Wiadomości Statystyczne. The Polish Statistician, 2023, *68*(1), 23–38 DOI: 10.5604/01.3001.0016.2361
Statystyka w praktyce / Statistics in practice
Zgłoszony/submitted: 17.02.2022, poprawiony/revised: 5.12.2022, zaakceptowany/accepted: 17.12.2022

# Principal component analysis of older people registered as unemployed in public employment offices

Karolina Bolesta[a]

**Abstract.** The determinants of registering as an unemployed person in the public employment office may be of both a socio-demographic and legal character. Although every individual has their own motivation to register as unemployed, it is still possible to analyse the phenomenon on a group level. The purpose of this study is to show the similarities and differences of older people registering as unemployed and to identify the factors that were key to professional deactivation. The research is based on data from the Polish Central Analytical and Reporting System concerning 1,276,555 people born in the years 1940–1965, who were at least once registered as unemployed in a public employment office. The study uses principal component analysis (PCA) to identify the factors which influence to the largest extent the decision to deactivate professionally. The Kaiser-Meyer-Olkin measure of sampling adequacy and Bartlett's test of sphericity proved the feasibility of the PCA. The number of principal components was determined on the basis of Kaiser's criterion. The varimax factor rotation was applied to simplify the relation between the variables and to enhance the interpretation of the obtained results. The analysis included five groups: pensioners, disability pensioners, people who reached retirement age, people who received pre-retirement benefits and the total population. For each group three to four components were identified which combined different variables. Education and occupation in the last place of work formed the only common component for the five groups which influences the most critical decisions in the labour market. This component demonstrates the level of competence and may determine the moment of professional deactivation. The research shows that economic mechanisms are more important than legal conditions in all the analysed groups.

**Keywords:** pensioners, disability pensioners, retirement age, pre-retirement benefit, unemployed, professional deactivation, principal component analysis, PCA, Kaiser-Meyer-Olkin measure, Bartlett's test of sphericity

**JEL:** J26

[a] Szkoła Główna Handlowa w Warszawie, Kolegium Analiz Ekonomicznych, Polska / SGH Warsaw School of Economics, Collegium of Economic Analysis, Poland. ORCID: https://orcid.org/0000-0002-9668-5782. E-mail: kboles@sgh.waw.pl.

# Analiza głównych składowych populacji osób starszych zarejestrowanych w powiatowych urzędach pracy jako bezrobotni

**Streszczenie.** Czynniki decydujące o zarejestrowaniu się w urzędzie pracy jako osoba bezrobotna mogą mieć charakter społeczno-demograficzny lub prawny. Podjęcie takiej decyzji jest kwestią indywidualną, niemniej jednak można przeanalizować to zjawisko na poziomie grupy. Celem badania omawianego w artykule jest wykazanie podobieństw i różnic osób rejestrujących się jako bezrobotni oraz identyfikacja czynników, które mają największe znaczenie dla dezaktywizacji zawodowej. W badaniu wykorzystano dane z Centralnego Systemu Analityczno-Raportowego dotyczące 1 276 555 osób urodzonych pomiędzy 1940 r. a 1965 r., które co najmniej raz były zarejestrowane w powiatowym urzędzie pracy jako bezrobotni. W celu identyfikacji komponentów najsilniej wpływających na decyzję o dezaktywizacji zawodowej przeprowadzono analizę głównych składowych (ang. *principal component analysis* – PCA). Wyniki pomiaru adekwatności Kaisera-Meyera-Olkina oraz testu sferyczności Bartletta potwierdziły słuszność zastosowania tej analizy. Na podstawie kryterium Kaisera określono liczbę głównych składowych. Przeprowadzono rotację czynników varimax, aby uprościć relację między zmiennymi i lepiej zinterpretować uzyskane wyniki. Analiza dotyczyła pięciu grup: emerytów, rencistów, osób, które osiągnęły wiek emerytalny, osób, które pobierały świadczenie przedemerytalne, oraz całej populacji. Dla każdego zbioru danych zidentyfikowano od trzech do czterech składowych łączących różne zmienne. We wszystkich grupach stwierdzono jeden wspólny komponent – łączący wykształcenie i zawód wykonywany w ostatnim miejscu pracy – który wpływa na podejmowanie kluczowych decyzji na rynku pracy. Obrazuje on kompetencje pracowników i może determinować moment dezaktywizacji zawodowej. Wyniki badania wskazują na większe znaczenie mechanizmu ekonomicznego niż uwarunkowań prawnych we wszystkich analizowanych grupach.

**Słowa kluczowe:** emeryci, renciści, wiek emerytalny, świadczenie przedemerytalne, bezrobotni, dezaktywizacja zawodowa, analiza głównych składowych, PCA, współczynnik Kaisera-Meyera-Olkina, test sferyczności Bartletta

## 1. Introduction

The decision to deactivate professionally depends on several factors determined by demographic characteristics and by one's professional background. While every individual may have their own motivation to deactivate professionally, it is crucial to check what exactly contributes to this decision on a group level. It may be a personal decision or the result of a company's actions (Armstrong-Sassen, 2008). This topic is becoming increasingly important, especially in Poland, as the population of older people is growing. Although the absolute value of the global population is rising, the study should also take into consideration the structure of a given population. In 1950, the Polish post-working population constituted 7.1% of the entire population, while in 2011 it was 17.5% and is predicted to continue to increase, especially within a longer period (Żołędowski, 2012). By 2050, Poland is likely to have one of the

oldest societies in Europe (Przybyłka, 2017). Age is one of the most important factors taken into account when labour-related decisions are made at the national level, as they have a significant impact on the overall state of the economy (Wyszkowska, 2020). The working age population shapes the national economy to the largest extent. The concept of the silver economy, on the other hand, assumes that all economic activities are to be designed to consider the fact that a large portion of the population consists of older people. Although the topic of older people is commonly studied in the sphere of economics, it tends to take the form of theoretical considerations of the determinants of labour-related decisions, while practical analyses based on recent data are crucial to taking relevant action. This article is an empirical attempt to bridge this research gap. The most important research question addressed in this paper is how different the groups of the population are and whether the same determinants lead these particular groups to professional deactivation.

There are several laws in Poland which significantly impact deactivation in the labour market. The Act of 20th April 2004 on the promotion of employment and on labour market institutions (Pol. Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy)[1] defines the responsibilities of the state in the sphere of employment protection, professional activation and mitigating the effects of unemployment. The other important law relates to old-age and disability pensions from the Social Insurance Fund.[2] This law indicates the conditions which have to be fulfilled to be eligible for pensions and disability pensions. The labour market and the pension system strongly interact with each other, as the former can be treated as the next stage for the latter. The decisions concerning the labour market have an ultimate influence on the lives of pensioners. Extending the period of professional activity improves the state of the labour market. It also determines how one's income is allocated throughout their life (Góra & Rutecka, 2013). The disability pension system underwent changes in 1999 as a result of its deteriorating financial state (Góra, 2014). When the reform was introduced, the population was divided into 3 groups: those born before 1st January 1949, between 1st January 1949 and 31st December 1968 and after 1st January 1969 (Błoch, 2018). The year of birth determined the system that an individual belonged to and the extent of the modifications that could be introduced to it. In addition, the amount of the pre-retirement benefit depended on the amount of the received unemployment benefit; factors which were taken into account included one's professional status and the

[1] See: https://www.ilo.org/dyn/natlex/natlex4.detail?p_isn=68870&p_lang=en.

[2] Act of 17th December 1998 on old-age and disability pensions from the Social Insurance Fund (Pol. Ustawa z dnia 17 grudnia 1998 r. o emeryturach i rentach z Funduszu Ubezpieczeń Społecznych). See: https://www.ilo.org/dyn/natlex/natlex4.detail?p_lang=en&p_isn=71898.

time for which the unemployment benefit was received. The passive labour market policies, having a protective effect, play a major role in this area (Cichowicz et al., 2022).

Even though each group was subject to a different set of legal regulations, it has not been checked how socio-demographic factors influence the choice to deactivate professionally. The purpose of this study is to show the similarities and differences of older people registering as unemployed and to identify the factors that were key to professional deactivation. These issues were studied on the basis of the principal component analysis (PCA). The identified components allowed important criteria to be distinguished which determined the decision to deactivate professionally. The population was divided into five groups described below, out of which four were analysed to help understand which factors proved most influential for each group.

## 2. Research method

### 2.1. Data

The research included all people born between 1940 and 1965 in Poland, who were at least once registered as unemployed in a public employment office.[3] The period selected for analysis resulted from data availability in the Polish Central Analytical and Reporting System (CeSAR). In December 2021, the research group consisted of 1,276,555 participants, out of which four groups were analysed: 136,642 were pensioners, 478,253 were disability pensioners, 149,827 were people who reached retirement age and 511,833 were people who received pre-retirement benefits. The fifth group was formed by people who did not meet the criteria to be classified in any of the four groups above. In Poland, reaching retirement age does not oblige an individual to deactivate professionally; however, if the person postpones retiring, then they cannot be in the official registry (Act of 17th December 1998 on old-age and disability pensions from the Social Insurance Fund, Act of 20th April 2004 on promotion of employment and on labour market institutions).

The studied four groups obtained different kinds of benefits having met the specific for their group legal requirements. The following part of the paper focuses on how certain socio-demographic characteristics observed among those groups determined their choice to deactivate professionally. The article presents an analysis showing in what way the groups differ from each other. The less they differ in terms

[3] In Poland, a person of a working age (18–59 years – women, 18–64 years – men; the upper age limit is determined by the official retirement age) can be registered as unemployed in the public employment office.

of socio-economic characteristics, the less crucial the legal conditions of professional deactivation become and the more significant the background economic mechanism is.

The study is based on individual administrative data obtained from the CeSAR. The level of detail expressed by the data is gmina, i.e. the Polish territorial unit equivalent to a commune. The data relate to the labour market and are managed by the Ministry of Labour and Social Policy. After structuring the data, the sociodemographic characteristics of registered people were analysed. Although there are several dozen variables for each of the groups, only nine were selected for the PCA. The set of variables is identical for each group. The following points present the variables used in the PCA with the abbreviation of each one (later used in the article) given in the parentheses:

- employment office (*empl. office*) – a unique ID of each poviat (i.e. the Polish territorial unit equivalent to a county) labour office, indicating the location of the office in Poland. The highest number of registered unemployed is reported in Śląskie (16.5%), Mazowieckie (10.1%) and Dolnośląskie (8.2%) Voivodships;
- sex (*sex*) – there are 47.5% females and 52.5% males in the whole dataset:
  - pensioners: 47.2% women and 52.8% men,
  - disability pensioners: 43.6% women and 56.4% men,
  - people who reached retirement age: 65.1% women and 34.9% men (in 2022, the retirement age is 60 for women and 65 for men),
  - people who obtained pre-retirement benefits: 58.1% women and 41.9% men;
- level of education (*education*) – there are several levels (date of last registration in public employment offices: December 2021):
  - 0 – no education (14.2%),
  - 1 – primary (30.1%),
  - 2 – vocational (25.8%),
  - 3 – secondary vocational (17.7%),
  - 4 – secondary general (5.7%),
  - 5 – post-secondary (2.7%),
  - 6 – tertiary (3.8%);
- last occupation group (*last occupation*) – groups of jobs classified according to the required skills on the basis of the International Standard Classification of Occupations (ISCO) in force in a given year (date of last registration in public employment offices: December 2021):
  - 1 – representatives of public authorities, senior officials and managers (2.5%),
  - 2 – specialists (4.0%),
  - 3 – technicians and other associate professionals (8.4%),
  - 4 – office workers (9.4%),
  - 5 – service workers and vendors (22.5%),

  - 6 – farmers, gardeners, foresters and fishermen (1.3%),
  - 7 – industrial workers and craftsmen (21.2%),
  - 8 – plant and machine operators and assemblers (10.1%),
  - 9 – workers doing simple jobs (20.6%);
- major occupation group (*major occupation*) – the structure is the same as above, it indicates the most frequent occupation taken up by a person:
  - 1 – representatives of public authorities, senior officials and managers (1.9%),
  - 2 – specialists (4.0%),
  - 3 – technicians and other associate professionals (8.1%),
  - 4 – office workers (10.0%),
  - 5 – service workers and vendors (23.1%),
  - 6 – farmers, gardeners, foresters and fishermen (1.3%),
  - 7 – industrial workers and craftsmen (26.1%),
  - 8 – plant and machine operators and assemblers (9.2%),
  - 9 – workers doing simple jobs (16.3%);
- years of professional experience (*experience*) – total tenure of the person at the moment of their last registration in the employment office. The average tenure was 26.1 years;
- pre-retirement benefit (*pre-ret. benefit*) – it has two values: 0 if someone was not granted a pre-retirement benefit (97.2%) and 1 if someone obtained a pre-retirement benefit at least once (2.8%);
- year of birth (*year of birth*) – the selected range of data includes people born between 1940 and 1965. The most numerous groups are formed by individuals born in 1952 (8.2%), 1954 (8.1%) and 1951 (8.0%);
- period counted (*period counted*) – period when a person did not work, included in the period of employment, which results from a specific legal provision; it is important for determining the scope of employee rights. The periods included in this article are:
  - ZATR – period of employment (0.3%),
  - ALMP – active labour market policies (3.4%),
  - ZAS – different kinds of benefits obtained from the social assistance office (4.6%),
  - INO – all different periods that can be counted to the period of employment (91.7%).

The data were analysed using a software package for computing statistics and performing visualisation.

### 2.2. Principal component equations

PCA was performed to increase the interpretability of the data. It is an unsupervised multivariate statistical method (Gnanadesikan, 1977). It was recognised as one of the most important methods applied when the analysis of big datasets is concerned (Mishra et al., 2017). It can be used in different contexts, from education (Wojnar, 2020) and sustainable development (Drastichová & Filzmoser, 2019) to chemometrics (Bro & Smilde, 2014). It is, however, mainly used in the social sciences, for instance for determining GDP growth (Górecki & Łaźniewska, 2013) and analysing financial markets (Majewska, 2016). The main purpose of this method is to reduce data complexity by determining important variables which are transformed into interpretable factors (Cattell, 1973). The method was introduced in 1901 by Pearson, who focused mainly on regression (Pearson, 1901). The current concept of the method was independently developed by Hotelling (1933) as the idea of linear combinations. The analysis involves the examination of the structure of the internal dependencies in multivariate observations with the purpose of determining any possible relationships. The variables from an existing database are grouped into new orthonormal, representative factors. In other words, the said method transforms the group of variables into new interpretable factors called components which are not correlated with one another. Through this transformation the data have fewer dimensions expressed by a minimum number of components. However, it retains a similar set of information to the original database. It allows the simplification of the descriptive analysis to major components (Sanchez, 2006).

PCA identifies the components as linear combinations of observable variables (Bro & Smilde, 2014):

$$\begin{aligned} P_1 &= a_{11}X_1 + a_{12}X_2 + \cdots + a_{1j}X_j, \\ P_2 &= a_{21}X_1 + a_{22}X_2 + \cdots + a_{2j}X_j, \\ &\cdots \\ P_i &= a_{i1}X_1 + a_{i2}X_2 + \cdots + a_{ij}X_j, \end{aligned} \tag{1}$$

where:

$P_k$ – principal components, where $k = 1, 2, 3, \ldots, i$,

$a_{kj}$ – loadings, where $k = 1, 2, 3, \ldots, i$, $j = 1, 2, 3, \ldots, i$,

$X_j$ – variables, where $j = 1, 2, 3, \ldots, i$.

The factors were rotated by means of the varimax method, whose main purpose is to receive a minimum number of factors with a maximum set of information (Thompson, 2004). The variables are categorical, thus, the subsystems can be solved for any combination of the variable values.

### 2.3. The Kaiser-Meyer-Olkin measure and Bartlett's test of sphericity

The verification of the proper use of the PCA can be based on two measurements: the Kaiser-Meyer-Olkin (KMO) measure and Bartlett's test of sphericity (Bramer, 2013). The KMO measure (being in fact an index, not a typical statistical test) provides information on sampling adequacy. The range of KMO is from 0 to 1. The closer to 1, the higher probability that the PCA is used properly, and exceeding 0.5 justifies the application of the PCA. While it is possible to calculate KMO for every single variable, the aggregated version of the indicator is sufficient. The formula for KMO is presented below (Tabachnick & Fidell, 2013):

$$KMO_j = \frac{\sum_{i \neq j} R_{ij}^2}{\sum_{i \neq j} R_{ij}^2 + \sum_{i \neq j} U_{ij}^2}, \tag{2}$$

where:
$R_{ij}^2$ – correlation matrix,
$U_{ij}^2$ – partial covariance matrix.

The other measure, i.e. Bartlett's test of sphericity verifies the null hypothesis in terms of whether the variables are not correlated (orthogonal). On the other hand, the alternative hypothesis checks whether the variables lack orthogonality. The significant value of less than 0.05 suggests that the PCA could be a good fit for the dataset. The null hypothesis claims that the variables in the original dataset are not correlated. The formula for Bartlett's test of sphericity is provided below (Guttman, 1954):

$$\chi^2 = -\left(n - 1 - \frac{2p + 5}{6}\right) \times \ln|R|, \tag{3}$$

where:
$p$ – number of variables,
$n$ – sample size,
$R$ – correlation matrix.

## 3. Discussion on the results

In order to check the feasibility of performing PCA, the KMO measure and Bartlett's test of sphericity were used. The results are presented in Table 1.

**Table 1.** The value of KMO and Bartlett's test of sphericity, which checks the feasibility of the PCA

| Specification | Pensioners | Disability pensioners | People who reached retirement age | People who had pre-retirement benefit | Total population |
|---|---|---|---|---|---|
| KMO | 0.741 | 0.670 | 0.643 | 0.675 | 0.630 |
| Test: $\chi^2$ | 139,000 | 2,390,000 | 1,570,000 | 3,690,000 | 620,000 |
| Df | 36 | 36 | 36 | 36 | 36 |
| $p$-value | 0.000 | 0.000 | 0.000 | 0.000 | 0.000 |

Source: author's calculations based on the CeSAR database and STATA software.

The KMO measure for each of the four groups and total population is greater than 0.5. As regards the Bartlett test of sphericity, the significance for all groups is 0.000. Both of these measurements confirm the feasibility of the PCA.

**Table 2.** Eigenvalues for the components for all the analysed groups as the result of the PCA

| Variable | Pensioners | | | | Disability pensioners | | | |
|---|---|---|---|---|---|---|---|---|
| | eigen-value | diff. | pro-portion | cumm | eigen-value | diff. | pro-portion | cumm |
| *empl. office* | **2.286** | 0.887 | 0.254 | 0.254 | **1.839** | 0.462 | 0.204 | 0.204 |
| *sex* | **1.398** | 0.287 | 0.155 | 0.409 | **1.377** | 0.334 | 0.153 | 0.357 |
| *education* | **1.111** | 0.093 | 0.123 | 0.533 | **1.043** | 0.047 | 0.116 | 0.473 |
| *last occupation* | **1.018** | 0.050 | 0.113 | 0.646 | 0.996 | 0.024 | 0.111 | 0.584 |
| *major occupation* | 0.968 | 0.073 | 0.108 | 0.753 | 0.972 | 0.100 | 0.108 | 0.692 |
| *experience* | 0.895 | 0.323 | 0.099 | 0.853 | 0.872 | 0.029 | 0.097 | 0.789 |
| *pre-ret. benefit* | 0.572 | 0.048 | 0.064 | 0.917 | 0.843 | 0.184 | 0.093 | 0.883 |
| *year of birth* | 0.524 | 0.297 | 0.058 | 0.974 | 0.659 | 0.261 | 0.073 | 0.956 |
| *period counted* | 0.228 | . | 0.025 | 1.000 | 0.398 | . | 0.044 | 1.000 |

(cont.)

| Variable | People who reached retirement age | | | | People who had pre-retirement benefit | | | |
|---|---|---|---|---|---|---|---|---|
| | eigen-value | diff. | pro-portion | cumm | eigen-value | diff. | pro-portion | cumm |
| *empl. office* | **1.972** | 0.414 | 0.219 | 0.219 | **2.130** | 0.607 | 0.237 | 0.237 |
| *sex* | **1.558** | 0.476 | 0.173 | 0.392 | **1.524** | 0.355 | 0.169 | 0.406 |
| *education* | **1.081** | 0.068 | 0.120 | 0.512 | **1.169** | 0.157 | 0.130 | 0.536 |
| *last occupation* | **1.013** | 0.039 | 0.113 | 0.625 | **1.012** | 0.075 | 0.112 | 0.648 |
| *major occupation* | 0.974 | 0.034 | 0.108 | 0.733 | 0.937 | 0.165 | 0.104 | 0.752 |
| *experience* | 0.939 | 0.238 | 0.104 | 0.837 | 0.772 | 0.127 | 0.086 | 0.838 |
| *pre-ret. benefit* | 0.701 | 0.304 | 0.078 | 0.915 | 0.645 | 0.186 | 0.072 | 0.910 |
| *year of birth* | 0.397 | 0.032 | 0.044 | 0.959 | 0.458 | 0.104 | 0.051 | 0.961 |
| *period counted* | 0.365 | . | 0.041 | 1.000 | 0.354 | . | 0.039 | 1.000 |

Note. Eigenvalue – variance of the factor; diff. – the difference between the previous and the current factor; proportion – the share in explaining the factor; cumm – the sum of the eigenvalues. The eigenvalues higher than 1 have been marked in bold.

Source: author's calculations based on the CeSAR database.

**Table 3.** Component matrix for the four analysed groups that shows which variables belong to which components

| Variable | Pensioners | | | | Disability pensioners | | |
|---|---|---|---|---|---|---|---|
| | comp1 | comp2 | comp3 | comp4 | comp1 | comp2 | comp3 |
| *empl. office* | 0.004 | –0.044 | –0.041 | **–0.791** | 0.012 | 0.030 | **0.442** |
| *sex* | –0.153 | **0.650** | 0.180 | –0.081 | –0.253 | **0.431** | 0.025 |
| *education* | **–0.517** | –0.075 | –0.001 | 0.047 | **–0.444** | –0.022 | –0.055 |
| *last occupation* | **0.595** | 0.036 | 0.125 | 0.014 | **0.579** | 0.167 | 0.010 |
| *major occupation* | **0.584** | 0.007 | 0.105 | 0.024 | **0.624** | 0.142 | –0.015 |
| *experience* | –0.045 | **–0.705** | 0.128 | 0.075 | 0.044 | **–0.647** | 0.136 |
| *pre-ret. benefit* | 0.027 | 0.217 | –0.302 | **0.521** | 0.014 | 0.046 | **0.631** |
| *year of birth* | –0.082 | –0.090 | **0.639** | 0.285 | –0.096 | **0.585** | 0.010 |
| *period counted* | –0.065 | 0.136 | **0.651** | –0.087 | 0.055 | –0.047 | **–0.620** |

(cont.)

| Variable | People who reached retirement age | | | | People who had pre-retirement benefits | | | |
|---|---|---|---|---|---|---|---|---|
| | comp1 | comp2 | comp3 | comp4 | comp1 | comp2 | comp3 | comp4 |
| *empl. office* | 0.013 | 0.069 | –0.232 | **0.716** | –0.152 | –0.103 | **–0.580** | 0.182 |
| *sex* | 0.305 | **0.634** | 0.004 | –0.037 | **0.472** | 0.256 | –0.363 | 0.060 |
| *education* | **0.455** | –0.142 | 0.219 | –0.083 | **0.422** | –0.100 | 0.396 | –0.080 |
| *last occupation* | **–0.506** | 0.263 | 0.204 | 0.008 | –0.277 | **0.606** | 0.108 | –0.072 |
| *major occupation* | **–0.557** | 0.254 | 0.178 | 0.003 | –0.337 | **0.584** | 0.095 | –0.031 |
| *experience* | 0.122 | –0.340 | **0.564** | 0.102 | –0.326 | –0.314 | **0.415** | 0.121 |
| *pre-ret. benefit* | 0.040 | 0.047 | 0.458 | **0.617** | 0.173 | 0.133 | **0.418** | 0.204 |
| *year of birth* | 0.336 | **0.562** | 0.138 | –0.023 | **0.499** | 0.293 | 0.084 | 0.049 |
| *period counted* | –0.054 | 0.074 | **0.528** | –0.294 | –0.018 | 0.057 | 0.032 | **0.945** |

Note. Comp – component. The values in bold indicate which variables belong to a component.
Source: author's calculations based on the CeSAR database.

Table 2 presents the eigenvalues and the variance for each component. The eigenvalue corresponds with the correlation matrix. The number of eigenvalues is the same as the number of components. The critical part of the PCA is the selection of the proper number of components. The literature presents many methods of achieving the above, but the Kaiser criterion tends to be used most often. It involves considering only those variables whose initial eigenvalues are greater than one, which guarantees that they provide more information than the initial variables. The variation of more than one initial variable is thus explained (Kaiser, 1960), as the component can be selected only when it explains the same amount of or more information compared to the initial variable.

The sum of the variance of all the initial variables is equal to the sum of the variables of the principal components. Since the first few components contain most of the information, the number of components can be reduced. The first principal component is most significant, as it provides the maximum possible amount of information. The second principal component is the second most significant, as it

contains the maximum left information from the data, and so on. Every new component creates a new analysis axis.

The main purpose of Table 2 is to determine the number of components. The eigenvalue is greater than one for four components, except for the disability pensioners group. As a result, there are four components for pensioners, three components for disability pensioners, four components for people who reached retirement age and four components for people with pre-retirement benefits.

Table 3 was designed to determine which initial variables belong to every single component, which was achieved by assigning the highest variable loading to each component.

For pensioners, the first component is associated with gaining knowledge. There are three variables which constitute this component: education, last occupation group with the same sign and major occupation group with opposite sign. This component can be understood as an individual's 'professional path'. The second component encompasses sex and year of professional experience with the opposite direction and so may be described as 'the sex in the labour market'. The third component includes year of birth and period counted with not only the same direction, but also the same value. It can be named as 'decisions on the labour market based on age'. The fourth component includes employment office and pre-retirement benefits, hence it can be referred to as 'inclination to receive pre-retirement benefits based on location'. The first three components depict the situation before professional deactivation and the fourth one is mainly connected with pre-retirement. It means that socio-demographic variables are important for pensioners when deciding to cease working.

When it comes to disability pensioners, the first component includes the same variables with the same direction as for pensioners, but with slightly different values. This 'professional path' depends mainly on the last occupation group and major occupation group, while the level of education has the lowest impact. The second component consists of mainly experience and year of birth, but also sex, thus, it can be labelled as the 'trajectory in the labour market from the perspective of the sex'. The third component includes the poviat labour office information about pre-retirement benefits, but the period counted variable belongs with the opposite direction. The name for it can be 'direction in the labour market'. This group has the lowest number of components, which may incline that among all the analysed groups, this one is the most homogenous.

The structures of variables for people who reached pre-retirement age is similar to those of pensioners, as are the values of the analysed variables. The first component includes the same variables (level of education, last occupation group, major occupation group), but the direction of them is exactly opposite. Therefore, the

following variables are of similar values: education, last occupation, major occupation. The second component includes year of birth and sex. It can be perceived as 'importance of age and sex'. The third component includes experience and period counted with the same value and direction, thus it can be called the 'trajectory in the labour market'. The last component focuses on the employment office and pre-retirement benefits, so it can be described as 'location and pre-retirement benefit'.

The structure of the components for people who received pre-retirement benefits is slightly different. The first component involves the following variables: sex, education and age with the same value and direction, thus it can be understood as the 'relation between age and gender in the context of the labour market'. The second component includes two variables solely describing occupation: last occupation group and major occupation group, so it can be referred to as the 'impact of the standard and previous job'. This implies a strong connection between the current job and the previous one. In this group, major occupation is most stable among all the studied groups. The third component is based on data on the poviat labour office with one direction, while experience and pre-retirement benefits on the opposite. It may be called 'location and benefits in the labour market'. The fourth component includes only one variable, i.e. period counted, therefore the name of the component is the same. Its value is the highest for all the components for all the groups (0.945).

Table 4 and Table 5 present information relating to the total population. Table 4 is sorted by decreasing eigenvalues and Table 5 shows how components are displayed for all the variables.

**Table 4.** Eigenvalues for the components relating to the total population

| Variable | Eigenvalue | Diff. | Proportion | Cumm |
|---|---|---|---|---|
| *empl. office* | **1.701** | 0.266 | 0.189 | 0.189 |
| *sex* | **1.435** | 0.357 | 0.159 | 0.349 |
| *education* | **1.078** | 0.048 | 0.120 | 0.468 |
| *last occupation* | **1.030** | 0.093 | 0.115 | 0.583 |
| *major occupation* | 0.937 | 0.097 | 0.104 | 0.687 |
| *experience* | 0.840 | 0.040 | 0.093 | 0.780 |
| *pre-ret. benefit* | 0.801 | 0.098 | 0.089 | 0.869 |
| *year of birth* | 0.702 | 0.228 | 0.078 | 0.947 |
| *period counted* | 0.475 | . | 0.053 | 1.000 |

Note. As in Table 2.

Source: author's calculations based on the CeSAR database and STATA software.

**Table 5.** Component matrix for the total population

| Variable | Comp1 | Comp2 | Comp3 | Comp4 |
|---|---|---|---|---|
| *empl. office* | –0.139 | –0.190 | 0.418 | **0.674** |
| *sex* | 0.387 | 0.244 | 0.085 | **0.498** |
| *education* | **0.500** | 0.051 | –0.056 | –0.350 |
| *last occupation* | –0.345 | **0.524** | 0.099 | –0.113 |
| *major occupation* | –0.471 | **0.475** | 0.119 | –0.072 |
| *experience* | –0.126 | **–0.463** | 0.313 | –0.255 |
| *pre-ret. benefit* | **0.431** | 0.393 | 0.145 | 0.025 |
| *year of birth* | 0.190 | 0.176 | **0.405** | –0.119 |
| *period counted* | –0.060 | 0.043 | **–0.713** | 0.280 |

Note. As in Table 3.

Source: author's calculations based on the CeSAR database and STATA software.

The first component includes information on education and eligibility for pre-retirement benefits, so it can be called the 'correlation of the level of education and the professional path in the labour market'. It shows information both from the beginning and the end of the labour market trajectory. The second component includes data on professional activity, namely the last occupation and major occupation group in one direction and experience in the opposite direction. It can be labelled as the 'theoretical and practical background'. It encompasses all the information gathered throughout an individual's lifecycle. The third component includes two variables in the same direction: year of birth and period counted. It can be called 'how age matters in the period counted'. The fourth component involves the employment office number and sex variables, so it can be referred to as 'sex--based differentiation in poviat labour offices'.

In all of the groups, occupation and education are interconnected. The way in which professional deactivation occurred proved not so important. Although the analysed data relates to people who in most cases finished their formal education 20 and more years ago, this education remains important even in the last stage of an individual's professional life. The most volatile variable was period counted – it was connected with different variables in each of the subgroups. In every subgroup not only the content of the components was different, but also the values of the variables. It shows the heterogeneity among the subgroups, which results in different policies being applied to the particular groups.

## 4. Conclusions

The analysed dataset related to older people born between 1940 and 1965, registered at least once in a public employment office. The purpose of the research was met as the most important factors behind professional deactivation were identified. The

PCA was conducted to extract important information from the original variables showing maximum variability. The correct application of the method was verified and confirmed by the KMO measure and Bartlett's test of sphericity. The results of this study allowed determining which variables were most considered when those registered as unemployed in public employment offices in Poland faced the decision to deactivate professionally. The study included five groups: pensioners, disability pensioners, people who reached retirement age, those who received pre-retirement benefits, and the total population. The analysis enabled focusing only on those variables which proved relevant in the decision-making process concerning each specific group of people, which policymakers are responsible for. The economics mechanism tends to be more important than the legal conditions.

The identification of the factors which determine the decision on professional deactivation allows policymakers to focus on those specific characteristics which matter the most within each group. Policymakers can therefore create more insightful programmes and opportunities, suited to the needs of the specific groups, thus contributing to a higher quality of life. The wide scope of historical data enables making predictions as to which programmes are likely to be needed in the future. The created components tend to provide two types of information: demographic characteristics and geographical traits. Most of the components remain different for all of the groups in terms of their structure and number. Even if some components include the same variables, the direction and especially the value of the variable differs. The only component that is common for all of the analysed groups is that relating to the education and occupation of a person. These variables are clearly important when planning to deactivate professionally. This information should be prioritised when working on programmes aiming to benefit older people.

The limitations of this study largely result from the scope of the dataset. Had there been more variables available, the content of the components would have provided more information on the population structure. The determinants of professional deactivation depend not only on the components indicated in the article; nevertheless, all the available ones were analysed. The other factors which are relevant include health or the family situation of an individual. The further direction for this study can be the measurement of the effectiveness of programmes focusing on the needs of older people. It is necessary to consider the regulations in effect during the period relating to the data from the analysed database. The conclusions of the analysis may be helpful in improving the regulations concerning older people that would apply during their professional deactivation.

## Acknowledgements

This article was prepared within the *Registered unemployment as a non-traditional route to non-participation of older workers. Recurrent event longitudinal data analysis* project (UMO-2018/30/E/HS4/00335), financed by the National Science Centre, Poland.

## References

Armstrong-Sassen, M. (2008). Organisational Practices and the Post-retirement Employment Experience of Older Workers. *Human Resource Management Journal*, *18*(1), 36–53. https://doi.org/10.1111/j.1748-8583.2007.00057.x.

Błoch, W. (2018). Niespójność systemu emerytalnego i systemu rentowego w Polsce. *Gospodarka w Praktyce i Teorii*, *53*(4), 27–39. https://doi.org/10.18778/1429-3730.53.02.

Bramer, M. (2013). *Principles of Data Mining* (2nd edition). Springer. https://doi.org/10.1007/978-1-4471-4884-5.

Bro, R., & Smilde, A. K. (2014). Principal component analysis. *Analytical Methods*, *6*(9), 2812–2831. https://doi.org/10.1039/c3ay41907j.

Cattell, R. B. (1973). *Factor analysis. An Introduction and Manual for the Psychologist and Social Scientist*. Greenwood Press.

Cichowicz, E., Flaszyńska, E., Gałecka-Burdziak, E., & Bolesta, K. (2022). Zasiłki i świadczenia przedemerytalne – przedwczesna dezaktywizacja zawodowa osób nie w późnym wieku?. *Praca i Zabezpieczenie Społeczne*, (7), 14–23. https://doi.org/10.33226/0032-6186.2022.7.4.

Drastichová, M., & Filzmoser, P. (2019). Ocena zrównoważonego rozwoju za pomocą analizy skupień i analizy głównych składników. *Problemy Ekorozwoju. Problems of Sustainable Development*, *14*(2), 7–24.

Gnanadesikan, R. (1977). *Methods for Statistical Data Analysis of Multivariate Observations* (2nd edition). John Wiley & Sons. https://doi.org/10.1002/9781118032671.

Góra, M. (2014). Redesigning pension systems. *IZA World of Labor*, (51), 1–10. https://doi.org/10.15185/izawol.51.

Góra, M., & Rutecka, J. (2013). System emerytalny. In M. Kiełkowska (Eds.), *Rynek pracy wobec zmian demograficznych* (pp. 162–179). Instytut Obywatelski.

Górecki, T., & Łaźniewska, E. (2013). Funkcjonalna analiza składowych głównych PKB. *Wiadomości Statystyczne*, *58*(4), 23–34. https://ws.stat.gov.pl/Article/2013/4/023-034.

Guttman, L. (1954). Some necessary conditions for common-factor analysis. *Psychometrika*, *19*(2), 149–161. https://doi.org/10.1007/BF02289162.

Hotelling, H. (1933). Analysis of a complex of statistical variables into principal components. *Journal of Educational Psychology*, *24*(6), 417–441. https://doi.org/10.1037/h0071325.

Kaiser, F. H. (1960). The Application of Electronic Computers to Factor Analysis. *Educational and Psychological Measurement*, *20*(1), 141–151. https://doi.org/10.1177/001316446002000116.

Majewska, E. (2016). Zastosowanie metody głównych składowych do analizy integracji rynków finansowych. *Finanse, Rynki Finansowe, Ubezpieczenia*, (4 cz. 2), 227–236. https://doi.org/10.18276/frfu.2016.4.82/2-18.

Mishra, S. P., Sarkar, U., Taraphder, S., Datta, S., Swain, D. P., Saikhom, R., Panda, S., & Laishram, M. (2017). Multivariate Statistical Data Analysis – Principal Component Analysis (PCA). *International Journal of Livestock Research*, *7*(5), 60–78.

Pearson, K. (1901). On Lines and Planes of Closest Fit to System of Points in Space. *Philosophical Magazine*, *2*(11), 559–572. https://doi.org/10.1080/14786440109462720.

Przybyłka, A. (2017). Starzenie się ludności w Polsce jako wyzwanie dla systemu ochrony zdrowia. *Studia Ekonomiczne. Zeszyty Naukowe Uniwersytetu Ekonomicznego w Katowicach*, (309), 179–190. https://www.ue.katowice.pl/fileadmin/user_upload/wydawnictwo/SE_Artyku%C5%82y_291_320/SE_309/16.pdf.

Sanchez, J. (2006). Handbook of Univariate and Multivariate Data Analysis and Interpretation with SPSS. *Journal of Statistical Software, Book Reviews*, *16*(4), 1–3. https://doi.org/10.18637/jss.v016.b04.

Tabachnick, B. G., & Fidell, L. S. (2013). *Using multivariate statistics* (6th edition). Pearson.

Thompson, B. (2004). *Exploratory and Confirmatory Factor Analysis. Understanding Concepts and Application*. American Psychological Association. https://doi.org/10.1037/10694-000.

Ustawa z dnia 17 grudnia 1998 r. o emeryturach i rentach z Funduszu Ubezpieczeń Społecznych (Dz.U. 1998 nr 162, poz. 1118 ze zm.).

Ustawa z dnia 20 kwietnia 2004 r. o promocji zatrudnienia i instytucjach rynku pracy (Dz.U. 2004 nr 99, poz. 1001 ze zm.).

Wojnar, J. (2020). Zróżnicowanie wykorzystania technologii informacyjno-komunikacyjnych w krajach Unii Europejskiej. *Wiadomości Statystyczne. The Polish Statistician*, *65*(8), 39–56. https://doi.org/10.5604/01.3001.0014.3526.

Wyszkowska, D. (2020). Senior economy in Poland. *Wiadomości Statystyczne. The Polish Statistician, 65*(5), 9–26. https://doi.org/10.5604/01.3001.0014.1330.

Żołędowski, C. (2012). Starzenie się ludności – Polska na tle Unii Europejskiej. *Problemy Polityki Społecznej*, (17), 29–43. http://www.problemypolitykispolecznej.pl/Starzenie-sie-ludnosci-n-Polska-na-tle-Unii-Europejskiej,123771,0,1.html.