Graphical representation of the relationships between qualitative variables concerning the process of hospitalization in the gynaecological ward using correspondence analysis

Anna Justyna Milewska[1], Dorota Jankowska[1], Urszula Górska[1], Robert Milewski[1], Sławomir Wołczyński[2]

[1] Department of Statistics and Medical Informatics, Medical University of Bialystok, Poland

[2] Department of Reproduction and Gynaecological Endocrinology, Medical University of Bialystok, Poland

Abstract. Correspondence analysis is an exploratory method that allows for the analysis of categorical variables. Its aim is to graphically present the relationship between the variables shown in a contingency table, using points in two- or three-dimensional space, with a minimum loss of information about the diversity of rows and columns. Correspondence analysis developed from the 40's of the 20th century, but the name in its current form gained popularity through the work of O. Hill, J.-P. Benzécrie and M. Greenacre, among others. Now, thanks to wide access to statistical packages, carrying out calculations does not create problems for researchers, and the interpretation of the results is simple and generally consistent with earlier intuitive assumptions. Correspondence analysis is used in many fields. The use of correspondence analysis for data describing the hospitalization process shows interesting relationships. It helped to depict issues such as an aging population, shorter hospital stay, or the migration of patients from other provinces.

Introduction

The development of technology causes that more and more often we are dealing with large databases containing many variables. Researchers often do not have full knowledge of the existence of multiple relationships and structures which are hidden in them. The development of modern methods of exploration and increasing performance of computers shortens the time needed to explore even the most complex and invisible at first sight relationships. More and more common is the use of data mining tools and techniques to reduce the variables. Their skilful use allows to explore a variety of databases containing information from different fields of science. It is also worth paying attention to the fact that discovery of dependencies is only the first step of the analysis. Thousands or even millions of variables generate the

ISBN 978–83–7431–336–0 ISSN 0860-150X

existence of hundreds of potential relationships. Selecting the most interesting and valuable can be labour-intensive as well as time consuming. Often their interpretation is cumbersome and requires deep knowledge of the field. Graphical methods designed specifically for the presentation of the results of individual analyzes may be helpful. Correspondence analysis gives this possibility. It allows the transparent and clear presentation of many relationships on one illustration.

Correspondence Analysis

Correspondence analysis is a technique that allows the examination of the relationship between qualitative variables (nominal and ordinal), which are common in medicine. The analysis of such data, which can be found in the contingency table, is started from the verification of the hypothesis showing no relationship between the characteristics using the χ^2 test. This procedure provides information only about the importance of the relationship between the variables. To indicate what is the nature of relations correspondence analysis can be used. On the other hand, it is an exploratory technique of reviewing large data sets. This tool enables the detection of associations between the two features with a graphical representation of the collected data [7]. This is done by creating the so-called correspondence map presenting relationships between variables. It therefore allows defining hypotheses, which can then be verified in a more formal way.

Correspondence analysis has developed since the 40's of the 20$^{\text{th}}$ century but the name in its current form gained popularity through the work of O. Hill, J.-P. Benzécrie and M. Greenacre, among others. Its idea is to create maps that graphically illustrate the contingency table which summarizes the collected data, where each row and each column is represented by one point. The main aim of this method is to show the set of points in the space of a maximum three dimensions, with a minimum loss of information on the diversity of rows and columns [27].

Correspondence analysis is carried out according to a certain schema [5]. It will be presented with a fictional example. [Tab. 1] presents information about the smoking status, depending on the level of physical activity which have been collected among 200 respondents.

The analysis of collected data begins with the creation of the correspondence matrix P. It is obtained by converting the contingency table in

Tab. 1. The relationship between smoking status and physical activity level of respondents

Smoking status	Physical activity level			
	(1) none	(2) seldom	(3) regularly	TOTAL
(a) none	12	22	35	**69**
(b) light	6	14	18	**38**
(c) medium	25	16	4	**45**
(d) heavy	28	20	0	**48**
TOTAL	**71**	**72**	**57**	**200**

a matrix of relative frequencies (dividing the number in each cell by the total number):

$$P = \begin{bmatrix} 0.060 & 0.110 & 0.175 \\ 0.030 & 0.070 & 0.090 \\ 0.125 & 0.080 & 0.020 \\ 0.140 & 0.100 & 0.000 \end{bmatrix}$$

The next step is to determine the matrix of row profiles (and column profiles). It is created by dividing the relative frequencies in each row (column) of correspondence matrix by the sum of relative frequencies across the row (column):

$$W = \begin{bmatrix} 0.17 & 0.32 & 0.51 \\ 0.16 & 0.37 & 0.47 \\ 0.56 & 0.36 & 0.09 \\ 0.58 & 0.32 & 0.00 \end{bmatrix} \qquad K = \begin{bmatrix} 0.17 & 0.31 & 0.61 \\ 0.08 & 0.19 & 0.32 \\ 0.35 & 0.22 & 0.07 \\ 0.39 & 0.28 & 0.00 \end{bmatrix}$$

Obtained row (column) profiles can be graphically represented in the space generated by the columns (rows) of the correspondence matrix. Particular frequencies in the profile can be understood as other coordinates in the considered space. In the analyzed example there will be illustrated row profiles in 3-coordinates system corresponding to physical activity categories [Fig. 1]. An analogous presentation of the column profiles is not possible due to the four-dimensional space generated by the rows of the matrix P.

Summed relative frequencies in each row of correspondence table is called the row mass (similarly for columns). This measure provides information about the significance of the rank of each row. An important parameter for the correspondence analysis is the average row (column) profile. It is

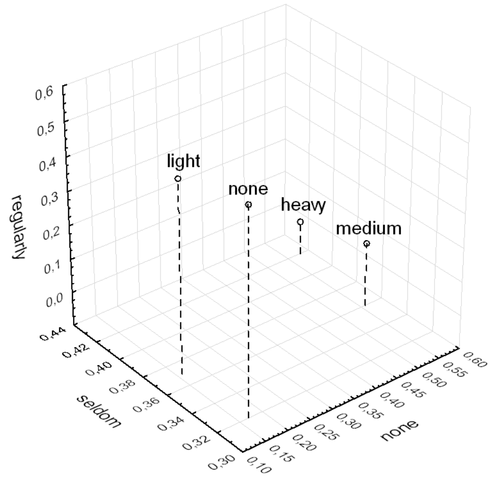

Fig. 1. Points represent row profiles in the space generated by the three levels of physical activity

obtained as a part of a summary row (column) of the contingency matrix in the total number of investigated population. In this example, we have:

– average row profile:

$$\begin{array}{ccc} (1) & (2) & (3) \\ 0.335 & 0.36 & 0.285 \end{array}$$

– average column profile:

$$\begin{array}{cc} (a) & 0.345 \\ (b) & 0.190 \\ (c) & 0.225 \\ (d) & 0.240 \end{array}$$

It is obvious that individual coordinates of average row (column) profile are the masses of corresponding columns (rows). The average profile is the center of mass of analyzed profiles.

Comparing and analyzing the profiles is done by determining the distance between them using χ^2 metric:

$$\chi^2 = d^2(p_i, p_i') = \sum_{i=1}^{k} \frac{(p_i - p_i')^2}{\bar{p}_i}$$

where:

k – dimensions of profiles

p_i, p_i' – subsequent profiles coordinates, where the distance is calculated

$\bar{p}_i$ – subsequent average profiles coordinates

It is worth noting that in this way, there can be only compare profiles of different categories of the same variable. Profiles scattering around the average profile is determined by the so-called total inertia which is calculated from the formula:

$$\Lambda^2 = \sum_{j=1}^{w} m_j d_j^2(p_j, \bar{p})$$

where:

w – number of rows (columns)
m_j – mass of j-row (column)
$\bar{p}$ – average row (column) profile
$d_j^2(p_j, \bar{p})$ – distance from row (column) profile to average profile measured with a χ^2 metric

Inertia for rows and for columns are equal. Its maximum value is $\min(w, k) - 1$ where w and k are the number of rows and columns of the contingency table respectively. The small value of inertia indicates slight differences between the profiles and the average profile. In such a situation we have in the analyzed example, because $\Lambda^2 = 0.299505$. In addition, the smaller the inertia is, the smaller the chance of occurrence of statistically significant relationships between the studied characteristics is. This follows from the relation:

$$\chi^2 = \Lambda^2 n$$

where:

χ^2 – value of the χ^2 test statistic
n – the total sample size

Another key step in the algorithm, by which correspondence analysis runs, is the projection of rows and columns profiles matrix for up to a three-dimensional space. At the same time the largest part of the information on the diversity of rows and columns should be maintained. This step is a response to a problem with graphical presentation of large number of dimensions. This is done by the method called singular value decomposition (SVD) [22, 27]. It runs in the following way:

– Symmetrical standardization of correspondence matrix P:

$$P = [p_{ji}] \rightarrow A = [a_{ji}]$$

where: $a_{ji} = \frac{p_{ji} - p_{j.}p_{.i}}{\sqrt{p_{j.}p_{.i}}}$

– Presentation of formed matrix A as a product of the following three matrixes:

$$A_{wxk} = U_{wxr} \cdot D_{\lambda_{rxr}} \cdot V_{rxk}^T$$

where:

- w, k – numbers of rows and columns respectively
- A – matrix with rank equal to r
- D_λ – matrix that has the diagonal with nonzero singular values AA^T in non-decreasing order
- U – matrix whose columns are the orthonormal eigenvectors corresponding to eigenvalues $\lambda_1^2, \lambda_2^2, \ldots$ of AA^T matrix
- V – matrix whose columns are the orthonormal eigenvectors corresponding to eigenvalues $\lambda_1^2, \lambda_2^2, \ldots$ of $A^T A$ matrix

In such distribution the columns of U are orthonormal basis for the columns of the matrix A. Thus they form the principal axes of subspace projection of category stored in columns. Similarly, the columns of the matrix V gives orthonormal basis for the transposed rows of the matrix A and hence generate a subspace projection of the principal axes of categories stored in rows.

There is the following relationship between the eigenvalues of the matrix $A^T A$ and the total inertia:

$$\Lambda^2 = \sum_{i=1}^{r} \lambda_i^2$$

where:

- r – $\operatorname{rank} A$, $r = \min(w, k) - 1$
- w, k – number of rows and columns in A matrix respectively

This relationship allows to determine how much of the total inertia is explained by the i-th factor, thanks to the determination of the quotient $\frac{\lambda_i^2}{\Lambda^2} \cdot 100\%$. What more it gives the possibility to choose the dimension of space, which will take place in the projection of the analyzed correspondence matrix. It is the minimum n such that:

$$\sum_{i=1}^{n} \frac{\lambda_i^2}{\Lambda^2} \cdot 100\% \geq m$$

where m is a predetermined level of explanation of the total inertia (eg 75% or 80%). Such proceeding provides the best selection of space that will represent the considered correspondence matrix in the fullest manner and will ensure the least loss of information resulting from the reduction of dimension. In the analyzed example $\lambda_1^2 = 0.296857$. It allows to explain 99.12% of inertia. In the case of two-dimensions space it reproduces full value of inertia, as is shown in [Tab. 2].

Tab. 2. The cumulative percentage of inertia explained by the individual eigenvalues

Number of dimensions	Total inertia = 0.29951; χ^2 = 59.901; *df* = 6; *p* = 0.0000				
	Singular Values	Eigenvalues	Percent of inertia	Cumulative percent	χ^2
1	0.544845	0.296857	99.11556	99.11556	59.37130
2	0.051468	0.002649	0.88444	100.00	0.52979

After selecting the dimension of the projection space remains only the computation of new coordinates of rows and columns profiles [4–5]:

– principal coordinates for row profiles $F = D_r^{-1} \cdot U \cdot D_\lambda$
– principal coordinates for column profiles $G = D_c^{-1} \cdot V \cdot D_\lambda$

where:

D_r – diagonal matrix in which coefficients lying on the diagonal are sums of frequencies of appropriate rows of the correspondence matrix P

D_c – diagonal matrix in which coefficients lying on the diagonal are sums of frequencies of appropriate columns of the correspondence matrix P

To create an n-dimensional map of correspondence, the first n columns of the matrix F is used to determine the coordinates of rows and analogously the first n columns of the matrix G to determine the coordinates of columns [8]. Such map in one coordinate system illustrates the best representations of row and column profiles, despite the fact that these profiles exist in different spaces [9]. It allows to analyze and interpret the distance between the point imaging one of the categories of analyzed features and the center of projection or between other points representing the various categories of the same variable. It also provides the ability to draw conclusions about the coexistence of different categories of comparable qualitative features. Row and column profiles lying close to each other create a combinations of category occurring together more often than would result from independence of variables.

2-dimensional map of the correspondence for the analyzed example is shown on [Fig. 2]. From the arrangement of points we may conclude that people practicing sport regularly usually do not smoke. Moreover, among medium smokers are those who show lack of physical activity.

To assess the accuracy of our correspondence map a parameter is used which refers to the quality of individual points. It is defined as the quotient

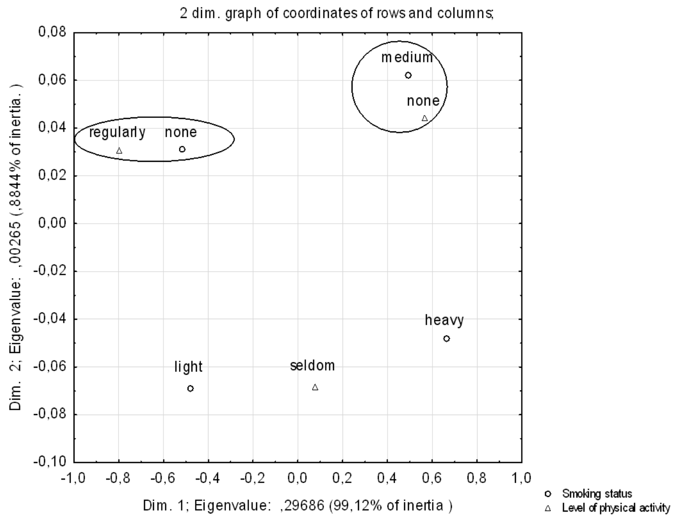

Fig. 2. Correspondence map, which represents the relationship between smoking status and physical activity level of respondents

of the square of the distance between a certain point and the beginning of the selected coordinate system, by the square of the distance between this point and beginning of the coordinate system, with the largest number of dimensions in the considered situation. The quality of representation of row or column profiles in low-dimensional space is perfect when the measure is equal to 1 [27].

Application of correspondence analysis in scientific research

The method discussed in the current work is used to analyze qualitative data. Many scientific disciplines describe the objects of their interest with a nominal scale. It is worth especially noting that medical data describe cases of patients who have multiple units and many of the symptoms of disease at the same time. We can describe trials and populations of this type using contingency tables. However, it appears that many labels of the individual variables (e.g. identified disease entities) and a large number of characteristics lead to the creation of many tables. Their complexity makes it impossible to observe hidden dependencies. Correspondence analysis allows to look at the same time at many aspects. In addition, by reducing dimensions and graphical presentation using two- or three-dimensional space, it is easier to interpret complex relationships. It is worth noting that the structures detected by correspondence analysis take into account several variables at the same time.

Correspondence analysis is used in many fields. Literature provides many examples of the use of this method in marketing research [3, 10, 26]. As one of them can replace the analysis of professional activity by age and gender. The use of correspondence analysis allows to show on one illustration several distinctive groups. It can be observed that age categories are closely related to work full-time, and which are particularly exposed to the problem of unemployment. An interesting idea is also the use of correspondence analysis to test students' knowledge about teaching and learning. Discovering patterns makes it possible to design better and more efficient educational programs [1]. There is also an attempt to use this tool in describing political preferences, as well as in the analysis of Parliament members speeches [2]. It allows to present transcripts containing over 100 million words using correspondence maps. Thanks to them we can see how different the language used in the Sejm and Senate is, as well as in subsequent years, the vocabulary change.

Attention should also be paid to the use of correspondence analysis in medical research. One of the more known works using the method with the correspondence maps shows the relationship between the type of headache and the age of patients [7]. The author also presents another interesting dependencies, such as the relationship between personality type and belonging to different diagnostic groups. In [25] correspondence analysis was used to identify the relationship between the incidence of the back pain symptom and a variety of factors, such as age, BMI, smoking, drinking alcohol. Correspondence maps may present which of these variables has a significant relationship with the occurrence of the disease.

The use of correspondence analysis in medicine on example of hospitalization on the gynaecological ward

Saving information in the process of hospitalization we get a typical example of a large data set. Extremely useful in such cases are data mining methods. Their application in medicine is becoming more popular; such as the use of artificial neural networks to predict the outcome of infertility treatment [15, 20], the use of feature selection algorithms to reduce dimensionality of the original data set of women treated for infertility [16–17], or cluster analysis in the process of recruitment for competitive swimming [24]. Data mining methods are useful even in such issues as the analysis of microarray data [18], or the analysis of the information gathered in the process of deep sequencing [6].

In this paper we present the use of correspondence analysis on data from hospital cards from the Department of Gynaecology. In this case, the database contains more than eight thousand cases and several features describing the treatment. For the analysis were selected 14 qualitative variables such as: age group, the NHF (National Health Fund) district, cause of hospitalization, categorized length of hospital stay, etc. The aim of the analysis was performed through graphical representation of the relationships between qualitative variables referring to the hospitalization process. To perform the analysis, statistical package Statistica 10.0, StatSoft was used.

Because of assumptions of the analysis, numerical data were grouped. Ranges describing the age are left-closed, and the length of treatment are presented by class: 1 day, 2 days, ..., 6 days, 1 week, 1–2 weeks, 2–3 weeks, more than three weeks. Disease entities were coded according to the international statistical classification of diseases and health problems ICD-10 [4]. Patients were also classified into five groups of diagnoses, created on the basis of the main causes of hospitalization: gynaecology, obstetrics, infertility, insemination, IVF ET. Place of residence was defined within the NHF district, which covers the patient.

The first pair of analyzed variables were age and the NHF district. Presentation of the relationship of these variables is presented in two-dimensional space, which explained 98.8% of the total inertia [Fig. 3].

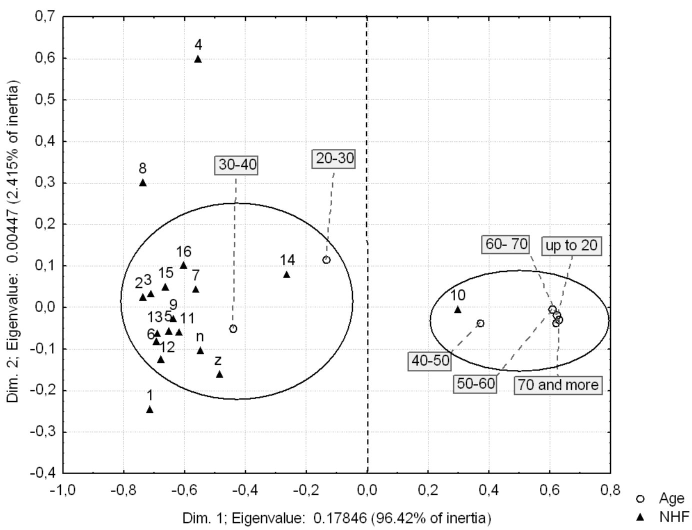

Fig. 3. Graphical representation of the relationship of residence (NHF) and the age of patients

The first dimension clearly distinguished NHF = 10 (Podlaskie Province), which is located on the right of the center axis in relation to the other NHF-s which are on the left side. In the figure we selected two areas of focus points. We note that the older patients (over 40 years) are residents of the 10th NHF, but younger patients (20–40 years old) live in areas outside the Podlaskie Province. Note that the interpretation of the clusters on the maps present the coexistence correspondence characteristics without giving information about the strength of this relationship. Observations obtained in this way need to be confirmed by other statistical methods.

In case of the presentation relationship between length of hospitalization and the place of residence (NHF), two-dimensional correspondence map explains 98.5% of the total inertia [Fig. 4]. The first dimension clearly differentiates between short and long hospitalizations. In this figure, we see the same as in the previous: two areas of focus. This time, a short, one-day hospitalizations coexisted with the place of residence other than the Podlaskie Province and longer hospitalizations were related to patients from the Podlaskie Province (NHF = 10).

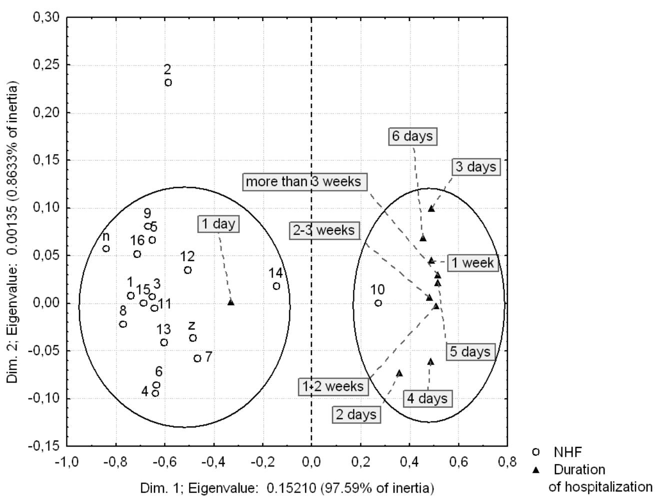

Fig. 4. Graphical representation of the relationship of residence (NHF) and the length of hospitalization

The next graph shows the relationship between characteristics: length of hospitalization and age [Fig. 5]. In this case, the two-dimensional space has been used, which explains 88.3% of the total inertia. We can see three interesting concentrations: short hospitalization (1 day) on young patients (20–40 years old), over one week hospitalization on 60–70 aged women, the longest hospital stays on the oldest patients. These dependencies seem to be natural, because the older the age – the worse the health, and the treatment is often complicated by comorbidities [11].

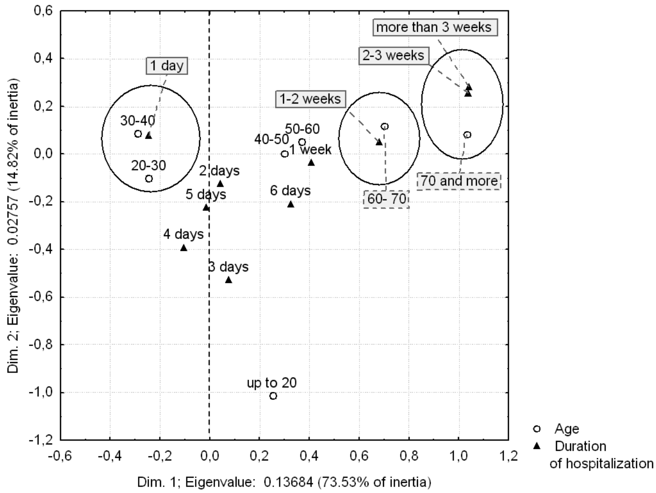

Fig. 5. Graphical representation of the relationship between age of patients and length of hospitalization

The main cause of hospitalization on the gynaecological ward is very diverse. For this reason, patients were assigned to the more homogeneous diagnoses subgroups. Another figure [Fig. 6] clearly shows the relationship between the two groups: the cause of treatment and the age of patients.

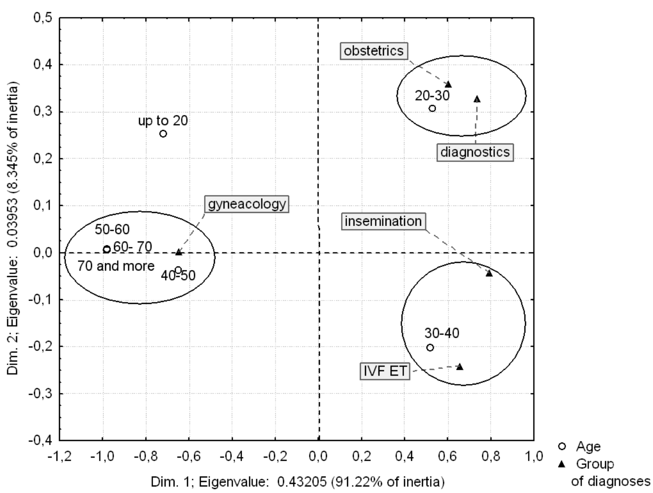

Fig. 6. Graphical representation of the relationship between the age of patients and the cause of hospitalization

Two-dimensional correspondence map explains almost all of the general inertia – 99.56%. The first dimension is stretched by gynaecology on the left side of the axis and the other groups on the right side. The second dimension differentiates patients treated by IVF ET from patients hospitalized with obstetric and infertility diagnostic reasons. The graph shows three areas of focus. The first tells us that the gynaecological patients were over 40. The second presents that the diagnosis and obstetric problems relate to patients aged 20–30 years, and the third focus shows that women aged 30–40 years were treated using IVF ET method or using insemination procedure. This is the age group in which a largest number of women decide to use the assisted reproduction methods in infertility treatment. It is limited from below by the continuously progressive deposition of reproductive decisions, and from above by drastically decreasing the effectiveness of treatment in women over 40 years old [19].

Further analysis of grouped causes of hospitalization showed a clear dependence on the area [Fig. 7]. The two-dimensional space explains 99.17% of the total inertia. The first dimension clearly indicates the focus of the NHF 10 and gynaecological-obstetric causes, which are located on the right of the axis center. On the other hand, we see that the diagnosis and treatment of infertility caused migration of patients from other provinces. These observations were analyzed with traditional statistical methods and presented in [14], as well as developed by basket analysis in [12].

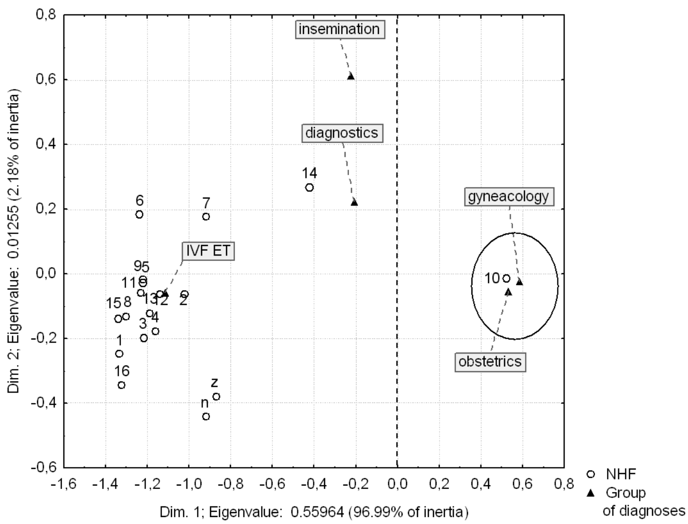

Fig. 7. Graphical representation of the relationship between the residence of patients and the cause of hospitalization

Another analysis of the grouped causes of hospitalization and the length of treatment is shown in a two-dimensional correspondence map, which explains 95.25% of the total inertia [Fig. 8]. The first dimension differentiates one-day hospitalizations from longer hospitalizations. Three areas of focus show the duration of hospitalization depending on the cause of hospitalization. One-day hospitalizations concern the treatment of infertility. Obstetric problems require a 2–5 day stay in the hospital, but the gynaecological problems need the longest hospitalization.

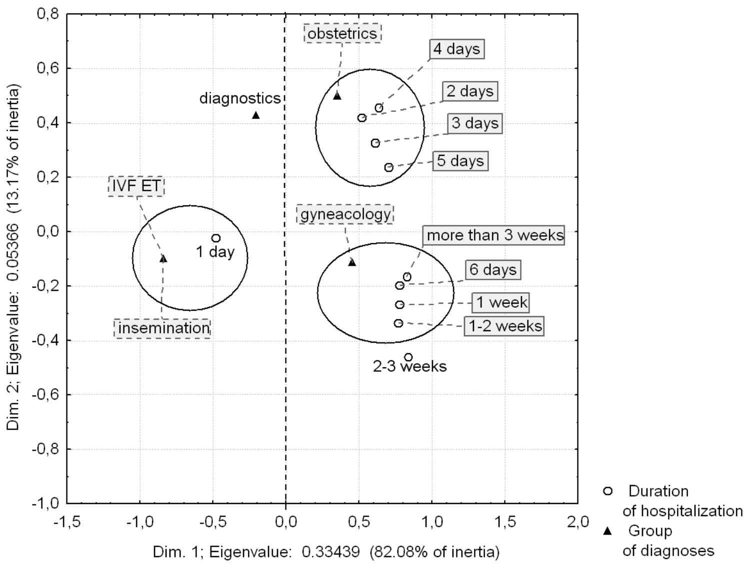

Fig. 8. A graphical representation of relationship between the length of the treatment and cause of hospitalization

Analyzed database concerns patients hospitalized in 1996, 2000 and 2004. The choice of particular years resulted from the subsequent stages of the reform of the health care system. The year 1996 refers to the period in which the health service was still financed directly from the state budget. The year 2000 was the first full year, in which the Health Management Organization was established. The year 2004 was the first full year after creating the institution of the National Health Fund.

Subsequent correspondence maps present relationship between the hospitalization year, the length of treatment and the age of patients. The analysis was performed only for patients whose main cause of hospitalization were gynaecological problems. In both cases the two dimensional space explained 100% of the total inertia, as this is the maximum dimension of the space.

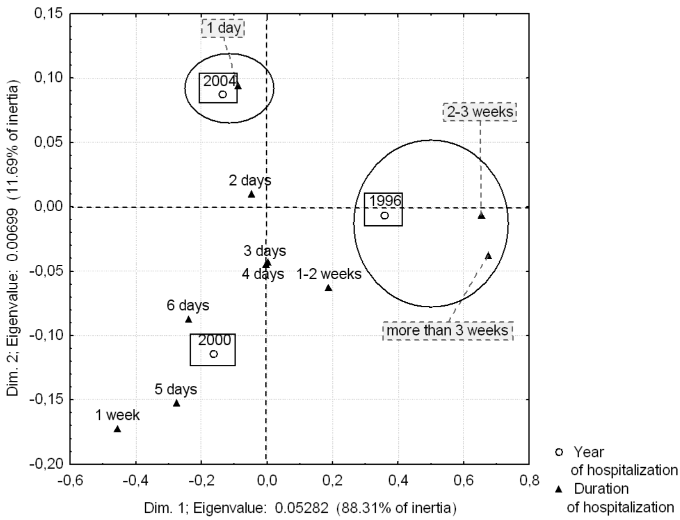

Fig. 9. Graphical representation of the relationship between the length of hospitalization and the year of treatment

[Fig. 9] shows the relationship of particular years, and different lengths of hospitalization. We note that the first dimension differentiates shorter (1–7 days) from longer hospitalizations. The year 1996 lies on the right side of the axis in contrast to the years 2000 and 2004 which are on the left side of axis. We can see that the longest hospitalizations (2–3 weeks, more than 3 weeks) coexists with the year 1996, while the one-day hospitalizations – with the year 2004. This relationship is consistent with the changes which have taken place in the health care system under the influence of the reform. The requirement of the financial account was introduced, which has resulted in a significant shortening of duration of hospitalization times [28]. The development of new treatment methods has also contributed to the reduction of hospitalization time [21, 23].

Another correspondence map shows the relationship between age of patients and the year of hospitalization [Fig. 10]. Over the analyzed years, we can see the effects of aging. It is shown, that the oldest group of patients (more than 70 years old) is closest to the year 2004. The results of the demographic changes observed on the basis of these data were presented in [13].

For patients hospitalized for gynaecological reasons the analysis of coexistence of treatment length and disease entities was performed, which are the cause of hospitalization [Fig. 11]. Two-dimensional space explains 84.2% of the total inertia. The first dimension differentiates the shorter and longer

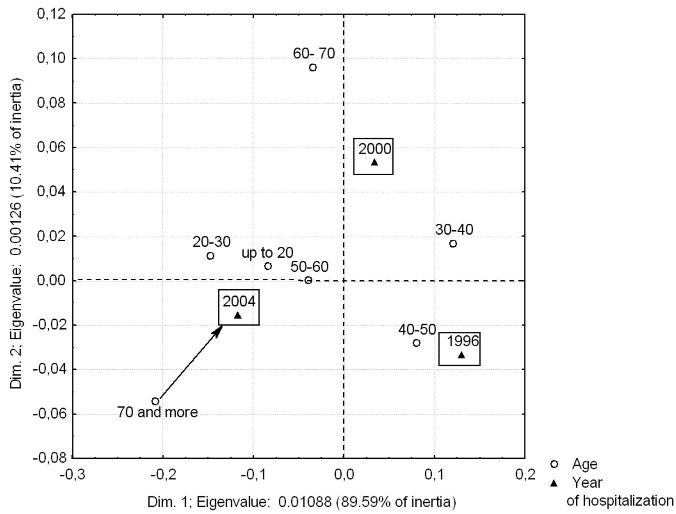

Fig. 10. Graphical representation of the relationship between the year of hospitalization and the age of patients

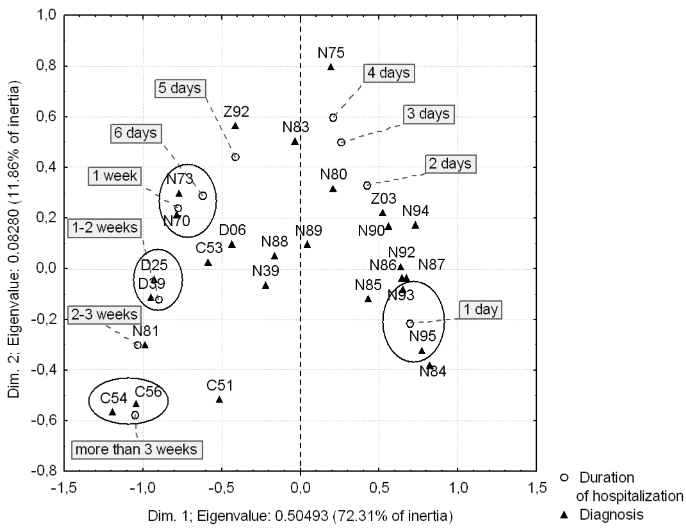

Fig. 11. Graphical representation of the relationship between length of hospitalization and the major diagnosis

hospitalizations. We can distinguish several clusters on the correspondence map:

- one-day hospitalizations relate to patients with diagnoses that require short treatment because of bleeding: N95, N93,

– 6–7 days hospitalizations relate to treatment of inflammation: N70, N73,
– 1–2 weeks hospitalizations relate to patients requiring surgical treatment: D39, D25,
– the longest hospitalizations (more than 2 weeks) relate to patients with cancer: C54, C56.

Conclusions

Correspondence analysis is a method being used increasingly in scientific research. Thanks to access to statistical packages performing calculations does not create problems to researchers, and the interpretation of the results is simple and generally consistent with prior intuition about the expected relationship between the features. Correspondence analysis provides researchers with many possibilities. First, it allows for the discovery of structures and patterns hidden in large databases. Secondly, it can reduce the variables to facilitate interpretation (taking into account the fact that some of the variables are redundant) and creates clear images, called correspondence maps that show related variables in the form of clusters. Observation of the cluster areas often shows the possibility of obtaining interesting results on the basis of observed data in subgroups.

The use of correspondence analysis to data concerning the process of hospitalization showed many interesting relationships. The analysis has outlined problems such as the population aging, shorter hospital stay, or the migration of patients from other provinces. However, the next step is to carry out further analysis using traditional statistical methods.

REFERENCES

[1] Askell-Williams H., Lawson M. J., A Correspondence Analysis of Child-Care Students' and Medical Students' Knowledge about Teaching and Learning, International Education Journal, 5 (2), pp. 176–204, 2004.
[2] Biecek P., Poszukiwanie struktury danych na przykładzie analizy korespondencji, pp. 8–13, Sulejów, 2010 (odczyt „Do czego to się przydaje")
[3] Bendixen M., A Practical Guide to the Use of Correspondence Analysis in Marketing Research, Marketing Bulletin, 14, 2003.
[4] Bartkowski S., Międzynarodowa statystyczna klasyfikacja chorób i problemów zdrowotnych, Rewizja dziesiąta, Tom 1, Vesalius, Kraków, 2006.
[5] Clausen S. E., Applied Correspondence Analysis: An Introduction, SAGE Publications, Thousand Oaks, CA, 1998.

[6] Czerniecki J., Wołczyński S., Deep sequencing – a new method and new requirements of gene expression analysis, Studies in Logic, Grammar and Rhetoric, 25 (38), pp. 41–47, 2011.

[7] Greenacre M., Correspondence analysis in medical research, Statistical Methods in Medical Research, 1, 1992.

[8] Greenacre M., The use of correspondence analysis in the exploration of health survey data, Documentos de Trabajo, 5, 2002.

[9] Greenacre M., Jorg B., Correspondence Analysis in Social Sciences, Academic Press, New York, London, 1994.

[10] Hoffman D. L., Franke G. R., Correspondence analysis: Graphical representation of categorical data in marketing research, Journal of Marketing Research, XXIII, pp. 213–227, 1986.

[11] Leowski J., Polityka zdrowotna a zdrowie publiczne, Wydawnictwo CeDeWu, Warszawa, 2004.

[12] Milewska A. J., Górska U., Jankowska D., et al., The use of the basket analysis in a research of the process of hospitalization in the gynecological ward, Studies in Logic, Grammar and Rhetoric, 25 (38), pp. 83–98, 2011.

[13] Milewska A. J., Milewski R., Mnich S. Z., Karpińska M., Wołczyński S., Wpływ starzenia się społeczeństwa na strukturę chorobowości w ginekologii, Przegląd Menopauzalny, 5, pp. 330–334, 2010.

[14] Milewska A. J., Milewski R., Wołczyński S., Analiza zjawiska migracji pacjentów na Podlasie na przykładzie Kliniki Ginekologii, Polityka Zdrowotna, 7, pp. 71–76, 2008.

[15] Milewski R., Jamiołkowski J., Milewska A. J. et al., Prognosis of the IVF ICSI/ET procedure efficiency with the use of artificial neural networks among patients of the Department of Reproduction and Gynecological Endocrinology, Ginekologia Polska, 80 (12), pp. 900–906, 2009.

[16] Milewski R., Malinowski P., Milewska A. J., et al., Nearest neighbor concept in the study of IVF ICSI/ET treatment effectiveness, Studies in Logic, Grammar and Rhetoric, 25 (38), pp. 49–57, 2011.

[17] Milewski R., Malinowski P., Milewska A. J., et al., The usage of margin-based feature selection algorithm in IVF ICSI/ET data analysis, Studies in Logic, Grammar and Rhetoric, 21 (34), pp. 35–46, 2010.

[18] Milewski R., Milewska A. J., Czerniecki J., Oligonucleotide microarrays in biomedical sciences – the use and data analysis, Studies in Logic, Grammar and Rhetoric, 25 (38), pp. 29–40, 2011.

[19] Milewski R., Milewska A. J., Domitrz J., et al., In vitro fertilization ICSI/ET in women over 40, Przegląd Menopauzalny, 2 (36), pp. 85–90, 2008.

[20] Milewski R., Milewska A. J., Jamiołkowski J., et al., The statistical module for the system of electronic registration of information about patients treated for infertility using the IVF ICSI/ET method, Studies in Logic, Grammar and Rhetoric, 21 (34), pp. 119–127, 2010.

[21] Neis K. J., Brandner P., Wagner S., Laparoskopische Operationsverfahren in der Gynakologie, Gynakologe, 39, 87–104, 2006.

[22] Nenadić O., Greenacre M., Correspondence Analysis in R, with Two- and Three-dimensional Graphics: The ca Package, Journal of Statistical Software, 20 (3), 2007.

[23] Obrzut B., Granice laparoskopii operacyjnej, Ginekologia Praktyczna, 15 (4), pp. 7–11, 2007.

[24] Roczniok R., Zastosowanie analizy skupień w procesie naboru do pływania sportowego, Zeszyty Metodyczno-Naukowe AWF, Katowice, 21, pp. 167–175, 2006.

[25] Sanittham K., Plubin B., Bookkamana P., Correspondence analysis in back pain symptom of employees in the Factory, Lampang Province, Thailand.

[26] Stanimir A., Wykorzystanie analizy korespondencji w badaniach marketingowych, Zastosowanie metod statystycznych w badaniach marketingowych III, StatSoft Polska, pp. 337–346, 2008.

[27] Stanisz A., Przystępny kurs statystyki z zastosowaniem STATISTICA PL na przykładach z medycyny, Tom 3. Analizy wielowymiarowe, StatSoft Polska, Kraków, 2006.

[28] Stelmach W., Olas-Janaszkiewicz E., Maniecka-Bryła I., et al., Średni czas pobytu i wykorzystanie łóżek na oddziałach chirurgii ogólnej w okresie funkcjonowania różnych systemów finansowania ochrony zdrowia, Wiadomości Lekarskie, 60 (7–8), pp. 341–345, 2007.