

# Sample size calculation to data analysis of a correlation study in Microsoft Excel®: A hands-on guide with example

Himel Mondal, Shaikat Mondal<sup>1</sup>

Department of Physiology, Maharaja Krishna Chandra Gajapati Medical College, Ganjam, Odisha, <sup>1</sup>Department of Physiology, Medical College and Hospital, Kolkata, West Bengal, India

## Abstract

**Background and Aim:** Investigators use correlation study to determine the extent to which two or more variables are related among a single group of people. One big question frequently encountered by investigators is How to calculate sample size? After data collection, it is essential to test the data statistically. Hence, we planned to provide hands-on guide for sample size calculation and statistical analysis of data in Microsoft Excel® 2010 commonly needed for a correlation study.

**Methods:** An example of a correlation study of body mass index and peak expiratory flow rate was taken. Its execution process from sample size calculation to statistical analysis of data has been explained step by step in Excel® 2010. All formula and functions were tested in Excel® 2007.

**Conclusion:** Excel® as a tool for statistical analysis of data would facilitate researchers to do all statistical analyses in a single file. Along with correlation study, analysis by *t*-test and ANOVA was also provided. With this discussion, new investigators would get a fair knowledge about sample size calculation and data analysis of correlation study in Excel® and can use it for their future studies.

**Key words:** Correlation study, Excel 2010, sample size, statistical analysis, *t*-test

*Received:* 13<sup>th</sup> October, 2016; *Revised:* 5<sup>th</sup> December, 2016; *Accepted:* 15<sup>th</sup> December, 2016

## INTRODUCTION

### Correlation study

Correlation study is used to determine the relationship between two or more continuous variables among a single group of people. A correlation has direction, and correlation coefficient (*r*) can be either positive or negative. Zero indicates no relationship between the two variables, and  $r = 1$  or  $r = -1$  indicate a perfect relationship. The strength can be anywhere between 0 and  $\pm 1$ .<sup>[1]</sup> Correlation gives the degree and direction of relationship between the two variables, whereas the regression analysis predicts the values of one variable on the basis of the other variable. When there is perfect correlation ( $r = +1$  or  $-1$ ), the two regression lines will coincide or become one straight line.<sup>[2]</sup>

### Essential terminology

#### Null hypothesis

The null hypothesis states that there is no association between the predictor and outcome variables in the population.

#### Alternative hypothesis

The proposition that there is an association is called the alternative hypothesis. The alternative hypothesis cannot be tested directly; it is accepted by default if the test of statistical significance rejects the null hypothesis.

**Address for correspondence:** Dr. Himel Mondal, Department of Physiology, MKCG Medical College, Ganjam - 760 004, Odisha, India.  
E-mail: himelmckg@gmail.com

### Access this article online

#### Quick Response Code:



**Website:**  
www.ijcep.org

**DOI:**  
10.4103/2348-8832.196896

This is an open access article distributed under the terms of the Creative Commons Attribution-NonCommercial-ShareAlike 3.0 License, which allows others to remix, tweak, and build upon the work non-commercially, as long as the author is credited and the new creations are licensed under the identical terms.

**For reprints contact:** reprints@medknow.com

**How to cite this article:** Mondal H, Mondal S. Sample size calculation to data analysis of a correlation study in Microsoft Excel®: A hands-on guide with example. Int J Clin Exp Physiol 2016;3:180-9.

**Type I error**

A type I error (false positive) occurs if an investigator rejects a null hypothesis that is actually true in the population.

**Type II error**

Type II error (false negative) occurs if the investigator fails to reject a null hypothesis that is actually not true in the population [Table 1].<sup>[3]</sup>

**Effect size**

Effect size is a quantitative reflection of the magnitude of some phenomenon that is used for the purpose of addressing a question of interest. The investigator should first try to find data from prior studies in related areas to estimate reasonable effect size. When data are not available, a small pilot study may be done.<sup>[4]</sup>

**α:** The significance level of a test; the probability of rejecting the null hypothesis when it is true (or the probability of making a type I error).

**β:** The probability of failing to reject the null hypothesis when it is false (or the probability of making type II error).

**Confidence level**

The probability than an estimate of population parameter is within certain specified limits of the true value; commonly denoted by “1 – α.”

**Power of test**

The probability of correctly rejecting the null hypothesis when it is false; commonly denoted by “1 – β.”<sup>[5]</sup>

**MATERIALS AND METHODS**

Commonly, researchers save raw data in Excel® and do statistical analysis in other software. It may be beneficial for them if they can do statistical analysis in the same Excel® sheet. Hence, we have discussed about the procedure of statistical analysis of data in Excel®. Screenshots of Windows were captured by Greenshot 1.2.7 Build 2 software for Windows. Though the screenshots were taken while executing analysis in Excel® 2010, we have tested all formula and functions in Excel® 2007, and all functions and formulae were compatible with Excel® 2007.

**Table 1:** Null hypothesis, Type I and Type II error

Truth in population	Decision	
	Null hypothesis retained	Null hypothesis rejected
True	Correct (1-α)	Type I error (α)
False	Type II error (β)	Correct (1-β)

**Designing a correlation study**

Our research question was – whether body mass index (BMI) is correlated with peak expiratory flow rate (PEFR)? Similar previous studies with correlation of BMI and PEFR were available and had been studied beforehand. Among those studies, one study was almost similar regarding the age, race, and other characteristics of participants of our aimed study. That study found a correlation coefficient (*r*) = 0.30. Hence, we had anticipated in our study that BMI would be correlated with PEFR as that study.

**Our null hypothesis**

There is no correlation between BMI and PEFR in the study population.

**Alternative hypothesis**

There is a correlation between BMI and PEFR in the study population.

**α:** (two sided) we needed only 5% chance of rejecting null hypothesis when it is true. Hence, “α” value was 0.05. Confidence level was 1 – 0.05 = 0.95 or 95%.

**β:** We needed 5% probability of failing to reject null hypothesis when it is false. Hence, “β” value was 0.05 and power of test was 1 – 0.05 = 0.95 or 95%.

**Effect size**

From the previous study, we found correlation coefficient (*r*) = 0.30 between BMI and PEFR. Hence, our effect size was |0.30| = 0.30.

The formula for sample size calculation for correlation study is as follows:<sup>[6]</sup>

$$N = ([Z_{\alpha} + Z_{\beta}]/C)^2 + 3$$

where, *N* = number of subjects required

*Z*<sub>α</sub> = the standard deviation for α

*Z*<sub>β</sub> = the standard deviation for β

$$C = 0.5 * \ln ([1 + r]/[1 - r])$$

*r* = expected correlation coefficient

**Sample size calculator**

A new Excel® sheet was opened. Three cells were selected for texts – “α=,” “β=,” and “expected *r* =.” Immediate right adjacent cells were colored for easy identification. We used those cells for inserting values for α, β, and *r*. On the right side of those cells, a cell text was written as “*Z*<sub>α</sub>=,” and the adjacent right cells’ formula was written as

"=NORMSINV((1-(cell address of  $\alpha/2$ )))". In our example, it was "=NORMSINV((1-(B4/2)))" [Figure 1]. As " $\alpha$ " was "two sided," we had divided it by 2. In similar fashion, " $Z_\beta$ " cells were selected and the formula was written as "=NORMSINV((1 - cell address of  $\beta$ ))." In the example, it was "=NORMSINV((1 - B6))." Another cell text was written as "C=" and the adjacent right cells' formula was written as "=0.5\*LN ((1 + cell address of expected  $r$ )/(1 - cell address of expected  $r$ ))." In our example data sheet, it was "=0.5\*LN ((1 + B8)/(1 - B8))" [Figure 2]. For the cell right side to "Sample size=", formula was written as "= (((cell address of  $Z_\alpha$  + cell address of  $Z_\beta$ )/cell address of C) ^ 2) +3." In the said example, it was "= (((E4 + E6)/E8) ^ 2) +3" [Figure 3]. For making the number as an integer, the formula is "=ROUND (cell address of sample size, 0)." In the example of Figure 3, it was "=ROUND (E11, 0)." With this calculator, after inputting values of  $\alpha$ ,  $\beta$ , and  $r$ , we found that the number of participants required for our study was 139.

This calculator can be saved and used for sample size calculation of any correlation study with the effect size being calculated by correlation coefficient ( $r$ ). Only  $\alpha$ ,  $\beta$ , and  $r$  values are required in specified cells, and the sample size will be calculated according to the formula we used.

### Data entry

We had taken the following data (1) height in cm to the nearest 0.1 cm, (2) weight in kg to the nearest 0.1 kg, and (3) PEFR in L/min. In an Excel® sheet, the 1<sup>st</sup> column

(column B) label was written as "Weight (kg)," in first row and in the first row, the 2<sup>nd</sup> column label was written as "Height (cm)," the 3<sup>rd</sup> as BMI, and the 4<sup>th</sup> as "PEFR (L/min)," [Figure 4] and the data were entered accordingly. For ease of description, only eight sets of data were entered.

### Calculating body mass index, mean, standard deviation and standard error

For calculation of BMI according to Quetelet's equation,<sup>[7]</sup> we had to convert the "Height in cm" to "Height in meter." A cell below the "BMI" cell was selected and the formula was written as "=cell address of weight in kg/(cell address of height (cm)/100) ^ 2." In the example of Figure 4, it was "=B2/(C2/100) ^ 2."

Arithmetic mean is the average. Hence, for the calculation of mean of the data of the column B, the cell equation was written as "=AVERAGE (starting cell address of column B: last cell address of column B)." For cell B10, it was "=AVERAGE (B2:B9)" [Figure 5]. For calculation of standard deviation (SD), the cell equation used with STDEV function was "=STDEV (starting cell address of column: Last cell address of column)." Here, it was "=STDEV (B2:B9)" in cell B11. For calculation of standard error (SE), we have to use the formula  $SE = SD/\sqrt{N}$ ,<sup>[1]</sup> where  $N$  is the sample size. If we use this formula in Excel® for the data of column B in the example of Figure 5, it will be "=STDEV (B2:B9)/SQRT ((COUNT (B2:B9)))." We can do it more simply as we already calculated the SD and we know the  $N$ . We can obtain the SE by writing a cell formula as "=B11/SQRT(8)" also.

	E4		$f_x$	=NORMSINV((1-(B4/2)))	
	A	B	C	D	E
1					
2	Input Data		Output data		
3					
4	$\alpha$ =	0.05		$Z_\alpha$ =	1.959964
5					
6	$\beta$ =	0.05		$Z_\beta$ =	1.644854
7					
8	expected $r$ =	0.3		C=	0.30952
9					
10					
11				Sample size =	138.6408
12				or	139

Figure 1: Screenshot showing calculation of  $Z_\alpha$  in cell E4 with the value of cell B4 in Excel® 2010

E8		fx =0.5*LN((1+B8)/(1-B8))			
	A	B	C	D	E
1					
2	Input Data		Output data		
3					
4	$\alpha =$	0.05		$Z_{\alpha} =$	1.959964
5					
6	$\beta =$	0.05		$Z_{\beta} =$	1.644854
7					
8	expected $r =$	0.3		C =	0.30952
9					
10					
11				Sample size =	138.6408
12				or	139

Figure 2: Calculation of C in cell E8 with the value of cell B8

E11		fx =(((E4+E6)/E8)^2)+3			
	A	B	C	D	E
1					
2	Input Data		Output data		
3					
4	$\alpha =$	0.05		$Z_{\alpha} =$	1.959964
5					
6	$\beta =$	0.05		$Z_{\beta} =$	1.644854
7					
8	expected $r =$	0.3		C =	0.30952
9					
10					
11				Sample size =	138.6408
12				or	139

Figure 3: Calculation of sample size for correlation study in cell E11 with the calculated value of cells E4, E6, and E8 based on the equation:  $N = ([Z_{\alpha} + Z_{\beta}] / C)^2 + 3$

**Calculation of Pearson’s correlation coefficient (r)**

For calculation of Pearson’s correlation coefficient (r), we had written a cell text as “r=,” and the formula for

adjacent right cell was written with PEARSON function as: “=PEARSON (first cell address of first column: Last cell address of first column, first cell address of second

D2				fx		=B2/(C2/100)^2	
	A	B	C	D	E		
1		Weight(kg)	Height(cm)	BMI	PEFR(L/min)		
2		54	173	18.0427	225.42		
3		42.4	151	18.59568	224.5		
4		50.6	152.5	21.75759	239.7		
5		42	151	18.42024	221.2		
6		66.7	170	23.07958	238.77		
7		69.2	178.5	21.71849	240.92		
8		62	170	21.45329	192.68		
9		74.3	159.5	29.20569	267.09		
10	Mean =	57.65	163.1875	21.53416	231.285		
11	SD =	12.2588743	11.009533	3.622402	21.29438692		
12							

Figure 4: Calculation of body mass index in cell D2 with the value of cells B2 and C2 on the basis of Quetelet's equation: body mass index = Weight in kg/height in m<sup>2</sup>

B10				fx		=AVERAGE(B2:B9)	
	A	B	C	D	E		
1		Weight(kg)	Height(cm)	BMI	PEFR(L/min)		
2		54	173	18.0427	225.42		
3		42.4	151	18.59568	224.5		
4		50.6	152.5	21.75759	239.7		
5		42	151	18.42024	221.2		
6		66.7	170	23.07958	238.77		
7		69.2	178.5	21.71849	240.92		
8		62	170	21.45329	192.68		
9		74.3	159.5	29.20569	267.09		
10	Mean =	57.65	163.1875	21.53416	231.285		
11	SD =	12.2588743	11.009533	3.622402	21.29438692		

Figure 5: Calculation of mean of values of cell B2 through B9 in the cell B10 with the "AVERAGE" function

column: Last cell address of second column)." Here, it was "=PEARSON (D2:D9, E2:E9)." In the example of Figure 6, the value for "r" was calculated in cell H2 and the value was 0.675689.

Alternatively, we could use the "Data Analysis Tools" to get the correlation coefficient. For correlation analysis, we have to follow the steps: Data tab (on ribbon)→ Data

Analysis Tools (on the extreme right side under data menu)→Correlation→Ok→ Input range (\$D\$2:\$E\$9 for the example) and output range (any cell where the result to be shown, e.g., \$A\$13) → Ok.

Under "Data" tab (on ribbon), if "Data Analysis Tools" is not shown, we have to follow the steps: File → Option → Add Ins → Select "Analysis ToolPak" → Click on "Go"

→ Click “Ok” in new window [Figure 7]. Now, we have enabled data analysis tool for Excel®.<sup>[8]</sup>

**Calculation of “P value”**

We needed to test whether the result was statistically significant or not. For calculation of “P value” from the “r” value, we had used the formula “=TDIST ((r\*SQRT (n-2))/(SQRT (1 - r^2)), n - 2,2)” for two-tailed “P value.” Calculation of “P” from “r” is not allowed in Excel®, so we had incorporated conversion of “r” to “t” in the Excel® formula according to the conversion formula:  $T = r(n - 2)/\sqrt{1 - r^2}$ <sup>[1]</sup>

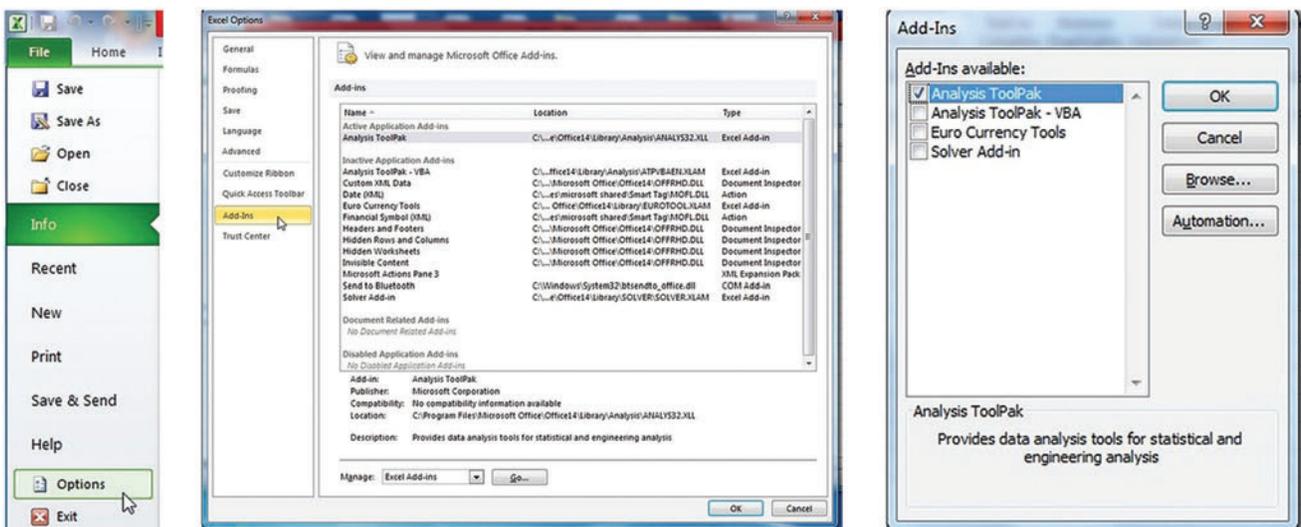
where *r* = correlation coefficient, *n* = number of participants, SQRT = square root function as written in Excel®, last 2 in formula was for 2 tail *P*.

In our example, *n* = 8 and *r* = 0.675689, we had not put the value of “r,” we had written the cell number of “r” instead. In the example of Figure 8, it was H2. We had written a cell text as “P value=,” and on the right adjacent cell, we had written the formula as “=TDIST((H2\*SQRT (6))/(SQRT (1-H2 ^ 2)),6,2).” Calculated two-tailed “P value” for the example data was 0.065879 in cell H4.

Solution for problems with negative “r” value: Where the “r” value is negative, we have to convert it to its absolute value of number (ignoring the negative sign) before using it in the calculation of *P* from *r*. To do this in any cell with ABS function, we have to write the formula for the cell as “=ABS (cell address of r).” This will show the absolute value in the selected cell.

H2		fx = =PEARSON(D2:D9,E2:E9)						
	A	B	C	D	E	F	G	H
1		Weight(kg)	Height(cm)	BMI	PEFR(L/min)			
2		54	173	18.0427	225.42		r =	0.675689
3		42.4	151	18.59568	224.5			
4		50.6	152.5	21.75759	239.7			
5		42	151	18.42024	221.2			
6		66.7	170	23.07958	238.77			
7		69.2	178.5	21.71849	240.92			
8		62	170	21.45329	192.68			
9		74.3	159.5	29.20569	267.09			
10	Mean =	57.65	163.1875	21.53416	231.285			
11	SD =	12.2588743	11.009533	3.622402	21.29438692			
12								

**Figure 6:** Calculation of Pearson’s correlation coefficient (*r*) of body mass index (column D) and peak expiratory flow rate (column E) data in cell H2 with the “PEARSON” function



**Figure 7:** Serial screenshots showing from left to right: After clicking “File,” “Options” to be clicked; in “Add-Ins,” “Analysis ToolPak” is to be selected and “Go” button to be clicked; “Analysis ToolPak” is ticked and “Ok” is to be clicked

Solution for scientific expression of number: When the value of “P” is far less than zero (frequently encountered), the number in Excel® is expressed in scientific format (e.g., 4.567E-23). To show it in decimal format, we have to right click the cell where the number is present and have to select “Format Cells...” Under “Number” tab, we have to select “Number” (below “General”) and increase the “Decimal places” until the number is shown after zeros in “Sample” and have to click “Ok.” If the cell width is automatically increased to accommodate the number, it is ok. Otherwise, increase the width manually to view the number.

### Regression analysis

In experimental science, after having understood the correlation between two variables, there are situations when it is necessary to do regression analysis.<sup>[1]</sup> For regression analysis, we had selected “Data→ Data Analysis Tools→Regression→Ok→ Input range and output range → Ok.”

In the example, we put “BMI” in “X range” and “PEFR” in “Y range.” For that, we had selected the “BMI” data column in “X range” as: “\$D\$2:\$D\$9” and “PEFR” data column in “Y range” as: “\$E\$2:\$E\$9” [Figure 9]. We could change the “Confidence Level,” but it was 95% by default, so we kept it. We selected “Output range” in cell I1. The regression analysis result is shown in Figure 10. The equation  $y = \alpha + \beta x$  is defined as the linear regression equation, where  $\alpha$  is the intercept and  $\beta$  is the regression coefficient.

In the example, we found:  $R^2 = 0.4566$ ; SE: 16.956;  $F: 5.041$ ; significance of  $F: 0.0659$ ; and the equation was “ $Y = 145.75 + 3.972 * X$ .”

### The Student’s t-test

The t-test is commonly used to determine whether the mean value of a continuous outcome variable in one group differs significantly from that in another group.<sup>[9]</sup> The t-test was not suitable for the study example we were discussing, but learning t-test procedure is a worth learning for any statistical analysis. Procedure is same as any “Data Analysis” we discussed above. We just need to select “t-test” from scrolling down after selecting “Data Analysis Tools” and input the range of data column in “Variable 1 Range” and “Variable 2 Range.” In the example shown in Figure 11, we wanted to carry out t-test for BMI of male (column A) and female (column B). We input “\$A\$2:\$A\$9” in “Variable 1 Range” and “\$B\$2:\$B\$9” in “Variable 2 Range.” We have selected “\$D\$3” as output range. Values for t-Stat and P have been highlighted with color in the result of t-test in Figure 11.

If we want to get only “P value” without other data analysis parameters, we can use the Excel® function “TTEST” and write the formula as “=TTEST (A2:A9, B2:B9,2,2)” as shown in the example provided in Figure 11. Here, the “P value” has been calculated in cell E1. See the formula; after the array 1 and array 2, the first “2” is for two-tailed test and the second “2” is for t-test assuming two samples of equal variance. We can change it according to our choice. If it is one tail and assuming two samples with unequal variance, the formula would be “=TTEST (A2:A9, B2:B9,1,3).”

### ANOVA

In the example shown in Figure 12, we have systolic blood pressure data of three groups of patients of underweight,

H4		fx =TDIST((H2*SQRT(6))/(SQRT(1-H2^2)),6,2)						
	A	B	C	D	E	F	G	H
1		Weight(kg)	Height(cm)	BMI	PEFR(L/min)			
2		54	173	18.0427	225.42		r=	0.675689
3		42.4	151	18.59568	224.5			
4		50.6	152.5	21.75759	239.7		p value=	0.065879
5		42	151	18.42024	221.2			
6		66.7	170	23.07958	238.77			
7		69.2	178.5	21.71849	240.92			
8		62	170	21.45329	192.68			
9		74.3	159.5	29.20569	267.09			
10	Mean =	57.65	163.1875	21.53416	231.285			
11	SD =	12.2588743	11.009533	3.622402	21.29438692			
12								

Figure 8: Calculation of two-tailed “P value” in cell H4 with the help of calculated value of “r” of cell H2 with the Excel® function “TDIST”

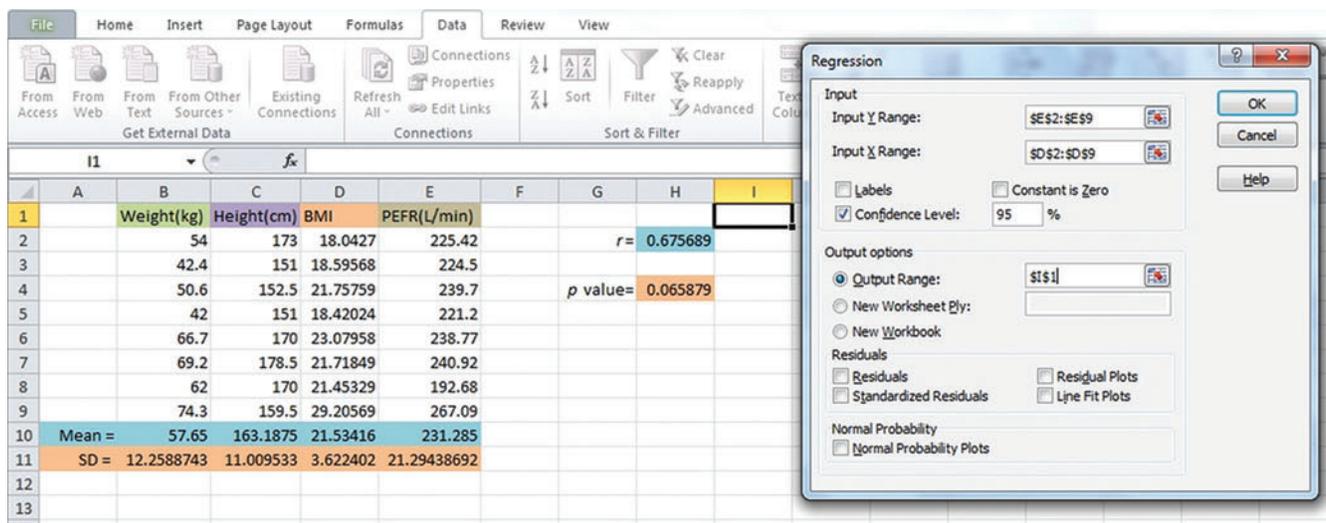


Figure 9: Screenshot during regression analysis of data of column D (Input Range X) and column E (Input Range Y) with 95% confidence interval. Output range is on cell I1

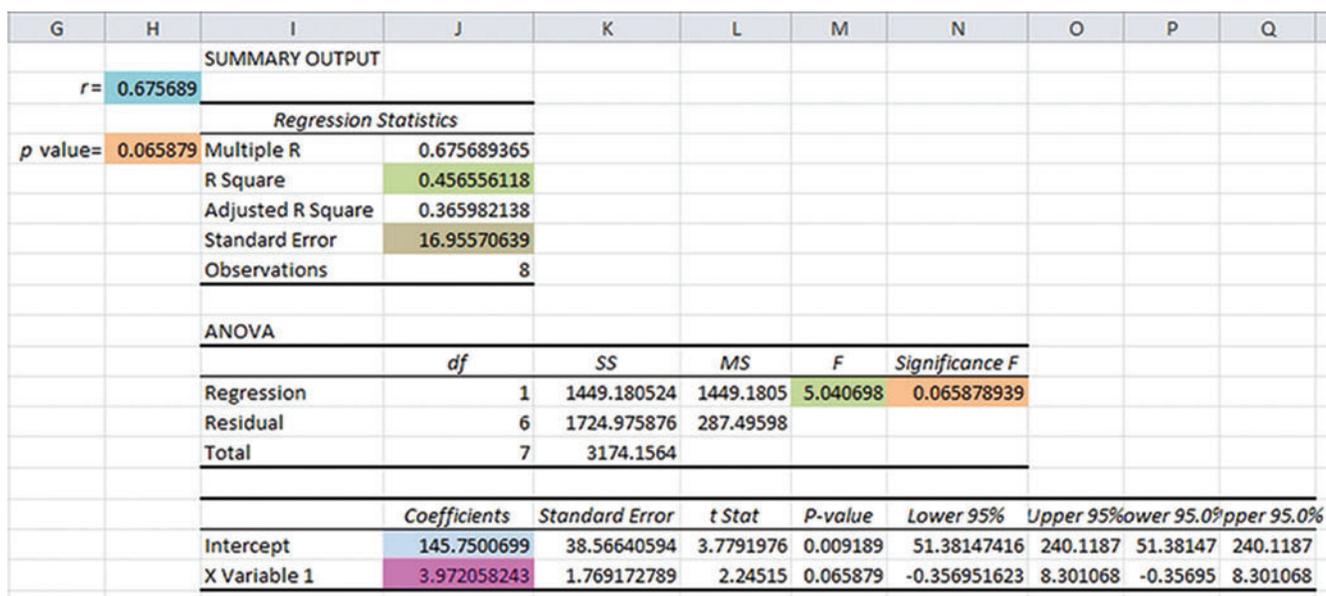


Figure 10: Summary output of regression analysis. Important parameters are colored. Equation:  $Y = 145.75 + 3.972 * X$  can be written from the data of J17 and J18

normal weight, and overweight. We have to use the Data Analysis Tools for ANOVA test. Here, we have selected the "ANOVA: Single Factor" and selected input range as "\$A:\$C" and output range as \$E\$1. As our first row contained labels, we had ticked the "Labels in the first row." Result is shown on the sheet as shown in Figure 12. This example is for one-way ANOVA. For two-way ANOVA, we have to select "ANOVA: Two-Factor with Replication." For two-way ANOVA, we have to tell Excel® about the "Rows per sample." Rest of the procedure is the same.

### Regression scatter plot

Graphical representation of regression analysis data gives a better understanding of the trend. In the example

of Figure 13, we have shown a scatter plot of regression analysis of column A (BMI) and column B (Fat%) data. For inserting the scatter plot on Excel® sheet, we have to click "Insert" tab on ribbon and in "Charts" section, "Scatter" to be clicked and "Scatter with only Markers" is to be selected. A blank graph plot will appear on the excel sheet. "Right click" on any area of the blank plot will give us option to "Select Data...." Clicking on "Select Data" will lead us to "Chart Data Range" selection. Column A and Column B have to be selected (Press and hold "Ctrl" and click on A and B column), and the "Chart Data Range" will appear as "=Sheet1!\$A:\$A, Sheet1!\$B:\$B." Alternatively, we can input the range as "=Sheet1!\$A\$1:\$B\$9" for the data in the example. Clicking on "Ok" will show the graph. Now,

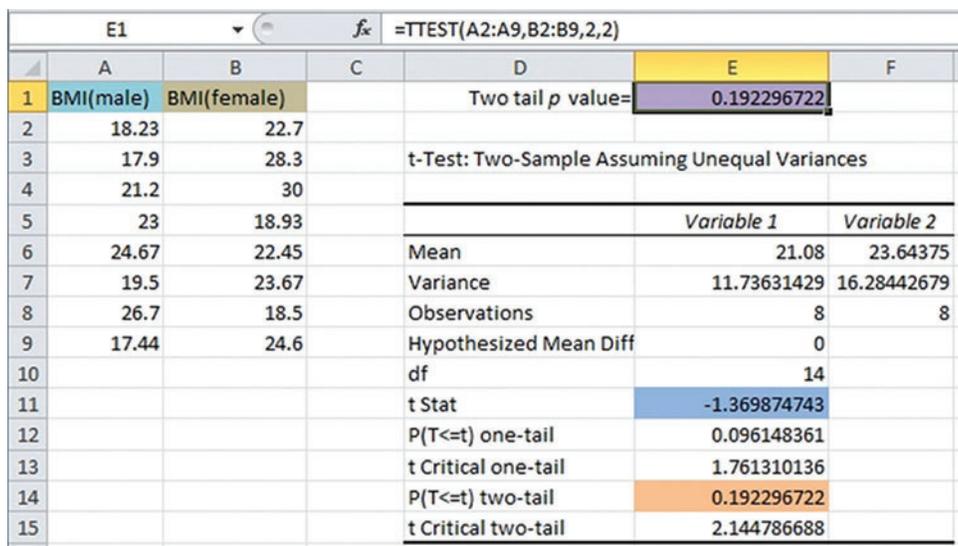


Figure 11: Screenshot of result of t-test of column A and column B by using Data Analysis Tool (output range is D3). Calculation of only two-tailed “P value” by Excel® function “TTEST” is shown in cell E1

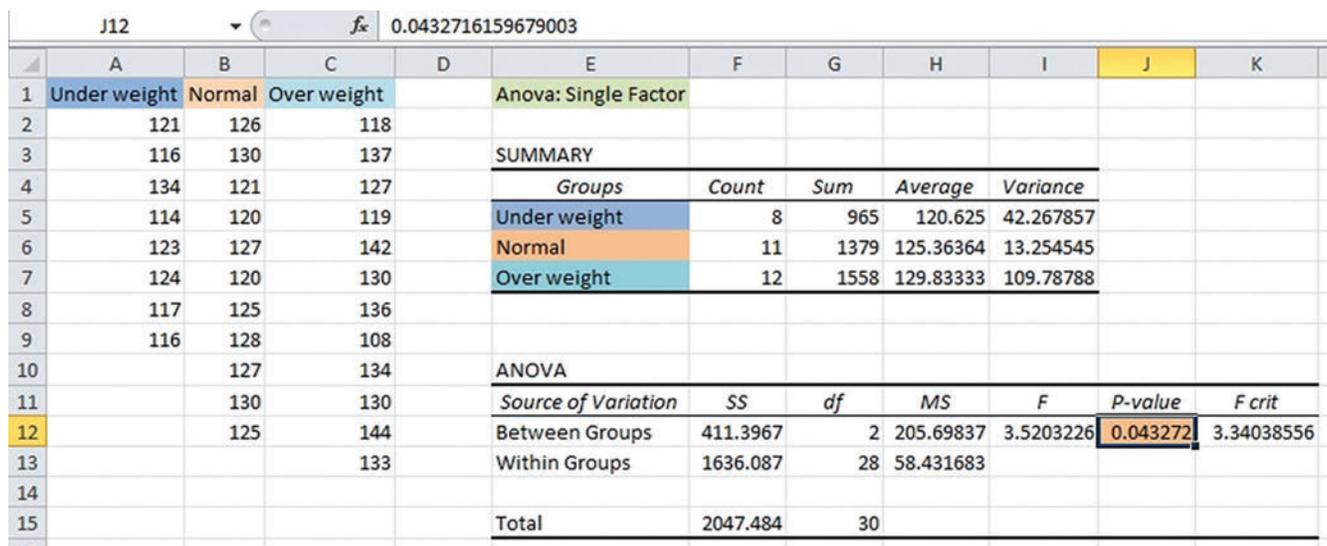


Figure 12: Analysis result of one-way ANOVA for the data of columns A, B, and C (output range is E1)

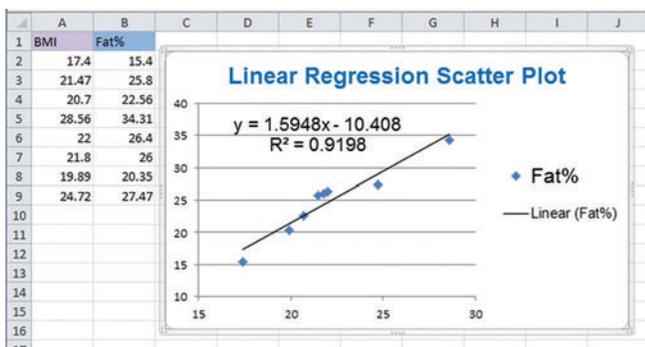


Figure 13: Regression scatter plot for column A and B data, body mass index is on X axis and Fat% is on Y axis

we have to right click on any series point (blue diamond) and select “Add Trend line...” and in “Trend line option,”

“Linear” is selected; “Display equation on chart” and “Display R-squared value on chart” are selected and the window is closed. Regression scatter plot with trend line will be shown with equation and  $R^2$  value. In this example of data [Figure 13], BMI is on X axis and Fat% is on Y axis. The linear relation equation is “ $Y = 1.5948 * X - 10.408$ .”

## DISCUSSION

Adequate sample size has paramount importance in assessing the accurate results, while less or more than required sample size has many disadvantages.<sup>[10]</sup> Sample size must be calculated during the planning stage of the study. Moreover, it should be done with reference to previous studies or a pilot study.<sup>[11]</sup> There are websites

which help in the calculation of sample size<sup>[12,13]</sup> free of cost. However, the commonly used software for statistical analysis such as IBM SPSS Statistics (New Orchard Road Armonk, New York, USA)<sup>[14]</sup> and GraphPad Prism (GraphPad Software, Inc., CA, USA)<sup>[15]</sup> are paid software. It may be difficult for researchers in small settings to get that software. On the other hand, Microsoft Office® is widely used software in educational institutions and offices. Excel® is a part of Microsoft Office® software package. Knowledge of Function, Formula, and Data Analysis Tools in Excel® is necessary for the analysis of data. We have discussed a simple example of a correlation study for better understanding of the procedure of a correlation study.

## CONCLUSION

This article would help new researchers to calculate the sample size for correlation study and statistical analysis without using any additional statistics software. They get hands-on experience on making sample size calculator, calculation and analysis of data in Excel®. Additional description of *t*-test and ANOVA would help them in data analysis of other studies. They can utilize this article as a reference tutorial for future studies.

## Financial support and sponsorship

Nil.

## Conflicts of interest

There are no conflicts of interest.

## REFERENCES

1. Khanal AB. Correlation and regression. In: Mahajan's Methods in Biostatistics for Medical Students and Research

- Workers. 8<sup>th</sup> ed. New Delhi, India: Jaypee Brothers Medical Publishers (P); 2016. p. 222-31.
2. Machin D, Campbell MJ, Walters SJ. Correlation, linear and logistic regression. In: Medical Statistics: A Textbook for the Health Sciences. 4<sup>th</sup> ed. England: John Wiley & Sons Ltd.; 2007. p. 150.
3. Hulley SB, Cummings SR, Browner WS, Grady DG, Newman TB. Getting ready to estimate sample size: Hypotheses and underlying principles. In: Designing Clinical Research. 3<sup>rd</sup> ed. Philadelphia, PA: Lippincott Williams and Wilkins; 2001. p. 55-9.
4. Kelley K, Preacher KJ. On effect size. *Psychol Methods* 2012;17:137-52.
5. Lwanga SK, Lemeshow S. Definitions of commonly used terms. In: Sample Size Determination in Health Studies: A Practical Manual. Geneva: World Health Organization; 1991. p. 21-2.
6. Zou GY. Sample size formulas for estimating intraclass correlation coefficients with precision and assurance. *Stat Med* 2012;31:3972-81.
7. Eknoyan G. Adolphe Quetelet (1796-1874) – The average man and indices of obesity. *Nephrol Dial Transplant* 2008;23:47-51.
8. McFedries P. Analyzing excel data. In: Teach Yourself Visually™ Excel® 2010. Indiana: Wiley Publishing Inc.; 2010. p. 242-3.
9. Gravetter FJ, Wallnau LB. Introduction to the t statistic. In: Essentials of Statistics for the Behavioral Sciences. 8<sup>th</sup> ed. USA: Cengage Learning; 2014. p. 252.
10. Kumar GS. Importance of sample size in clinical trials. *Int J Clin Exp Physiol* 2014;1:10-2.
11. Kim J, Seo BS. How to calculate sample size and why. *Clin Orthop Surg* 2013;5:235-42.
12. California: UCSF Clinical & Translational Science Institute. Available from: <http://www.sample-size.net/correlation-sample-size/>. [Last updated on 2016 Jun 02; Last cited on 2016 Sep 12].
13. Hong Kong: Centre for Clinical Research and Biostatistics, the Chinese University of Hong Kong. Available from: <https://www2.ccrb.cuhk.edu.hk/stat/other/correlation.htm>. [Last cited on 2016 Sep 12].
14. IBM SPSS Statistics for Windows, Version 22.0. Armonk, NY: IBM Corp.; 2013 Release. Available from: <http://www.ibm.com/analytics/us/en/technology/spss>. [Last cited on 2016 Sep 13].
15. GraphPad Prism Version 6.00 for Windows. La Jolla California USA: GraphPad Software. Available from: <http://www.graphpad.com>. [Last cited on 2016 Sep 13].