

MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE
West Ukrainian National University

Methodical instructions for conducting practical lessons
in the discipline
«ECONOMETRICS»

Ternopil – 2022

Approved at the meeting of the Department of Applied Mathematics, protocol № 1 of 26 august 2022.

UDK 519.2

Reviewers:

S.V. Martyniuk – PhD, Candidate of physical and mathematical sciences, associate professor of the Department of Mathematics and Teaching Methods at Volodymyr Hnatyuk Ternopil National Pedagogical University.

O.S. Bashutska – PhD, Candidate of Economics, associate Professor of the Department of Economic Cybernetics and Informatics of the West Ukrainian National University;

approved at the meeting of the Department of Applied Mathematics, protocol № 1 of 26.08.2022.

Examples of solution of the tasks in the course «Econometrics» are considered in methodical instructions on the base of necessary theoretical knowledge. Methodical instructions to study the discipline «Econometrics» are proposed for students for helping them in solution of individual tasks.

Plaskon S.A. , Dzyubanovska N.V. Methodical instructions for conducting practical lessons in the discipline «Econometrics». - Ternopil: WUNU, 2022.- 22 p.

UDK 519.2

Responsible for release: O.M. Martyniuk, PhD, Candidate of physical and mathematical sciences, associate professor of the Department of Applied Mathematics,

Head of the Department of Applied Mathematics

Topic . ONE-FACTOR LINEAR ECONOMETRIC MODELS

Task. For ten enterprises of the region, for a certain conditional period, the numerical values of two economic indicators are known: gross production y (million dollars) and the cost of the main production assets x (million dollars), (Table 1). To study the characteristics of the influence of the cost of the main production assets (x) on the output of gross products (y) of the enterprise:

1. Make a model specification.
2. Find statistical estimates of the parameters of the linear regression equation and construct an estimation line.
3. Calculate the value of the point estimate of the dispersion of disturbances.
4. For the level of significance $\alpha = 0,05$ check the significance of the regression coefficients α_0 and α_1 .
5. Find the confidence intervals of the regression coefficients with reliability $\gamma = 0,95$.
6. Calculate the values of the sample: coefficient of determination, coefficient of correlation.
7. Find and construct the confidence interval of the regression function with reliability $\gamma = 0,95$.
8. Check the adequacy of the constructed econometric model. If the model is adequate, then find the forecast value of the gross output for the value of the main production assets in the amount of 12 million dollars, with reliably $\gamma = 0,95$ build a confidence interval for this forecast value.

Table 1

№ firms	Gross output, mln. dol., y_i	Main production assets, mln. dol., x_i
1	4,1	2,3
2	6,3	3,4
3	7,6	4,2
4	8,9	3,7
5	7,8	4,3
6	9,7	5,4
7	11,4	6,2
8	9,5	7,8
9	12,2	8,4
10	14,3	9,8

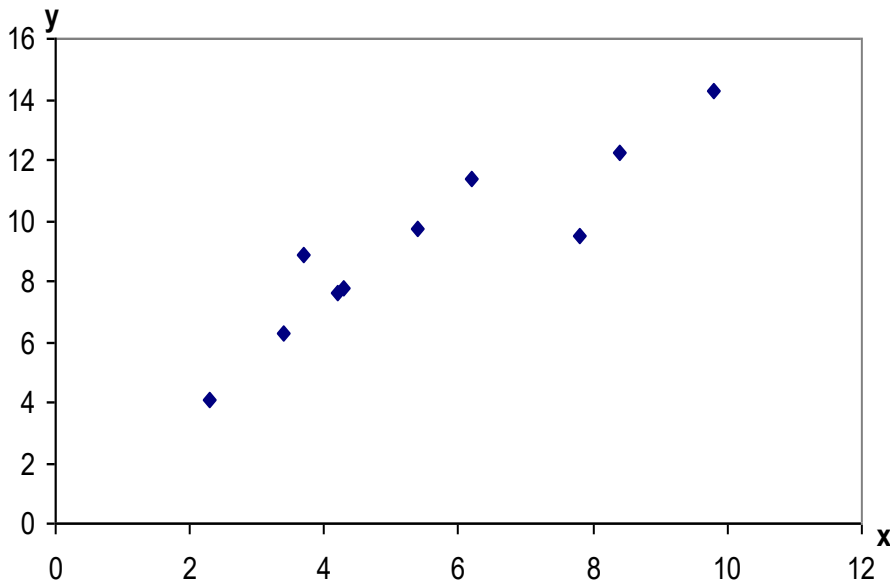
◆*Solution.*

1. In order to determine the form of dependence between two variables, a scatter diagram is used, which is a graphical representation of information in a rectangular coordinate system. The abscissa indicates the value of the independent variable (x), on the y-axis – the value of the dependent variable (y). The result of each observation (x_i, y_i) of some economic process is represented by a point on the plane. The set of these points forms a cloud that shows the relationship between the two variables.

By the width of the spread of points, it is possible to draw a conclusion about the closeness of the connection of the population. If the points are placed close to each other (in the form of a narrow strip), then it can be argued that there is a relatively close connection. If the points on the plot are widely scattered, there is a weak or no relationship between the variables.

The figure shows the main forms of dependencies.

Based on the appearance of the cluster of points, it is possible to hypothesize the linearity or non-linearity of the relationship between the variables. Thus, on diagram 1 (a, r), we have clearly expressed linear trends of clustering of points. Let's build a scatter diagram of the dependence of gross output (y) on the cost of the enterprise's main production assets (x):



The placement of points on the scatter diagram makes it possible to make an assumption about the existence of a linear form of the relationship in the form of a function:

$$\hat{y} = a_0 + a_1x, \quad (1)$$

where \hat{y} – estimated volume of gross output, mln. dol.; x – the cost of the main production assets, mln. dol.

2. Statistical estimates of the parameters of the linear regression equation can be found by the method of least squares (LSM), which is based on the requirement to minimize the sum of squares of the deviations of the empirical values of the variable y from the values calculated by the equation of the straight line. With its help, such estimates of the parameters of the regression equation are found that minimize the selected measure of dispersion. These estimates can be found by the system of normal

$$\begin{cases} na_0 + \left(\sum_{i=1}^n x_i \right) a_1 = \sum_{i=1}^n y_i, \\ \left(\sum_{i=1}^n x_i \right) a_0 + \left(\sum_{i=1}^n x_i^2 \right) a_1 = \sum_{i=1}^n x_i y_i. \end{cases} \quad (2)$$

Dividing each of the equations by n , we will get:

$$\begin{cases} a_0 + \bar{x}a_1 = \bar{y}, \\ \bar{x}a_0 + \bar{x}^2a_1 = \overline{xy}, \end{cases} \quad (3)$$

where $\bar{x} = \sum x_i/n$, $\bar{y} = \sum y_i/n$, $\overline{xy} = \sum x_i y_i/n$, $\bar{x}^2 = \sum x_i^2/n$.

Solution of the system (3) find by Kramer's rule:

$$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\bar{x}^2 - (\bar{x})^2}, \quad a_0 = \bar{y} - a_1 \bar{x}. \quad (4)$$

The value of the same estimates can be found using formulas due to the deviation from the average:

$$a_1 = \frac{\sum_{i=1}^n (\Delta x_i \Delta y_i)}{\sum_{i=1}^n (\Delta x_i)^2}. \quad (5)$$

The value of the assessment a we find from the formula (1.7):

$$a_0 = \bar{y} - a_1 \bar{x}, \quad (6)$$

where $\Delta y_i = y_i - \bar{y}$ and $\Delta x_i = x_i - \bar{x}$ – this is the variance of the variables y and x from their average values.

To simplify calculations when finding estimates a_0 and a_1 let's build a table of parameters of the econometric model:

Table 2

№	y_i	x_i	x_i^2	y_i^2	$x_i \cdot y_i$	Δx_i	$(\Delta x_i)^2$	Δy_i	$\Delta x_i \cdot \Delta y_i$
1	4,1	2,3	5,29	16,81	9,43	-3,25	10,56	-5,08	16,51
2	6,3	3,4	11,56	39,69	21,42	-2,15	4,62	-2,88	6,19
3	7,6	4,2	17,64	57,76	31,92	-1,35	1,82	-1,58	2,13
4	8,9	3,7	13,69	79,21	32,93	-1,85	3,42	-0,28	0,52
5	7,8	4,3	18,49	60,84	33,54	-1,25	1,56	-1,38	1,73
6	9,7	5,4	29,16	94,09	52,38	-0,15	0,02	0,52	-0,08
7	11,4	6,2	38,44	129,96	70,68	0,65	0,42	2,22	1,44

8	9,5	7,8	60,84	90,25	74,1	2,25	5,06	0,32	0,72
9	12,2	8,4	70,56	148,84	102,48	2,85	8,12	3,02	8,61
10	14,3	9,8	96,04	204,49	140,14	4,25	18,06	5,12	21,76
Сума	91,8	55,5	361,71	921,94	569,02	0	53,69	0	59,53

Let's calculate the average values:

$$\bar{x} = \frac{\sum_{i=1}^n x_i}{n} = \frac{55,5}{10} = 5,55; \quad \bar{y} = \frac{\sum_{i=1}^n y_i}{n} = \frac{91,8}{10} = 9,18.$$

$$\overline{x^2} = \frac{\sum_{i=1}^n x_i^2}{n} = \frac{361,71}{10} = 36,17; \quad \overline{xy} = \frac{\sum_{i=1}^n x_i y_i}{n} = \frac{569,02}{10} = 56,9.$$

Let's
write
down

the system (3):

$$\begin{cases} a_0 + 5,55a_1 = 9,18, \\ 5,55a_0 + 36,17a_1 = 56,9, \end{cases}$$

and solve it using the formulas (4):

$$a_1 = \frac{\overline{xy} - \bar{x} \cdot \bar{y}}{\overline{x^2} - (\bar{x})^2} = \frac{56,9 - 5,55 \cdot 9,18}{36,17 - 5,55^2} = \frac{5,95}{5,37} = 1,11,$$

$$a_0 = \bar{y} - a_1 \bar{x} = 9,18 - 1,11 \cdot 5,55 = 3,02.$$

Let's find the same estimates using the formulas due to the deviation from the average.

First, let's calculate the deviation of the variables from their average values:

$$\begin{aligned} \Delta x_1 &= x_1 - \bar{x} = 2,3 - 5,55 = -3,25; & \Delta y_1 &= y_1 - \bar{y} = 4,1 - 9,18 = -5,08; \\ \Delta x_2 &= x_2 - \bar{x} = 3,4 - 5,55 = -2,15; & \Delta y_2 &= y_2 - \bar{y} = 6,3 - 9,18 = -2,88; \\ \dots & & \dots & \\ \Delta x_{10} &= x_{10} - \bar{x} = 9,8 - 5,55 = 4,25. & \Delta y_{10} &= y_{10} - \bar{y} = 14,3 - 9,18 = 5,12. \end{aligned}$$

Then

$$a_1 = \frac{\sum_{i=1}^n (\Delta x_i \Delta y_i)}{\sum_{i=1}^n (\Delta x_i)^2} = \frac{59,53}{53,69} = 1,11,$$

$$a_0 = \bar{y} - a_1 \bar{x} = 9,18 - 1,11 \cdot 5,55 = 3,02,$$

So, we obtained the estimation equation of the econometric model

$$\hat{y} = 3,02 + 1,11x.$$

Let's construct this estimation straight line.

3. Unbiased point estimate S_u^2 of the unknown variance of disturbances σ_u^2 we find by the formula:

$$S_u^2 = \frac{1}{n-2} \sum_{i=1}^n (y_i - \hat{y}_i)^2.$$

First, let's calculate the estimated values \hat{y}_i according to the estimation equation of the econometric model $\hat{y} = 3,02 + 1,11x$, then deviation $u_i = y_i - \hat{y}_i$ and $u_i^2 = (y_i - \hat{y}_i)^2$.

Table 3

№	y_i	x_i	\hat{y}_i	$y_i - \hat{y}_i$	$(y_i - \hat{y}_i)^2$
1	4,1	2,3	5,58	-1,48	2,18
2	6,3	3,4	6,80	-0,50	0,25
3	7,6	4,2	7,68	-0,08	0,01
4	8,9	3,7	7,13	1,77	3,14
5	7,8	4,3	7,79	0,01	0,00
6	9,7	5,4	9,01	0,69	0,47
7	11,4	6,2	9,90	1,50	2,25
8	9,5	7,8	11,67	-2,17	4,73
9	12,2	8,4	12,34	-0,14	0,02
10	14,3	9,8	13,89	0,41	0,17
Сума	91,8	55,5	91,8	0	13,2

Using the obtained calculations, we get:

$$S_u^2 = \frac{1}{10-2} \cdot 13,2 = 1,65.$$

4. Significance of regression coefficients a_0 та a_1 we check using the null hypothesis $H_0(\alpha_m = 0)$, the content of which and alternative to it $H_1(\alpha_m \neq 0)$ is as follows: if the inequality holds

$$\left| \frac{a_m}{S_{a_m}} \right| > t_{кр.},$$

where $t_{кр.} = t_{\text{двост.кр.}}(\alpha, n-2)$, $\alpha = 1-\gamma$, – critical point of the Student's distribution, then the hypothesis is accepted at the level of significance α the hypothesis H_1 is accepted, that is, it is considered that $\alpha_m \neq 0$.

Values S_{a_0} and S_{a_1} we will find using the formulas:

$$S_{a_0} = \sqrt{\frac{S_u^2 x^2}{n\sigma_x^2}} = \frac{S_u}{\sigma_x} \sqrt{\frac{x^2}{n}} = \frac{\sqrt{1,65}}{\sqrt{x^2 - (\bar{x})^2}} \sqrt{\frac{x^2}{10}} = \frac{\sqrt{1,65}}{\sqrt{36,17 - (5,55)^2}} \sqrt{\frac{36,17}{10}} = 1,05;$$

$$S_{a_1} = \sqrt{\frac{S_u^2}{n\sigma_x^2}} = \frac{S_u}{\sigma_x \sqrt{n}} = \frac{\sqrt{1,65}}{\sqrt{10[36,17 - (5,55)^2]}} = 0,17.$$

Then the empirical values of the criterion:

$$\left| \frac{a_0}{S_{a_0}} \right| = \frac{3,02}{1,05} = 2,88, \quad \left| \frac{a_1}{S_{a_1}} \right| = \frac{1,11}{0,17} = 6,53.$$

Critical point for the two-sided critical region

$t_{кр.} = t_{\text{двост.}}(\alpha, k)$ at the values $\alpha = 0,05$, $k = n - 2 = 8$ we find from the table of critical values of the Student's distribution $t_{кр.} = 2,306$.

Since $2,88 > t_{kp.} = 2,306$ and $6,53 > t_{kp.} = 2,306$, then at the level of significance $\alpha = 0,05$ we conclude that $\alpha_0 \neq 0$ and $\alpha_1 \neq 0$.

5. Confidence intervals with reliability γ for unknown regression parameters a_0 and a_1 have the appearance:

$$a_m - t_m(\gamma, k)S_{a_m} < \alpha_m < a_m + t_m(\gamma, k)S_{a_m},$$

where $m = 0,1$, $t_m = t_m(\gamma, k)$ – root of equation $P(|t_m| < t) = \gamma$, t_0 and t_1 – random variables distributed according to Student's law.

In our case $\gamma = 0,95$, the number of degrees of freedom $k = n - 2 = 8$.

We find by the table $t_0(0,95;8) = t_1(0,95;8) = 2,306$. Then, taking into account the found values $S_{a_0} = 0,1055$, $S_{a_1} = 0,1132$ we will get:

$$3,02 - 2,306 \cdot 1,05 < \alpha_0 < 3,02 + 2,306 \cdot 1,05,$$

$$1,11 - 2,306 \cdot 0,17 < \alpha_1 < 1,11 + 2,306 \cdot 0,17$$

or

$$0,6 < \alpha_0 < 5,44,$$

$$0,72 < \alpha_1 < 1,5.$$

6. Coefficient of determination R^2 we find by the formula:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n\sigma_y^2}.$$

From the table 3 $\sum_{i=1}^{10} (y_i - \hat{y}_i)^2 = 13,02$.

Calculate σ_y^2 , using the calculations of item 2:

$$\sigma_y^2 = \overline{y^2} - (\overline{y})^2 = 921,94/10 - (9,18)^2 = 7,92.$$

Then

$$R^2 = 1 - \frac{13,02}{10 \cdot 7,92} = 0,836.$$

Thus, the variation of the dependent variable Y on 83,6% explained by the variation of the explanatory variable.

The sample correlation coefficient can be found using the formula:

$$r = \sqrt{R^2} = \sqrt{0,836} = 0,914.$$

At the same time, the positive sign of this number is chosen due to the fact that $a_1 > 0$.

7. In order to construct a confidence interval for the regression function, it is necessary to place points with coordinates on the coordinate plane $\{x_i; \hat{y}_i - t(\gamma, n - 2)S_{\hat{y}_i}\}$, $i = \overline{1, n}$ and connect adjacent (by index) points with straight lines, and then carry out a similar procedure for points

$$\{x_i; \hat{y}_i + t(\gamma, n - 2)S_{\hat{y}_i}\}.$$

Let's calculate the value $S_{\hat{y}_i}$ according to the formula:

$$S_{\hat{y}_i} = S_u \sqrt{\left[1 + \frac{(x_i - \bar{x})^2}{\sigma_x^2}\right] \frac{1}{n}}.$$

We will use the value found in point 3

$$S_u = \sqrt{1,65} = 1,28 \text{ та } \sigma_x^2 = \overline{x^2} - (\bar{x})^2 = 36,17 - 5,55^2 = 5,37, \text{ we will get:}$$

$$S_{\hat{y}_1} = 1,28 \sqrt{\left[1 + \frac{(2,3 - 5,55)^2}{5,37}\right] \frac{1}{10}} = 0,698;$$

$$S_{\hat{y}_1} = 1,28 \sqrt{\left[1 + \frac{(3,4 - 5,55)^2}{5,37}\right] \frac{1}{10}} = 0,552;$$

$$S_{\hat{y}_3} = 0,469; \quad S_{\hat{y}_4} = 0,518; \quad S_{\hat{y}_5} = 0,46; \quad S_{\hat{y}_6} = 0,406;$$

$$S_{\hat{y}_7} = 0,421; \quad S_{\hat{y}_8} = 0,564; \quad S_{\hat{y}_9} = 0,462; \quad S_{\hat{y}_{10}} = 0,846.$$

We find by the table $t(0,95;8) = 2,306$. Using the value \hat{y}_i from table 3 and values

$S_{\hat{y}_i}$, we will get the ordinates of the points of the lower limit of the confidence zone:

$$\hat{y}_1 - tS_{\hat{y}_1} = 5,58 - 2,306 \cdot 0,698 = 3,97;$$

$$\hat{y}_2 - tS_{\hat{y}_2} = 6,8 - 2,306 \cdot 0,552 = 5,52;$$

$$\hat{y}_3 - tS_{\hat{y}_3} = 7,68 - 2,306 \cdot 0,469 = 6,6;$$

$$\hat{y}_4 - tS_{\hat{y}_4} = 7,13 - 2,306 \cdot 0,518 = 5,93;$$

$$\hat{y}_5 - tS_{\hat{y}_5} = 7,79 - 2,306 \cdot 0,46 = 6,73$$

$$\hat{y}_6 - tS_{\hat{y}_6} = 9,01 - 2,306 \cdot 0,406 = 8,08;$$

$$\hat{y}_7 - tS_{\hat{y}_7} = 9,9 - 2,306 \cdot 0,421 = 8,93;$$

$$\hat{y}_8 - tS_{\hat{y}_8} = 11,67 - 2,306 \cdot 0,564 = 10,97;$$

$$\hat{y}_9 - tS_{\hat{y}_9} = 12,34 - 2,306 \cdot 0,462 = 10,86;$$

$$\hat{y}_{10} - tS_{\hat{y}_{10}} = 13,89 - 2,306 \cdot 0,846 = 11,94.$$

Then the ordinates of the points of the upper limit of the confidence zone take the following:

$$\hat{y}_1 + tS_{\hat{y}_1} = 5,58 + 2,306 \cdot 0,698 = 7,18;$$

$$\hat{y}_2 + tS_{\hat{y}_2} = 6,8 + 2,306 \cdot 0,552 = 8,07;$$

$$\hat{y}_3 + tS_{\hat{y}_3} = 7,68 + 2,306 \cdot 0,469 = 8,76;$$

$$\hat{y}_4 + tS_{\hat{y}_4} = 7,13 + 2,306 \cdot 0,518 = 8,32;$$

$$\hat{y}_5 + tS_{\hat{y}_5} = 7,79 + 2,306 \cdot 0,46 = 8,86$$

$$\hat{y}_6 + tS_{\hat{y}_6} = 9,01 + 2,306 \cdot 0,406 = 9,95;$$

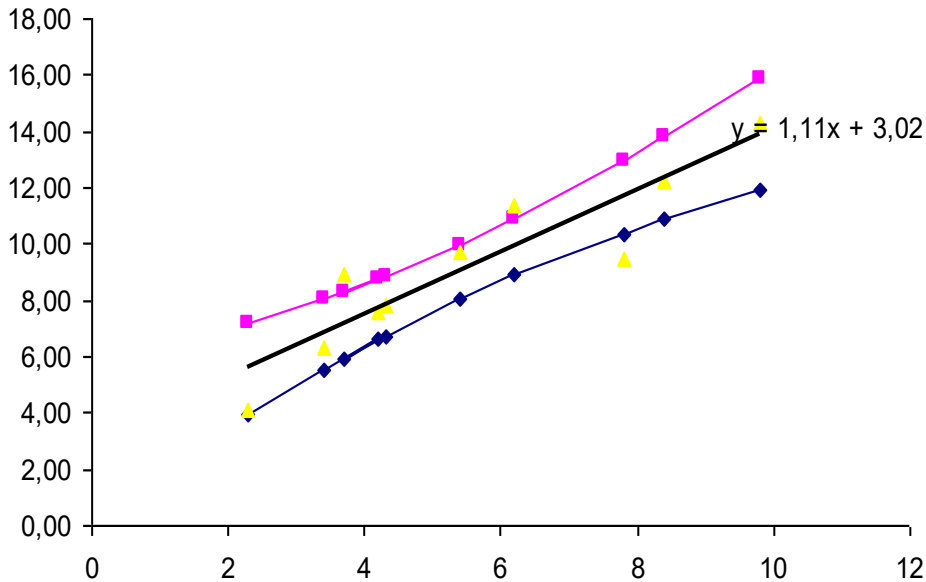
$$\hat{y}_7 + tS_{\hat{y}_7} = 9,9 + 2,306 \cdot 0,421 = 10,87;$$

$$\hat{y}_8 + tS_{\hat{y}_8} = 11,67 + 2,306 \cdot 0,564 = 12,98;$$

$$\hat{y}_9 + tS_{\hat{y}_9} = 12,34 + 2,306 \cdot 0,462 = 13,98;$$

$$\hat{y}_{10} + tS_{\hat{y}_{10}} = 13,89 + 2,306 \cdot 0,846 = 15,84.$$

Confidence interval with reliability 0,95 for the regression function has the form:



8. The adequacy of the constructed econometric model can be checked using the coefficient of determination. If its value is close to one, it can be considered that the obtained econometric model is adequate. In this case, the change in the value of the outcome variable y depends linearly on the change in the explanatory variable x , and not due to the influence of random factors. If the value of the coefficient of determination is close to zero, then the model is considered inadequate, that is, there is no linear relationship between y and x . If the value of the coefficient of determination is unclear, that is, close to 0.5, then the Fisher F . test is used to check the adequacy of the econometric model.

We calculate the empirical value of the F . parameter using the formula:

$$F_{emn} = \frac{R^2(n-m-1)}{m(1-R^2)},$$

where m – the number of independent variables (for simple regression $m=1$).

After that, we find the value from the table F_{kp} – critical value F - Fisher distribution with 3 ($k_1=m=1, k_2=n-m-1$) degrees of freedom and level of significance α . For example, $\alpha=0,05$, then we can be wrong in 5 % cases, and in 95 % of cases our conclusions will be correct.

If the value calculated by us $F_{emn} > F_{kp}$, then the econometric model built by us is adequate to reality.

In the opposite case, that is, if calculated $F_{emn} \leq F_{kp}$, then in this case the econometric model we built is inadequate to the real reality.

Let's calculate:

$$F_{emn} = \frac{R^2(n-m-1)}{m(1-R^2)} = \frac{0,836 \cdot (10-1-1)}{1 \cdot (1-0,836)} = 40,78.$$

Let's find the tabular value of this criterion (F_{kp}) for the reliability level $p=0,95$ and the number of degrees of freedom $k_1=m=1$, $k_2=n-m-1=10-1-1=8$:

$$F_{kp} = 5,32.$$

Since $F_{emn} > F_{kp}$, then the estimation equation of the econometric model obtained by us

$$\hat{y} = 3,02 + 1,11x$$

adequate to the real reality and on its basis it is possible to make forecasts, i.e. predict the ways of development of the studied phenomena and processes for the near future. The forecast can be point or interval.

Point forecast for the next one $n+1$ we get the period when we substitute the value of the explanatory variable x_{n+1} into the estimation equation of the econometric model. The forecast value of the gross production output for the cost of the main production assets in the amount of 12 million dollars will be:

$$\hat{y}_{n+1} = 3,02 + 1,1 \cdot 12 = 16,22.$$

An interval forecast is an interval in which with a given probability $p=1-\alpha$ will get a valid value of the result variable y . Confidence interval for the predicted value y_{n+1} with reliability $\gamma = 0,95$ looks like:

$$\hat{y}_{n+1} - t(\gamma; n-2)S_{u_{n+1}} < y_{n+1} < \hat{y}_{n+1} + t(\gamma; n-2)S_{u_{n+1}},$$

where

$$S_{u_{n+1}}^2 = S_u^2 \left[1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n\sigma_x^2} \right].$$

The lower limit of this interval is called a pessimistic forecast, and the upper limit is called an optimistic forecast. First for $x_{n+1}=12$ find:

$$S_{u_{n+1}} = S_u \sqrt{1 + \frac{1}{n} + \frac{(x_{n+1} - \bar{x})^2}{n\sigma_x^2}} = 1,28 \sqrt{1 + \frac{1}{10} + \frac{(12 - 5,55)^2}{10 \cdot 5,37}} = 1,75.$$

Then the desired confidence interval has the form

$$16,22 - 2,306 \cdot 1,75 < y_{n+1} < 16,22 + 2,306 \cdot 1,75$$

or

$$12,18 < y_{n+1} < 20,26.$$

©

Topic. One-factor nonlinear econometric models

Task 2.

Using the sample data of task 1, select the curve that most fully describes the trend using Excel (TREND LINE). For a better selection of the trend line, try several variants of the curves: linear; polynomial (degree 2); polynomial (degree 3); logarithmic; exponent; exponential.

Topic. Classical linear multivariate model

Construction of an econometric model with three variables by the OLS method

Task. Let consider the sample of data Y, X_1, X_2 in the next table

Table

Y_i	4	5	6	8	11	11	12	12	13	14
X_1	3	4	5	7	9	11	10	12	11	12
X_2	7	9	11	12	12	15	18	21	22	24

Determine the grades $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$ using OLS for linear econometric model.

Solution. Find $\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2$

$$\begin{cases} \sum_{i=1}^n Y_i = n\hat{\beta}_0 + \hat{\beta}_1 \sum_{i=1}^n X_{1i} + \hat{\beta}_2 \sum_{i=1}^n X_{2i} \\ \sum_{i=1}^n Y_i X_{1i} = \hat{\beta}_0 \sum_{i=1}^n X_{1i} + \hat{\beta}_1 \sum_{i=1}^n X_{1i}^2 + \hat{\beta}_2 \sum_{i=1}^n X_{2i} X_{1i} \\ \sum_{i=1}^n Y_i X_{2i} = \hat{\beta}_0 \sum_{i=1}^n X_{2i} + \hat{\beta}_1 \sum_{i=1}^n X_{1i} X_{2i} + \hat{\beta}_2 \sum_{i=1}^n X_{2i}^2 \end{cases}$$

№	Y_i	X_{1i}	X_{2i}	X_{1i}^2	X_{2i}^2	$Y_i X_{1i}$	$Y_i X_{2i}$	$X_{1i} X_{2i}$
1	4	3	7	9	49	12	28	21
2	5	4	9	16	81	20	45	36
3	6	5	11	25	121	30	66	55
4	8	7	12	49	144	56	96	84
5	11	9	12	81	144	99	132	108
6	11	11	15	121	225	121	165	165
7	12	10	18	100	324	120	216	180
8	12	12	21	144	441	144	252	252
9	13	11	22	121	484	143	286	242
10	14	12	24	144	576	168	336	288
Σ	96	84	151	810	2589	913	1622	1431

$$\begin{cases} 96 = 10\hat{\beta}_0 + 84\hat{\beta}_1 + 151\hat{\beta}_2 \\ 913 = 84\hat{\beta}_0 + 810\hat{\beta}_1 + 1431\hat{\beta}_2 \\ 1622 = 151\hat{\beta}_0 + 1431\hat{\beta}_1 + 2589\hat{\beta}_2 \end{cases}$$

$$\begin{pmatrix} 10 & 84 & 151 \\ 84 & 810 & 1431 \\ 151 & 1431 & 2589 \end{pmatrix}$$

$$\begin{pmatrix} 0.848978 & -0.02401 & -0.03625 \\ -0.02401 & 0.053163 & -0.02798 \\ -0.03625 & -0.02798 & 0.017968 \end{pmatrix}$$

$$\begin{pmatrix} 0.848978 & -0.02401 & -0.03625 \\ -0.02401 & 0.053163 & -0.02798 \\ -0.03625 & -0.02798 & 0.017968 \end{pmatrix} \times \begin{pmatrix} 96 \\ 913 \\ 1622 \end{pmatrix} = \begin{pmatrix} 0.79 \\ 0.84 \\ 0.11 \end{pmatrix}$$

$$\hat{\beta}_0 = 0.79$$

$$\hat{\beta}_1 = 0.84$$

$$\hat{\beta}_2 = 0.11$$

$$\hat{Y} = 0.79 + 0.84X_1 + 0.11X_2.$$

$$\hat{\beta}_1 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_{1i} - \bar{X}_1) \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 - \sum_{i=1}^n (Y_i - \bar{Y})(X_{2i} - \bar{X}_2) \sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 \sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 - \left(\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2) \right)^2}$$

$$\hat{\beta}_2 = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_{2i} - \bar{X}_2) \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 - \sum_{i=1}^n (Y_i - \bar{Y})(X_{1i} - \bar{X}_1) \sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{1i} - \bar{X}_1)}{\sum_{i=1}^n (X_{2i} - \bar{X}_2)^2 \sum_{i=1}^n (X_{1i} - \bar{X}_1)^2 - \left(\sum_{i=1}^n (X_{2i} - \bar{X}_2)(X_{1i} - \bar{X}_1) \right)^2}$$

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}_1 - \hat{\beta}_2 \bar{X}_2$$

№	Y_i	X_{1i}	X_{2i}	\bar{Y}	\bar{X}_1	\bar{X}_2	$Y_i - \bar{Y}$	$X_{1i} - \bar{X}_1$	$X_{2i} - \bar{X}_2$	$(X_{1i} - \bar{X}_1)^2$	$(X_{2i} - \bar{X}_2)^2$	$\frac{(X_{1i} - \bar{X}_1) \times (Y_i - \bar{Y})}{(X_{2i} - \bar{X}_2)}$	$\frac{(Y_i - \bar{Y}) \times (X_{1i} - \bar{X}_1)}{(X_{2i} - \bar{X}_2)}$	$\frac{(Y_i - \bar{Y}) \times (X_{2i} - \bar{X}_2)}{(X_{2i} - \bar{X}_2)}$
1	4	3	7	9,6	8,4	15,1	-5,6	-5,4	-8,1	29,16	65,61	43,74	30,24	45,36
2	5	4	9				-4,6	-4,4	-6,1	19,36	37,21	26,84	20,24	28,06
3	6	5	11				-3,6	-3,4	-4,1	11,56	16,81	13,94	12,24	14,76
4	8	7	12				-1,6	-1,4	-3,1	1,96	9,61	4,34	2,24	4,96
5	11	9	12				1,4	0,6	-3,1	0,36	9,61	-1,86	0,84	-4,34
6	11	11	15				1,4	2,6	-0,1	6,76	0,01	-0,26	3,64	-0,14
7	12	10	18				2,4	1,6	2,9	2,56	8,41	4,64	3,84	6,96

8	12	12	21				2,4	3,6	5,9	12,96	34,81	21,24	8,64	14,16
9	13	11	22				3,4	2,6	6,9	6,76	47,61	17,94	8,84	23,46
10	14	12	24				4,4	3,6	8,9	12,96	79,21	32,04	15,84	39,16
Σ	96	84	151				0	0	0	104,4	308,9	162,6	106,6	172,4

$$\sum_{i=1}^{10} (Y_i - \bar{Y})^2 = (-5.6)^2 + (-4.6)^2 + (-3.6)^2 + (-1.6)^2 + (1.4)^2 + (1.4)^2 + (2.4)^2 + (2.4)^2 + (3.4)^2 + (4.4)^2 =$$

$$= 31.36 + 21.16 + 12.96 + 2.56 + 1.96 + 1.96 + 5.76 + 5.76 + 11.56 + 19.36 = 114.4$$

$$\hat{\beta}_1 = \frac{106.6 \cdot 308.9 - 172.4 \cdot 162.6}{104.4 \cdot 308.9 - (162.6)^2} \approx 0.8427$$

$$\hat{\beta}_2 = \frac{172.4 \cdot 104.4 - 106.6 \cdot 162.6}{104.4 \cdot 308.9 - (162.6)^2} \approx 0.1145$$

$$\hat{\beta}_0 = 9.6 - 0.84 \cdot 8.4 - 0.1145 \cdot 15.1 = 0.7919$$

$$\hat{Y} = 0.79 + 0.84X_1 + 0.11X_2.$$

Topic. Econometric model with three variables: hierarchy of correlation coefficients

Task. Calculate the coefficients of hair, patial and multiplying correlation.

Solution.

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2,$$

$$Y = \alpha' + \beta' X_1$$

$$Y = \alpha'' + \beta'' X_2$$

$$X_1 = \alpha''' + \beta''' X_2$$

$$\therefore r_{YX_1} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_{1i} - \bar{X}_1)}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (X_{1i} - \bar{X}_1)^2}}, \quad r_{YX_2} = \frac{\sum_{i=1}^n (Y_i - \bar{Y})(X_{2i} - \bar{X}_2)}{\sqrt{\sum (Y_i - \bar{Y})^2 \sum (X_{2i} - \bar{X}_2)^2}},$$

$$r_{X_1 X_2} = \frac{\sum_{i=1}^n (X_{1i} - \bar{X}_1)(X_{2i} - \bar{X}_2)}{\sqrt{\sum (X_{2i} - \bar{X}_2)^2 \sum (X_{1i} - \bar{X}_1)^2}}$$

$$r_{YX_1} = \frac{106.4}{\sqrt{114.4 \cdot 104.4}} = \frac{106.4}{109.28} \approx 0.97, \quad r_{YX_2} = \frac{172.4}{\sqrt{114.4 \cdot 308.9}} = \frac{172.4}{187.98} \approx 0.92,$$

$$r_{X_1 X_2} = \frac{162.6}{\sqrt{104.4 \cdot 308.9}} = \frac{162.6}{179.58} \approx 0.91$$

$$|r| = \begin{pmatrix} 1 & 0.97 & 0.92 \\ 0.97 & 1 & 0.91 \\ 0.92 & 0.91 & 1 \end{pmatrix}$$

$$r_{YX_2.X_1} = \frac{r_{YX_2} - r_{YX_1} \cdot r_{X_1 X_2}}{\sqrt{1 - r_{YX_1}^2} \cdot \sqrt{1 - r_{X_1 X_2}^2}} = \frac{0.92 - 0.97 \cdot 0.91}{\sqrt{1 - (0.97)^2} \cdot \sqrt{1 - (0.91)^2}} = 0.4$$

$$r_{YX_1.X_2} = \frac{r_{YX_1} - r_{YX_2} \cdot r_{X_1 X_2}}{\sqrt{1 - r_{YX_2}^2} \cdot \sqrt{1 - r_{X_1 X_2}^2}} = \frac{0.97 - 0.92 \cdot 0.91}{\sqrt{1 - (0.92)^2} \cdot \sqrt{1 - (0.91)^2}} = 0.83$$

$$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 X_1 + \hat{\beta}_2 X_2.$$

$$R^2 = \frac{r_{YX_1}^2 + r_{YX_2}^2 - 2r_{YX_1}r_{YX_2}r_{X_1 X_2}}{1 - r_{X_1 X_2}^2}$$

$$R^2 = \frac{(0.97)^2 + (0.92)^2 - 2 \cdot 0.97 \cdot 0.92 \cdot 0.91}{1 - (0.91)^2} = 0.95$$

$$R = \sqrt{R^2} = \sqrt{0.95} = 0.97$$

Topic. A matrix approach to a linear multivariate model

Task

Let consider the statistical data y , x_1 , x_2 in the table. Calculate econometric model in matrix form.

y	4	5	6	8	11	11	12	12	13	14
x_1	3	4	5	7	9	11	10	12	11	12
x_2	7	9	11	12	12	15	18	21	22	24

Find vector of estimation of coefficients $\hat{\beta} = \begin{pmatrix} \hat{\beta}_1 \\ \hat{\beta}_2 \\ \hat{\beta}_3 \end{pmatrix}$ and matrix of dispersion of estimation $\text{var}(\hat{\beta})$.

Solution.

$$\hat{y} = \hat{\beta}_1 + \hat{\beta}_2 x_1 + \hat{\beta}_3 x_2$$

$$\hat{\beta} = [X]^T [X]^{-1} [X]^T Y, \text{ де}$$

$$[X] = \begin{pmatrix} 1 & 3 & 7 \\ 1 & 4 & 9 \\ 1 & 5 & 11 \\ 1 & 7 & 12 \\ 1 & 9 & 12 \\ 1 & 11 & 15 \\ 1 & 10 & 18 \\ 1 & 12 & 21 \\ 1 & 11 & 22 \\ 1 & 12 & 24 \end{pmatrix}$$

$$Y = \begin{pmatrix} 4 \\ 5 \\ 6 \\ 8 \\ 11 \\ 11 \\ 12 \\ 12 \\ 13 \\ 14 \end{pmatrix}$$

$$[X]^T [X] = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 7 & 9 & 11 & 10 & 12 & 11 & 12 \\ 7 & 9 & 11 & 12 & 12 & 15 & 18 & 21 & 22 & 24 \end{pmatrix} \times \begin{pmatrix} 1 & 3 & 7 \\ 1 & 4 & 9 \\ 1 & 5 & 11 \\ 1 & 7 & 12 \\ 1 & 9 & 12 \\ 1 & 11 & 15 \\ 1 & 10 & 18 \\ 1 & 12 & 21 \\ 1 & 11 & 22 \\ 1 & 12 & 24 \end{pmatrix} = \begin{pmatrix} 10 & 84 & 151 \\ 84 & 810 & 1431 \\ 151 & 1431 & 2589 \end{pmatrix}.$$

$$[[X]^T [X]]^{-1} = \begin{pmatrix} 0,848978 & -0,02401 & -0,03625 \\ -0,02401 & 0,053163 & -0,02798 \\ -0,03625 & -0,02798 & 0,017968 \end{pmatrix}.$$

$$[X]^T Y = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 3 & 4 & 5 & 7 & 9 & 11 & 10 & 12 & 11 & 12 \\ 7 & 9 & 11 & 12 & 12 & 15 & 18 & 21 & 22 & 24 \end{pmatrix} \times \begin{pmatrix} 4 \\ 5 \\ 6 \\ 8 \\ 11 \\ 11 \\ 12 \\ 12 \\ 13 \\ 14 \end{pmatrix} = \begin{pmatrix} 96 \\ 913 \\ 1622 \end{pmatrix}.$$

$$\hat{\beta} = \begin{pmatrix} 0,848978 & -0,02401 & -0,03625 \\ -0,02401 & 0,053163 & -0,02798 \\ -0,03625 & -0,02798 & 0,017968 \end{pmatrix} \times \begin{pmatrix} 96 \\ 913 \\ 1622 \end{pmatrix} = \begin{pmatrix} 0,79 \\ 0,84 \\ 0,11 \end{pmatrix}.$$

$$y = 0,79 + 0,84x_1 + 0,11x_2.$$

$$\text{VAR}(\hat{\beta}) = \sigma_u^2 [[X^T][X]]^{-1} = \begin{pmatrix} \sigma_{\hat{\beta}_1}^2 & \sigma_{\hat{\beta}_1\hat{\beta}_2}^2 & \sigma_{\hat{\beta}_1\hat{\beta}_3}^2 \\ \sigma_{\hat{\beta}_2\hat{\beta}_1}^2 & \sigma_{\hat{\beta}_2}^2 & \sigma_{\hat{\beta}_2\hat{\beta}_3}^2 \\ \sigma_{\hat{\beta}_3\hat{\beta}_1}^2 & \sigma_{\hat{\beta}_3\hat{\beta}_2}^2 & \sigma_{\hat{\beta}_3}^2 \end{pmatrix},$$

where $\sigma_u^2 = \frac{e'e}{n-k}$,

$$\hat{y}_1 = 0,79 + 0,84 \cdot 3 + 0,11 \cdot 7 = 4,12;$$

$$\hat{y}_2 = 0,79 + 0,84 \cdot 4 + 0,11 \cdot 9 = 5,19;$$

and so on

$$\hat{y}_{10} = 0,79 + 0,84 \cdot 12 + 0,11 \cdot 24 = 13,65.$$

$$e_1 = 4 - 4,12 = -0,12; \quad e_2 = 5 - 5,19 = -0,19; \quad \text{and so on } e_{10} = 14 - 13,65 = 0,35.$$

$$e' e = (-0,12 \quad -0,19 \quad -0,26 \quad -0,06 \quad 1,25 \quad -0,78 \quad 0,72 \quad -1,31 \quad 0,42 \quad 0,35) \cdot \begin{pmatrix} -0,12 \\ -0,19 \\ -0,26 \\ -0,06 \\ 1,25 \\ -0,78 \\ 0,72 \\ -1,31 \\ 0,42 \\ 0,35 \end{pmatrix} = 4,82.$$

$$\sigma_u^2 = 4,82 / (10 - 3) = 0,69.$$

$$\text{var}(\hat{\beta}) = 0,69 \times \begin{pmatrix} 0,848978 & -0,02401 & -0,03625 \\ -0,02401 & 0,053163 & -0,02798 \\ -0,03625 & -0,02798 & 0,017968 \end{pmatrix} = \begin{pmatrix} 0,585795 & -0,01657 & -0,02501 \\ -0,01657 & 0,036682 & -0,01931 \\ -0,02501 & -0,01931 & 0,012398 \end{pmatrix}.$$

Topic. Multicollinearity

Task

Investigate the problem of multicollinearity for the next statistical sample

x_1	x_2	x_3
1,43	9,18	8,97
3,92	10,94	9,8
5,49	11,3	9,6
8,17	13,99	8,02
9,68	14,59	10,51
14,42	15,73	12,39
15,92	17,97	16,36
18,04	19,45	9,32
20,69	21,49	12,72
22,68	21,8	16,22

$$[R] = \begin{bmatrix} 1 & 0,99 & 0,70 \\ 0,99 & 1 & 0,66 \\ 0,70 & 0,66 & 1 \end{bmatrix}. \quad \det[R] = 0,0077.$$

$$\chi_p^2 = -\left[n - 1 - \frac{1}{6}(2m + 5)\right] \ln \det[R] = -\left[10 - 1 - \frac{1}{6}(6 + 5)\right] \ln 0,0077 = 34,85.$$

$$\chi_{kp}^2(0,95;3) = 7,8.$$

$$[Z] = [R]^{-1} = \begin{bmatrix} 72,48 & -67,98 & -5,83 \\ -67,98 & 65,54 & 4,29 \\ -5,83 & 4,29 & 2,25 \end{bmatrix}$$

$$t_{ij} = \frac{r_{ij}^* \sqrt{n - m - 1}}{\sqrt{1 - r_{ij}^{*2}}},$$

where $r_{ij}^* = \frac{-z_{ij}}{\sqrt{z_{ii}z_{jj}}}$; z_{ij}, z_{ii}, z_{jj} – elements of matrix $[Z]$.

$$r_{12}^* = \frac{-(-67,98)}{\sqrt{72,48 \cdot 65,54}} = 0,986; \quad t_{12} = \frac{0,986 \cdot \sqrt{10 - 3 - 1}}{\sqrt{1 - 0,986^2}} = 14,639;$$

$$r_{13}^* = \frac{-(-5,83)}{\sqrt{72,48 \cdot 2,25}} = 0,456; \quad t_{13} = \frac{0,456 \cdot \sqrt{10 - 3 - 1}}{\sqrt{1 - 0,456^2}} = 1,256;$$

$$r_{23}^* = \frac{-4,29}{\sqrt{65,54 \cdot 2,25}} = -0,353; \quad t_{23} = \frac{-0,353 \cdot \sqrt{10 - 3 - 1}}{\sqrt{1 - 0,353^2}} = -0,923.$$

$k = n - m - 1 = 10 - 3 - 1 = 6$ and $p = 0,95$ $t(0,95;6) = 2,227$. For x_1 and x_2 $t_{12} > t(0,95;6)$.

Multicollinearity exist for pair of factors x_1 and x_2 with significance $p = 0,95$.

Factor x_1 is deleted from the next analysis.

For x_2 and x_3 : $[R] = \begin{bmatrix} 1 & 0,66 \\ 0,66 & 1 \end{bmatrix}.$

$$[Z] = \begin{bmatrix} 1,785 & -1,184 \\ -1,184 & 1,785 \end{bmatrix}, \det[R] = 0,560.$$

Value $\chi_p^2 = -\left[10 - 1 - \frac{1}{6}(4 + 5)\right] \ln 0,560 = 4,346$ is lesser $\chi_{kp}^2(0,95;2) = 6,0$.

Multicollinearity between x_2 and x_3 – absent.