

Jelena ZVIZDOJEVIĆ, Milica VUKOTIĆ¹

APPLICATION OF STATISTICAL METHODS IN ANALYSIS OF AGRICULTURE - CORRELATION AND REGRESSION ANALYSIS

SUMMARY

The understanding and proper use of statistics is an integral part of everyday business environment. Statistics is a scientific discipline that deals with the collection, analysis and interpretation of the data of the observed phenomena, and it is usually associated with a numerical indicator. As a scientific discipline, it can be divided into descriptive and inferential statistics. Descriptive statistics describes the statistical data, and within it, the most commonly applied time series and graphs. Inferential statistics includes statistical methods to test hypotheses, determines the relationship between variables and predicts trends of the observed phenomena.

This paper analyses statistical methods, and special emphasis in this paper is given to the correlation and regression analysis (multiple regressions). The paper has two objectives. The first is to use statistical methods and statistical analysis, to determine the relationship between selected independent variables and the dependent variable (the value of agricultural production). Another aim of the paper is that using Chow tests the significance of individual factors on agricultural production is tested, by looking at two groups of countries.

Keywords: Statistics, statistical methods, correlation, multiple regressions

INTRODUCTION

Statistics is a science that has its own methodology, which is used in research and analysis of phenomena around us. Phenomena are in fact subjects of statistical research. The subjects of statistical researches are mass phenomena that are inherently variable, and should be viewed in a number of cases and based on these observations there are made conclusions. Therefore, the statistics are often defined as a scientific method of quantitative research of mass phenomena (Zizic *et al.*, 1996).

Variability that is observed in the mass of cases occurs because different causes and circumstances, and they are greater if the circumstances are more complex. Therefore, the statistics makes it possible to identify and explain the general relations of movement, as well as the causal connection between the

¹ Jelena ZVIZDOJEVIĆ, (corresponding author: jelena.zvizdojevic@udg.edu.me), Faculty of International Economics, Finance and Business, PhD Studies, "International Economics", Podgorica, University of Donja Gorica (UDG), Montenegro; Milica VUKOTIĆ Faculty of Information Systems and Technologies, University of Donja Gorica (UDG), Podgorica, Montenegro.

observed phenomena. Today there is no longer any branch of science in which it would not be able to successfully apply the statistical research method. These are primarily economic and social sciences in general, and then agriculture. Thus, the agricultural impact of climatic factors and agricultural practices on the yield of crops explores the observation of statistical samples. Research and analysis of agriculture in this paper is reflected in the definition of the shape of the link between the value of agricultural production and the factors that may affect it, as well as the determination of the research method of phenomena.

The first part of paper emphasizes the implementation of statistical methods, especially correlation and regression analysis, while an overview is provided on the implementation of indices and statistical series that are included in domain of descriptive statistics. In addition to the theoretical framework of correlation and regression analysis, the paper presents the data collected for the agricultural sector, performed correlation analysis, simple regression, and multiple regression, based on which we reach conclusions about the strength and direction of links between dependent and independent variables, as well as predictions about the scaling of the dependent variable, depending on the change of size of independent variables. The second part of the paper refers to the application of CHOW test. CHOW test helps to determine whether the same independent factors have the same effect on the dependent variable, in developed and less developed countries.

MATERIAL AND METHODS

Some of the modern methods of statistical analysis were grounded for more than a century. Statistical methodology consists of four main phases: data collection, presentation, analysis, and interpretation of numerical data/method are a manner to decipher the facts and their interpretation. As a scientific method, the statistics origins simultaneously from Germany and England in the seventeenth century. In Germany, the task of statistics was based primarily on the description, whereas in England began introducing the mathematical processing of data in order to detect regularities in the behaviour of the observed phenomena.

The application of statistical methodology allows, not only to detect the general characteristics of variable phenomena, but also to detect regularities in the tendency of these phenomena. The task of statistics is to make a selection of those statistical methods appropriate to the nature of economic processes and phenomena. Statistical survey method phenomena can be divided into two main groups. First method involves the collection, sorting, presentation and describing the basic characteristics of statistical data series, and it belongs to the domain of descriptive statistical analysis.

The second group consists of methods of statistical analysis dealing with relationship issues, links (correlation), and conclusions on the basis of sample. These methods are included in the area of analytical statistics, although there is no strict underlined line between these two groups of methods.

Statistical analysis is one of the most useful resources of measurement variables for testing hypotheses about the interrelation between two or more variables, the evaluation of the numerical values of functions by which they are expressed, as well as for verifying survey results, comparing actual and assumed status (Pejanovic, 2007). The statistical analysis commonly uses the following: statistical series (series structure and time series); chain and base indices; graph and tabular representations; correlation analysis; regression analysis; econometric models, etc.

Statistical series (time series, structural series), chain and base indices, graphical and tabular views belong to the domain of descriptive statistics. Statistical series represent an ordered set of variations of the characteristics of the observed statistical phenomenon.² According to the type of features they are characterized with, and depending on what they are showing, they are divided into: (i) series of structures; (ii) temporal (chronological) series; (iii) the geographical series. Series of structures show the distribution of statistical units according to the modalities or values of attributes, and depending on the type of features, based on which the grouping of data is made, there are a series of structures with an attributive and a series of structures with numeric feature.³ Temporal (chronological) series are series of statistics that show the variation of the observed phenomena over time. An example of temporal series is presented in the following graph, which also shows a tendency of a decrease both in the total number of agricultural population, and in active agricultural population for period from 2003 to 2010, (Figure 1).

When the variables within the time series are compared in order to determine the relative change we talk about indexes. Indexes may be chained and basic. At the basic index value/size at a given time were taken as the basis for all comparisons with other values in time (e.g. value of agricultural production in 2012, 2011, 2010, 2009, compared to the value of agricultural production in 1986). They show the changes of the absolute values in relation to one fixed size between them, which is called a base (Radulovic, 2003).

² Statistical series consists of two columns. In the first column was given feature by which the groupings were made (attributive or numerical feature, place or time). The second column shows the number of units of certain groups in the series. Depending on the number of features exist simple and complex series? Simple series are those for which data are reported only for one characteristic of the observed phenomena (level of education by gender). Complex series are those in which the data are expressed on several characteristics of the observed phenomena (educational level by sex, municipalities, age, etc.).

³For example, Series with attributive characteristic: sex: male, female; marital status: married, single, divorced, widowed; agricultural land: used (fields, meadows, pastures, ponds), unused etc. Series of numerical characteristics: number of household members, age, etc.

$$I^b = \frac{P_i}{P_o} \times 100$$

I^b –Basic index; P_i - the level of occurrence in a given i -period; P_o –the level of occurrence in the base period.

The base size is denoted by the value of 100 which is compared to other sizes. Data collected for all EU countries, show that the value of agricultural production increased in period from 2003 to 2010, with 2003 as a base year. Thus, the value of agricultural production in 2005 was 43.1 percent higher than 2003, in 2006 the value was higher by 30.9 percent, while in 2010 compared to 2003, the value almost doubled.⁴

In the chain index, value/size of variables of each of the previous period is taken as the base for the next value (e.g. average earnings in September compared to August; the average earnings in August compared to July, etc.).

$$I^l = \frac{P_i}{P(i-1)} \times 100$$

Chain indices show a proportional level of changes from period to period, and using them we can obtain a rate of phenomena that is observed (Radulovic, 2003). Comparing again the value of agricultural production, but its movement from year to year, the data show that in 2004, compared with 2003, the value of agricultural production was higher by 38.1 per cent, while in 2005 compared to year 2004 was higher by 3.6 percent. Also, comparing 2009 with 2008, the value of agricultural production decreased by 21.6 per cent, while in 2010 compared to 2009 it was higher by 6.4 per cent.

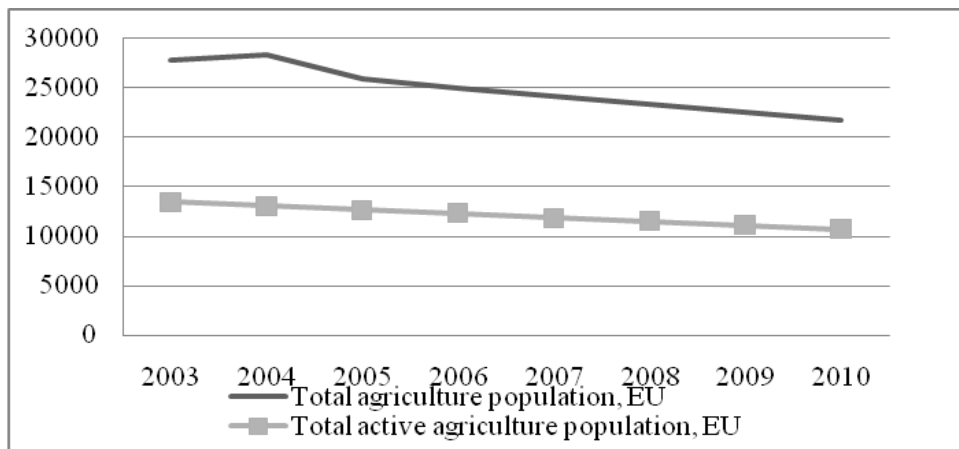


Figure 1: Trends of total and active agricultural population in the EU, (in 1 000)
Source: Food and agriculture organization of United Nations (FAO) database, <http://faostat.fao.org/>

⁴ Data source: <http://faostat3.fao.org/faostat-gateway/go/to/home/E>

The interdependence of phenomena which are observed (at the micro or macro level) mathematically is done mostly by correlation and regression analysis.

RESULTS AND DISCUSSION

Correlation analysis

The correlation represents the relationship of two or more phenomena. Correlation analysis comprises in applying methods that show the strength of the statistical relationships between observed phenomena or statistical variables.⁵ The correlation is measured by correlation coefficient, which ranges from -1 to +1. The correlation coefficient of -1 indicates a perfect negative correlation, which means that the fall of one variable always follows the same fall of other variable. On the other hand, if the coefficient of variation has value +1, then it indicates a perfect positive correlation. The growth of a single variable follows the same growth of other variable.

The correlation coefficient 0 means that there is no linear relationship between two variables. If the correlation coefficient is in the interval from 0 to +1, this means that between the two variables exists a correlation, and that, as a rule, growth of one leads to the growth of the other one.

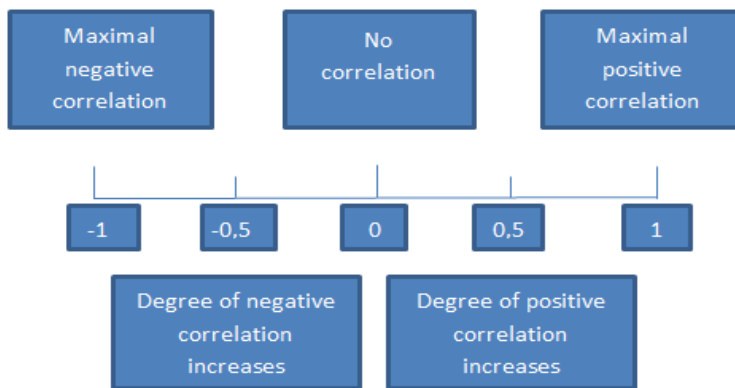


Figure 2. Degrees of positive and negative correlation

The lower the correlation coefficient is, it implies that there are more discrepancies. Correlation coefficients are divided into four groups for the purpose of easier interpretation:

- From 0 to 0.2 –a weak correlation. There is a connection, but it is very weak and it is often not statistically significant;

⁵ "Correlation does not measure causes of link, the causes of interdependence! Only gives a signal that some appearances mutually want or not! Do you remember that correlation does not measure causality between the two phenomena? Measures only similarity, agreement, concordance " (Vukotic V., 2010, p. 55)

- From 0.2 to 0.45 –medium correlation. There is a strong regularity in the movement of the variables that are observed, but there are exceptions too;
- From 0.45 or more indicates a strong correlation between the two variables;
- 1 indicates a perfect correlation.

Test of statistical significance of the correlation coefficient indicates whether coefficient is statistically significant. In this test the statistical significance of H_0 (null hypothesis) is that the coefficient is not statistically significant and that we cannot claim that there is linear relationship between two variables. Alternative hypothesis H_1 is that the coefficient is statistically significant, i.e. there is a linear relationship between two variables.

As an example of strong correlations and significant correlation coefficient is the ratio of the variable value of agricultural production and total active agricultural population.⁶

Table 1. Ratio of the variable value of agricultural production and total active agricultural population

		The value of agricultural production in euro	Total active agricultural population
The value of agricultural production in euro	Pearson Correlation	1	.464**
	Sig. (2-tailed)		.007
	N	32	32
Total active agricultural population in thousands	Pearson Correlation	.464**	1
	Sig. (2-tailed)	.007	
	N	32	32

** . Correlation is significant at the 0.01 level (2-tailed).

As Sig. is less than 0.05, we accept the alternative hypothesis H_1 . The coefficient is statistically significant, and there is a linear relationship between two variables –the value of agricultural production and the total number of active agricultural population.

Correlation analysis is necessary because the regression analysis begins by calculating simple correlation coefficient (bivariate correlations) for all pairs of variables, and all calculations require a linear relationship between the pairs of variables.

Regression analysis⁷ and correlation examine the connection of dependent and independent variables, except that regression analysis in addition to

⁶ Value of agricultural production and the value of active agricultural population, are data series for the EU member states, as well as the values for countries: Iceland, Bosnia and Herzegovina, Serbia, the Former Yugoslav Republic of Macedonia and Montenegro.

establishing the existence of links between two or more variables also achieves prediction of change in dependent variable regards to changes in the independent variables. In other words, by use of the regression analysis the type and intensity of the connections are defined, as well as the quantity of change in one variable in relation to a unit change of the second variable. The regression is analysis of causality. How we can get to the cause of what is now emerging as a consequence?⁸

In regression analysis, there are at least two variables: the criterion (dependent) variable and the predictor (independent) variable. The dependent variable is a variable of interest, or variable that we want to examine, explain, and predict. Independent variable attempts to explain the dependent variable. When there is only one independent variable, then we are talking about a simple regression, and when there are several independent variables, then we are talking about multiple regression.⁹ The goal of regression analysis is to estimate the regression model that minimizes the total distances of the dependent variable from the regression line (Horvat and Mijoc, 2012).

Multiple regression

When the dependence of a phenomenon with two or more independent variables is examined, then we talk about multiple regressions. Name of multiple linear regression means: (i) multiple –there are more independent variables X; (ii) linear –the regression function is linear by the coefficients B1, B2, ... Bn; (iii) regression –the regression function as the best prediction of Y based on Xi, i = 1, ..., n is used.

The goal of multiple regressions is to determine the intensity of the relationships as more independent variables and the dependent variable. One of the assumptions for the use of multiple regression analysis is the existence of a linear dependence between the variables. Multiple regression has the form $Y = A + B1X1 + B3X3 + b2x2 + \dots + BnXn$, where Y is the dependent variable, and the X1, X2, X3, Xn independent variables, and it is represented by Venn diagram (Figure 3).

Understanding the correlation can be easily explained by Venn diagrams. The overlaps that can be observed (interference field) are indicating a correlation

⁷ The name of regression was introduced by statistician Sir Frances Galton (1822-1911) and it means returning back. Examining the links between the height of parents and children he found that senior parents tend to have high child, but not as much as they do. This is the reason for the name of back space, i.e. regression according to the average. The test result is the fact that the child will almost always be high between the height of parents, male children closer to his father's height, but female children closer to the mother's height.

⁸ Vukotic, V. 2010, p. 55

⁹ Independent variables are selected according to the opinion of researchers, i.e. which or what variables the researcher considered the most important variables in describing the dependent variable.

between the variables and indicate which part of one variable explains another variable. What the interfering field is greater, the greater is the level of correlation, or if it is lower, the lower the correlation is.

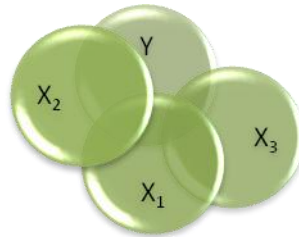


Figure 3. Venn diagram

Multiple regressions provide answers to the following questions:

- How well the independent variables can explain the dependent variable (R^2)?
- What is the relative importance of each independent variable in explaining the change in the dependent variables (beta coefficients), under the condition that there is no significant multi co linearity?
- What change of the dependent variable is expected for each unit change of each independent variable (which shows the simple correlation coefficients)? If X_1 increases by one unit, what is the expected change in Y (response gives B_1) assuming that does not change the impact of other explanatory variables X_2, X_3, \dots, X_n ?

An important goal while creating a regression model is to explain the largest possible percentage of the dependent variable. In simple regression, also multiple regression, the percentage of the explained dependent variable is labelled as R^2 (R -squared) and it tells what percentage of the dependent variable is explained by the included independent variables.

Which variables will be taken as predictor variables or independent variables depends on the aim of survey, as well as on the data availability. When creating the model, it is important to examine the problem of multi-co linearity between independent variables. Namely, if there is a problem of multi-co linearity, then small changes on independent variables can cause large changes in the dependent variable, which does not provide a realistic picture of the model. There are several statistical indicators that examine multi-co linearity: correlation coefficients between the independent variables, Eigen values and condition index, tolerance, VIF, etc. Eigen values shows how many different dimensions among the independent variables, while condition index calculated as the square root relationship largest Eigen values and Eigen values for a given independent variable. If there are multiple independent variables in which the Eigen value close to 0, there is probably a problem of multi-co linearity. It is similar if the condition index

greater than 15, if it is greater than 30, then multi-co linearity is a serious problem in the regression model.

Tolerance and VGF (variance growth factor)¹⁰ are associated indicators, because the one is calculated on the basis of another. Tolerance shows a part of the variance of the independent variable which is not included through other independent variables. The high level of tolerance, over 0.8, means that the variable is relatively uncorrelated with other variables. The low level of tolerance, up to 0.2, is indicating high multi-co linearity and that the variable contributes little to explaining the dependent variable in the model.

Tolerance level: $1 - R^2_i$

VGF is a reciprocal of the tolerance; high values indicate that there is very little contribution to the model. Variance Growth Factor:

$$VGF = \frac{1}{1 - R^2}$$

One of the solutions for the problem of multi-co linearity in the model is the use of stepwise regression. Multiple regressions provide a model that includes all variables with which the analysis was initiated, regardless of their different importance, and also in the case when there is large multi-co linearity. Stepwise regression allows solving the problem of multi-co linearity, with independent variables which are of little importance. The Stepwise regression allows eliminating the variables that overlap with each other, and therefore they have little or no contribution to the prediction accuracy of the model.¹¹

The rating of model is achieved by using analysis of residuals. The residuals represent the difference between the prediction and the original variables, i.e. the part of the dependent variable which failed to be explained by the independent variables. A good model exists when the difference between the sums of squares model and the residual sum of squares (F test) is higher possible (Gegaj, 2011).

Multiple regression model in this study was applied through analysis of the influence of the following independent variables on the dependent the variables – the value of agricultural production: share of agricultural land in total area; the total number of people employed in agriculture; the total number of agricultural population; percentage of agricultural land in relation to the total area; total number of active agricultural population; investment in research and development of agriculture; investment in land development, investment in livestock, investments in agricultural machinery and equipment, investment in plant crops, the total investment, the total number of tourists, share of the cost of

¹⁰ VIF – Variance Inflation Factor

¹¹As a result of the stepwise analysis is obtained new model with mostly smaller number of independent variables, which is just as good as the model containing all independent variables.

food in total expenditure, imports of agricultural products, the consumption of pesticides.¹²

The variables that were shown to be statistically significant in this analysis, i.e. the variables that entered into the model were: percentage of agricultural land in relation to the total area; investment in research and development of agriculture¹³ and the total active population in agriculture. Correlation coefficients of selected variables show that there is a correlation of all three independent variables with the dependent variable. Also, all three variables have a positive impact on the dependent variable, i.e. increase of each individual leads to increased values of the dependent variable.

Examining the existence of autocorrelation is done using the Durbin–Watson test. The autocorrelation implies the existence of correlation between individual observations of random value, i.e. existence of correlation between alternating levels of the observed time series (Vukotic, 1985). Durbin–Watson (DW) statistic has a value of 0 to 4, and if D–W is between 0 and 2, then there is need to do test of positive autocorrelation, while if the DW between 2 and 4, it is necessary to do the test of negative autocorrelation.

For the mentioned model, the D–W is 2.34 ($d = 2.34$), and it is necessary to further investigate the existence of negative autocorrelation. Table value for $n = 31$, $k = 3$, is: $d_l = 1.23$, $d_u = 1.65$. As the d value less than " $4 - d_u$ " concludes that there is no autocorrelation.

Three specified independent variables explain 89 per cent of the variance of the dependent variable (the value of agricultural production) as presented in the following table.

Table 2: Model Summary^b

Model	R	R Square	Adjusted R Square	Std. Error of the Estimate	Durbin–Watson
1	.943 ^a	.890	.878	8512.2	2.341

a. Predictors: (Constant), Total active agricultural population in thousands, investment in research and development of agriculture, millions; the percentage of agricultural land in relation to the total area

b. Dependent Variable: The value of agricultural production, million euro

Considering that R^2 is higher than 80 per cent, it means that the dependent variable was excellently explained by the independent variables. This confirms the sum of squared residuals and the sum of squares of regression, where the sum

¹² Data for all these variables were collected from the official websites of Eurostat, FAO as well as the National Statistical Institutes.

¹³ Series of data for variable investment in research and development of agriculture are from 2008 (because these are the latest published data for this variable), while for the other variables is 2009.

of squares regression is greater than the sum of the squared residuals as presented in the following table.

Table 3: ANOVA test

Model		Sum of Squares	df	Mean Square	F	Sig.
1	Regression	16432883726.16	3	5477627908.722	75.596	.000 ^b
	Residual	2028850663.303	28	72458952.261		
	Total	18461734389.46	31			

a. Dependent Variable: The value of agricultural production, million euros

b. Predictors: (Constant), Total active agricultural population in thousands, investment in research and development of agriculture, millions; the percentage of agricultural land in relation to the total area

Also, in order to test further the model quality, the hypothesis is tested—that at least one independent variable explains the dependent variable.

1. $H_0: \beta_1 = \beta_2 = \beta_3 = 0$

H_1 : at least one independent variable explains the dependent variable

2. $\alpha = 0.05$

3. F (regression) = 75.596; F boundary for regression (3) and residual(28) is $F(3, 28) = 2.95$

Value of calculated Fisher's statistics (F) is greater than the value of Fisher's statistics read from the table. For this reason it is rejected $H_0: \beta_1 = \beta_2 = \beta_3 = 0$, i.e. at least one independent variable explains the dependent variable. Based on the information from the following table, the regression equation has the following form:

$$Y' = -4716,1 + 141,7X_1 + 47,4X_2 + 6,3X_3$$

Presented regression equation shows that if:

- The percentage of agricultural land in relation to total area increased by one unit it leads to an increase in the value of agricultural production for 141.7 units;
- Investment in research and development of agriculture increased by one unit it leads to an increase in the value of agricultural production by 47.4 units and
- The total active agricultural population increased by one unit leads to an increase in the value of agricultural production by 6.4 units.

There are cases even when the independent variables are well chosen, that there is a problem when they depend on each other. In that case it points out the multi co linearity.

Multicollinearity implies such a case/model when among the observed variables there is more than one interdependence¹⁴. There are very rare cases where there is no correlation between the independent variables; however, high correlations can lead to inaccurate assessment of model. In this case, the previously mentioned table indicates that there is no problem of multicollinearity, which concludes on the basis of $VIF < 10$ and $Tolerance > 0.1$.

Table 4: Regression equation

Model	Unstandardized Coefficients		Standardized Coefficients	t	Sig.	Collinearity Statistics	
	B	Std. Error	Beta			Tolerance	VIF
(Constant)	-4716.1	4393.3		1.073	.292		
The percentage of agricultural land in relation to the total area	141.7	102.2	.095	1.387	.176	.843	1.186
Investment in research and development of agriculture, millions	47443	3771	.849	12582	.000	.863	1159
Total active agricultural population in thousands	6398	2964	.150	2158	.040	.813	1230

a. Dependent Variable: The value of agricultural production, million euros

Also, the problem of multicollinearity was tested by using Eigenvalue and Condition Index (not exceeding 15), confirming that there is no problem of multicollinearity.

Further work on the model examines the significance of each variable individually, i.e. raises the question –whether each variable is statistically significant individually. For the mentioned examination/testing the t–test is used.

$$1. H_0: \beta B_1 = 0; B_2 = 0; B_3 = 0$$

$$H_1: B_1 \neq 0; B_2 \neq 0; B_3 \neq 0$$

$$2. \alpha = 0.05$$

$$3. t_1(\text{regression}) = 15,156; t_2(\text{regression}) = 3.560; t_3(\text{regression}) = 3.699;$$

¹⁴ Vukotic, V. (1985), p. 119: „We can extract two types of multi–collinearity. The first type refers to the problem, which is reflected in explaining whether a model that consists of an equation is theoretically based, or that equation must be considered together with a series of regression equations. Multi–collinearity of the second type refers to a model consisting of one equation, when two or more of the independent variables have a high correlation so that practically there are one or more dependencies among some or all of independent parameters.”

Tabular: t (two-tailed, df_{31}) = 2.03

Based on the data presented the hypothesis $H_0: B_1 = 0; B_2 = 0; B_3 = 0$ is rejected, i.e. well above mentioned coefficient sex plant he dependent variable very well.

Histogram, which is shown in the continuation of the work shows that the distribution of residuals significantly deviates from the standard normal errors.

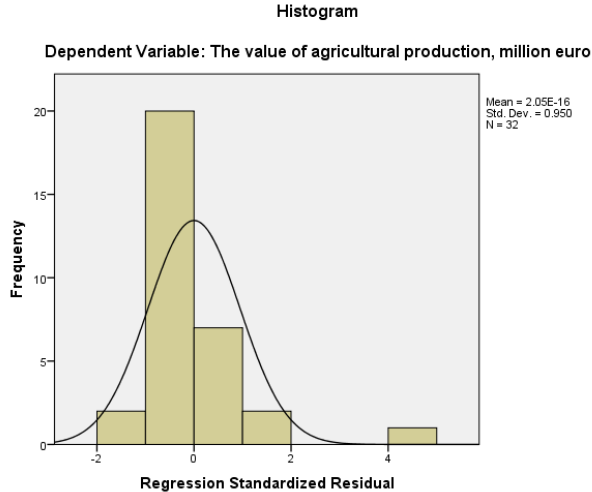


Figure 4. Histogram - Distribution of residuals

One of the assumptions of a good model is the assumption of invariance or homoscedasticity of variance. The assumption is that the variations of random deviations from the regression do not depend on the values of the variables X , i.e. they do not differ in different areas of the value of the independent variable.

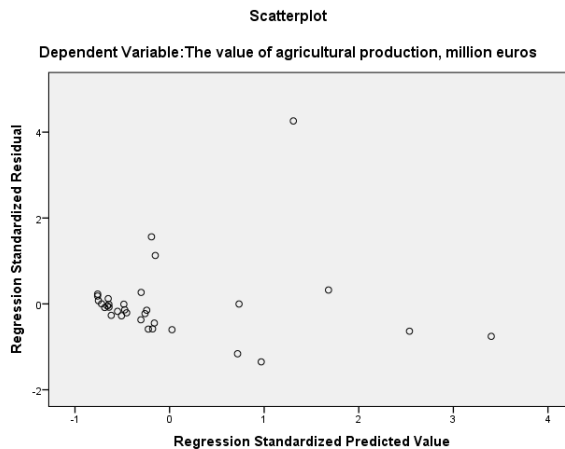


Figure 5. Scatter diagram - Homoscedasticity of model

Homoscedasticity is present in the case where the variance of the errors (residuals) is constant, i.e. Homoscedasticity simply says that the variance of the

errors which of course are random variables is always the same. When the error variance is not constant appears heteroskedasticity. Scatter diagram shows the dispersion of dots, which means that homoscedasticity appears - the variance of errors is constant.

CHOW test

The question that arises, and that Chow test provides an answer is whether the independent variables have the same impact on the dependent variable, in developed and less developed countries. In order to obtain answers to the question about the impact of the same independent variables on the value of agricultural production, countries are divided into two groups on the basis of the gross domestic product (GDP) per capita. The first group comprises countries with a GDP of less than 17 400 euro per capita, while the second group comprises countries with a GDP of more than 17 400 euro per capita. Within these two groups of countries, the hypothesis of equality of regression equations is tested using the Chow test.

Previously tested model has the form: $Y = A + B_1X_1 + B_2X_2 + B_3X_3$
Now that model, i.e. a group of countries which is divided into two groups, examines the following models:

First group: $Y_1 = A_1 + B_{11}X_1 + B_{21}X_2 + B_{31}X_3$

and second group: $Y_2 = A_2 + B_{12}X_1 + B_{22}X_2 + B_{32}X_3$

H_0 hypothesis of CHOW test states that: $A_1 = A_2$; $B_{11} = B_{12}$; $B_{21} = B_{22}$ and $B_{31} = B_{32}$, while H_1 hypothesis is $A_1 \neq A_2$; $B_{11} \neq B_{12}$; $B_{21} \neq B_{22}$ and $B_{31} \neq B_{32}$.

After dividing, we obtained the following regression equation:

$$\text{First: } Y' = -3875.089 + 162.469X_1 + 0.695X_2 + 8.231X_3$$

$$\text{Second: } Y' = -3570.607 + 41.106X_1 + 33.989X_2 + 38.858X_3$$

The sum of the squared residuals obtained by evaluating two separate regressions based on a set of observations n_1 and n_2 , with the one obtained from the initial regression based on all n observations are compared:

$$F(\text{calculated}) = \frac{(a - b)/p}{b/(n - 2p)}$$

Where:

a = sum of squared residuals in the overall model

b = the sum of the sum of squares of residuals of model 1 and model 2

n = number of observations in the overall model

p = number of parameters

$$(a-b) / p = (2,028,850,663 - 976,890,793.5) / 4$$

$$F(\text{calculated}) = 6.19$$

$$b / (n-2p) = 976,890,793.5 / (31 - 2 \times 4)$$

F tabular for $\alpha = 0.05$ is $F(4.23) = 2.80$. As calculated F is higher than F tabular, H_0 is rejected, i.e. variations in the independent variables differently affect the value of the dependent variable in the two groups of countries.

CONCLUSIONS

Statistical method as a bridge between our minds and the reality around us (Vukotic, 2002), is a very complex process, which is reflected in answers to the questions: How statistical data are collected? How do you classify them? What is the manner of their processing? How economic models are created on the basis of these data?

Part of the statistical process is presented in this work through the analysis of statistical methods, with a particular application of correlation and regression analysis on statistics of the agricultural sector. To a model that is presented in this paper, it was reached through a large number of independent variables, which were selected on the basis of theoretical postulates about the factors that influence the value of agricultural production in a country. The model showed that resources of one country, which are reflected in agricultural areas, as well as investment in research and development of agriculture, have the highest impact on the value of agricultural production. This model and others provide a clearer understanding of the reality around us, i.e. the reality in the agriculture.

Also, the models which were obtained, by division into two groups, and then examined the influence of three independent variables on the dependent variable, indicate that the same factors have different intensity of the impact on agricultural production in developed and less developed countries.

REFERENCES

- Vas, A. (2011): "Modelling of Multiple Linear Regression", Master thesis, Faculty of Science, Department of Mathematics and Informatics;
- Vranjes-Soljan, B. (1982): "Quantitative methods in Historical Researches" -reference to the "New Applications of Quantitative Methods in Economic and Social History", University of Zagreb -Institute of Croatian History, Eighth International Congress of Economic Historians- Budapest16-20. August1982; Papers, vol. 15;
- Reinert, E.S. (2006): "The global Economy-How Rich Become Rich and Why the Poor Are Becoming Poorer," Cigoja, Belgrade
- FAO Statistical Yearbook 2012; 2013; <http://faostat.fao.org/>
- Krstic, G., Soskic, D. (2012): "Economic Statistics", Publishing Centre, Faculty of Economics in Belgrade
- Chow, G.C. (1960): „Tests of Equality Between Sets of Coefficients in Two Linear Regressions“, *Econometrica* Vol. 28, No. 3 (Jul., 1960), pp. 591-605, Published by: The Econometric Society; <http://www.jstor.org/stable/1910133>
- Radulovic, J. (2003): "Statistics", second edition, Faculty of Economics, Podgorica
- Horvat, J., Mijoc, J. (2012): "Fundamentals of Statistics"; Ljevak doo, Zagreb
- Petrovic, L.J. (2010): "Theoretical Statistics-Theory of Statistical Conclusion", Publishing Centre, Faculty of Economics in Belgrade

- Zizic, M., Lovric, M., Pavlicic, D. (1996): "Methods of Statistical Analysis", Faculty of Economics, Belgrade
- Gajic, M. (2004): "Agrarian Relations and Agrarian Structure as Factors of Agricultural Development in Serbia", PhD thesis, University of Belgrade, Faculty of Economics
- Gegaj, P. (2011): "Methodology of Social and Political Sciences", Department of Political Science, Podgorica
- Pejanovic, R. (2007): "Some methodological phases and procedures in economy (and agricultural economy)", Yearbook of scientific papers, vol. 31, no. 1, pp. 174-186, University of Novi Sad - Faculty of Agriculture
- Statistical Office of Montenegro, Monthly Statistical Review
- Statistical Office of Montenegro; Statistical Yearbook 2011, 2012
- Statistical Office of Montenegro (2011): Publications of the Census of Agriculture 2010
- The World Bank (2008): World Development Report – Agriculture for Development
- Vukotic, V. (1985): "Statistical Analysis of Labour Productivity", University word, Titograd
- Vukotic, V. (2010): "Principles of Economics -Network of Interdependence", UDG
- Vukotic, V. (2010): "Life and Word" University, Faculty of International Economics, Finance and Business, Donja Gorica