In class, we've briefly described one set of Excel tools that can be used to create, run, and evaluate the results of Monte Carlo simulations. This note reviews what tools are used; it also identifies some additional functions in Excel that can be useful in performing the same tasks "from step one" so you can see the process that Excel is automating for us. This note assumes you have already read the textbook section on simulation and that you understand the spreadsheet functions and commands the chapter describes.

The Process of Simulation

The first step in creating a simulation is to *model the system* we wish to analyze and to define the elements of the system's behavior which we would like to describe. In *Monte Carlo* simulations, which model *stochastic* systems (systems in which randomness exists), an essential element of the system's description is an explicit model of the nature of uncertainty. Specifically, we use *random variables* (RV's) to define the stochastic elements in the system.

After constructing the simulation model, the second step is to *run the simulation* a (large) number of times. Each time we run the simulation, we sample from each of the random variables used to describe the system and recalculate the effect on the elements whose behavior we wish to characterize. Sometimes we call each run of the simulation a "trial."

The last step in the process is to *analyze the results*. In this phase, we view the result of each trial of the simulation as if it were a sample from an infinite population of simulation trials that we might have run. Accordingly, we can use standard statistical techniques to calculate best estimates of the population mean and of the standard deviation of the true population mean from our best estimate:

number of simulation trials: n

result of the i^{th} trial of the simulation: x_i

sample mean: $\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i$

sample standard deviation: $\hat{s} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(x_i - \bar{x}\right)^2}$

standard error of our estimate of the population mean: $\hat{s}_{pop} = \hat{s} / \sqrt{n}$

confidence level: *e.g.* 95% confidence $\Rightarrow$ z = 1.96

confidence interval for the population mean: $\bar{x} \pm z\,\hat{s}_{pop}$

Note that the larger the confidence interval, the less certain we are about the accuracy of our estimate of the expected value of the output variable that we've analyzed.

A frequency histogram provides a complementary way in which we can characterize the result of running a large number of simulation trials. While it does not provide a numerical measure of the accuracy of our estimated mean, it does provide a full view of the range of outcomes present in our simulation trials. Indeed, the shape of the distribution of outcomes often gives us a information concerning the possibility of big gains or losses. This is information that estimates of the mean alone do not provide, and an important reason why simulation might give us a richer analysis than an analytical model of the same system.

A Review of the Tools Covered in Class

In class we have showed one set of tools that Excel provides for the three-step process we've just described—built-in functions originally intended for statistical analysis, but which are also useful in simulation. We briefly describe these tools here, for your reference; for a fuller description of *how* they work, you must trudge through Excel's on-line help or user manual.

The [Tools | Data Analysis | Random Number Generation] Command: The first command cluster is [Tools | Data Analysis | Random Number Generation] ([T|D|R]). When you select Data Analysis from Excel's *Tools* menu, you will be given a list of options, including *Random Number Generation*. (You may need to inform Excel that it should load the Data Analysis add-in.) You will find that the Random Number Generation menu offers the option of sampling from Uniform, Normal, and Bernoulli distributions, as well as from Binomial, Poisson, and general Discrete distributions. (It even has something called a "Patterned" distribution, which we have not yet figured out!) The [T|D|R] command will generate one (or more) columns filled with samples from the distribution of your choosing.

Unlike the case when using the `=RAND()`function, the result of using the [T|D|R] command is a column of cells filled with *numerical values*, (rather than formulas that get recalculated each time the spreadsheet is updated). For example, to perform 500 trials of a simulation that requires 20 normally-distributed random variables, we can use the Normal Distribution menu within [T|D|R]. Each trial's random returns might occupy the first 20 columns of one row in the spreadsheet.

Finally, there is one last advantage to using the [T|D|R] command: repeatability. Each time you use the [T|D|R] command, you may select a *random seed*, a number that determines the entire set of random numbers the [T|D|R] command will generate. Each time you use the same random seed to sample a set of random variables in [T|D|R], you get the exact same *sequence* of random numbers. This is to say that every time the "dice get rolled," they always "come up the same" as long as you're using the same random seed. Different random seeds, of course, generate different sequences of random numbers.

The ability to regenerate the same sequence of random numbers can be useful when we want to do a careful comparison of related systems. In particular, this ability to compare related systems on a trial-by-trial basis is most useful when we wish to compare stochastic systems whose behavior differs only because of some factor, or *policy*, we control. Indeed, simulation is often used to compare the effectiveness of different policies that a manager is considering implementing to manage a stochastic system. In just these situations, this trial-by-trial comparison can be quite handy.

The Descriptive Statistics and Histogram Commands

Descriptive statistics and histograms are located under the Tools menu, specifically at *[Tools | Data Analysis | Descriptive Statistics]* and *[Tools | Data Analysis |Histogram]*. The *Descriptive Statistics* subcommand generates calculations of the mean, median, variance, and similar quantities about your simulation data, whereas the Histogram command generates a graph of the relative frequencies of outcomes from your simulation. Excel gives you several options for customizing your histograms or descriptive statistics; refer to its on-line help facility for specific details while the *Histogram* or *Descriptive Statistics* dialog box is open.

Extra for Experts : **Doing Simulations in Excel from Scratch**

We briefly describe two refinements to the process of constructing simulations. The first is a direct method of generating random numbers that lets you control the actual sequence of numbers being generated. The second is a method of generating "live" frequency histograms that can change as the results of your simulation change. We briefly describe the two only briefly; for a fuller description of how they work, you must trudge through Excel's help windows or user manual.

For the first step we could use the `=RAND()` function in Excel to generate a random variable Uniformly distributed between 0 and 1. It follows that the formula `=IF(RAND()<p, "TRUE", "FALSE")` can be used to sample a *Bernoulli* random variable, which models a TRUE-FALSE event which is TRUE with probability *p*. For example, we could use the Bernoulli random variable to model getting "Heads" when tossing a coin, where TRUE corresponds to "Heads" and p=0.5, or the process of getting 6 on a single die, where TRUE corresponds to "the '6' face is up" and p=1/6. Similarly, we can use the formula `=NORMINV(RAND(), μ, σ)` to sample from a Normal RV with mean μ and standard deviation σ.

There are two ways in which the outcome of a simulation model can be replicated to produce a large number of trials when using Excel's `=RAND()` function. The first way is to [1] create the model within one row of the spreadsheet; [2] copy that single row "down" in the spreadsheet to create the desired number of trials. This technique works because each time the original row is copied to a new row, all of the cells in the row that contain `=RAND()` functions calculate new samples from the Uniform distribution.

The second way to replicate the outcome of a simulation model is to use Excel's [Data | Table] command. Using this technique, we tell the [Data | Table] command which cell in the simulation model represents the "result" whose behavior we want to characterize, as well as a range of cells into which the result should be recalculated. The trials for both a simple Bernoulli simulation can be calculated in either manner. We might create a single cell that contains the formula `=IF(RAND()<0.5, 1, 0)` and then use the [Data | Table] command to replicate the result the desired number of times.

In the first part of this note, we emphasized the repeatability property when using the [T|D|R] menu in Excel. With the `=RAND()` function, by contrast, you have little control over how the random numbers are generated. The only way you can fix the sequence of random variables is to copy their values; otherwise every time the spreadsheet is updated their values change. Furthermore, there is no way to regenerate the same sequence of random numbers. If you want to use the same sequence or random samples more than once, you must *be sure to save your random numbers* from one time to the next; otherwise, they will vanish.

Finally, for the third step, the first part of this note presented two commands in the [Tools | Data Analysis] menu. *Descriptive statistics* automatically calculates the confidence interval for the estimated sample mean, and *Histogram* automatically creates a chart of the underlying frequency distribution of trial results of the simulation. Another way to generate a histogram is to use the `=FREQUENCY()` array function.

The =FREQUENCY() Array Function allows us to automatically regenerate frequency histograms from a data table each time the table's number are updated. More specifically, suppose that we do not want to perform the [Edit | Copy] and [Edit | Paste Special | Values] commands that turn [Data | Table] based simulation results into a column of numbers. Instead, we leave the `=TABLE()` array function intact in the cells which display the simulation trials. We then use the `=FREQUENCY()` array function to create a frequency distribution for the trials. Each time the spreadsheet is recalculated: [1] the `=TABLE()`functions used in the simulation recalculate a new set of results for the simulation (resampling from the `=RAND()` functions that generate the results); [2] the `=FREQUENCY()` functions recalculate, in turn, a new frequency distribution for the results which is based on the new trials calculated by the `=TABLE()`functions.

The automatic updating of the analysis can be further extended. If we create a histogram chart of the `=FREQUENCY()` functions, the chart will also be updated automatically each time the spreadsheet is recalculated. To automatically recalculate the sample mean, as well as its confidence interval, we can use the formulas presented in this note (and in class), rather than using the [Tools | Data Analysis | Descriptive Statistics] command. Again, by using the formulae, each time the spreadsheet is updated, a new sample mean and confidence interval will automatically be calculated.