# Generating Nice Integer Data Sets For The Teaching Of Correlation And Regression

Kenneth H. Sutrick, (E-mail: ken.sutrick@murraystate.edu), Murray State University

Statistics instructors always have a choice in teaching statistical concepts. The choice is between using real data sets or data that has been constructed. For those statistics courses, whose students have good mathematical skills, real data is preferable demonstrating the usefulness and the applicability of the subject of statistics. However, students with less mathematical skills sometimes spend more time handling or worrying about calculations than learning the concepts taught. We believe that having data sets with integer values for averages and standard deviations in calculation intensive topics like correlation and regression allow the students to spend more time learning the concepts. The techniques we present also yield integer coefficients in regression and rational number values for the correlation. Hundreds of data sets can be constructed. A second benefit of this paper is in the taking and correcting of exams. Real data sets are unnecessary for exams (probably even undesirable). If a student on an exam gets a regression equation of the form: $\hat{Y} = 5 + 4X$ , they are confident that they have the correct answer. For the professor it is easier to grade correct answers.

Two techniques for generating data sets with nice numbers in correlation and regression are detailed. The techniques for finding integer solutions are based on first generating sets of uncorrelated variables and then using simple search algorithms on a computer. Section One has the first technique, Section Two has the second method which is based on permutations on the order of the data, Section Three discusses autocorrelation using techniques similar to Section One.

## SECTION ONE

This section finds data sets for correlation and regression based on the formulas:

$$r = \frac{SSXY/n}{\sigma_x \sigma_y}, \quad \hat{Y} = a + bX , \quad b = r\frac{\sigma_y}{\sigma_x}, \; a = \bar{Y} - b\bar{X} ,$$

where $SSXY = \sum (X - \bar{X})(Y - \bar{Y})$, $\sigma_X = \sqrt{\sum (X - \bar{X})^2 / n}$, $\sigma_Y = \sqrt{\sum (Y - \bar{Y})^2 / n}$ .

The technique will produce integer data; the parameters a, b, $\sigma_x$, $\sigma_y$ will be integers; r will then be a rational number. We illustrate by finding an integer data set with an integer standard deviation for n=4 pairs of data. Letting i, j, k be integers, then the deviations from the average $D_x$ satisfy $X_1 - \bar{X} = i,\ X_2 - \bar{X} = j,\ X_3 - \bar{X} = k,\ X_4 - \bar{X} = -i-j-k.$ To have an integer standard deviation $\sigma_x$ the equation $i^2+j^2+k^2+(i+j+k)^2 = 4*m_x^2$ must be satisfied for some integer $m_x$ . A sequence of four nested 'do loops' generates the required data. The Excel Macro Code "GetSDX" in the appendix generates the following sets of deviations $D_x^T$ ={i, j, k, -i-j-k}:

#1: $D_x^T$ = {-1, -1, 1, 1} with $\sigma_x$ =1
#2: $D_x^T$ = {-3, -1,-1, 5} with $\sigma_x$ =3
#3: $D_x^T$ = {-7, -1, 1, 7} with $\sigma_x$ =5
#4 $D_x^T$ = {-11, -1, 5, 7} with $\sigma_x$ =9
#5 $D_x^T$ = {-15, 1, 7, 7} with $\sigma_x$ =11.

The variable Y will be generated from deviations $Y-\bar{Y}=$ $D_y$ of the form $D_y = b*D_x + f*O_3 + g*O_4$ , where $O_3$, $O_4$ are orthogonal to $D_x$ and the unit vector $I^T = \{1,1,1,1\}$ ( $^T$ denotes transpose), and b, f, g are integer constants to be determined so that $D_y$ has an integer standard deviation. The vectors $O_3$, $O_4$ are easily determined with a little tweaking and trial and error from any linear algebra program that outputs rational numbers from the Gram-Schmidt process. [The vectors $O_3$, $O_4$ can also be determined from multiple regression since the residuals are orthogonal to the independent variables in a regression.] As an example, if the deviations are $D_x^T = \{ -3, -1, -1, 5\}$ then a possible basis is:

$$[\mathrm{I}\ \mathrm{D_X}\ \mathrm{O_3}\ \mathrm{O_4}] = \begin{bmatrix} 1 & -3 & 0 & 3 \\ 1 & -1 & -1 & -2 \\ 1 & -1 & 1 & -2 \\ 1 & 5 & 0 & 1 \end{bmatrix}$$

Now, define the inner product in the usual way: $(x, y) = \sum x_i y_i$ , so that orthogonal vectors have $(x, y) = 0$, and $\sigma_{D_x} = \sqrt{(D_x, D_x)/n}$ . Since $(D_x, D_x) = 36$, $(O_3, O_3)=2$, $(O_4, O_4)=18$, to have an integer value for $\sigma_y$, $D_y$ must have $(D_y, D_y)$ of the form $36b^2+2f^2+18g^2 = 4*m_y^2$ for some integer $m_y$. This requires only a slight modification of the code for the Excel Macro "GetSDX". The modification is given in the appendix as Excel Macro "GetSDY". GetSDY generates 68 possible values for f, g, and h. Two of those 68 are summarized directly below along with the correlation between $D_x$ and $D_y$, $r = (D_x, D_y) / [\sqrt{(D_x, D_x)}\ \sqrt{(D_y, D_y)}]$:

b=4, f=9, g=3 with $\sigma_y = 15$, $D_y^T = \{-3,-19,-1,23\}$ r=.8

b=2, f=3, g=3 with $\sigma_y = 9$, $D_y^T = \{3,-11,-5,13\}$ r=2/3.

The X variable can be any affine transformation of the deviations: $X= c + d*D_x$ , where c, d are integer constants. The variable Y can now be generated as $Y=a +b*X +d*f*O_3+d*g*O_4 = a +b*c +b*d*D_x +d*f*O_3+d*g*O_4$. Given the orthogonality it is easy to compute that $SSXY=b*d^2(D_x, D_x)$, $\sigma_x=d\sqrt{(D_x,D_x)/n}$ , $\sigma_y=d*m_y$. The correlation is, of course, unchanged by the linear transformation of the variables, and is the same as reported above. Also, since by construction $O_3$, $O_4$ are orthogonal to X, the regression of Y on X yields the regression equation $\hat{Y}=a+bX$, with the same a, b as in the definition of Y.

This section is finished by reporting two produced data sets.

The quantities $D_x^T = \{-3,-1,-1,5\}$, c=4, d=1, b=4, f=9, g=3 bring about a data set:

| X | Y |
|---|---|
| 1 | 18 |
| 3 | 2 |
| 3 | 20 |
| 9 | 44. |

$\bar{X} = 4$, $\sigma_x = 3$, $\bar{Y} = 21$, $\sigma_y = 15$, r = .8, $\hat{Y}=5+4X$,

$\mathrm{RMSE}=\sqrt{\sum (Y - \hat{Y})^2 / n} = 9$.

The quantities $D_x^T = \{-3,-1,-1,5\}$, c=5, d=1, b=3, f=12, g=4 bring about the data set:

| X | Y |
|---|---|
| 2 | 28 |
| 4 | 2 |
| 4 | 26 |
| 10 | 44. |

$\bar{X} = 5$, $\sigma_x = 3$, $\bar{Y} = 25$, $\sigma_y = 15$, $r = .6$, $\hat{Y}=10+3X$, RMSE = 12.

## SECTION TWO

This section presents a second method for generating regression data sets. It starts with two sets of deviations $D_x$, $D_y$ fashioned as in Section One[$D_x$, $D_y$ can be the same]. The technique fixes the order of the $D_x$ deviations, then computes every permutation of the order of the $D_y$ deviations, followed by the calculated correlation between $D_x$ and the permuted $D_y$. A visual perusal of the resulting list will probably yield sets of deviations with clean correlations. Affine transformations of $D_x$, $D_y$ create regression data sets with integer parameters.

We illustrate with n=5 and formulas based on computing standard deviations with n-1:

$$r = \frac{SSXY/(n-1)}{s_x s_y}, \quad \hat{Y} = a + bX, \quad b = r\frac{s_y}{s_x}, \quad a = \bar{Y} - b\bar{X},$$

where $s_x = \sqrt{\sum (X-\bar{X})^2/(n-1)}$, $s_y = \sqrt{\sum (Y-\bar{Y})^2/(n-1)}$.

Two sets of integer sample standard deviation deviations were generated from GETSDX with n replaced by n-1: $D_x^T = \{-3,-1,1,1,2\}$, $s_x=2$, $D_y^T = \{-5,-2,-1,3,5\}$, $s_{y=}4$. The permutations were generated from an algorithm found in Knuth[1], and the Excel Macro code for the permutations of the integers 1, 2, 3, 4, 5 is given in the appendix. After that the VLOOKUP function can be used in Excel to permute the deviations $D_y$. Some of the 5!=120 permutations of $D_y$ along with their correlations with $D_x$ are given below:

$D_x^T = \{ -3 \quad -1 \quad 1 \quad 1 \quad 2 \}$

Permuted $D_y^T$ =

$\{ -1 \quad 5 \quad -5 \quad 3 \quad 1 \}$ $r = -1/4 = -.25$
$\{ 3 \quad 5 \quad -5 \quad -1 \quad -2 \}$ $r = -3/4 = -.75$
$\{ -5 \quad 5 \quad 3 \quad -1 \quad -2 \}$ $r = 1/4 = .25$
$\{ -1 \quad 5 \quad -2 \quad 3 \quad 5 \}$ $r = -11/32 = -.34375$
$\{ 3 \quad 5 \quad -2 \quad -1 \quad -5 \}$ $r = -27/32 = -.84375$
$\{ -5 \quad -3 \quad 1 \quad 2 \quad 5 \}$ $r = 31/32 = .96875$
$\{ -1 \quad -5 \quad 5 \quad 3 \quad -2 \}$ $r = 3/8 = .375$
$\{ 3 \quad -5 \quad 5 \quad -1 \quad -2 \}$ $r = -1/8 = -.125$
$\{ -5 \quad -1 \quad 5 \quad 3 \quad 2 \}$ $r = 5/8 = .625.$

The X,Y variables are formed as $X=c+d*D_x$, $Y=f+g*D_y$ yielding a regression equation $\hat{Y}=a+bX$, where

$$b=\frac{g}{d}\frac{(D_x,D_y)}{(D_x,D_x)}, \quad a=f-b*c.$$

Choosing c, d, f, g appropriately will make a and b integers. [Obviously, if $g=(D_x,D_x)k$ and $d=(D_x,D_y)$ for some integer k then a and b will be integers.]

We finish this section by using the technique to create two more data sets.

Take $D_x^T = \{ -3 \;\; -1 \;\; 1 \;\; 1 \;\; 2 \}$, $D_y^T = \{ -1 \;\; 5 \;\; -5 \;\; 3 \;\; -2 \}$ and choose c=5, d=1, f=22, g=4 to get the data set:

| X | Y |
|---|---|
| 2 | 18 |
| 4 | 42 |
| 6 | 2 |
| 6 | 34 |
| 7 | 14. |

$\bar{X} = 5$, $s_x = 2$, $\bar{Y} = 22$, $s_y = 16$, r = -.25, $\hat{Y}=32-2X$, RMSE=13.856.

Suppose $D_x^T = \{ -3 \;\; -1 \;\; 1 \;\; 1 \;\; 2\}$, $D_y^T = \{ -3 \;\; -5 \;\; 1 \;\; 5 \;\; 2 \}$. Choosing c=7, d=2, f=23, g=4 and the first set of deviations gives a data set:

| X | Y |
|---|---|
| 1 | 11 |
| 5 | 3 |
| 9 | 27 |
| 9 | 43 |
| 11 | 31. |

$\bar{X} = 7$, $s_x = 4$, $\bar{Y} = 23$, $s_y = 16$, r = .75, $\hat{Y}=2+3X$, RMSE=9.466.

## SECTION THREE

In this part the methods of Section One are used to derive data sets that have autocorrelations with rational values. Assume the time series model: Y= a +bT +E. The autocorrelation is computed from the residuals of the regression of Y on T. We shall produce a set of residuals with nice autocorrelations. Since the residuals from the regression are orthogonal to I and T, we shall again find a basis for the subspace of residuals. The complete set of possible regression residuals is then determined from linear combinations of the basis. A simple search allows one to find residuals with autocorrelations that are rational numbers.

As an example suppose n=7, T = { 1, 2, 3, 4, 5, 6, 7 }$^T$ . The subspace of residuals is orthogonal to:

$$I = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} \qquad T-\bar{T} = \begin{bmatrix} -3 \\ -2 \\ -1 \\ 0 \\ 1 \\ 2 \\ 3 \end{bmatrix}$$

One possible orthogonal basis for the residuals is:

$$O_3=\begin{bmatrix}0\\0\\1\\-2\\1\\0\\0\end{bmatrix}\quad O_4=\begin{bmatrix}0\\1\\-2\\0\\2\\-1\\0\end{bmatrix}\quad O_5=\begin{bmatrix}1\\-1\\0\\0\\0\\-1\\1\end{bmatrix}\quad O_6=\begin{bmatrix}3\\3\\-4\\-4\\-4\\3\\3\end{bmatrix}\quad O_7=\begin{bmatrix}5\\-6\\-3\\0\\3\\6\\-5\end{bmatrix}.$$

Y can then be determined as Y = a + bT+E = a + bT + $iO_3$ + $jO_4$ + $kO_5$ + $mO_6$ + $pO_7$, where a, b are arbitrary, and i, j, k, m, p are determined by a search. GetSDX() can be easily modified to accomplish this. Instead of checking if $i^2+j^2+k^2+(i+j+k)^2 = 4*m_x^2$ , one forms $iO_3 + jO_4 + kO_5 + mO_6 + pO_7$, then codes and prints the autocorrelation and supporting calculations like standard deviations. Six of the resulting data sets are given below:

Choosing a= 2, b=4, i=1, j=0, k=0, m=0, p=0 gives the data set below with a regression equation: $\hat{Y}=2+4X$, and residuals E:

| T | $Y_T$ | $E=Y_T-\hat{Y}_T$ |
|---|---|---|
| 1 | 6 | 0 |
| 2 | 10 | 0 |
| 3 | 15 | 1 |
| 4 | 16 | -2 |
| 5 | 23 | 1 |
| 6 | 26 | 0 |
| 7 | 30 | 0. |

Let $E_1$ be the first (n-1) components of E and let $E_2$ be the last (n-1) components of E, then the autocorrelation is corr($E_1$,$E_2$).

$$E_1=\begin{bmatrix}0\\0\\1\\-2\\1\\0\end{bmatrix}\quad E_2=\begin{bmatrix}0\\1\\-2\\1\\0\\0\end{bmatrix}.$$

Thus: $\bar{E}_1=0$, $\bar{E}_2=0$, $\sigma_{E_1}=1$ $\sigma_{E_2}=1$, the autocorrelation is corr($E_1$,$E_2$)= -2/3=-.666667.

Choosing a= 5, b=2, i=0, j=0, k=0, m=2, p=0 gives a second data set:

$Y^T$ = {13, 15, 3, 5, 7, 23, 25}, regression equation: $\hat{Y}=5+2X$, residuals:

$E^T$ = {6, 6, -8, -8, -8, 6, 6}, $\overline{E}_1=-1$, $\overline{E}_2=-1$, $\sigma_{E_1}=7$, $\sigma_{E_2}=7$, and the autocorrelation is corr($E_1$,$E_2$)= 1/3=.333334.

Choosing a= 1, b=3, i=4, j=0, k=-3, m=1, p=0 gives a third data set:

$Y^T$ = {4, 13, 10, 1, 16, 25, 22}, regression equation: $\hat{Y}=1+3X$, residuals:

$E^T$ = {0, 6, 0, -12, 0, 6, 0}, $\overline{E}_1=0$, $\overline{E}_2=0$, $\sigma_{E_1}=6$, $\sigma_{E_2}=6$, and the autocorrelation is corr($E_1$,$E_2$)= 0.

Choosing a= 3, b=5, i=-2, j=-3, k=-6, m=0, p=0 gives a fourth data set:

$Y^T$ = {2, 16, 22, 27, 20, 42, 32}, regression equation: $\hat{Y}=3+5X$, residuals:

$E^T$ = {-6, 3, 4, 4, -8, 9, -6}, $\overline{E}_1=1$, $\overline{E}_2=1$, $\sigma_{E_1}=6$, $\sigma_{E_2}=6$, and the autocorrelation is corr($E_1$,$E_2$)= -77/108= -.712963.

Choosing a= -1, b=1, i=1, j=1, k=0, m=0, p=0 gives a fifth data set:

$Y^T$ = {0, 2, 1, 1, 7, 4, 6}, regression equation: $\hat{Y}=-1+X$, residuals:

$E^T$ = {0, 1, -1, -2, 3, -1, 0}, $\overline{E}_1=0$, $\overline{E}_2=0$, $\sigma_{E_1}=\sqrt{8/3}$, $\sigma_{E_2}=\sqrt{8/3}$, and the autocorrelation is corr($E_1$,$E_2$)= -.5

Choosing a= 25, b=-2, i=5, j=-3, k=6, m=-2, p=0 gives the sixth data set:

$Y^T$ = {23, 6, 38, 15, 22, 4, 11}, regression equation: $\hat{Y}=25-2X$, residuals:

$E^T$ = {0, -15, 19, -2, 7, -9, 0}, $\overline{E}_1=0$, $\overline{E}_2=0$, $s_{E_1}=12$, $s_{E_2}=12$, and the autocorrelation is corr($E_1$,$E_2$)= -5/9= -.555556

## CONCLUSIONS

With a small amount of work one can generate a large number of data sets with integer standard deviations, integer regression coefficients, and correlations that are rational numbers.

## REFERENCES

1. Knuth, Donald E. (1973) The Art of Computer Programming, Volume One, *Fundamental Algorithms*, Second Edition, p. 44, Addison-Wesley.

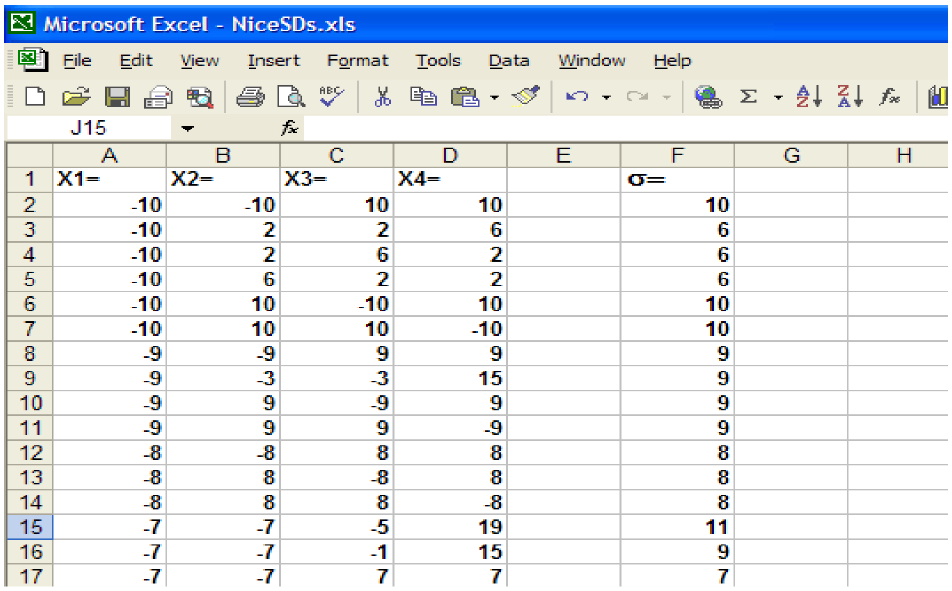
**APPENDIX**

**Code Set #1**

The first code set generates deviations giving integer values of the standard deviations, $\sigma_x$ for the case n=4:

```
Sub GetSDX()
Dim i As Integer, j As Integer, k As Integer
Dim count As Integer

count = 1
For i = -10 To 10
For j = -10 To 10
For k = -10 To 10
For m = 1 To 40
If i ^ 2 + j ^ 2 + k ^ 2 + (i + j + k) ^ 2 = 4 * m ^ 2

Then

count = count + 1
Cells(count, 1).Value = i
Cells(count, 2).Value = j
Cells(count, 3).Value = k
Cells(count, 4).Value = -i - j - k
Cells(count, 6).Value = m
End If
Next m
Next k
Next j
Next i
End Sub
```

Below is a screen shot of part of the output of GetSDX( ):


Microsoft Excel - NiceSDs.xls
File Edit View Insert Format Tools Data Window Help
J15

| | A | B | C | D | E | F | G | H |
|---|---|---|---|---|---|---|---|---|
| 1 | X1= | X2= | X3= | X4= | | σ= | | |
| 2 | -10 | -10 | 10 | 10 | | 10 | | |
| 3 | -10 | 2 | 2 | 6 | | 6 | | |
| 4 | -10 | 2 | 6 | 2 | | 6 | | |
| 5 | -10 | 6 | 2 | 2 | | 6 | | |
| 6 | -10 | 10 | -10 | 10 | | 10 | | |
| 7 | -10 | 10 | 10 | -10 | | 10 | | |
| 8 | -9 | -9 | 9 | 9 | | 9 | | |
| 9 | -9 | -3 | -3 | 15 | | 9 | | |
| 10 | -9 | 9 | -9 | 9 | | 9 | | |
| 11 | -9 | 9 | 9 | -9 | | 9 | | |
| 12 | -8 | -8 | 8 | 8 | | 8 | | |
| 13 | -8 | 8 | -8 | 8 | | 8 | | |
| 14 | -8 | 8 | 8 | -8 | | 8 | | |
| 15 | -7 | -7 | -5 | 19 | | 11 | | |
| 16 | -7 | -7 | -1 | 15 | | 9 | | |
| 17 | -7 | -7 | 7 | 7 | | 7 | | |

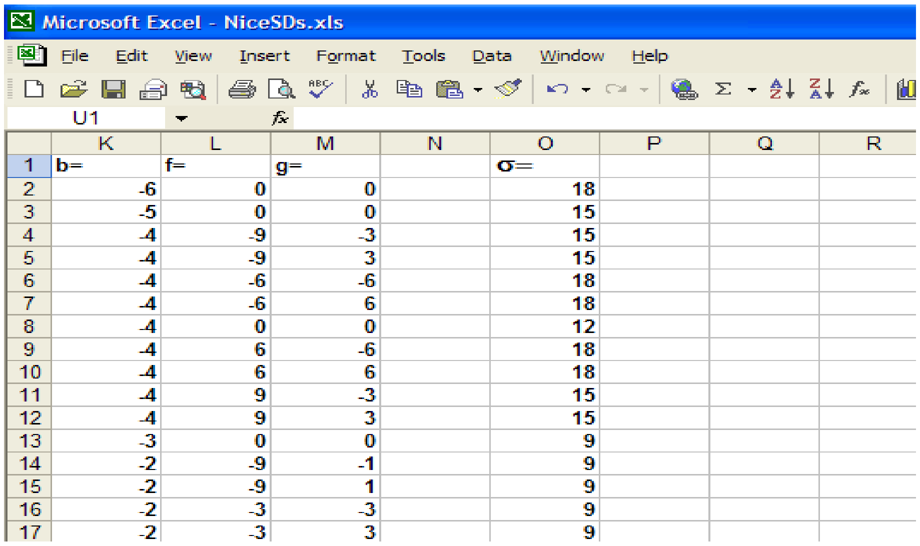
The data -10, -10, 10, 10 and -9, -9, 9, 9 are the same data set except for a scale factor. It is possible to write code that will report only one set of this type of data. The code is somewhat long and in our opinion not worth the effort.

**Code Set#2**

This set of code computes the sample standard deviation, $s_y$ for the case n=5:

```
Sub GetSDY()
Dim i As Integer, j As Integer, k As Integer
Dim count As Integer

count = 1
For i = -10 To 10
For j = -10 To 10
For k = -10 To 10
For m = 1 To 20
If 36 * i ^ 2 + 2 * j ^ 2 + 18 * k ^ 2 = 4 * m ^ 2

Then

count = count + 1
Cells(count, 11).Value = i
Cells(count, 12).Value = j
Cells(count, 13).Value = k
Cells(count, 15).Value = m
End If
Next m
Next k
Next j
Next i
End Sub
```

Below is a screen shot of part of the output of GetSDY( ):



Microsoft Excel - NiceSDs.xls
File Edit View Insert Format Tools Data Window Help
U1

| | K | L | M | N | O | P | Q | R |
|---|---|---|---|---|---|---|---|---|
| 1 | b= | f= | g= | | σ= | | | |
| 2 | -6 | 0 | 0 | | 18 | | | |
| 3 | -5 | 0 | 0 | | 15 | | | |
| 4 | -4 | -9 | -3 | | 15 | | | |
| 5 | -4 | -9 | 3 | | 15 | | | |
| 6 | -4 | -6 | -6 | | 18 | | | |
| 7 | -4 | -6 | 6 | | 18 | | | |
| 8 | -4 | 0 | 0 | | 12 | | | |
| 9 | -4 | 6 | -6 | | 18 | | | |
| 10 | -4 | 6 | 6 | | 18 | | | |
| 11 | -4 | 9 | -3 | | 15 | | | |
| 12 | -4 | 9 | 3 | | 15 | | | |
| 13 | -3 | 0 | 0 | | 9 | | | |
| 14 | -2 | -9 | -1 | | 9 | | | |
| 15 | -2 | -9 | 1 | | 9 | | | |
| 16 | -2 | -3 | -3 | | 9 | | | |
| 17 | -2 | -3 | 3 | | 9 | | | |

**Code Set#3**

The next program generates permutations of the integers 1, 2, 3, …, n with an algorithm found in Knuth[1]. It does so iteratively by starting with all permutations of the integers 1, 2, …, i-1 then using this set to generate all permutations of the integers 1, 2, …, i-1, i. For each permutation $a_1a_2a_3…a_{i-1}$ on i-1 elements, form i others by inserting the number i in all possible places, obtaining: i $a_1$ $a_2$ $a_3$ … $a_{i-1}$, $a_1$ i $a_2$ $a_3$ … $a_{i-1}$, $a_1$, $a_2$ i $a_3$ … $a_{i-1}$, ……, $a_1$ $a_2$ $a_3$ … $a_{i-1}$ i.

The code below generates the following sequence of permutations: (only n=1, n=2, and n=3 are shown)

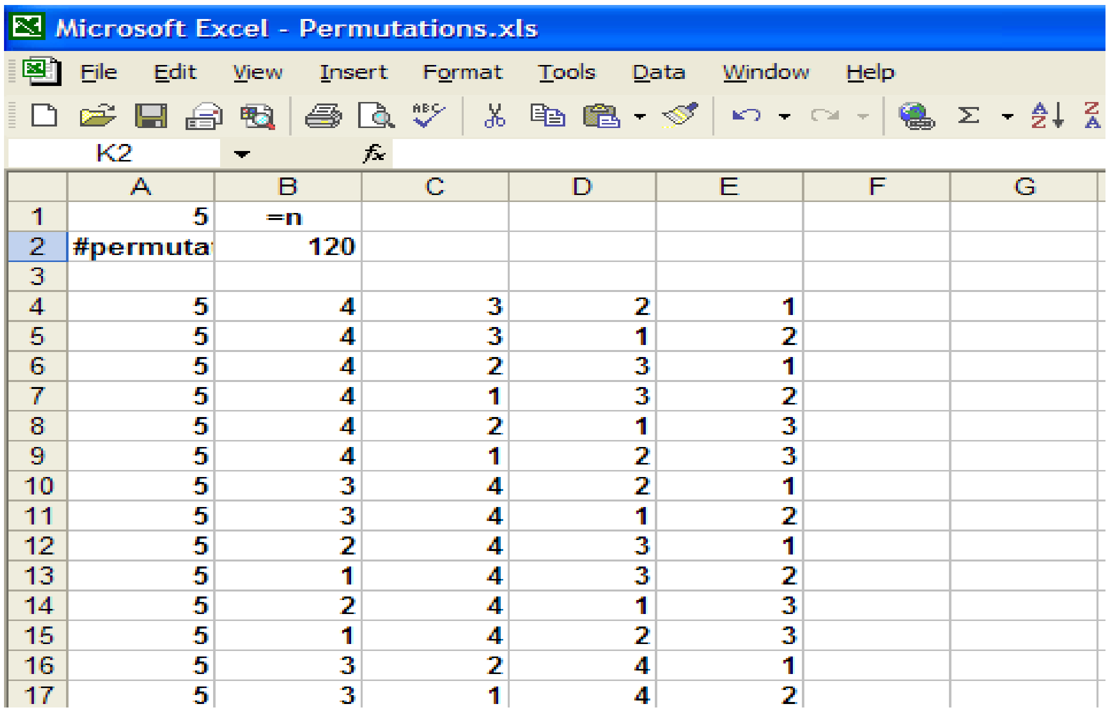
```
Permutations of one interger:       1

Permutations of two integers:        2  1
                                     1  2
Permutations of three integers:      3  2  1
                                     3  1  2
                                     2  3  1
                                     1  3  2
                                     2  1  3
                                     1  2  3
```

The code uses array A for the set of permutations of 1, 2, …, i-1,and array B for the next set of permutations of 1, 2, …, i-1, i.

```
Sub Perm()
Dim n2 As Integer
Dim n As Integer
Dim A(1 To 120, 1 To 6) As Integer
Dim B(1 To 720, 1 To 7) As Integer
n = Cells(1, 1).Value       ‘ Read the value of n from the spread sheet
Cells(4, 1).Value = 1       ‘ Start the algorithm with the permutations of one element(=1)
ntot = 1

For i = 2 To n
ntot = ntot * i               ‘ ntot is the total number or permutations of  i elements
ntotpre = ntot / i            ‘ ntotpre is the total number or permutations of  i-1 elements

For j = 1 To ntotpre
For m = 1 To i - 1
A(j, m) = Cells(3 + j, m).Value    ‘ Read the current set of permutations from the spreadsheet
Next
Next

For j = 1 To ntotpre           ‘ Generate permutations of the form:   i a1 a2 a3 … ai-1
B(j, 1) = i                    ‘ Put i into first column of B, A into last i-1 columns of B
For m = 1 To i
B(j, m + 1) = A(j, m)
Next
Next

For k = 2 To i - 1             ‘ Generate permutations of the form:  a1 i a2 a3 … ai-1,
For j = 1 To ntotpre             a1, a2 i a3 … ai-1, …
For m = 1 To k - 1
```

```
B(j + (k - 1) * ntotpre, m) = A(j, m)       ‘ Put the first k-1 columns of A in B
Next
B(j + (k - 1) * ntotpre, k) = i             ‘ Put i into the k-th column of B
For m = k + 1 To i
B(j + (k - 1) * ntotpre, m) = A(j, m - 1)     ‘ Put the last i-k columns of A into the
Next                                           ‘ last i-k columns of B
Next
Next

For j = 1 To ntotpre                ‘ Generate permutations of the form:  a1 a2 a3 … ai-1 i.
For    m = 1 To i                   ‘ Put A into first i-1 columns of B, put i into i-th column
B(j + (i - 1) * ntotpre, m) = A(j, m)      ‘ of B
Next
B(j + (i - 1) * ntotpre, i) = i
Next

For j = 1 To ntot                   ‘ Print the new set of permutations B into the spreadsheet
For m = 1 To i
Cells(3 + j, m).Value = B(j, m)
Next
Next
Next
Cells(2, 2).Value = ntot
End Sub
```

Below is a screen shot of part of the output of the Macro  Perm( ):


Microsoft Excel - Permutations.xls
File Edit View Insert Format Tools Data Window Help
K2

| | A | B | C | D | E | F | G |
|---|---|---|---|---|---|---|---|
| 1 | 5 | =n | | | | | |
| 2 | #permuta | 120 | | | | | |
| 3 | | | | | | | |
| 4 | 5 | 4 | 3 | 2 | 1 | | |
| 5 | 5 | 4 | 3 | 1 | 2 | | |
| 6 | 5 | 4 | 2 | 3 | 1 | | |
| 7 | 5 | 4 | 1 | 3 | 2 | | |
| 8 | 5 | 4 | 2 | 1 | 3 | | |
| 9 | 5 | 4 | 1 | 2 | 3 | | |
| 10 | 5 | 3 | 4 | 2 | 1 | | |
| 11 | 5 | 3 | 4 | 1 | 2 | | |
| 12 | 5 | 2 | 4 | 3 | 1 | | |
| 13 | 5 | 1 | 4 | 3 | 2 | | |
| 14 | 5 | 2 | 4 | 1 | 3 | | |
| 15 | 5 | 1 | 4 | 2 | 3 | | |
| 16 | 5 | 3 | 2 | 4 | 1 | | |
| 17 | 5 | 3 | 1 | 4 | 2 | | |