SOFTWARE APPLICATIONS FOR PEARSON CORRELATIONS ANALYSIS

H. VÎRGOLICI[#], BOGDANA VÎRGOLICI, MARIA MOHORA, V.L. PURCĂREA

"Carol Davila" University of Medicine and Pharmacy,
8 Eroilor Sanitari blvd., Bucharest, România, [#]e-mail: hvirgolici@gmail.com

Abstract. In order to facilitate the calculations when analysed data are in an Excel data base, we propose two applications. The first application calculates at once Pearson correlations between many contiguous columns in a data table and shows the significantly correlated parameters with a chosen parameter. The second application is useful when the user has already the data containing Pearson correlations and wants to see, in a friendly way, what parameters are significantly correlated with a chosen parameter. The applications are developed in Visual Basic for Applications.

Key words: Pearson correlations, Excel, VBA.

INTRODUCTION

In medicine it is important to know how two measurable variables taken from the same group of subjects relate to each other, in effect how they co-vary. Pearson's correlation coefficient (r) is used to quantify the direction and strength of the linear relationship between two variables.

Our aim is to offer two useful tools for physicians and clinician researchers in order to calculate with a click Pearson correlations between as many as possible contiguous columns in a data table and to see (in a list) what are the parameters (columns) correlated with a chosen parameter, respectively.

MATERIALS AND METHODS

We made two macros in Excel using VBA (Visual Basic for Applications). In order to use the first application (which calculates Pearson correlations and shows the correlated parameters with a chosen parameter) the user has to import (or to introduce) in a macro-enabled Excel book the data containing all the parameters

Received: June 2015;
in final form September 2015.

ROMANIAN J. BIOPHYS., Vol. 25, No. 3, P. 201–208, BUCHAREST, 2015

values for all the patients. In order to use the second application (which shows the correlated parameters with a chosen parameter), the user has to import (or to introduce) in a macro-enabled Excel book the data table containing only Pearson correlations.

RESULTS

FIRST APPLICATION

Let's suppose we have a data table with blood markers for 102 patients (Table 1).

Table 1

Blood markers

	A	B	C	D	E	F	G	H	I	J	K	L
1	ID	Glycemia	Urea	Uric acid	eFGR	Creatinine	ALT	AST	Cholesterol	Triglycerides	HDL-C	LDL-C
2	M.R.	93	34,1	1,6	101,73	0,59	27	38	243	41	110	
3	P.D.	100	35,2	3,6	74,42	0,95	19	17	210	56	70	128
4	U.G.	100	32,7	3,7	101,93	0,6	41	33	235	60	65	157
5	C.M.	81	27,7	4	83,05	0,8	32	26	200	61	110	
6	D.S.	251	37,9	2,9	92,1	0,73	27	18	231	65	82	136
...												
100	H.MF.	99	32,2	5,7	97,36	0,9	41	31	197	334	42	88
101	T.AM	96	30,6	6,2	65,53	1,01	48	24	265	351	31	164
102	G.I.	119	37,4	6,3	66,51	0,93	40	40	318	525	33	
103	C.C.	197	29,8	4,9	83,99	0,81	35	23	271	525	36	

The user has to introduce the row and the column of the upper left cell (containing the name of each parameter) and the row and the column of the bottom right cell of the desired domain of cells. An example is given in Fig. 1, suitable for data shown in Table 1.

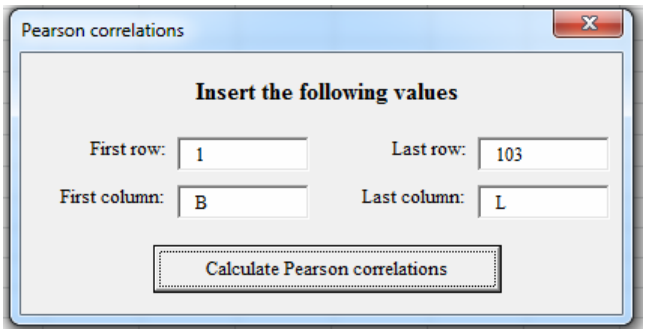

Fig. 1. Required values.

If a required value is missing or is introduced in a wrong format, a message box is displayed (an example is given in Fig. 2).

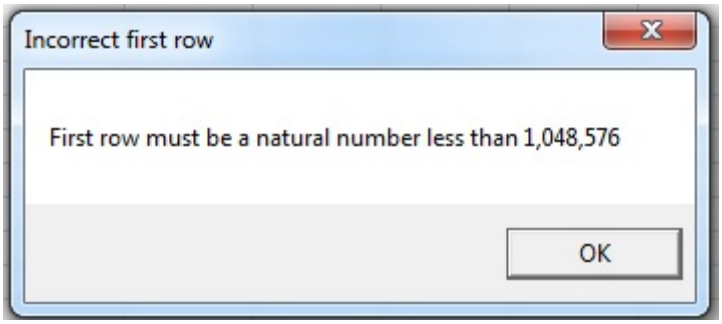

Fig. 2. A message box indicating a wrong value (for the first row).

By clicking ***Calculate Pearson correlations*** button, the correlations between the parameters are calculated and displayed (by default) three rows below the initial data table, spreading from the column of the upper left cell of the domain to the column of the right bottom cell of the domain. The names of parameters involved are shown above and also in the rightmost column, formatted in a different way (Table 2).

Table 2

Pearson correlations

	A	B	C	D	E	F	G	H	I	J	K	L	M
103	C.C.	197	29,8	4,9	83,99	0,81	35	23	271	525	36		
104													
105													
106		**Glycemia**	**Urea**	**Uric acid**	**eFGR**	**Creatinine**	**ALT**	**AST**	**Cholesterol**	**Triglycerides**	**HDL-C**	**LDL-C**	
107		1	0,049	-0,043643	-0,0641	0,0044217	0,2	0,1	-0,163544	0,308030818	-0,233	-0,24	**Glycemia**
108		0,049386	1	0,2424476	-0,45445	0,4019085	-0,1	0	-0,0188066	-0,03144687	-0,108	0,034	**Urea**
109		-0,04364	0,242	1	-0,56441	0,479351	0	-0	-0,1218461	0,281498598	-0,421	-0,11	**Uric acid**
110		-0,0641	-0,454	-0,564415	1	-0,8880361	0,1	0	0,14649087	-0,02732853	0,1865	0,103	**eFGR**
111		0,004422	0,402	0,479351	-0,88804	1	-0	0	-0,2462777	0,022808309	-0,187	-0,207	**Creatinine**
112		0,174058	-0,116	0,0241561	0,117753	-0,0432315	1	0,7	-0,0574674	0,140648387	-0,288	0,003	**ALT**
113		0,14456	0,048	-0,01621	0,027995	0,0084573	0,7	1	-0,0110288	0,102302481	-0,112	-0,023	**AST**
114		-0,16354	-0,019	-0,121846	0,146491	-0,2462777	-0,1	-0	1	0,100155984	0,2641	0,923	**Cholesterol**
115		0,308031	-0,031	0,2814986	-0,02733	0,0228083	0,1	0,1	0,10015598	1	-0,536	-0,271	**Triglycerides**
116		-0,23299	-0,108	-0,421093	0,18646	-0,1872788	-0,3	-0,1	0,26412802	-0,53649373	1	0,18	**HDL-C**
117		-0,23997	0,034	-0,109689	0,10263	-0,2066436	0	-0	0,92265445	-0,27112952	0,1801	1	**LDL-C**

If one or more columns of the original domain of cells contain(s) no values or data are in a wrong format, the corresponding cells containing Pearson correlations will contain an error message (for example *#DIV/0!*).

The user is next asked to choose a parameter from the parameters list (created by application from the original data table and shown in a combo list), to choose the desired type of correlations (positive or negative or both) and then to push ***Show correlated parameters*** button. The program will compute r_{min}, based on number of rows in the domain of cells, and will display in a text box, one parameter in a row, the parameters which are significantly correlated with the chosen parameter (*i.e.* the parameters which have $|r| \geq r_{min}$). An example is given in Figure 3.

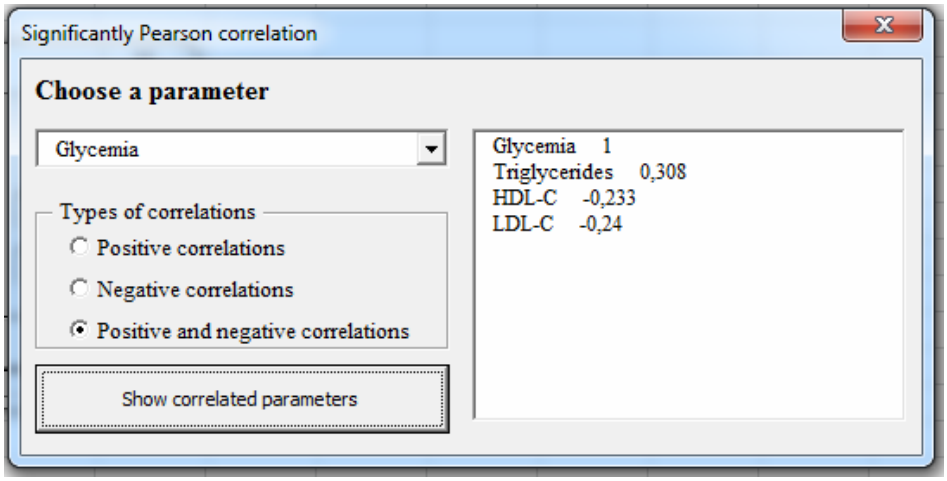

Fig. 3. The list of parameters which are positive and negative correlated with Glycemia.

THE SECOND APPLICATION

This application is useful when data containing Pearson correlations already exist in a macro-enabled Excel book. The user is asked, as a beginning step, to introduce the number of rows (observations, on which Pearson correlations were calculated), the row containing the names of the parameters, the first column (containing the data for the first parameter) and the last column (containing the data for the last parameter) and then to push ***Validate*** button. An example is given in Fig. 4, which is suitable for data containing Pearson correlations shown in Table 2.

If the information is correct, the user can choose a parameter from the combo list, the type of correlation and then push ***Show correlated parameters*** button. An example is shown in Fig. 6. (We observe that the significantly correlated parameters with Glycemia are the same as those shown in Fig. 3, which is obvious!).

If the information is not correct, a message box is displayed (an example is shown in Fig. 5).

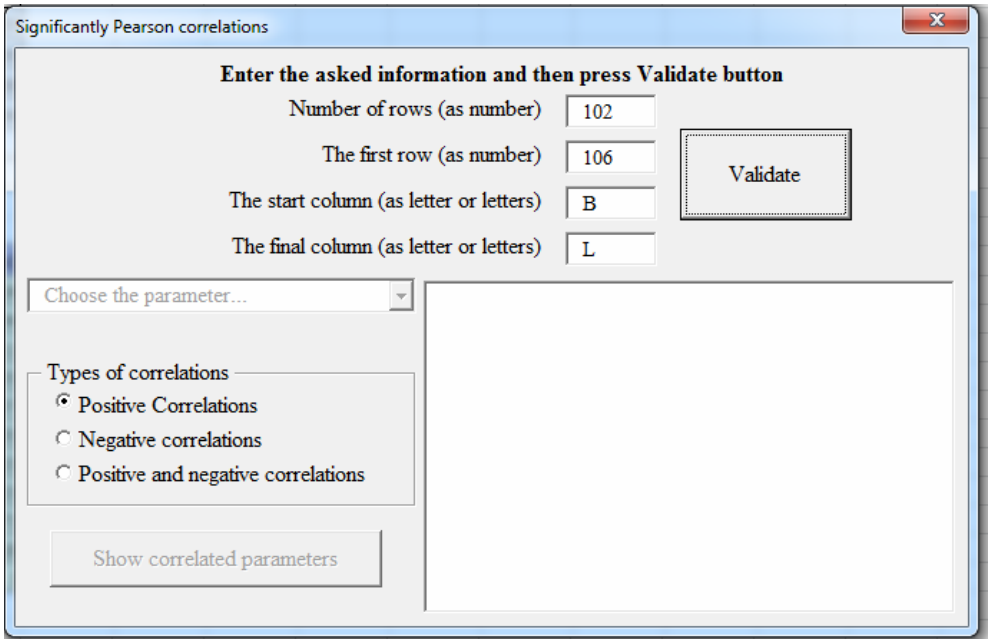

Fig. 4. Significantly Pearson correlations window.

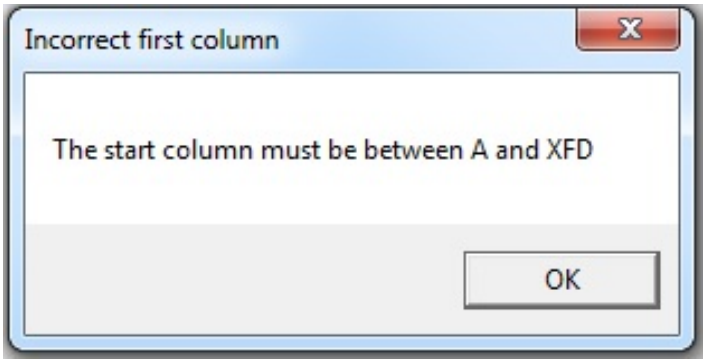

Fig. 5. A message box indicating that the first (start) column is incorrect.

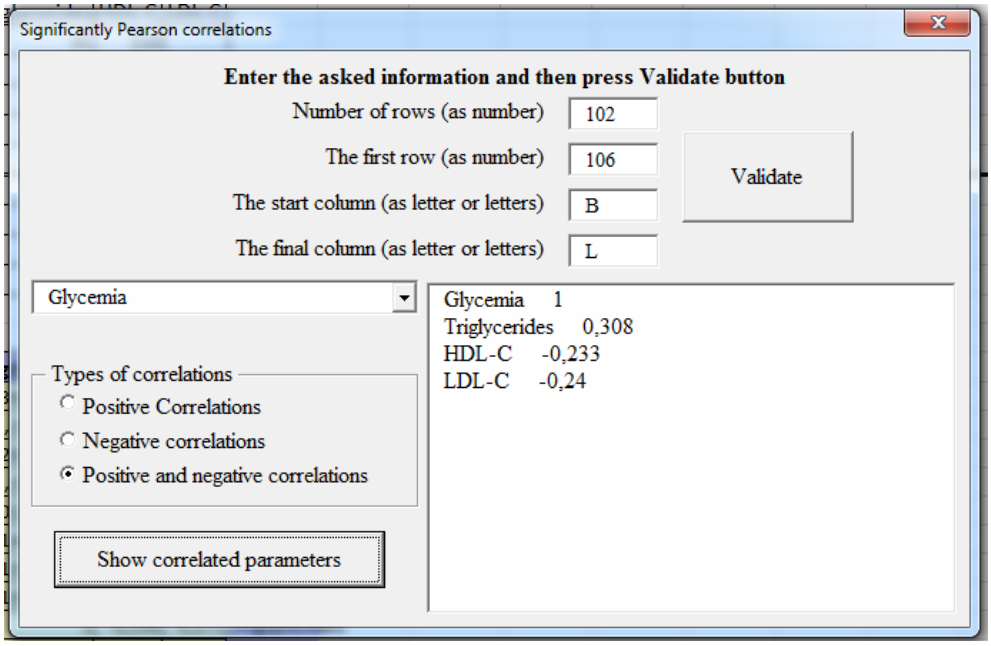

Fig. 6. The significantly positive and negative correlated parameters with *Glycemia.*

DISCUSSION AND CONCLUSIONS

The results are very useful and help physicians to think the complications of a disease and save money for the future blood tests of the patient. These applications can be used in clinical trials on a large number of patients for whom we record a large number of parameters. The results are obtained almost instantly. These applications can help validate some medical correlations, which could be standardized in the future. The cost for medical analyses could be reduced. On PubMed, there are many articles focused on correlations between different parameters. For example, our research team has often used these applications for medical studies [1–4]. We also recommend these applications for any field of research.

REFERENCES

1. BACĂNU, E.V., D. LIXANDRU, I. STOIAN, B. VÎRGOLICI, M. MOHORA, N. MITREA, C. IONESCU TÎRGOVI□TE, Correlations between obesity anthropometric markers, adipocytokines and monocytes oxidative stress status in type 2 diabetic patients, *Farmacia*, 2012, **60**(1), 21–32.

2. CASARIU, E.D., B. VÎRGOLICI, D. LIXANDRU, M. STANCU, M. GREABU, A. TOTAN, D. MIRICESCU, M. MOHORA, Relations for APO B/APO A-I ratio with different cardiovascular risk factors, in obese children, *Romanian J. Biophys.*, 2012, **22**(2), 83–94.
3. LIXANDRU, D., E.V. BĂCANU, B. VÎRGOLICI, H. VÎRGOLICI, A. PETRUȚA, M.E. BĂCANU, P. GAGNIUC, C. IONESCU-TÎRGOVIŞTE, C. SERAFINCEANU, Changes in the serum proinflammatory cytokines in patients with elevated homa-ir and type 2 diabetes mellitus, *Farmacia*, 2015, **63**(1), 132–139.
4. VÎRGOLICI, B., M. MOHORA, V. RĂDOI, D. LIXANDRU, I. STOIAN, L. GAMAN, A. COMAN, B.Y. MANUEL, Correlations between dysglycemia and markers of oxidative stress and inflammation in diabetic foot patients, *Farmacia*, 2011, **59**(2), 216–226.
5. WALKENBACH, J, *Excel 2010 Power Programming with VBA*, Wiley Publishing, Inc., Indianapolis, Indiana, 2010.