УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ" - ШТИП
ФАКУЛТЕТ ЗА ИНФОРМАТИКА

ISSN:1857-8691

ГОДИШЕН ЗБОРНИК 2014

YEARBOOK 2014

ГОДИНА 3 VOLUME III

GOCE DELCEV UNIVERSITY - STIP
FACULTY OF COMPUTER SCIENCE

УНИВЕРЗИТЕТ „ГОЦЕ ДЕЛЧЕВ“ – ШТИП
ФАКУЛТЕТ ЗА ИНФОРМАТИКА

ГОДИШЕН ЗБОРНИК
2014
YEARBOOK
2014

ГОДИНА 3 | ЈУНИ, 2015 | VOLUME III

GOCE DELCEV UNIVERSITY – STIP
FACULTY OF COMPUTER SCIENCE

ГОДИШЕН ЗБОРНИК
ФАКУЛТЕТ ЗА ИНФОРМАТИКА
YEARBOOK
FACULTY OF COMPUTER SCIENCE

Редакција и администрација	**Address of the editorial office**
Универзитет „Гоце Делчев“-Штип	Goce Delcev University – Stip
Факултет за информатика	Faculty of Computer Science
ул. „Крсте Мисирков“ 10-А	Krste Misirkov 10-A
п. фах 201, 2000 Штип	PO box 201, 2000 Štip,
Р. Македонија	R. of Macedonia

СОДРЖИНА
CONTENT

VALUATION OF FACTORS AFFECTING THE UNEMPLOYMENT RATE OF YOUNG PEOPLE IN REPUBLIC OF MACEDONIA

Tatjana Atanasova Pacemska[1], Elena Mitreva[1]

[1]Faculty of computer science, Goce Delcev University, Stip, Macedonia
(tatjana.pacemska, elena.210145)@ugd.edu.mk

Abstract
The ue nemployment as a social and economic category causes negative changes in the development of every community. The unemployment of young people is a long – term problem which leaves consequences to the individuals, to the community, to the economy and society in general. In this paper, we analyzed the problem of unemployment young population in Republic of Macedonia. The research will be conducted using factor analysis, and the factors which influence to the unemployment of young people will be derived. The results obtained from this analysis identify the key elements which should develop new strategies and polices to resolve the problem of unemployment of young people in Republic of Macedonia.
Key words: *factor analysis, regions, variables, rotation, correlation, mathematical model*

1. Introduction

The reduction of the number of unemployed young people is one of the biggest problems in all the countries in the world and modern economies. The unemployment as one of the main macroeconomic elements is a mover for the economy of the country. The decrease of the unemployment rate leads to economic growth, an increase the overall satisfaction of the people in it. All the governments around the world are trying to find new economic and mathematical models that attempt to discover the factors which lead to unemployment and therefore also to find new ways and models to employ unemployed population.

Factor analysis is one of the most popular multivariate methods which aim is to identify and understand the common features of a number of variables and reduce their numbers based on their connectivity. These features are called common factors. The identified factors represented the basic idea of the important components. For one who makes the survey is much easier to focus on several important characteristics which represent the factors than on all possible characteristics which are considered. Thus, factor analysis provides a good basis for understanding of the most important, essential dimensions or ideas that are related to the investigated appearances.

In this paper we determine and identify the most significant factors that affect to the rate of unemployment of young people in Republic of Macedonia. For this purpose eight variables were selected: regional budget, the unemployment rate, proportion of unemployed in the total population at regional level, the proportion of the population with no education or primary education at regional level, GDP at regional level, young unemployed people up to 29 years, length of duration for job hinting longer than a year, opened available work places. It should be noted that the data analysis of this variables are related to regional level and they are obtained from the National Institute of Statistics and Employment agency in Republic of Macedonia. By applying of factor analysis on these variables we will separate and identify the most important factors which effect and which are the key for reducing of the unemployment rate for young working population in Republic of Macedonia.

2. Material and methods

2.1. Methodological basis of factor analysis

The factor analysis as a set of statistical – mathematical procedures suitable for data analysis of correlation between observed appearances, is proved as useful in all situations where appear many variables that are correlated each other and determine the main sources of covariance between the data [1]. The factor analysis allows to go into details in mutuality and relationships between appearances. One of the main reasons for using factor analysis is reducing of the number of variables [2]. That is the purpose of factor analysis, the interrelationship between the growing number of variables to explain by the less number of fundamental or latent variables, or dimensions, or sources of covariance [3-4]. These fundamental variables are called factors.
The factor analysis is implemented through several steps:

1) Calculation of correlation coefficients between all of the original variables;
2) Calculation of factor loading from the matrix of correlation coefficient;
3) Rotation of the common factors;
4) Evaluation and eventual redefinition of the model;
5) Interpretation of common factors, including choosing the right name;
6) Calculation of the factor scores;

General mathematical model of factor analysis

Let x is a vector which consist of p random variables $x_1, x_2, \dots x_p$ with means $E(x_1)=\mu_1,\ E(x_2)=\mu_2,\ \dots, E(x_p)=\mu_p$. The variables can be represented as a linear function $m(m<p)$ from hypothetical random variables $f_1, f_2, \dots f_m$ which are called common factors. Thus, the general model of factor analysis says:

$$x_1 = \mu_1 + a_{11}f_1 + a_{12}f_2 + \cdots + a_{1m}f_m + e_1 \tag{1}$$
$$x_2 = \mu_2 + a_{21}f_1 + a_{22}f_2 + \cdots + a_{2m}f_m + e_2 \tag{2}$$
$$x_p = \mu_p + a_{p1}f_1 + a_{p2}f_2 + \cdots + a_{pm}f_m + e_p \tag{3}$$

Where a_{jk} is a constant called factor coefficient, and e_j, $j=1,2,\dots p$ is an error. The error e_j is called specific factor, because e_j is specific for the variable x_j, while f_j are common to all variables x_j.

x - Value of variable with arithmetic mean zero and variance one
p - A number of the variables
F - Factors which are mutually dependent
m - A number of the factors
a - Factor loading (coefficient)
e - Specific factor only connected with given variable
In matrix entry, the factor model is:

$$x - \mu = Af + e \tag{4}$$

At first view, the model of factor analysis looks like multiple regression. However, there is a difference in the number of variables that are registered, because in the factor model p deviations $x_1-\mu_1, x_2-\mu_2,\ \dots, x_p-\mu_p$ are expressed through $m+p$ random variables $f_1, f_2, \dots f_m$ and $e_1, e_2, \dots e_p$ which are not registered, in contrast to multiple regression where the independent variables are registered. For every random variable x_k, $k=1,2,\dots p$ a sample with volume n central observations $(x_{k1}, x_{k2}, \dots, x_{kn})$, $k=1,2,\dots p$ and gets data matrix $\boldsymbol{X}$, which applies:

$$\boldsymbol{X}=\boldsymbol{FA'}+\boldsymbol{E} \tag{5}$$

where $\boldsymbol{F}$ is a matrix factor, $\boldsymbol{A}$ is the coefficient matrix, $\boldsymbol{E}$ is the matrix of residual. Thus, any element x_{ij} of matrix $\boldsymbol{X}$ can be calculated as:

$$x_{ij} = \sum_{k=1}^{m} f_{ik}a_{ik} + e_{ij} \tag{6}$$

The general form of the factor model is not standard, because most authors introduce assumptions to get another form. A special case of the model is when $\mathbf{\Phi}=\mathbf{I}$, i.e. is getting orthogonal model in which the factors are independent. By considering of the observed orthogonal method $(\mathbf{\Phi}=\mathbf{I})$ the new model of the factor analysis is:

$$\Sigma = \boldsymbol{AA'} + \Psi \tag{7}$$

where, Σ is covariate matrix, $\boldsymbol{A}$ is matrix of factor loadings and Ψ is diagonal matrix of specific factors.

The interpretation of the results obtained with the factor analysis often applies rotation of the common factors. Let $\boldsymbol{T}$ is an orthogonal matrix with which will be made orthogonal rotation of the common factors. The basic model is:

$$\boldsymbol{x} = \boldsymbol{Af} + \boldsymbol{e} = \boldsymbol{ATT'f} + \boldsymbol{e} = \boldsymbol{A}^{*}\boldsymbol{f}^{*} + \boldsymbol{e} \tag{8}$$

Further, the assumptions are examined with those helps to get the model of factor analysis. Under these assumptions, the new transformed model is:

$$\Sigma = \mathbf{AA}' + \Psi = \mathbf{A}TT'\mathbf{A}' + \Psi = \mathbf{A}^{*}\mathbf{A}^{*\prime} + \Psi \tag{9}$$

This matrix shows that there is no single solution to the matrix A and F, because by using of the orthogonal transformation matrix T can be changed (rotated) in new solution [5].

3. Results and discussion

The data in this paper are taken from the National Institute of Statistics and Employment Agency that refer to the current situation of unemployment and its dependence on certain variables. The data analysis of the study is performed by using of the statistical package SPSS Statistics 20. This statistical package can measure the frequency, percentage representation of data, measures of central tendency (median, standard deviation, dispersion (variance)). The descriptive statistics is shown in Table 1 and Table 2.

Table 1. Descriptive statistics

Descriptive Statistics

	N	Minimum	Maximum	Mean	Std. Deviation	Variance
X1	8	2228035580,00	9219221697,00	3750010369,1250	2327557740,77601	5417525036646298600,000
X2	8	18,80	44,90	29,3500	9,05176	81,934
X3	8	6969,00	20447,00	14392,1250	4628,76690	21425482,982
X4	8	1981,00	5797,00	3783,0000	1145,58382	1312362,286
X5	8	25511000,00	194823000,00	57473500,0000	56094391,75685	3146580786571428,500
X6	8	1811,00	4926,00	3452,5000	1100,55908	1211230,286
X7	8	442,00	1912,00	968,8750	467,81450	218850,411
X8	8	288,00	1896,00	752,0000	506,97422	257022,857
Valid N (listwise)	8					

Table 2. Descriptive statistic – Skewness and Kurtosis (asymmetric and bulge)

Descriptive Statistics

	N	Skewness		Kurtosis	
	Statistic	Statistic	Std. Error	Statistic	Std. Error
X1	8	2,306	,752	5,740	1,481
X2	8	,446	,752	(,501)	1,481
X3	8	(,318)	,752	(,907)	1,481
X4	8	,204	,752	,769	1,481
X5	8	2,712	,752	7,495	1,481
X6	8	,063	,752	(,952)	1,481
X7	8	1,143	,752	1,616	1,481
X8	8	1,932	,752	4,356	1,481
Valid N (listwise)	8				

In the statistical data analysis the method of factor analysis was applied. As input to the factor analysis were used 8 variables.
Variables:
X1 – Regional budget
X2 – The unemployment rate
X3 - Proportion of the unemployed in the total population at regional level
X4 - The proportion of the population with no education or primary education at regional level
X5 - GDP at regional level
X6 - Young unemployed people to 29 years

X7 - Length of duration for job hinting longer than a year
X8 - Opened available work places.

After selecting the input variables and standardization of their values, it is necessary to examine the reasonableness of the application of the analysis, to make a decision which method of factor analysis will be used. The calculation of factor analysis is based on correlation matrix which contain coefficients of simple linear correlation of each pair of variables. Based on the correlation matrix, the groups of related variables are identified. If in the correlation matrix the correlated variables which forming one or more groups are discernible, then there is a common factor of variables of a given group. Table 3 shows the correlation matrix of variables typical for the unemployment.

Table 3. Correlation matrix

	Budget	*Unemployment rate*	*Number of unemployed*	*A low level of education*	*GDP*	*Unemployed to 29 years*	*Duration of job seeking*	*Available work places*
Budget	1,00							
Unemployment rate	0,01	1,00						
Number of unemployed	0,72	0,29	1,00					
A low level of education	0,23	0,36	0,73	1,00				
GDP	0,95	-0,11	0,53	-0,07	1,00			
Unemployed to 29 years	0,35	0,26	0,87	0,90	0,09	1,00		
Duration of job seeking	0,40	0,49	0,62	0,60	0,21	0,60	1,00	
Available work places	0,82	-0,24	0,42	-0,23	0,93	-0,01	0,02	1,00

The coefficients of the linear correlation in the correlation matrix have different values and different sign. It could be noticed high positive correlation form (0.95) between *GDP at regional level* and *regional budget*. Slightly weaker positive correlation from (0.36) exists between *the proportion of the population with no education or primary education at regional level* and *the unemployment rate*. Weak negative correlation from (-0.23) exists between *opened available work places* and *the proportion of the population with no education or primary education at regional level*.

The procedure of the factor analysis is made by the method of principal components and the varimax rotation is applied on the separated factors. So, the number of the factors will be determined. The eigenvalue criteria of the factor are applied, i.e. the number of factors is the number of eigenvalues greater than 1. Table 4 shows that eigenvalue of first factor is 3.944, of second factor is 2.625 and the eigenvalue of third factor is 0.864. Consequently, from the application of eigenvalue criteria with eigenvalue greater than 1, two factors are separated. Based on the shown percentage of variance of each factor, we obtain that each factor explains less variance compared to the previous or first factor explains 49.29% of total variance and the second factor explains 32.81% of total variance. The separation of factors stops when the next factor explains less proportion of variance. The result is cumulative percentage of variance, and the two factors explains 82.11% of total variance.

Table 4. Number of factors for extraction

Total Variance Explained

Component	Initial Eigenvalues			Rotation Sums of Squared Loadings		
	Total	% of Variance	Cumulative %	Total	% of Variance	Cumulative %
1	3,944	49,294	49,294	3,391	42,393	42,393
2	2,625	32,812	82,106	3,177	39,713	82,106
3	,864	10,802	92,909			
4	,380	4,748	97,657			
5	,134	1,681	99,338			
6	,043	,533	99,871			
7	,010	,129	100,000			
8	9,985E-018	1,248E-016	100,000			

Extraction Method: Principal Component Analysis.

Because the number of factors is determinate, it is necessary to interpret it. The interpretation of separated factors is based on the matrix of factor structure. The matrix of factor structure contains factor loadings which present the correlation coefficients between factors and variables, suggesting the importance of variables for each factor. Because the matrix of factor structure doesn't have simple structure (some factors are common for multiple factors) interpretation is difficult and begin the process of rotation of factors. Varimax rotation is applied. Varimax rotation results with a simplification of the columns in the factor matrix, i.e. simplifying the factors. Table 5 shows the matrix of factor structure after the varimax rotation of factors.

Table 5. Rotated factor loadings

Rotated Component Matrix[a]

	Component	
	1	2
X1	,291	,936
X2	,571	(,207)
X3	,800	,551
X4	,932	(,061)
X5	,011	,989
X6	,909	,132
X7	,788	,169
X8	(,158)	,963

Extraction Method: Principal Component Analysis.
Rotation Method: Varimax with Kaiser Normalization.
a. Rotation converged in 3 iterations.

From the matrix of factor structure, after the rotation of factors, it could be seen that each variable has significant factor loading with only one factor. That structure allows better interpretation of factors concerning to the initial factor matrix. The interpretation of the factors is obtained from identification of variables which have high absolute factor loading on the same factor.

The first factor has the highest factor loading at the variables: the proportion of the population with no education or primary education at regional level and young unemployed people to 29 years.

The second factor has the highest factor loading at the variables: regional budget, GDP at regional level and opened available work places.

Setting hypothesis

H_0 : There is no statistically significant difference between budget and the unemployment rate as variables that affect unemployment

H_1 : There is a statistically significant difference between budget and the unemployment rate as variables that affect unemployment

Table 6. T - Statistics

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	X1 - X2	3750010339,77500	2327557740,71149	822915931,03017	1804123372,23901	5695897307,31099	4,557	7	,003

Table 6 shows the relationships between the two variables X1 and X2, i.e. the relationship between budget and the unemployment rate as variables that affect to the rate of unemployment. According t – test, which is shown in Table 6, it can be seen that the value is $t = 4,557$. In this case, it is evident that the significant level is less than 0.05, i.e. significant level is $p < 0,05$. Thus, the table shows that the significance level is $p = 0,003$, which means that $p = 0,003 < 0,05$. We can conclude that the null hypothesis H_0 is rejected, which means that the alternative hypothesis H_1 is accepted, i.e. that there is statistically significant difference between budget and the unemployment rate as variables that affect unemployment.

Table7. Analyze of variance

Anova: Single Factor

SUMMARY

Groups	*Count*	*Sum*	*Average*	*Variance*
Column 1	8	30000082953	3750010369	5,41753E+18
Column 2	8	234,8	29,35	81,93428571
Column 3	8	115137	14392,125	21425482,98
Column 4	8	30264	3783	1312362,286
Column 5	8	459788000	57473500	3,14658E+15
Column 6	8	27620	3452,5	1211230,286
Column 7	8	7751	968,875	218850,4107
Column 8	8	6016	752	257022,8571

ANOVA

Source of Variation	*SS*	*df*	*MS*	*F*	*P-value*	*F crit*
Between Groups	9,80299E+19	7	1,40043E+19	20,66795793	1,99785E-13	2,178155555
Within Groups	3,79447E+19	56	6,77584E+17			
Total	1,35975E+20	63				

ANOVA is statistical method which is used to determine if there are differences between the means of the variables. In this case, shown in Table 7, $F = 20.66 > F_{crit} = 2.17$, so we rejected the null hypothesis and accept the alternative hypothesis.

4. Concluding remarks

The obtained data in the paper identify the main factors or reasons which lead to the increasing of the regional unemployment rate of young population in Republic of Macedonia. In order to be obtained these results, the method of factor analysis was used. On this way, were selected only those variables which are considered like the most affecting factors to the unemployment problem. By the application of factor analysis it is obtained that from eight initially selected variables; only two variables are significant and important for the unemployment of young people, the regional budget and the unemployment rate. The significant difference between these two variables is considered by testing of the hypothesis. By applying of the method ANOVA and t – test, we conclude that there is a significant difference between these two variables which affect to the regional unemployment.

References

[1] Fulgosi, A. (1988). *Faktorska analiza*. Treće dopunjeno izdanje. Školska knjiga. Zagreb

[2] Gauch, H.G., (2003). *Scientific Method in Practice*. Cambridge: Cambridge University Press

[3] Harman, H.H., (1967). *Modern factor analysis*. Chicago: University of Chicago Press

[4] Kim, J.O., Mueller, W.C. (1986). *Factor Analysis: Statistical Methods and Practical Issues.* Beverly Hills/London: Sage Publications

[5] Kovačić, J. Z., (1994). *Multivarijaciona analiza. Faktorska analiza*. Univerzitet u Beogradu. Beograd (215-223).

[6] Richard, A.R., Joreskog, K.G. (1996). *Applied Factor Analysis in the Natural Sciences*. Cambridge University Press.

[7] Агенција за вработување на Република Македонија (2014). Превземено 2014. http://www.avrm.gov.mk

[8] Државен завод за статистика (2014). *Анкета на работна сила*. Превземено 2014. http://www.stat.gov.mk