Chapter 915

Probability Calculator

Introduction

Most statisticians have a set of probability tables that they refer to in doing their statistical work. This procedure provides you with a set of electronic statistical tables that will let you look up values for various probability distributions.

To run this option, select Probability Calculator from the Other menu of the Analysis menu. A window will appear that will let you indicate which probability distribution you want to use along with various input parameters. Select the Calculate button to find and display the results.

Many of the probability distributions have two selection buttons to the left of them. The first (left) button selects the inverse probability distribution. An inverse probability distribution is in a form so that when you give it a probability, it calculates the associated critical value. The second (right) button selects the regular probability distribution which is formulated so that when you give it a critical value, it calculates the (left tail) probability.

Probability Distributions

Beta Distribution

The beta distribution is usually used because of its relationship to other distributions, such as the t and F distributions. The noncentral beta distribution function is formulated as follows:

$$\Pr(0 \le x \le X|A,B,L) = I_X(A,B,L) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\sum_{k=0}^{\infty}\frac{(L/2)^k e^{-L/2}}{k!}\int_0^X t^{A+k-1}(1-t)^{B-1}dt$$

where

$0 < A$, $0 < B$, $0 \le L$, and $0 \le x \le 1$

When the noncentrality parameter (NCP), L, is set to zero, the above formula reduces to the *standard* beta distribution, formulated as

$$\Pr(0 \le x \le X|A,B) = \frac{\Gamma(A+B)}{\Gamma(A)\Gamma(B)}\int_0^X t^{A-1}(1-t)^{B-1}dt$$

When the inverse distribution is selected, you supply the probability value, and the program solves for *X*. When the regular distribution is selected, you supply *X*, and the program solves for the cumulative (left-tail) probability.

Binomial Distribution

The binomial distribution is used to model the counts of a sequence of independent binary trials in which the probability of a success, *P*, is constant. The total number of trials (sample size) is *N*. *R* represents the number of successes in *N* trials. The probability of exactly R successes is:

$$\Pr(r = R|N, P) = \binom{N}{R} (P)^R (1 - P)^{N-R}$$

where

$$\binom{N}{R} = \frac{N!}{R!\,(N - R!)}$$

The probability of from 0 to R successes is given by:

$$\Pr(0 \le r \le R|N, P) = \sum_{r=0}^{R} \binom{N}{r} P^r (1 - P)^{N-r}$$

When the inverse distribution is selected, you supply the probability value, and the program solves for *R*. When the regular distribution is selected, you supply *R*, and the program solves for the cumulative (left-tail) probability.

Bivariate Normal Distribution

The bivariate normal distribution is given by the formula

$$\Pr(x < h, y < k|r) = \frac{1}{2\pi\sqrt{1 - r^2}} \int_{-\infty}^{h} \int_{-\infty}^{k} e^{\frac{-x^2+2rxy-y^2}{2(1-r^2)}} dx\, dy$$

where *x* and *y* follow the bivariate normal distribution with correlation coefficient *r*.

Chi-Square Distribution

The Chi-square distribution arises often in statistics when the normally distributed random variables are squared and added together. DF is the degrees of freedom of the estimated standard error.

The noncentral Chi-square distribution function is used in power calculations. The noncentral Chi-square distribution is calculated using the formula:

$$\Pr(0 \le x \le X|df, L) = \sum_{k=0}^{\infty} \frac{(L/2)^k e^{-L/2}}{k!} P(X|df + 2k)$$

where

$$\Pr(X|df) = \frac{1}{2^{df/2}\Gamma\left(\frac{df}{2}\right)} \int_0^X t^{\frac{df}{2}-1} e^{-t/2} dt$$

When the noncentrality parameter (NCP), L, is set to zero, the above formula reduces to the (central) Chi-square distribution.

When the inverse distribution is selected, you supply the probability value, and the program solves for *X*. When the regular distribution is selected, you supply *X*, and the program solves for the cumulative (left-tail) probability.

Correlation Coefficient Distribution

The correlation coefficient distribution is formulated as follows:

$$\Pr(r \leq R|n,\rho) = \int_{-1}^{R} \frac{2^{n-3}}{\pi(n-3)!}(1-\rho)^{(n-1)/2}(1-r)^{(n-4)/2}\sum_{i=0}^{\infty}\Gamma^2\left(\frac{n+i-1}{2}\right)\frac{(2\rho r)^i}{i!}dr$$

where

$|r| < 1$, $|\rho| < 1$, and $|R| < 1$

When the inverse distribution is selected, you supply the probability value, and the program solves for *R*. When the regular distribution is selected, you supply *R*, and the program solves for the cumulative (left-tail) probability.

F Distribution

The F distribution is used in the analysis of variance and in other places where the distribution of the ratio of two variances is needed. The degrees of freedom of the numerator variance is DF1 and the degrees of freedom of the denominator variance is DF2.

The noncentral-F distribution function is used in power calculations. We calculate the noncentral-F distribution using the following relationship between the F and the beta distribution function.

$$\Pr(0 \leq f \leq F|df_1, df_2, L) = I_X\left(\frac{df_1}{2}, \frac{df_2}{2}, L\right)$$

where

$$X = \frac{F(df_1)}{F(df_1) + df_2}$$

When the noncentrality parameter (NCP), L, is set to zero, the above formula reduces to the *standard* F distribution

When the inverse distribution is selected, you supply the probability value, and the program solves for *F*. When the regular distribution is selected, you supply *F*, and the program solves for the cumulative (left-tail) probability.

Hotelling's T2 Distribution

Hotelling's *T*-Squared distribution is used in multivariate analysis. We calculate the distribution using the following relationship between the *F* and the *T*2 distribution function.

$$\Pr\left(0 \le x \le \frac{(df-k+1)}{k(df)} T^2_{k,df} \middle| k, df\right) = \Pr\left(0 \le x \le F_{k,df-k+1} \middle| k, df\right)$$

where *k* is the number of variables and *df* is the degrees of freedom associated with the covariance matrix. When the inverse distribution is selected, you supply the probability value, and the program solves for *T*2. When the regular distribution is selected, you supply *T*2, and the program solves for the cumulative (left-tail) probability.

Gamma Distribution

The Gamma distribution is formulated as follows:

$$\Pr(0 \le g \le G|A,B) = \frac{1}{\mathrm{B}^A \Gamma(A)} \int_0^G x^{A-1} e^{-X/B} dx$$

where

$$\Gamma(A) = \int_0^\infty x^{A-1} e^{-x} dx$$

$0 < A, 0 < B$, and $0 \le \mathrm{G}$

When the inverse distribution is selected, you supply the probability value, and the program solves for *G*. When the regular distribution is selected, you supply *G*, and the program solves for the cumulative (left-tail) probability.

Hypergeometric Distribution

The hypergeometric distribution is used to model the following situation. Suppose a sample of size R is selected from a population with N items, M of which have a characteristic of interest. What is the probability that X of the items in the sample have this characteristic.

The probability of exactly X successes is:

$$\Pr(x = X|N, M, R) = \frac{\binom{M}{X}\binom{N-M}{R-X}}{\binom{N}{R}}$$

where

$$\binom{N}{R} = \frac{N!}{R!\,(N-R!)}$$

$$\text{Maximum}(M, R-N+M) \leq X \leq \text{Minimum}(M, R)$$

Negative Binomial Distribution

The negative binomial distribution is used to model the counts of a sequence of independent binary trials in which the probability of a success, P, is constant. The total number of trials (sample size) is N. R represents the number of successes in N trials. Unlike the binomial distribution, the sample size, N, is the variable of interest.

The question answered by the negative binomial distribution is how many tosses of a coin (with probability of a head equal to P) is necessary to achieve R heads and X tails.

The probability of exactly R successes is:

$$\Pr(x = X|R, P) = \binom{X+R-1}{R-1} P^R (1-P)^X$$

where

$$\binom{N}{R} = \frac{N!}{R!\,(N-R!)}$$

Normal Distribution

The normal distribution is formulated as follows:

$$\Pr(x \leq X|\mu,\sigma) = \frac{1}{\sqrt{2\pi\sigma^2}} \int_{-\infty}^{X} e^{-\frac{(x-\mu)^2}{2\sigma^2}} dx$$

When the mean is 0 and the variance is 1, we have the standard normal distribution. The regular normal distribution uses the variable X. The standard normal distribution uses the variable Z. Any normal distribution may be transformed to the standard normal distribution using the relationship:

$$z = \frac{x-\mu}{\sigma}$$

When the inverse distribution is selected, you supply the probability value, and the program solves for *R*. When the regular distribution is selected, you supply *R*, and the program solves for the cumulative (left-tail) probability.

Poisson Distribution

The Poisson distribution is used to model the following situation. Suppose the average number of accidents at a given intersection is 13.5 per year. What is the probability of having 2 accidents during the next half year?

The probability of exactly X occurrences with a mean occurrence rate of M is:

$$\Pr(x = X|M) = \frac{e^{-M}M^X}{X!}$$

Studentized Range Distribution

The studentized range distribution is used whenever the distribution of the ratio of a range and an independent estimate of its standard error is needed. This distribution is used quite often in multiple comparison tests run after an analysis of variance. DF is the degrees of freedom of the estimated standard error (often the degrees of freedom of the MSE). K is the number of items (means) in the sample. The distribution function is given by:

$$\Pr(0 \leq r \leq R|df,k) = \int_0^\infty \frac{2^{-\frac{df}{2}+1} df^{\frac{df}{2}} s^{df-1}}{\Gamma\left(\frac{df}{2}\right)} e^{-dfs^2/2} \mathrm{P}(Rs|n)\, dx$$

where *P(Rs|n)* is the probability integral of the range.

When the inverse distribution is selected, you supply the probability value, and the program solves for *R*. When the regular distribution is selected, you supply *R*, and the program solves for the cumulative (left-tail) probability.

Student's t Distribution

The t distribution is used whenever the distribution of the ratio of a statistic and its standard error is needed. DF is the degrees of freedom of the estimated standard error.

The noncentral-t distribution function is used in power calculations. We calculate the noncentral-t distribution using the following relationship between the t and the beta distribution function.

$$\Pr(-\infty \leq t \leq T|df, L) = 1 - \sum_{k=0}^{\infty} e^{-L^2/2} \frac{(L^2/2)^k}{2k!} I_X\left(\frac{df}{2}, \frac{1}{2}, 0\right)$$

where

$$X = \frac{df}{df + T}$$

When the noncentrality parameter (NCP), L, is set to zero, the above formula reduces to the (central) Student's t distribution

When the inverse distribution is selected, you supply the probability value, and the program solves for *T*. When the regular distribution is selected, you supply *T*, and the program solves for the cumulative (left-tail) probability.

Weibull Distribution

The Weibull distribution is formulated as follows:

$$\Pr(t \leq T|\lambda, \gamma) = 1 - e^{-(\lambda T)^{\gamma}}$$

When gamma (γ) equal to one, the distribution simplifies to the exponential distribution.

When the inverse distribution is selected, you supply the probability value, and the program solves for *T*. When the regular distribution is selected, you supply *T*, and the program solves for the cumulative (left-tail) probability.