Exploiting Excel's Data Table Creatively in the Study of Mathematics

Graham Supiri and Deane Arganbright
gsupiri@dwu.ac.pg & argandeane@yahoo.com
Divine Word University
P.O. Box 483, Madang, Papua New Guinea

Abstract
A spreadsheet program, such as Microsoft Excel, has proven to be an excellent tool for the study, implementation, and visualization of a wide range of mathematical concepts, as well as for mathematical modeling of diverse applications. We focus on the use of Excel's powerful Data Table tool to present new and creative ways to study and teach a wide range of mathematical topics. We draw our topics from algebra, calculus, linear algebra, probability, statistics, and numerical analysis, as well as from business and scientific fields. Our examples have been used successfully in classroom teaching. Each is accompanied by interactive and animated graphics created in Excel. We also present additional ideas for educators and their students to pursue.

Introduction

Ever since the initial spreadsheet program, *Visicalc,* was launched in 1980, and especially following the inclusion of added features and graphic capabilities in subsequent spreadsheets, the educational community has constantly found innovative ways to use spreadsheets for the study and teaching of mathematics [1,3,5]. In this paper we provide some examples for the creative use of the Data Table tool of *Excel* and other spreadsheets for the study and teaching of mathematics, and provide suggestions for further areas of application of this valuable tool.

Example 1. Geometric Growth

To describe the use of the Data Table command, our first example is a standard compound interest model [3]. It is often advantageous to introduce a new topic through the use of a familiar example that we can also examine using other methods, thereby allowing us to both verify and compare the approaches. Our model assumes that we enter a one-time deposit that earns annual compound interest at a given rate for 10 years.

In Figure 1(a) we first enter the principal and annual interest rate in Cells B1:B2. We use Column A to count years, while we find the annual starting balance and interest in Columns B-C. In Figures 1(a) and 1(b) we see the resulting output and the underlying formulas in Columns A:C. Cell B15 contains the resulting 10^{th} year balance.

Now, suppose that we want to create a summary of the 10^{th} year balance for different annual interest rates. Of course, we could simply change cell B2 repeatedly, and manually write down the resulting balance. However, the Data Table command can do this for us automatically. We create the data table in the Block E5:F15. In Column E we leave the top cell E5 blank, and then enter in the rest of Column E the interest rates that we desire to examine. While we have used formulas to increment the rates in steps of 1%, we can enter whatever rates are desired. In Cell F5, we enter a formula, =B15, that reproduces the 10-year balance for the current rate.

To issue the Data Table command, we first use our selection device to select the Block E5:F15, as shown in Figure 2. We then select the *Excel* command options Data, What If Analysis, Data Table. In the resulting Data Table display, we click in the Column Input cell box, and then click on the rate cell, B2. After we then click on OK, *Excel* repeatedly inserts each of the rate values in Column E into Cell B2, and records the corresponding 10^{th} year balances in Column C, as shown in Figure 1(c).

	A	B	C	D	E	F
1	prin	1000				
2	rate	0.05				
3					data table	
4	year	balance	interest		rate	10-year
5	0	1,000.00	50.00			1,628.89
6	1	1,050.00	52.50		1%	
7	2	1,102.50	55.13		2%	
8	3	1,157.63	57.88		3%	
9	4	1,215.51	60.78		4%	
10	5	1,276.28	63.81		5%	
11	6	1,340.10	67.00		6%	
12	7	1,407.10	70.36		7%	
13	8	1,477.46	73.87		8%	
14	9	1,551.33	77.57		9%	
15	10	1,628.89			10%	

Figure 1(a). Savings Layout

	A	B	C	D	E	F
1	prin	1000				
2	rate	0.05				
3					data table	
4	year	balance	interest		rate	10-year
5	0	=B1	=$B$2*B5			=B15
6	=1+A5	=B5+C5	=$B$2*B6		0.01	
7	=1+A6	=B6+C6	=$B$2*B7		=0.01+E6	
8	=1+A7	=B7+C7	=$B$2*B8		=0.01+E7	
9	=1+A8	=B8+C8	=$B$2*B9		=0.01+E8	
10	=1+A9	=B9+C9	=$B$2*B10		=0.01+E9	
11	=1+A10	=B10+C10	=$B$2*B11		=0.01+E10	
12	=1+A11	=B11+C11	=$B$2*B12		=0.01+E11	
13	=1+A12	=B12+C12	=$B$2*B13		=0.01+E12	
14	=1+A13	=B13+C13	=$B$2*B14		=0.01+E13	
15	=1+A14	=B14+C14			=0.01+E14	

Figure 1(b). Savings Formulas

	E	F
1	rounding used	
2		
3	data table	
4	rate	10-year
5		1,628.86
6	1%	1,104.59
7	2%	1,218.96
8	3%	1,343.88
9	4%	1,480.21
10	5%	1,628.86
11	6%	1,790.80
12	7%	1,967.10
13	8%	2,158.87
14	9%	2,367.31
15	10%	2,593.72

Figure 1(c). Data Table

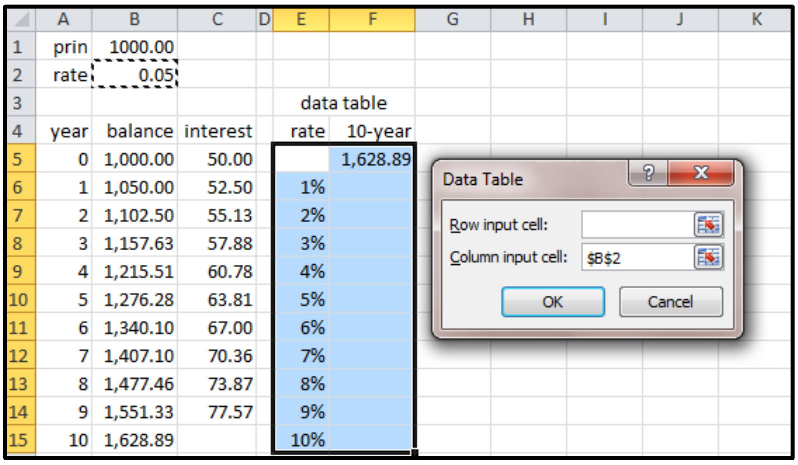


Figure 2. Data Table Display

If this is the first time for someone to use the Data Table, it is a good exercise to create an additional column of the 10th year values by using the closed formula $P(1+r)^{10}$ or one of *Excel*'s built-in financial functions. We shall then see that the result agrees with our Data Table model.

Example 2. Cramer's Rule

Cramer's Rule [7] is a well-known procedure that uses determinants to solve a linear system of n equations in n unknowns. Unfortunately, doing this procedure by hand is both tedious and very inefficient for $n>3$. However, using *Excel's* determinant and matrix functions, together with the Data Table tool, makes the process quite accessible and efficient, even for rather large values of n.

For an $n \times n$ system,

$$
\begin{array}{ccccccc}
a_{11}x_1 & + a_{12}x_2 & & + a_{1k}x_k & & + a_{1n}x_n & = c_1 \\
a_{21}x_1 & + a_{22}x_2 & & + a_{2k}x_k & & + a_{2n}x_n & = c_2 \\
 & & & & & & \\
a_{i1}x_1 & + a_{i2}x_2 & & + a_{ik}x_k & & + a_{in}x_n & = c_i \\
 & & & & & & \\
a_{n1}x_1 & + a_{n2}x_2 & & + a_{nk}x_k & & + a_{nn}x_n & = c_n
\end{array}
$$

Cramer's Rule states that the value, x_k, of the k-th component of the solution, can be found by dividing the determinant that results by replacing the entries of the k^{th} column of the $n \times n$ matrix of coefficients with the column of constants by the determinant of the coefficients, or

$$x_k = \begin{vmatrix} a_{11} & a_{12} & c_1 & a_{1n} \\ a_{21} & a_{22} & c_2 & a_{2n} \\ \\ a_{i1} & a_{i2} & c_i & a_{in} \\ \\ a_{n1} & a_{n2} & c_n & a_{nn} \end{vmatrix} \Bigg/ \begin{vmatrix} a_{11} & a_{12} & a_{1k} & a_{1n} \\ a_{21} & a_{22} & a_{2k} & a_{2n} \\ \\ a_{i1} & a_{i2} & a_{ik} & a_{in} \\ \\ a_{n1} & a_{n2} & a_{nk} & a_{nn} \end{vmatrix}$$

In Figure 3 we illustrate our procedure for the 5x5 case. We enter the system in the block A4:F8 with Row 3 used as a counter for k. We then enter a value for k in Cell D1. In the block A10:F14 the entries in Column k of A are replaced by the column of constants to produce the matrix A_k. To do this, in Cell A10 we enter =IF(A$3=$D$1,$F4,A4), and then copy it throughout the Block A10:E14. We compute the determinant of A in Cell B1 by =MDETERM(A4:E8) and the determinant of A_k in Cell F1 by =MDETERM(A10:E14).

We next create the Data Table in the Block I3:J8, entering the formula =F1/B1 for $|A_k|/|A|$ in Cell J3, and values for k = 1, 2, 3, 4, 5 down Column I. We then issue the Data Table command, selecting the Block I3:J8 and Column Input Cell, D1. Figure 3(b) shows the resulting display. Modifying the design for larger systems is straight-forward. We have used this successfully for systems of as many as 20 variables.

	A	B	C	D	E	F	G	H	I	J
1	|A|	160	k	2	|A$_k$|	-320			data table	
2	coefficient matrix A								k	x$_k$
3	1	2	3	4	5	con				-2
4	3	2	1	8	-3	13		x_1	1	
5	5	-2	1	1	-3	22		x_2	2	
6	2	0	2	-3	-1	9		x_3	3	
7	1	1	-1	2	3	7		x_4	4	
8	-1	3	1	0	3	-1		x_5	5	
9	matrix A$_k$									
10	3	13	1	8	-3					
11	5	22	1	1	-3					
12	2	9	2	-3	-1					
13	1	7	-1	2	3					
14	-1	-1	1	0	3					

Figure 3(a). Cramer's Rule Model

	H	I	J
1		data table	
2		k	x$_k$
3			-2
4	x_1	1	4
5	x_2	2	-2
6	x_3	3	3
7	x_4	4	1
8	x_5	5	2
9			
10			
11			
12			
13			
14			

Figure 3(b). Cramer Data Table

Example 3. Numerical Integration via Trapezoidal Rule

Here we present the use of the Data Table tool to significantly condense lengthy series of calculations. In calculus we examine a variety of ways to approximate a definite integral as the area beneath a curve. Using the trapezoidal rule [2,6], we approximate the area under equally-spaced segments of a continuous curve by trapezoids, as shown in Figure 4(a), using the formula $A \approx (b-a)\left(f(a)+f(b)\right)/2$.

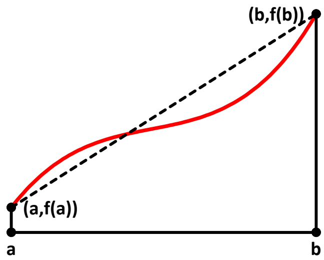


Figure 4(a). Trap: $n = 1$

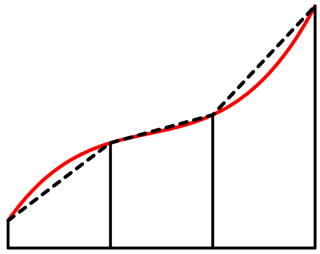


Figure 4(b). Trap: $n = 3$

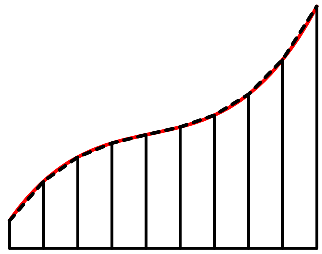


Figure 4(c). Trap: $n = 9$

We use the trapezoidal rule to obtain an approximation for an integral by dividing the interval $[a,b]$ into n equal subdivisions. As n increases we will get increasingly better approximations until round-off error intrudes. However, it can be inconvenient or impossible to increase the number of rows in a spreadsheet model beyond a certain number of divisions. We can overcome much of this difficulty by using a Data Table. For our illustration, we use the function: $f(x) = 1/x^2$, $1 \le x \le 3$. $f(x) = 1/x^2, 1 \le x \le 3$.

In Column B of Figure 5(a) we enter the number, n, of major subdivisions (here $n = 10$) and the values of a and b. We also compute the interval width $b-a$, the subdivision width or gap as $gap = (b-a)/n$. We divide each major subdivision into 1000 small divisions, using a step size of $step = gap/1000$. Next, in Columns D:I we compute the trapezoidal areas for the small segments of the interval starting with the entry in Cell E3. We leave details left to readers. The basic Trapezoidal Rule formulas are given in Figure 6.

We now create a 10-step Data Table in Columns L:M. Column L consists of the major interval starting points. In Cell M3 we compute the sum of Column I as =SUM(I3:I1002). We then issue the Data Table Command, using E3 as the Column Input cell. After this the overall resulting area approximation is found in Cell M15 as =SUM(M4:M13). We then reproduce this in Column A where we compare it with exact area found by integration. Figure 5(b) shows the initial Data Table.

	A	B	C	D	E	F	G	H	I	J	K	L	M
1				Trapezoidal Rule							data table		
2	input/output			k	x_l	x_r	$f(x_l)$	$f(x_r)$	area		n	x	area
3	num	10		1	1.0	1.0002	1	0.9996	0.000200				0.166667
4	a	1		2	1.0002	1.0004	0.9996	0.9992	0.000200		1	1.0	0.166667
5	b	3		3	1.0004	1.0006	0.9992	0.9988	0.000200		2	1.2	0.119048
6	intv	2		4	1.0006	1.0008	0.9988	0.9984	0.000200		3	1.4	0.089286
7	gap	0.2		5	1.0008	1.001	0.9984	0.998	0.000200		4	1.6	0.069444
8	step	0.0002		6	1.001	1.0012	0.998	0.9976	0.000200		5	1.8	0.055556
9				7	1.0012	1.0014	0.9976	0.9972	0.000199		6	2.0	0.045455
10	area	0.666667		8	1.0014	1.0016	0.9972	0.9968	0.000199		7	2.2	0.037879
11	exact	0.666667		9	1.0016	1.0018	0.9968	0.9964	0.000199		8	2.4	0.032051
12	error	6.42E-09		10	1.0018	1.002	0.9964	0.996	0.000199		9	2.6	0.027473
13				11	1.002	1.0022	0.996	0.9956	0.000199		10	2.8	0.023810
14				12	1.0022	1.0024	0.9956	0.9952	0.000199				
15				13	1.0024	1.0026	0.9952	0.9948	0.000199			Area	0.666667
16				14	1.0026	1.0028	0.9948	0.9944	0.000199				
1002				1000	1.1998	1.2	0.6947	0.6944	0.000139				

Figure 5(a). Trapezoidal Model

	K	L	M
1	data table		
2	n	x	area
3			0.166667
4	1	1.0	
5	2	1.2	
6	3	1.4	
7	4	1.6	
8	5	1.8	
9	6	2.0	
10	7	2.2	
11	8	2.4	
12	9	2.6	
13	10	2.8	
14			
15		Area	0.000000
16			

Figure 5(b). Data Table

	D	E	F	G	H	I
2	k	x_l	x_r	$f(x_l)$	$f(x_r)$	area
3	1	1	=E3+B$8	=1/E3^2	=1/F3^2	=$B$8*(G3+H3)/2
4	=1+D3	=E3+$B$8	=E4+B$8	=1/E4^2	=1/F4^2	=$B$8*(G4+H4)/2

Figure 6. Trapezoidal Formulas

There are various other numerical integration algorithms (*e.g.*, Simpson's method) that we can implement similarly.

Example 4. Eigenvalues and Characteristic Polynomial

A major topic in linear algebra is that of eigenvalues and eigenvectors of matrices [7]. A real number, λ, is an eigenvalue of an $n \times n$ real matrix A if there is a non-zero vector v for which $Av = \lambda v$. Any vector v that satisfies this equation is called an eigenvector corresponding to λ. To find the real eigenvalues of a square matrix A, we note that if $Av = \lambda v = (\lambda I)v$, and $(A - \lambda I)v = 0$. This happens when the determinant $|A - \lambda I| = 0$. We use this fact to find the real eigenvalues of a $n \times n$ matrix A.

We illustrate our process in Figure 7 with a 4×4 matrix, and employ a data table to evaluate points of the characteristic polynomial $f(x) = |A - xI|$. The zeroes, λ, of this function then are the eigenvalues of A. We enter the values of A in the Block B2:E5, and formulas to compute the entries of $A - \lambda I$ in the block B7:E10. We compute $|A - \lambda I|$ in Cell E1.

We next use the Data Table command to compute the (x,y) coordinates of $f(x)$, with steps in the x-values of size 0.1, over the interval $-4 \le x \le 4$. Cell B2 is the column input cell for the data table.

	A	B	C	D	E	F	G	H
1	λ	2		|A–λI| =	-38		x	f(x)
2	A	1	5	1	1			-38
3		1	-2	1	0		-4.0	
4		1	1	0	1		-3.9	
5		1	1	1	1		-3.8	
6							-3.7	
7	A–λI	-1	5	1	1		-3.6	
8		1	-4	1	0		-3.5	
9		1	1	-2	1		-3.4	
10		1	1	1	-1		-3.3	
11							-3.2	

Figure 7(a). Char. Polynomial (layout)

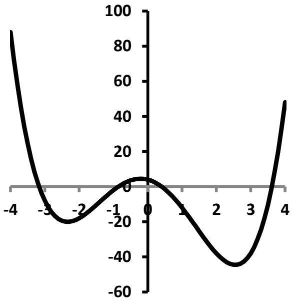
	A	B	C	D	E	F	G	H
1	λ	2		|A–λI| =	=MDETERM(B7:E10)		x	f(x)
2	A	1	5	1	1			=E1
3		1	-2	1	0		-4	
4		1	1	0	1		=0.1+G3	
5		1	1	1	1		=0.1+G4	
6							=0.1+G5	
7	A–λI	=B2-B1	=C2	=D2	=E2		=0.1+G6	
8		=B3	=C3-B1	=D3	=E3		=0.1+G7	
9		=B4	=C4	=D4-B1	=E4		=0.1+G8	
10		=B5	=C5	=D5	=E5-B1		=0.1+G9	
11							=0.1+G10	

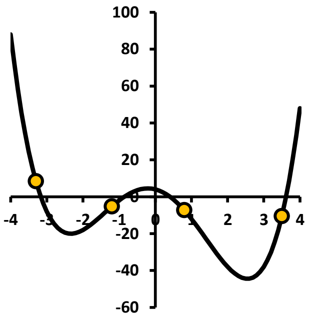
Figure 7(b). Char. Polynomial (Formulas)

In Figure 8 we see the initial values of the resulting output. From this we can use Columns G:H to generate the *xy*-graph, shown at the left in Figure 9(a).

	A	B	C	D	E	F	G	H
1	λ	2		|A–λI| =	-38		x	f(x)
2	A	1	5	1	1			-38
3		1	-2	1	0		-4.0	88
4		1	1	0	1		-3.9	72.32
5		1	1	1	1		-3.8	58.23
6							-3.7	45.64
7	A–λI	-1	5	1	1		-3.6	34.44
8		1	-4	1	0		-3.5	24.56
9		1	1	-2	1		-3.4	15.91
10		1	1	1	-1		-3.3	8.412
11							-3.2	1.978

Figure 8. Characteristic Polynomial from Data Table





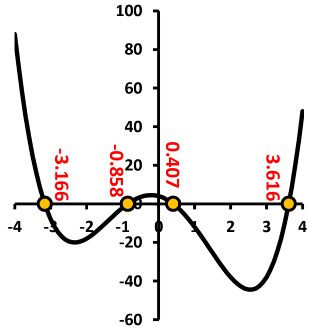


Figure 9(a) Char. Polynomial Figure 9(b). Estimate of Zeroes Figure 9(c). Zeroes

From the graph of Figure 9(a), we can estimate manually the values of the eigenvalues as the points where the curve crosses the *x*-axis as in Figure 10(a). We then use the Data Table command again to get the *y*-values corresponding to these points, as shown in Figure 10(b) where we have used rough estimates of the zeroes of the function. We incorporate these into our graph as markers only in Figure 9(b). We could use estimates of *x*-values other than those in our given list.

	I	J	K
1	x	y	y^2
2		-38	
3	-3.3		
4	-1.2		
5	0.8		
6	3.5		
7			
8		sse	

Figure 10(a).Estimates

	I	J	K
1	x	y	y^2
2		-38	
3	-3.3	8.412	70.76
4	-1.2	-5.21	27.11
5	0.8	-7.27	52.86
6	3.5	-10.4	108.9
7			
8		sse	259.7

Figure 10(b). Data Table

	I	J	K
1	x	y	y^2
2		-38	
3	-3.166	1E-03	1E-06
4	-0.858	0.002	6E-06
5	0.407	-0	2E-06
6	3.616	-0	1E-07
7			
8		sse	9E-06

Figure 10(c). Solver

We next use *Excel*'s Solver command to find a better fit, as illustrated in Figure 10(c). To do this we compute the sum of the squares of the *y*-values that result from our estimates for *x*. In the Solver we set the goal of making the sum of the squares (Cell K8) to 0 by varying the *x*-values (I3:I6). We set the Solver type to GRG nonlinear. To ensure that no two of the values converge to the same zero, we include constraints such as, I3 ≤ -3 and I3 ≥ -3.5 in the Solver's Constraints section. We then press Solve to generate the estimates in Figures 9(c) and 10(c). As another valuable project, we also can implement traditional algorithms from numerical analysis [2] for computing eigenvalues and compare them with our results.

Example 5. Calculus: Newton's Method

Newton's Method [2,6] provides us with a means for estimating the zeroes of a differentiable function. Thus, if $y = f(x)$ is such a function, then we start with a reasonable estimate, x_0, of a zero. Next, as illustrated in Figure 11(a), we find where the tangent line to the curve at the point $(x_0, f(x_0))$ intersects the *x*-axis, $x_1 = x_0 - f(x_0)/f'(x_0)$. This will generally be a better approximation. We then repeat this process, which usually converges quickly to the desired zero. However, sometimes this method may fail to converge. Also, in some cases, even a small changes in the choice of x_0 can produce very different points of convergence, as shown in Figures 11(b) and 11(c), using $f(x) = \cos x$. We investigate this phenomenon in an *Excel* model using a data table.

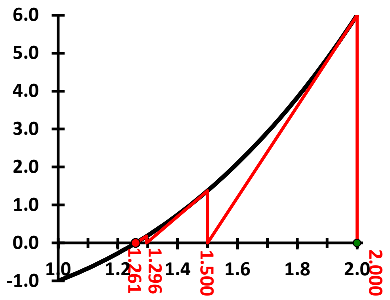


Figure 11(a). Newton

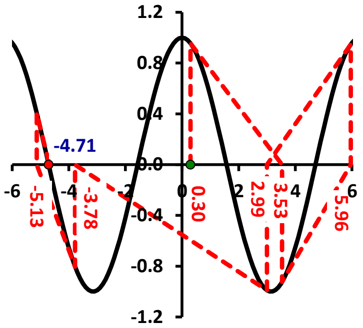


Figure 11(b). Newton (x_0=0.3)

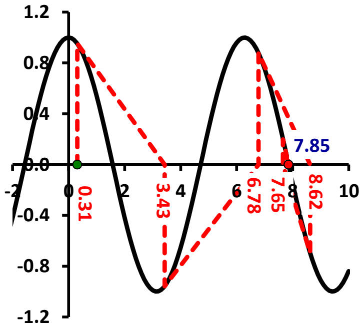


Figure 11(c). Newton (x_0=0.31)

Figure 12 presents a model for Newton's Method in Columns A:D with $f(x) = \cos x$ and $f'(x) = -\sin x$. We enter an initial estimate x_0 in Cell B1. This becomes the first value for x in Cell B4. We find $f(x)$ and $f'(x)$ in Columns C:D and the next value of x in Column B of the following row. Our implementation extends through an extensive number of rows, by which time Newton's Method usually will have converged, say in Cell B24, to the resulting zero. Then in Columns F:G we create a data table for various initial x_0-values, using the resulting approximation from Cell B24 as the returned value.

We can see that if our initial estimate is $x_0 = 0.30$ then the algorithm will converge to $x \approx -4.71$ (*i.e.*, -3π/2), while if $x_0 = 0.31$ then it converges to $x \approx 7.85$ (*i.e.*, 5π/2). We can study this phenomenon further by using a data table in Columns F:G. We generate a range of the initial values, x_0, down Column F with the resulting point of convergence x_1 in Column G. We produce the source of the x_1 values for the Data Table tool in Cell G4 as =B24.

	A	B	C	D	E	F	G
1	x_0	0.31				data table	
2						points	
3	n	x	f(x)	f'(x)		x_0	x_1
4	0	=B1	=COS(B4)	=-SIN(B4)			=B24
5	=1+A4	=B4-C4/D4	=COS(B5)	=-SIN(B5)		-7	
6	=1+A5	=B5-C5/D5	=COS(B6)	=-SIN(B6)		=0.01+F5	
7	=1+A6	=B6-C6/D6	=COS(B7)	=-SIN(B7)		=0.01+F6	
8	=1+A7	=B7-C7/D7	=COS(B8)	=-SIN(B8)		=0.01+F7	
9	=1+A8	=B8-C8/D8	=COS(B9)	=-SIN(B9)		=0.01+F8	
10	=1+A9	=B9-C9/D9	=COS(B10)	=-SIN(B10)		=0.01+F9	

Figure 12(a). Newton's Method (Formulas)

	A	B	C	D	E	F	G
1	x_0	0.31				data table	
2						points	
3	n	x	f(x)	f'(x)		x_0	x_1
4	0	0.31	0.9523	-0.31			7.854
5	1	3.432	-0.9582	0.286		-7	-7.85
6	2	6.78	0.8790	-0.48		-6.99	-7.85
7	3	8.623	-0.6958	-0.72		-6.98	-7.85
8	4	7.655	0.1979	-0.98		-6.97	-7.85
9	5	7.857	-0.0027	-1		-6.96	-7.85
10	6	7.854	0.0000	-1		-6.95	-7.85

Figure 12(b). Newton's Method (Data Table)

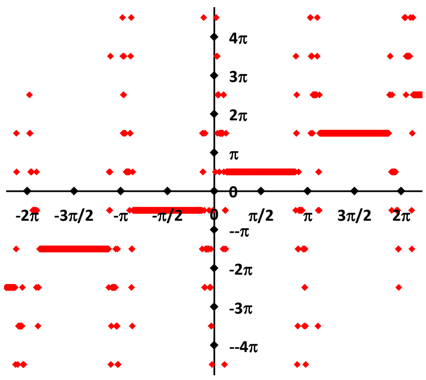


Figure 13. Points of Convergence for Newton's Method

To complete our model, we create an *xy*-graph from Columns F:G plotting markers to get the picture of Figure 13 showing how the point of convergence varies for a range of initial estimates x_0.

Example 6 – Random Model of Coin Flip

We can use *Excel's* RANDOM function to simulate events and concepts that are included in the study of probability and statistics [8]. When combined with the Data Table tool, we are able to produce summaries of sets of repeated random trials. Here we create a simulation of 1000 sets of 100 flips of a fair coin (*i.e.*, with probability $p = 0.5$ of obtaining a head).

In the model of Figure 14 we count the flips in Column A and generate random numbers between 0 and 1 in Column B. We set the probability of success p (here $p = 0.5$) in Cell C2. Next, in Cell C5 we use the =IF function to generate a head, "H" if the current random number is less than p, and "T" otherwise. We then copy this expression down Column C. Next, in Cell C1 we determine the number of heads using the =COUNTIF function. Finally, we create a data table in the Block E4:F1004 to find the number of heads in each of 1000 repetitions. In Cell F1 of the data table we use the =AVERAGE function to compute the mean number of heads in the 1000 flips.

It is also instructive to look at the distribution of the number of heads, n, obtained in the 1000 flips. We create this summary in Columns H:I using the =COUNTIF function. Here we have computed this for $30 \leq n \leq 70$ in order to produce the graph of Figure 16 that is broad enough to include all outcomes. We can simulate other binomial events by changing the value of p in Cell C2. In this case we may need to modify the range of our graph.

	A	B	C	D	E	F	G	H	I
1		heads	53		mean	#####		mean	0
2		p	0.5					sum	0
3	n	rand	coin		rep	head		head	num
4						53			
5	1	0.716	T		1			30	0
6	2	0.261	H		2			31	0
7	3	0.411	H		3			32	0
8	4	0.366	H		4			33	0

Figure 14(a). Coin Flip (layout)

	A	B	C	D	E	F	G	H	I
1		heads	53		mean	49.89		mean	49.89
2		p	0.5					sum	1000
3	n	rand	coin		rep	head		head	num
4						53			
5	1	0.716	T		1	46		30	0
6	2	0.261	H		2	59		31	0
7	3	0.411	H		3	46		32	0
8	4	0.366	H		4	44		33	0

Figure 14(b). Coin Flip (Data Table)

	A	B	C	D	E	F	G	H	I
1		heads	=COUNTIF(C5:C104,"H")		mean	=AVERAGE(F5:F1004)		mean	=SUMPRODUCT(H5:H45,I5:I45)/1000
2		p	0.5					sum	=SUM(I5:I45)
3	n	rand	coin		rep	head		head	num
4						=C1			
5	1	=RAND()	=IF(B5<C$2,"H","T")		1	=TABLE(,A2)		30	=COUNTIF($F$5:$F$1004,H5)
6	=1+A5	=RAND()	=IF(B6<C$2,"H","T")		=1+E5	=TABLE(,A2)		=1+H5	=COUNTIF($F$5:$F$1004,H6)
7	=1+A6	=RAND()	=IF(B7<C$2,"H","T")		=1+E6	=TABLE(,A2)		=1+H6	=COUNTIF($F$5:$F$1004,H7)
8	=1+A7	=RAND()	=IF(B8<C$2,"H","T")		=1+E7	=TABLE(,A2)		=1+H7	=COUNTIF($F$5:$F$1004,H8)

Figure 15. Coin Flip Simulation (formulas)

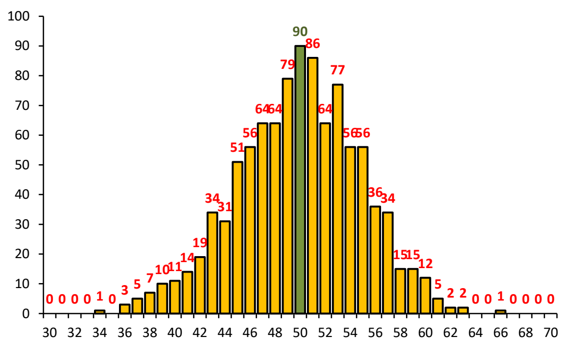


Figure 16. Typical Coin Flip Simulation (graph)

Example 7. Two-Dimensional Data Table (Generalized Birthday Problem)

The Data Table examples that we have presented so far have examined changes in a single variable. However, we also can vary two variables by using a two-dimensional data table. Here we look at a generalized version of the classical birthday problem [3,8]. As typically presented, in a group of *m* people chosen at random, we determine the probability that at least two people will share the same birthday. Here we present a model that will determine the smallest number, *m*, of people that must be selected for the probability that least one duplicate exceeds $p = 0.5$.

In fact, we generalize the process to the case of selecting *m* positive integers at random from the set of the first *n* positive integers ($n = 365$ gives the birthday problem), and finding the value of *m* that ensures that the probability of obtaining at least one duplicate value among the *m* selections exceeds a given probability, *p*. This is useful for illustrating the ideas of the birthday problem with a small class of size *m* by having the students in the class choose individual integers at random from the set {1, 2, …, *n*} instead of using birthdays. We start with the model of Figure 17.

First, we enter values for *n* (Cell A3) and *p* (Cell B3). In Column A we let *k* be a counter for the numbers of items that we are selecting. In Column B we compute the probability, P(*k*) that all of the first *k* selections are different. Clearly, P(1) = 1. Then, in computing P(*k*+1), we notice that the previous values must all be different, and that *n*–*k* will remain. Thus, the probability that the first *k*+1 are all different is $\mathrm{P}(k)(n-k)/n$. We enter this formula in Cell B7 and copy the current entries down their respective columns. The probability that there is at least one repeated value among the first *k* is computed in Column D as $1-\mathrm{P}(k)$. We find the desired number, *m*, of values needed for the probability to exceed *p* in Cell C3 by using the form of a table lookup function as in Figure 17b.

	A	B	C
1	Input		Output
2	n	p	num
3	365	0.5	23
4			
5	k	all diff	dups
6	1	1	0
7	2	0.9973	0.0027
8	3	0.9918	0.0082
9	4	0.9836	0.0164
10	5	0.9729	0.0271

Figure 17(a). Birthday (output)

	A	B	C
1	Input		Output
2	n	p	num
3	365	0.5	=LOOKUP(B3,C6:C105,A6:A105)+1
4			
5	k	all different	duplicates
6	1	1	=1-B6
7	=1+A6	=($A$3-A6)*B6/$A$3	=1-B7
8	=1+A7	=($A$3-A7)*B7/$A$3	=1-B8
9	=1+A8	=($A$3-A8)*B8/$A$3	=1-B9
10	=1+A9	=($A$3-A9)*B9/$A$3	=1-B10

Figure 17(b). Birthday (formulas)

Here =LOOKUP(B3,C6:C105,A6:A105)+1 looks for the value of B3 in the Block C6:C105, and returns the corresponding value in the Block A6:A105. This returns the value of *m* that gives the last probability that is less than or equal to the probability, *p*, that is being sought. Consequently, we add 1 in order to find the first value that exceeds *p*.

Now we create a 2-dimensional data table as shown in Figure 18(a). In Column F we create ranges of values of *n* down Column F and probabilities *p* in Row 2. We then use *Excel's* Data Table tool (see Figure 19) to fill in the number of integers that must be selected to ensure the probability that there is at least one duplicate exceeds *p*. First, in Cell F2 we enter the formula =C3, to obtain that number from our choices for *n* and *p* in Cells A3:B3.

	E	F	G	H	I	J	K	L
1			Probability of at least one duplicate					
2		23	0.40	0.50	0.60	0.70	0.80	0.90
3	Numbers available	50						
4		60						
5		70						
6		80						
7		90						
8		100						
9		200						
10		300						
11		365						
12		400						

Figure 18(a). 2-D Birthday (start)

	E	F	G	H	I	J	K	L
1			Probability of at least one duplicate					
2		23	0.40	0.50	0.60	0.70	0.80	0.90
3	Numbers available	50	8	9	10	12	13	15
4		60	9	10	11	13	14	17
5		70	9	11	12	14	15	18
6		80	10	11	13	14	17	19
7		90	10	12	14	15	17	21
8		100	11	13	14	16	18	22
9		200	15	17	20	23	26	31
10		300	18	21	24	27	32	37
11		365	20	23	27	30	35	41
12		400	21	24	28	32	36	43

Figure 18(b). 2-D Birthday (Data Table)

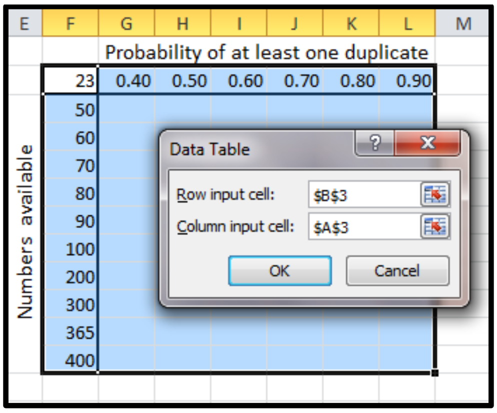


Figure 19. Two Dimensional Data Table Setup

As in Figure 19, we then use our selection device to highlight the array F2:L12 and select the Data Table command as before, this time entering the values for both n and p as shown in Figure 19. We obtain the output of Figure 18(b). For example, the number of integers that must be selected at random from a set of $n = 100$ integers that will ensure that the probability of at least one duplicate exceeds $p = 0.70$ is 16.

Example 8. Euler's Phi Function

The field of number theory provides many opportunities to employ a data table approach to classical topics. In this example we look at Euler's phi function [4]. For a positive integer, n, $\varphi(k)$ is the number of positive integers that are less than k and are relatively prime to k (*i.e.*, that have no common divisors with k other than 1. Thus $\varphi(6) = 2$ since only 1 and 5 have no divisors other than 1 in common with 6. In our example we use *Excel*'s greatest common divisor function, = gcd()..

In our model of Figure 20(a) we enter a value for k in Cell A2. We then generate consecutive positive integers, i, down Column A and the value of gcd(i,k) in Column B. In Column C for each i we produce the number 1 if and only if only k and i are relatively prime and $i < k$. Then, in Cell B2, we compute $\varphi(k)$ as the sum of the entries in Column C. The formulas appear in Figure 20(b).

We then use the Data Table command using Columns E:F where the formula in Cell F2 is simply =B2. We augment our model to indicate prime integers in Column H as those integers, n, for which $\varphi(n) = 1$. Figure 21 is a graph of the Euler Phi Function, formed from Columns E:F of our model, with points $(n,\varphi(n))$ for $n \leq 1000$.

	A	B	C	D	E	F
1	k	ϕ(k)			n	ϕ(n)
2	9	6				6
3	i	gcd(k,n)	count		2	
4	1	1	1		3	
5	2	1	1		4	
6	3	3			5	
7	4	1	1		6	
8	5	1	1		7	
9	6	3			8	
10	7	1	1		9	
11	8	1	1		10	
12	9	9			11	

Figure 20(a). Phi Initial.

	A	B	C
1	k	ϕ(k)	
2	9	=SUM(C4:C1003)	
3	i	gcd(k,n)	count
4	1	=GCD(A4,$A$2)	=IF(AND(A4<$A$2,B4=1),1,"")
5	=1+A4	=GCD(A5,$A$2)	=IF(AND(A5<$A$2,B5=1),1,"")
6	=1+A5	=GCD(A6,$A$2)	=IF(AND(A6<$A$2,B6=1),1,"")
7	=1+A6	=GCD(A7,$A$2)	=IF(AND(A7<$A$2,B7=1),1,"")
8	=1+A7	=GCD(A8,$A$2)	=IF(AND(A8<$A$2,B8=1),1,"")
9	=1+A8	=GCD(A9,$A$2)	=IF(AND(A9<$A$2,B9=1),1,"")
10	=1+A9	=GCD(A10,$A$2)	=IF(AND(A10<$A$2,B10=1),1,"")
11	=1+A10	=GCD(A11,$A$2)	=IF(AND(A11<$A$2,B11=1),1,"")
12	=1+A11	=GCD(A12,$A$2)	=IF(AND(A12<$A$2,B12=1),1,"")

Figure 20(b). Phi Formulas.

	A	B	C	D	E	F	G	H
1	k	ϕ(k)			n	ϕ(n)		
2	9	6				6		prime
3	i	gcd(k,n)	count		2	1		2
4	1	1	1		3	2		3
5	2	1	1		4	2		
6	3	3			5	4		5
7	4	1	1		6	2		
8	5	1	1		7	6		7
9	6	3			8	4		
10	7	1	1		9	6		
11	8	1	1		10	4		
12	9	9			11	10		11

Figure 20(c). Phi Data Table

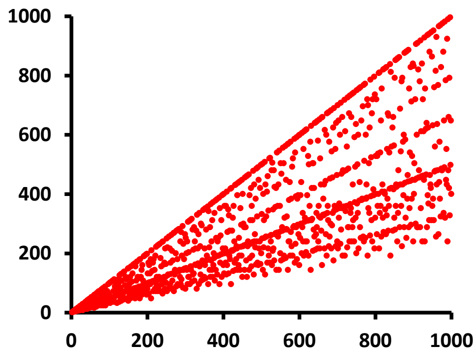


Figure 21. Euler Phi Function Graph

Example 9: Greatest Common Divisor

In Example 8 we used *Excel*'s built-in greatest common devisor function [4]. However, we can use also use a data table together with Euclid's GCD algorithm to produce these greatest common values. That algorithm can be expressed in a sequence of operations that we can implement naturally in a spreadsheet: Suppose that m and n are positive integers with $m > n$. Then, by the standard division algorithm, there are non-negative integers q_i and r_i so that

$$m = q_1 n + r_1 \text{ with } 0 \leq r_1 < n$$
$$n = q_2 r_1 + r_2 \text{ with } 0 \leq r_2 < r_1$$
$$r_1 = q_3 r_2 + r_3 \text{ with } 0 \leq r_3 < r_2$$

and so on. Since the r_i values are continually decreasing, eventually there is a first value of k for which $r_{k+1} = 0$. and r_k is the sought for greatest common divisor of m and n. See [4].

In Figures 22(a) and 22(b) we present an *Excel* model for this algorithm. Using this in Figure 22(c) we use the Data Table tool to find the greatest common divisors of m and n for n = 1, 2, 3, …

	A	B	C	D
1	Input	m	n	
2		36	10	
3	Output	locate	gcd(m,n)	
4	Locate	5	2	
5				
6	Large	Small	Rem	
7	36	10	6	
8	10	6	4	
9	6	4	2	
10	4	2	0	
11	2	0	#DIV/0!	
12	0	#DIV/0!	#DIV/0!	
13	#DIV/0!	#DIV/0!	#DIV/0!	

Figure 22(a). Euclid

	A	B	C
1	Input	m	n
2		36	10
3	Output	locate	gcd(m,n)
4	Locate	=MATC	=INDEX(B7:B28,B4-1)
5			
6	Large	Small	Rem
7	=B2	=C2	=MOD(A7,B7)
8	=B7	=C7	=MOD(A8,B8)
9	=B8	=C8	=MOD(A9,B9)
10	=B9	=C9	=MOD(A10,B10)
11	=B10	=C10	=MOD(A11,B11)
12	=B11	=C11	=MOD(A12,B12)
13	=B12	=C12	=MOD(A13,B13)

Figure 22(b). Formulas

	A	B	C	D	E	F
1	Input	m	n			
2		36	10			
3	Output	locate	gcd(m,n)			
4	Locate	5	2			
5					n	gcd(m.n)
6	Large	Small	Rem			2
7	**36**	**10**	6		1	1
8	10	6	4		2	2
9	6	4	2		3	3
10	4	2	0		4	4
11	2	0	#DIV/0!		5	1
12	0	#DIV/0!	#DIV/0!		6	6
13	#DIV/0!	#DIV/0!	#DIV/0!		7	1

Figure 22(c). GCD Data Table

Final Remark

In this paper we have presented only a few topics that can be explored creatively using the Data Table tool. In particular, examples such as creating Bezier curves for numerical analysis, implementing the Sieve of Eratosthenes in number theory, and executing legislative apportionment algorithms are discussed in [3]. Another interesting application lies in the computation of the probabilities of obtaining false positives in drug testing.

References

[1] Abramovich, Sergei. *Exploring Mathematics with Integrated Spreadsheets in Teacher Education*, World Scientific, Singapore, 2016.

[2] Burden, Richard, A. Burden, and J. D. Faires. *Numerical Analysis,* 10th ed., Cengage, Boston, 2016.

[3] Neuwirth, Erich and Deane Arganbright. *The Active Modeler: Mathematical Modeling with Microsoft Excel*, Cengage, Boston, 2004.

[4] Ore, Oystein. Number Theory and Its History, McGraw-Hill, New York, 1948.

[5] Ragsdale, Cliff. *Spreadsheet Modeling and Decision Modeling*, 8th ed., Cengage, Boston, 2018.

[6] Stewart, James. *Calculus: Early Transcendentals*, 8thed., Cengage, Boston, 2016.

[7] Strang, Gilbert. *Introduction to Linear Algebra*, 5th ed., Wellesley-Cambridge, Wellesley, 2016.

[8] Triola, Mario: *Elementary Statistics*, 12th ed., Pearson, London, 2012.