MINISTRY OF EDUCATION AND SCIENCE OF UKRAINE

SIMON KUZNETS KHARKIV NATIONAL UNIVERSITY OF ECONOMICS

Guidelines
to self-study work
on the academic discipline
"ECONOMETRICS"
for full-time students
of training direction 6.030601 "Management"

Харків. Вид. ХНЕУ ім. С. Кузнеця, 2014

Затверджено на засіданні кафедри економіки, організації та планування діяльності підприємства.

Протокол № 11 від 03.03.2014 р.

Compiled by K. Tonieva

G94 Guidelines to self-study work on the academic discipline "Econometrics" for full-time students of training direction 6.030601 "Management" / compiled by K. Tonieva. – Kh. : Publishing House of S. Kuznets KhNUE, 2014. – 32 p. (English)

Tasks for self-study on the academic discipline, guidelines to their implementation and questions to consolidate knowledge are given.

Recommended for full-time students of training direction 6.030601 "Management".

Надано завдання до самостійної роботи з навчальної дисципліни, методичні рекомендації до їх виконання та питання для закріплення знань.

Рекомендовано для студентів напряму підготовки 6.030601 "Менеджмент" денної форми навчання.

Introduction

The results of enterprises' commercial activities are affected by a large number of factors and phenomena. This effect is varied: some factors and phenomena have a major impact, others are weaker, and some of them have no effect. To develop an effective policy it is necessary to assess the relationship between the factors and phenomena, and their impact on the enterprises' performance.

Mathematical methods are essential for selection and formalization of the relationship between economic indicators, assessment of the form and parameters of their dependencies. Accordingly, the production manager should know and be able to apply this method.

A future manager should also know economic features, organization and production planning, taking them into account while developing mathematical models of economic activity.

Economic processes and phenomena are studied by means of Econometrics methods. Therefore, an expert in management should know construction principles and econometric model analysis to be able to justify quantitatively and solve current problems. To solve economic problems by means of econometric methods, a production manager must also work with a personal computer and specialized applications of software packages.

The purpose of the academic discipline "Econometrics" is the formation of future managers' theoretical knowledge and practical skills in econometrics, the application of methods that optimize the tasks of management, organization and production planning. The aim of the academic discipline is to apply econometrics so as to test economic theories concerning factual and quantitative optimal solutions.

The subject of the academic discipline is economic and mathematical methods and tools for studying the economic phenomena and processes.

The academic discipline "Econometrics" is taught within training direction 6.030601 "Management" of full-time form of study.

Assignment 1

A company receives orders for production of an original part of an aggregate machine. The task is to assess the value of this order. The main economic indicator, based on which one can estimate the enlargement cost of the order, is the laboriousness of its production, which depends on the technical parameters of the order.

Studies were conducted which showed that the laboriousness (denoted by L) of the original part production for the aggregate manufacturing machine is mostly affected by the following technical parameters: weight (W), the number of operations (P), the number of machined surfaces (F), the number of parts (N). These data are in order and they can be used to justify the laboriousnes of fulfilling the order.

The objective of the assignment is to build an econometric model of the order fulfillment laboriousness dependence on one of its technical parameters.

To do this you need:

1. Select an econometric model of laboriousness dependence on the technical parameter based on the economic and graphical analyzes. Build the dependence of laboriousness chart (for all the 25 initial data points), which will serve a variable dependent (Y), on the technical parameter which is an independent variable or factor argument (X). Select linear dependence of the type $Y = a + b\,X$ as an econometric model.

2. Solve this model by the method of least squares.

3. Build a theoretical regression line on the chart of the dependence of laboriousness on the technical parameter. Give an economic analysis of the resulting model.

4. Rate closeness of the relationship between the indicators. For this you need to calculate the correlation coefficient and the correlation ratio. Check the correlation coefficient materiality by ′Z Fisher. Give a mathematical validation of linear forms of relationship between these indicators.

5. Calculate the residual variability and the coefficient of variation, and suggest a possibility of using this model in practice for determining the value of the labor input of fulfilling the order.

Do the task using a personal computer. To do this, use the program Excel and calculate all the necessary parameters.

The variants of assignments are determined by the last number of the student's record-book: for example, the last number is 5, the variant is 5. For this and other variants use Table 1, to determine an array of the initial data and the name of the technical parameter (*X*), while the value of the labor input (*У*) and technical parameters are given in Table 2.

The assignment should be done in accordance with the title page which includes the name of the university, the department (in this case the Department of Management), the name of the academic discipline, the name and patronymic of the student, the code of the test book and a detailed home address. At the end of the work it if necessary to put the date, the signature, provide a list of references.

Table 1

Variants of assignment

Variant number	Array of initial data (from Table 2)	Pair model	
		Dependent variable (У)	Quotient argument (X)
1	1 – 25	L	W
2	1 – 25	L	P
3	1 – 25	L	F
4	3 – 27	L	W
5	3 – 27	L	P
6	3 – 27	L	F
7	3 – 27	L	N
8	5 – 29	L	P
9	5 – 29	L	W
10	5 – 29	L	F

Table 2

Labor input of manufacturing and technical parameters of the product

Number	L (labor input in norm-hours)	W (weight, kg)	P (number of operations)	F (number of machined surfaces)	N (number of parts)
1	54.1	60.8	221	524	53
2	72.4	176.20	239	449	52
3	72.7	226.0	387	567	44
4	73.4	148.3	366	688	65
5	75.5	183.6	431	329	37
6	78.5	253.4	272	720	120
7	89.0	217.3	263	942	51
8	91.5	147.4	345	676	89
9	93.0	120.5	404	701	94
10	95.5	146.1	394	921	78
11	98.8	131.8	309	748	44
12	105.8	226.8	284	917	45
13	107.2	214.9	419	835	98
14	107.9	187.5	462	1008	71
15	114.8	146.1	302	947	106
16	122.7	167.1	582	1384	106
17	134.6	245.2	510	1070	91
18	147.6	294.0	510	1309	118
19	160.3	161.2	620	1233	111
20	170.7	161.8	634	1186	93
21	192.1	197.5	656	1425	154
22	194.6	300.0	782	1508	207
23	198.8	299.0	692	1454	120
24	202.7	304.1	628	1297	175
25	206.1	220.3	1070	1777	233
26	206.8	277.3	748	1665	166
27	215.5	314.6	1076	2085	234
28	220.7	249.1	849	1662	230
29	229.0	289.2	838	1608	243

The graph should be made on the computer. If this is not possible, plot the graph on the markup paper or plain paper.

Methodological guidelines for Assignment 1
1. The essence of correlation and regression

A dialectical approach to the study of nature and society requires consideration of phenomena and processes in their relationship and perpetual change. In the study of social phenomena, identification and quantification of relationships between them are of paramount importance. In nature and social life relationships are very different. Some signs are called factors and effect others denoted by *X*. Others are the result of these factors' action. They are called functions and are denoted by *Y*. Diversity of relationships require their classification. Relationships are divided into functional and statistical.

A functional relationship implies such links when each value of a factor corresponds to one value of the function. Quantitative measurement of functional relationships is conducted in the course of mathematical analysis. For example, the area of a circle is equal to $S = \pi * R^2$. In social life, daily worker's wage based on piecework pay and constant pricing is functionally related to the amount of the manufactured products. Functional relationships are almost never encountered in mass phenomena of social life.

Statistical relationships are connections between random variables. Random variables are values that change their values in the same production conditions. An example of statistical relationships is the relationship between the volume of a printed book and its price. Two books of the same volume can cost differently. The price of a book does not only depend on the volume but on other factors as well. If we consider a number of books, it appears that a certain volume of a book corresponds to a variety rather than one price value and vice versa.

Statistical relationships in which the change of the average value of one attribute results in the change of the average value of another attribute and which occur in a large number of observations, are referred to as correlation relationships. *Correlation* is a Latin word that means *rati*, relationship. In mathematical statistics, this concept was introduced by the British scientist Galton.

Examples of correlation relationships are the dependence of wages on productivity; unit cost on labor productivity; labor productivity of a worker on his length of service, etc. Labor productivity increases with the increasing skill of a worker. This is due to the fact that productivity growth is also influenced by other factors: technical equipment of labor, mechanization and automation of production processes, work organization. These factors act in parallel with increasing qualifications. The effect of these factors is such that the trend of labor productivity growth in relation to qualification increase manifests itself only as an average tendency which can be revealed in a mass of observations.

The reason for the existence of correlations is that the economic indicator is affected by many factors, but only the relationship of some of them is studied. The effect of the recorded factors causes a change of the dependent variables on average against a large number of vibrations, and these vibrations are caused by the effect of unaccounted factors. Therefore, if we take, for example, a factory, the relationship between the cost and productivity may not appear clearly, and a large mass of phenomena of random fluctuations mutually cancel each other, and the connection will be seen more clearly.

Correlations are studied by the method of mathematical statistics – correlation analysis, which makes it possible to quantify the strength of the relationship between the phenomena. Correlation analysis requires both interrelated phenomena to be expressed in random numbers.

Correlations can be of different nature. Among them, one can distinguish three groups: dependence, interdependence and relationship arising under the influence of "third factors".

Examples of relationships. Dependence: the relationship between productivity and skills of workers. Here there is a one-sided relationship.

Interdependence: productivity affects the cost and is an important factor in its decline and salary increase. In turn, wage growth has a stimulating effect on the increase in labor productivity, i.e., a two-way relationship can be traced.

The relationship arising under the influence of "third factors". It is known that an increase in weight of similar finished parts tends to increase their laboriousness, if the same equipment and a processing method are used. A direct causal relationship between weight and laboriousness is difficult to establish. Workpiece weight has a direct impact on the time of installation and removal of parts. However, such a relationship exists, and it is caused by the following reasons. The laboriousness of processing is directly affected by the surface and the larger it is, the higher laboriousness. But the increase of the surface of similar parts is associated with the increase in the their linear dimensions and thus weight. Thus, the weight effects the laboriousness.

Correlations can be direct and inverse. Direct relationship is such in which Y increases with increasing X, e.g. dependence of the volume of production on fixed assets, wages on qualifications. Inverse correlation is such in which a factor-dependent variable argument decreases, e.g. the relationship between productivity and the unit cost of production.

The relationships, by analytical expression, can be straight and curved, single-factor and multifactor.

The purposes of applying the correlation analysis in the economy are: to determine the essentiality of the relationship between the studied phenomena, to construct mathematical formulas relating these phenomena.

Correlation studies are also sometimes called regressive. Thus, correlation means connection between random variables, and regression is such relations when the dependent variable is random, and one or more factors are non-random. An example of such relations is the research into productivity, cost and other economic indicators' dependence on the size of the fixed assets, the share of wages, the level of specialization and so on. Methods of analysis in general are similar. But regression analysis imposes less stringent requirements on the initial information, and hence the lack of reliability of some results.

2. Building a correlation model

The study of economic processes and phenomena by mathematical methods consists of the following stages:

economic setting of a problem;

building econometric models;

solution of the model;

evaluation of the solution.

These steps are interrelated and influence each other. The least developed of these stages is the stage of econometric modeling. During the correlation modeling a number of issues are solved, the most difficult one is the choice of factors and forms of relation.

The correlation model has the form:

$$У = f (X_1, X_2, \dots X_n),$$

where *Y* is a dependent variable,

X is factors-arguments,

n is the number of factors-arguments, (when n = 1 it is a pair model, when n >1 it is a multifactor model),

f is the form of relation.

The problem of modeling is to choose and find a formula, which stands under the sign of *f*. To do this, follow these steps:

1) based on the objectives of the study, choose the dependent variable to determine the appropriate form of communication;

2) make a selection of factors-arguments;

3) accept the hypothesis of the forms of the relation between indicators (the choice of forms of relations);

4) prepare the initial data to solve the model.

The choice of the dependent variable is determined by the purpose of the research, for example, the definition of the factors affecting the cost of production. The dependent variable in this case is cost, where costs can be expressed as the cost of production, unit costs, the cost reduction index.

Selection of factors-arguments. Factors-arguments should reflect objective characteristics and production conditions. Establishing a relationship between the indicators is based on the economic way rather than a mathematical one. Productivity selection of factors-arguments is hard because no factor can be considered whose changes will lead to a change in the dependent variable.

Direct selection of factors-arguments is based on a qualitative theoretical economic analysis. The disadvantage of this method is that it does not allow us to answer the question about the quantitative nature of the relations between economic indicators. This contradiction can be eliminated through a two-step selection of process factors. The first stage is the theoretical justification of factors. The second one is the final selection with a special additional quantification and qualitative analysis.

In this case graphical analysis of pair relations is widely used. For this purpose, a graph of dependence is build, where the *X*-axis is the factor and the *Y*-axis is the dependent variable. According to the location of points on the graph, you can answer the question not only about the existence of a connection, but also about its form. The model should include only the main factors. The presence of extra factors complicates the calculation as it increases the number of calculations, creates the illusion of accounting a large number of factors. Usually 4–5 factors are taken.

Requirements to factors-arguments

Factors-arguments must be quantitatively comparable. Qualitative attributes must be transformed into quantitative ones by ranking, ball count etc.

Factors-arguments should not be in the functional dependence on each other. The presence of a linear relationship between the factors is called multicollinearity. This means that the factors describe the same phenomenon. For example, the volume of commodity output depends on the quantity. Therefore, if the model has a marketable output, labor productivity, the number should not be entered.

Construction of indicators-arguments: indicators should be calculated on the basis of statistical reporting and should not require additional calculations and studies to obtain them; there should be strictly defined boundaries of the period for which the dependent variable and independent figures are calculated. It is preferable to take a year since random fluctuations (disposable cancellation of materials to order, delinquency, etc.) are less affected by annual rates than monthly ones.

The choice of the relation form

The most critical and complicated issue is the choice of the mathematical form of relation. The mathematical form of relation refers to the type of mathematical functions used to express the relationship between indicators.

Requirements for selection of functions:

a function should reflect the nature of economic relations and the form of communication should correspond to the logic of economic analysis;

the form of relation should be simple, it follows from the requirement of convenience. It is possible to construct a function that has as many parameters as the initial data, but this model will not express any consistent pattern.

Ways to build forms of relationship:

1. A model based on the economic analysis.
2. A graphical analysis, i.e. building a chart of dependence.

3. Analysis and use of well-known correlation models.

4. Sorting out a method of known mathematical functions.

In the economic calculations, the following functions are implemented: linear ($Y = a + b\ X$), power ($Y = a\ X^{b}$), logarithmic ($Y = a + b\ lg\ X$), parabola ($Y = a + b\ X + c\ X^2$), hyperbole ($Y = a + b / X$), and others.

Preparation of initial data for correlation analysis:

1. The number of observations must be over 30.

2. Data must be homogeneous. If they belong to different and heterogeneous aggregates, it is impossible to find a pattern.

3. Outliers are forbidden where one or more observations are very different from others. There are formulas and criteria by which you can justify the need for ignoring some observations.

4. To increase the number of observations plant years and workshop months are used. Between these parameters there can be autocorrelation (correlation between individual values from time series).

5. Data is obtained by the timing, inventory, pictures, calculations by the proposed method, questioning etc.

3. Solving the correlation model

Once the model is made, it must be solved. Solving the model means determining the numerical values of the parameters of the dependence under consideration. The correlation model parameters can be determined in various ways: the method of least squares, the point method, the graphical method, the linear programming techniques, and others. The most widely used method is the method of least squares. The method consists in the following: find such parameters of the equation, which minimize the sum of squared deviations of the calculated values of the dependent variable from their actual values: $Z = \sum(Y - Y_p)^2 \rightarrow \min$.

For the linear function У = a + b X the system of equations has a form:

$$\begin{aligned} \sum У &= a\, N + b \sum X; \\ \sum У\, X &= a \sum X + b \sum X^2. \end{aligned} \tag{1}$$

The resulting system is called the system of normal equations. Nonlinear forms are solved as well, but we need to introduce additional variables that lead the system to a linear form. After the construction of the system it is necessary to solve it by one of the known methods, e.g. by Gause.

For the pair model there are formulas, obtained by the method of substitution which are convenient for remembering:

$$b = D\,yx / D\,xx; \quad a = (\sum У - b \sum X) / N;$$
$$Dyx = \sum УX - (\sum X \sum У) / N; \qquad D\,xx = \sum X^2 - (\sum X)^2 / N.$$

Consider an example. You need to find the dependence of the cost of reconstruction of a workshop on its capacity. The initial data are shown in Table 3 (col. 1 – 3). The first out-turn was accepted as a unit, then all the others were recalculated with relation to it.

Solving the problem

Choose the dependent variable: *Y* is renovation costs, *X* is conventional out-turn.

Build a chart of dependence of these indicators.

Select a form of relationship between indicators based on the economic and graphical analyzes. Accept the linear dependence of the form $У = a + b X$. This dependence corresponds to the logic of the economic analysis: with increasing out-turn, costs of reconstruction should grow, but not in proportion to the growth of this out-turn. Parameter *"a"* evaluates the impact of other factors on the costs (buildings, structures, communications, engineers, etc.).

Solve the problem by the least squares method. To do this, make a system of equations:

$$\sum У = a N + b \sum X; \qquad (2)$$
$$\sum У X = a \sum X + b \sum X^2.$$

Calculation of the necessary amounts is made in Table 3. Based on its data solve the system and find parameters *a* and *b*:

$$Dyx = 50.25 - 13.39 \cdot 12 / 4 = 10.05; \quad Dxx = 50 - 12 \cdot 12 / 4 = 14;$$
$$Dyy = 52.1839 - 13.39^2 / 4 = 7.361;$$

where $b = Dyx / Dxx = 10.05 / 14 = 0.718$; $a = (13.39 - 0.718 \cdot 12) / 4 = 1.194$.

Thus, the equation of costs depending on the out-turn has the form:

$$У_t = 1.194 + 0.718\,X. \qquad (3)$$

This model is consistent with the logic of economic analysis: with increasing out-turn costs are rising, but not proportionally, as the cost is influenced by other factors that do not depend on the out-turn.

Determination of the calculated values of the dependent variable ($У_p$) for the found model.

For this, we substitute the actual values of technical parameters (X) in model (1).

Checking the accuracy of calculations $\sum У = \sum У_p$. In our example, the differences are small, then the parameters are calculated correctly. Based on the data from column 6, Table 3, build on two points a theoretical regression line on the graph of the dependence of productivity on the technical parameter. If the parameters are calculated correctly, the line must pass through an array of points. In this example it happens.

4. Evaluation of relation's closeness

Evaluation of the results of solving the problem is made by using qualitative and quantitative analysis. Qualitative analysis means verification compliance with the direction and nature of the influence of individual factors on the dependent variable. Quantitative analysis: implies finding the closeness of the connection parameters. The close connection in correlation analysis means the share of explainable variability of *Y* in all its variability. In mathematical statistics, variability is measured using indicators of dispersion and is given by:

$$\sigma_y^{\,2} = \sum(Y - \bar{Y})^2 / N. \tag{4}$$

All the variability can be represented as a sum of two terms: variability due to factor *X* (σ_x^2) and netbook variability (σ^2_{net}):

$$\sigma^{\,2} = \sigma^2_x + \sigma^2_{net}. \tag{5}$$

There are two indicators used to evaluate closeness of relation: the correlation ratio $\eta^2_{y/x}$ and the correlation coefficient *r*.

The correlation ratio $\eta^2_{y/x}$ is the share of the explainable variability, i.e. the variability of the dependent variable due to this factor-argument.

The correlation ratio $(\eta^2_{y/x})$ varies from zero to one, a positive value, the higher the ratio, the higher the effect of X on Y:

$$\eta^2_{y/x} = 1 - \sum (У - Уt)^2 / \sum(У - \bar{У})^2. \tag{6}$$

The correlation ratio is calculated to measure all forms of relations. The square root from the correlation ratio is the index of correlation $\eta_{y/x}$.

The correlation coefficient r is the second indicator to estimate the closeness of the relation. It varies from –1 to +1. When the coefficient is zero $(r = 0)$, the correlation is not observed (no linear relationship), but there may be statistical or non-linear (functional dependency). If $r = \pm 1$, it is the functional relationship. The correlation coefficient does not characterize any relationship between X and Y, but the degree of closeness of the linear relationship between indicators.

In the equation of the regression coefficient *"b"* shows how much the average change in $У$ per unit change in X. A value of *"b"* is dimensional and depends on the units X and $У$. Therefore, there is a problem to evaluate the closeness of the relationship without accounting dimensions X and $У$. A system has been proposed, where $t_x = (X - \bar{X}) / \sigma_x$, $t_y = (Y - Y) / \sigma_y$, and the equation has the form $t_y = r\ t_x$. Then the regression coefficient *"b"* in a standardized scale will show by how many sigmas (σ_y) the average change of $У$ will be when X changes by one sigma (σ_x). This standardized regression coefficient is called the coefficient of correlation. The formula for calculating the correlation coefficient:

$$r = b \cdot \sigma_x / \sigma_y \ ; \qquad r = \frac{D_{yx}}{\sqrt{D_{xx} D_{yy}}}. \tag{7}$$

The sign of the correlation coefficient coincides with the sign of regression *"b"* and if X increases there is an increase of Y, then the correlation coefficient and the parameter *"b"* are positive; if Y decreases with increasing X, then the correlation coefficient and the parameter *"b"* will be negative.

For linear relationship, the correlation coefficient is also a measure of the closeness of the relation. Squared correlation coefficient r^2 is called the coefficient of determination, which shows how much of the changes *У* go due to changes of *X.*

With the nonlinear dependence, the value of the correlation coefficient may be considerably less than the theoretical correlation ratio. The correlation coefficient can be close to zero, even when the correlation coefficient is close to one.

The correlation coefficient allows us to characterize the degree of approximation to the study of linear correlation and estimate closeness of the relationship. For linear dependence the correlation coefficient is used. If the coefficient of determination is equal to the correlation ratio $(|r^2 - \eta^2| < 0.1)$, then the linear form of relation is correct. The differences between these indicators are not different by form of calculation, but by a different form of relation.

For our example:

$$r = \frac{10.05}{\sqrt{14 \cdot 7.361}} = 0.9899 = 0.99;$$

$$\eta^2_{y/x} = 1 - \sum(Y - Yp)^2 / \sum(Y - Y)^2 = 0.9;$$

$$r^2 - \eta^2_{y/x} = 0.98 - 0.9 = 0.08 < 0.1.$$

Then the linear form of relation is correct.

Determining the essentiality of the relation

For solving the problems only some of the data is used while the results apply to the entire data set. Is it possible? Yes, this method is used in all kinds of scientific activity. The problem is that there is a pattern found in the selection. You need to check whether there is a relation in the general aggregate.

Determining essentiality of the relation is performed using various criteria. Consider one of them – criteria Z' *Fisher's*. For the test, a coefficient of correlation is used. The essence of the test is the following. The value is calculated:

Z' = 1.151 lg (1 + |r|) / (1 – |r|) or Z' = 1/2 ln (1+|r|) / (1 – |r|), then the standard deviation $S_{z'} = Z' / \sqrt{N-3}$ and their ratio, which is called quantile $Up = Z' / S_{z'}$.

This quantile is compared with the table value for a given probability. If *Up* > *Um*, with the given probability *P* it is said, that the relationship exists in the universal set and the correlation coefficient in it is not a zero, i.e. it is significant or existent. If *Up* < *Um*, then with a given probability it can not be argued that the relationship is significant.

If the probability is P = 0.90, the quantile Uт = 1.64; P = 0.95, Uт = 1.96; P = 0.99, Uт = 2.58.

In our example r = 0.99: Z' = 1.151 lg (1 + 0.99)/(1 – 0.99) = 2.55;

Ut = 2.55 / $\sqrt{4-3}$ = 2.55 > 1.96 for the probability P = 0.95.

Hence, the correlation coefficient is significant, and there is a relation between the indicators and the general aggregate.

5. Evaluation of the model reliability

The most important element of the correlation analysis is the evaluation of the results. It allows you to answer the question: can we use this model for analysis and planning of the production? In the case of contradiction the reasons should be identified and eliminated. The reason for discrepancies may be the following factors: the model does not includes all the factors, mistakes were made in the selection of the initial material, the aggregate is not heterogeneous, the form of relation was chosen incorrectly

Qualitative analysis consists in verifying compliance with the directions in the influence of individual factors on the dependent variable and estimating the accuracy of evaluations made by the model. The variability calculated for this netbook (measured in the same units as *Y*):

$$\sigma_{net} = \sqrt{\frac{\sum (У - У_t)^2}{N}}. \qquad (8)$$

For our example $\sigma_{net} = \sqrt{\frac{0.1440}{4}}$ = ± 0.19 mln UAH.

To get the relative value of the variability – the coefficient of variation, the netbook variability is divided by the average value of the dependent variable: $V = (\pm\sigma_{net} / \bar{Y})\, 100$. If the variability of the calculated values $У_p$ from the actual *У (V)* does not exceed 5 – 10 %, it is considered that the resulting model can be used. If the variability is higher, certain factors have not been taken into account in the model, and accounting them in a model will lead to the construction of a multi-factor correlation model.

In our example V = (± 0.19 / 3.35) 100 = ± 5.65 % < 10 % ($\bar{У}$ = ∑ У / N = 13.39 / 4 = 3.35). So you can use the model in practice to justify the cost for reconstruction depending on the out-turn of the object.

Table 3

Calculation table

No.	У	X	УX	X^2	$У^2$	Уt	$(У – Уt)^2$	$(У – \bar{У})^2$
1	2	3	4	5	6	7	8	9
1	1.75	1	1.75	1	3.0625	1.91	0.0256	2.56
2	2.59	2	5.18	4	6.7081	2.63	0.0016	0.5776
3	3.67	3	11.01	9	13.4889	3.35	0.1024	0.1024
4	5.38	6	32.28	36	28.9444	5.50	0.0144	4.1209
∑	13.39	12	50.22	50	52.1839	13.39	0.1440	7.3609

Assignment 2

Evaluation of the relative importance of the parameters of a new product

The content of the task. A company regularly updates its product range. The strategic plan of the company is the launch of a new product. It is necessary to evaluate the importance of its parameters to develop an effective advertising campaign.

The estimation of the importance of the parameters can be carried out with the help of expert methods. Various ranking techniques are avaliable. In this paper, it is necessary to estimate the importance of the parameters with a simple ranking method and pairwise comparisons. *n* experts were invited (Table 1 shows the number of specialists in the variants of the task).

4 parameters of the product are ranked: A, B, C and D. The experts should evaluate the importance of these parameters using a simple ranking method. These data are shown in Table 2.

On the basis of these data, using the methods of pairwise comparisons it is necessary to estimate the reliability of expert evaluations and rank parameters in order of their importance.

Table 1

Initial data for the assignment – the number of experts (n)

Number of experts	Number of group							
	1	2	3	4	5	6	7	8
n	7	8	9	10	11	12	5	6

Table 2

Initial data for the assignment – ranks set by experts

Expert	Parameter			
	A	B	C	D
1	3	2	1	4
2	1	2	3	4
3	3	1	2	4
4	1	2	3	4
5	3	1	2	4
6	3	1	2	4
7	3	2	4	1
8	3	4	1	2
9	2	4	1	3
10	2	1	3	4
11	2	1	2	2
12	1	1	3	1

The order of doing the task

1. Based on the initial data (Table 1 and Table 2) calculate average grades in the parameters of the group consisting of *n* experts. Draw conclusions about the importance of the parameters.

2. Create a matrix of pairwise comparisons (matrix A).

3. On the basis of matrix A justify the share of cases of preferences parameter *i* to parameter *j* (matrix *P*).

4. Create a differences matrix (matrix *t*) between parameters *i* and *j* in the standardized deviations using the tables of integral distribution function (Table 7) and the matrix *P*.

5. Estimate the importance of the parameters.

6. Check the evaluation of importance of parameters for inconsistency.

7. Draw conclusions.

Methodological guidelines for doing the task

1. To determine the number of cases, when parameter *i* is more important than parameter *j* matrix A is constructed (Table 3).

Table 3

Matrix A

Parameter *i*	Parameter *j*			
	A	B	C	D
A	–			
B		–		
C			–	
D				–

2. Grounding the cases of preference of parameter *i* to parameter *j* is done by means of matrix B (Table 4).

Table 4

Matrix B (actual cumulative probability Pij)

Parameter *i*	Parameter *j*			
	A	B	C	D
A	–			
B		–		
C			–	
D				–

3. For determination of the difference of experts' opinions matrix *t* is built (Table 5) using the data from Tables 2 and 7.

Table 5

Matrix t

Parameter *i*	Parameter *j*				Sum of *t* for the line	Average value of $\bar{t}$
	A	B	C	D		
A	0					
B		0				
C			0			
D				0		

4. Evaluation of the indexes of relative importance is given in Table 6.

Table 6

Calculation of parameters

Parameter	Average value *t*	Value of integral function $F(\overline{t})$	Ranking in order of importance (1 is the most important)
A			
B			
C			
D			

5. Checking the consistency of the results.

To check the consistency it necessary to know the difference between the calculated percentage interrelation of the number of cases when parameter *i* is defined as more important than parameter *j*, and the actual number of cases when parameter *i* exceeds parameter *j*. We denote this difference by Δij.

Each Δij is compared with the average value of the absolute values of all of these differences. The resulting value will be considered as a deviation from the expected *Pij*.

Validation should be performed using Table 7.

Table 7

Verification of the data consistency

Δij = $\overline{t_i} - \overline{t_j}$	Estimated cumulative probability *Pij* for given Δij
$\overline{t_A} - \overline{t_B}$	
$\overline{t_A} - \overline{t_C}$	
$\overline{t_A} - \overline{t_D}$	
$\overline{t_B} - \overline{t_C}$	
$\overline{t_B} - \overline{t_D}$	
$\overline{t_C} - \overline{t_D}$	

Assingment 3

Topic "Models of consumer choice and dynamics"

The content of the task. Fluctuations of supply and demand, as well as exchange rates have a different character. To create a stable economic situation, the National Bank carries out certain policies conducive to stabilization of the hryvnia on the currency market.

The objective of the assignment. Determine the scope of intervention of the National Bank to maintain the hryvnia exchange rate at a predetermined level.

The initial data for doing the task are as follows : the volume of demand (*Yd*) and the demand price (*Xd*), the volume of supply (*Ys*) and the supply price (*Xs*) for a certain period (in dollars or euros). You should use at least 40 – 50 observations on the selected currency. These data are collected by students themselves.

The order of doing the task

1. Build and solve a model of supply and demand (*Y*) depending on the currency exchange rate (*X*) by the least squares method.

2. Determine the equilibrium quantity and the equilibrium price of the model supply.

3. Determine the scope of intervention of the Central Bank to maintain the given rate of hryvnia, using the model of demand.

Methodological guidelines for Assignment 3

As econometric models use the model form:

$$Y = B_0 + B_1X, \qquad (9)$$

where *Y* is the volume of demand or supply,

X is the rate of demand or supply of a particular currency.

The rate is supported by the Central Bank.

This task is done by students themselves.

Assingment 4

Topic "Production functions and elasticity"

The content of the task. To develop an optimal strategy for rubber production it is necessary to know on what macroeconomic indicators its volume mostly depends on. Preliminary analysis showed that the production of rubber is most significantly affected by such indicators as car production and income per capita. To establish an econometric relationship between these parameters, data on the production of rubber (*Y*), the production of cars (X_1) and income per capita (X_2) that are listed in the index form (%) to account the inflation were collected (Table 1).

The objective of the assignment: with the production functions of the form

$$Y = B_0 + B_1 X_1 + B_2 X_2, \qquad (10)$$

evaluate the elasticity of impact on the volume of rubber production of such factors as the production of cars (X_1) and income per capita (X_2).

The order of doing the task

1. Build a chart of dependence of *Y* on X_1 and *Y* on X_2 and select models as a function $Y = B_0 + B_1 X_1 + B_2 X_2$.

2. Solve the model by the least squares method.

3. For the model:

a) estimate the impact of all the factors using the multiple correlation coefficient;

b) test the effect of X_1: calculate the coefficient of pair correlation $r_{y.x2}$; determine the local correlation coefficient and verify its materiality;

c) test the effect of X_2: calculate the coefficient of pair correlation $r_{y.x1}$; determine the local correlation coefficient and verify its materiality;

d) estimate the reliability of the model using the coefficient of variation;

d) determine the elasticity of the influence of factors.

4. Draw conclusions.

Table 1

Initial data in percent for Assignment 4

Number of observation	Production of rubber (Y)	Production of cars (X_1)	Income per capita (X_2)
1	2	3	4
1	90.9	128.7	98.7
2	125.2	128.1	106.4
3	94.7	78.7	100.7
4	102.2	79.6	101.2
5	104.4	139.2	102.9
6	90.5	89.4	99.3
7	121.9	140.0	104.7
8	92.3	72.1	102.4
9	100.1	103.2	100.3
10	91.6	68.5	99.3
11	117.3	129.1	102.7
12	93.8	117.0	100.1
13	96.5	81.7	101.4
14	110.6	123.1	103.2

Table 1 (the end)

1	2	3	4
15	101.1	108.6	102.0
16	108.0	110.1	105.3
17	119.0	103.2	106.4
18	108.4	103.4	107.1
19	115.5	106.5	108.3
20	125.4	108.6	110.8

Methodological guidelines for Assignment 4

1. Parameters of the model $Y = B_0 + B_1 X_1 + B_2 X_2$ are determined by the least square method by solving the system of equations:

$$\begin{aligned} \sum Y &= B_0 N + B_1 \sum X_1 + B_2 \sum X_2; \\ \sum Y X_1 &= B_0 \sum X_1 + B_1 \sum X_1^2 + B_2 \sum X_2 X_1; \\ \sum Y X_2 &= B_0 \sum X_2 + B_1 \sum X_1 X_2 + B_2 \sum X_2^2. \end{aligned} \tag{11}$$

To determine parameters B_0, B_1, B_2 we need to preliminarily calculate measures $Dy\,x_j$ and $Dx_i\,x_j$ (j = 1.2, i = 1.2):

$$\begin{aligned} Dyx_1 &= \sum YX_1 - \frac{\sum Y \sum X_1}{N}, \\ Dyx_2 &= \sum YX_2 - \frac{\sum Y \sum X_2}{N}, \\ Dx_1 x_2 &= \sum X_1 X_2 - \frac{\sum X_1 \sum X_2}{N}, \\ Dx_1 x_1 &= \sum X_1^2 - \frac{\sum X_1 \sum X_1}{N}, \\ Dx_2 x_2 &= \sum X_2^2 - \frac{\sum X_2 \sum X_2}{N}, \\ Dyy &= \sum Y^2 - \frac{\sum Y \sum Y}{N}. \end{aligned} \tag{12}$$

$$\overline{Y} = \frac{\sum Y}{N}, \qquad \overline{X_1} = \frac{\sum X_1}{N}, \qquad \overline{X_2} = \frac{\sum X_2}{N}. \tag{13}$$

Then $B_1 = \dfrac{Dyx_1 - \dfrac{Dx_1x_2Dyx_2}{Dx_2x_2}}{Dx_1x_1 - \dfrac{(Dx_1x_2)^2}{Dx_2x_2}}$,

$$B_2 = \frac{Dyx_2 - \dfrac{Dx_1x_2Dyx_1}{Dx_1x_1}}{Dx_2x_2 - \dfrac{(Dx_1x_2)^2}{Dx_1x_1}}, \tag{14}$$

$$B_0 = \overline{Y} - B_1\overline{X_1} - B_2\overline{X_2}.$$

2. Calculation of the coefficient of multiple correlation $Ry.x_1x_2$:

$$Ry.x_1x_2 = \frac{1}{\sigma_y\sqrt{N}}\sqrt{B_1Dyx_1 + B_2Dyx_2}, \tag{15}$$

where $\sigma_y = \sqrt{\dfrac{\sum(Y - \overline{Y})^2}{N}}$.

3. The coefficient of local correlation (r_{local}) and checking its essentiality by the criteria Z' Fisher's (the influence of X_2 is checked):

$r_{local} = r_{yx2.x1} = \sqrt{1 - \dfrac{1 - R_{y.x_1x_2}^{\;2}}{1 - r_{y.x_1}^{\;2}}}$, $r_{y.x1}$ is taken from the pair model;

$$U_p = \frac{Z'}{S_{z'}}, \tag{16}$$

where $S_{z'} = \dfrac{1}{\sqrt{N - 3 - g}}$;

g is the number of residual factors in the model.

4. Calculation of netbook variability $\sigma_{net} = \sigma_y\sqrt{1 - R^2}$ and the coefficient of variation $V = \dfrac{\sigma_{net}}{\overline{Y}}100$.

5. Elasticity of influence of the *i*-th factor is determined by the formula

$$E_{Xi} = Y'\frac{X_i}{Y}, \tag{17}$$

where Y' is the first derivative of the function;

or $E_{xi} = b_i \sum X_i / \sum Y$. (18)

Assignment 5

Topic "Dispersion analysis in the economy"

The content of the task. A large company has a network of shops (kiosks) and is interested in increasing the sales revenue, which depends on many factors: the quality of goods, buyers' habits, the location of shops (kiosks), the presence of nearby competitors, as well as properly designed and conducted marketing policy.

Fluctuations in the level of revenue from the sale of certain goods by weekdays and shops are listed in Table 2. In this task, we investigate a two-factor model: one factor (*A*) is the day of the week, the second (*B*) is the store location.

The objective of the assignment. Estimate the effect of the location of a shop and the day of the week on the fluctuations in revenues from sales of a specific product using the dispersion analysis.

Tables 2 and 3 show the initial data and the variation reference.

The order of doing the task

1. Calculate the variation of the total revenue from sales and by factors.
2. Determine the variance by factors to evaluate the significance of the dispersion using the criterion F-Fisher (Table 4).
3. Calculate the net dispersion of factors and give the results of an economic analysis of the solution.

Guidelines for Assignment 5

The total variability of the process (*V*) in the two-factor analysis (factors *A* and *B*) can be divided into variabilities by reasons:

$$V = V_A + V_B + V_{net},$$

where V_A is variability due to factor *A* (the day of the week);
V_B is variability due to factor *B* (the location of the shop);
V_{net} is net variability due to other factors.

Variability is defined by the following formulas:
total:

$$V = \sum_{I=1}^{n}\sum_{J=1}^{P} X^2_{ij} - \frac{T^2}{N}; \qquad (19)$$

variability due to factor *A:*

$$V_A = \frac{\sum_{J=1}^{P} A_J{}^2}{n} - \frac{T^2}{N}; \tag{20}$$

variability due to factor B:

$$V_B = \frac{\sum_{1}^{n} B_i^2}{P} - \frac{T^2}{N}; \tag{21}$$

netbook variability:

$$V_{net} = V - (V_A + V_B), \tag{22}$$

where $T = \sum A_j = \sum B_i$;

n is variants of factor A (the number of columns);

p is variants of factor B (the number of initial data lines).

Dispersions due to the factors are determined by the following formulas:

total dispersion:

$$S^2 = \frac{V}{np-1}; \tag{23}$$

dispersion of factor *A*:

$$S^2{}_A = \frac{V_A}{p-1}; \tag{24}$$

dispersion of factor *B*:

$$S^2{}_B = \frac{\nu_B}{n-1}; \tag{25}$$

net dispersion:

$$S^2{}_{net} = \frac{\nu_{net}}{(n-1)(p-1)}. \tag{26}$$

net dispersion due to factor *A* equals

$$\sigma^2_A = (S_A^2 - S_{net}^2)\frac{1}{p-1}; \tag{27}$$

net dispersion due to factor *B*:

$$\sigma^2_b = (S_B^2 - S_{net}^2)\frac{1}{n-1}; \tag{28}$$

net: $\sigma^2_{net} = S^2_{net}$;

total net dispersion $\sigma^2 = \sigma^2_A + \sigma^2_b + \sigma^2_{net.}$

Table 1

Deviation of the actual revenue from the planned one, thousand UAH

Days	Numbers of shops										
	1	2	3	4	5	6	7	8	9	10	11
1	1	– 1	0	0	– 1	1	0	1	1	0	1
2	1	– 2	1	1	0	2	1	3	1	– 2	– 1
3	2	0	2	1	– 2	3	2	4	2	– 3	– 2
4	2	– 3	3	2	– 3	4	5	6	3	– 4	– 5
5	4	– 6	5	7	– 4	6	6	7	4	– 5	– 7
6	5	– 6	6	6	– 7	7	7	4	6	– 7	– 6

Table 2

Variants of Assingment 5

Variant	Number of shop	Variant	Number of shop
1	2	3	4
1	1, 2, 3, 4, 5	15	6, 7, 8, 9, 10
2	2, 3, 4, 5, 6	16	7, 8, 9, 10, 11
3	3, 4, 5, 6, 7	17	4, 5, 9, 10, 11
4	1, 2, 4, 5, 6	18	6, 7, 9, 10, 11
5	1, 3, 4, 5, 6	19	6, 8, 9, 10, 11
6	1, 2, 5, 6, 7	20	5, 7, 8, 10, 11
7	1, 2, 4, 5, 8	21	1, 2, 7, 10, 11
8	2, 4, 5, 6, 8	22	5, 7, 8, 9, 10
9	2, 3, 5, 6, 7	23	4, 7, 8, 9, 10

Table 2 (the end)

1	2	3	4
10	1, 2, 3, 6, 7	24	3, 7, 8, 9, 10
11	1, 4, 5, 6, 7	25	2, 7, 8, 9, 10
12	1, 4, 5, 6, 10	26	1,7, 8, 9, 10
13	1, 3, 5, 7, 9	27	1, 3, 5, 8, 10
14	1, 3, 5, 6, 7	28	2, 3, 6, 7, 8

Table 3

Valuation of criterion F-Fisher with probability 0.95

$К_{net}$	The number of degrees of freedom for greater variance dispersion ($К_f$)									
	1	2	3	4	5	11	14	20	30	60
1	2	3	4	5	6	7	8	9	10	11
1	161.4	199.5	215.7	224.6	230.2	243	245	248	250	252
2	18.51	19.00	19.16	19.25	19.3	19.33	19.35	19.37	19.38	19.48
3	10.13	9.55	9.28	9.117	9.014	8.941	8.887	8.845	8.812	8.58
4	7.709	6.944	6.591	6.388	6.256	6.163	6.094	6.041	5.999	26.35
5	6.608	5.786	5.410	5.192	5.050	4.950	4.876	4.818	4.773	4.44
6	5.987	5.143	4.757	4.534	4.387	4.287	4.207	4.147	4.099	9.24
7	5.591	4.737	4.347	4.120	3.972	3.866	3.787	3.726	3.677	3.32
8	5.318	4.459	4.066	3.838	3.688	3.581	3.501	3.438	3.388	5.85
9	5.117	4.257	3.863	3.633	3.482	3.374	3.293	3.230	3.179	2.80
10	4.965	4.103	3.708	3.478	3.326	3.217	3.136	3.072	3.020	4.51
11	4.84	3.982	3.587	3.357	3.204	3.095	3.012	2.948	2.896	2.50
12	4.474	3.885	3.490	3.259	3.106	2.996	2.913	2.849	2.796	3.80
13	4.667	3.806	3.411	3.179	3.025	2.915	2.832	2.767	2.714	2.32
14	4.600	3.739	3.344	3.112	2.958	2.848	2.764	2.699	2.646	3.37
15	4.543	3.682	3.287	3.056	2.901	2.791	2.707	2.641	2.588	2.18
16	4.494	3.634	3.239	3.007	2.852	2.741	2.657	2.591	2.538	3.07
17	4.451	3.592	3.197	2.965	2.810	2.699	2.614	2.548	2.494	2.08
18	4.414	3.555	3.160	2.928	2.773	2.661	2.577	2.510	2.456	2.86
19	4.381	3.522	3.127	2.895	2.74	2.628	2.544	2.477	2.423	2.00
20	4.351	3.493	3.098	2.866	2.711	2.599	2.514	2.447	2.393	2.70
21	4.242	3.385	2.991	2.759	2.603	2.490	2.405	2.337	2.282	1.93
22	8.02	4.87	4.04	3.65	3.40	3.24	3.07	2.88	2.58	2.58
23	4.28	3.03	2.64	2.45	2.32	2.24	2.14	2.4	1.96	1.88
24	7.88	4.76	3.94	3.54	3.30	3.14	2.97	2.78	2.62	2.49

Table 3 (the end)

1	2	3	4	5	6	7	8	9	10	11
25	4.24	2.99	2.60	2.41	2.28	2.20	2.11	2.00	1.92	1.84
26	7.77	4.68	3.86	3.46	3.21	3.05	2.89	2.70	2.54	2.40
27	4.21	2.96	2.57	2.37	2.25	2.16	2.18	1.97	1.88	1.80
28	7.68	4.60	3.79	3.39	3.14	2.98	2.83	2.63	2.45	2.33
29	4.18	2.93	2.54	2.35	2.22	2.14	2.05	1.94	1.85	1.77
30	7.60	4.54	3.73	3.33	3.08	2.92	2.77	2.57	2.41	2.27

Recommended bibliography

Main

1. Доугерти К. Введение в эконометрику / К. Доугерти ; пер. с англ. – М. : Инфра-М, 2001. – 402 с.

2. Замков О. Эконометрические методы в макроэкономическом анализе / О. Замков. – М. : ГУВШЭ, 2001. – 122 с.

3. Клебанова Т. С. Эконометрия : учеб.-метод. пособ. для самостоятельного изучения дисциплины / Т. С. Клебанова, Н. А. Дубровина, Е. В. Раевнева. – Х. : Изд. дом "ИНЖЭК", 2003. – 132 с.

4. Клебанова Т. С. Эконометрия : учеб.-метод. пособ. для самостоятельного изучения дисциплины / Т. С. Клебанова, Н. А. Дубровина, Е. В. Раевнева. – Х. : Изд. дом "ИНЖЭК", 2005. – 156 с.

Additional

5. Кулинич І. О. Економетрія : навч. посіб. / І. О. Кулинич. – Хмельницький : Вид. "Поділля", 1997. – 115 с.

6. Наконечний С. І. Економетрія : підручник / С. І. Наконечний, Т. О. Терещенко, Т. П. Романюк. – Вид. 2-ге, доп. перероб. – К. : КНЕУ, 2000. – 296 с.

7. Эконометрика : учебник / И. И. Елисеева, С. В. Курышева, Т. В. Катеева и др. ; под ред. И. И. Елисеевой. – 2-е изд., перераб. и доп. – М. : Финансы и статистика, 2005. – 576 с.

Information resources

8. State Agency of Ukraine for Investments and Innovations. – Access mode : http://www.in.gov.ua/.

9. State Statistics Committee of Ukraine. – Access mode : http://www.ukrstat.gov.ua/.

10. National Bank of Ukraine. – Access mode: http://www.bank.gov.ua/.

11. Press Service of the National Bank of Ukraine. – Access mode : http://pr.bank.gov.ua/ukr/.

EDUCATIONAL EDITION

Guidelines
to self-study work
on the academic discipline
"ECONOMETRICS"
for full-time students
of training direction 6.030601 "Management"

Compiled by K. **Tonieva**

Editorial director **O. Iastremska**

Editor **Z. Zobova**
Proof-reader **M. Promskiy**

НАВЧАЛЬНЕ ВИДАННЯ

Методичні рекомендації
до самостійної роботи
з навчальної дисципліни
"ЕКОНОМЕТРІЯ"
для студентів напряму підготовки 6.030601 "Менеджмент"
денної форми навчання
(англ. мовою)

Укладач **Тонєва** Кристина Валеріївна

Відповідальний за випуск **Ястремська О. М.**

Редактор **Зобова З. В.**
Коректор **Промський М. Н.**

План 2014 р. Поз. № 97.
Підп. до друку 29.12.2014 р. Формат 60 x 90 1/16. Папір MultiCopy. Друк Riso.
Ум.-друк. арк. 2,0. Обл.-вид. арк. 2,5. Тираж 50 прим. Зам. № 343.

Видавець і виготівник – видавництво ХНЕУ ім. С. Кузнеця, 61166, м. Харків, пр. Леніна, 9-А

Свідоцтво про внесення до Державного реєстру суб'єктів видавничої справи ***Дк № 481 від 13.06.2001 р.***