

A quick introduction to First Bayes

Lawrence Joseph*

October 1, 2003

1 Introduction

This document very briefly reviews the main features of the First Bayes statistical teaching package. For full details, please refer to the manual that comes with First Bayes, the file name is 1b.doc. First Bayes is a program written by Tony O'Hagan and can be downloaded from

<http://www.shef.ac.uk/~st1ao/1b.html>.

It is not intended for serious statistical analyses, but it is a very nice way to investigate the properties of all commonly used statistical distributions, and is an easy to use replacement for the need for statistical tables. One can also do seven types of Bayesian analyses, and learn about how these analyses work, in terms of seeing the prior, likelihood and posterior distributions all on one graph.

2 Installation

Installation follows five steps:

1. **Copy files into the right directory:** Copy all the files from 1b.zip (downloaded from the First Bayes WWW site) into a directory called c:\1b.
2. **Add a registry entry:** Add First Bayes to the Registry by running the following command (use the Run option from the Start menu at the bottom of the screen):

```
regedit c:\1b\1b.reg
```
3. **Install required fonts:** To install fonts, open "My Computer", then "Control Panel", then "Fonts". Select File/Install New Font. Within the Add Fonts dialog box, select the folder in which First Bayes files have been placed (i.e., c:\1b). You should see two names appear in the List of Fonts, "Dyalog Std" and "Dyalog Std TT". Click on Select All and then OK.

*Department of Epidemiology and Biostatistics, McGill University, 1650 Cedar Avenue, Montreal, Quebec, Canada H3A 1A2

4. **Create a shortcut from which the program is started:** To create a shortcut to First Bayes, open “My Computer”, then the hard drive and folder containing the First Bayes files (i.e., c:\1b). With the right mouse button, drag the 1stbayes.dws file onto the desktop. From the list of options in the menu which then opens, select Create Shortcut(s) Here. Now click with the right mouse button on the shortcut icon and from the resulting menu select Properties. Click on the Shortcut tab in the Properties dialog box. Two things need to be changed here. One is to alter the text in the Target field, to read

c:\1b\dyalogrt .\1stbayes inifile=Software\1stBayes

Then click on the Change Icon button, select Browse, then in the file dialog box locate the folder in which the First Bayes files are and select 1b.ico (or just 1b if you have file extensions switched off). Click on Open, then OK and OK again. The First Bayes icon should now appear as the icon for the shortcut on the desktop.

5. **Open First Bayes:** Once First Bayes has been installed it can be started by double-clicking on its icon, just like other Windows programs.

3 Using First Bayes

Obviously, it is best if you follow this brief tour with First Bayes open in front of you.

3.1 What’s available

Upon clicking on the First Bayes icon that should now be on your desktop, you are greeted with a blank screen, as shown in Figure 1. Note that you have six menus to choose from, but we will only ever use four of these. The help and options menus should be ignored. The four useful ones are:

1. The **File** menu item is to load data into First Bayes for analysis, and to quit the program.
2. The **Distributions** item is to investigate the properties of and plot any of thirteen statistical distributions.
3. The **Analyses** menu is to do any of seven types of Bayesian analyses.
4. The **Windows** menu arranges windows, as in all Windows programs.

Figure 2 illustrates the thirteen distributions available, while Figure 3 shows the seven types of analyses that can be done in First Bayes.

We will now take a brief tour of the **File**, **Distributions**, and **Analyses** menu options, starting with **Distributions**.

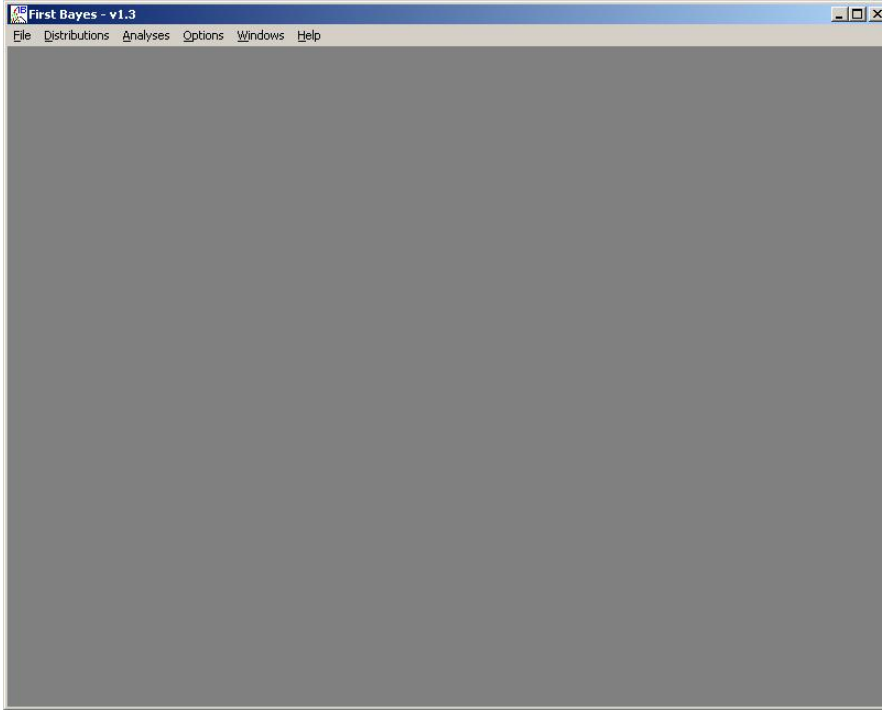


Figure 1: Menu options when opening First Bayes.

3.2 Distributions menu item

Picking any of the distributions under the **Distributions** menu brings up a new window that allows you to look at the properties and do calculations for that distribution. For example, if you click on the normal menu item under distributions, you will see a screen like that depicted in Figure 4.

In general, you enter information (mean, variance, weight) about the distribution on the right hand side of the screen, and information (distribution name, mean, median, variance, SD, mode, 25% and 75% quantiles) about your choice of distribution (and chosen parameters) pops up on the left hand side.

For purposes of our course, the weight parameter is always set to one. See the First Bayes manual for uses of this parameter, which essentially allows for arbitrary mixtures of distributions.

At any time, you can click on the “plot” function to see a plot of your distribution. You can also calculate the probability under the curve between any two points, and calculate 95% HDI (highest density interval) regions, by typing in your selections and hitting the “calc” button.

The t distribution is similar to the normal. Note that it is a three parameter family, so that the median, scale and weight must be filled in. To get standard t tables, fill in 0 for the median, 1 for the scale, and 1 for the weight. Typing in the degrees of freedom gives you the information for the corresponding t distribution. See Figure 5.

Another example appears in Figure 6, for the beta density. This distribution will be covered in Section 5 of our course. Here one can enter any two beta distribution parameters

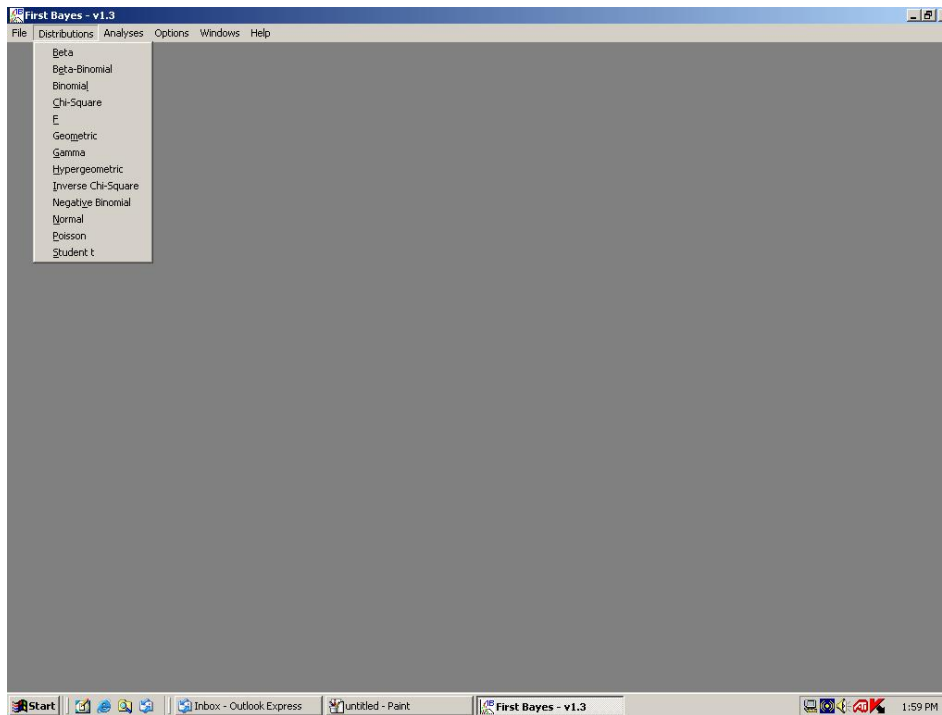


Figure 2: Thirteen different distributions are available.

(with weight one) and plot/investigate this distribution.

As one last example, Figure 7 shows a binomial distribution. Note that this is a discrete distribution, but everything works the same way. Note that this replaces the need for binomial tables, as one can calculate the probability for any set of values.

3.3 File menu - Entering data and Printing Plots

Any plot created in First Bayes can be sent to a printer by making that plot the active window (by clicking on it) and using the “Print Active Plot” menu item under **File**.

New data are entered in the **Data** menu item under **File**. You can select an existing data file (your list may differ from mine, as I have added data files over the years), change the values in these data files, create new data files, and load data for later analysis. You must choose a file name if you are creating new data before you can save it. One of the limitations of First Bayes is that data can only be created here, i.e., you cannot upload files from other programs. However, if you have a data set in ascii format, it can be cut and pasted into the new data window, and so entered without retyping. See Figure 8.

4 Analyses in First Bayes

We will briefly look at normal, binomial and regression analyses, as these will concern us most during the course.

Typically, one follows these steps in doing Bayesian analyses in First Bayes:

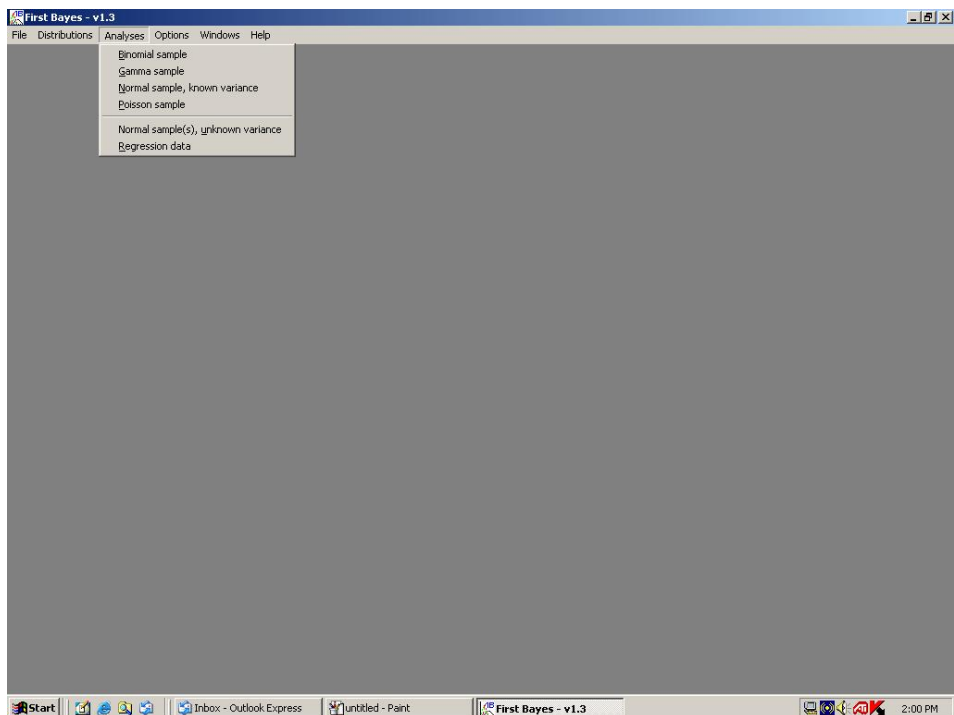


Figure 3: Seven different analyses can be carried out.

1. Enter some data, or use existing data. Make sure the data are loaded (should appear in the “Data Loaded” box).
2. Choose the analysis type you need to do from the **Analyses** menu.
3. Select a prior distribution. This is done from the analysis page. Click on the “Edit” box in the prior section on the right hand side of the screen to select your prior. You are automatically brought to a page with the corresponding *conjugate* prior distribution, whose parameters you must select. Note that for a normal mean, you must choose normal parameters, while for a binomial analysis, you choose beta parameters, and so on. When you are finished selecting your prior (you can go back anytime and change it), click on “Quit” to go back to the analysis. Remember that you can always look at the properties of your prior, to ensure it is appropriate. You can plot the prior, look at its HDI interval, and so on.
4. Once your prior is chosen, you need to load in your data. You do this simply by choosing one of the data sets you loaded (remember, you must load your data first).
5. Once you have selected your prior and loaded in your data, the posterior distribution is automatically calculated for you. You can then investigate its properties and calculate probabilities for it exactly as when investigating any distribution. For example, you can obtain point estimates (posterior mean) and 95% HDI’s for all parameters of interest.
6. Look at the triplot: Triplots display the prior distribution, the likelihood function, and the posterior distribution on a single graph. These are very useful for examin-

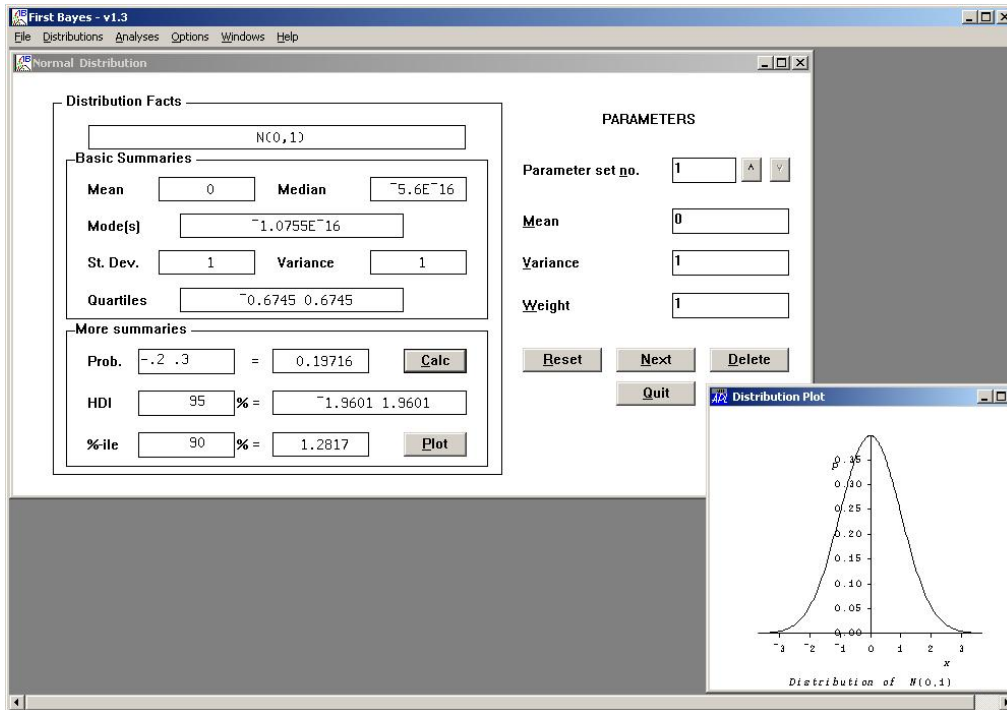


Figure 4: Examining the normal distribution

ing the effect of the data and the choice of prior on any posterior distribution. For example, for large data sets, the posterior usually resembles the likelihood function (most information comes from the data), while for smaller data sets, the prior has more influence.

7. If desired, make some predictions for future data. To do this, click on the “Show Predictive” button, which toggles the information on the right hand side of the screen from posterior to predictive. This is especially useful for regression (see below).

4.1 Normal Analyses

Following the above basic steps and choosing “Normal sample, known variance” from the analysis menu, one can do an analysis of data assumed to follow a normal distribution (i.e., a normal likelihood function). Figure 9 displays some results from a normal analysis, and Figure 10 displays the triplot for this analysis. Note that because the variance of the data is assumed known, you need to fill in a value for it at the top of the right hand side of the normal analysis screen.

4.2 Binomial Analyses

Similarly, following the above basic steps and choosing “Binomial sample” from the analysis menu, one can do an analysis of data assumed to follow a binomial distribution (i.e., a binomial likelihood function). Figures 11 and 12 show the results of a binomial analysis with

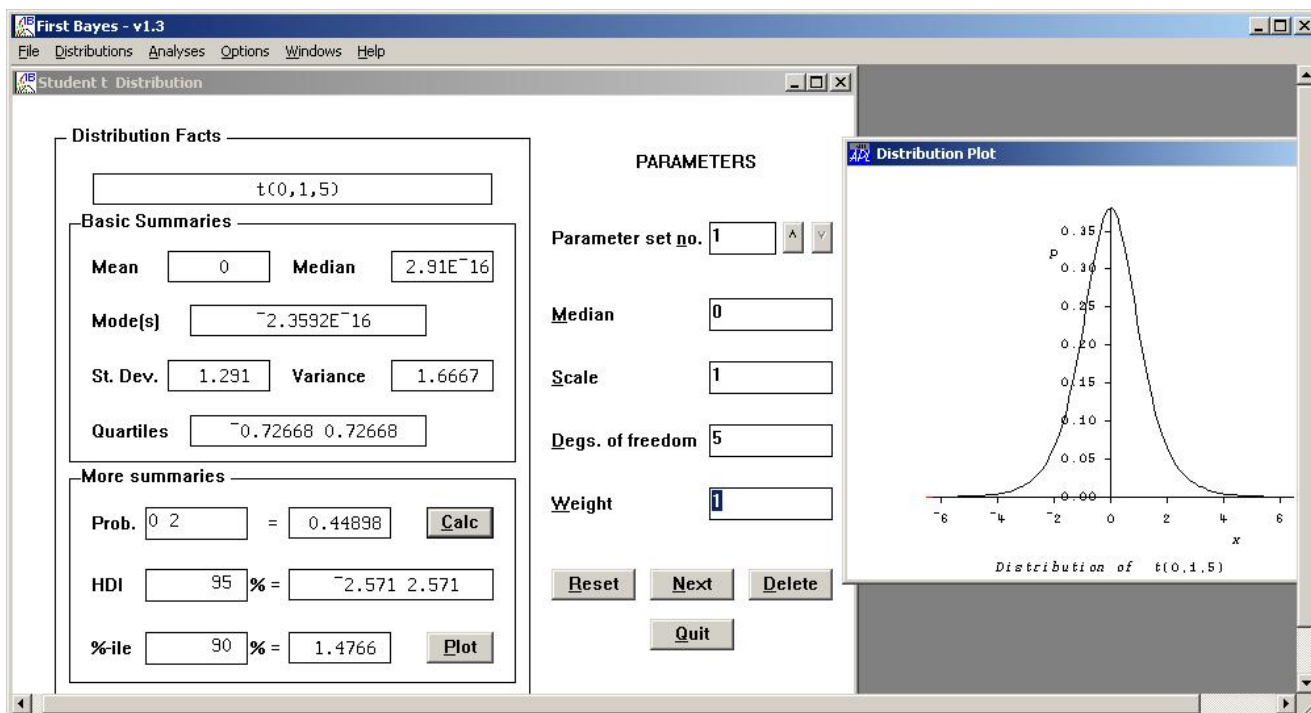


Figure 5: Examining the standard t distribution with 5 degrees of freedom.

corresponding triplot. Note that a beta prior distribution is used, as is standard for binomial problems.

4.3 Regression Analyses

Regression analyses differ from normal and binomial analyses in three important ways:

- First, there are three unknown parameters to estimate, as we have a regression slope (beta), an intercept to the regression line (alpha), as well as the residual variance about the regression line (sigma).
- Second, no prior information need be entered. First Bayes assumes noninformative priors by default, i.e., the posterior distributions are formed from the information contained in the data set, with little contribution from the prior, if any.
- Finally, two data sets need to be loaded and used, one for the dependent variable (Y) and one for the independent variable (X).

To see how regression analyses work, we will use two pre-existing data sets, CASTINGSX and CASTINGSY. Loading these two data sets in, choosing the regression menu item from the **Analysis** menu, and loading the two data sets in the data box in the regression window produces the window shown in Figure 13.

This provides (by default) the posterior distribution for the regression line slope, beta. Note that you can plot it, get 95% intervals, and so on, as usual. Clicking in the margin box

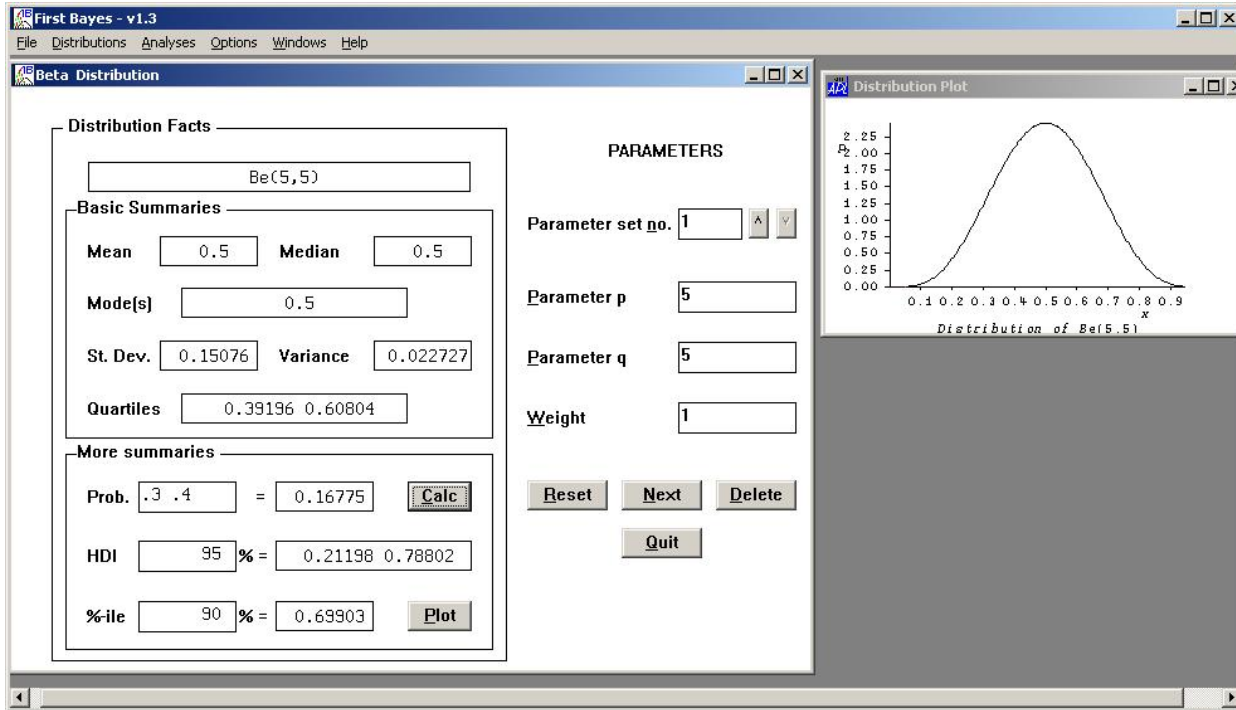


Figure 6: Examining the beta distribution

cycles you through the posterior distributions of the other parameters, alpha and sigma for the intercept and variance, respectively. Note how the left hand side of the page automatically updates each time. Note how the weight buttons are here set to one, by default, for whichever parameter you are currently looking at, and zero for the others.

To get predictions from the regression line, you need to use the weights button, and click on the “Show Predictive” button. For example, if you want to predict the next observation that occurs at $x = 70$, say, you need to have a weight of 70 for beta, a weight of 1 for alpha, and a weight of zero for sigma (because we want the prediction for $y = 1 \times \alpha + 70 \times \beta + 0 \times \sigma$). To enter these three weights, use the “up arrow” button near the right hand side of the screen, i.e., the one near the weight, not the one on the left side of the margin box. Use this up arrow to cycle through the three parameters and fill in the appropriate weight via typing in the box each time. If you accidentally hit the left rather than the right up arrow button, you will need to start over. See Figure 14 for the resulting posterior distribution of these predictions. Predicting for one observation shows you the posterior for this case, while increasing this number gets you the mean prediction for more than one observation.

One final note: You can click on “Scatter Plot” while looking at any regression analysis to see a plot of the data and two sets of predictive lines: The inner two lines are 95% HDI’s for the mean of regression line at any given x value, while the outer two are for predicting the next observation at that x value. See Figure 15. Similarly, you can also get a plot of the residuals. just click on “Resid”.

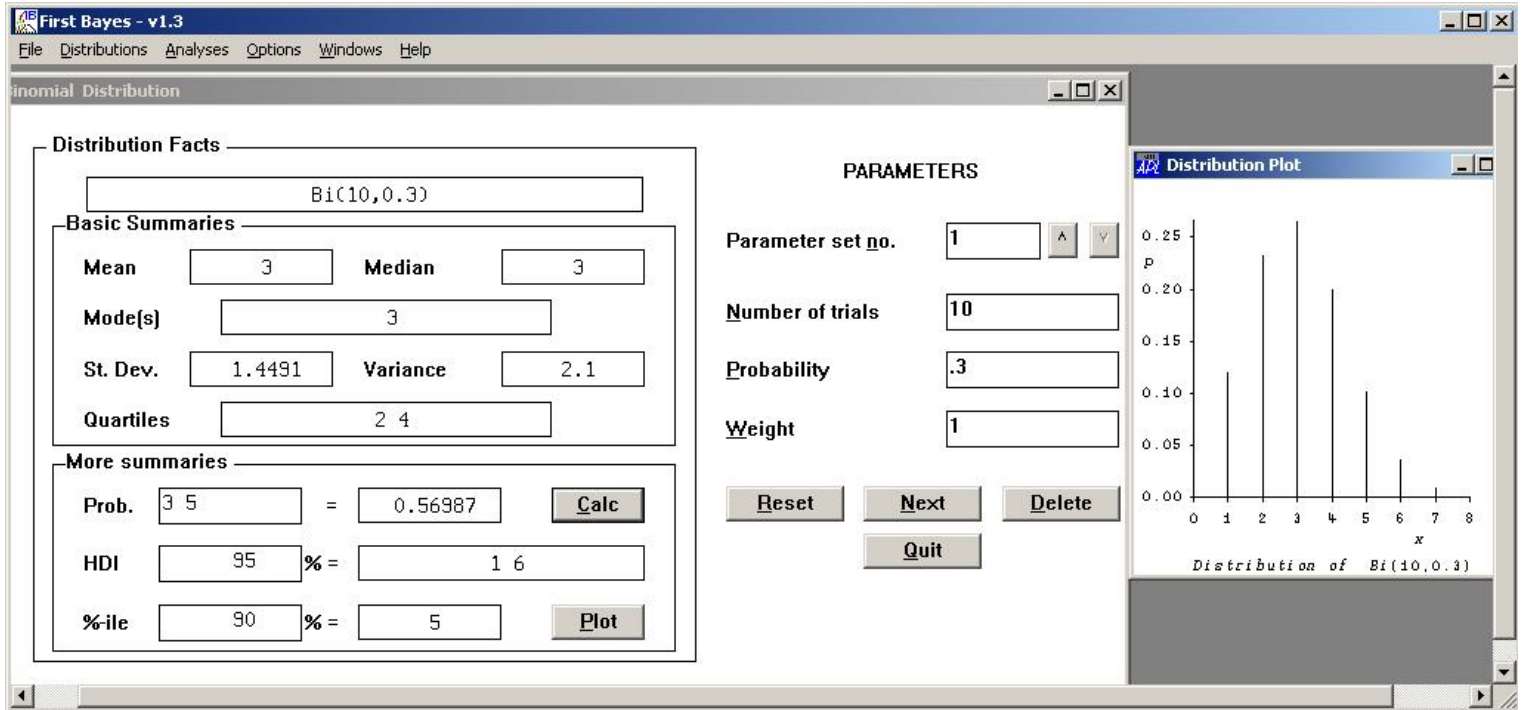


Figure 7: Examining the binomial distribution

5 Conclusion

This is of course only a very brief tour, and I have left out some nice features of First Bayes, including other distributions, other analyses, mixtures of distributions, generating random data from any distribution, and so on. Please refer to the First Bayes manual for more details.

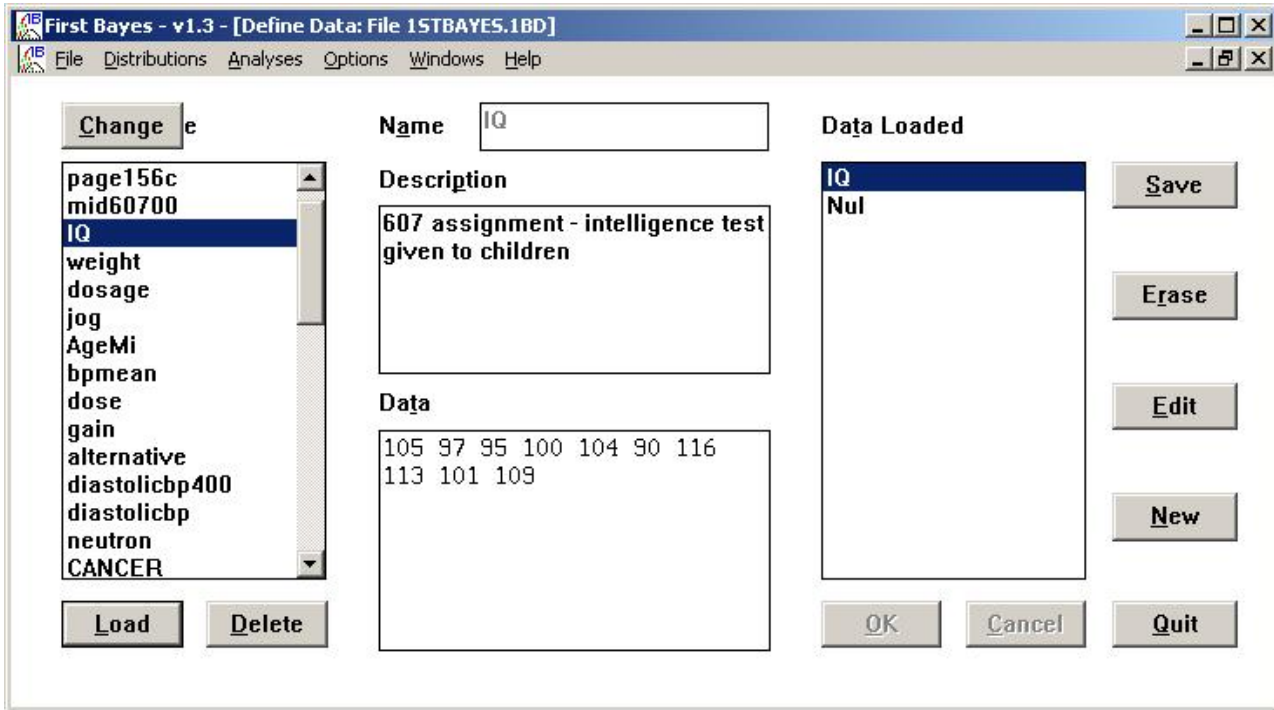


Figure 8: Entering new data or calling up previously entered data sets.

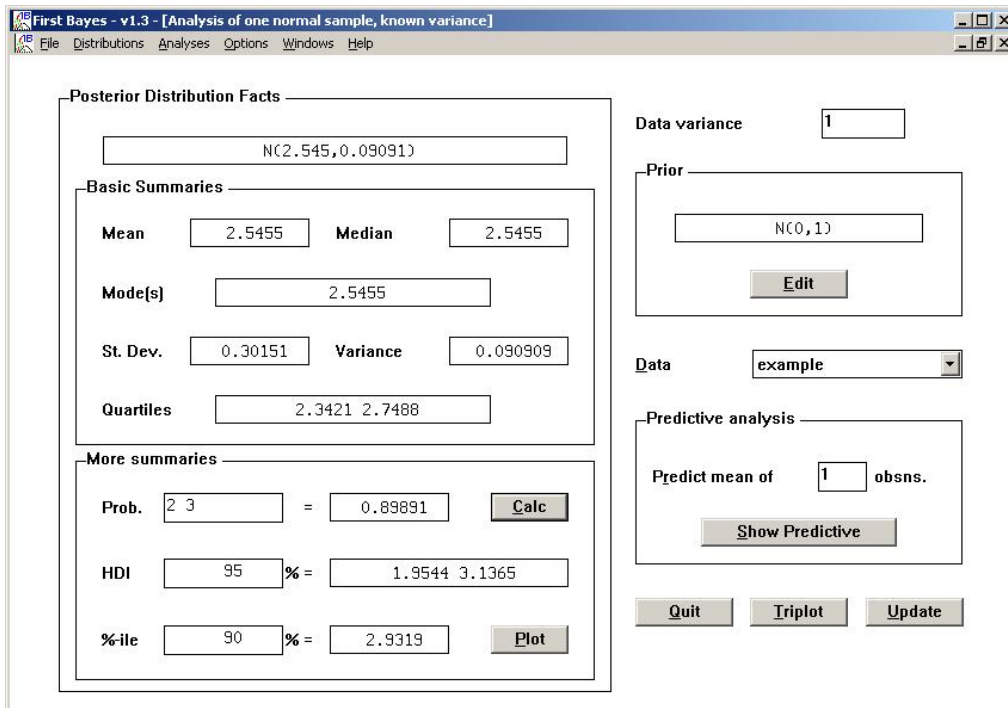


Figure 9: Analysis of a normal mean.

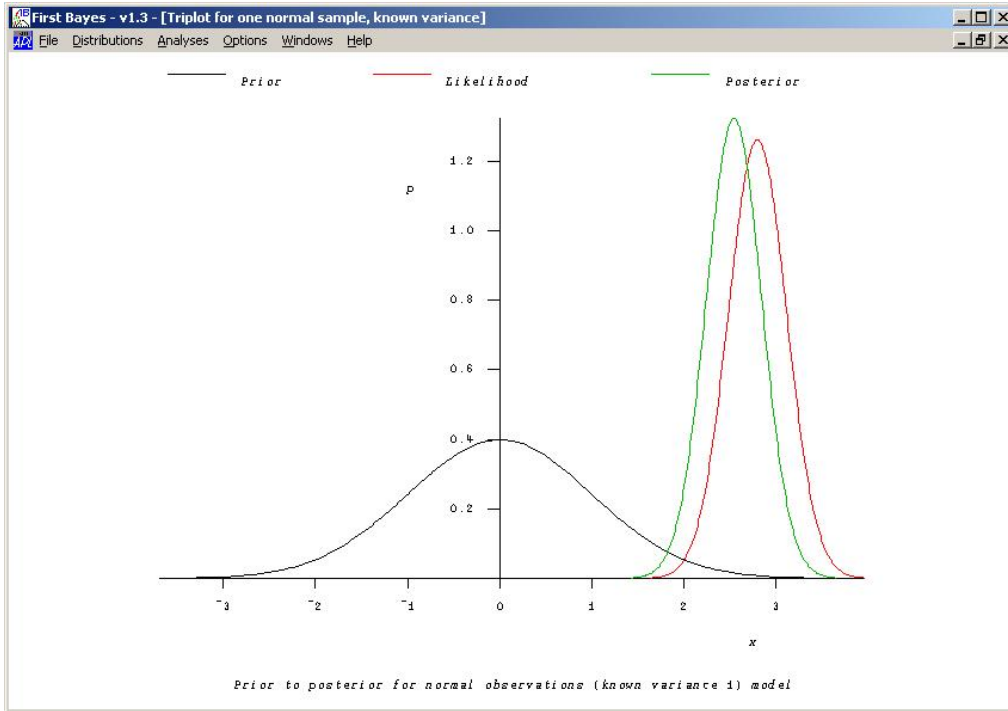


Figure 10: Plot of the prior, likelihood, and posterior distribution, a so-called tri-plot, for a normal analysis.

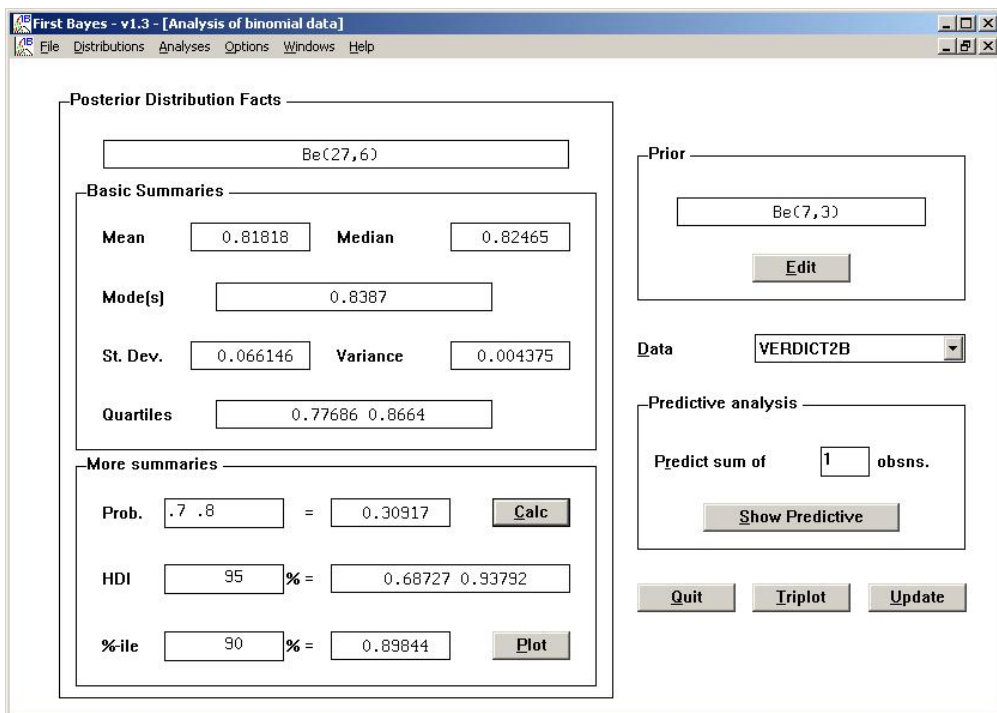


Figure 11: Analysis of a binomial mean.

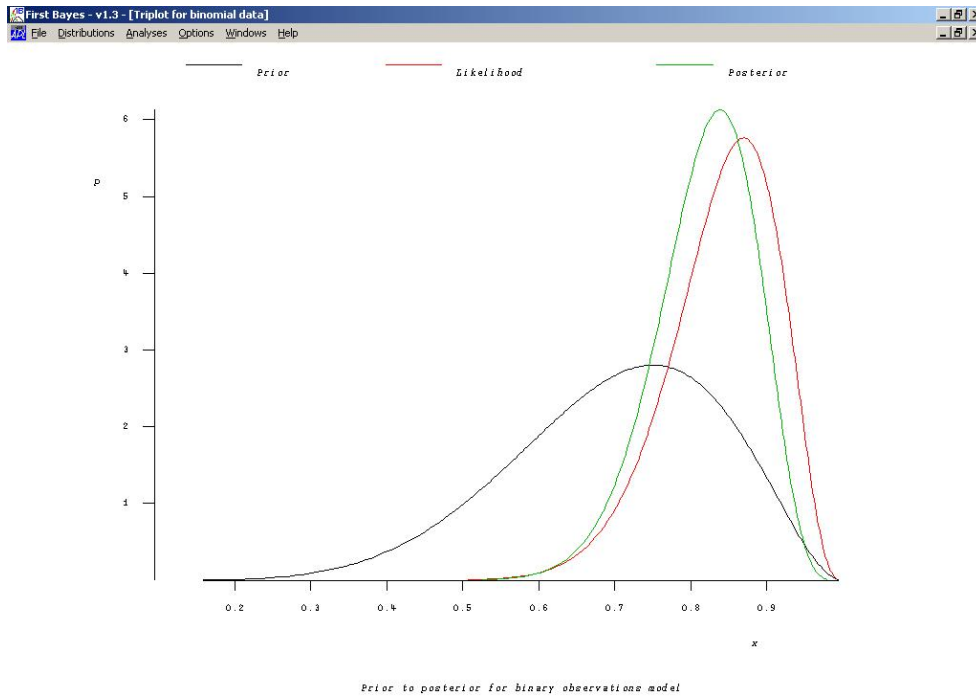


Figure 12: Plot of the prior, likelihood, and posterior distribution, a so-called tri-plot, for a binomial analysis.

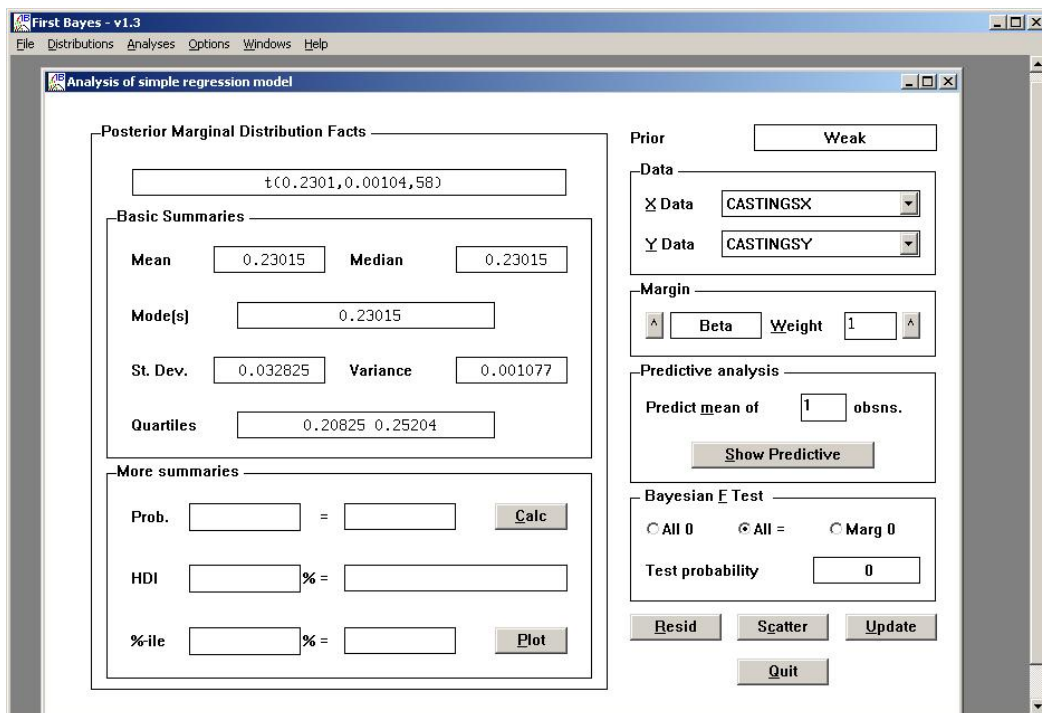


Figure 13: Regression analysis.

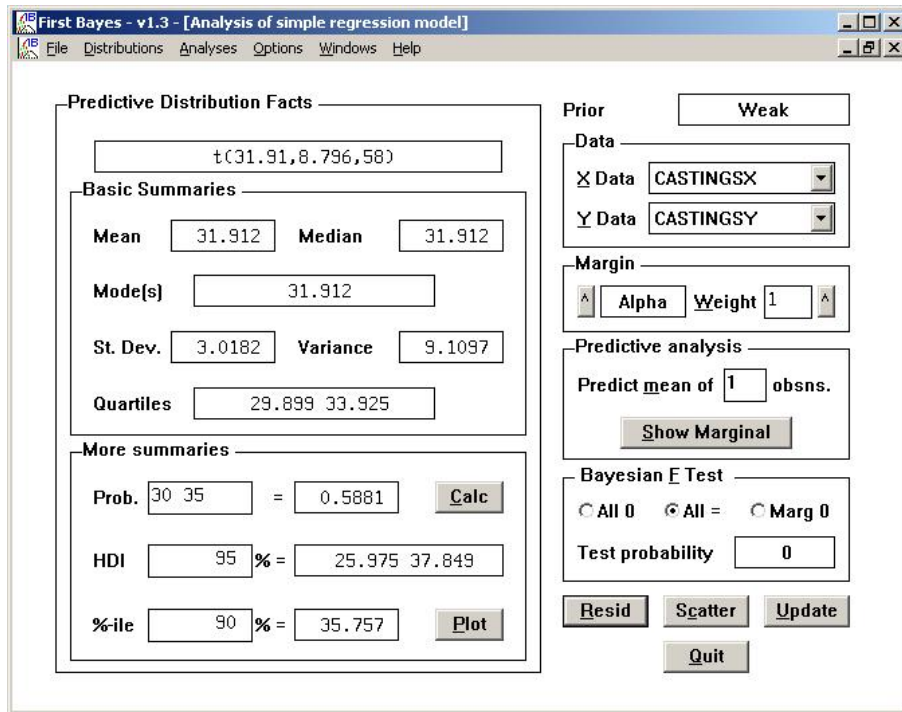


Figure 14: Predictions in regression analysis.

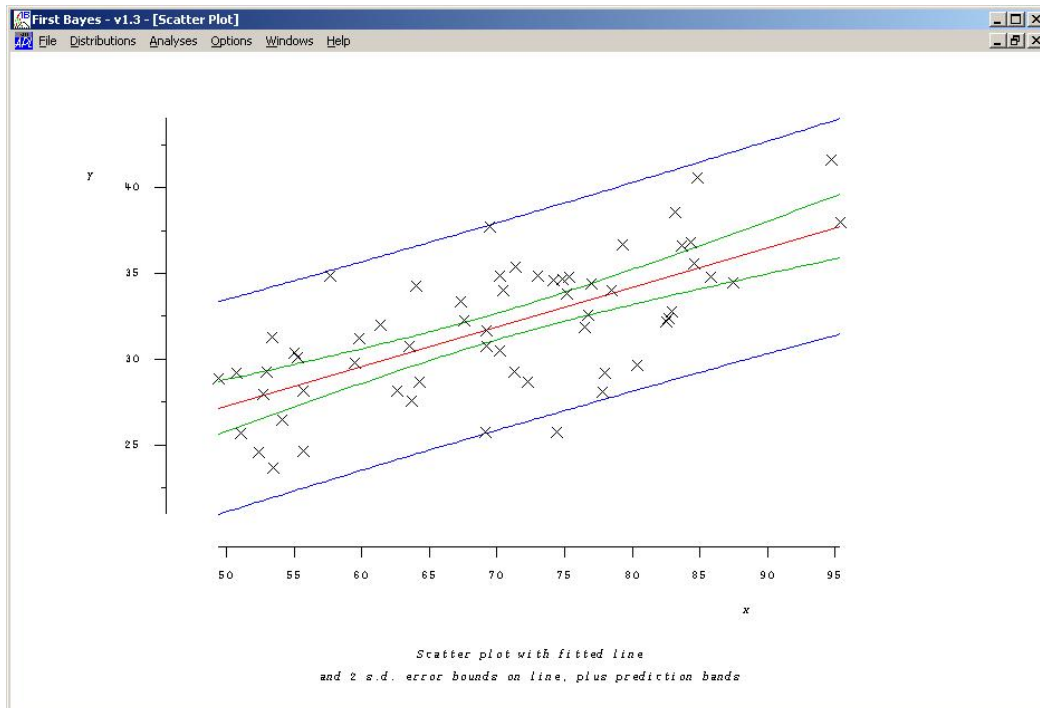


Figure 15: Regression analysis scatter plot.