DISTRIB

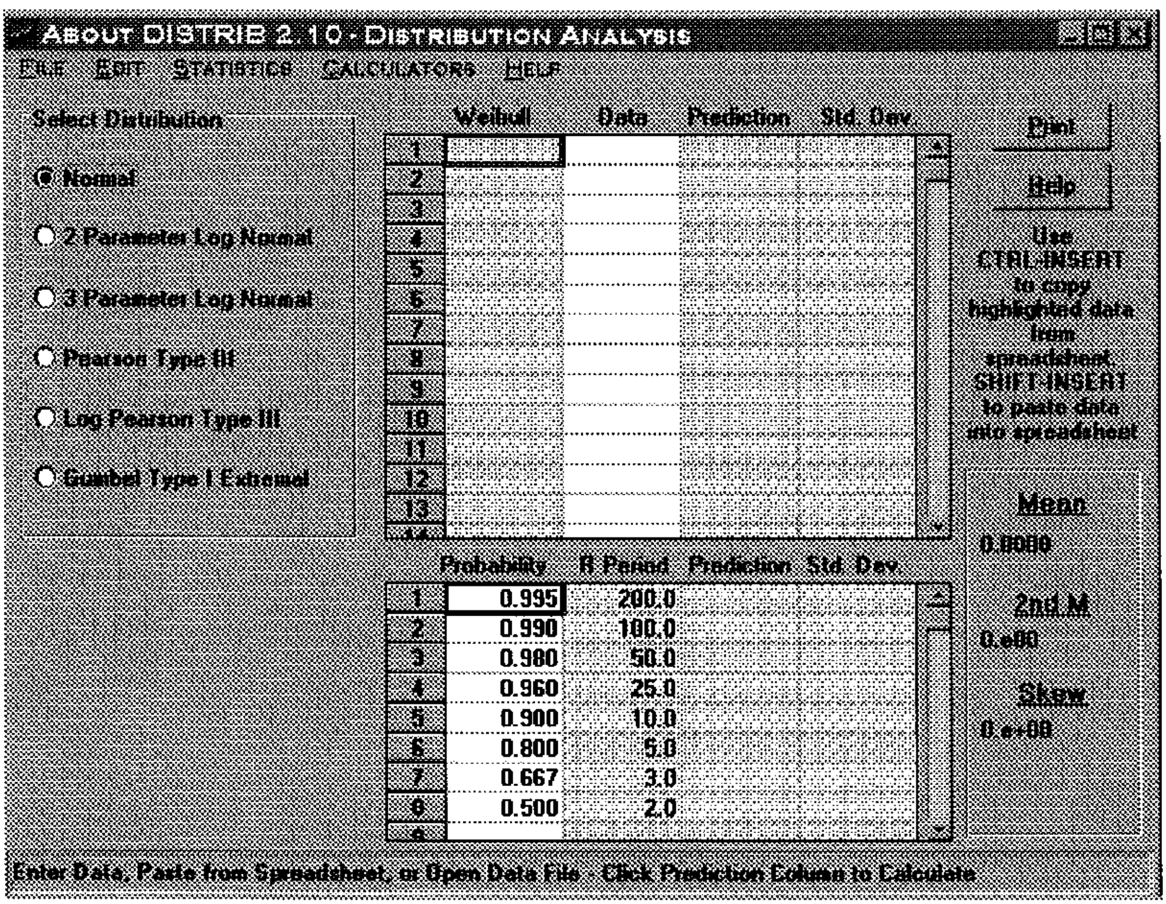
Probability Distribution Analysis

Introduction

DISTRIB is a program designed to easily allow univariate data to be fit to a probability distribution. The program interface is primarily contained in one dialog window, this screen contains all the analysis and entry tools which would normally be needed to perform all predictive analysis using probability distributions. Included with the program are a Poisson and a Binomial probability calculator.

I. The DISTRIB Main Dialog Window

The DISTRIB main dialog window contains all the features necessary to analyze a set of univariate data.



There are 5 major parts to the main dialog window:

1. Select Distribution frame
2. Data Entry Spreadsheet (upper spreadsheet with 'white' **Data** column)
3. Prediction Spreadsheet (lower spreadsheet with 'white' **Probability** column)
4. Results frame
5. Status Bar (bottom of Window)

Univariate data are directly entered into the **Data Entry Spreadsheet** (upper). The white portion of the spreadsheet allows for the entry of this data. Data should be entered with one entry per row.

Do not leave blank rows when entering data.

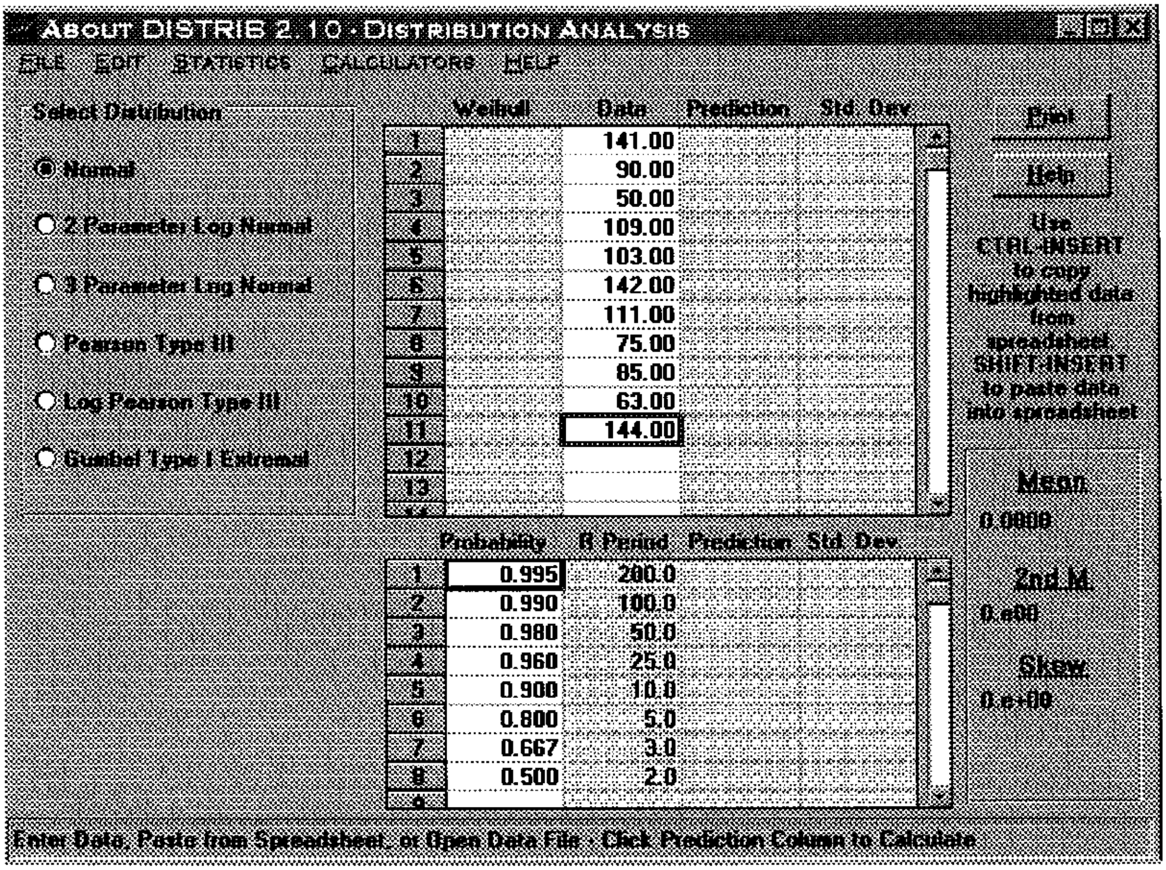
Once all data are entered, use the mouse to click in the **Prediction** column of the upper **Data Entry Spreadsheet** or in the **Select Distribution** column, select the distribution type you wish to "fit" the data.

*** DISTRIB Example 1 ***

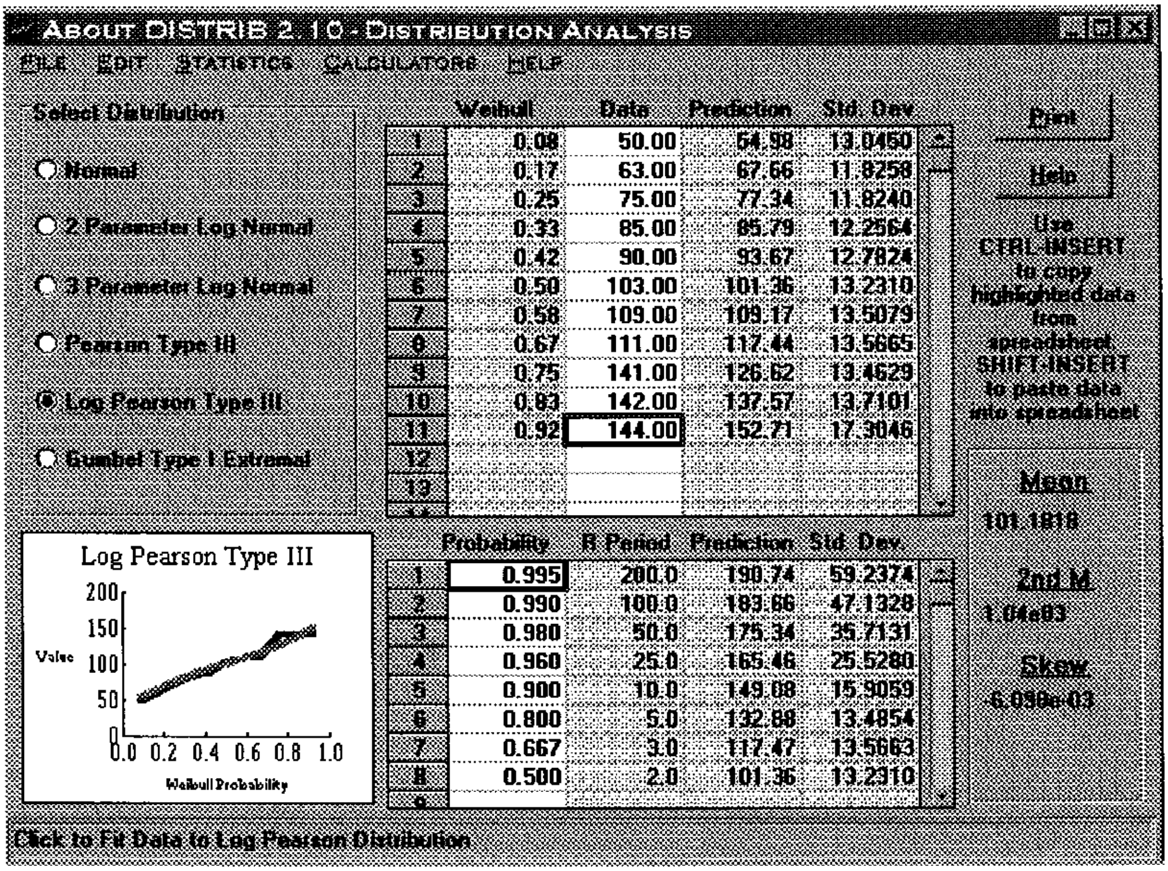
Using the maximum annual flowrate data as shown, predict the 25 year return period flow using the Log Pearson Type III distribution.

Year	Flowrate (cms)
1956	141
1957	90
1958	50
1959	109
1960	103
1961	142
1962	111
1963	75
1964	85
1965	63
1966	144

The data should be entered as shown:



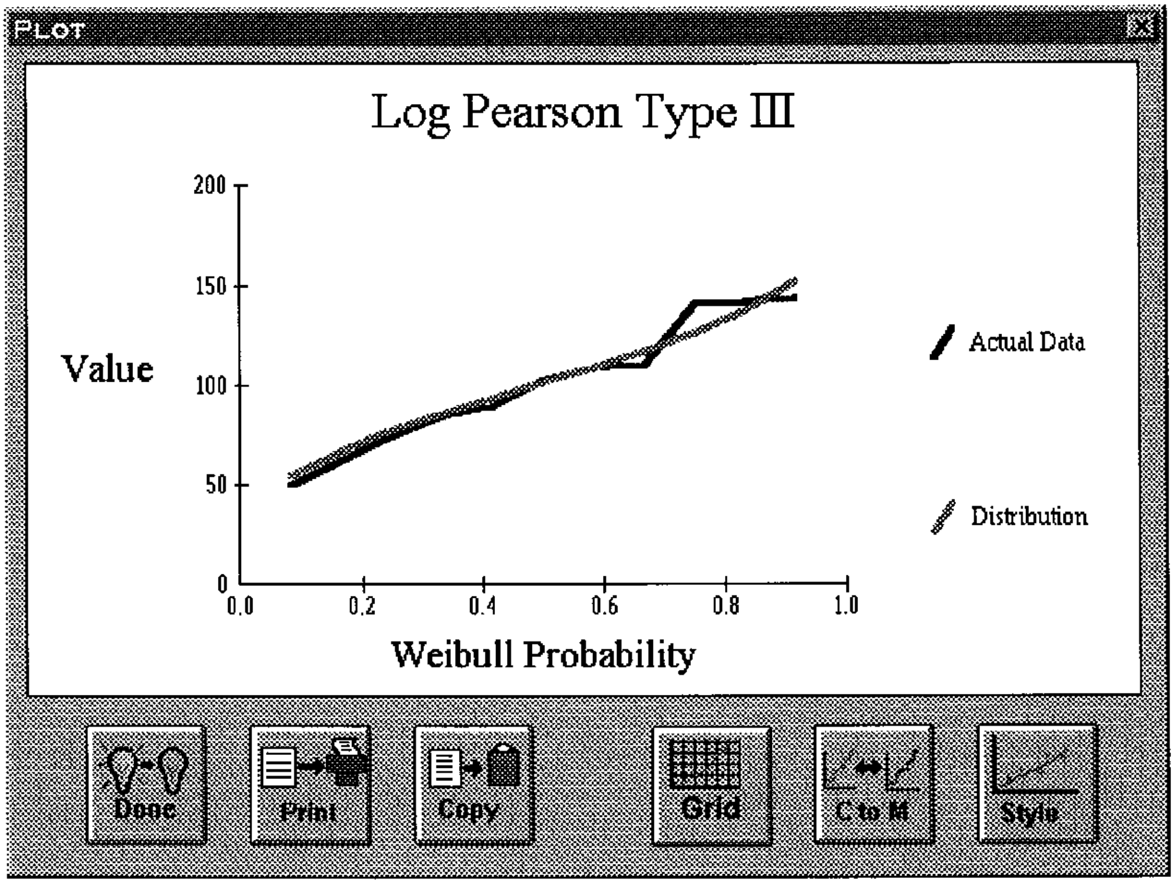
Next click the **Log Pearson Type III** button in the **Select Distribution** frame.



From the lower **Prediction Spreadsheet** we can see that the 25 year return period prediction is 165.46 cms.

*** End of Example 1 ***

The DISTRIB main dialog window allows for a number of options. To see a close-up of the plot, click on the actual distribution plot in the lower left corner of the main dialog window - an enlarged plot of the data will appear.



You may now **Print** or **Copy** (to clipboard) the plot. You may also change the grid displayed by clicking on the **Grid** button. Each consecutive click will add horizontal lines, vertical lines, or return the plot to the above shown empty grid.

The "**C to M**" button switches the plot between color and monochrome display. The **Style** button allows for various plot line displays.

Select the **Done** button to return to the main DISTRIB dialog window.

The **Prediction Spreadsheet** allows for the entry of different **Probabilities** and calculation of predictions based on different return periods. The return period can be calculated from probability using;

$$RP = \frac{1}{1-p}$$

where,

RP = Return Period

p = probability

Enter the desired probabilities in the **Probabilities** (white) column of the **Prediction Spreadsheet** (lower) and click on the **Return Period** heading to activate the calculation process.

DISTRIB allows for optional plotting position formulas. Clicking on the word **Weibull** in the upper corner of the **Data Entry Spreadsheet** will allow you to scroll through plotting position formula;

Weibull	m/(n+1)
California	m/n
Foster	(2m-1)/2n
Exceedence	(m-1)/n

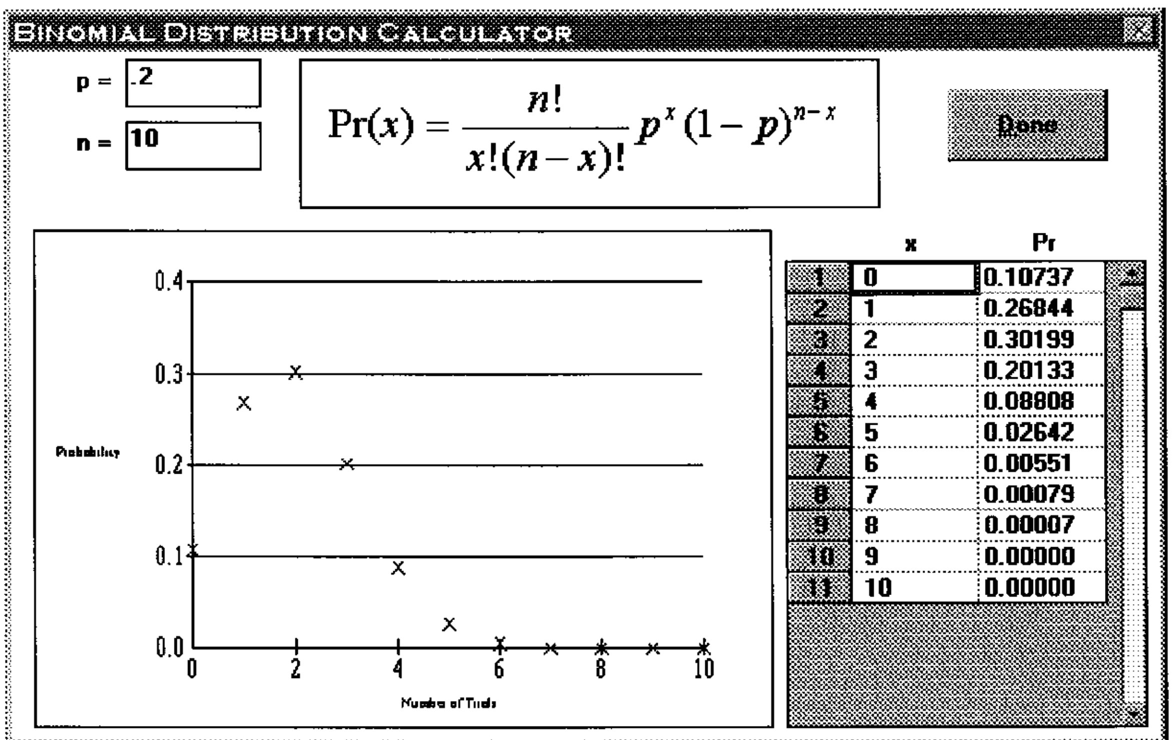
II. DISTRIB Options

A. Binomial Calculator

The **Binomial Calculator** is available from the **Calculators** menu. The **Binomial Calculator**, as shown below, allows for calculation of the probability of any discrete value using the binomial distribution.

*** DISTRIB Example 2 ***

If the probability of any a certain event is 0.2, what is the probability of that event occurring exactly 3 times in 10 tries?

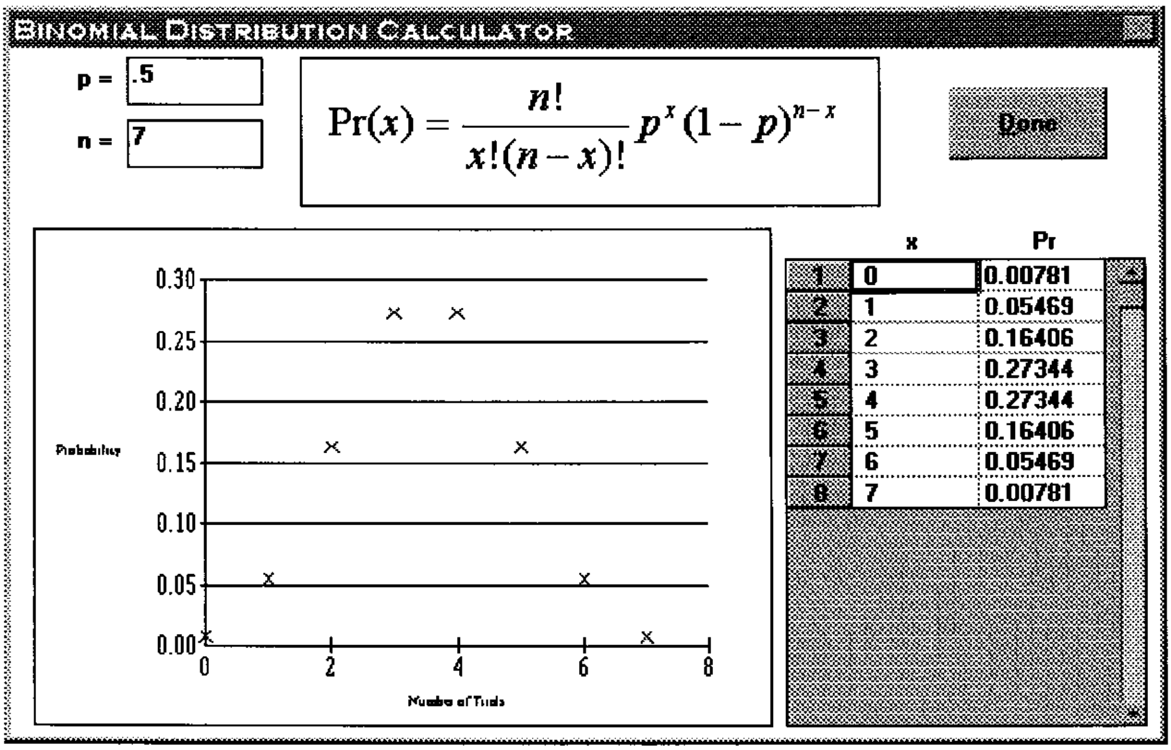


From the **Binomial Calculator** it can be seen that the probability is 0.20133 or 20.133 %.

*** End of DISTRIB Example 2 ***

* DISTRIB Example 3 *

What is the probability of getting less than 3 heads in a toss of 7 coins?



The probability of less than 3 heads, is the sum of 0, 1, and 2 heads in the toss.

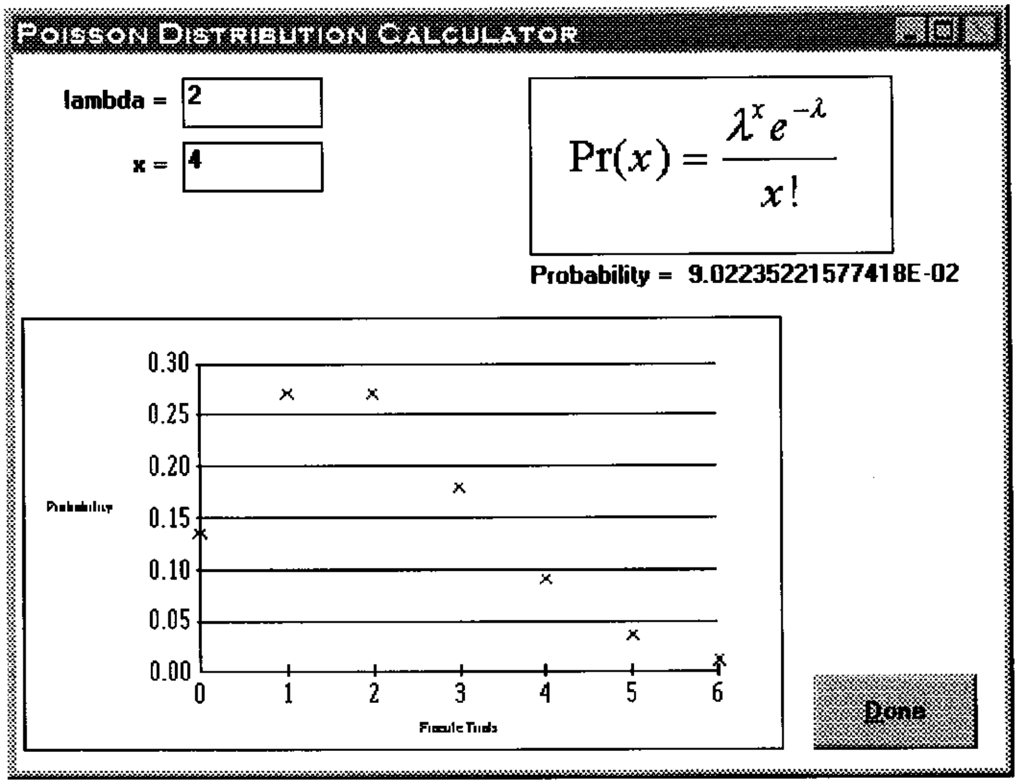
$P_R(0) = 0.00781$
$P_R(1) = 0.05469$
$P_R(2) = 0.16406$

$P_R(x < 3) = 0.22656$ or the probability of having less than 3 heads in a toss of 7 coins is 22.656 %.

*** End of DISTRIB Example 3 ***

B. Poisson Calculator

The **Poisson Calculator** is available from the **Calculators** menu. The **Poisson Calculator** performs calculations for the Poisson distribution.



C. Annualization Factor

The **Annualization Factor** is available from the **Edit** menu. The **Annualization Factor** allows for the use of partial duration series data in analysis using DISTRIB.

The **Annualization Factor** is the ratio of number of data points used in the analysis to the span of years the data covers. This factor is used in the prediction to put the return period predictions on an annual basis.

D. Statistics

The **Statistics** menu option allows you to activate calculation or plotting of the statistics without having to click on any of the **Results** frames (**Prediction** columns, **Standard Deviation** columns, or the **Return Period** column).

The **Fit Distribution** selection will calculate all of the **Results** frame data. The **Plot Distribution** will activate the enlarged plot of the distribution without having to click on the small plot in the lower left-hand corner of the main DISTRIB dialog window.

III. Types of Distributions

DISTRIB uses the common distribution types which are used in the field of water resources. These distribution are represented by probability density functions, and are shown below:

A. Normal Distribution

The probability density function for the Normal distribution is:

$$p_x(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2 \right]$$

where,

$p_x(x)$ = probability of event of magnitude less than x
μ = mean of population
σ = standard deviation of population

B. Log Normal Distribution

The probability density function of the Log Normal distribution is:

$$p_x(y) = \frac{1}{\sigma_y\sqrt{2\pi}} \exp\left[-\frac{1}{2}\left(\frac{y-\mu_y}{\sigma_y}\right)^2 \right]$$

$$y = \ln(x)$$

where,

$p_x(y)$ = probability of event of magnitude less than y
μ_y = mean of population of y
σ_y = standard deviation of population of y

C. Three parameter Log Normal

In the Three Parameter Log Normal distribution the parameter y of the log normal distribution is calculated as:

$$y = \ln(x - a)$$

where,

"a" is a constant which is determined in the analysis of the data.

D. Pearson Distribution

The probability density function of the Pearson distribution is:

$$p_x(x) = p_o\left(1 + \frac{x}{\alpha}\right)^{\alpha/\delta} e^{-x/\delta}$$

where,

δ = difference between mean and mode ($\delta = \mu - X_m$)
with X_m = mode of population x
α = scale parameter of distribution
p_o = value of $p_x(x)$ at mode

E. Log Pearson Distribution

The probability density function of the Log Pearson distribution is:

$$p_x(y) = p_{yo}\left(1+\frac{y}{\alpha}\right)^{\alpha/\delta_y} e^{-y/\delta_y}$$

where,

δy = difference between mean and mode ($\delta = \mu_y - Y_m$)
with Y_m = mode of population y
α = scale parameter of distribution
p_{yo} = value of $p_x(y)$ at mode

F. Gumbel Distribution

The probability density function of the Gumbel distribution is:

$$p_x(x) = \frac{\alpha}{\beta-\gamma}\left\{\frac{x-\gamma}{\beta-\gamma}\right\}^{\alpha-1} e^{-\left\{\frac{x-\gamma}{\beta-\gamma}\right\}^{\alpha}}$$

where,

α = scale parameter of the distribution
β = location parameter of the distribution

IV. Performing Distribution Analysis

A. Dealing with Zero Values

A number of methods have been used to handle 0.0 values. DISTRIB does not perform any of these conversions for you and **may crash** if you attempt to fit 0 data to a log distribution. To alleviate this problem you may:

1. Add 1.0 to all data
2. Add a small positive value to all data.
3. Substitute 1.0 in place of all 0 data.
4. Substitute a small positive number in the place of all zero readings.
5. Ignore all zero observations.
6. Consider the probability distribution as the sum of the probability mass at 0.0 and a probability distribution over the remainder of the range. This method is described in Jennings and Benson.

B. Selecting the "Best Fit" Distribution

In most real world cases a type of distribution will not be specified. In this situation the best distribution for the data must be selected. It is recommended that:

1. The data be fit to each distribution and plotted.
2. The plots be compared for best fit in the region of interest.
3. The selected best distribution be used for all similar data.

The "region of interest" is the portion of the fit which is going to be predicted. When making predictions of extreme events, the left side of the plot (high return periods) should fit well to the data. It is not important that the low end numbers reflect a good fit in this case.