

**Iwona Markowicz**  
**iwmar@szafir.univ.szczecin.pl**  
**Beata Stolorz**  
**stolorz@interia.pl**

**Department of Econometrics and Statistics**  
**University of Szczecin**

## **LEVEL OF EDUCATION AS ONE OF DETERMINANTS OF LABOUR SEEKING TIME**

### **Abstract**

In the economy where unemployment exists, proper level of education of an individual seeking job, suitable to employers' demands has great significance. Human capital is one of factors of economic development and level of education of society decides about its quality. Knowledge and its gaining are processes connected with the labour market at the stages of both performing work and searching it. The main objective of the article is to estimate the influence of level of education to unemployment time, especially considering interaction of level of education with gender and age on the basis of study of the characteristics of people registered in Local Labour Office in Szczecin who found workplaces in the last quarter of 2006. The aim of the article is also to present the possibility to apply methods of survival analysis to study the phenomenon of unemployment.

The study was carried out on the basis of data concerning registered unemployed provided by Local Labour Office in Szczecin. The cohorts are the individuals withdrawn from the register of the unemployed in the last quarter of year 2006. In this case survival function describes the time from the moment the individuals are registered as unemployed in Local Labour Office (PUP) until they find employment. The main objective is to establish if gender, level of education or age of the unemployed determines the length of their unemployment period until they find a job, and to find the correlation between the level of education, gender and age of the individuals and their influence for the length of the mentioned period.

**Key words:** semiparametric and nonparametric methods, model construction and estimation, unemployment: models, duration, incidence and job search, regional labour markets

### **Introduction**

In the economy where unemployment exists, proper level of education of an individual seeking job, suitable to employers' demands has great significance. Human capital is one of factors of economic development and level of education of society decides about its quality<sup>1</sup>. Knowledge and its gaining are processes connected with the labour market at the stages of both performing work and searching it. The main objective of the article is to estimate the influence of level of education to unemployment time, especially considering interaction of level of education with gender and age on the basis of study of the characteristics of people registered in Local Labour Office in Szczecin who found workplaces in the last quarter of

---

<sup>1</sup> See Kopycińska (ed. 2006).

2006.

The aim of the article is also to present the possibility to apply methods of survival analysis to study the phenomenon of unemployment. The survival analysis is a series of statistical procedures for which time between specified events or time of the event is the random variable. The event effects in the transition of the unit from one state to another (e.g. death of a person, breakdown of a device, fall of a firm<sup>2</sup>, unregistering from Local Labour Office). The period between the initial state and the moment of occurrence of the event is called the survival time. The survival function is determined by calculating the probability that an unit will survive through consecutive values of time  $t$ . These functions determined for two or more samples can be compared. An example of the random variable in this case is registering time of unemployment in Local Labour Office. The influence of gender, age and level of education of unemployed individual for labour seeking time has been studied as well as correlations between these variables.

### Characteristics of data used in the study

The study was carried out on the basis of data concerning registered unemployed provided by Local Labour Office in Szczecin. The cohorts are the individuals withdrawn from the register of the unemployed in the last quarter of year 2006. In this case survival function describes the time from the moment the individuals are registered as unemployed in Local Labour Office (PUP) until they find employment. The main objective is to establish if gender, level of education or age of the unemployed determines the length of their unemployment period until they find a job, and to find the correlation between the level of education, gender and age of the individuals and their influence for the length of the mentioned period. Individuals withdrawn from the register for other reasons are treated as censored observations. To make calculations, the authors used the classification used by Local Labour Office distinguished according to level of education. These groups were assigned numbers for study purposes as shown in table 1.

Tab. 1. Group numbering according to the level of education and gender

Characteristic	Group	Number
Education	• lack or incomplete primary	1
	• primary,	
	• lower secondary,	
	• basic vocational,	2
	• general secondary,	3
Education	• 4-year vocational secondary,	4
	• vocational secondary,	
	• post-secondary,	
Education	• higher (tertiary) or college.	5
	Gender	• males
	• females	1

Source: own study based on Szczecin Local Labour Office classification.

The data of 5074 individuals in total were used in the analysis. Quantitative analysis of groups

<sup>2</sup> See Markowicz, Stolorz (2006a, 2006b).

under study is shown in table 2.

Tab.2. Quantitative characteristic of groups under study

Characteristic	Groups	Size of group	Censored observations		Uncensored observations	
			numbers	percentage	numbers	percentage
Gender	Females	2469	1277	51,72	1192	48,28
	Males	2605	1652	63,42	953	36,58
Age	Continuous variable	5074	2929	57,73	2145	42,27
Education	<ul style="list-style-type: none"> <li>• lack or incomplete primary</li> <li>• primary,</li> <li>• lower secondary,</li> </ul>	1755	1226	69,86	529	30,14
	<ul style="list-style-type: none"> <li>• basic vocational,</li> </ul>	1055	628	59,53	427	40,47
	<ul style="list-style-type: none"> <li>• general secondary,</li> </ul>	488	273	55,94	215	44,06
	<ul style="list-style-type: none"> <li>• 4-year vocational secondary,</li> <li>• vocational secondary,</li> <li>• post-secondary,</li> </ul>	941	485	51,54	456	48,46
	<ul style="list-style-type: none"> <li>• higher (tertiary) or college.</li> </ul>	835	317	37,96	518	62,04

Source: own study based on Szczecin Local Labour Office classification.

### Gehan's nonparametric test

Nonparametric models for a random variable of survival time are used if an analytic nature of distribution is unknown. The theory of estimation of certain functions in this case is very complex and extensive. An example of these is Kaplan-Meier Product-Limit-Estimation method<sup>3</sup> used for the construction of life tables. This is a nonparametric method, including the occurrence of censored observations, and thus, the incomplete data. In this case there is no need to construct the intervals for time variable, but only to put in order the episodes according to time duration. Every point in time when at least one event occurred is ascribed risk value.

The population under study can be divided into groups on the basis of studied characteristics. Then it is possible to estimate survival function for each of these groups and to study the importance of group differences.

Survival times can be compared in two or more samples. Since their distribution is unknown, nonparametric test is needed. These tests are based on rank order of survival times. In the case of two groups the following tests can be applied: Gehan's generalization of Wilcoxon test (Gehan, 1965a, Gehan, 1965b), Cox-Mantel's test, F. Cox's test, log-rank test and Peto and Peto generalization of Wilcoxon test (Namboodiri, Suchindran, 1987, pp. 71-91, Cox, Oakes,

<sup>3</sup> See Hosmer, Lemeshow (1999), pp. 28-39.

1984, pp. 123-125). Tests to compare a number of groups are also available. Unfortunately, there are no universally accepted methods of test selection in a given situation. It all depends on the number of samples, the occurrence of censored data, the variables' distribution (Lawless, 1982, pp. 425-427). In most tests calculated statistical data for a big sample drift asymptotically towards normal distribution. This fact is used to test statistical significance of differences between samples. Most tests give reliable results only for big samples, whereas effectiveness of tests for smaller samples is less known. The authors of the article used Gehan's test for two samples<sup>4</sup> to verify the hypothesis of the equality of survival functions  $H_0: S_1(t) = S_2(t)$ . To compare several samples the test in Statistica 6 set, equivalent to Gehan's test in case of two samples only, was used.

The method described above is used for survival and failure time analysis in case some data are censored<sup>5</sup>. Hence, it is widely used in demography, biology, social sciences, engineering, technology, and, in particular, in medicine (Hozer, ed., 2002) where patients under study cannot be contacted any longer.

### **Analysis of labour seeking time in relation to the level of education**

The starting point in the conducted study was to set Kaplan-Meier survival function. Then, using Gehan's test, the hypothesis about identical survival functions for the unemployed in groups under study was verified (fig.1). The test was carried out for all the groups of a given characteristic and then for particular group pairs. To do this Statistica 6 program was used. It enables you to estimate survival function for each of the subgroups and to study the significance of differences between them. Setting the level of significance at 0,05 it can be stated that subgroups of the unemployed differ in their survival time. This characteristic affect the withdrawal process from the unemployed registered in Szczecin Local Labour Office. The results are shown in table 3.

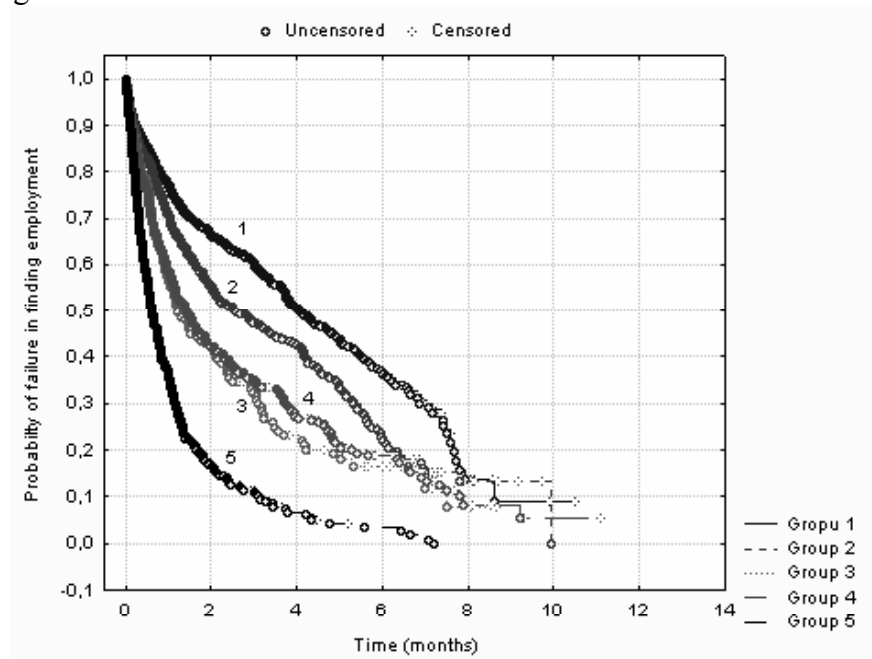
The test allowing the comparison of survival functions for many samples carried out for age groups showed significant differences between them (fig.2). After accurate study of particular subgroups (table 5) it turns out there are no significant differences between groups 4 and 6 and between groups 5 and 6. However, there is an important difference between groups 4 and 5. It seems that the main cause of this fact are the small numbers of group 6 where only 11 uncensored and 41 censored observations are present. Reaching retirement age by the unemployed registered in Szczecin Local Labour Office in this case often causes censoring. Studying the curves of survival for particular age groups it is easy to notice that the probability of finding employment decreases with age. Namely, individuals aged 18-25 accepted jobs offered by Szczecin Local Labour Office much quicker than those aged 60-65, where the situation was the worst.

---

<sup>4</sup> Based on Namboodiri, Suchindran, 1987, p. 75.

<sup>5</sup> Censored observations are those for which the exact survival time is unknown. The reason could be lack of occurrence of an event before the end of study or lost contact with an individual under study. In literature the term "censored" data is used, Domański, Pruska, 2000, pp. 203-204.

Fig. 1 Kaplan-Meier survival function - the risk of not finding employment by the unemployed according to the criterion of education



Source: own compilation with the application of Statistica program.

Table 3. Gehan's test results for groups selected by education

Groups (education)	Gehan's test result	Probability $p$ where differences are significant
1 / 2 / 3 / 4 / 5	330,888	0,00000
1 / 2	3,488620	0,00049
1 / 3	6,658588	0,00000
1 / 4	8,386546	0,00000
1 / 5	17,04635	0,00000
2 / 3	3,963809	0,00000
2 / 4	4,771897	0,00000
2 / 5	13,05308	0,00007
3 / 4	0,0799918	0,93624
3 / 5	6,705415	0,00000
4 / 5	8,432880	0,00000

Source: own calculations with the application of Statistica program.

The analysis of groups isolated on the basis of education level showed significant differences between them (table 6). After the study of group pairs it can be seen that Gehan's test showed no significant differences between group 3 and 4, namely, no differences between individuals with general secondary education and vocational secondary education. The analysis of graph 3 leads to more conclusions. Finding a job is closely related to the level of education. The unemployed with the higher (tertiary) education have the best chances of finding employment and those with lower secondary education – the least chances.

### Cox's proportional hazard model

Varied regression models can not be used to study the influence of potential variables for labour seeking time due to the facts of unawareness with dependent variable distribution and existence of censored observations.

Cox's proportional hazard model assumes that hazard function is independent variable function (Frątczak, Gach-Ciepiela, Babiker, 2005, pp. 37-38) and no other assumptions about its nature or form are considered. This model can be described as follows (Stanisz, 2005, p. 355):

$$h(t : x_1, x_2, \dots, x_n) = h_0(t) e^{\sum_{i=1}^n \beta_i x_i} = h_0(t) \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_n x_n) \quad (1)$$

where:

$h(t : x_1, x_2, \dots, x_n)$  – outcome hazard (chance) at given  $n$  independent variables  $x_1, x_2, \dots, x_n$  and appropriate survival time

$h_0(t)$  – hazard of reference or hazard's line zero,

$\beta_1, \beta_2, \dots, \beta_n$  – model factors,

$t$  – time of observation.

Hazard of reference  $h_0(t)$  has value for which all independent variables are equal to zero.

If variables being the products of two other variables present in the model are included as independent variables in this model, Cox's interaction model<sup>6</sup> will be obtained. Product of hazards between classes of studied characteristics can be set according the following formula:

$$HR = \exp(\beta_1 + \beta_3 x_2) , \quad (2)$$

where:

$x_2$  – independent variable of the model,

$\beta_1$  – model's factor corresponding to variable  $x_1$ ,

$\beta_3$  – model's factor corresponding to variable  $x_1 \times x_2$ .

### Level of education versus chance to find a job

Level of education is essential determinant of labour seeking time, although not the only one. Very significant for employers are gender and age of employees. It is interesting to find out the correlations between these variables.

#### Unifactoral analysis

Unifactoral Cox's model has been used to determine the significance and following this – the possibility of inclusion of studied data to multifactoral model. Gender and level of education are categorized variables. The first one is divided by nature to two groups and the other one, considering Local Labour Office classification, has been divided into five categories. The age is constant characteristic. In table 4 results of unifactoral Cox's models estimations are shown.

<sup>6</sup> See. Kleinbaum, Klein pp.100-103.

Table 4. Results of unifactoral Cox's models parameters estimation for studied characteristics

Variable	Parameter beta	Index beta	Value <i>p</i>
Education	0,294394	1,342313	0,000000
Gender	0,170428	1,185812	0,000089
Age	-0,039938	0,960850	0,000000

Source: own study with the application of Statistica program.

Value  $p(\leq 00,5)$  indicates statistically significant influence of particular variables for labour seeking time. These variables can be therefore used in further analysis.

### Multifactoral analysis

Multifactoral Cox's proportional hazard model enables estimation of simultaneous influence of many variables for event occurrence time. Models with interaction will be made for study the correlation between the level of education and age of unemployed individual.

#### 1. Model with categorized variables: level of education and gender

Table 5. Results of Cox's regression model estimation – categorized variables: gender and level of education ( $\chi^2 = 401,384, p = 0,0000$ )

Variables	Parameter beta	Standard error	Value <i>t</i>	Index beta	Wald statistics	<i>p</i>
Education	0,296009	0,015031	19,69362	1,344483	387,8388	0,000000
Gender	-0,021157	0,044665	-0,47367	0,979066	0,2244	0,635738

Source: own study with the application of Statistica program.

In the model (table 5) encoding used in table 1 has been used for level of education and gender. The results can be interpreted as follows:

- together with growth of level of education by one category relative chance to find a job rises average 34,4%,
- chance to find a job by a woman is average 2,1 % less than by a man (significant variable at level 0,635738).

#### 2. Model with categorized variables: level of education, gender and interaction between them.

Table 6. Results of Cox's regression model estimation – categorized variables: gender, level of education and interaction between them. ( $\chi^2 = 402,247, p = 0,0000$ )

Variables	Parameter beta	Standard error	Value <i>t</i>	Index beta	Wald statistics	<i>p</i>
Education	0,311206	0,022161	14,04285	1,365071	197,2016	0,000000
Gender	0,059619	0,097653	0,61052	1,061432	0,37270	0,541522
Education × Gender	-0,027519	0,029585	-0,93014	0,972857	0,8652	0,352304

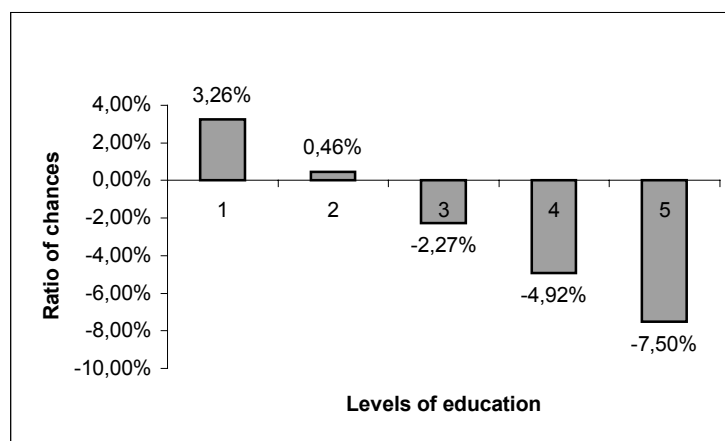
Source: own study with the application of Statistica program.

In the model (table 6) both gender and level of education are considered as categorized variables and encoding presented in table 1 has been used.

The results (Fig.2) can be interpreted as follows:

- in the first level the chance to find a job by a woman is for 3,26% higher than by a man,
- in the second level chances are similar for both genders,
- in the third, fourth and fifth level chance to find a job by a woman is less respectively for 2,27%, 4,92% and 7,5%.

Fig. 2 Ratio of chances (HR) of finding a job for women and men for various levels of education



Source: own study.

Table 7. Ratio of chances (*HR*) to find a job together with growth of level of education for women and men

Gender	Ratio of chances: Women/Men
Women	1,328017
Men	1,365070

Source: own study.

The results can be interpreted as follows:

- in case of women growth of level of education effects with increase of chances to find a job for 32,8%,
- for men – growth of level of education effects with increase of chances to find a job for 36,5%.

### 3. Model with categorized variable: level of education and continuous variable: age

In the table 8 model level of education is categorized variable and age is continuous variable. Encoding presented in table 1 has been used. The results can be interpreted as follows:

- together with growth of level of education for one category chance to find a job by un-



- employed in any age increases average for 27,3%,
- together with growth of age for one year chance to find a job by unemployed decreases average for 3,3%.

Table 8. Results of Cox's regression model estimation – categorized variable: level of education and continuous variable: age ( $\chi^2 = 670,132, p = 0,0000$ )

Variables	Parameter beta	Standard error	Value <i>t</i>	Index beta	Wald statistics	<i>p</i>
Education	0,241503	0,014827	16,1882	1,273161	265,5592	0,000000
Age	-0,034029	0,002146	-15,8606	-0,966543	251,5592	0,000000

Source: own study with the application of Statistica program.

4. Model with categorized variable: level of education, continuous variable: age and interaction between them.

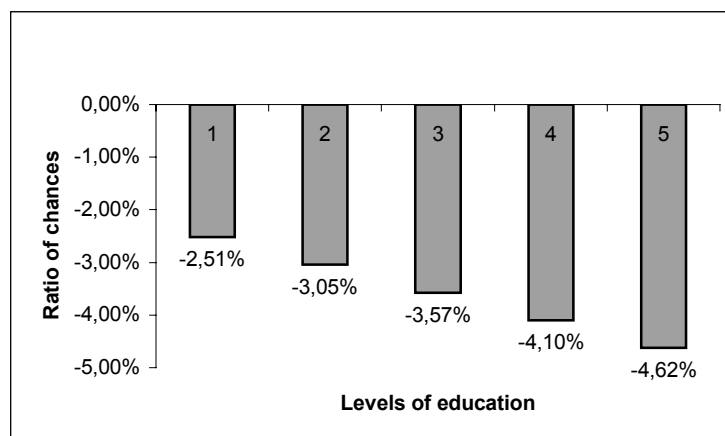
Table 9. Results of Cox's regression model estimation – categorized variable: level of education, continuous variable: age and interaction between them ( $\chi^2 = 684,255, p = 0,0000$ )

Variables	Parameter beta	Standard error	Value <i>t</i>	Index beta	Wald statistics	<i>p</i>
Education	0,423101	0,050698	8,34557	1,526689	69,64859	0,000000
Age	-0,020019	0,004265	-4,69323	0,980180	22,02636	0,000003
Education×Age	-0,005452	0,001458	-3,74096	0,994562	13,99480	0,000184

Source: own study with the application of Statistica program.

In table 9 model both gender and level of education are considered as categorized variables and encoding presented in table 1 has been used.

Fig. 3. Ratio of chances (*HR*) to find a job for unitary age growth for various education levels

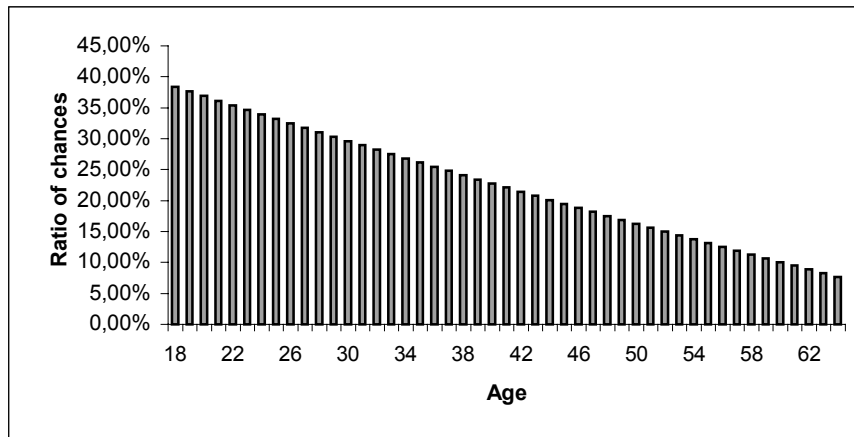


Source: own study.

Following conclusions can be drawn:

- in case of age growth for one year chances to find a job decrease in all education levels respectively for 2,51%, 3,05%, 3,57%, 4,10%, 4,62%,
- these chances decrease more and more quickly, i.e. the least decrease is noted in the first level, the largest in the case of the highest level of education.

Fig. 4. Ratio of chances to find a job together with growth of level of education for various age groups



Source: own study

Following conclusions can be drawn from figure 4:

- in case of growth of education level for one category, chance to be employed increase in each age,
- this increase is becoming less together with age, i.e. the largest is for young people (for 18 years old it is 38,4%) and the least is for older people (for 64 years old it is 7,8%).

### Alterations of relative chance to find a job together with growth of level of education

There is possibility to estimate the relative chance for transition from one group to another together with growth of level of education using the binary encoding of variables variants for estimation of Cox's regression model. Appropriate beta parameter value is assigned as relation of proportional hazard function for compared categories of given variable, considering constancy of other explanatory variables.

Alterations of the relative chance to find a job together with growth of level of education have been calculated on the basis of the formula:

$$Educ(i, j) = \frac{Educ(i, 1)}{Educ(j, 1)} = \frac{e^{\beta_i}}{e^{\beta_j}} = e^{\beta_i - \beta_j}, \text{ for } i, j = 1, \dots, 5, \quad (6)$$

And results are shown in table 10.

Table 10. Relative chance at transition of unemployed person to another group according to level of education

Compared groups	Parameter beta	Index beta	Value <i>p</i>
<i>Educ</i> (2,1)	0,28140	1,32498	32,5%
<i>Educ</i> (3,2)	0,20620	1,22900	22,9%
<i>Educ</i> (4,3)	0,07788	1,08099	8,1%
<i>Educ</i> (5,4)	0,60519	1,83160	83,2%

Source: own study with the application of Statistica program.

Together with growth of level of education chances to find a job by unemployed individual in Szczecin increase and the largest increase takes place when obtaining university education. As presented in table 10 – transition from vocational secondary education to general secondary education effects with increase of chances to be employed for as much as 83,2%. One has to remark that it is not obligatory to pass between consecutive groups (table 11). Therefore unemployed individual being until now in the group of people with general secondary education (group 3) can pass directly to group of university education (group 5) and his chances find a job are increased much more, for 98%.

Table 11. Relative chance at transition of unemployed to any group according the level of education

<i>Educ</i> (i,j)	1	2	3	4	5
1	1	-	-	-	-
2	32,50%	1	-	-	-
3	62,84%	22,90%	1	-	-
4	76,03%	32,85%	8,10%	1	-
5	222,42%	143,34%	97,99%	83,20%	1

Source: own study.

## Conclusions

In the paper results of studies concerning influence of level of education of unemployed for labour seeking time with regards to their gender and age has been presented. Due to specific of studied phenomenon, especially occurrence of censored data, methods of survival analysis were used for identification of determinants and for study of their influence to labour seeking time. The study comprised 5074 persons registered as unemployed in Local Labour Office in Szczecin and withdrawn in the 4<sup>th</sup> quarter of year 2006 due to various reasons. 42.3% of surveyed found job. In general it can be stated that persons better educated found job faster and gender as well as age had influence to this phenomenon. The largest chance to find a job had young people regardless from level of education. In the lowest level of education the employers preferred women. Together with growth of level of education men had more chances to find job and from the level of general secondary education these chances were bigger for men than for women.

**REFERENCES:**

1. Cox D. R., Oakes D., (1984), *Analysis of Survival Data*, Chapman and Hall, London.
2. Domański C., Pruska K., (2000), *Nieklasyczne metody statystyczne*, PWE, Warszawa.
3. Frątczak E., Gach-Ciepiela U., Babiker H., (2005), *Analiza historii zdarzeń. Elementy teorii, wybrane przykłady zastosowań*, SGH, Warszawa.
4. Frątczak E., Józwiak J., Paszek B., (1996), *Zastosowania analizy historii zdarzeń w demografii*, SGH, Warszawa.
5. Gehan, E. A., (1965a), A generalized Wilcoxon test for comparing arbitrary single-censored samples, *Biometrika*, no 52, pp. 203-223.
6. Gehan, E. A. (1965b): A generalized two-sample Wilcoxon test for double-censored data, *Biometrika*, no 52, pp. 650-653.
7. Hosmer D.W., Lemeshow S., (1999), *Applied Survival Analysis. Regression Modeling of Time to Event Data*, John Wiley & Sons, INC, New York.
8. Hozer J., ed.,(2002), *Badania statystyczne w ubezpieczeniach*, Wydawnictwo Naukowe Uniwersytetu Szczecińskiego, Szczecin.
9. Kleinbaum D. G., Klein M., (2005), *Survival Analysis, Second Edition*, Springer, New York.
10. Kopycińska D., ed., (2006), *Kapitał ludzki jako czynnik rozwoju społeczno-gospodarczego*, Katedra Mikroekonomii, Uniwersytet Szczeciński, Szczecin.
11. Lawless J. F., (1982), *Statistical Models and Methods for Lifetime Data*, John Wiley & Sons, New York.
12. Markowicz I., Stolorz B., (2006a), *Analysis of the Survival Function of Firms*, in: *Baltic Business Development. SME Management International Degree Programmes*. Faculty of Economics and Management, Szczecin University, Szczecin.
13. Markowicz I., Stolorz B., (2006b), *Wykorzystanie analizy historii zdarzeń do konstrukcji tablic żywotności firm*, *Wiadomości Statystyczne* no 4/2006, GUS, PTS, Warszawa.
14. Namboodiri K., Suchindran C. M., (1987), *Life Table Techniques and Their Applications*, Academic Press Inc., New York.
15. Stanisław A., ed., (2005), *Biostatystyka*, Wydawnictwo Uniwersytetu Jagiellońskiego, Kraków.