

EXCEL TUTORIAL

This tutorial will introduce you to some essential features of Excel that we will be using throughout MS&E 121 to solve linear systems of equations, analyzing data, and simulating random variables. You will learn how to

- use some of the functions commonly used in math/statistics,
- represent a linear system of equations in an Excel worksheet,
- use the matrix functions MMULT and MINVERSE to solve linear systems of equations, and
- generate various types of random variables.

For additional Excel assistance, try consulting the Excel Help files or numerous online resources.

Parts of this tutorial borrow from Albert Whangbo's *Excel Solver Tutorial*.

1 Commonly used Excel functions

In many probability applications we will commonly need to analyze data by computing quantities such as mean, variance, standard deviation, correlation, confidence intervals, etc., and Excel provides us with easy to use functions to do such things. All the functions can be accessed through the *insert function* button, *fx*, or directly by typing them on the formula bar.

Naming Vectors and Matrices in Excel

Excel allows you to *name* ranges of cells in the worksheet. This feature is especially convenient for doing matrix calculations and for setting up linear systems of equations. To name a range of cells, select the entire range with the mouse and use the *Insert* → *Name* → *Define...* menu. Alternatively, select the range and enter its name in the Name Box adjacent to the Formula Bar.

General use functions

One of the most useful functions is $SUM(range)$, which sums the numbers in an array of cells. This array can either be a vector (vertical or horizontal) or a matrix. This function can also be used by highlighting first the desired vector of numbers and then pressing the Σ button, in which case the answer is automatically placed at the end of the vector.

Two other useful functions are $SUMPRODUCT(range1, range2)$ and $SUMSQ(range)$. The arguments of SUM , $SUMPRODUCT$, and $SUMSQ$ can be either ranges of cells or range names. For a pair of ranges named S and T , entering $SUMPRODUCT(S,T)$ into the Formula Bar returns

$$\sum_{i,j} s_{ij}t_{ij},$$

i.e., the sum of the products of corresponding elements of S and T. Note that SUMPRODUCT will return an error if the two arrays do not have the same dimensions. Entering SUMSQ(S) returns

$$\sum_{i,j} s_{ij}^2,$$

i.e., the sum of the squared elements of S.

Statistical functions

The function AVERAGE(*range*) computes the arithmetic mean of the range of numbers, i.e.,

$$\bar{x} = \frac{1}{n} \sum_{i=1}^n x_i.$$

The function VAR(*range*) computes the sample variance of the range of numbers, i.e.,

$$S^2 = \frac{1}{n-1} \sum_{i=1}^n (x_i - \bar{x})^2.$$

And the function STDEV(*range*) the sample standard deviation, i.e.,

$$S = \sqrt{S^2}.$$

To compute the correlation between two data sets, use the function CORREL(*range1*, *range2*). If X and Y are two vectors with elements x_i and y_i , respectively, then CORREL(X, Y) returns

$$\frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2} \sqrt{\sum_{i=1}^n (y_i - \bar{y})^2}}.$$

To close this section, it is important to mention that Excel counts with several cumulative distribution functions, CDFs, and inverse CDFs for some of the most commonly used probability distributions. The inverse CDFs are especially useful for computing confidence intervals, quantiles, and simulating random variables through the *inverse transformation method*.

| <i>Prob. Distribution</i> | <i>CDF</i> | <i>inverse CDF</i> |
|---------------------------|--------------|--------------------|
| Beta | BETADIST | BETAINV |
| Binomial* | BINOMDIST | |
| Chi Square | CHIDIST | CHIINV |
| Exponential* | EXPONDIST | |
| F | FDIST | FINV |
| Gamma* | GAMMADIST | GAMMAINV |
| Hypergeometric | HYPGEOMDIST | |
| Lognormal | LOGNORMDIST | LOGINV |
| Negative Binomial | NEGBINOMDIST | |
| Normal* | NORMDIST | NORMINV |
| Standard Normal | NORMSDIST | NORMSINV |
| Poisson* | POISSON | |
| Student's T | TDIST | TINV |
| Weibull* | WEIBULL | |

The distributions marked with * have the added feature that their CDF functions have an option to evaluate both the CDF and the PDF (probability density function). For example,

$$\begin{aligned} \text{EXPONDIST}(3, 2, \text{TRUE}) &= 1 - e^{-2(3)} && \rightarrow \text{the CDF} \\ \text{EXPONDIST}(3, 2, \text{FALSE}) &= 2e^{-2(3)} && \rightarrow \text{the PDF.} \end{aligned}$$

The PDFs are specially useful for simulating random variables through the *Acceptance/Rejection method*.

2 Setting Up Linear Systems of Equations In Excel

In several applications throughout the quarter, we will be dealing with the problem of finding a solution to a linear system of equations. A linear system of equations can be represented mathematically as

$$Ax = b$$

where the matrices A and b comprise the problem *data*. Our goal is to find a solution x to the above equation.

Such systems of equations can have either

- one unique solution,
- infinitely many solutions, or
- no solution.

In the examples that we will be dealing with, we will typically have n equations and n unknowns, that is, the matrix A will have n rows and n columns. If the matrix A is invertible, then the solution to the system of equations is unique and is given by

$$x = A^{-1}b.$$

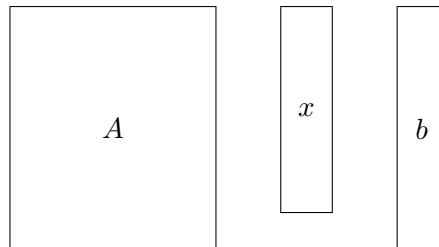
Entering Matrices into Excel

It is easy to display a linear system of equations in their natural way in Excel, i.e., as a range of cells. For example, if A is a 5×5 matrix, then each element of A can be entered into its corresponding cell in a 5×5 range of cells.

| | | | | | | |
|--|----------|----------|----------|----------|----------|--|
| | | | | | | |
| | a_{11} | a_{12} | a_{13} | a_{14} | a_{15} | |
| | a_{21} | a_{22} | a_{23} | a_{24} | a_{25} | |
| | a_{31} | a_{32} | a_{33} | a_{34} | a_{35} | |
| | a_{41} | a_{42} | a_{43} | a_{44} | a_{45} | |
| | a_{51} | a_{52} | a_{53} | a_{54} | a_{55} | |
| | | | | | | |

Each matrix element occupies a cell in the worksheet.

To keep your worksheet organized, you might consider arranging the matrices as shown below.



A sample layout of a linear system of equations in an Excel worksheet.

Matrix Calculations: MMULT and MINVERSE

We will frequently use the MMULT and MINVERSE commands in Excel when solving systems of linear equations.

| Syntax | Description |
|--|--------------------------------|
| MMULT(<i>range1</i> , <i>range2</i>) | ordinary matrix multiplication |
| MINVERSE(<i>range</i>) | matrix inversion |

The arguments of MMULT, and MINVERSE can be either ranges of cells or range names. It is more concise and descriptive to use names. For a pair of ranges named S and T, entering MMULT(S,T) into the Formula Bar returns the matrix whose ij^{th} element is $\sum_k s_{ik}t_{kj}$, i.e., normal matrix multiplication; and entering MINVERSE(S) will return S^{-1} .

MMULT(S,T) will return an error unless the number of columns of S equals the number of rows of T. If you need to take the transpose of a matrix, use the TRANSPOSE function. Similarly, MINVERSE will return an error if the range of cells does not constitute a square matrix, or if the matrix is not invertible. In the first case the error will be #VALUE!, and in the second case #NUM!

Important note: If a matrix function returns a matrix larger a single element, you must first use the mouse to select the cells that will contain the result, then type the matrix function into the Formula Bar, and finally type CTRL SHIFT ENTER (COMMAND ENTER on a Mac) to evaluate the matrix function. For example, if S and T are 3×5 and 5×4 matrices, respectively, then the result of MMULT(S,T) will be a 3×4 matrix. To perform this calculation and display the result in the worksheet, first highlight a 3×4 range of cells, type MMULT(S,T) into the Formula Bar, then type CTRL SHIFT ENTER to display the result in the highlighted area.

An Example

Consider the following system of linear equations:

$$\begin{aligned} 3x_1 + 3x_2 - 4x_3 + x_4 &= 8 \\ 4x_1 + x_2 + 2x_3 + 8x_4 &= 0 \\ x_1 + 2x_3 + 5x_4 &= 6 \\ 3x_1 - x_2 + 4x_3 &= -1 \end{aligned}$$

This problem can easily be represented in matrix notation as

$$Ax = b$$

where data matrices A and b are

$$A = \begin{bmatrix} 3 & 3 & -4 & 1 \\ 4 & 1 & 2 & 8 \\ 1 & 0 & 2 & 5 \\ 3 & -1 & 4 & 0 \end{bmatrix} \quad b = \begin{bmatrix} 8 \\ 0 \\ 6 \\ -1 \end{bmatrix}$$

In an Excel worksheet, identify ranges of cells to accommodate A , b , and x , and name the ranges appropriately. In this example we will assume these ranges have been named A, b, and x, respectively. Select the range where the solution x will appear, and type in the formula bar MMULT(MINVERSE(A),b)¹.

¹Remember that since x is a 4×1 matrix, you will first have to highlight a 4×1 range of cells and press CTRL SHIFT ENTER to evaluate MMULT(MINVERSE(A),b).

The vector

$$x = \begin{bmatrix} -11.4 \\ 39.6 \\ 18.2 \\ -3.8 \end{bmatrix}$$

is the unique solution to the equation $Ax = b$ in this example.

Simulating Random Variables in Excel

Excel comes with a random number generator that can be used through the function `RAND()`. This function generates a *uniform random variable* on the interval $[0, 1]$, which can be used to generate many other random variables.

For example, if we want to generate a uniform random variable on the interval $[a, b]$, we use the formula

$$\text{RAND()} * (b - a) + a.$$

Note: If you want to use `RAND` to generate a random number but you do not want the number to change every time the cell is calculated, you can enter `=RAND()` in the formula bar and then press F9 to change the formula for a fixed random number.

To compute a random number in the set $\{1, 2, \dots, n\}$ according to a discrete uniform distribution, we can generate a random number in $[0, n]$ as we did in the first example and then compute the ceiling of that number (i.e., if the random number is 3.4167, then its ceiling is the integer obtained by rounding-up, in this case, 4):

$$\text{CEILING}(\text{RAND()} * n, 1).$$

If we want to generate a random number having distribution $\text{Normal}(\mu, \sigma^2)$ we can do it by the inverse transformation method. As we will see in class, if $F(x)$ is the CDF of a random variable, $F^{-1}(y)$ is its inverse, and U is a uniform random number in $[0, 1]$, then $F^{-1}(U)$ is a random observation from the distribution F . In Excel we can do this by typing

$$\text{NORMINV}(\text{RAND}(), \mu, \sigma).$$

An Example *Taken from Hillier, Lieberman, Introduction to Operations Research*

Suppose you want to play a game that has the following rules:

1. Each play of the game involves repeatedly flipping an unbiased coin until the *difference* between the number of heads tossed and the number of tails is 3.
2. If you decide to play the game, you are required to pay \$1 for each flip of the coin. You are not allowed to quit during a play of the game.

- You receive \$8 at the end of each play of the game.

Thus, you win money if the number of flips required is fewer than 8, but you lose money if more than 8 flips are required. How would you decide whether to play this game?

We will use simulation to determine what is the probability that you will win money if you play this game. First we need to generate a sequence of random observations to represent the tosses of a coin. We will denote a Head with a "0" and a Tail with a "1", and what we need is to generate a sequence of 0s and 1s where the probability of a 0 is 1/2 and the probability of a 1 is 1/2 (since we are assuming that we will be using a fair coin). To simulate each flip of the coin we can use the formula

$$\text{CEILING}(\text{RAND}() * 2, 1).$$

Now we generate a large number of flips, say 50, and we use the following spreadsheet to record how many flips were necessary to perform before the game ended.

| Flip | Result | Total Heads | Total Tails | Difference | Stop? |
|------|--------|-------------|-------------|------------|-------|
| 1 | 0 | 1 | 0 | 1 | |
| 2 | 1 | 1 | 1 | 0 | |
| 3 | 1 | 1 | 2 | 1 | |
| 4 | 0 | 2 | 2 | 0 | |
| 5 | 0 | 3 | 2 | 1 | |
| 6 | 1 | 3 | 3 | 0 | |
| 7 | 0 | 4 | 3 | 1 | |
| 8 | 1 | 4 | 4 | 0 | |
| 9 | 0 | 5 | 4 | 1 | |
| 10 | 0 | 6 | 4 | 2 | |
| 11 | 0 | 7 | 4 | 3 | Stop |
| 12 | 0 | 8 | 4 | 4 | |
| 13 | 1 | 8 | 5 | 3 | |
| 14 | 0 | 9 | 5 | 4 | |
| ⋮ | ⋮ | ⋮ | ⋮ | ⋮ | |

In this particular simulation, 11 flips were performed before the game ended, in which case you would lose \$3. By repeating this experiment many times and counting the number of simulations in which the game stops before the 8^{th} flip, you can approximate the probability of making money in this game with the formula

$$\frac{\# \text{ of simulations in which the game stops before the } 8^{th} \text{ flip}}{\text{total } \# \text{ of simulations performed}}.$$

Just remember that for any inference made out of a simulation to be reliable, you must usually repeat the experiment many times. In class we will see how many simulations are required in general to obtain a reliable approximation to the quantity we are computing, and we will also learn to provide confidence intervals for such approximation.