Statistics: Continuous Methods

STAT452/652, Spring 2013

Computer Lab 2

Thursday, February 7, 2013
DMS 106
1:00-2:15PM

Empirical cdf, quantiles,
probability plots, quantile-quantile plots with

Instructor: Ilya Zaliapin

Topic: Empirical cdf, quantiles, probability plots, q-q plots

Goals: Learn how to

- construct and interpret empirical cdf,
- find theoretical and empirical quantiles of a rv,
- construct and interpret probability plots,
- construct and interpret quantile-quantile plots.

Assignments:

Use the data file Lab2_data_sets.MTW from the lab webpage. It consists of five samples (column-wise): Normal, Exponential, Uniform, $F(x) = x^2$, and one that is neither of above.

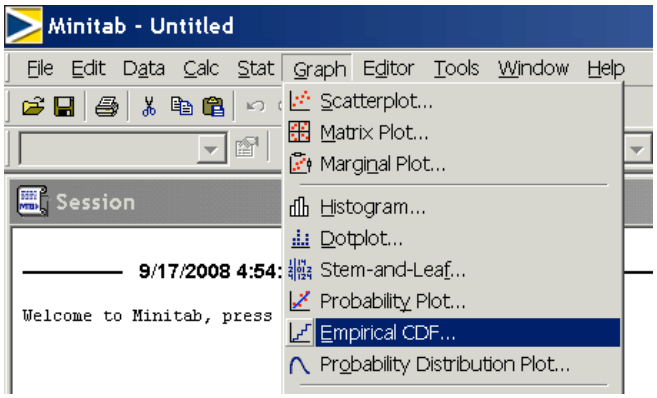
1. Use the ecdf approach to find the Normal and exponential samples;
2. Use the probability plot approach to find the Normal and exponential samples (be sure your results are consistent with that of assignment 1);
3. Use the quantile-quantile plot approach to find the Normal and exponential samples (be sure your results are consistent with that of assignments 1,2);
4. Find the theoretical 0.7 quantile of the exponential distribution with parameter 3; find the empirical 0.7 quantile of an exponential sample with the same parameter. Compare, explain and illustrate the difference in terms of the ecdf plot.
5. Generate 100 rvs with cdf $F(x) = 1-(1-x)^3$, $x \in [0, 1]$. Show the respective ecdf and the theoretical cdf in the same axes.

Report:

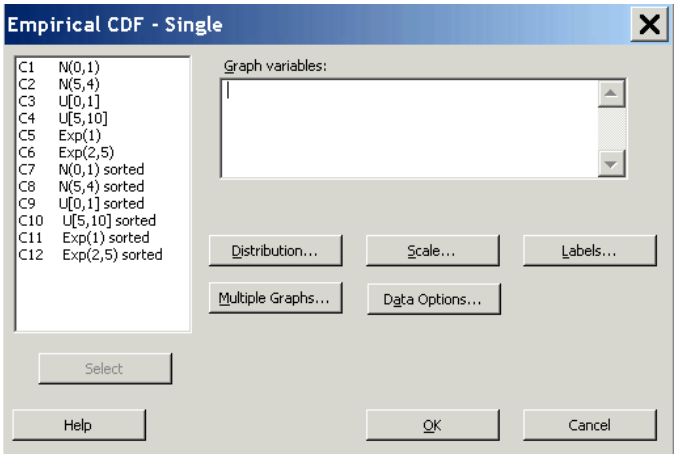
A printed report for this Lab is due on **Thursday, February 14** in class. BW printouts are OK. Reports will not be accepted by mail.

1. Empirical cumulative distribution function (ECDF)

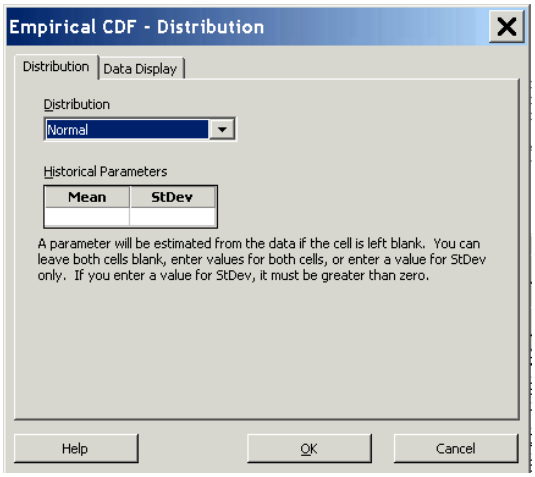
To compute ecdf for a given data set, use the menu **Graph/Empirical CDF…**



… choose the variable(s) to use in the following submenu…

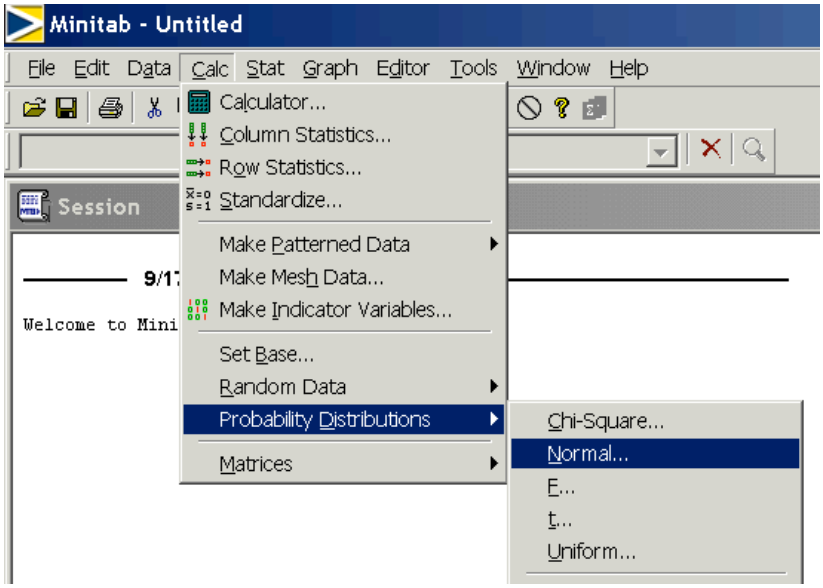


… and specify the "**Distribution…**" to which you want to compare your ecdf:

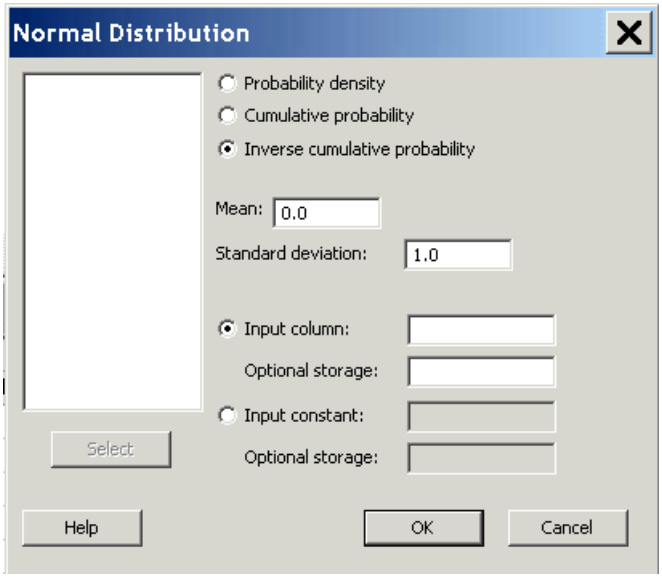


2. Theoretical quantiles

To find *theoretical* quantiles for one of the standard distributions, go to **Calc/Probability Distributions** and choose a cdf to work with:



In the following submenu, hit **Inverse cumulative probability** button and choose distribution parameters:

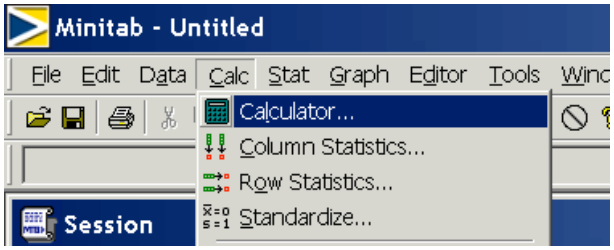


You can use several p-values stored in the data worksheet (option **Input column**), or enter a single p-value (option **Input constant**).

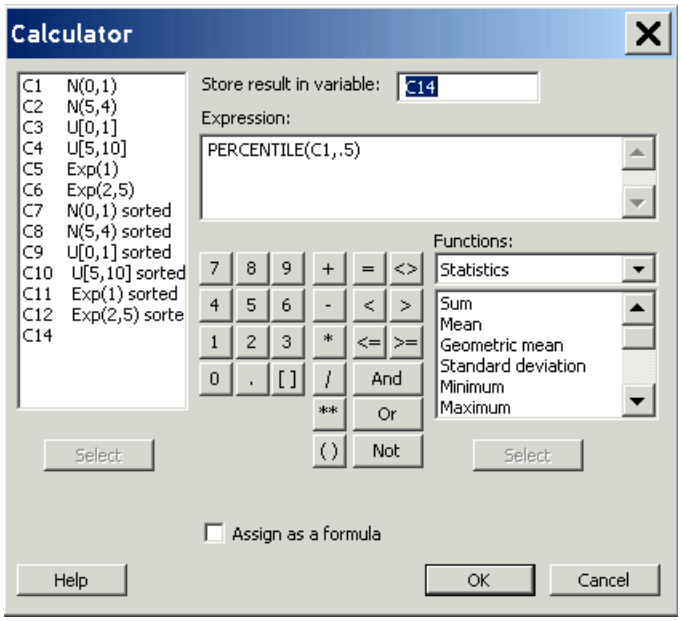
3. Empirical quantiles

To find *empirical* quantiles for a data set in the worksheet, go to

Calc/Calculator:



..and use the function **PERCENTILE(variable, probability)**

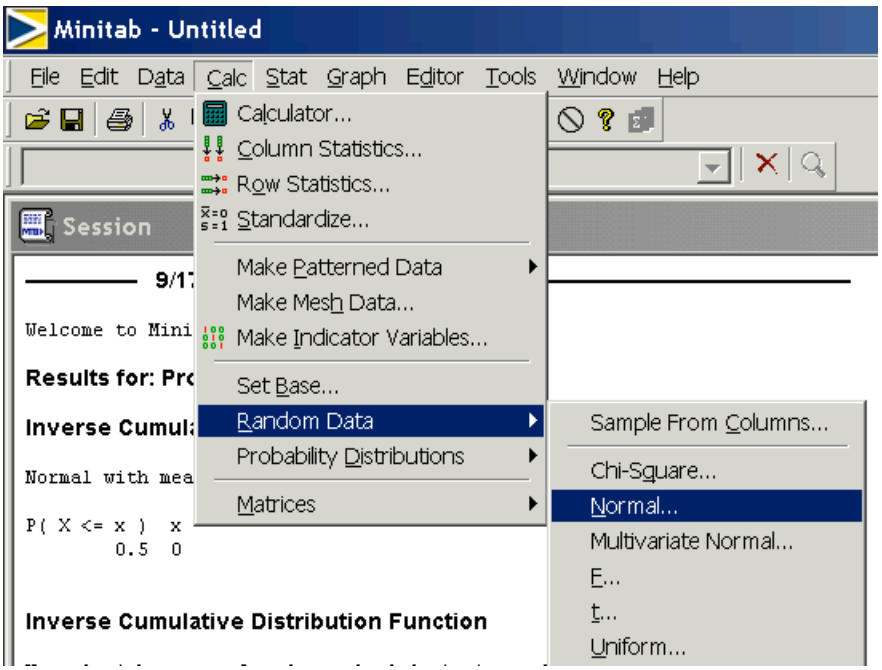


The output will be stored in the worksheet. You can use a single p-value by entering it in the function, or several p-values from a column in the worksheet.

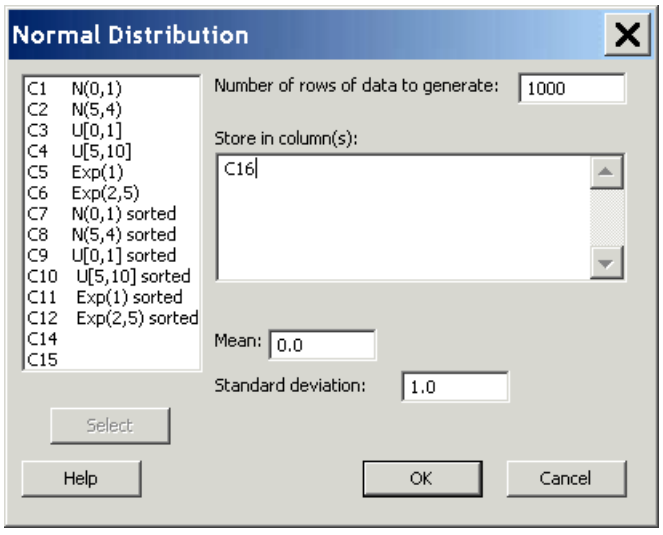
Do not get confused: the function is called **percentile**, BUT it asks for **probability**, NOT **percentage** as its second argument!

4. Generating rvs using standard Minitab routine

To generate an iid sample from one of the standard distributions, go to menu **Calc/Random Data** and choose the distribution to use…



… specify the number of values to generate, the column to store the data, and distribution parameters in the next window:



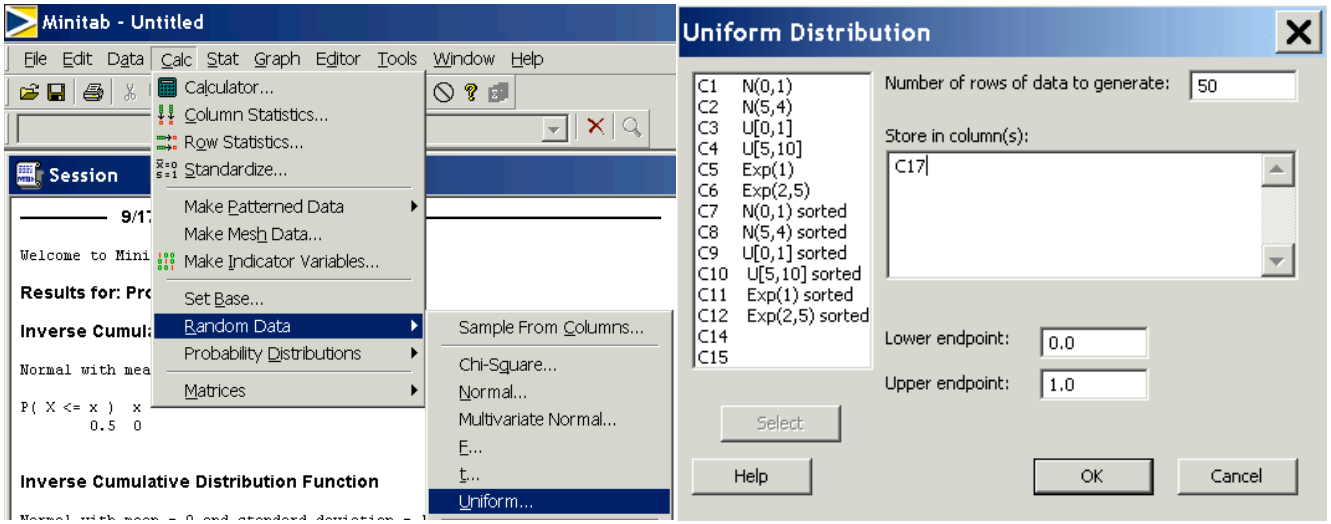
5. Generating rvs using the inverse cdf method

To generate N random variables from a (non-standard) cdf $F(x)$

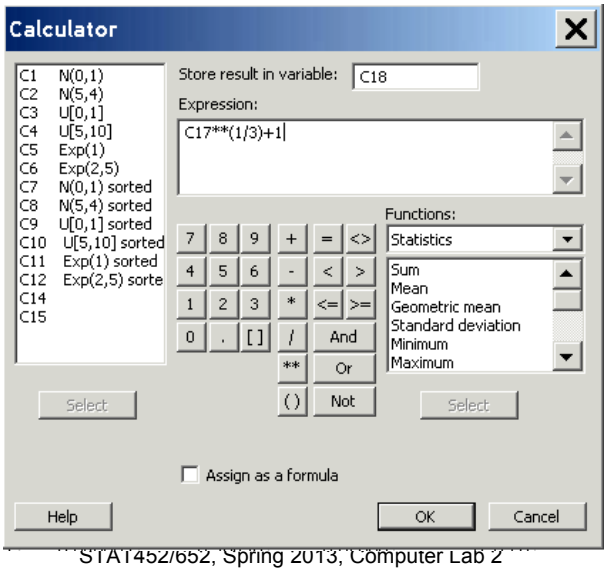
1. Generate N random variables U_i from the uniform distribution on [0,1]
2. Find the inverse cdf (quantile function) $F^{-1}(p) = Q(p)$
3. Compute $Q(U_i)$ using the menu **Calc/Calculator**

Example: Generate 50 rvs X_i with cdf $F(x) = x^3$

1. Go to **Calc/Random Data/Uniform** and choose appropriate parameters of the uniform distribution:

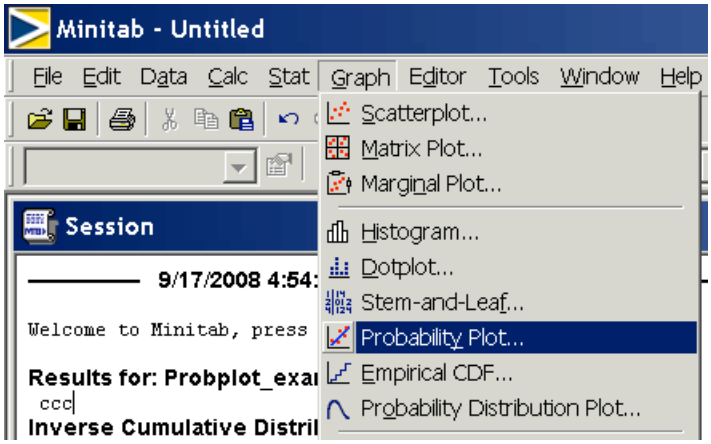


2. Find the inverse cdf $F^{-1}(p) = Q(p) = p^{1/3}$
3. Go to **Calc/Calculator** and calculate the values of X_i using U_i (stored in C17)

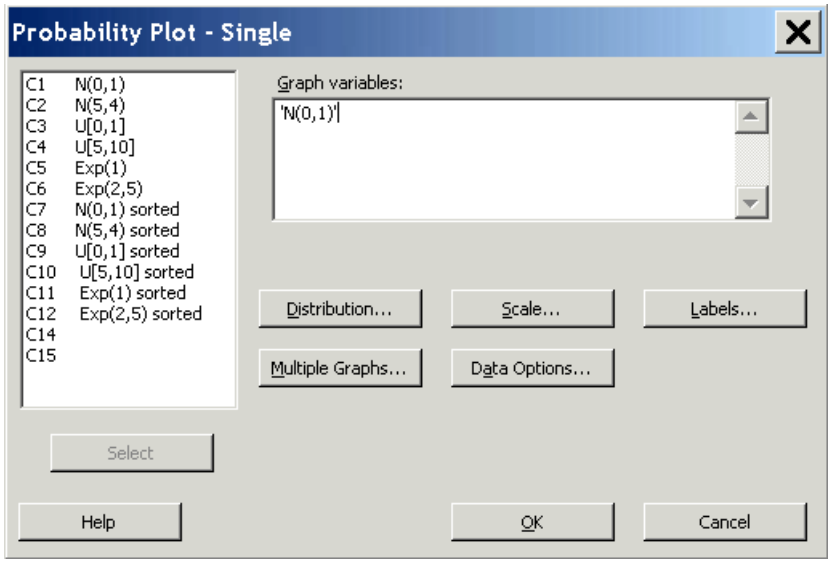


6. Probability plot

To create a probability plot, go to **Graph/Probability Plot…**



… choose the variable to work with and the "**Distribution…**" for the probability plot:



7. Quantile-quantile plot

Minitab does not do quantile-quantile (qq) plot. However, a simple version of qq-plot can be done by using **Graph/Scatterplot** option:

1. Sort two data sets with the same number of observations using **Data/Sort** option:

2. Use **Graph/Scatterplot** to plot the **sorted** data vs each other.

The linear shape of the qq-plot indicates that two data sets may be coming from the same distribution.