World Applied Sciences Journal 31 (6): 1168-1172, 2014
ISSN 1818-4952
© IDOSI Publications, 2014
DOI: 10.5829/idosi.wasj.2014.31.06.14346

Statistics and Econometrics from the Point of View Methodology Mathematics

Tatjana Vasiljevna Kapustina, Aleksandr Vasilevich Popyrin and Lyubov Nikolaevna Savina

Elabuga Institute of Kazan (Volga Region) Federal University, Kazanskaya st.89,
Elabuga 423600, Russia

Submitted: Jan 17, 2014; **Accepted:** Mar 6, 2014; **Published:** Mar 23, 2014

Abstract: The authors consider teaching statistics and econometrics with the use of mathematical methods. The advantages of step-by-step solution of problems and carrying out of laboratory works with the use of computer technologies (EXCEL, Mathematica) are discussed. The example of 2 variants of linear (4-factor) multiple regression in EXCEL and in Mathematica (with the use of single line programs in functional style) is given. Apart from professional program software (EViews, etc.) which give only summary table problem solution performed step-by-step can be used in teaching.

Key words: Statistics · Econometrics · Multiple regression · Heteroscedasticity

INTRODUCTION

All over the world statistics and econometrics belong to the number of basic disciplines in modern economics. Statistics as a discipline appeared together with formation of economics departments. Apart from it econometrics is relatively new discipline for future economists. Econometrics is integrity of 3 components: statistics, economic theory and mathematics.

Term econometrics was introduced by R. Frish for the first time in 1926, it was the name of the journal issued by him. The subject of econometrics is quantitative analysis of real economic phenomena for which mathematics is used: linear algebra, elements of mathematical analysis, probability theory and mathematical statistics. Methods of teaching of econometrics are only being formed. Existing textbooks on econometrics (about 40 in the world) are very diversified by style and contents. Books of problems for solutions are even less numerous. And there are no generally accepted guide-books of laboratory works.

Both in statistics and economics for correct solution of problems it is necessary to inculcate the learners with the following skills: firstly, ability to choose quickly necessary formula from a set of formulas and substitute data in this formula in appropriate way, secondly, to make correct economic conclusions and forecasts.

Since solution of specific econometric and statistical problems is related to cumbersome calculations the use of computer programs is imperative. And here almost anarchical situation is observed: some people use spreadsheets (Excel), the others prefer professional computer packages on applied economics (Statistica, Econometric Views and others) and almost everywhere in teaching of econometrics computer mathematics system are used. Since disciplines of statistics and econometrics are interchangeable to a great extent we found it appropriate to use Excel for statistics and Mathematica for econometrics. The purpose of this article is an attempt to analyze methodical-mathematical aspect of this problem.

In the course of mathematical statistics where abstract look at observed phenomenon and processes is needed and only numbers must be in focus, students learn simplified methods of calculations, for example, by introduction of conventional zero calculation of mathematical expectation and variance, primary and central moments can be simplified. Then on the base of these characteristics obtained numbers can be easily transformed into necessary ones. Let us consider different techniques of work with data in Excel through the example of multiple regression which is broadly used in solution of problems of demand, profitability of shares, in studies of production costs functions, in micro-economic calculations etc. The main purpose of multiple regression

Corresponding Author: Popyrin, Elabuga Institute of Kazan (Volga Region) Federal University, Kazanskaya st.89, Elabuga 423600, Russia.

is to build a model with big number of factors establishing the separate influence of each of them and aggregate influence on the modeled indicator. For teaching purposes multiple regression is taught through example of 2-factor model as it is practiced in textbooks [1, 2]. It is explained partly by the uniformity of texts which go from book to book since those times when there were no computers and all calculations must be done by hand. On the other hand, it is necessary, maybe, for learner to master the transition from 2 to 4 factors [3]. Optimizing the process of learning we found time to do laboratory work on this topic. We warn you that if using computer (Excel) for solution of this problem you can choose different ways of calculations - from detailed, step-by-step ones which clarify the contents and meaning of the formula and facilitate its memorizing to the those which give ready results. You also can use built-in tools and statistical functions, such as REGRESSION, CORRELATION, DESCRIPTIVE STATISTICS. The use of the latter allows to display all calculated characteristics and residual plots on the screen.

Laboratory Work on Statistics Multiple Regression:

Table 1: Observations data

#	y	x1	x2	x3	x4
1	40	25	3	19	8
2	78	23	7	15	8
3	57	20	7	16	6
4	73	14	5	14	13
5	111	28	10	21	13
6	75	21	8	20	12
7	113	25	6	12	16
8	59	22	5	18	9
9	64	20	9	20	9
10	41	12	3	13	9
11	53	18	3	13	8
12	47	18	5	14	5
13	52	19	7	20	8
14	62	22	8	14	4
15	95	19	8	7	9
16	73	14	4	10	13
17	96	29	7	19	11
18	43	18	4	15	8
19	63	18	7	14	5
20	63	22	7	19	9
21	81	27	7	14	7
22	48	21	6	19	7
23	62	16	8	17	8
24	56	23	1	12	9
25	44	16	4	17	9
26	77	28	4	22	15
27	43	20	2	11	5
28	75	24	5	19	12
29	87	25	6	11	10
30	61	23	4	12	6

The task: to calculate coefficients of linear equation of multiple regression, paired linear coefficients and correlations and build correlation matrix.

The equation of linear 4-factor regression has the following form:

$$\hat{y}_{x_1x_2x_3x_4} = a_0 + a_1x_1 + a_2x_2 + a_3x_3 + a_4x_4$$

For calculation of coefficients by least squares method we shall use the following simultaneous equations: [2]

$$\begin{cases} a_0 + a_1\overline{x_1} + a_2\overline{x_2} + a_3\overline{x_3} + a_4\overline{x_4} = \overline{y} \\ a_0\overline{x_1} + a_1\overline{x_1^2} + a_2\overline{x_1x_2} + a_3\overline{x_1x_3} + a_4\overline{x_1x_4} = \overline{yx_1} \\ a_0\overline{x_2} + a_1\overline{x_1x_2} + a_2\overline{x_2^2} + a_3\overline{x_2x_3} + a_4\overline{x_2x_4} = \overline{yx_2} \\ a_0\overline{x_3} + a_1\overline{x_1x_3} + a_2\overline{x_2x_3} + a_3\overline{x_3^2} + a_4\overline{x_3x_4} = \overline{yx_3} \\ a_0\overline{x_4} + a_1\overline{x_1x_4} + a_2\overline{x_2x_4} + a_3\overline{x_3x_4} + a_4\overline{x_4^2} = \overline{yx_4} \end{cases}$$

First we need to calculate all the coefficients in Excel. Then we must enter data into calculation table column by column. Then we calculate the following indicators:

$$yx_1, yx_2, yx_3, yx_4, x_1x_2, x_1x_3, x_1x_4, x_2x_3, x_2x_4, x_3x_4, x_1^2, x_2^2, x_3^2, x_4^2$$

We shall calculate averages which are coefficients of considered simultaneous equations by the tool COMPARISON OF VALUES, or using formulas:

$$\overline{y} = \frac{\sum y_i}{n}, \overline{x_1} = \frac{\sum x_i}{n}, \ldots$$

The results will be put into the following order:

1	$\overline{x_1}$	$\overline{x_2}$	$\overline{x_3}$	$\overline{x_4}$	$\overline{y}$
$\overline{x_1}$	$\overline{x_1^2}$	$\overline{x_1x_2}$	$\overline{x_1x_3}$	$\overline{x_1x_4}$	$\overline{yx_1}$
$\overline{x_2}$	$\overline{x_1x_2}$	$\overline{x_2^2}$	$\overline{x_2x_3}$	$\overline{x_2x_4}$	$\overline{yx_2}$
$\overline{x_3}$	$\overline{x_1x_3}$	$\overline{x_2x_3}$	$\overline{x_3^2}$	$\overline{x_3x_4}$	$\overline{yx_3}$
$\overline{x_4}$	$\overline{x_1x_4}$	$\overline{x_2x_4}$	$\overline{x_3x_4}$	$\overline{x_4^2}$	$\overline{yx_4}$

In Our Case Excel Page Displays the Following: Let us calculate the matrix which is reversed to system matrix using mathematic tool REVERSED MATRIX, filling up the arguments with first 5 columns of this table, having determined output massive as 5x5:

Then we shall calculate coefficients of equation of linear regression using function MATRIX MULTIPLICATION, filling up the arguments of function of reversed matrix (Table 3) and the last column of the Table 3 (Column of free terms):

This solution is obtained with the use of reversed matrix. These simultaneous equations can be solved by the Kramer's rule after calculation of 6 determinants with the use of function MATRIX DETERMINANT using the necessary columns from Table 2 for argument of the function:

Table 2: System's coefficient

1	21,00	5,67	15,57	9,03	66,40
21,00	459,20	121,00	332,47	192,57	1438,80
5,67	121,00	36,67	90,77	51,47	398,73
15,57	332,47	90,77	255,97	143,00	1030,83
9,03	192,57	51,47	143,00	90,30	635,87

Table 3: Reversed matrix.

35,86	-0,82	-0,51	-0,61	-0,59
-0,82	0,07	-0,02	-0,02	-0,02
-0,51	-0,02	0,25	-0,04	0,01
-0,61	-0,02	-0,04	0,09	-0,02
-0,59	-0,02	0,01	-0,02	0,12

Table 4: Coefficients of regression equation

a0	-0,77
a1	2,06
a2	5,32
a3	-2,75
a4	4,05

$$det = \begin{vmatrix} 1 & 21{,}00 & 5{,}67 & 15{,}57 & 9{,}03 \\ 21{,}00 & 459{,}20 & 121{,}00 & 332{,}47 & 192{,}57 \\ 5{,}67 & 121{,}00 & 36{,}67 & 90{,}77 & 51{,}47 \\ 15{,}57 & 332{,}47 & 90{,}77 & 51{,}47 & 143{,}00 \\ 9{,}03 & 192{,}57 & 51{,}47 & 143{,}00 & 90{,}30 \end{vmatrix} = 7022{,}57$$

Accordingly,

$$det1 = 5415{,}77, det2 = 14446{,}95, det3 = 37379{,}93, det4 = -19293, det5 = 28432{,}12$$

Continuing the process we shall get the same values of coefficients (Table 4).

We shall calculate the coefficients of paired regression and position them in the form of matrix form instead of corresponding markers [2]. Matrix of paired coefficients correlation of variables can be calculated with the use of CORRELATION tool from Data analysis.

Table 5: Example of correlation matrix

	y	x1	x2	x3	x4
y	1	r_{yx1}	r_{yx2}	r_{yx3}	r_{yx4}
x1	r_{x1y}	1	r_{x1x2}	r_{x1x3}	r_{x1x4}
x2	r_{x2y}	r_{x2x1}	1	r_{x2x3}	r_{x2x4}
x3	r_{x3y}	r_{x3x1}	r_{x3x2}	1	r_{x3x4}
x4	r_{x4y}	r_{x4x1}	r_{x4x2}	r_{x4x3}	1

Table 6: Correlation matrix

	y	x1	x2	x3	x4
y	1,00	0,53	0,54	-0,04	0,63
x1	0,53	1,00	0,22	0,35	0,23
x2	0,54	0,22	1,00	0,32	0,04
x3	-0,04	0,35	0,32	1,00	0,22
x4	0,63	0,23	0,04	0,22	1,00

Now we shall calculate the coefficients of linear regression equation with the use of function LINEAR, which allows to print regression statistics in the following order:

Value a_4	Value a_3	Value a_2	Value a_1	Value a_0
Mean root square deviation a_4	Mean root square deviation a_3	Mean root square deviation a_2	Mean root square deviation a_1	Mean root square deviation a_0
Determination coefficient R^2	Mean root square deviation y			
F-statistics	Number of degrees of freedom			
Regressive sum of squares	Residual sum of squares			

Table 7: Results of LINEAR function

4,05	-2,75	5,32	2,06	-0,77
0,27	0,23	0,38	0,19	4,51
0,96	4,13	#Н/Д	#Н/Д	#Н/Д
160,42	25,00	#Н/Д	#Н/Д	#Н/Д
10937,08	426,12	#Н/Д	#Н/Д	#Н/Д

We see that coefficients of linear regression are the same.

Finally we shall calculate (using tool REGRESSION) the results of regression statistics, variance analysis and intervals of confidence, we shall get residuals and the graphs for fitting of regression lines, residuals and normal probability entering the ranges of cells which contain data of resultative attribute Y and factorial attributes: X_1, X_2, X_3, X_4 placing necessary flags in dialogue window and marking left top cell of future massive as output interval.

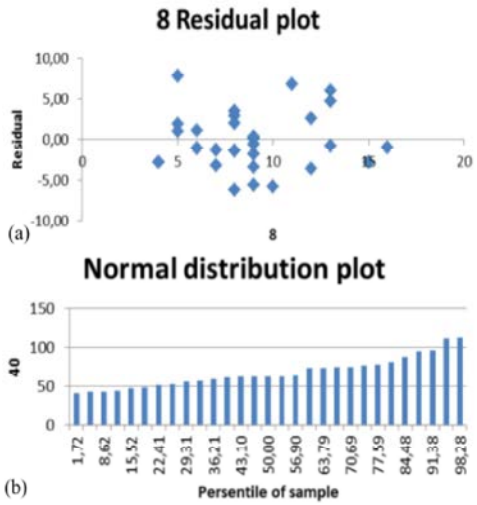


Fig. 1: Results of regression analysis

Our experience of laboratory works on econometrics shows that best choice is mathematics systems (Mathematica or Maple). Of course it is necessary to teach students to use professional computer economic packages of Econometric Views type because they will need them in their future professional activity. But the main task of every laboratory class is deliberate (reasoned) use of theoretical material [4], which is not possible in Econometric Views because this package is designed in such a way that you must enter original data and get the answer at once in the form of table containing all needed characteristics. The way of their calculation is not clear, calculations are deep inside the program.

Quite different situation is observed in Mathematica environment. Interactivity of this program allows to perform work step by step and see the result of every step, to understand every entry (formulas in cells have the same form as in traditional mathematical notation), if necessary you can get graph illustration of regression models – in 2D and 3D form. Of course, plans of laboratory works (typical calculations and commentary) must be prepared by teachers beforehand.

Let us consider the fragment of e-version of one of laboratory works. Entry cells will be printed on the screen with bold type, output cells - with lighter and smaller type. In the beginning of work the table of observations is given (it is skipped because all data will be entered) and the task is formulated: to build the model of multiple linear regression (test the model for heteroscedasticity [5-8], in case of its availability to correct the model, test it one more time and make sure that heteroscedasticity is eliminated.

Topic: Finding and elimination of heteroscedasticity.

We need to set the sample size *n* and the number *k* of explaining variables:

{n=30, k=4}
{30, 4}

Form sample for X_1 (enter observation data; N denotes approximate meanings of entered values [9, 10]):
x1=N[{25,23,20,14,28,21,25,22,20,12,18,18,19,22,19,14,29,18,18,22,27,21,16,23,16,28,20,24,25,23}]
{25.23.20.14.28.21.25.22.20.12.18.18.19.22.19.14.29.18.18.22.27.21.16.23.16.28.
20.24.25.23.}

Form samples for X_2, X_3 и X_4 :
x2=N[{3,7,7,5,10,8,6,5,9,3,3,5,7,8,8,4,7,4,7,7,7,6,8,1,4,4,2,5,
6,4}]
{3.7.7.5.10.8.6.5.9.3.3.5.7.8.8.4.7.4.7.7.7.6.8.1.4.4.2.5.6.4.}
x3=N[{19,15,16,14,21,20,12,18,20,13,13,14,20,14,7,10,19,15
,14,19,14,19,17,12,17,
22,11,19,11,12}]
{19.15.16.14.21.20.12.18.20.13.13.14.20.14.7.10.19.15.14.19
.14.19.17.12.17.22.
11.19.11.12.}
x4=N[{8,8,6,13,13,12,16,9,9,9,8,5,8,4,9,13,11,8,5,9,7,7,8,9,9,
15,5,12,10,6}]
{8.8.6.13.13.12.16.9.9.9.8.5.8.4.9.13.11.8.5.9.7.7.8.9.9.15.5.1
2.10.6.}

Form sample for *Y*:
y=N[{40,78,57,73,111,75,113,59,64,41,53,47,52,62,95,73,96,
43,63,63,81,48,62,56,
44,77,43,75,87,61}]
{40.78.57.73.111.75.113.59.64.41.53.47.52.62.95.73.96.43.6
3.63.81.48.62.56.44.
77.43.75.87.61.}
Enter vector *i*:
i=Table[1.{n}]
{1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.}

Regression equation in matrix form: $Y=Xb+\epsilon$, ãÏå X — matrix from columns i, x_1, x_2, x_3, x_4:
X=Transpose[{i,x1,x2,x3,x4}]
{{1.25.3.19.8.},{1.23.7.15.8.},{1.20.7.16.6.},{1.14.5.14.13.}
,{1.28.10.21.13.},{1.21.8.20.12.},{1.25.6.12.16.},{1.22.5.18.
9.},{1.20.9.20.9.},{1.12.3.13.9.},{1.18.3.13.8.},
{1.18.5.14.5.},{1.19.7.20.8.},{1.22.8.14.4.},{1.19.8.7.9.},{1.
14.4.10.13.},{1.29.7.
19.11.},{1.18.4.15.8.},{1.18.7.14.5.},{1.22.7.19.9.},{1.27.7.
14.7.},{1.21.6.19.7.},{1.16.8.17.8.},{1.23.1.12.9.},{1.16.4.1
7.9.},{1.28.4.22.15.},{1.20.2.11.5.},{1.24.5.19.
12.},{1.25.6.11.10.},{1.23.4.12.6.}}

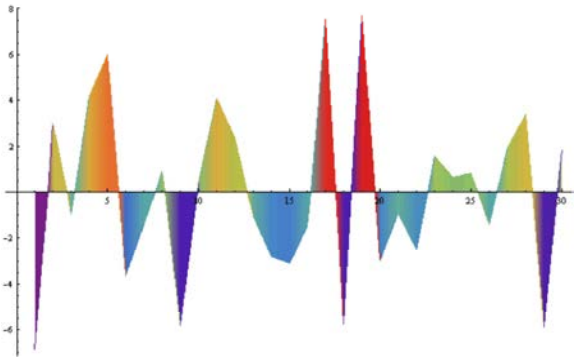
Assume that *Xt* is transport matrix *X'*, we shall get:

Xt={i,x1,x2,x3,x4}
{{1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.1.},{
25.23.20.14.28.
21.25.22.20.12.18.18.19.22.19.14.29.18.18.22.27.21.16.23.16
.28.20.24.25.23.},
{3.7.7.5.10.8.6.5.9.3.3.5.7.8.8.4.7.4.7.7.7.6.8.1.4.4.2.5.6.4.},
{19.15.16.14.21.

20.12.18.20.13.13.14.20.14.7.10.19.15.14.19.14.19.17.12.17.
22.11.19.11.12.},{8.8.6.13.13.12.16.9.9.9.8.5.8.4.9.13.11.8.5.
9.7.7.8.9.9.15.5.12.10.6.}}

Calculate the parameters of regression
(vector β=$(X'(X))^{-1}$ X' Y):
β= Inverse[Xt.X].(Xt.y)
{-0.771196,2.05722,5.32283,-2.74729,4.04868}
{β_0=Part[β,1], β_1=Part[β,2], β_2=Part[β,3], β_3=Part[β,4], β_4=Part[β,5]}
{-0.771196, 2.05722, 5.32283, 2.74729, 4.04868}
The formula for predicted value y (we use postponed acquisition, there will be no output cell):
y[u_,v_,w_,z_] :=$\beta_0 + \beta_1$ u+β_2 v+β_3 w+β_4 z
Residue vector $e_t = Y_t - Y_t$ can be calculated by this equation:

e_t= y-y [x1,x2,x3,x4]
{-6.81865,3.01529,-0.9684,4.18523,6.00106,-3.65136,-1.40763,0.91137, -5.77093,0.392759,4.09813,
2.3458,-1.01937,-2.80289,-3.10569,-1.48111,7.51511,-5.73012,7.70013,-2.987,-0.912197,
-2.50959,1.58758,0.661722,0.830222,-1.41202,1.95797,3.39818,-5.86285,1.83927}
Build residue graph:
gr1=ListLinePlot[e_t,ColorFunction®"Rainbow",Filling®Axis]



Given above information is only a part of laboratory work just to demonstrate how easily multiple regression model can be built in Mathematica system. Then test of Goldfeld-Kuandt is done to find out heteroscedasticity and using the method of weighted least squares the variances of estimates will be reduced, in other words heteroscedasticity is eliminated. All calculations are visually accessible and there is opportunity to experiment, to try ways of heteroscedasticity elimination.

Use of computer mathematics system will provide highest degree of visual impression in teaching econometrics, not only by plots but by fixing of every step in calculations and detailed commentaries. A student performs a task copying typical example, enters his data, arrives at his own conclusions. In such a way activity approach to teaching is provided, interactivity and increase in information level take place.

REFERENCES

1. Gusarov, V., 2003. Statistics: Textbook. Moscow: UNITY-DANA, pp: 463.
2. Sizova, T., 2005. Statistics: Textbook. St. Petersburg: St. Petersburg GUITMO, pp: 190.
3. Efimova, M.P., O.I. Ganchenko and E.V. Petrova, 2004. Practicum on general theory of statistics: Textbook. Moscow: Finances and Statistics, pp: 336.
4. Shanchenko, N., 2011. Econometrics: laboratory practicum: Textbook. Ulyanovsk: UlSTU, pp: 117.
5. Magnus, J.R. and J. Durbin, 1999. Estimation of regression coefficients of interest when order regression coefficients are of no interest. Econometrica, 67: 639-543.
6. Magnus, J.R., P.K. Katyshev and A.A. Peresetsky, 2007. Econometrics. Beginners' course: Textbook. Moscow: Delo, pp: 576.
7. Verbeek, M., 2008. A Guide to Modern Econometrics, 3rd Edition. Wiley, New York, pp: 488.
8. Magnus, J.R. and H. Neudecker, 2007. Matrix Differential Calculus with Applications in Statistics and Econometrics, 3rd Edition. Wiley, New York, pp: 468.
9. Wolfram, S., 2003. The Mathematica Book. 5th Edition. Mathematica Version 5. Cambridge University Press, pp: 1301.
10. Wolfram Mathematica Documentation Center. Wolfram Research, 2009. Date Views 07.03.14 reference.wolfram.com/mathematica/guide/Mathematica.html.