
5. Two additional probability functions commonly used in reliability or lifetime analysis are the survival function and the hazard function. Although DATAPLOT does not provide functions directly for these, they are straightforward to generate using the cdf and pdf functions. The survival function is the probability that an individual survives until time t while the hazard function specifies the rate of failure at time t. That is:

S(t) = Pr(T>=t)
h(t) = f(t)/S(t)

For continuous distributions, the survival function is simply 1 minus the cdf function. For discrete distributions, the survival function is 1 minus the cdf function plus the pdf function. The hazard function can then be defined in terms of the pdf and survival functions. Some analysts prefer to work with the cumulative hazard. For continuous distributions, the cumulative hazard at time t is the integral of the hazard function from negative infinity to t. This turns out to be the negative logarithm of the survival function (i.e., -ln(S(t))). Some analysts prefer to plot the log of the cumulative hazard function (i.e., ln(-ln(S(t)))). Analogs for the discrete case are found by taking the logs of the survivor function in the same way.

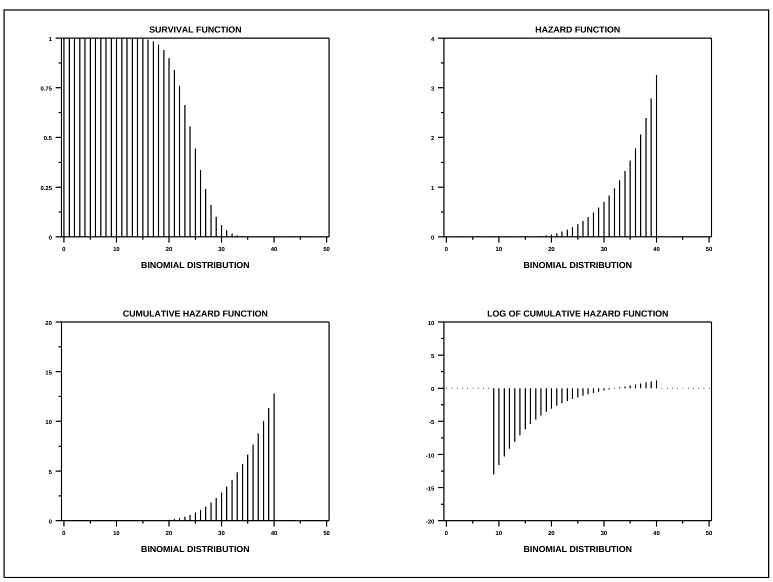
For example, to plot these functions for the extreme value type I distribution, do the following:

```
TITLE SIZE 3; TITLE SURVIVAL FUNCTION
MULTIPLOT 2 2; MULTIPLOT CORNER COORDINATES 0 0 100 100
X1lABEL EV1 DISTRIBUTION (MINMAX = 2)
SET MINMAX 2
PLOT 1 - EV1CDF(X) FOR X = -2 0.1 4
TITLE HAZARD FUNCTION
PLOT EV1PDF(X)/(1 - EV1CDF(X)) FOR X = -2 0.1 4
TITLE CUMULATIVE HAZARD FUNCTION
PLOT -LN(1 - EV1CDF(X)) FOR X = -2 0.1 4
TITLE LOG OF CUMULATIVE HAZARD FUNCTION
PLOT LN(-LN(1 - EV1CDF(X))) FOR X = -2 0.1 4
END OF MULTIPLOT
```



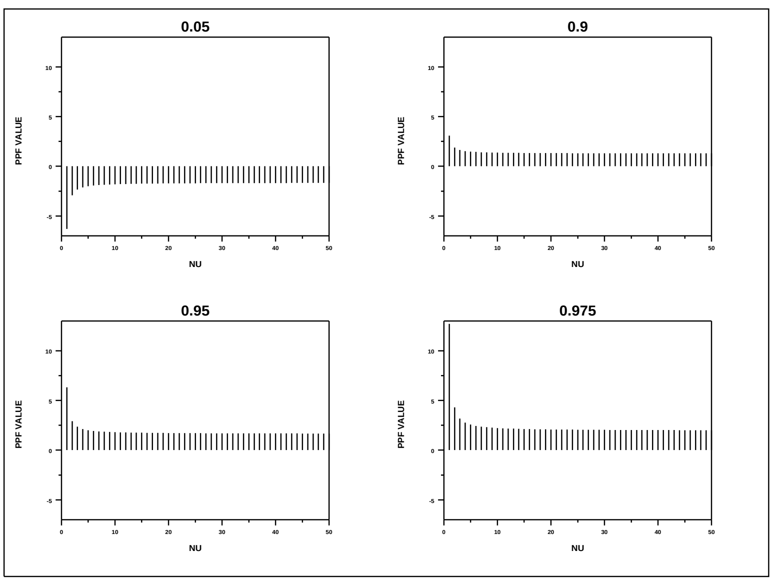
To plot these functions for the binomial distribution, do the following:

```
LET N = 50
LET P = 0.5
TITLE SIZE 3
LET X = SEQUENCE 0 1 N
LET SURV = 1 - BINCDF(X,0.5,N)
LET TEMP = BINPDF(X,0.5,N)
LET HAZARD = 0 FOR I = 1 1 N
LET CUMHAZ = 0 FOR I = 1 1 N
LET LOGCHAZ = 0 FOR I = 1 1 N
LET HAZARD = TEMP/SURV SUBSET SURV > 0
LET CUMHAZ = -LN(SURV) SUBSET SURV > 0
LET LOGCHAZ = LN(HAZARD) SUBSET HAZARD > 0
.
LINE OFF; SPIKE ON
XLIMITS 0 N; XTIC OFFSET 0.5 0.5
.
MULTUPLOT 2 2; MULTIPLOT CORNER COORDINATES 0 0 100 100
X1lABEL BINOMIAL DISTRIBUTION
TITLE SURVIVAL FUNCTION
PLOT SURV X
TITLE HAZARD FUNCTION
PLOT HAZARD X
TITLE CUMULATIVE HAZARD FUNCTION
PLOT CUMHAZ X
TITLE LOG OF CUMULATIVE HAZARD FUNCTION
PLOT LOGCHAZ X
END OF MULTIPLOT
```



6. When using the percent point function with a distribution that has a family parameter, it is most common to leave this family parameter fixed. However, an interesting plot can be generated by leaving the probability value fixed and varying the family parameter. This is demonstrated for the t distribution, which has the parameter NU which specifies the degrees of freedom.

```
LET NU = SEQUENCE 1 1 50
MULTIPLOT 2 2; MULTIPLOT CORNER COORDINATES 0 0 100 100
X1LABEL NU; Y1LABEL PPF VALUE
YLIMITS -5 10; YTIC OFFSET -2 3
LINE BLANK; SPIKE ON
.
LET P = .05; TITLE ^P; LET PPF = TPPF(P,NU)
PLOT PPF VS NU
LET P = .90; TITLE ^P; LET PPF = TPPF(P,NU)
PLOT PPF VS NU
LET P = .95; TITLE ^P; LET PPF = TPPF(P,NU)
PLOT PPF VS NU
LET P = .975; TITLE ^P; LET PPF = TPPF(P,NU)
PLOT PPF VS NU
END OF MULTIPLOT
```



This plot can be used to show the values that produce a statistically significant result for various values of NU.