# Albright’s DADM\_Tools Add-In

I created this add-in for our Business Analytics: Data Analysis and Decision Making textbook as a “lightweight” alternative to Palisade’s DecisionTools Suite. Based on market research, we have found that a sizable percentage of our users prefer not to use the Palisade software, for a variety of reasons. Perhaps some of you will choose to use the DADM\_Tools add-in instead, for the following reasons:

* It is free to everyone and can be downloaded from <https://kelley.iu.edu/albrightbooks/Free_downloads.htm>.
* Although it has many fewer features than the Palisade software, its features are sufficient for the book.
* Because it has fewer features, it is very simple to use, with virtually no learning curve.
* It works with Excel for Mac, whereas the Palisade software does not.

The rest of this document discusses and illustrates the tools in the add-in.

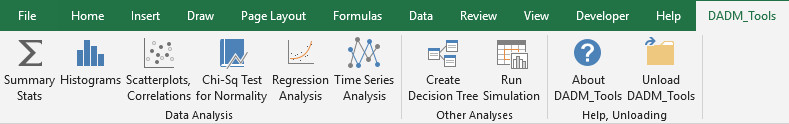
* [Loading and Unloading DADM\_Tools](#_Loading_and_Unloading)
* [Common Features](#_Common_Features)
* [Summary Stats](#_Summary_Stats)
* [Histograms](#_Histograms)
* [Scatterplots, Correlations](#_Scatterplots/Correlations)
* [Chi-Sq Test for Normality](#_Chi-Sq_Test_for)
* [Regression Analysis](#_Regression_Analysis)
* [Time Series Analysis](#_Time_Series_Analysis)
* [Create Decision Tree](#_Decision_Trees)
* [Run Simulation](#_Run_Simulation)
* [Random Functions](#_Random_Functions)

## [Loading and Unloading DADM\_Tools](#_Albright’s_DADM_Tools_Add-In)

You load DADM\_Tools like any add-in. The easiest way is to first make Excel’s Developer ribbon visible, if it isn’t already, as follows:

* **Excel for Windows:** Right-click any ribbon, select Customize the Ribbon, and check the Developer item in the right pane.
* **Excel for Mac:** Select Preferences from the Excel menu, then View, then check the Developer tab button.

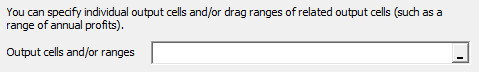
Then click Excel Add-Ins on the Developer ribbon to see the add-ins list. If DADM\_Tools is in the list, check it. Otherwise, click Browse to find the DADM\_Tools.xlam file. Once DADM\_Tools is checked in the add-ins list, you will see a DADM\_Tools tab with the following ribbon:



Each of the first eight buttons on this ribbon runs a program, as discussed in the sections below. The Unload button lets you unload the add-in, which you can also do by unchecking the DADM\_Tools item in the add-ins list.

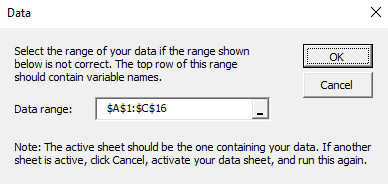
## [Common Features](#_Albright’s_DADM_Tools_Add-In)

* Except for the decision tree program, each of the DADM\_Tools programs creates a new worksheet for each run, with a name corresponding to the program and the source data sheet. For example, if you run Summary Stats based on data in a sheet called Data, the results will be placed on a sheet named Data\_Summ1. If you run another Summary Stats based on the same data, the results will be placed on a sheet named Data\_Summ2, and so on. (The exception to this, the Decision Trees program, places the decision tree on the same sheet as the source data.) You can rename these results sheets as you like.
* The results sheets just discussed sometimes use range names. However, these are all created as *worksheet-level* range names, so they don’t interfere with any existing range names in your workbook.
* Whenever possible (the regression program is an exception), the results are linked to the source data with regular Excel formulas. For example, the Decision Tree program uses formulas on its branches that indicate exactly how the folding-back process works. Such straightforward formulas are useful for learning purposes.
* The DADM\_Tools dialog boxes often ask you to select a cell or a range, such as the following:



To select the range by clicking or dragging (the preferred way), you must *first* click the small “dash” button on the rightand *then* click the cell or drag the range. (**Technical note:** This is a modified version of Excel’s “RefEdit” control. RefEdit controls have been known to misbehave in Excel for Windows, and they don’t work at all in Excel for Mac. Therefore, I created this modified version to work in Windows and the Mac.)

* Each of the six programs in the Data Analysis group first asks you for a data range, as in the following dialog box:



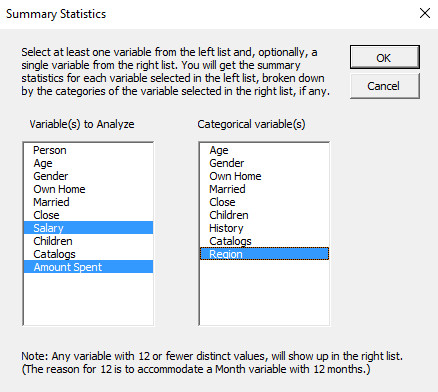
The program guesses the data range based on the current selection. Although you can change this data range, it is a good idea to select a cell inside your data set before running the program. If you do so, the guess is usually correct. (**Note:** Unlike Palisade’s StatTools add-in, there are no “defined” data sets in DADM\_Tools. Therefore, to run any of the data analysis programs, you must first activate the data sheet.)

* Several of the data analysis programs internally identify categorical variables for various purposes. These are defined as either text or numeric variables with 12 or fewer distinct values. (The reason for choosing 12 is to accommodate Month variables with 12 distinct months.) More will be said about the roles of categorical variables in the following sections.

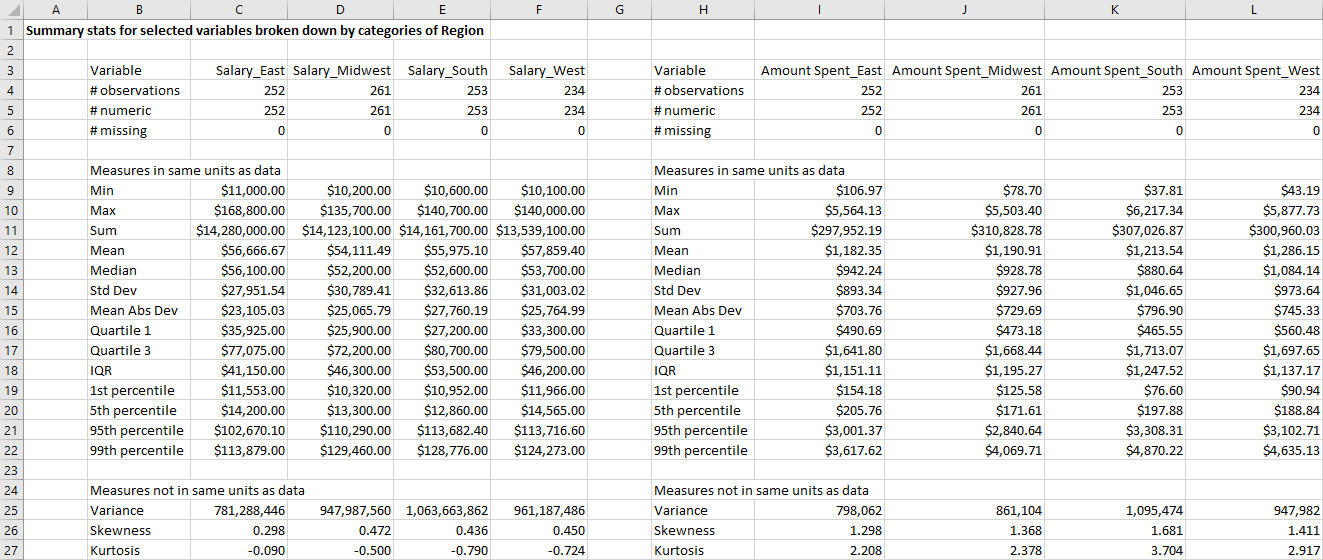
## [Summary Stats](#_Albright’s_DADM_Tools_Add-In)

This program uses built-in Excel functions to calculate summary measures for selected numeric variables in a data set, possibly broken down by a categorical variable. The steps are:

1. Click the Summary Stats button on the DADM\_Tools ribbon and then select the data set range.
2. Select one or more numeric variables from the left pane in the following dialog box:



1. If you want to break the summary stats down by a categorical variable, select one from the list on the right.
2. With the above choices, the results are as follows:



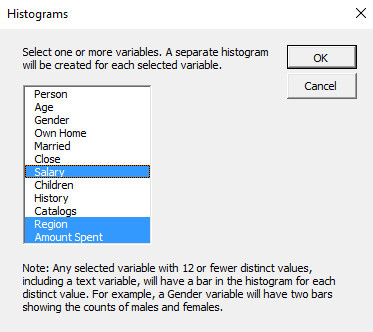
**Notes:**

* You have no choice for the summary measures included, but you can delete rows you don’t want.
* The formatting of the results is based on the formatting of the original data variables, but you can change it as you like.
* If you don’t break down by a categorical variable, the result formulas are straightforward, using built-in Excel functions such as AVERAGE, MEDIAN, STDEV.S, and so on. However, if you break down by a categorical variable, the formulas are more complex. They still use built-in Excel functions, but they are *array* formulas, basically of the form =STDEV.S(IF(*condition*,*data\_range*)). For example, if the condition is that gender is male, this finds the standard deviation of the males only. (No such array formulas would be necessary if Excel had functions like MEDIANIF, STDEVIF, and others like its AVERAGEIF function, but unfortunately, no such functions yet exist.)

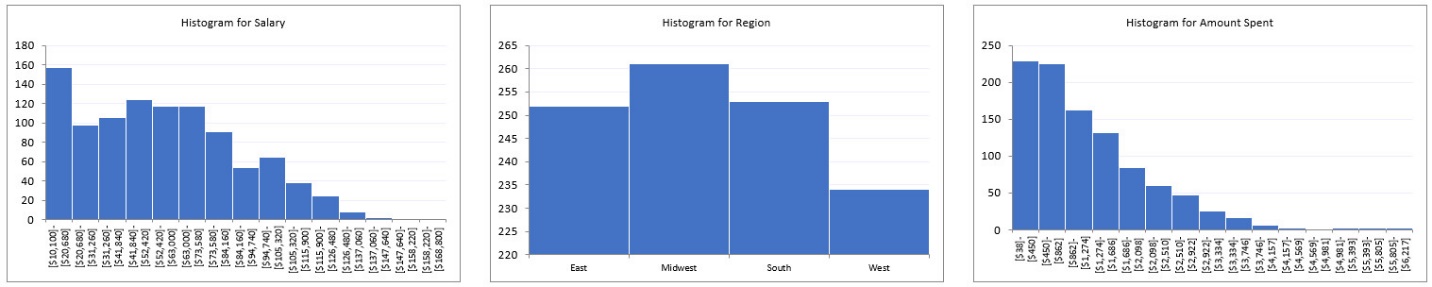
## [Histograms](#_Albright’s_DADM_Tools_Add-In)

This program creates a histogram for each selected numeric variable in a data set. For any selected variables identified as categorical (12 or fewer distinct values), it creates a column bar chart with a bar for each distinct value. These categorical variables can be text or numeric. The steps are:

1. Click the Histograms button on the DADM\_Tools ribbon and then select the data range.
2. Select one or more variables from the following dialog box:



1. With the above choices, the results are as follows. The first and third charts are histograms in the usual meaning of the term. The middle chart is a column bar chart of counts in the various regions.



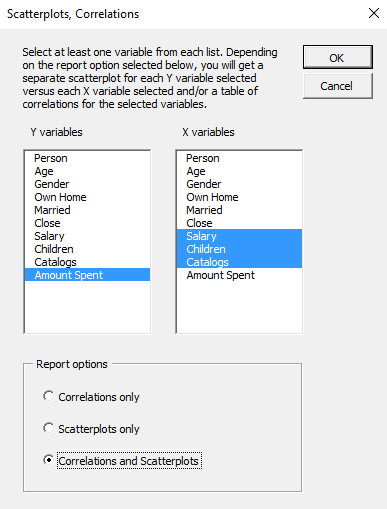
**Notes:**

* For numeric (non-categorical) variables, the bins are chosen automatically by the well-known Freedman-Diaconis rule from the statistics literature. The maximum number of bins is 15.
* The bin information is hidden behind the charts in white font. However, if the data in the data set change, the bins, and hence the charts, change automatically.

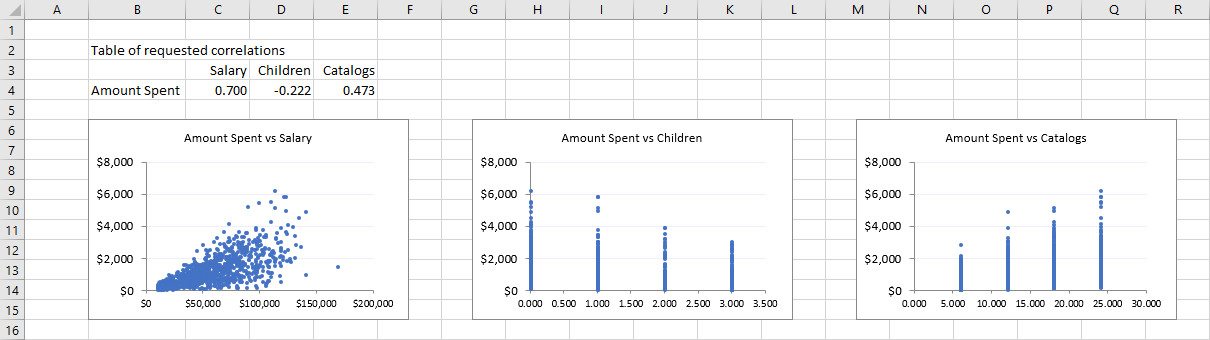
## [Scatterplots, Correlations](#_Albright’s_DADM_Tools_Add-In)

This program creates a scatterplot for each pair of selected Y and X variables and/or a table of correlations between the selected variables. The steps are:

1. Click the Scatterplots, Correlations button on the DADM\_Tools ribbon and then select the data range.
2. Select one or more Y and X variables from the lists in the following dialog box.



1. With the above selections, the results are as follows:



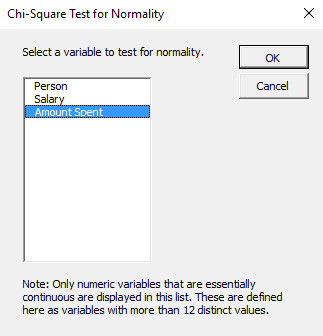
**Notes:**

* The correlations are created in the usual way with Excel’s CORREL function.
* You can select the *same* Y variables and X variables to obtain a square “matrix” of scatterplots and/or correlations.
* If you ask for correlations only, the suffix on the results sheet will be Corr, as in Data\_Corr1. Otherwise, it will be Scat, as in Data\_Scat1.

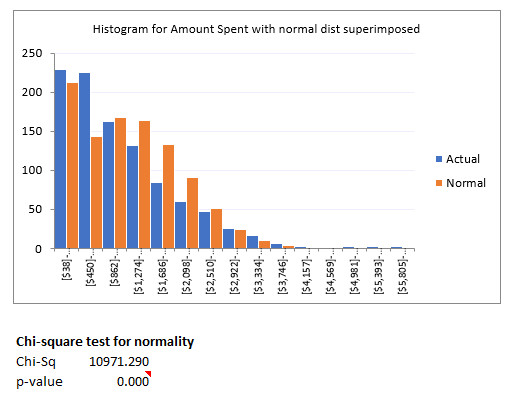
## [Chi-Sq Test for Normality](#_Albright’s_DADM_Tools_Add-In)

This program extends the Histograms program by comparing a histogram of the data with the histogram of data from a normal distribution with the same mean and standard deviation and reporting the results of a chi-square goodness-of-fit test for normality. The steps are:

1. Click the Chi-Sq Test for Normality button on the DADM\_Tools ribbon and then select the data range.
2. Select a single numeric variable from the following dialog box:



1. With the above selection, the results are as follows:



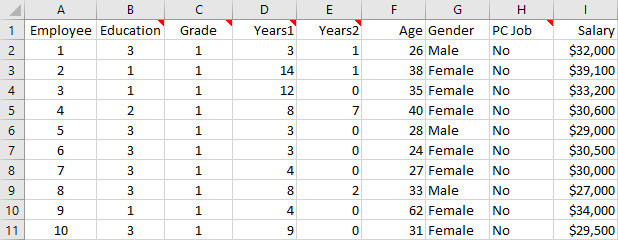
**Notes:**

* As with the Histograms program, the information about bins is hidden in white font behind the chart.
* This is the usual chi-square test for normality, where a small p-value indicates that the data are *not* normally distributed.

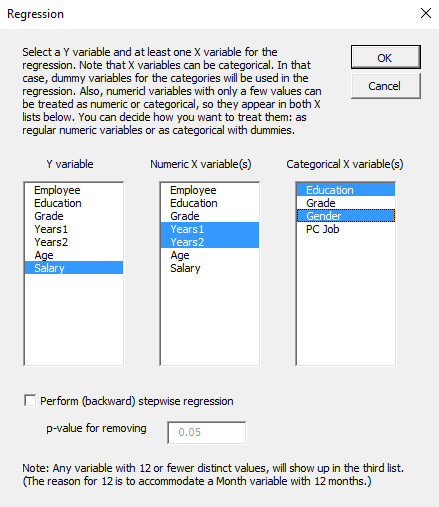
## [Regression Analysis](#_Albright’s_DADM_Tools_Add-In)

This program runs a regression of a selected Y variable versus one or more selected X variables. The Y variable must be numeric and have at least two distinct values (so binary Y variables don’t qualify). The X variables must be numeric, or they can be categorical (numeric or text) with 12 or fewer distinct values. If a variable is numeric with 12 or fewer distinct values, the program can enter it in the regression as a “usual” numeric X variable, or it can create dummies for the distinct values and enter all but one of them in the regression. This is your choice. Similarly, if a variable is text with 12 or fewer variables, the program can create dummies for the distinct values and enter all but one of them in the regression. The program also allows you to run a backward stepwise regression. The steps are:

1. Click the Regression Analysis button on the DADM\_Tools ribbon and then select the data range. For this example, the first 10 rows of the data set are the following:



1. Select the regression variables from the following dialog box. Note, for example, that Education is numeric with five distinct values (1 to 5), so it can be treated as a numeric variable or a categorical variable with dummies, but not both.



1. With these selections, the results are as follows, with many bottom rows not shown. (The output also contains a correlation matrix, not shown here, of all Y and X variables.)

A close up of text on a white background

Description automatically generated

1. If the backward stepwise option is checked, with the same variable selections, the top part of the output is as follows:

A close up of text on a white background

Description automatically generated

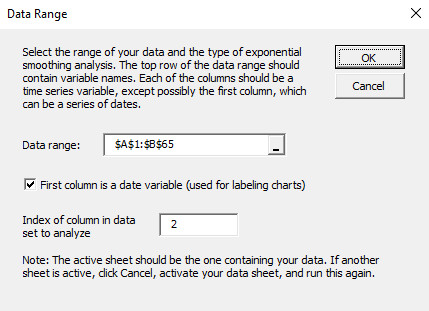
**Notes:**

* As usual, if an X variable is treated as categorical with *n* distinct values, only *n*-1 dummies are entered in the regression. The omitted category, the reference category, is the last in alphabetical order.
* In addition to the usual regression output from standard statistical packages, the VIF and R-square columns are added as multicollinearity checks. The cell comments explain these.
* If there is perfect multicollinearity, where at least one X variable is a perfect linear combination of others, the program issues an error message to this effect and no regression is run.

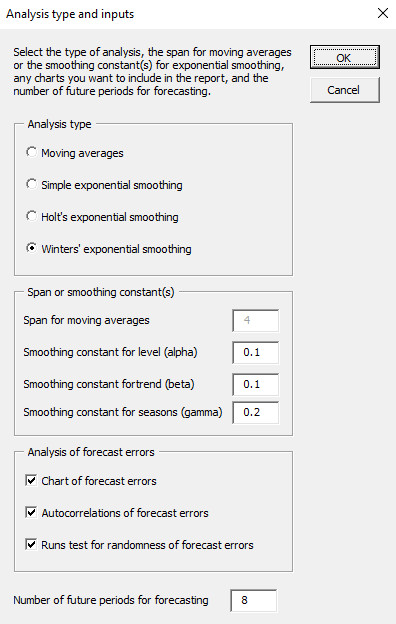
## [Time Series Analysis](#_Albright’s_DADM_Tools_Add-In)

This program runs a time series analysis for a selected times series variable. There are four methods available: moving averages, simple exponential smoothing, Holt’s exponential smoothing method for trend, and Winters’ exponential smoothing method for seasonality. Although you must choose the span (for moving averages) or the smoothing constant(s) (for exponential smoothing), the program then varies these and shows how the corresponding error measures vary. This allows you to see which span or smoothing constant(s) provide the best fit to the historical data. The steps are:

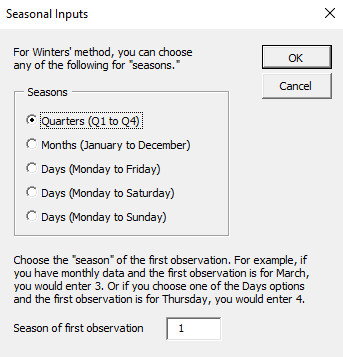
1. Click the Time Series Analysis button on the DADM\_Tools ribbon. Then select a time series variable and, optionally, a date variable for labeling in the following dialog box. The data for this example are quarterly sales with seasonality. The quarter labels are in column A and the sales values are in column B.



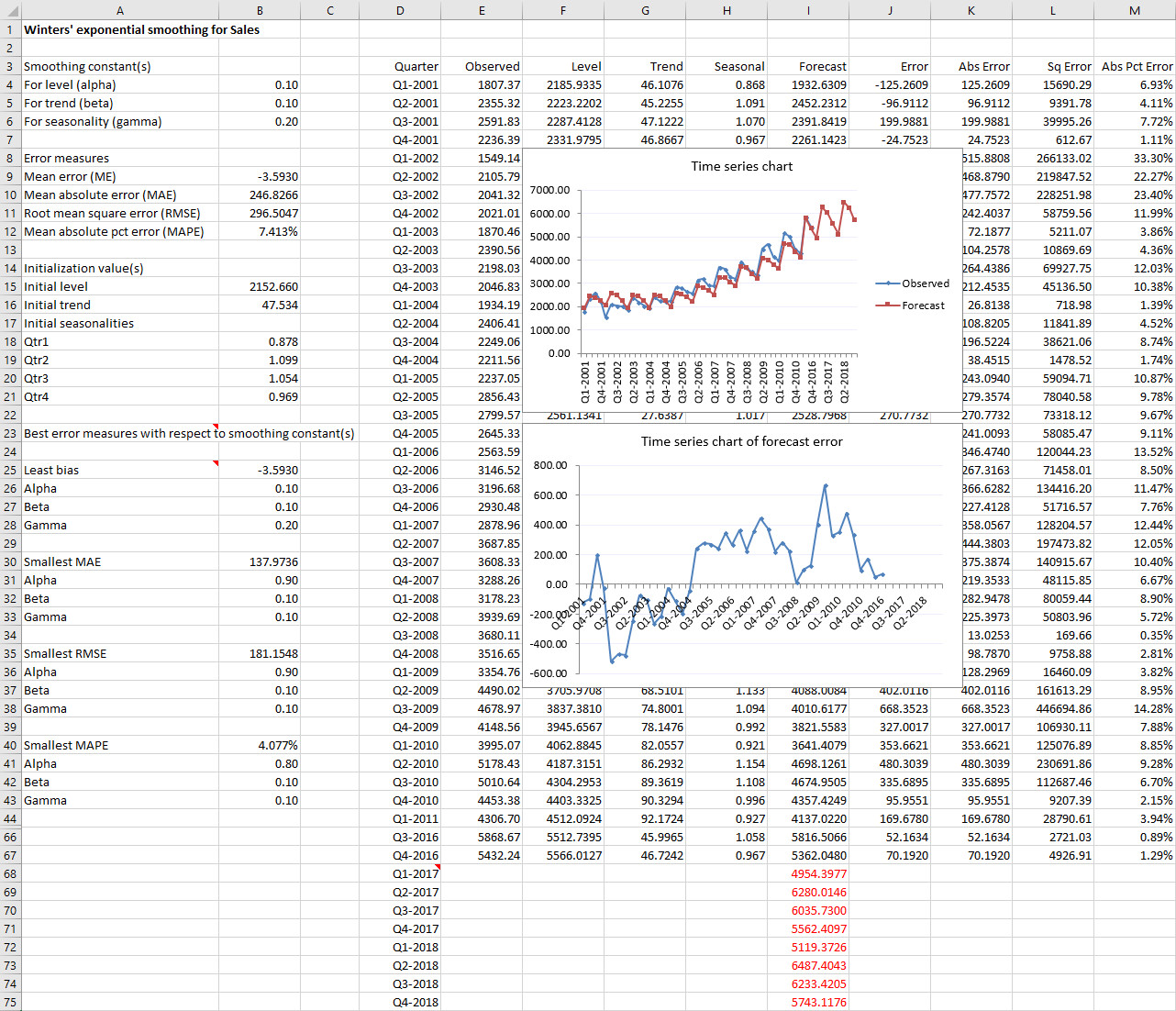
1. Choose the method and parameters in the following dialog box. In this case, Winters’ method is used with reasonable smoothing constants, and 8 quarters of future forecasts are requested.



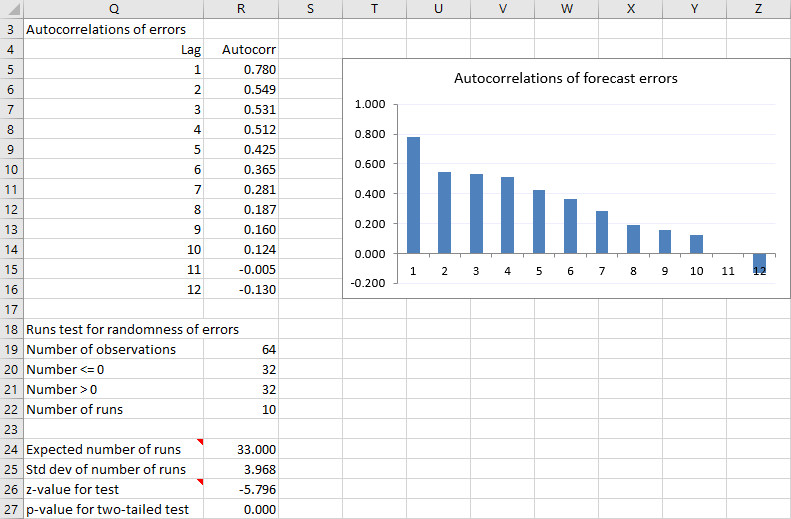
1. If Winters’ method is used, indicate information about seasons in the following dialog box. In this case, the quarterly sales data start in quarter 1.



1. With these selections, the main results are as follows, with a few hidden rows, and future forecasts appear at the bottom in red font. As you can see in the lower sections of columns A and B, it would be better to use a larger value of the Alpha smoothing constant in terms of minimizing MAE, RMSE, or MAPE. In fact, you can then substitute these “suggested” values in cells B4 to B6, and everything will update automatically.



In addition, because optional autocorrelations and an optional runs test were requested, the following results are also returned. There is evidently significant *positive* autocorrelation in the forecast errors.



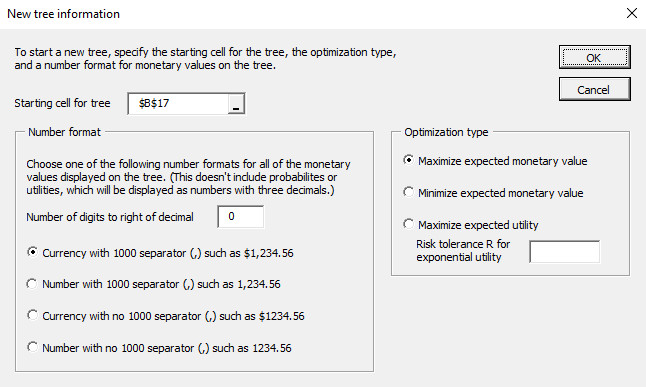
**Notes:**

* All calculations in the body of the results (columns F to M in the above screenshot) are implemented with straightforward copyable Excel formulas.
* The initialization for exponential smoothing methods, particularly Winters’ method, is somewhat complex and you don’t see formulas in the relevant cells in column B. Different statistical packages use different initialization methods, with similar results. The initialization used in DADM\_Tools is the method used in the Stata software package.
* The “best error measures” results from row 23 down in the first screenshot above are found by varying the smoothing constants through a grid of values: from 0.05 to 0.95 in increments of 0.05 for simple and Holt’s methods, and from 0.10 to 0.90 in increments of 0.10 for Winters’ method. For the moving averages method, the span is varied from 1 to 12 in increments of 1.

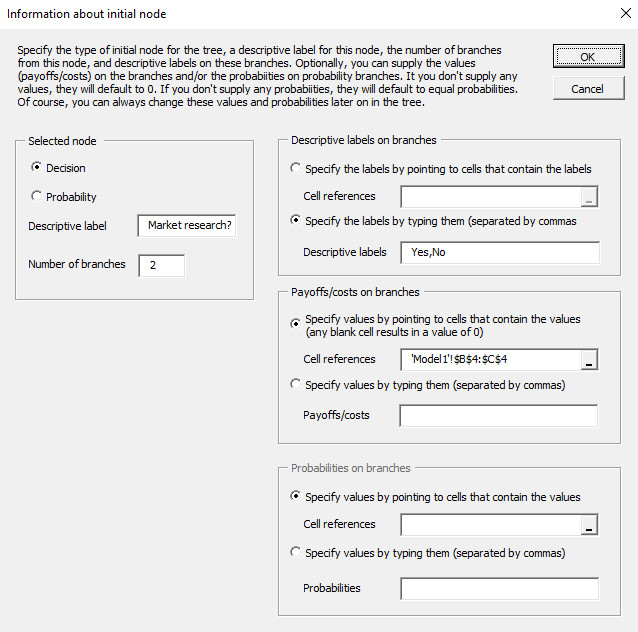
## [Decision Trees](#_Albright’s_DADM_Tools_Add-In)

This program let you build a decision tree from scratch on an Excel worksheet. It offers five basic functions: (1) start a new tree in a selected cell; (2) expand the tree from any existing end node; (3) copy the structure of any subtree emanating from a selected node; (4) paste a copied subtree to any existing end node; and (5) delete any subtree, including the entire tree. The steps are:

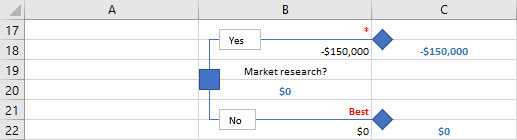
1. Click the Decision Trees button on the DADM\_Tools ribbon.
2. Fill out the following dialog box with basic information about the tree.



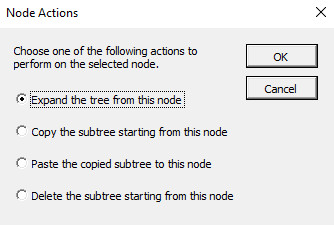
1. Fill out the following dialog box for starting the tree. For this example, the first decision is whether to purchase market research, yes or no. The cost of the research, as a negative number, is stored in cell B4, and the cost of no market research is 0, in cell C4. (Actually, cell C4 is blank, but this evaluates to 0.)



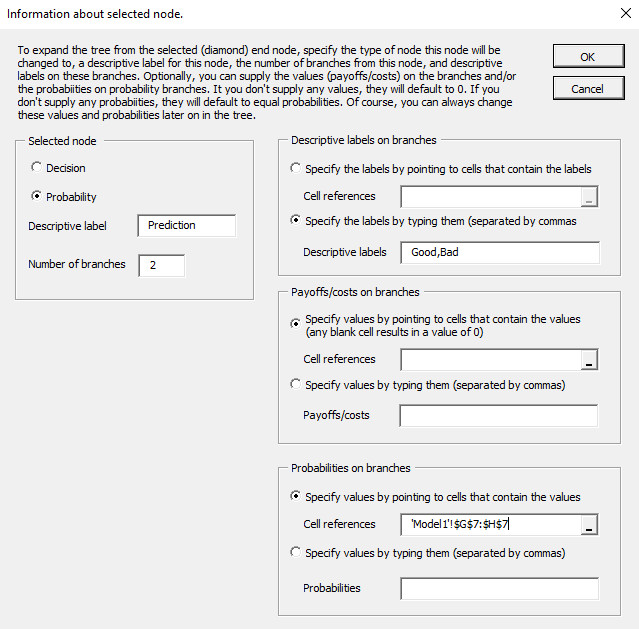
This produces the following tree. Throughout, the best decision branch is labeled “Best” and all other decision branches are labeled with an asterisk. As usual, squares correspond to decision nodes, circles correspond to chance nodes, and diamonds correspond to end nodes. All user input values are in black font, and all values calculated by the program are in blue font.

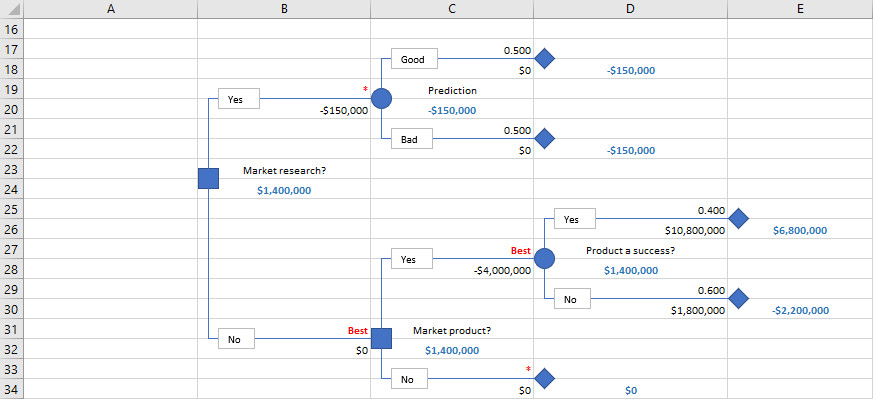


1. From here, click the top end node to get the following dialog box. (You get these same four options any time you click *any* node.)

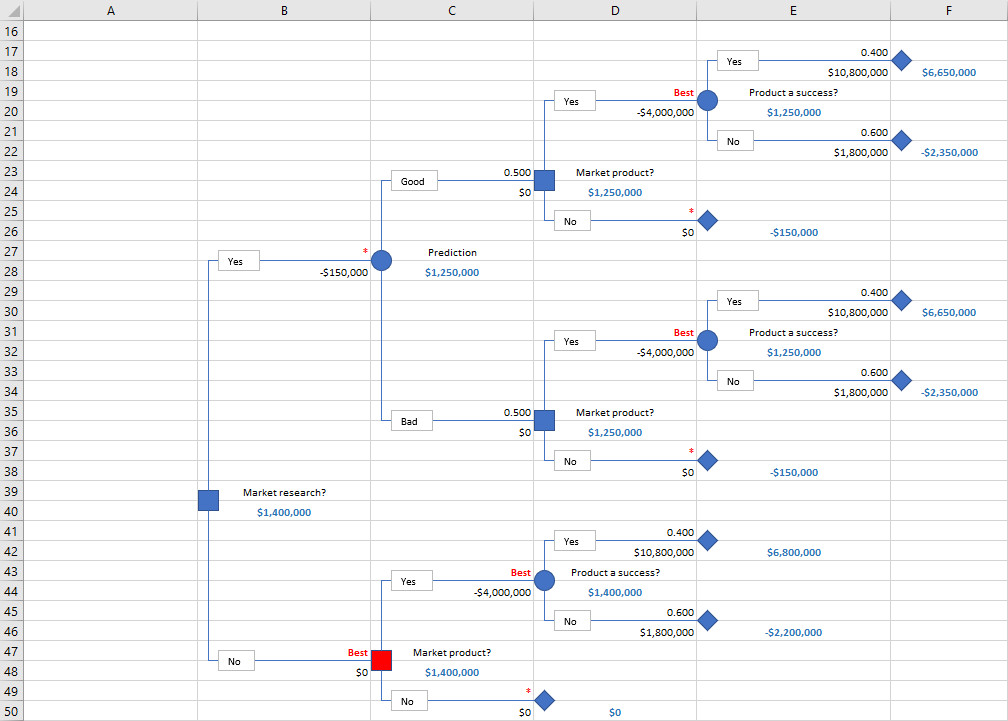


1. If you choose the “expand” option, you get the same dialog box shown earlier. If you fill it out as follows, you get the top part of the expanded tree shown below. You get the bottom part of the tree by further uses of the expand option.





1. Here is a time-saver. Click the bottom decision node between rows 31 and 32 and choose the “copy subtree” option. For this example, the structure of the subtree from this node to the right should be copied to each of the end nodes in the top part of the tree. When you choose the “copy subtree” option, the node you click turns red, meaning that its structure is ready to be copied. Now click each of the end nodes in the top part of the tree and for each, select the “paste subtree” option. The tree then becomes the following. This tree is probably not quite correct because cell references to monetary values and probabilities are copied in the usual Excel way. For example, the cell reference in cell E17 probably needs to be changed. But this tree has the correct structure, and fixing the cell references takes only a little time.



1. Finally, click the red node to return it to its original blue color, indicating that it is no longer ready to be copied.

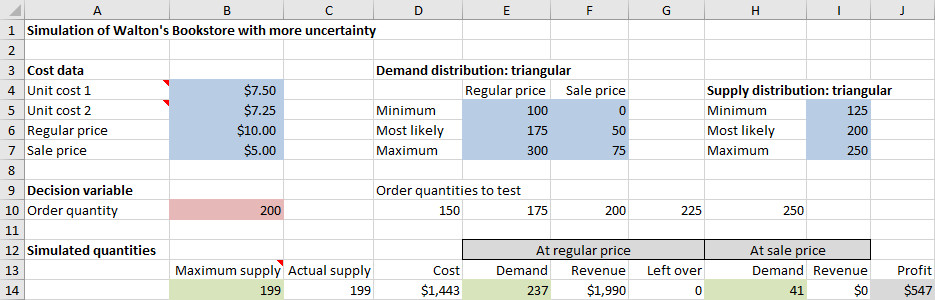
**Notes:**

* The program allows you to create only one decision tree on a given worksheet.
* To change any branch label, click the corresponding text box and type a new label.
* The cells with blue font contain formulas that implement decision tree logic: (1) sums of monetary values next to end nodes; (2) maximums (or minimums for a cost minimization tree) next to decision nodes; and (3) sums of products of monetary values and probabilities next to chance nodes. You can study these formulas to understand how decision trees work, but don’t change them! The only cells you should change are those with black font.
* In the initial dialog box, you can choose the “maximize expected utility” option, along with a risk tolerance value for the exponential utility function. (The higher the risk tolerance, the less risk averse you are.) The same procedure is used, except that all values in blue font are utility (or expected utility) values.
* Only one node can be red (i.e., in copy mode) at a time. If one node is red and you click another node to copy it, the first node reverts back to blue.
* The program creates a hidden worksheet to store the structure of the tree. (This is how it enables copying, pasting, and deleting subtrees.) You can unhide this worksheet and examine it, but you should *not* change it or delete it. This hidden sheet takes its name from the decision tree sheet. For example, if the decision tree sheet is named Model1, the hidden sheet is named Model1DT, that is, “DT” is appended. Therefore, if you decide to rename the decision tree sheet, you should rename the hidden sheet in the same way.

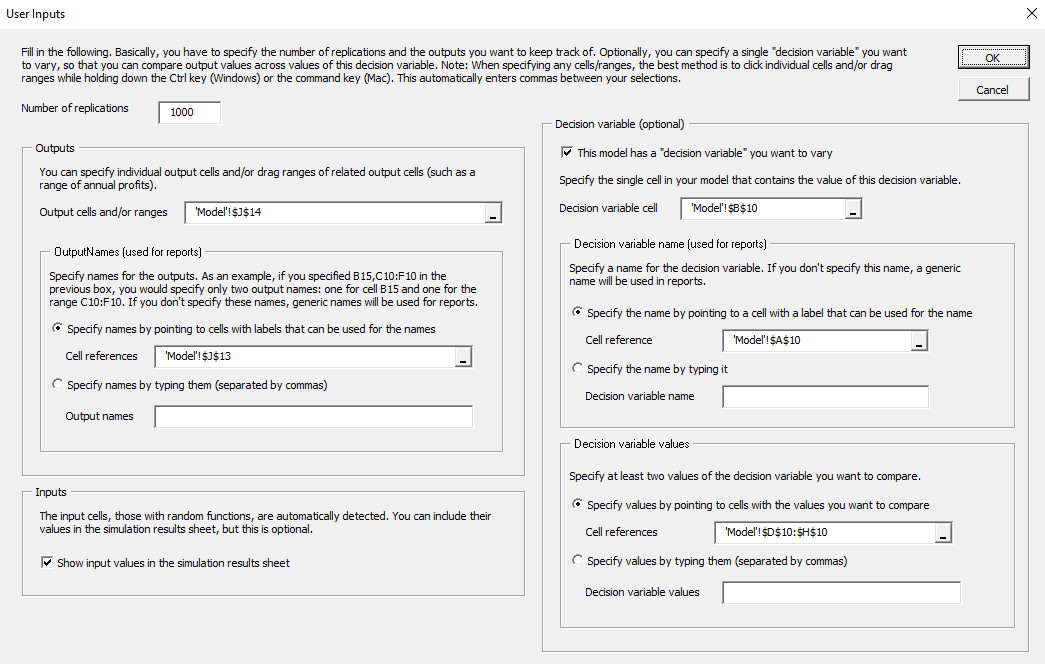
## [Run Simulation](#_Albright’s_DADM_Tools_Add-In)

This program starts with a simulation model, that is, a spreadsheet model with at least one random input cell. It then runs as many replications of this model as you like, each using different random numbers, and it returns the results for output cells you designate. It can also vary a “decision variable” cell over a range you specify and, for each value in this range, run the simulation and report the results for the output cells. In this case, it uses common random numbers (CRN) for a fair comparison. That is, for each random input cell, it uses the *same* random numbers for each value of the decision variable. The steps are:

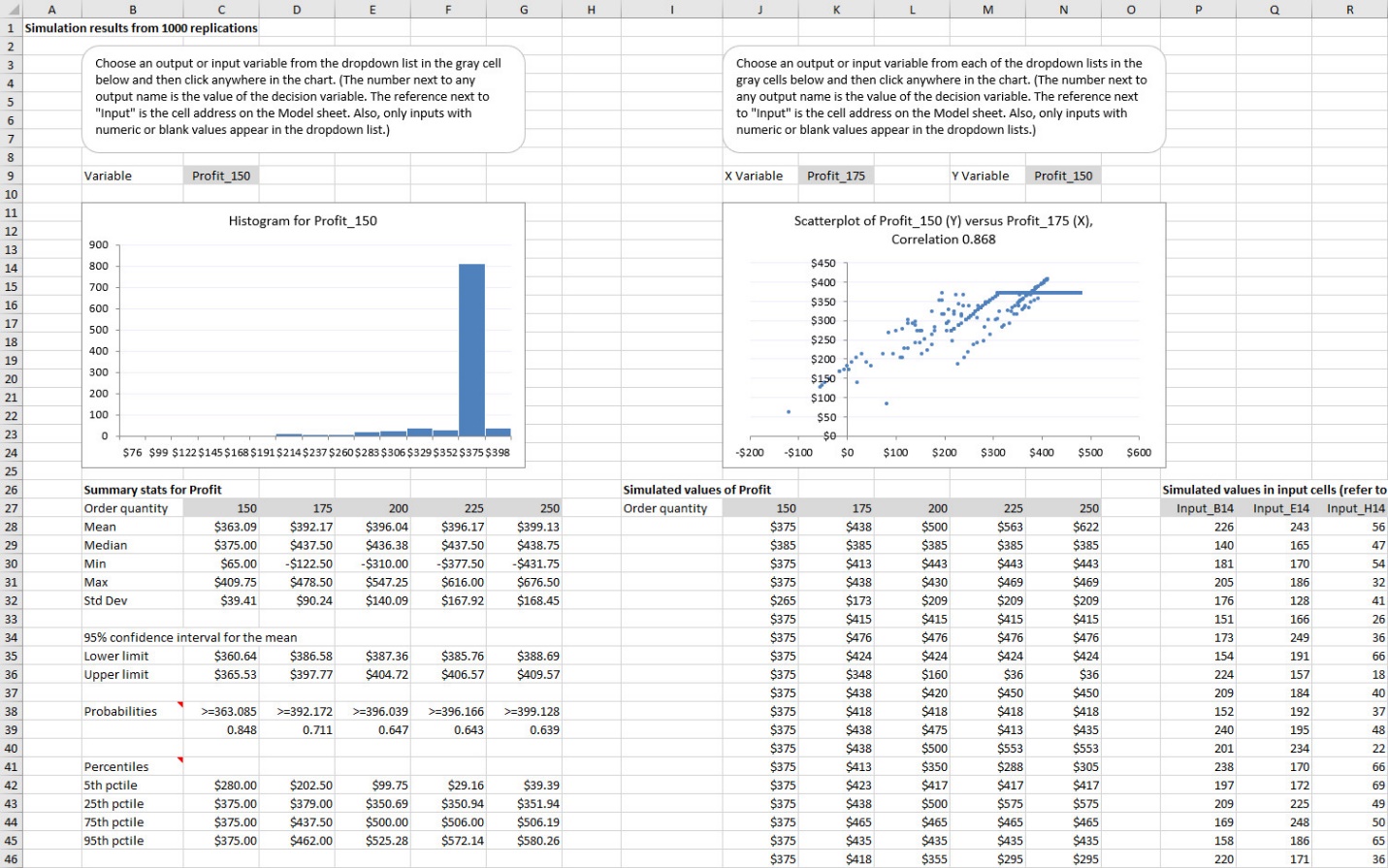
1. Start with a simulation model. The following model is used for this example. The three green cells are the random input cells, each containing a random number from a triangular distribution. (**Note:** Read the next section of this document, Random Functions, about entering random numbers.) Only one cell, the gray profit cell, will be designated as an output cell, but any number of cells could be designated as output cells. This model also contains a decision variable cell, the red order quantity cell, and it will be varied over the values to its right.



1. Click the Run Simulation button on the DADM\_Tools ribbon and fill out the dialog box as shown below. Note that the right side is relevant only if the decision variable option is checked. Also, you should check the bottom left option only if you want to see the simulated values of the random input cells in the results.

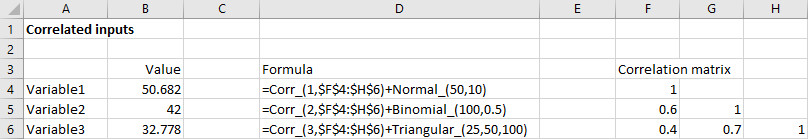


1. The results eventually appear in a new worksheet, as shown below. It contains the following: (1) summary stats for each output, broken down by the values of the decision variable if there is one; (2) values of each output, broken down by the values of the decision variable if there is one; (3) values of the inputs if requested; (4) a histogram of the variable in cell C9; and (5) a scatterplot of the variables in cells K9 and N9. For the histogram, you can change the variable in cell C9 through a dropdown list in this cell and then click the chart to refresh the histogram. The scatterplot works similarly. (See the text box explanations above the charts.) You can also change the conditions in row 38 to get different probabilities. For example, if you change the condition in cell C38 to <100, the probability below it will change automatically. Finally, note that the input values in columns P-R are *not* broken down by the values of the decision variable. This is because of CRN; these *same* random input values are used for each value of the decision variable.



**Notes:**

* The program creates a hidden sheet that stores your settings (those in the above dialog box) from the previous run, if any. It then applies these to the next run so that you don’t have to start from scratch.
* The values in the green cells of this model, the random input cells, use the Triangular\_ function. This is one of 14 custom functions available for your use. Each generates a random number from a well-known distribution. (See more information in the Random Functions section below.)
* There is one other custom function, Corr\_, which can be used to correlate random inputs. The following screenshot illustrates its use. The three variables are to be correlated according to the correlation matrix in columns F-H. The required formulas in column B are spelled out in column D. Each formula appends a Corr\_ function to a usual random function. The Corr\_ function takes two arguments, an index of the variable (1 to 3 in this example) and a reference to the correlation matrix. Note that the distributions being correlated could all be the same distribution, but this is not required. In this example, one is normal, one is binomial, and one is triangular. If you run a simulation with these inputs and request a list of the input values, you can check that they not only have the correct distributions (normal, binomial, and triangular), but that they have approximately the requested correlations.



* Two other things about correlations: First, as shown in the screenshot, you only need to supply the correlations below the diagonal. Second, not all correlation matrices are valid, where “valid” means that they *could* be generated by actual data. For example, if each correlation in the above example were replaced by its negative, the correlation matrix would be invalid (do you see why?), and the program would give you an immediate error message to this effect. You would then have to adjust the correlations to make the matrix valid.
* When there is a decision variable cell, the program uses CRN. This is a good thing because it provides a fair comparison across values of the decision variable. However, it imposes a constraint on the model. Namely, the random input cells should not depend, directly or indirectly, on the decision variable cell. An obvious case where this constraint is violated is if the decision variable cell contains the *mean* of one or more input distributions. In general, dependencies might not be this obvious, but you can check by using Excel’s Trace Dependents tool (in the Formula Auditing on the Formulas ribbon) to find the dependents (and their dependents, etc.) of the decision variable cell. In any case, if you try to run the simulation when there are dependencies, you will get an error message. In this case, you will need to run a separate simulation for each value of the decision variable cell—and you won’t get to take advantage of CRN.

## [Random Functions](#_Albright’s_DADM_Tools_Add-In)

The simulation program just discussed requires random numbers from various distributions in random input cells. The question then is how you generate a random number from a given probability distribution in Excel. For some distributions, you can use built-in Excel functions to do this. For example, the formula **=NORM.INV(RAND(),100,20)** generates a normally distributed random number with mean 100 and standard deviation 20. However, such formulas are hard to remember, and for some distributions, such “simple” formulas don’t even exist. (A good example is the triangular distribution.) Therefore, I created several custom functions, each ending in an underscore (\_), to generate random numbers from the most well-known distributions. These are listed in the following table.

|  |  |  |  |
| --- | --- | --- | --- |
| Distribution | Function and arguments | Restrictions on arguments | Comments |
| Bernoulli | bernoulli\_(p) | 0<=p<=1 | 1 if a success, 0 otherwise, where success has probability p |
| Binomial | bnomial\_(n,p) | n>=0, 0<=p<=1 | Number of successes in n independent trials, where p is the probability of success on each trial |
| Poisson | poisson\_(mean) | mean>0 | Nonnegative integer-valued, often used for the number of events in some amount of time or place |
| Discrete | discrete\_(values,probs) | # of values must match # of probs, and probs must sum to 1 | General discrete distribution where values is any list of possible values and probs is the corresponding list of probabilities |
| Uniform | uniform\_(min,max) | min<=max | Flat distribution, where any value between min and max is equally likely |
| Normal | normal\_(mean,stdev) | stdev>0 | Famous symmetric bell-shaped distribution with given mean and standard deviation |
| Triangular | triangular\_(min,mostlikely,max) | min<=mostlikely<=max | Distribution bounded by min and max, with peak at mostlikely value |
| Pert | pert\_(min,mostlikely,max) | min<=mostlikely<=max | A “rounded” version of the triangular distribution |
| Beta | beta\_(alpha1,alpha2,min,max) | alpha1>0, alpha2>0, min<=max | Bounded by min and max, shape determined by alpha1 and alpha2 |
| Exponential | exponential\_(mean) | mean>0 | Nonnegative “memoryless” distribution with given mean and mode at 0 |
| Erlang | erlang\_(n,beta) | n>0, n integer, beta>0 | Right-skewed nonnegative distribution with integer shape parameter n and mean equal to n\*beta |
| Gamma | gamma\_(alpha,beta) | alpha>0, beta>0 | Generalization of Erlang distribution where the shape parameter alpha can be any nonnegative value. Mean is alpha\*beta. |
| Lognormal | lognormal\_(mean,stdev) | mean>0, stdev>0 | Right-skewed nonnegative distribution with given mean and standard deviation |
| Weibull | weibull\_(alpha,beta) | alpha>0, beta>0 | Right-skewed nonnegative distribution with shape parameter alpha and scale parameter beta. Mean is a complex function of alpha and beta. |

These custom functions are often called user-defined functions, or UDFs. It is easy to include these UDFs in the DADM\_Tools add-in, and I did so at first. With this setup, assuming DADM\_Tools is loaded, you could generate a random input in any cell with a formula like **=normal\_(50,10)**. However, this creates the following problem. If you created simulation files this way and you then shared them with another person who opened them on another computer, the random input cells would contain hard-coded paths to wherever DADM\_Tools is stored on *your* computer, which might not be where it is stored on the other person’scomputer, and this would cause errors. The other person could then do a search and replace to get rid of these hard-coded paths, but that’s more work than should be necessary.

I was able to bypass this hard-coded path problem for Windows users, but unfortunately not for Mac users. First, the DADM\_Tools add-in does *not* include the random function UDFs, so whether you are using Windows or a Mac, if you install DADM\_Tools only, a formula such as **=normal\_(0,1)** wont’ be recognized. The workarounds are then as follows.

**Windows users:** I created the random function UDFs in a file called an XLL. This is a type of dynamic-linked library (DLL) that Excel understands. You don’t need to understand the technical details. All you need to understand is that if you install this XLL on your computer, the UDFs in the above table are available to you at any time, independently of whether the DADM\_Tools add-in is loaded. Installing this XLL on your Windows computer is easy, as explained at <https://kelley.iu.edu/albrightbooks/Free_downloads.htm>, and you should do so before using the simulation program in DADM\_Tools.

**Mac users:** Unfortunately, XLLs don’t work in Excel for Mac, so I instead created another add-in called **Random Functions for the Mac.xlam** that contains the random functions in the above table. To run a simulation on the Mac, you should load this add-in *and* the DADM\_Tools add-in. However, just be aware of two downsides. First, the hard-coded paths to the random functions mentioned above will still be an issue. After all, the XLL was developed to bypass this issue, and XLLs don’t work in Excel for Mac. Second, my experience is that simulations run considerablyslower on the Mac than in Windows. The bottom line is that if you have a Mac and you plan to run a lot of simulations, you should consider installing Windows emulation software (Boot Camp or Parallels) and then go the Windows route.