

| Justyna Brzezińska

A MULTIVARIATE STATISTICAL ANALYSIS OF THE INFORMATION SOCIETY IN POLAND

Abstract

Widespread access to high-speed Internet, user-friendly public e-services and the increasing digital competence of society are the main goals for the coming years according to the latest reports published by the Central Statistical Office in Poland. These goals are included in the Operational Programme Digital Poland. This technological development is also connected with the development of economic areas and public services. The rapidly increasing significance of information and electronic services, and thus the application of information and communication technologies (ICT) in the economy, public administration (central and local), and in the everyday life of citizens has triggered a new transformation trend – a transformation towards the information society. This term describes a society for which the processing of information with the use of ICT solutions creates significant economic, social and cultural value.

In this paper we present the current state, main aspects, vision and mission of the information society in Poland and carry out a statistical analysis of the information society in Poland using multivariate statistical methods. All the calculations are based on data from the Central Statistical Office and they are conducted using R software.

Keywords: information society in Poland, multivariate statistical analysis, categorical data analysis, R software.

JEL Classification: C30, C35, C40.

1. Introduction

Information technology (IT) has a huge impact on people, society, and the economy. We briefly present some facts on the information society in 2015 in Poland that will help us understand the impact of IT on the economy.

| Justyna Brzezińska, University of Economics in Katowice, Faculty of Finance and Insurance, Department of Economic and Financial Analysis, 1 Maja 50, 40-287 Katowice, Poland, e-mail: justyna.brzezinska@ue.katowice.pl

The information and communication technology (ICT) sector plays an important role in the modern economy, directly affecting the growth of productivity, employment, productivity and, indirectly, other branches of the economy.

The development of modern enterprises is inextricably linked with the use of ICT. Companies generally use computers with broadband Internet access. The dynamic growth of Internet use in business stems from its essential characteristics, such as: lack of bureaucratic restrictions, the possibility of international cooperation among specialists, global reach and flexibility. As a result, the Internet has become a key business tool implemented in companies engaged in services, sales, information exchange, or acquiring suppliers in the financial field, and thus has become tool for building competitive advantage. In an age of broad technological development, ICT information and the knowledge linked to it is becoming a strategic product. The ability to gain such knowledge and data and its transmission through the network often determines market success. Information also is changing the business world by creating broad opportunities for business contacts, the conduct of transactions and the acquisition of market news around the world.

From 2012, the percentage of large companies with Internet access was close to 100%, which indicates saturation among this group of entities. In 2015, 92.7% of companies had Internet access and mostly used broadband (91.9%). Mobile broadband was used by over 61.5% of companies. The proportion of enterprises with their own website in 2015 was 65.4%. Nearly two-thirds of companies used their own websites to present the products and services they offer. In 2014, every fifth company received orders via computer networks, and every tenth company received orders via the Internet. In 2015, almost half of large enterprises used social media. One quarter of large enterprises used the services of cloud computing in 2015 (www.gus.pl).

The financial perspective for the 2014 to 2020 period has opened up new possibilities for supporting the development of the information society. Widespread and easy access to the Internet, user-friendly public e-services, and the increasing digital competence of society are the aims for the coming years. ICT has a bigger and bigger influence on our daily lives. In particular, the Internet is creating a new layer of participation for individuals in social and economic life. New services on the Internet affect a wide range of activities, such as information retrieval, communication, training and education, commerce and finance, participation in social networks and

leisure activities. The growing pervasiveness of ICT leads to user-friendlier applications on the one hand but requires access to ICT and basic Internet and computer skills on the other.

The goal of this paper is to provide knowledge about the information society in Poland with the use of statistical analysis as well as to identify qualitative development trends in the process of transforming Poland's society. Furthermore, we provide a survey of new trends, processes and phenomena, focusing on those groups and economic sectors of the information society in Poland which are at the stage of highest development. The statistical analysis includes correspondence analysis, hierarchical clustering using the Ward method and modern visualising tools for categorical data. We present a comprehensive description of the dynamic change of the information society in Poland and use multivariate statistical analysis to study the strengths and weaknesses of the information society in Poland using up-to-date empirical data. The main hypothesis is: there are sectors of the information society in Poland that can be clustered into similar sectors. The hypothesis will be tested using multivariate statistical methods and visualising tools for categorical data. All the calculations are conducted using R software.

2. Facts and Figures on the ICT Sector and ICT Products

In 2014, the number of enterprises hiring 10 or more persons in the ICT sector amounted to 2146 (a 6.3% increase compared to the previous year), of which 89.1% offered ICT services. Almost three quarters of ICT service enterprises provided IT services. In comparison with 2011, the number of ICT enterprises steadily increased and was higher by 24.5% (including service enterprises, which increased 29.2%). The number of persons employed in the sector amounted to 196,400 (an increase of 6.5% compared to the previous year and 10.7% compared to 2011), with persons hired in ICT services constituting over three quarters. IT services was also the field of activity in which enterprises hired the biggest number of persons of all those employed in ICT services (66.1%). The value of net revenues from sales in the ICT sector increased by 8.8% compared to 2011 and amounted to over PLN 132 billion in 2014 (only in 2013 was a slight decrease noted in comparison with 2012 – 0.7%). Services, in particular telecommunications, made the biggest contribution to the generation of revenues in the ICT sector. In 2014, ICT manufacturing enterprises earned almost two thirds of their revenue from export sales, while ICT service enterprises earned

only 15.4%. In 2011–2013, these revenues steadily decreased in ICT manufacturing enterprises but in 2014 they recorded an annual growth of 8.3%. In service enterprises, revenues from exports were continually increased. Enterprises providing IT services had the biggest share in these revenues (63.5% in 2014). An increase of expenditure on R&D in the ICT sector was noted in the years 2011–2014 (by PLN 553 million). Enterprises offering ICT services incurred approximately 90% of their expenditure on R&D in each surveyed year. In the analysed period, ICT sector enterprises were more innovative than enterprises in the economy as a whole (all manufacturing and service enterprises). The ICT sector most frequently implemented new or improved products.

According to the annual report published by the Central Statistical Office entitled *Information Society in Poland. Results of Statistical Surveys in the Years 2012–2016*, in 2015 the number of enterprises hiring 10 or more persons in the ICT sector amounted to 2045 (a 4.7% decrease compared to the previous year), of which 88.5% offered ICT services. Almost three quarters of ICT service enterprises provided IT services. The number of persons employed in the sector amounted to 214,200 (an increase of 9.1% compared to the previous year and 18.5% compared to 2012), with persons hired in ICT services constituting over three quarters of that figure. IT services was also the field of activity in which enterprises hired the biggest number of persons of all those employed in ICT services (70.2%). The value of net revenues from sales in the ICT sector increased by 9.8% compared to 2012, up to PLN 139 billion in 2015 (only in 2013 was a slight decrease noted in comparison with 2012 – 0.7%). Services, in particular telecommunications, made the biggest contribution to the generation of revenues in the ICT sector. In 2015, ICT manufacturing enterprises earned almost three quarters of their revenue from export sales, while ICT service enterprises earned under one fifth. In the analysed period, these revenues in ICT manufacturing enterprises slightly fluctuated, while in service enterprises revenues from exports were steadily on the rise. Enterprises providing IT services had the biggest share in these revenues (67.0% in 2015). An increase of expenditure on R&D in the ICT sector (by PLN 654 million) was noted in the years 2012–2015. Enterprises offering ICT services incurred over 90% of their expenditure on R&D in the ICT sector in each surveyed year. In the analysed period, ICT sector enterprises were more innovative than enterprises in the economy as a whole (all manufacturing and service enterprises). The ICT sector most frequently implemented new or significantly improved products. In 2015, the value of sold production

of ICT products decreased by PLN 1 billion compared to 2014. Over the analysed period, the value of exports and imports of ICT products steadily increased. In the years 2012–2015, exports increased by 24.4% and imports by 14.3%.

There are many papers on this topic in the literature: UNDP (2002), A. M. J. Skulimowski (*Transfer...* 2006), L. Z. Karvalics (2007) and A. Szewczyk (2007).

3. The Information Society

The Polish government, having in mind the good of Poland and its inhabitants, is striving to ensure rapid and sustainable economic growth and social development that will improve the living conditions of citizens. One of the key stimulants of economic growth is citizens' ability to acquire, accumulate and use information as a result of the dynamic development of ICT. The importance of this factor for economic growth is confirmed by various research, which concludes that ICT accounts for approximately one quarter of GDP growth and 40% of productivity growth in the European Union. The rapidly increasing significance of information and electronic services, and thus the application of information and communication technologies (ICT) in the economy, public administration (central and local), and in the everyday life of citizens has triggered a new transformation trend – a transformation towards the information society. The term “information society”, as adopted for the purposes of this paper, is defined as a society for which the processing of information with the use of ICT solutions creates significant economic, social and cultural value. This strategy is sectoral and, as such, defines the vision and mission for the development of the information society in Poland until 2013. Within each of its three areas – Human, Economy and State – it maps out strategic directions and determines the objectives that should be accomplished in order to achieve the desired development status for the information society in Poland in 2013. The creation of the strategy was preceded by a series of extensive consultations with experts representing organisations and institutions most competent to express views on the issue of information society development. The outcomes of the consultations and the strategic development directions adopted by the Polish government were the basis for the formulation in 2013 of the vision of the information society in Poland and the core principles of its development. Within the framework of the partial subject areas, the vision of the information society in Poland in 2013 is presented in detail.

This paper presents the most vital activities initiated thus far and selected directions for further action required to accomplish the adopted objectives.

Research on the links between the diffusion of ICTs and social and economic development has been underway for decades. Evidence of links between social and digital engagement, particularly with respect to the Internet, has been the focus of many studies conducted by academic as well as government institutions. These studies have consistently shown that individuals who have access to ICTs, from the telephone to the Internet, tend to have more schooling, higher incomes, and higher status occupations than do those who do not have access. To analyse this study in the next part of this paper we present a statistical analysis based on data from the Central Statistical Office on the information society in Poland. According to a report published by the Central Statistical Office entitled *Information Society in Poland. Results of Statistical Surveys in the Years 2010–2014*, ICT usage in enterprises is as follows: in 2014, 94.4% of enterprises, including almost all large enterprises, used computers. The value of this indicator in 2013 placed Poland slightly below the EU average. Since 2010, the percentage of enterprises with access to the Internet has exceeded 90%, with almost all large enterprises having such access. In 2014, the highest level was recorded in the Dolnośląskie voivodship (95%) and the lowest in the Zachodniopomorskie voivodship (89.6%). The number of enterprises using broadband (fixed and mobile 3G) connections increased in 2014 compared to the previous year. The biggest rise in Internet access using broadband connections occurred among small enterprises. The value of this indicator increased annually in the majority of voivodships, with the biggest values being in the Mazowieckie and Śląskie voivodships. The systematic growth in the share of employees using computers, including ones with access to the Internet, was noticed in the analysed period. However, the value of this indicator varied considerably depending on the region and type of activity. In 2014, over one half of enterprises equipped their employees with devices enabling mobile access to the Internet. As for large enterprises, 9 out of 10 offered their employees the possibility to use mobile devices. In 2014, 10.2% of enterprises in Poland employed ICT specialists, with the biggest activity in this regard among large enterprises. In the years 2010–2013, an increase in the percentage of enterprises receiving and sending orders via computer networks by 1.7 and 7.2 percentage points was recorded. As for the automatic sharing of information within the enterprise in 2014, the percentage of ERP and CRM system usage was almost identical and amounted to 22.0% and 21.8%, respectively. The percentage of enterprises using the Internet for

interaction with public authorities slightly decreased, reaching 88% in 2013. In the same year more than one third of enterprises incurred expenditure on purchasing ICT equipment, mostly IT equipment. Large enterprises prevailed in this group. Entities classified in the financial and insurance activities and electricity, gas, steam and conditioning supply sections were the ones doing purchasing most often.

Information published on the ITC usage of households is as follows: in 2014, 77.1% of households were equipped with at least one computer. This percentage has steadily increased in recent years and is significantly higher in households with children. The number of regular computer users also increased over the 2010–2014 period. Three quarters of households had access to the Internet in 2014, which is a 2.9 percentage point increase in comparison with the previous year. This indicator was also on the rise during the analysed period. The percentage varied depending on the type of household, degree of urbanisation, place of residence and region. Households with children more frequently owned a computer and had access to the Internet. This percentage was also higher in big cities and in Central Poland. Reported reasons for the lack of Internet access in households varied over the surveyed period, although no need for Internet access at home was mentioned most often. In the years 2010–2014, the share of households with broadband access to the Internet increased faster than access to the Internet in general and reached 71.1% in 2014. As in the previous cases, having children and a place of residence were advantageous factors. A total of 63.0% of persons aged 16–74 used the Internet on a regular basis in 2014. However, differentiation depending on age, economic activity, level of education and place of residence was also observed. The highest number of regular users was found among pupils and students (98.6%), the self-employed (85.0%) as well as residents of big cities and persons with tertiary education. The number of regular Internet users was higher among the residents of Central Poland than other regions. The Pomorskie voivodship was dominant when it came to households with computers and access to the Internet as well as persons using computers and the Internet (including regular users) in 2014. As for the purposes of using the Internet in Poland, e-mail users constituted the biggest share. In 2014, the share of e-mail users in the total population aged 16–74 amounted to 52.9%, while the share of e-mail users among Internet users was 79.5%. In 2014, over 34% of Poles aged 16–74 purchased goods and services over the Internet. Men made such purchases more frequently than women. The highest number of persons ordering or purchasing goods or services for private use in 2014 was in the

Mazowieckie voivodship. In 2014, the share of persons using e-government services in the previous 12 months amounted to 26.9%. Compared to the previous year, there was an increase in the percentage of persons using these services by 4.3 percentage points. Searching for information on the websites of public authorities was the most popular way of using e-government. The highest percentage of e-government service users was found in the Mazowieckie voivodship. In 2014, the ICT usage survey was supplemented with an additional module concerning the use of cloud computing services (storage space) and advanced ICT. The results show that nearly 9% of persons aged 16–74 used storage space on the Internet in the previous 3 months. The most common reason for using cloud computing services was the possibility to use files from several devices or locations. In 2014, over one quarter of surveyed persons used such advanced ICT as a smartphone and 14% of persons had a smart TV set at home, enabling them to access the Internet and use advanced functions. In 2014, an additional survey covered persons aged 12–15, among whom almost every person (99.2%) had used a computer of any type in the past, with regular computer users amounting to 97.2%. A total of 98.6% of 12–15 year-olds used the Internet, with regular Internet users amounting to 95.4%, which is 32.4 percentage points more than among persons aged 16–74. Use of social media was the most popular purpose of using the Internet for 77.2% of 12–15 year olds, which is 40.4 percentage points more than among persons aged 16–74. Persons using the Internet to search for information in order to facilitate learning also accounted for a substantial share (59.2%) of 12–15 year olds.

4. Research Methodology

The primary objectives of the research required commencing work on theoretical and empirical characteristics. Various statistical research methods were applied in the paper in order to test the research hypothesis.

We present a multivariate statistical analysis of the information society in Poland with the use of R software. We use data from the report published by the Central Statistical Office entitled *Information Society in Poland*. All the calculations were conducted in R software using: `vcd`, `vcdExtra`, `Mass` and `ca` packages.

5. The Survey Results and Application in R

We present the analysis using data on the number of enterprises and employees in the ICT sector in 2011–2014 (Table 1).

First, we conduct a correspondence analysis for the number of enterprises in the ICT sector in 2011–2014. The analysis is based on data from the Central Statistical Office on the number of enterprises in the ICT sector in 2011–2014.

We present the results of the correspondence analysis in Table 1.

Table 1. Number of Enterprises in the ICT Sector

Specification	2011	2012	2013	2014
ICT production	245	239	225	235
ICT wholesale	190	207	230	235
Telecommunications	219	231	258	289
IT services	1070	1181	1305	1387

Source: Central Statistical Office.

The percentage being explained by the first dimension is 94.22%, and for the second dimension – 5.38%. These two dimensions explain 99.6% of the total inertia. The total inertia is 0.0017, which shows that there is a very weak association between two variables: year and specification. The row and column masses, chi-square distance and inertia for both categories of row and column are presented below:

Principal inertias (eigenvalues):

	1	2	3
Value	0.001629	9.3e-05	7e-06
Percentage	94.22%	5.38%	0.4%

Rows:

	ICT production	ICT wholesale	Telecommu- nications	ICT services
Mass	0.121869	0.111283	0.128712	0.638136
ChiDist	0.107785	0.015622	0.030357	0.016220
Inertia	0.001416	0.000027	0.000119	0.000168
Dim. 1	-2.669651	0.091573	0.490427	0.394954
Dim. 2	0.207310	-1.431880	2.366810	-0.267273

Columns:

	2011	2012	2013	2014
Mass	0.222566	0.239866	0.260522	0.277046
ChiDist	0.062238	0.023124	0.032682	0.040783
Inertia	0.000862	0.000128	0.000278	0.000461
Dim. 1	-1.532660	-0.524810	0.765233	0.966059
Dim. 2	0.664981	-0.877842	-1.068300	1.230398

We can display the results of the simple correspondence analysis in the form of a perceptual map (Fig. 1).

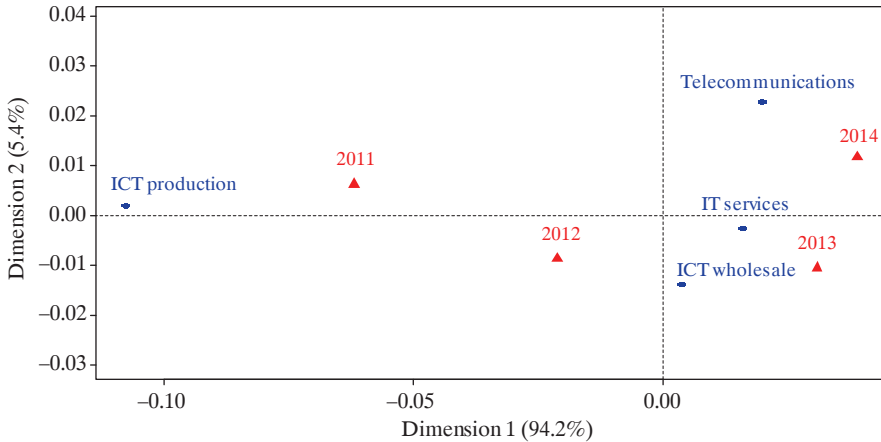


Fig. 1. Two-dimensional Perceptual Map for Correspondence Analysis for the Number of Enterprises in the ICT Sector

Source: author's own elaboration.

Looking at the graph (Fig. 1) we see that ICT production is situated very close to the year 2011. ICT wholesales and IT services are related with the years 2012 and 2013. Finally, telecommunications is situated very near to the year 2014. This graphical presentation may suggest that there is a trend moving ICT production from 2011 to telecommunications services in 2014.

To present graphically a tree diagram for categories of rows for the number of enterprises in the ICT sector (ICT production, ICT wholesale, Telecommunications, and IT services), we will apply agglomerative hierarchical clustering. What follows is a dendrogram of the results of running this data through the Ward clustering algorithm (Fig. 2).

We can see that there are three clusters of ICT specifications separated. One cluster contains two categories: ICT production and ICT wholesale, the second cluster contains only one category – Telecommunications, and last cluster also contains one category – IT services. The analysis is based on data from the Central Statistical Office on the number of employees in the ICT sector in 2011–2014.

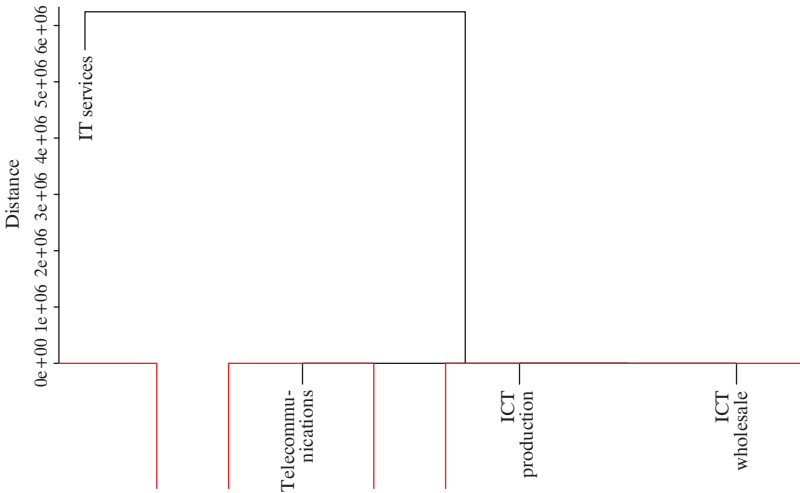


Fig. 2. Dendrogram for the Number of Enterprises in the ICT Sector Using the Ward Method

Source: author’s own elaboration.

A correspondence analysis was also conducted for the number of employees in the ICT sector (Table 2).

Table 2. Number of Employees in the ICT Sector

Specification	2011	2012	2013	2014
ICT production	44,930	41,150	36,892	39,337
ICT wholesale	10,363	10,598	11,372	11,496
Telecommunications	46,516	43,890	42,634	41,786
IT services	75,539	85,178	93,422	103,739

Source: Central Statistical Office.

The percentage being explained by first dimension is 97.72%, and for the second dimension – 2.27%. These two dimensions explain 99.99% of the total inertia. The total inertia is 0.0067, which shows that there is a very weak association between two variables: year and specification. The row and column masses, chi-square distance and inertia for both categories of row and column are presented below:

Principal inertias (eigenvalues):

	1	2	3
Value	0.006547	0.000152	1e-06
Percentage	97.72%	2.27%	0.01%

Rows:

	ICT production	ICT wholesale	Telecommu- nications	ICT services
Mass	0.219680	0.059321	0.236622	0.484377
ChiDist	0.099928	0.023115	0.076194	0.080015
Inertia	0.002194	0.000032	0.001374	0.003101
Dim. 1	-1.216980	0.110858	-0.921128	0.988340
Dim. 2	1.378733	-1.703701	-1.282066	0.209650

Columns:

	2011	2012	2013	2014
Mass	0.240035	0.244729	0.249471	0.265765
ChiDist	0.122501	0.028559	0.054708	0.089983
Inertia	0.003602	0.000200	0.000747	0.002152
Dim. 1	-1.513489	-0.351812	0.631948	1.097723
Dim. 2	0.224847	0.119762	-1.578475	1.168343

We can display the results of the simple correspondence analysis in the form of a perceptual map (Fig. 3).

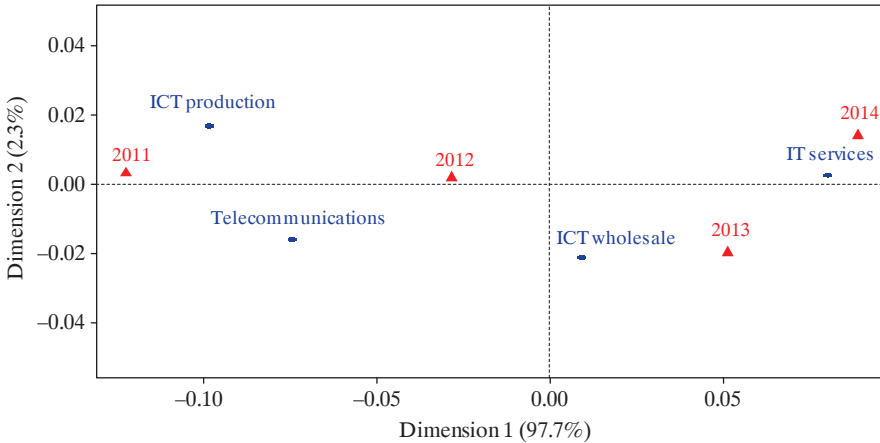


Fig. 3. Two-dimensional Perceptual Map for Correspondence Analysis for the Number of Employees in the ICT Sector

Source: author's own elaboration.

Looking at the two-dimensional perceptual map above (Fig. 3), we can see that the points are situated in a different way to the analysis conducted for enterprises. ICT production is situated very close to the year 2011. Telecommunications are related with the year 2012. ICT wholesale is situated very near to the category 2013, and IT services are plotted near to the year 2013.

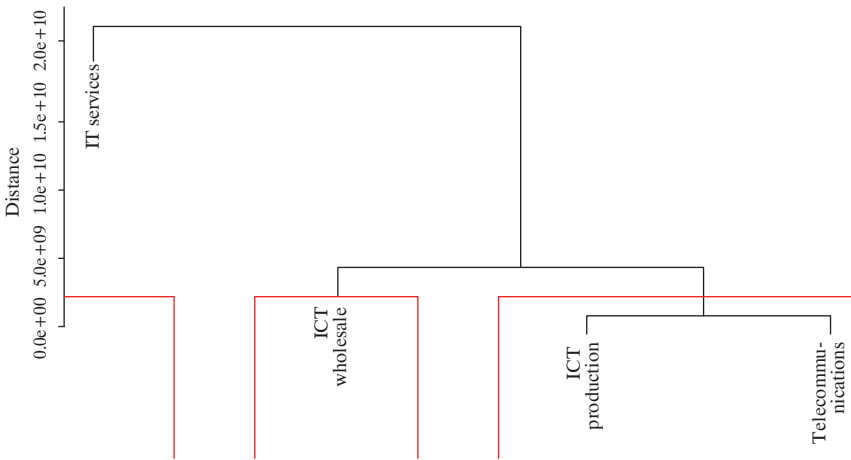


Fig. 4. Dendrogram for the Number of Employees in the ICT Sector Using the Ward Method

Source: author's own elaboration.

Looking at the dendrogram for the number of employees in the ICT sector (Fig. 4), we can see that there are three separated clusters. The first cluster contains two categories: ICT production and Telecommunications, the second cluster contains one category – ICT wholesale, and finally the last cluster also contains one category – IT services. Using multivariate methods, we can see which categories of analysed variables belong to the cluster of objects that are similar to each other.

6. Visualising the ICT Sector

We can also provide a graphical presentation for categorical data in R using the `vcd` library. Here we provide basic plots for visualising categorical data such as: mosaic, sieve, association and double-decker plot.

The mosaic plot was first introduced by J. A. Hartigan and B. Kleiner (1981, 1984), and M. Theus and R. W. Lauer (1999), and is one of the most popular and useful methods of log-linear modelling. Mosaic plots are a generalisation of multi-way tables and a direct tool for testing log-linear models. A mosaic represents each cell of the table by a rectangle (or tile) whose area is proportional to the cell count. The mosaic is constructed by dividing a unit square vertically by one variable, then horizontally by another. Further variables are introduced by recursively subdividing each tile by the conditional proportions of the categories of the next variable in each cell, alternating on the vertical and horizontal dimensions of the display.

First, we apply the `mosaicplot` function to present the mosaic plot for our dataset, the number of enterprises and number of employees, respectively (Fig. 5).

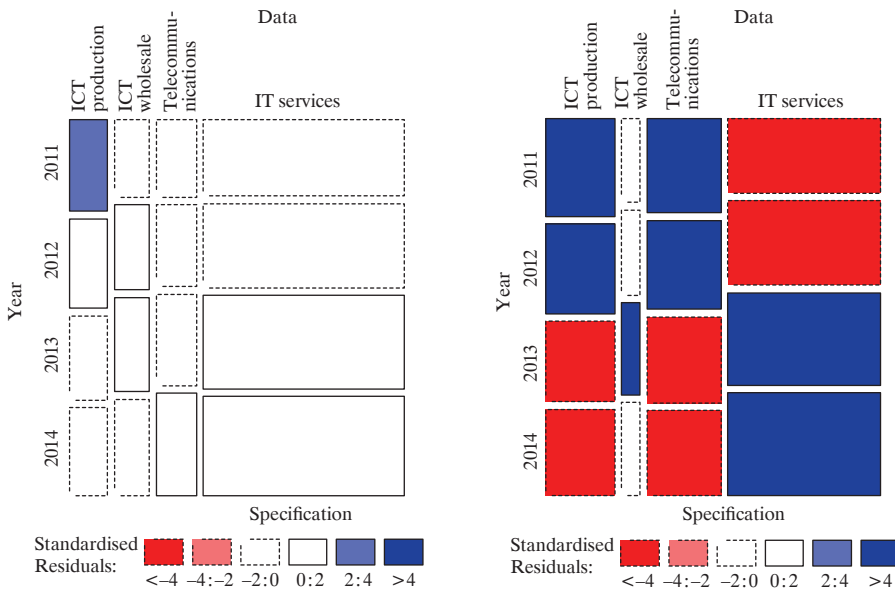


Fig. 5. Mosaic Plot for the Number of Enterprises and Employees in the ICT Sector
Source: author's own elaboration.

The mosaic plots show that there are strong differences between the ICT sector for enterprises and employees. From the mosaic plot we can conclude that blue means that there are more observations in that cell than would be expected under the null model (independence). Red means that there are

fewer observations than would have been expected. White cells in the first plot indicate very small Pearson’s residuals for enterprises, which means that the difference between the observed and theoretical cell counts is very small and the model is close to independence. By contrast, in the second picture there are large Pearson’s residuals shown in blue (positive) and red (negative) for employees. This means that the dependence between the analysed variables is strong.

H. Riedwyl and M. Schüpbach (1983, 1994) proposed a sieve diagram (later called a parquet diagram) based on this principle. In this display, the area of each rectangle is proportional to expected frequency, and observed frequency is shown by the number of squares in each rectangle. Hence, the difference between observed and expected frequency appears as the density of shading, using colour to indicate whether the deviation from independence is positive or negative. In the next step we apply the sieve function to present the sieve plot for the number of enterprises and number of employees, respectively (Fig. 6).

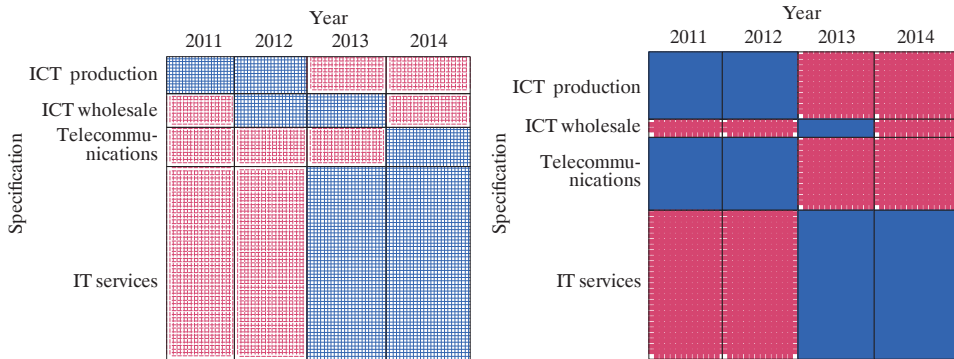


Fig. 6. Sieve Plot for the Number of Enterprises and Employees in the ICT Sector
 Source: author’s own elaboration.

Each rectangle in the sieve plot is then cross-ruled based on the observed frequency. The deviations from independence are reflected in the density of the shading. Denser shading indicates that the observed frequency is greater than expected, while sparse shading indicates that the observed frequency is less than expected. As an additional cue, positive and negative departures from independence can be coded with different colours.

From the analysis of Figure 6 we can observe stronger differences for employees compared to enterprises. The differences for IT production and

services in 2011–2014 are the same sign, but the density is stronger for the number of employees, which indicates the stronger dependence of variables.

Another plot available in R software that can be used for visualising categorical data is `assoc` in the `vcd` package (Meyer, Zeileis & Hornik 2005). It produces a Cohen-Friendly association plot that indicates deviations from the independence of rows and columns in a two-dimensional contingency table. If the observed frequency of a particular cell is greater than the expected one, the box rises above the baseline and is shaded in the colour specified by the first element of `col`, which defaults to black; otherwise, the box falls below the baseline and is shaded in the colour specified by the second element of `col`, which defaults to red.

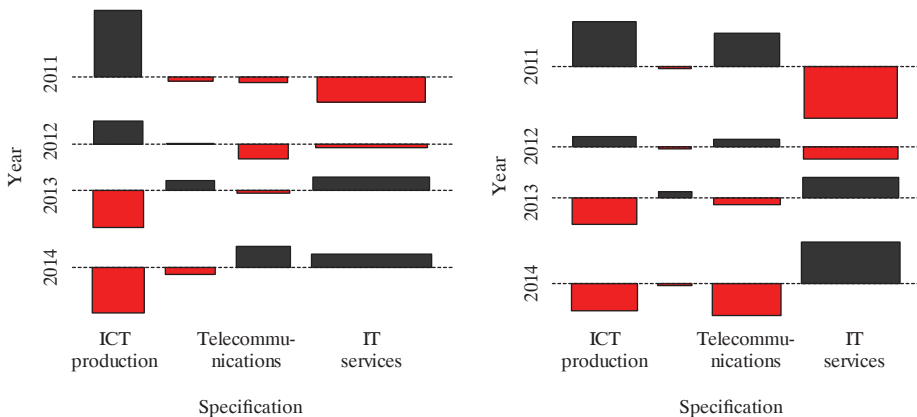


Fig. 7. Association Plot for the Number of Enterprises and Employees in the ICT Sector

Source: author's own elaboration.

From the analysis of Figure 7 we can see that the comparison between the number of enterprises and employees in the ICT sector is very similar in term of deviations for ICT production and IT services during the years 2011–2014. However, it varies for ICT wholesale (second column) and Telecommunications (third column). The differences in such ICT areas are very similar to those observed in the sieve plot (Fig. 6), indicating differences in particular specification.

In order to optimise the comparison of all crossings of all input variables, a doubledecker plot is most efficient. This plot can be done using the `vcd` package in R software. In a doubledecker plot the conditional layout of the mosaic plot is linearised by putting all category crossings side by side just

as in a spine plot. This plot is in grey, with a different colour shade for each category.

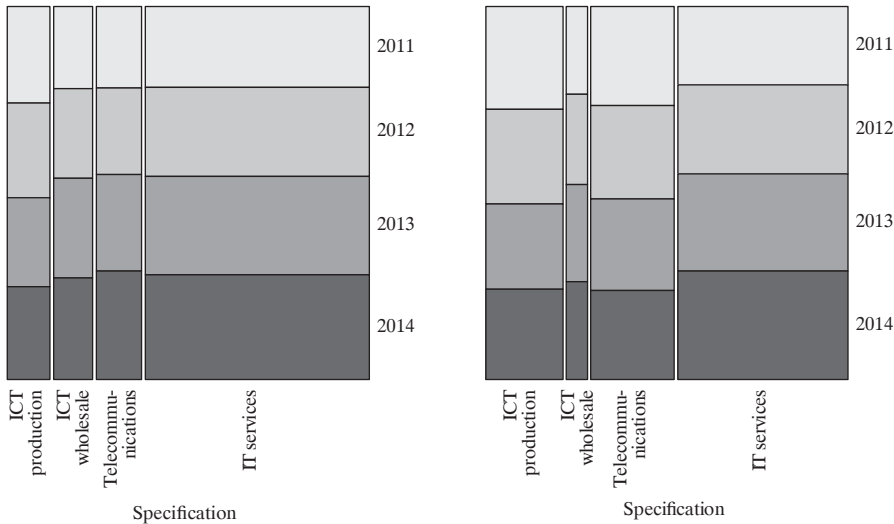


Fig. 8. Doubledecker Plot for the Number of Enterprises and Employees in the ICT Sector

Source: author's own elaboration.

From the plots in Figure 8 we can see the structure of our datasets. For the number of enterprises in the ICT sector we see that the highest number of enterprises falls in the IT services category, and the largest area is for 2014. Smaller numbers are visible for enterprises and employees in the ICT production, ICT wholesale and Telecommunications. There is a slight difference for the number of employees in the ICT sector, where the largest areas are also for IT services, and the smallest area is for ICT wholesale.

7. Conclusions

The statistical methods applied in the empirical part of this paper allow us to see the deeper structure of the number of enterprises in the ICT sector as well as the number of employees in the ICT sector in 2011–2014. The analysis is based on data from the Central Statistical Office. Statistical methods for analysing categorical data were applied, such as: correspondence analysis and hierarchical clustering analysis using the Ward method. As a result, a perception map and dendrogram were obtained for

a number of enterprises as well as employees in the ICT sector in 2011–2014. Such methods allow us to see which of the analysed categories are similar to each other in successive years. We also used the `vcd` package to visualise categorical data. Mosaic, sieve, association and doubledecker plots were presented for the number of employees and enterprises in 2011–2014 and an interpretation was provided.

The analysis allows us to conclude that, using correspondence analysis for enterprises, three clusters can be distinguished: ICT production, telecommunication, and IT services and ICT wholesale in one cluster. For the employees group, there are also three clusters: ICT production, telecommunication, and ICT wholesale and IT services in one cluster. Similar results were obtained in the hierarchical analysis. Three clusters were separated for enterprise data: IT services, telecommunication, and in last cluster ICT production with ICT wholesale. For employee data there were also three clusters, but with different categories: IT services, ICT production, and in the last cluster ICT production with telecommunication. The analysis conducted in the paper shows different business areas that are similar in the years 2011–2014. It also shows the trend over time, which areas are more developed and on top in successive years, and how the passage of time in business enables a change from one area of activity to another. The conducted analysis allowed the hypothesis to be tested and showed that there may be some clusters in different sectors of the Polish economy in the area of the information society. The conducted research may be used in future studies to analyse the information society in terms of sectors and future trends and economic situation in order to predict and analyse clusters and groups similar to each other in terms of characteristics. Also graphical presentation, which is unique as regards analysing the information society problem, may be used for a detailed interpretation of results. The use of computer software and the application of R software as well as the use of advanced visualising methods are added value and a unique feature of research in this area.

Bibliography

- Hartigan, J. A. and Kleiner B. (1981) “Mosaics for Contingency Tables” in W. F. Eddy (ed.) *Computer Science and Statistics: Proceedings of the 13th Symposium on the Interface*. Springer: New York, 268–73, https://doi.org/10.1007/978-1-4613-9464-8_37.
- Hartigan, J. A. and Kleiner, B. (1984) “A Mosaic of Television Ratings”. *The American Statistician* 38(1): 32–35, <https://doi.org/10.2307/2683556>.

- Karvalics, L. Z. (2007) *Information Society – What Is It Exactly?* Budapest: Network for Teaching Information Society.
- Meyer, D., Zeileis, A. and Hornik, K. (2005) “The Strucplot Framework: Visualizing Multi-way Contingency Tables with vcd”. *Report 22*. Research Report Series. Department of Statistics and Mathematics. Vienna: Wirtschaftsuniversität Wien.
- Riedwyl, H. and Schüpbach, M. (1983) “Siebdiagramme: Graphische Darstellung von Kontingenztafeln”. Technical Report No. 12, Institute for Mathematical Statistics. Bern, Switzerland: University of Bern.
- Riedwyl, H. and Schüpbach, M. (1994) “Parquet Diagram to Plot Contingency Tables” in F. Faulbaum (ed.) *Softstat '93: Advances in Statistical Software*. New York: Gustav Fischer, pp. 293–99.
- Szewczyk, A. (2007) *Spoleczeństwo informacyjne. Problemy rozwoju* [Information society. Problems of development]. Warsaw: Difin.
- Theus, M. and Lauer, R. W. (1999). “Technical Report 12, Visualizing Loglinear Models”. *Journal of Computational and Graphical Methods* 8 (3): 396–412.
- Transfer technologii w informatyce i automatyce* (2006) [Technology transfer in ICT] ed. A. M. J. Skulimowski. Kraków: Progress & Business Publishers.
- UNDP (2002) “Poland and the Global Information Society: Logging on. Human Development Report”. Warsaw: United Nations Development Programme, http://www.scholze-simmel.at/it_star/ws1/poland.pdf (accessed: 3 November 2017).

Abstract

Wielowymiarowa analiza statystyczna społeczeństwa informacyjnego w Polsce

Powszechny dostęp do szybkiego internetu, przyjazne dla użytkownika e-usługi publiczne i rosnące kompetencje cyfrowe społeczeństwa są głównymi celami na najbliższe lata według najnowszych raportów opublikowanych przez Główny Urząd Statystyczny w Polsce. Cele te zawarte są w Programie Operacyjnym Polska Cyfrowa. Rozwój technologiczny wiąże się również z rozwojem obszarów gospodarczych i usług publicznych. Szybko rosnące znaczenie informacji i elektroniki usług, a tym samym zastosowania technologii informacyjnych i komunikacyjnych (Information and Communication Technologies – ICT) w gospodarce, administracji publicznej (centralnej i lokalnej), a także w codziennym życiu obywateli spowodowało nowy trend transformacji – transformację w kierunku społeczeństwa informacyjnego. Termin ten opisuje społeczeństwo, dla którego przetwarzane informacje z wykorzystaniem rozwiązań ICT stwarzają istotną ekonomiczną, społeczną oraz kulturową wartość. W artykule przedstawiono obecny stan, główne aspekty oraz wizję i misję społeczeństwa informacyjnego w Polsce. Przeprowadzono statystyczną analizę społeczeństwa informacyjnego w Polsce za pomocą wielowymiarowych metod statystycznych. Wszystkie obliczenia oparto na danych pochodzących z Głównego Urzędu Statystycznego i wykonano z wykorzystaniem programu R.

Słowa kluczowe: społeczeństwo informacyjne w Polsce, wielowymiarowa analiza statystyczna, analiza danych jakościowych, program R.