European Research Studies Journal
Volume XXIV, Issue 2, 2021

pp. 714-747

The Standard of Living of Inhabitants and the Scientific and Technological Potential in Selected European Union Regions

Submitted 25/03/21, 1st revision 18/04/21, 2nd revision 09/05/21, accepted 20/05/21

Mariusz Malinowski[1]

Abstract:

Purpose: *This paper aims to determine the relationship between the standard of living of the inhabitants of selected regions of the European Union and the scientific and technological potential.*
Design/Methodology/Approach: *The study covered 60 regions – NUTS 2 – in 13 European Union Member States. The data concerned the year 2018. Due to the multidimensional character of the analyzed categories, a canonical analysis was used as a generalization of multiple linear regression into two sets of variables. In order to evaluate the statistical significance of the analyzed canonical variables, the Wilks' lambda significance test was conducted. As part of the canonical analysis, canonical correlations, total redundancy, and extracted variances were calculated.*
Findings: *Based on results of the classical correlation analysis, it can be concluded that there is a positive, high, and statistically significant correlation dependence (Spearman's rank correlation coefficient was nearly 0.66) between the standard of living of inhabitants from selected EU regions and S&T potential (measured by synthetic measures constructed based on the TOPSIS method). Five statistically significant canonical variables (components) were identified in the canonical analysis. The value of the largest and most statistically significant canonical correlation was over 0.97. For the last (fifth) statistically significant canonical variable, this value was over 0.72.*
Practical Implications: *The results of the conducted research (among others ranking of countries according to the standard of living of the inhabitants) may be indirectly used by the central and local authorities responsible for local and regional development (including undertaking pro-social and pro-innovation activities) in the context of the choice of the direction for the socio-economic restructuring of particular countries and local government units.*
Originality/Value: *A rarely used multivariate in socio-economic research, canonical analysis is a valuable tool for assessing the relationships between two compiled, multi-faceted categories.*

Keywords: *Standard of living, technological potential, multidimensional analysis, canonical analysis, linear ordering, classification of objects.*

JEL code: *C38, I30.*

Paper type: *Study research.*

[1]*Department Economics, Faculty of Economics, University of Life Sciences, Poznań, Poland, mariusz.malinowski@up.poznan.pl;*

1. Introduction

The issues concerning the standard of living are very diverse and require the use of a variety of both quantitative and qualitative methods. The standard of living is difficult to measure directly, and one can only try to describe it with several partial indicators and then, based on them, try to assess the studied phenomenon using synthetic measures. Research on the degree to which needs are met has been and is being carried out by researchers worldwide. Such research concerns the inhabitants of particular countries, particular social groups (e.g., farmers, residents of rural areas), or families with different income levels.

Studies carried out for many years show that life satisfaction in industrialized countries has not increased for 40 years despite a significant increase in income. What does it take to be happy, then? The challenges of modern living require an individual to be active and creative in both thinking and doing. More and more often, it is emphasized that the basic requirements of the modern world are, quality of work, competence, creativity, adaptability, love for socializing and working in teams, transfer of skills, independence, and the ability to cope with unpredictable conditions (Midor and Wieczorek, 2016). The standard of living of the inhabitants (or the welfare in the broadest sense of the term) is nowadays determined by, among others, safety, interpersonal relations, and durability of ties with people, intellectual and emotional preparation, and the way of fulfilling social roles, etc. Modern technologies play a specific role in this context. At the beginning of the third decade of the 21st century, modern technologies (especially ICT) have become a permanent feature of the human environment in many countries (especially the wealthier ones).

Despite numerous analyses, it seems that both the quantification of the standard of living and identification of factors contributing to the increase in the level of the analysed phenomenon remains a problem that has not been fully resolved. This applies both to the selection of diagnostic variables used for the formation of aggregate measures, methods for measuring this phenomenon, and methods for identifying relationships between the standard of living and determinants.

This article aims to estimate the relationship between the standard of living of the inhabitants of selected EU countries and the level of S&T potential. One of the advanced methods of multidimensional statistical analysis – canonical analysis – was used for this purpose. The intention of the author of this article was not to seek ideal definitions of "standard of living," "technology," or "scientific and technological potential" (S&T potential). Therefore, attempts have been made to avoid entering considerations regarding their information capacity (which is often the case primarily in analyses concerning living standard, welfare, quality of life, and living conditions).

However, the focus was not on quantifying these phenomena based on publicly available indicators and investigating the relationship between these multi-faceted categories. The classic TOPSIS method was applied to quantify the analysed phenomena, while Ward's method and – applied less frequently in practice – the

FANNY method was used to group EU regions. The selection of partial variables was made based on substantive, statistical (appropriate differentiation and degree of correlation), and formal criteria (primarily completeness and data availability for the studied objects). The analysis was conducted based on EUROSTAT data and concerned mainly the year 2018. In this study, the research objects were regions in selected EU countries that acceded to the Community in 2004 (Cyprus, the Czech Republic, Estonia, Lithuania, Latvia, Malta, Poland, Slovakia, Slovenia, Hungary) and later Bulgaria and Romania in 2007, Croatia in 2013). The establishment of such collectivity primarily resulted from the timeliness and comparability of available statistics. For some countries (Malta and Latvia), due to the frequent lack of regionally aggregated data, country-wide data were accepted (but only when there was one NUTS unit per country).

2. Living Standard and Modern Technologies

To a large extent, the concept of living standard, due to its multifacetedness, does not have any single, universally accepted definition in the literature. The semantic capacity of this term is vast and varied, which is associated with the fact that it is subject to analysis in many sciences (e.g., sociology, philosophy, economics, physiology, psychology). On the one hand, there is a broad interdisciplinary research perspective, while on the other hand, there is a problem of operationalizing this research category.

The issue of "living standard" is primarily based on the need theory. A *need* can be defined as a perceived state of absence of something by an individual, while social needs are needs whose fulfillment requires the existence and action of various social institutions for the intended purposes. A characteristic feature of needs is their variability over time, which is less related to basic needs (e.g., food, shelter) and more to the higher (luxurious) ones (Słaby, 2007).

In the scientific literature, individual authors in their research have distinguished different groups of human needs, among which can be distinguished those applied, among others, by:

- Max-Neef'a (1991): 1. Subsistence; 2. Protection; 3. Affection, 4. Understanding; 5. Participation; 6. Idleness; 7. Creation; 8. Identity; 9. Freedom.
- Central Statistical Office (2004): 1. Income; 2. Household expenses; 3. Food consumption; 4. Housing conditions; 5. Equipping households with durable goods; 6. Health and social care; 7. Education; 8. Culture and recreation.
- Ding, Jiang, and Riloff (2018): 1. Physiological Needs; 2. Physical Health and Safety Needs; 3. Leisure and Aesthetic Needs; 4. Social, Self-Worth, and Self-Esteem Needs; 5. Finances, Possessions, and Job Needs; 6. Cognition and Education Needs; 7. Freedom of Movement and Accessibility Needs.

The above presentation of the classification of groups of needs is not a closed list, but to some extent, it shows the evolution of taking into account the extent of satisfaction of needs in empirical research, also in the context of the research category, which is the living standard of inhabitants. A relatively broad definition of the "standard of living" is proposed by J. Berbeka. According to her, the standard of living in the condition and availability of goods and services and the conditions by which an individual (municipal community) can satisfy their material and spiritual needs and the scope of their use (Berbeka, 2006). Hansen and Grubb (2002) define the standard of living as utility or happiness derived from consumption. In this sense, consumption is broadly understood as any good, activity, or obtained condition that people can acquire/gain. In terms of a global non-profit organization - One Global Economy - using modern technologies to combine underpaid community around the world with important information and online resources, the standard of living is determined primarily by three categories (Mourad *et al.*, 2014):

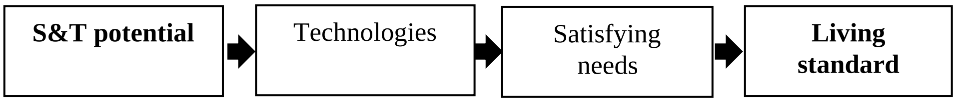
- income (changes in annual income, savings, employment and career, entrepreneurship)
- education (completion of secondary school, admission to university)
- health (availability of health care system, disease management programs, preventive medicine; health (availability of health care system, disease management programs, preventive medicine (i.a. prenatal care, sanitation services, vaccinations).

It seems that today, in the context of stimulating the standard of living, the importance of S&T potential and technologies will increase in individual countries (regions). It is not without reason that technology, alongside science and innovation, is considered by OECD documents to be one of the most critical factors for productivity growth and a driver for long-term economic growth and widely perceived prosperity (OECD, 2015).

The term "technology" means processing materials into various products, the science of means of natural resources processing (Latusek, 2004). By contrast, "technological potential" in the economic literature is most often considered in the context of business competitiveness and is usually treated as the set of technologies available to individual economic operators. In this sense, it consists of codified knowledge (projects, formulas, sketches, production instructions), the knowledge possessed by individual persons and teams working in the company (some of which are contained in the company's procedures and organization), as well as machines and equipment (Wiśniewska, 2012).

However, technological potential means more than a set of technologies currently available to economic operators. The technological potential is based on knowledge (i.a., research personnel and the level and quality of research). It is difficult to disagree with Z. Chyba, who believes that the technological potential includes, among others, the effectiveness of R&D activity, creativity, and entrepreneurship of employees, their critical abilities and competencies, willingness to learn, etc., (Chyba, 2014). Technological potential can therefore be understood as all possibilities and abilities to create and then use technology. Scientific achievements and scientists thus create the foundation of technological potential.

Figure 1. *Scientific and technological (S&T) potential and the standard of living of inhabitants*



Source*: Author's own study.*

Table 1. *Selected areas of the impact of technology on human activity*

Need/service	Example
Safety areas	
Control (monitoring) systems	Aircraft, car, train, air traffic, highway traffic and road traffic control systems, as well as industrial monitoring system, production management systems, forest and tunnel monitoring
E-government	
Development of direct democracy; Reduction of administrative costs; Electronic recording of all types of official data; Computerisation of land and mortgage registers; Possibility of vehicle registration	Online voting; e-offices at all levels
Education	
Need for access to information resources and knowledge	Possibility to browse folders, upload files, exchange information
Popularisation of knowledge	Social knowledge base
Tele-education; training	Teaching regardless of place and time
Need for development courses that help with re-training or acquiring new competences	Improvement of qualifications, competencies
Learning	
Modelling capabilities for processes requiring huge processing power	Weather forecasting including e.g. tornadoes
Biology and computer biocybernetics; population biology	Access to biological databases
Universalisation of the means and content of information transfer – universality	Information is provided instantly in several languages
Work	
Need for professional activity (e-work); activation of disabled people and women	The possibility to work remotely, at different times, on a part-time basis
Medicine	
Needs for efficient, quick diagnosis of the patient without the need to keep them at the medical facility; fast and non-invasive telediagnostics; online registration and reservation; e-prescribing; prevention	Access to high-class specialist physicians via Internet; remote ECG; electronic prescribing of medicines and electronic transmission of prescriptions; health and care websites

Source: *Author's own study based on Świeboda and Sienkiewicz (2010).*

The standard of living is defined by many variables that represent different groups of human needs, including those that can be met to a significant extent with the support of modern technologies (primarily CIT), such as (Świeboda and Sienkiewicz, 2010):

- social needs – satisfied by specific institutions (e-health, e-safety, e-education, e-policy);
- economic needs – satisfied by economic activity (e-services, e-work);
- information needs – satisfied by access to up-to-date information and knowledge as well as by the development of social communication networks;
- safety needs – satisfied by the development of ICT security systems, data protection systems, monitoring, etc.

Technological change not only contributes to the creation of new methods of treatment and the development of advanced medicines, or the creation of new working and learning conditions but also causes a lifestyle change. Modern technologies make it easier to do shopping (including the flow of current shopping trends across continents), move or complete many formalities without leaving home.

These technologies (CIT) make it possible to visit museums, rooms of art, and digital libraries, shape civic attitudes and behaviour in society, and develop one's interests and skills practically at any place and time. New technologies can also be a factor in facilitating effective professional and social advancement for socially excluded groups of people. The influence of new technologies, particularly CIT (which can be treated as the flow of knowledge-based economies), covers virtually all spheres of human functioning. However, it seems that the actual availability of new technologies in human life is determined primarily by three aspects: (a) physical availability of technological equipment (the percentage of inhabitants using the technology can be taken as an indicator of availability of new technologies); (b) costs of using technologies (equipment and infrastructure); (c) skills in using and usefulness in daily life.

Technological progress, especially the evolution of CIT, undoubtedly contributes to improving working conditions and increasing the efficiency of companies from many industries, and machines are increasingly replacing people. It should be borne in mind that modern technologies contribute to the emergence of a kind of pathology associated with the performance of professional duties, which is linked to the dependence on modern technological solutions (e.g., smartphones, smartwatches, the Internet) as well as the appearance of civilization diseases associated with working in front of a monitor (e.g., obesity, eye, and spinal diseases).

In the 21st century, modern technologies have become a factor enhancing social inequality (also in the European Union). They even contribute to the formation of social groups that cannot (e.g., due to lack of financial resources or adequate knowledge) keep up with the progress of changes. Consequently, this can have an impact on the exclusion of many people from social and professional life and thus lower the standard of living.

Frey and Osborne note that over 700 professions will cease to exist in the next decade or two. The most endangered professions include telemarketers, insurers, cargo and freight agents, librarians. According to their estimations, as much as 47% of total employment in the United States is seriously threatened by increasing automation (Frey and Osborne, 2013). Undoubtedly, the progressing changes will have a significant relationship to the development of artificial intelligence. The possibilities of artificial intelligence are increasingly mentioned not only in the context of analyzing large amounts of data, improving and automating business and production processes but also in the context of using it to create texts on one's own (which in consequence will contribute to replacing, among others, journalists, and translators).

The artificial intelligence and Heliograph tool used by the Washington Post should be mentioned here. Heliograph is equipped with AI algorithms that allow it to create notes based on data provided. In the first year, the tool created approximately 850 articles and won the "Post" award for "Excellence in the use of bots" for its work on election reports in 2016 (Strømmen-Bakhtiar, 2020).

From the perspective of the considerations conducted, it is worth mentioning the quantification of the multidimensional phenomenon: the standard of living and the measures used for this purpose. In the literature (Słaby, 2007; Grzega 2015), there is no single, universally accepted measure (indicator) nor set of measures (indicators) to measure living standards (Table 2). The difference between the terms "measure" and "indicator" concerns their role in the financial information system.

According to Słaby (2007), the measure is a concrete, empirical numerical evaluation, whereas the indicator is the evaluation applied to interpret changes in (e.g., social) conditions. A different approach to defining these terms is presented by Borys, who believes that the measure determines the size, quality, value of an object or phenomenon. On the other hand, the indicator is understood as a number expressing the level of a given phenomenon, presented in relative or absolute form. In economic theory, these terms are used interchangeably (Borys, 1999).

Table 2. *Standard-of-living measures.*

Traditional measures	Alternative measures
I. of a subjective nature, which are opinions of individuals about the extent to which their needs are met; II. of an objective nature (expressed in value or in natural units), including: - partial measures used for single, narrow feature of the living standard or group of needs - synthetic measures used for a wider group of needs (e.g., HDI – Human Development Index).	Most frequently, these are measures based on non-value categories. Among these types of measures, one can distinguish biological and anthropometric ones. They are based on physical characteristics of human body (e.g., height, weight and body mass).

In this group, global measures (e.g., average life expectancy index) and composite ones (e.g., synthetic standard of living index calculated using the Geneva method).	

Source: *Author's own study based on Słaby (2007), Grzega (2015).*

In analysing complex spatial structures, linear ordering methods based on synthetic measures provide a high cognitive value. The essence of these methods is to order the objects from the best to the worst using an adequately constructed taxonomic meter. Such an approach allows for a synthetic look at the level of complexity of a phenomenon (it allows for examining the similarities and structural differences of objects characterized by similar or different levels of complexity) (Korol and Szczuciński, 2009). A synthetic measure (indicator) of a function of variable values of a set of characteristics. Each ordering corresponds to some ordering function, e.g., some linear combination of variables or a regression function between some mainly selected and ordering variables (Balicki, 2009).

When analysing both the standard of living of the inhabitants and the scientific and technological potential in the EU regions, it is necessary to compare many research facilities described using a large set of variables. It is not easy to express these phenomena using a single measurable one. This fact results in the use of multidimensional statistical analysis methods based on synthetic taxonomic measures, which replace the description of objects using several variables with a description using a single aggregate figure to quantify the lives of inhabitants and technological potential, as well as to study the relationships between these phenomena.

According to Piasny (1993), synthetic measures, rather than partial indicators, are a more appropriate measure of the inhabitants' standard of living. However, some limitations on synthetic measures should be kept in mind (Karpińska-Mizielska and Smuga, 1999) a subjective selection of diagnostic variables used in constructing a synthetic measure and a subjective selection of weights for individual variables in the aggregation formula.

From the point of view of the specificity of the multidimensional phenomenon, i.e., the standard of living, it should also be noted that synthetic measures do not consider the qualitative and immeasurable aspects related to the satisfaction of human needs.

In recent years, quantifying the standard of living has been discussed in many publications. The linear ordering or classification of objects according to the inhabitants' standard of living was carried out by researchers at various levels of aggregation – local, regional, and international. It is worth mentioning at this point the extensive taxonomic analysis of the standard of living of the inhabitants of European Union Member States carried out by Zelias *et al.* (2004), which, among other things, identified groups of countries with a similar level of the analysed phenomenon, or used the correspondence analysis as a tool of comparative analysis.

Taxonomic methods to measure the standard of living were used, among others, by Warzecha (2009) to compare the standard of living in Poland and EU Member States based on the Hellwig and Ward method, by Majka (2015) to classify Polish provinces by Janusz (2014) for the analysis of spatial differentiation of the standard of living at the level of districts in Warmińsko-Mazurskie Province (based on the Hellwig method), and also by Liang, Changdi, and Liming (2017), who used the TOPSIS method for linear ordering of the main cities of Guizhou Province in terms of the standard of living of its inhabitants, or by Malinowski and Smoluk-Sikorska (2020) to quantify the standard of living of the inhabitants of Polish districts based on the TOPSIS method.

3. Materials and Methods

In this analysis, the selection of diagnostic variables used to construct synthetic measures (both the standard of living of inhabitants and the scientific and technological potential) and canonical analysis was based on substantive, formal, and statistical criteria. During the first stage of the study, diagnostic variables relevant in the context of the phenomenon under consideration were selected. During the second stage, based on statistical criteria, the set of primary variables was reduced.

As suggested by Zelias (2000), in terms of substantive and formal criteria, the choice of partial variables should be made, considering issues such as universality - variables should have universally recognised weight and importance:

- measurability - the variables must be directly or indirectly measurable and expressed in absolute or relative terms,
- availability of figures - obtaining full numerical information on each variable included in the study,
- data quality - it is necessary to check whether the collected data are not riddled with a significant number of accidental errors (e.g. (recording errors), and whether they are sufficiently accurate,
- cost-efficiency - the cost of collecting data should be considered,
- interpretability - the variables should have a clear interpretation,
- the way the variables (stimulant, destimulant or nominant) interact.

As a result, based on substantive and formal premises, 26 potential diagnostic variables related to the standard of living were proposed (S14, S18, NT4-NT6 concerned 2017. For the NT2 and NT7 variables, the data were from 2016. The other variables used concerned 2018.), which were then divided according to substantive criteria into seven thematic groups (Słaby, 2007; Zeliaś (ed.) 2004):

- DEMOGRAPHY: S1 - Average life expectancy at birth; S2 - Natural increase.
- S3 - Population density; S4 - Infant mortality rate; S5 - Total fertility rate; S6 - Average age of mothers at birth.

- EDUCATION: S7 - Pupils enrolled in early childhood education (pre-school education); S8 - Youth not working and not receiving education (15-24 years); S9 - Students of lower-secondary schools; S10 - Ratio of higher education institution students to population ratio; S11 - Completed higher education in the 25-64 age group (The variables S10 and S11 are similar, but it is assumed that not every student completes his/her studies and receives a degree.).
- ECONOMY: S12 - Real growth rate of regional gross value added (GVA) at basic prices; S13 - Average employee wage; S14 - Regional gross domestic product (PPS per capita).
- LABOUR MARKET: S15 - Long-term unemployment rate (12 months and more); S16 - Labour force participation rates (persons aged 15-74); S17 - Unemployment rate.
- HEALTH: S18 - Beds available in hospitals (per 100,000 inhabitants); S19 - Dentists (per 100,000 inhabitants)[2]; S20 - Physicians (per 100,000 inhabitants).
- TOURISM: S21 - Nights spent at tourist accommodation establishments (per 100,000 inhabitants); S22 - Arrivals at tourist accommodation establishments (per 100,000 inhabitants); S23 - Net occupancy rate of beds places and bedrooms in hotels and similar accommodation; S24 - Number of establishments and bed places (per 100,000 inhabitants).
- TRANSPORT: S25 - Length of motorways per 100 km^2, S26 - Number of cars per 100,000 inhabitants.

In order to estimate the level of technological potential, the following set of diagnostic variables was used: NT1 - High-tech employment per 100,000 inhabitants; NT2 - Registered Community designs (RCD) per 100,000 inhabitants; NT3 - Human resources in science and technology (HRST) per 100,000 inhabitants; NT4 - Total R&D personnel per 100,000 inhabitants; NT5 - Researchers per 100,000 inhabitants; NT6 - Gross domestic expenditure on R&D (GERD) per 100,000 inhabitants; NT7 - Number of applications for a European Union trade mark (EUTM) per 100,000 inhabitants.

Efforts have been made to ensure that the included partial variables had an indicative character (e.g., given per 100,000 inhabitants) and not an absolute one in both analysed data sets. This approach has made it possible to avoid distortions because objects (EU regions) had specific characteristics (e.g., much larger geographical area than other objects). The sets of characteristics thus distinguished were subjected to a further selection process based on statistical premises. Namely, those that do not have a high discriminatory and adequate information potential have been eliminated from the sets of primary variables. Such a selection is sometimes treated as a particular case of variable weighting, as the eliminated variables are assigned a value of 0 and included in further analysis 1.

---

[2] *For the Pest and Budapest regions, due to the lack of regional data, a country-wide average is used for this variable.*

Multidimensional comparative analyses require that the individual partial variables, which will ultimately be included in the set of diagnostic variables, show adequate variability (in other words, the variable shows adequate discriminatory capacity) as a poorly differentiated variable brings little analytical value. For this reason, it was assumed that the set of primary variables would be reduced by those for which the absolute value of the classic coefficient of variation did not exceed an arbitrarily fixed, critical threshold value of 10% (such variables are quasi-stable, not carrying vital information about the phenomenon under consideration).

Apart from evaluating variability, an essential criterion for selecting partial variables is their degree of correlation (informational potential) with other variables. The informational potential is the more significant, the less correlated a variable is with other variables (in this sense, correlating means conveying the same information). To assess the informational value of the variables in question, the so-called inverse correlation matrix method was used. This method considers both direct and indirect links between the permitted diagnostic variables. The starting point is the construction of an asymmetric correlation matrix of potential diagnostic variables:

$$R = [r_{j,j'}],\ j,j' = 1,2,\ldots,m, \tag{1}$$

For each set of variables (and in the case of the standard of living for each of the seven thematic groups) an inverse matrix to the Pearson correlation matrix was calculated:

$$R^{-1} = \tilde{r}_{jj}, \qquad \text{for } \tilde{r}_{jj'} = \frac{(-1)^{j+j'}|R_{jj'}|}{|R|}, \tag{2}$$

where: $R_{jj'}$— the matrix reduced by deleting the j-th row and the j'-th column; $\lceil R \rceil$, $|R_{jj'}|$—determinants of the R and $R_{jj'}$ matrices respectively.

According to this method, from the set of primary variables, the variable for which the corresponding diagonal element of the inversed correlation matrix has the highest value, exceeding an arbitrarily fixed threshold value (often r*=10), should be eliminated. Variables that exceed a threshold value led to an ill-conditioned R matrix. The (already reduced) inverse correlation matrix is then re-determined, and it is checked whether the diagonal values do not exceed the established threshold value. The process is continued until no diagonal value exceeds the established threshold value (Młodak, 2006; Panek and Zwierzchowski, 2013).

According to the additivity requirement, the variables used in constructing synthetic measures may be presented in their primary or standardized form (according to the additivity requirement). The most common methods of data normalization include standardization, unitisation, and quotient transformation. For these analyses, to achieve comparability of the variables in question (the selected variables have different names and rows), a standardization process was carried out based on one of the most popular (Młodak, 2006) standardization formulae:

$$Vj = \frac{s_j}{\bar{x}_j} \cdot 100 \tag{3}$$

where: $\bar{x}_j$ — the arithmetic mean of the *j*-th variable; s_j — the standard deviation, j = 1, 2,…, *m*.

The purpose of standardisation is to obtain variables with a distribution with an average of 0 and a standard deviation of 1. As a result, the primary variables are deprived of natural units and the row of variables facilitates making comparisons.

To facilitate linear ordering of the selected regions of the European Union in terms of the standard of living of the inhabitants of the selected Member States and the scientific and technological potential, the classic TOPSIS method (*The Technique for Order of Preference by Similarity to Ideal Solution*) was used which is considered a model method of linear ordering. It is a somewhat modified version of the commonly used Hellwig development pattern method. In this method, the synthetic measure is constructed considering both the Euclidean distance from the pattern object and the anti-pattern object (in case of the aforementioned Hellwig development pattern method, only the said distance from the pattern object is considered). The following stages in the construction of a synthetic measure can be distinguished in this method (Hwang and Yoon, 1981):

1. Creation of a standardised decision matrix based on quotient transformation, for instance.

$$z_{ij} = \frac{x_{ij}}{\sqrt{\sum_{i=1}^{m} x_{ij}^2}} \text{ for } i = 1,2, \dots, m \text{ and } j = 1,2,\dots, n \tag{4}$$

where: x_{ij} - the observation of the *j*-th variable in the *i*-th object.

Of course, it is possible to use other formulae to standardise characteristics.

2. Where variable weighting is used, a weighting matrix must be constructed and then a weighted standard decision tree must be created:

$$v_{ij} = w_j \cdot z_{ij} \tag{5}$$

3. Based on a standardised decision matrix, the determination of the value vector for the pattern object (A^+) and the anti-pattern object (A^-) is made:

$$A^+ = (\max_i(v_{i1}), \max_i(v_{i2}), \dots, \max_i(v_{iN})), = (v_1^+, v_2^+, \dots, v_N^+) \tag{6}$$

$$A^- = (\min_i(v_{i1}), \min_i(v_{i2}), \dots, \min_i(v_{iN})), = (v_1^-, v_2^-, \dots, v_N^-) \tag{7}$$

4. Calculation for each object under analysis (in this case an EU region) of the Euclidean distance from the pattern object and the anti-pattern object:

$$s_i^+ = \sqrt{\sum_{j=1}^{N}\left(v_{ij} - v_j^+\right)^2};\ \ s_i^- = \sqrt{\sum_{j=1}^{N}(v_{ij} - v_j^-)^2}, i = 1,2,\dots,M, j = 1,2,\dots,N \quad (8)$$

5. Determination of the value of a synthetic variable that identifies the similarity of the considered objects to the "pattern" object according to an aggregation formula:

$$C_i = \frac{s_i^-}{s_i^+ - s_i^-} \text{ where } 0 \le C_i \le 1, i = 1,\dots,N. \quad (9)$$

The smaller the distance of a given object from the development pattern (and thus the more considerable distance from the development anti-pattern), the value of the synthetic measure is closer to 1.

To deepen the analyses, the regions were classified according to a similar standard of living of the inhabitants (and the level of scientific and technological potential) by two methods - Ward's method (Incremental Sum of Squares) and the FANNY method.

The Ward's method used in this analysis is quite widely described in the statistical literature (Młodak, 2006; Balicki 2009; Murtagh and Legendre, 2014). It is generally accepted that the effectiveness of detecting the actual data structure using Ward's method is much higher compared to other agglomerative methods. As simulations show, the Wards method is 40% better than the other methods (the second best was the farthest neighbor method) (Wysocki, 2010). However, this method tends to combine clusters with a small number of observations and generate clusters of similar size (Strahl, 2006; Stanisz, 2007; Młodak, 2006). Unlike many other clustering methods, Ward's method uses the variance analysis approach to estimate the distance between clusters. At each stage of merging, clusters of objects, out of all possible clusters of objects, are merged into one cluster, which creates a cluster of objects with the slightest variation due to the variables describing them. The measure of this variation is the error sum of squares (ESS) criterion, expressed by the formula:

$$ESS = \sum_{i=1}^{k}(x_i - \bar{x})^2 \quad (10)$$

where: x_i - the value of the variable being a segmentation criterion for the i-th object; k - the number of objects in the cluster.

In other words, according to the algorithm of this method, at each stage, an attempt is made to optimize the division obtained by combining the two elements by applying the criterion of minimal increase in the total intra-cluster sum of squares of deviations of all the variables for each object from their cluster averages. The clustering effects of Ward's method can be presented in the form of a linkage tree (the so-called dendrogram).

To perform a kind of corroborative analysis, the objects in question were grouped using the FANNY method, which is the so-called fuzzy classification of objects with a predetermined number of k clusters (in this case, the number of clusters is the same as in Ward's method). In this method, the object in question (in this case, the EU regions) is assigned a degree of affiliation (membership) of the object to clusters called the membership coefficient.

The algorithm of this method is based on the minimisation of the objective function, which can be presented as follows (cf. UNESCO, 2008):

$$F = \sum_{c=1}^{k} \frac{\sum_i \sum_j u_{ic}^2 u_{jc}^2 d_{ij}}{2\sum_j u_{jc}^2} \tag{11}$$

where: u_{ic} and u_{jc} are the membership coefficient (indicating the extent to which the *i*-th object belongs to class *c*) satisfying the following conditions:

$u_{ic} \geq 0\ i\ \sum_c u_{ic} = 1, dla\ i = 1,2,\dots,n, c = 1,2,\dots,k, k$ – the number of classes

The algorithm that minimises this objective function is an iterative algorithm and stops when the objective function does not decrease. The following two coefficients are calculated in this method to assess the quality of the classification:

- Dunn's partition coefficient illustrated by the following formula:

$$F_k = \sum_{i=1}^{n} \sum_{c=1}^{k} u_{ic}^2 / n \tag{12}$$

The smallest value of this coefficient can be $\frac{1}{k}$ for a completely fuzzy clustering – when all membership coefficients u_{ic} are the same and no object can be classified into one class. The highest value of F_k coefficient is 1 when all membership coefficients u_{ic} are 0 or 1. This is the case if the *i*-th object belongs completely to only one class *c* (u_{ic} = 1) and does not belong to the other classes at all, i.e., the membership coefficients for the other classes are 0.

- normalised Dunn's partition coefficient illustrated by the following formula:

$$F_k' = \frac{F_k - \left(\frac{1}{k}\right)}{1 - \left(\frac{1}{k}\right)} = \frac{kF_k - 1}{k - 1} \tag{13}$$

This is a standardized version of Dunn's partition coefficient, which always has values in the range [0,1].

In the next step, a correlation analysis was carried out to determine the strength and direction of the relationship between the taxonomic measures of the inhabitants' standard of living and technological potential. A correlation relationship is characterized by the fact that specific values of one variable (e.g., X) specific average values of the second variable are assigned (e.g., Y). A positive correlation occurs when an increase in the value of one variable corresponds to an increase in the average value of the other variable. A negative correlation occurs when an increase in the value of one variable is accompanied by a decrease in the average value of the other variable (Zeliaś, 2000). To level the influence of possible outliers on the results of the correlation analysis to some extent, the non-parametric Spearman's rank correlation coefficient was used:

$$r_s = 1 - \frac{6\sum_{i=1}^{n} d_i^2}{n^3 - n}, \quad (14)$$

where: d_i — the difference between the ranks of characteristic X and Y; n — the number of elements in the sample under consideration.

Spearman's rank correlation coefficient has values in the range [-1, 1]. The closer the r_s coefficient is to 1, the stronger is the analysed relationship. If the r_s coefficient = -1, then pairs of ranks are ranked in the opposite order. Therefore, its interpretation is similar to the Pearson correlation coefficient (the most commonly used measure of interdependence), but significant differences exist between these coefficients. The Pearson correlation coefficient is used to evaluate the linear relationship, while Spearman's rank correlation coefficient is used to evaluate the monotonous relationship - increasing or decreasing (not necessarily a linear one). If there is a linear relationship between the variables under examination, the value of both coefficients will be similar. On the other hand, if there is a curvilinear (but monotonous) relationship, the value of Spearman's rank correlation coefficient (the absolute value) will be greater than the value of the Pearson correlation coefficient (Szymczak, 2010).

Then, to present the relationship between the selected sets of variables (and not individual variables) relating to the standard of living of the inhabitants and the scientific and technological potential, a canonical analysis was used as one of the elements of multidimensional statistical analysis.

A canonical analysis is a more complex statistical inference procedure. It is a generalization of multiple linear regression (whereby the variability of one endogenous variable can be explained by the variability of a set of exogenous variables) into two sets of variables (endogenous and exogenous). The main idea of this method is to study the relationship between two sets of variables to analyse the relationship between the two new types of variables (the so-called canonical variables, also known as canonical components). These "new meta-variables" are the weighted totals of the first and second set, and the weights are selected so that the two weighted totals are as correlated as possible (the first type of variable is a linear function of the first set of variables, just as the second type of variable is a linear function of the second set).

In other words, the canonical variable is a secondary structure composed of primary characteristics; it is a set of primary variables correlated with each other and hierarchized according to their contributions to the new variable. It is influenced by an invisible factor hidden in the overtly existing primary variables (Ter Braak, 1990; Cavadias *et al*., 2001; Hardoon *et al.*, 2003; Piekut, 2008; Naylor *et al.*, 2010). When considering two linear combinations which can be treated as a vector of exogenous variables and (a vector of endogenous variables), the expression is maximized (Weenink, 2003, Hardoon *et al.*, 2003):

$$r_l = \frac{(w_x^T R_{xy} w_y)}{\sqrt{(w_x^T R_{xx} w_x w_y^T R_{yy} w_y)}}, \quad (16)$$

where: R_{xx} — the correlation matrix of exogenous variables, R_{yy} - the correlation matrix of endogenous variables, R_{xy} — the correlation matrix of both types of variables, w_x, w_y — weights for the first and second type of canonical variables, r_l — the canonical correlation coefficient.

This tool is not used very often in the context of standard of living or related issues. It would be prudent to mention here a study by Ebenezer (2012). He used canonical analysis to investigate to what extent selected poverty-related variables to correlate with literacy in Nigeria or a study by Grzeskowiak (2016) on the relationships between multidimensional sets of variables representing satisfaction from various aspects of life (including work and financial situation satisfaction) and other socio-economic indicators (e.g., percentage of people at risk of poverty or social exclusion) in Poland.

It can be presumed that the relatively rare use of this tool in economic analyses (in terms of the commonly used, e.g., classical correlation analysis or regression analysis) has at least two underlying causes. Firstly, the method itself is complicated (it requires knowledge of multiple regressions). Secondly, there are some difficulties in interpreting the results obtained (e.g., the multiplicity of indicators).

Given that the categories concerned are multi-faceted, using this multidimensional exploration technique to assess the relationship between them seems justified. Using multiple regression models for this purpose, for example, and analysing each endogenous variable separately could distort the analysis results. This is due to the loss of relevant information on the relationships in the set of endogenous variables. In turn, only the classical correlation analysis (e.g., Pearson's or Spearman's) between pairs of variables seems to be insufficient as it does not consider the links within the set of endogenous and exogenous variables.

One of the first steps, affecting the course of the whole canonical analysis, is to determine (by checking statistical significance) how many first pairs of canonical variables should be subjected to in-depth evaluation. The null hypothesis in the significance tests in the canonical correlation analysis is that there is no correlation between two sets of input variables. To verify the null hypothesis, a canonical correlation significance test, i.e., the Wilks' lambda test (Wilks' Λ), was applied. The verification of the significance of canonical component pairs was carried out based on test statistics for a set of s-p variable forms (Noble Jr., 2000; Panek and Zwierzchowski, 2013):

$$\Lambda_p = \prod_{l=p}^{s}(l - r_l^2) \tag{17}$$

where: s - the number of canonical components, r_l^2- the square of the canonical correlation coefficient of the *l*-th canonical variable.

The test statistic has a Wilks' lambda probability distribution of the number of degrees of freedom:

$$d_{f1} = m - p - s + 1 \text{ and } d_{f2} = n - m - p + s \tag{18}$$

In canonical analysis, one of the basic assumptions (often marginalised in socio-economic analyses) is that all partial variables included are characterised by normal distribution. Due to the difficulty of guaranteeing the normality of all analysed variables, the use of canonical correlation to analyse socio-economic phenomena is more justified for descriptive purposes than for statistical inference. Normality of the distribution of the considered variables was assessed based on the results of the Shapiro-Wilk test. To verify the H_0 *null hypothesis:* $F(x)=F_0(x)$, where $F_0(x)$ is the distribution function of normal distribution against the alternative hypothesis H_1: $F(x) \neq F_0(x)$, using the following statistics formula (Sobczyk, 2007):

$$W = \frac{[\sum_i a_i(n)(X_{n-i+1} - X_i)]^2}{\sum_{i=1}^{n}(X_j - \bar{X})^2}, \tag{19}$$

where: $a_i(n)$ – the fixed, tabulated value.

If the Shapiro-Wilk test results identified variables that did not meet the assumption of normal distribution, a Box-Cox transformation was used to approximate normal distribution. This transformation can be illustrated using the following formula (Box and Cox, 1964):

$$y_i^{(\lambda)} = \begin{cases} \frac{y_i^{\lambda} - 1}{\lambda} \; dla \; \lambda \neq 0, \\ \log y_i \; dla \; \lambda = 0\,. \end{cases} \tag{20}$$

where the choice of λ transformation parameter is made by the most reliable method.

As mentioned earlier in the canonical analysis, canonical weights are determined in such a way as to maximize the correlation between consecutive pairs of canonical variables. To facilitate the interpretation of canonical weights, it is recommended to use a standardized output data matrix (Panek and Zwierzchowski, 2013). For this reason, the output set of variables has undergone a standardization process (as already mentioned earlier).

As part of the canonical analysis, for each canonical variable, the values of the isolated variance have been calculated, determining what percentage of the variance of the input variables explain these canonical variables. It is determined by summing up the canonical squares of the factor loadings of the individual variables in the set for a given canonical component and then dividing it by the number of input variables, which the following formula can represent:

$$\overline{R^2_{u_l}} = \frac{1}{q}\sum_{j=1}^{q} c^2_{jl} \text{ or } \overline{R^2_{v_l}} = \frac{1}{m-q}\sum_{j=q+1}^{m} d^2_{jl}, l=1,2, \ldots, s, \quad (21)$$

where: q – the number of input variables; c_{jl} – is the canonical factor loadings of the *j-th* basic variable and the *l-th* canonical variable of the first type; d_{jl} – is the canonical factor loadings of the *j-th* basic variable and the *l-th* canonical variable of the second type. Then, by multiplying this mean by the square of canonical correlations, the redundancy index was obtained (for more details, see Thomson 1987). This index indicates how much average variance in one set is explained by a given canonical variable, for a given set of other variables. This index can be presented using the following formula:

$$R^2_{u_l,x^2} = \overline{R^2_{u_l}} \cdot \lambda_l \text{ or } R^2_{v_l,x^1} = \overline{R^2_{v_l}} \cdot \lambda_l, l=1,2, \ldots, s, \quad (22)$$

where: λ_l - the characteristic element of the square matrix of canonical correlation.

4. Results and Discussion

The multitude of variables (being carriers of various information) describing the analysed objects (EU regions) in multidimensional comparative analyses makes it necessary to choose the most important ones from the research point of view. Therefore, the construction of synthetic measures (Table 3) and canonical analysis was preceded by the reduction of the primary set of variables (created based on non-statistical (substantive and formal) criteria) through the evaluation of variability (discriminatory criterion) and the degree of correlation of individual variables (capacity criterion).

Consequently, considering the discriminatory criterion, three variables should be eliminated from the set of variables relating to the standard of living: S1 (coefficient of variation was 2.6843%); S6 (where the coefficient of variation had the value of 4.3070%), S16 (6.9547%). However, given the significant substantive value, especially of the S1 variable, all the variables in this set were further analysed. On the other hand, in the set of variables relating to technological potential, all variables were characterized by a coefficient of variation more significant than the adopted critical threshold of 10%. Therefore, also in this set, all variables were further analysed. However, following the evaluation of the information potential (based on the results obtained using the inversed correlation matrix method), the NT4 variable ((r*>10)) was eliminated from both sets under consideration.

The construction of synthetic measures requires the determination of the nature of individual variables - identification of the direction of influence on the analysed phenomena. Based on factual premises (or correlation analysis), it should be determined whether the selected variables are stimulants (desired high values from the point of view of the essence of the phenomenon under consideration), de stimulants (desired low values), or nominates (where the optimum value represents specific nominal values, and deviations from this value lead to more a negative assessment of the analysed phenomenon). In the set of variables relating to the standard of living, the destimulant set includes the following variables: S3 (Population density); S4 (Infant mortality rate); S6 (Average age of mothers at birth); S8 (Youth not working and not receiving education (15-24 years)); S15 (Long-term unemployment rate (12 months and more)) and S17 (Unemployment rate). The other variables are stimulants. This also applies to variables describing tourism. It was assumed that the higher the values of the variables considered, the greater the tourist attractiveness of the region and the more possibilities of spending leisure time. On the other hand, in the set of variables relating to technological potential, all variables were classified as stimulants.

Table 3. *Values of synthetic measures of the standard of living and scientific and technological potential.*

Region	I	II	Region	I	II
BG31:Severozapaden	0.3016	0.1903	PL43:Lubuskie	0.3621	0.2028
BG32:Severen tsentralen	0.3122	0.2100	PL51:Dolnoslaskie	0.3880	0.2040
BG33:Severoiztochen	0.3431	0.1943	PL52:Opolskie	0.3639	0.1918
BG34:Yugoiztochen	0.3734	0.1320	PL61:Kujawsko-Pom.	0.3735	0.1643
BG41:Yugozapaden	0.4192	0.2858	PL62:Warminsko-Mazur.	0.3360	0.1579
BG42:Yuzhen tsentralen	0.3699	0.1296	PL63:Pomorskie	0.4020	0.2263
CZ01:Praha	0.5554	0.5571	PL71:Lódzkie	0.3649	0.1582
CZ02:Strední Cechy	0.4379	0.3193	PL72:Swietokrzyskie	0.3480	0.2200
CZ03:Jihozápad	0.4291	0.2269	PL81:Lubelskie	0.3447	0.1522
CZ04:Severozápad	0.3803	0.1334	PL82:Podkarpackie	0.3529	0.1940
CZ05:Severovýchod	0.4225	0.2294	PL84:Podlaskie	0.3702	0.1830
CZ06:Jihovýchod	0.4317	0.3281	PL92:Mazowiecki reg.	0.3562	0.3230
CZ07:Strední Morava	0.4117	0.2491	RO11:Nord-Vest	0.4050	0.0594
CZ08:Moravskoslezsko	0.3887	0.2230	RO12:Centru	0.3583	0.0663
EE00:Eesti	0.4541	0.3307	RO21:Nord-Est	0.4411	0.3397
CY00:Kypros	0.4501	0.4418	RO22:Sud-Est	0.3215	0.0362
LV00:Latvija	0.4000	0.1438	RO31:Sud - Muntenia	0.3024	0.0452
LT01:Sostines regionas	0.5060	0.5017	RO32:Bucuresti - Ilfov	0.4917	0.2663
LT02:Vidurio ir vakaru Lietuvos regionas	0.4097	0.1626	RO41:Sud-Vest Oltenia	0.3254	0.0495
HR03:Jadranska Hrvatska	0.4974	0.1390	RO42:Vest	0.3720	0.1283
HR04:Kontinentalna Hrvatska	0.3489	0.1737	SI03:Vzhodna Slovenija	0.4044	0.2744
HU11:Budapest	0.4854	0.4438	SI04:Zahodna Slovenija	0.4760	0.4800
HU12:Pest	0.4077	0.2179	SK01:Bratislavský kraj	0.5155	0.5884
HU21:Közép-Dunántúl	0.3841	0.2167	SK02:Západné Slovensko	0.3527	0.1357
HU22:Nyugat-Dunántúl	0.3942	0.1918	SK03:Stredné Slovensko	0.3419	0.1473
HU23:Dél-Dunántúl	0.3453	0.1808	SK04:Východné Slovensko	0.3333	0.1356
HU31:Észak-Magyarország	0.3620	0.1818	**DIFFERENTIATION**		
HU32:Észak-Alföld	0.3550	0.1569	AA	0.3930	0.2278

HU33:Dél-Alföld	0.3674	0.1824	V_s [in %]	13.9494	57.2434
MT00:Malta	0.4384	0.6473	SD	0.0548	0.1304
PL21:Malopolskie	0.4156	0.3134	MED	0.3822	0.1929
PL22:Slaskie	0.3776	0.1619	Q1	0.3544	0.1510
PL41:Wielkopolskie	0.3989	0.2004	Q3	0.4200	0.2684
PL42:Zachodniopomorskie	0.3987	0.1435			

Note: *I – Synthetic measure of the standard of living, II - Synthetic measure of the scientific and technological potential, AA – arithmetic average,* V_s *– coefficient of variation, SD – standard deviation, MED – median, Q1 – first quartile, Q3 - third quartile*
Source: *Authors' own study.*

The calculations show that higher values for synthetic measures of the standard of living of inhabitants were characteristic of regions that included national capitals, where a significant part of the socio-economic potential of the whole country is accumulated (such as business environment institutions, cultural institutions).

Among ten regions with the highest value of the synthetic standard-of-living measure, such a situation occurred in 8 regions (Praha, Bratislavský Kraj, Sostines regions, București-Ilfov, Budapest, Zahodna Slovenija, Eesti, Kypros). The exceptions included Jadranska Hrvatska in Croatia (5th place given the value of the synthetic measure) and Nord-Est in Romania (10th place). In these regions, high values of variables relating, among others, to the population with higher education in the age group 25-64, average employee's remuneration, regional gross domestic product, labor force participation rate, number of dentists, and low values (which is essential for the considerations conducted) of the infant mortality rate and the rate relating to the number of young people, not in employment nor training (aged 15-24) were noted.

In turn, among the regions with the lowest values of the synthetic standard-of-living measure, the place among the ten lowest rankings was most frequently taken by regions from Bulgaria (3 times: Severozapaden, Severen centrale, Severoiztochen) and Romania (3 times: Sud-Muntenia, Sud-Est, Sud-Vest Oltenia). In addition, among 10 EU regions with the lowest score of the standard of living were two areas from Slovakia (Stredné Slovensko and Východné Slovensko) and Poland (Lubelskie and Warmińsko-Mazurskie Voivodeships). In these regions, relatively low values of partial variables relating, among others, to average life expectancy at birth, participants of pre-school education, average employee's salary, professional activity rate, number of nights spent in tourist accommodation facilities, the density of motorways were noted.

For the variables included, the synthetic measure of the standard of living of inhabitants was characterized by right-sided asymmetry, which means that values not exceeding the arithmetic mean prevailed (the classical skewness coefficient was 0.82). The classical variation coefficient was less than 14%, indicating a relatively weak differentiation of the analysed phenomenon. In the case of 75% of the analysed regions, the value of the synthetic measure did not exceed 0.42 (with a minimum value of 0.30 and a maximum value of 0.56).

The level of S&T potential in the analysed regions is much more diverse. The variation coefficient is over 57% for the variables included, and the standard deviation is over 0.13 (with an average value of approx. 0.23). As in the synthetic standard-of-living measure, the highest values of the synthetic measure of S&T potential have also been noted in the regions that include national capitals.

This was the case in 8 out of 10 regions with the highest values of the constructed measure (Malta, Bratislavský Kraj, Praha, Sostines regions, Zahodna Slovenija, Budapest, Kypros, Eesti). In the top ten, the Nord-Est region in Romania (8th place) and Jihovýchod in the Czech Republic (10th place) are the exceptions. These regions have high or extremely high values of variables relating mainly to human resources in science and technology (HRST), internal R&D expenditure, number of EU trademark applications. Among ten countries with the lowest scores in terms of S&T potential, six regions are from Romania (Sud-Est, Sud-Muntenia, Sud-Vest Oltenia, Nord-Vest, Center, Vest), two from Bulgaria (Yuzhen Tsentralen, Yugoiztochen) and one from the Czech Republic (Severozápad) and Slovakia (Východné Slovensko). The partial variables included in the synthetic measure formation had low or shallow values in those regions. This was particularly true for the variables describing the number of registered Community designs (RCDs), the total number of R&D personnel, and internal R&D expenditure.

The selected EU regions were classified using taxonomic similarity – Ward's method (with classical Euclidean distance) and the FANNY method to deepen the analyses. Separate clusters include regions with the standard of living and S&T potential, but the composition of a given group does not provide information on developing the analyzed phenomenon. Suppose Ward's or the FANNY method is used (in general, methods for non-linear ordering), it is impossible to establish a hierarchy of analysed multivariate objects. The result of grouping by these methods can be related to the results of linear ordering, but they may not be convergent ultimately.

A significant problem appearing in Ward's method (in general, in this type of grouping) is establishing a critical size of the distance at which dendrogram arms are cut off, and thus the clusters of the analysed objects are determined. The decision to set a threshold value is subjective. It is frequently determined top-down at level 4. For this reason, one of the supporting techniques was used to reduce subjectivity to some extent. Namely, to determine the critical value of the distance at which the dendrogram arms are cut off, the following formula was used (Panek and Zwierzchowski, 2013):

$$d_{i+1}^{*} > \bar{d} + ks_d$$

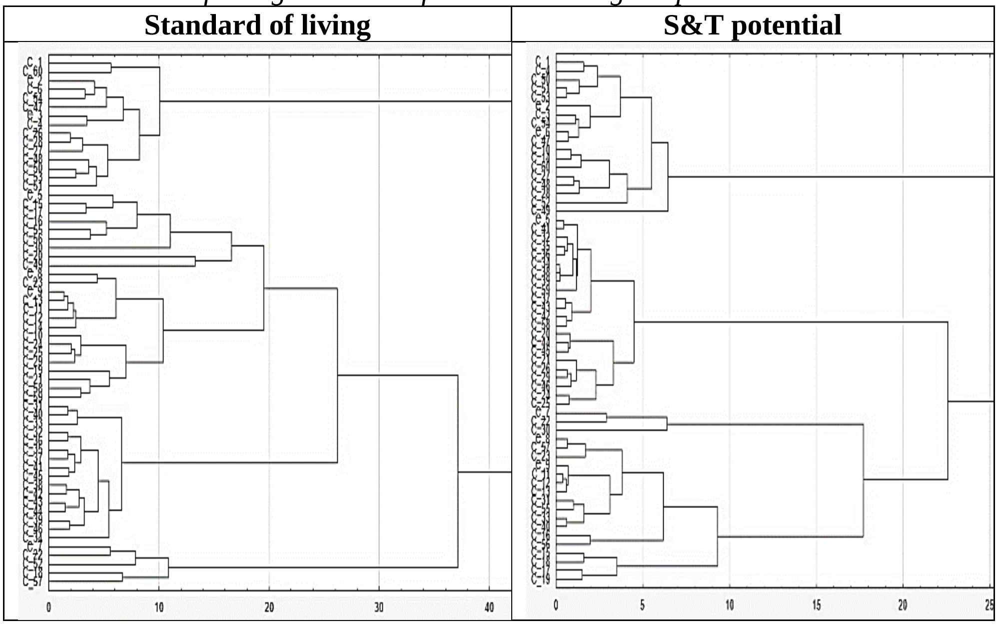
where: d_{i+1}^{*} - critical value of the distance corresponding to *i+1* of the branch length; $\bar{d}$, s_d - arithmetic mean and standard deviation of the tree branch length, k - parameter whose optimal value is within the range [2,5; 3,5]. Assumed $k = 3$.

The number of classes in the fuzzy classification can be determined in two ways, i.e., assuming the number of classes determined using the agglomeration classification or carrying out the classification with a different number of classes and then selecting the one for which the classification quality assessment index applied reaches an extreme level (Wysocki, 2010). The first approach has been applied in these analyses.

As a result, for the synthetic measure of the standard of living of inhabitants, the critical value of the distance at which the dendrogram arms were cut off was 26.62, while in S&T, potential – 18.50.

Based on the adopted criterion of determining the critical size of the distance at which dendrogram arms are cut off, three groups of regions should be created for the standard of living and two groups for the variables describing scientific and technological potential. To ensure comparability of the classification results, the same number of groups was adopted - 3.

Figure 2. *Dendrograms determined by Ward's method for sets of variables relating to the standard of living and scientific and technological potential*



Source: *Authors' own study.*

To facilitate interpretation, the results of the procedure for classifying a group of regions were numbered in descending order according to the arithmetic means of synthetic measures (obtained using the TOPSIS method) within a given cluster (Table 4). The use of different classification methods has resulted in heterogeneous grouping results. These incompatibilities may result, i.a. from a different way of calculating the distance between objects or the distance between clusters themselves.

Table 4. *Grouping results for selected EU regions by standard of living and S&T potential.*

	I	II	III
	Ward's method		
SL	CZ01:Praha; LT01:Sostines regionas; HU11:Budapest; RO32:Bucuresti – Ilfov; SK01:Bratislavský kraj	BG41:Yugozapaden; CZ02:Strední Cechy; CZ03:Jihozápad; CZ04:Severozápad; CZ05:Severovýchod; CZ06:Jihovýchod; CZ07:Strední Morava; CZ08:Moravskoslezsko; EE00:Eesti; CY00:Kypros; LV00:Latvija; LT02:Vidurio ir vakaru Lietuvos regionas; HR03:Jadranska Hrvatska; HR04:Kontinentalna Hrvatska; HU12:Pest; HU21:Közép-Dunántúl; HU22:Nyugat-Dunántúl;HU33:Dél-Alföld; MT00:Malta;PL21:Malopolskie; PL22:Slaskie; PL41:Wielkopolskie; PL42:Zachodniopomorskie; PL43:Lubuskie; PL51:Dolnoslaskie; PL52:Opolskie; PL61:Kujawsko-Pomorskie; PL62:Warminsko-Mazurskie; PL63:Pomorskie; PL71:Lódzkie; PL72:Swietokrzyskie; PL81:Lubelskie; PL82:Podkarpackie;PL84:Podlaskie; PL92:Mazowiecki regionalny;RO21:Nord-Est; SI03:Vzhodna Slovenija;SI04:Zahodna Slovenija;SK02:Západné Slovensko; SK03:Stredné Slovensko; SK04:Východné Slovensko	BG31:Severozapaden; BG32:Severen tsentralen; BG33:Severoiztochen; BG34:Yugoiztochen; BG42:Yuzhen tsentralen; HU23:Dél-Dunántúl; HU31:Észak-Magyarország; HU32:Észak-Alföld; RO11:Nord-Vest; RO12:Centru; RO22:Sud-Est; RO31:Sud – Muntenia; RO41:Sud-Vest Oltenia; RO42:Vest
TP	CZ01:Praha; CY00:Kypros; LT01:Sostines regionas HU11:Budapest; MT00:Malta; PL92:Mazowiecki regionalny; RO21:Nord-Est; SI04:Zahodna Slovenija; SK01:Bratislavský kraj	BG41:Yugozapaden; CZ02:Strední Cechy; CZ03:Jihozápad; CZ05:Severovýchod; CZ06:Jihovýchod; CZ07:Strední Morava CZ08:Moravskoslezsko; EE00:Eesti; HU21:Közép-Dunántúl; PL21:Malopolskie; RO32:Bucuresti – Ilfov; SI03:Vzhodna Slovenija	BG31:Severozapaden; BG32:Severen tsentralen; BG33:Severoiztochen; BG34:Yugoiztochen BG42:Yuzhen tsentralen; CZ04:Severozápad; LV00:Latvija; LT02:Vidurio ir vakaru Lietuvos regionas; HR03:Jadranska Hrvatska; HR04:Kontinentalna Hrvatska; HU12:Pest;

			HU22:Nyugat-Dunántúl; HU23:Dél-Dunántúl; HU31:Észak-Magyarország; HU32:Észak-Alföld HU33:Dél-Alföld; PL22:Slaskie; PL41:Wielkopolskie; PL42:Zachodniopomorskie; PL43:Lubuskie; PL51:Dolnoslaskie; PL52:Opolskie; PL61:Kujawsko-Pomorskie; PL62:Warminsko-Mazurskie; PL63:Pomorskie; PL71:Lódzkie; PL72:Swietokrzyskie; PL81:Lubelskie; PL82:Podkarpackie; PL84:Podlaskie; RO11:Nord-Vest; RO12:Centru; RO22:Sud-Est; RO31:Sud – Muntenia; RO41:Sud-Vest Oltenia; RO42:Vest; SK02:Západné Slovensko; SK03:Stredné Slovensko;SK04:Východné Slovensko
	FANNY		
SL	BG41:Yugozapaden; CZ01:Praha; EE00:Eesti; CY00:Kypros; LV00:Latvija; LT01:Sostines regionas;HU11:Budapest; HU12:Pest; MT00:Malta; RO21:Nord-Est; RO32:Bucuresti – Ilfov; SI04:Zahodna Slovenija; SK01:Bratislavský kraj;	CZ02:Strední Cechy; CZ03:Jihozápad; CZ05:Severovýchod; CZ06:Jihovýchod; CZ07:Strední Morava; CZ08:Moravskoslezsko; HU21:Közép-Dunántúl; HU22:Nyugat-Dunántúl; PL21:Malopolskie; PL22:Slaskie; PL41:Wielkopolskie; PL42:Zachodniopomorskie; PL43:Lubuskie; PL51:Dolnoslaskie; PL52:Opolskie; PL61:Kujawsko-Pomorskie; PL62:Warminsko-Mazurskie; PL63:Pomorskie; PL71:Lódzkie PL72:Swietokrzyskie; PL81:Lubelskie; PL82:Podkarpackie; PL84:Podlaskie; PL92:Mazowiecki regionalny; SI03:Vzhodna Slovenija; SK02:Západné Slovensko;	BG31:Severozapaden; BG32:Severen tsentralen; BG33:Severoiztochen; BG34:Yugoiztochen; BG42:Yuzhen tsentralen; CZ04:Severozápad; LT02:Vidurio ir vakaru Lietuvos regionas; HR03:Jadranska Hrvatska; HR04:Kontinentalna Hrvatska; HU23:Dél-Dunántúl; HU31:Észak-Magyarország; HU32:Észak-Alföld; HU33:Dél-Alföld; RO11:Nord-Vest; RO12:Centru; RO22:Sud-Est; RO31:Sud – Muntenia; RO41:Sud-Vest Oltenia; RO42:Vest; SK03:Stredné Slovensko; SK04:Východné Slovensko;

TP	BG41:Yugozapaden; CZ01:Praha; CZ02:Strední Cechy; CZ03:Jihozápad; CZ05:Severovýchod; CZ06:Jihovýchod; CZ07:Strední Morava; CZ08:Moravskoslezsko ; EE00:Eesti; CY00:Kypros; LT01:Sostines regionas; HU11:Budapest; HU21:Közép-Dunántúl; MT00:Malta; PL21:Malopolskie; PL63:Pomorskie; PL92:Mazowiecki regionalny; RO21:Nord-Est; RO32:Bucuresti – Ilfov; SI03:Vzhodna Slovenija SI04:Zahodna Slovenija SK01:Bratislavský kraj	PL22:Slaskie; PL41:Wielkopolskie; PL51:Dolnoslaskie; PL72:Swietokrzyskie; PL82:Podkarpackie;	BG31:Severozapaden; BG32:Severen tsentralen; BG33:Severoiztochen; BG34:Yugoiztochen; BG42:Yuzhen tsentralen; CZ04:Severozápad; LV00:Latvija; LT02:Vidurio ir vakaru Lietuvos regionas; HR03:Jadranska Hrvatska; HR04:Kontinentalna Hrvatska; HU12:Pest; HU22:Nyugat-Dunántúl; HU23:Dél-Dunántúl; HU31:Észak-Magyarország; HU32:Észak-Alföld; HU33:Dél-Alföld; PL42:Zachodniopomorskie; PL43:Lubuskie; PL52:Opolskie; PL61:Kujawsko-Pomorskie; PL62:Warminsko-Mazurskie; PL71:Lódzkie; PL81:Lubelskie; PL84:Podlaskie; RO11:Nord-Vest RO12:Centru; RO22:Sud-Est; RO31:Sud – Muntenia; RO41:Sud-Vest Oltenia; RO42:Vest; SK02:Západné Slovensko; SK03:Stredné Slovensko; SK04:Východné Slovensko;

Note: *SL – standard of living; TP – scientific and technological potential*
Source: *Author's own study.*

In the grouping results for selected EU regions by the standard of living of inhabitants using Ward's method, the most numerous groups of regions were group II (41 regions). For the grouping results by the scientific and technological potential using this method, the third group (39 objects) was the most numerous. Using the FANNY method, most of the regions were also included in group II (26 regions) by the standard of living of inhabitants and group III by S&T potential (33 regions). In both Ward's and FANNY methods, the group of regions with the highest standard of living of inhabitants (5 and 13 regions, respectively) was characterized by the relatively small size.

The use of both Ward's and FANNY methods to classify regions by the standard of living of inhabitants and S&T potential allowed to identify of certain spatial regularities, namely:

- regardless of the method used, in the group of regions with the highest standard of living of inhabitants, the areas with national capitals dominated;
- the largest number of regions belonging to the last group by the standard of living of inhabitants (regardless of the method used) was identified in the south-eastern part of the analysed area – Bulgaria, Hungary and Romania;
- given the level of scientific and technological potential, most of the regions with capitals of the analysed countries were classified in the first group regardless of the method used. In the case of Ward's method, out of a total of 9 regions classified in this group, there were 8 such regions (with capitals). Regarding the FANNY method, however, it was 10 out of 22 regions.
- at the level of individual countries (in the case of both methods), significant spatial heterogeneity of the standard of living of inhabitants is visible mainly in Hungary and Romania. Using Ward's method, in the case of Hungary, one region was included in group I and three regions in groups II and III while in Romania – one region in groups I and II and six regions in group III. However, for the grouping results using the FANNY method, in the case of Hungary, 2 regions in groups I and II and 4 ones in group III were identified, whereas in Romania – 2 regions in group I and 6 in group III;
- in the case of scientific and technological potential, based on the results of the classification using both methods, the greatest differentiation is visible mainly in Poland. Based on the grouping results using Ward's method, one region was classified to groups I and II each, and the rest (14) to group III. The FANNY method, on the other hand, identified a one-element group I, group II with 5 regions, and group III with the remaining 10 regions.

Then, to investigate the relationship between the standard of living of inhabitants and the S&T potential of selected EU regions, a correlation analysis was carried out, based on the non-parametric Spearman's rank correlation coefficient.

The rank coefficient is more resistant to outliers than the commonly used Pearson's correlation coefficient, but it is also recommended to apply it if the distribution of a sample does not meet the assumption of a normal distribution (Kopczewska, 2009). The value of Spearman's rank correlation coefficient between the synthetic measure of standard of living of inhabitants from selected EU regions and the scientific and technological potential was 0.6584, which allows assessing the strength of this relationship as high. The determined correlation coefficient was statistically significant at the significance level $p < 0.05$.

The next step was a canonical analysis. The number of all generated canonical variables is equal to the minimum number of variables included in any analysed sets. In this case, there are six canonical variables because this is the number of a reduced set of variables describing S&T potential.

The first pair of canonical variables, which synthetically illustrates the relationships between analysed sets of variables, explains most of the relationships between these sets. Consequently, most attention is paid to the correlation for the first canonical variable. However, the first pair of canonical variables do not fully explain the relationships between considered variables.

Therefore, it becomes necessary to determine successive pairs of canonical variables that explain relationships in other but less significant dimensions. These calculations are continued until all canonical variables are calculated, equal to the minimum number of variables in one of the sets. For this article, only statistically significant canonical variables were analysed in detail. To identify them, the already described Wilks' lambda test was carried out (Table 5).

Table 5. *Wilks' lambda test results*

Root removed	**Canonical correlation**	**x^2 test value**	**Number of degrees of freedom for χ^2 test**	**Probability level *p*for χ^2 test**	**Wilks' lambda statistic value**
0	0.9723	436.3491	156	0.0000	0.0000
1	0.9402	312.8461	125	0.0000	0.0006
2	0.9323	221.2953	96	0.0000	0.0055
3	0.8678	134.8347	69	0.0000	0.0419
4	0.8030	75.3999	44	0.0023	0.1696
5	0.7228	31.4040	21	0.0673	0.4776

Source: *Authors' own study.*

Based on the critical value of the materiality level, the first five canonical variables were included in further analysis. As already mentioned, each of the variables belonging to successive pairs of canonical variables is a linear function of the variables belonging to the first and second set of input variables. However, it is not correlated with any of the canonical variables of the same type since it explains the interactions between sets of input variables in other dimensions.

In the first stage of research, canonical weights are determined for the first pair of canonical variables with the most significant share in explaining the relationships between the analysed phenomena. Weights were then determined for statistically significant canonical variables. Canonical weights for standardized sets of input variables are equivalent to beta factors in multiple regression. They reflect the specific contribution of each variable to the generated weighted sum. The higher the absolute value, the more significant the contribution (positive or negative) to the canonical variable generation.

As the variables used for canonical analysis have undergone a standardization process, it is possible to directly compare absolute values of the canonical weights determined (Table 6). The calculations show that the biggest (absolute) values of weights for the first canonical variable have variables S18 0.7584) and NT5 0.376). Thus, it can be assumed that the correlation between the number of available beds in hospitals (per 100,000 inhabitants) and the number of researchers per 100,000 inhabitants was the major contributor to the generation of a first canonical variable.

For the determination of a second canonical variable, the most significant contribution was made by partial variables: S19 (1.4574) describing the number of dentists per 100,000 population and NT3 (0.7255) referring to human resources in science and technology (HRST) per 100,000 population. S10 made the most significant contribution to the generation of a third canonical variable (ratio of university students to the total population) and NT6 [gross domestic expenditure on R&D (GERD) per 100,000 inhabitants] variables, and to the generation of a fourth canonical variable – S10 and NT1 (employment in high-tech sectors per 100,000 inhabitants) variables. For the determination of a fifth canonical variable, S8 [young people not working or training (aged 15-24)] and NT2 (registered community designs (RCD) per 100,000 inhabitants) partial variables had the most significant contribution. Due to the number of variables applied and the number of statistically significant canonical variables, the results of the canonical analysis are presented in tabular form, not using canonical models.

In the next step, canonical factor loadings and redundancies were calculated (see Table 6). Factor loadings are identified with the correlation between canonical variables and interchangeable ones in each set. The larger they are (in terms of an absolute value), the more emphasis should be placed on this variable. According to T. Panek and J. Zwierzchowski (2013), It is recommended that those variables for which the square of this correlation coefficient is more significant than 0,5 be interpreted. For this article, a critical value for the squared correlation coefficient was assumed to be 0.4.

Table 6. *Canonical weights and factor loadings.*

Set 1	Standard-of-living factors									
	Canonical weights*					**Factor loadings**				
	I	II	III	IV	V	I	II	III	IV	V
S1	-0,01	-0,23	0,46	-0,65	0,24	0,05	-0,12	0,00	-0,05	-0,03
S2	0,25	-0,04	-0,45	-0,10	-0,69	0,68	0,21	0,11	0,12	0,28
S3	0,26	-0,01	-0,13	0,14	-0,25	0,33	0,47	-0,08	-0,12	-0,09
S4	-0,03	0,08	0,24	-0,34	0,37	-0,55	-0,39	0,18	-0,41	-0,12
S5	-0,02	0,01	0,59	-0,27	0,34	-0,29	0,00	-0,41	0,13	0,25
S6	-0,16	-0,13	-0,12	-0,18	0,34	0,54	0,48	0,06	0,18	0,17
S7	-0,10	0,12	-0,24	-0,47	-0,05	0,40	0,20	-0,02	0,19	0,01
S8	0,44	0,17	0,09	-0,17	-1,05	-0,55	-0,50	0,11	-0,30	0,06
S9	-0,01	0,00	-0,10	0,47	0,21	0,02	-0,16	-0,55	-0,18	0,32
S10	0,42	-1,14	0,86	-1,40	-0,14	0,38	0,56	-0,01	-0,51	0,02
S11	0,29	0,61	-0,63	-0,07	0,21	0,71	0,40	0,28	-0,15	-0,14
S12	0,30	-0,32	-0,50	-0,79	-0,46	0,14	0,09	0,08	0,07	0,08
S13	0,04	0,27	0,23	0,60	0,37	0,62	0,49	-0,08	-0,03	0,35
S14	0,31	0,30	0,11	0,84	0,47	0,62	0,46	-0,20	-0,15	0,18
S15	0,02	-0,05	-0,23	-0,26	-0,44	-0,15	-0,51	0,07	-0,26	0,06
S16	0,04	-0,32	0,20	-0,27	0,17	0,68	0,33	-0,02	0,07	0,26
S17	0,24	0,32	0,18	0,32	0,40	-0,21	-0,38	0,20	-0,16	0,05
S18	0,76	-1,37	-0,16	-0,82	-0,44	-0,01	0,25	-0,10	-0,39	-0,42
S19	-0,71	1,46	0,05	0,80	0,98	0,35	0,03	-0,26	-0,25	0,34
S20	-0,09	-0,05	0,16	-0,29	-0,16	0,62	0,08	-0,13	-0,36	0,18
S21	0,03	0,21	-0,14	0,72	-0,49	0,69	-0,08	-0,12	-0,14	0,27

S22	0,00	0,11	-0,54	0,27	0,35	0,37	0,45	-0,06	-0,09	0,48
S23	-0,03	-0,05	0,49	0,12	0,12	0,35	0,28	0,17	-0,18	0,32
S24	-0,01	-0,23	0,46	-0,65	0,24	0,34	-0,36	-0,33	0,08	0,09
S25	0,25	-0,04	-0,45	-0,10	-0,69	0,36	0,36	-0,46	-0,04	0,28
S26	0,26	-0,01	-0,13	0,14	-0,25	0,31	0,45	0,55	0,23	-0,16
Set 2	**Variables concerning S&T potential**									
	Canonical weights					**Factor loadings**				
	I	II	III	IV	V	I	II	III	IV	V
NT1	0,33	0,07	-0,61	1,54	-0,36	0,67	0,49	-0,46	0,26	-0,18
NT2	0,30	0,25	0,09	0,12	1,68	0,63	0,52	0,40	0,09	0,40
NT3	0,05	0,73	0,07	-1,50	-0,56	0,67	0,59	-0,22	-0,32	-0,22
NT5	0,64	-0,72	0,61	0,36	-0,67	0,90	-0,36	0,16	-0,15	-0,13
NT6	-0,08	0,25	0,67	0,22	-1,01	0,26	0,53	0,73	0,23	-0,02
NT7	0,08	-0,13	-0,48	-0,53	0,49	0,87	-0,17	-0,23	-0,22	0,09

Note: ** Statistically significant I, II, III, IV, V variables – first, second, third, fourth, fifth canonical variable, respectively*
Source: *Authors' own study.*

In the set of variables relating to the standard of living of inhabitants, the most significant factor loading for the first canonical variable is demonstrated by the S11 variable (0.7089), for the second one – S10 variable (0.5572), for the third one – S26 (0.5545), for the fourth one – S10 (-0.5099), and the fifth one – S22 (0.4802). In the case of the second set of variables for the first canonical variable, the most significant factor loading is determined by the NT5 variable (0.8967), for the second one – NT3 variable (0.5869), for the third one – NT6 (0.725), for the fourth one – NT3 (-0.3188), and the fifth one – NT2 (0.3975).

In the literature, opinions call for canonical factor loadings to interpret individual variables during the interpretation of results obtained based on a canonical analysis (Panek, 2009). This is because they are easy to understand intuitively. However, it should be borne in mind that the values of these coefficients indicate correlations of individual input variables with canonical variables and, unlike canonical weights, do not consider the effects of covariance within a given set of input variables. For this reason, the interpretation of canonical variables based on correlation coefficients may lead to different conclusions than a more complete "multidimensional" interpretation based on canonical weights (Panek and Zwierzchowski, 2013).

Based on the values of canonical weights and factor loadings, it can be concluded that the first statistically significant canonical root explained the following relationships:

- as the number of researchers increases, the population with higher education in the 25-64 age group rises;
- as employment in high-tech sectors and the number of researchers increase, the number of doctors is likely to decrease;
- the higher the number of researchers and people employed in high-technology sectors, the higher the average remuneration of employees and the higher the economic activity rate;

- the higher the number of European Union trademark (EUTM) and registered community design (RCD) applications, the higher the regional gross domestic product (PPS per capita);

When analysing the values of factor loadings for the remaining canonical roots, it is easy to see that for each partial variable – except for one case – the squared correlation coefficient between canonical and interchangeable variables in each set was much lower than 0.4. The only exception was the NT6 variable for the third canonical variable. For this reason, for the remaining canonical variables, the interpretation of factor loadings and canonical weights was abandoned.

Table 7. *Isolated variances and redundancies.*

Specification	Set of variables relating to S&T potential		A set of variables reflecting the standard of living	
	Isolated variance	Redundancy	Isolated variance	Redundancy
First canonical variable	0.4898	0.4630	0.2025	0.1914
Second canonical variable	0.2163	0.1912	0.1241	0.1097
Third canonical variable	0.1714	0.1490	0.0569	0.0495
Fourth canonical variable	0.0500	0.0377	0.0491	0.0370
Fifth canonical variable	0.0440	0.0284	0.0537	0.0346

Source: *Authors' own study.*

Then, for each statistically significant canonical variable, the mean of squares of factor loadings for each considered set was calculated, and thus the isolated variance was obtained. In turn, multiplying this average by the squared canonical correlation resulted in a redundancy value. The table below shows the values of isolated variances and redundancies (Table 7).

First – the most statistically significant – canonical variable distinguishes almost 49% of variances in the set of variables relating to S&T potential and over 20% in the second set (relating to the standard of living of inhabitants from selected EU regions). In the case of subsequent canonical variables, the degree of isolation of variances is much smaller. Fifth – the last statistically significant canonical variable – distinguishes just over 4% of the variance in the first set and over 5% in the second set.

For a set of input variables reflecting the standard of living of inhabitants from selected EU regions, we can explain, respectively, 19.1%, 11.0%, 5.0%, 3.7%, and 3.5% of the variance of the set of variables concerning conditions related to S&T potential. In turn, for a set of input variables concerning S&T conditions, we explain, respectively, 46.3%, 19.1%, 14.9%, 3.8%, and 2.8% of the variance based on the first five statistically significant canonical variables. Thus, the fourth and fifth statistically significant canonical variable already makes a small specific contribution to explaining this variability.

In the next step, total redundancy has been calculated, which is interpreted as the average percentage of the variance explained in one set of variables for a given second set, based on all canonical variables. The calculations show that with the knowledge of the values of variables describing S&T potential, it is possible to explain more than 88.41% of the variance of the variables from the set referring to the standard of living of inhabitants from selected EU regions. The calculated value of total redundancy can be evaluated as high. To obtain even better results, further research should be carried out in the future, i.a., with a different set of input variables and the changed number of these variables.

When analysing multidimensional relationships between the considered categories, high and, which is essential, statistically highly significant (Table 5), canonical correlation values should be noted. Canonical correlation cannot be interpreted as a classical correlation (e.g., Spearman's). These values shall be interpreted as correlations between the weighted sum values in each set and the weights calculated for subsequent canonical variables. The value of the largest and most statistically significant canonical correlation was over 0.97. For the last (i.e., fifth) statistically significant canonical variable, this value was over 0.72. The square of these canonical correlations is a measure of the degree of explanation by linear relationships of the variability of one set of variables, by the other of the input sets, by successive pairs of canonical variables. For the first statistically significant canonical variable, squared canonical correlation is approx. 0.95, while for the second one it is over 0.88. For the last statistically significant canonical variable, this coefficient is over 0.52. It can be assumed that this generated model describes the considered data sets relatively well.

5. Conclusions

The main objective of the research was to detect the relationships between sets of variables describing the standard of living of inhabitants from selected EU regions and S&T potential. A canonical analysis seems to be the most accurate technique that can be used to analyse multidimensional relationships between two sets of variables. In this type of research, the use of only classical correlation analysis or regression analysis – considering the multifacetedness above of the considered phenomena – seems to be insufficient. Therefore, in socio-economic analyses, the popularisation of multidimensional exploratory methods (such as e.g., canonical analysis) to identify the relationships between compiled, multifaceted categories is essential.

The canonical analysis was preceded by the construction of synthetic measures and the determination of a correlation coefficient between them. In the case of 75% of the analysed regions, the value of the synthetic standard-of-living measure for selected EU regions did not exceed 0.42 (with a minimum value of 0.30 and a maximum value of 0.56). The average value of the synthetic measure of S&T potential was 0.23. Based on results of the classical correlation analysis, it can be concluded that there is a positive, high, and statistically significant correlation dependence (Spearman's rank correlation coefficient was nearly 0.66) between the standard of living of inhabitants from selected EU regions and S&T potential (measured by synthetic measures constructed based on the TOPSIS method).

Five statistically significant canonical variables were identified in the canonical analysis. Based on the value of a redundancy coefficient determined within the canonical analysis, it can be concluded that, with the knowledge of the included variables that describe the scientific and technological potential, 88% of the variance of the variables from the set referring to the standard of living of inhabitants can be explained. In other words, almost four-fifths of the variability related to the standard of living of inhabitants from selected EU regions is determined by included partial variables relating to the scientific and technological potential. It should also be mentioned that high values of canonical correlation coefficients were identified for statistically significant canonical variables. For the most statistically significant canonical variable, this coefficient was 0.97, while for the least statistically significant – 0.72.

An issue of technology transfer between individual regions/countries was omitted in the analyses. In the future, to consider this aspect to a certain extent, it would be worthwhile to carry out modeling using spatial regression (taking into account spatial interactions between separate areas). In addition, due to the lack of relevant statistics, infrastructure issues were also omitted (e.g., a saturation of regions with universities, research institutes, laboratories). To increase the reliability of the canonical analysis, tests should be conducted on an increased number of analyzed elements. Analyses at the level of smaller spatial units would also be valuable, but it may not be easy to provide comparable statistical data.

References:

Balicki, A. 2009. Statystyczna analiza wielowymiarowa i jej zastosowanie społeczno ekonomiczne. Wydawnictwo Uniwersytetu Gdańskiego: Gdańsk.

Berbeka, J. 2006. Poziom życia ludności a wzrost gospodarczy w krajach Unii Europejskiej. Wydawnictwo Akademii Ekonomicznej w Krakowie: Kraków.

Borys, T. 1999. Wskaźniki ekorozwoju. Wydawnictwo Ekonomia i Środowisko: Białystok.

Box, G.E.P., Cox, D.R. 1964. An Analysis of Transformations. Journal of the Royal Statistical Society, 26(2), 211-243.

Cavadias, G.S., Ouarda, T.B.M.J., Bobee, B., Girard, C. 2001. A Canonical Correlation Approach to the Determination of Homogeneous Regions for Regional Flood Estimation of Ungauged Basins. Hydrological Sciences, 46(4), 499-512.

Central Statistical Office. 2004. Warunki życia ludności w 2004 r. GUS: Warszawa.

Ding, H., Jiang, T., Riloff, E. 2018. Why is an Event Affective? Classifying Affective Events based on Human Needs. AAI Workshops, 1-8.

Ebenezer, O.R. 2012. Canonical Correlation Analysis of Poverty and Literacy Levels in Ekiti State, NIGERIA. Mathematical Theory and Modeling, 2(6), 31-39.

EUROSTAT. Retrieved from: https://ec.europa.eu/eurostat/data/database.

Frey, C.B., Osborne, M. 2013. The Future of Employment: How Susceptible Are Jobs to Computerisation? Oxford Martin School: Oxford.

Gittins, R. 1985. Canonical Analysis - A Review with Applications in Ecology. Springer-Verlag: Berlin - Heidelberg - New York - Tokyo.

Grzega, U. 2015. Tradycyjne i alternatywne mierniki poziomu życia ludności w ujęciu teoretycznym. Zeszyty Naukowe Uniwersytetu Szczecińskiego. Współczesne Problemy Ekonomiczne. Globalizacja. Liberalizacja. Etyka, 10, 77-87.

Grzeskowiak, A. 2016. Satisfaction with chosen aspects of life in Poland - evaluation by canonical correlation methods. International Journal of Social Sciences, V(1), 60-71.

Hansen, M.E., Grubb F. 2002. Anthropometric Versus Income Measures of the Standard of Living: Issues of Theoretical Consistency. Retrieved from: http://www.humanics-es.com/anthro11-13-02.pdf.

Hardoon, D.R., Szedmak, S., Shawe-Taylor, J. 2003. Canonical correlation analysis; An overview with application to learning methods. University of London: London.

Hwang, C.L., Yoon, K. 1981. Multiple Attribute Decision Making: Methods and Applications. Springer: New York-Berlin/Heidelberg.

Chyba, Z. 2014. Potencjał technologiczny a kreowanie konkurencyjności przedsiębiorstw. Przegląd Organizacji, 2(889), 21-25.

Janusz, M. 2014. Poziom życia i jego przestrzenne zróżnicowanie w województwie warmińsko-mazurskim. Economics and Management, 4, 169-184.

Karpińska-Mizielska, W., Smuga, T. 1999. Instytucje infrastruktury rynkowej w kreowaniu przedsiębiorczości lokalnej. Fundacja Promocji i Rozwoju: Warszawa.

Korol, J, Szczuciński, P. 2009. Ekonometryczne modelowanie procesów gospodarki regionalnej opartej na wiedzy. Wydawnictwo Adam Marszałek: Toruń.

Latusek, A. 2004. Nowy Słownik Wyrazów Obcych. Wydawnictwo Zielona Sowa: Kraków.

Liang, Z., Changdi, S., Liming, L. 2017. Application of AHP Method and TOPSIS Method in Comprehensive Economic Strength Evaluation of Major Cities in Guizhou Province. International Conference on Computer Science and Application Engineering, CSAE14085, 687-94.

Majka, A. 2015. Taksonomiczna analiza zróżnicowania poziomu życia w Polsce. Research Papers of Wrocław University of Economics, 407, 354-63.

Malinowski, M., Smoluk-Sikorska, J. 2020. Spatial Relations between the Standards of Living and the Financial Capacity of Polish District-Level Local Government. Sustainability, 12(5)/1825, 1-27.

Max-Neef, M.A. 1991. Human Scale Development. Conception, Application and Further Reflections. The Apex Press: New York-London.

Midor, K., Wieczorek, A. 2016. Kształtowanie jakości życia pracowników w środowisku przemysłowym - studium literaturowe. In: Knosala, R. (Ed.) Innowacje w zarządzaniu i inżynierii produkcji. Oficyna WPTZP, Opole, 431-438.

Młodak, A. 2006. Analiza taksonomiczna w statystyce regionalnej. Difin: Warszawa.

Mourad, M., Perez, A., and Richardson, C. 2014. Digital Inclusion Social Impact Evaluation. Final Report, One Global Economy: Washington.

Murtagh, F., Legendre, P. 2014. Ward's Hierarchical Agglomerative Clustering Method: Which Algorithms Implement Ward's Criterion? Journal of Classification, 31, 274-295.

Naylor, M.G., Lin, X., Weiss, S.T., Raby, B.A., Lange, C. 2010. Using Canonical Correlation Analysis to Discover Genetic Regulatory Variants. PLoS ONE, 5(5), 1-7.

Noble, R.B. 2000. Multivariate Applications of Bayesian Model Averaging. Blacksburg-Virginia Polytechnic Institute, Virginia.

OECD. 2015. Directorate For Science, Technology and innovation committee for scientific and technological policy, DSTI/STP/MIN(2015)1/REV1, Daejeon.

Panek, T. 2009. Statystyczne metody wielowymiarowej analizy porównawczej. Szkoła Główna Handlowa w Warszawie: Warszawa.
Panek, T., Zwierzchowski, J. 2013. Statystyczne Metody Wielowymiarowej Analizy Porównawczej: Teoria i Zastosowania. Oficyna Wydawnicza SGH: Warszawa.
Piasny, J. 1993. Poziom i jakość życia ludności oraz źródła i mierniki ich określania. RUCH PRAWNICZY, EKONOMICZNY I SOCJOLOGICZNY, 2, 73-92.
Piekut, M. 2008. Polskie gospodarstwa domowe - dochody, wydatki i wyposażenie w dobra trwałego użytkowania. Wydawnictwo SGGW: Warszawa.
Słaby, T. 2007. Poziom i jakość życia. [In:] Panek T. (Ed.) Statystyka Społeczna. PWE: Warszawa, 99-130.
Sobczyk, M. 2010. Statystyka opisowa. Wydawnictwo C.H. Beck: Warszawa.
Stanisz, A. 2007. Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny. Tom III. StatSoft: Kraków.
Strahl, D. 2006. Gospodarka lokalna w teorii i praktyce. Prace Naukowe Akademii Ekonomicznej im. Oskara Langego we Wrocławiu: Wrocław.
Strømmen-Bakhtiar, A. 2020. Introduction to Digital Transformation: and its impact on socjety. Informing Science Press: Santa Rosa.
Świeboda, H., Sienkiewicz, P. 2010. Ekonomiczne i społeczne konsekwencje rozwoju społeczeństwa informacyjnego dla jakości życia. Zeszyty Naukowe Uniwersytetu Szczecińskiego. Ekonomiczne Problemy Usług, 57, 367-73.
Szymczak, W. 2010. Podstawy statystyki dla psychologów. Difin: Warszawa.
Ter Braak, C.J.F. 1990. Interpreting Canonical Correlation Analysis Through Biplots of Structure Correlations and Weights. PSYCHOMETRIKA, 55(3), 519-531.
Thompson, B. 1987. Fundamentals of Canonical Correlation Analysis: Basics and Three Common Fallacies in Interpretation. University of New Orleans and Louisiana State University Medical Center: New Orleans.
UNESCO. 2008. IDAMS. Internationally Developed Data Analysis and Management Software Package. UNESCO: Paris.
Warzecha, K. 2009. Poziom życia ludności Polski i pozostałych krajów Unii Europejskiej - analiza taksonomiczna. [In:] Pongsy-Kania S (Ed.) Gospodarka polska po 20 latach transformacji, osiągnięcia, problemy i wyzwania. IWiI, Warszawa, 1-16.
Weenink, D. 2003. Canonical Correlation Analysis. IFA Proceedings, 25, 81-99.
Wiśniewska, J. 2012. Zarządzanie zasobami technologicznymi przedsiębiorstw. [In:] Wiśniewska J., and Janasz, K. (Ed.) Innowacyjność organizacji w strategii inteligentnego i zrównoważonego rozwoju. DIFIN: Warszawa, 22-43.
Wysocki, F. 2010. Metody taksonomiczne w rozpoznawaniu typów ekonomicznych rolnictwa i obszarów wiejskich. Wydawnictwo Uniwersytetu Przyrodniczego w Poznaniu: Poznań.
Zeliaś, A. 2000. Dobór zmiennych diagnostycznych. [In] Zeliaś, A. (Ed.) Taksonomiczna analiza przestrzennego zróżnicowania poziomu życia w Polsce w ujęciu dynamicznym. Wydawnictwo Akademii Ekonomicznej w Krakowie: Kraków, 35-55.
Zeliaś, A. 2004. Poziom życia w Polsce i krajach Unii Europejskie. PWE: Warszawa.