COMPARISON OF MORTALITY CAUSED BY SERIOUS DISEASES WITHIN REGIONS OF THE CZECH REPUBLIC

Lucie Kopecká

Abstract: The various serious diseases affect the human population all over the world. Good population health is one of the indicators of prospering country and influences economic prosperity. Oncological and cardiovascular diseases belong to the most frequently occurring causes of death in Europe but also in the Czech Republic (CR). The main aim of this article is to provide and present results of multivariate statistical methods which include correlation analysis, factor analysis, cluster analysis and multidimensional comparative analysis. We are able to provide an overview of mortality caused by oncological and cardiovascular diseases by selected indicators of mortality by using multivariate statistical methods in different regions. This information on mortality which is caused by these serious diseases is useful for the health sector to take appropriate measures against these diseases in different regions, it is also useful for insurance companies to improve offer of critical illnesses insurance and useful for revealing socio-economic situation in individual regions within the CR. Data are obtained from the database of National Health Information System of the CR (NZIS CR).

Keywords: Serious Diseases, Mortality, Correlation Analysis, Factor Analysis, Cluster Analysis, Multidimensional Comparative Analysis.

JEL Classification: C38, I15.

Introduction

Serious diseases are the most often causes of death worldwide. Oncological and cardiovascular diseases belong to the most frequently occurring causes of deaths which primarily affect the population in the developed countries such as European countries. Specifically, the Czech Republic (CR) belongs to the countries which are more vulnerable. Mortality from the cardiovascular diseases is higher than mortality from the oncological diseases, which can be caused primarily by stress and poor diets in the CR according to the website of Institute of Health Information and Statistics (ÚZIS ČR, 2016). Although the Europe belongs to the most developed parts of the world, significant socio-economic disparities exist among European countries, which influence health and mortality in Europe as demonstrates (Staníčková, 2015).

Research in this field is important because risks which result from the occurrence of these diseases influence not only the health sector but also insurance companies which offer products of critical illnesses insurance as describes (Jindrová, 2013). As in the case of European countries, significant differences in mortality can exist within regions in the CR and they are the reason for research, see (Jindrová & Kopecká, 2017; Kopecká & Jindrová, 2017). Not only socio-economic disparities can influence mortality in the individual regions but also high mortality can cause bad socio-economic environment in these regions because they create together vicious circle. These following publications deal with socio-economic situation, health care
1. **Statement of a problem**

The main aim of this article is to provide and present the results of multivariate statistical methods and to provide an overview of mortality caused by oncological and cardiovascular diseases by selected indicators of mortality in different regions in the CR.

Oncological and cardiovascular diseases belong to the most frequently occurring causes of death in Europe but also in the CR where the mortality caused by these diseases is still higher than mortality in the more developed western European countries as mentioned (Jindrová & Kopecká, 2017; Kopecká & Jindrová, 2017).

However, cardiovascular diseases are more frequently causes of deaths than oncological diseases. Development of the mortalities caused by oncological and cardiovascular diseases has a decreasing character almost in all cases with a few exceptions, for example mortalities caused by malignant neoplasm of the cervix uteri, malignant melanoma of the skin and ischemic heart diseases. It is observed from 1994 to 2015. For instance, mortalities caused by malignant neoplasms of the larynx, trachea, bronchus and lung, malignant neoplasms of colon and rectum and malignant neoplasms of female breast belong to the most common mortalities. On the other hand, mortalities associated with malignant neoplasms of lymphoid hematopoietic and related issue, malignant neoplasms of stomach and malignant neoplasms of skin belong to less frequent mortalities as demonstrates database (ÚZIS ČR, 2016).

As mentioned above, significant differences in mortality exist not only among European countries but these differences can exist within regions of these counties as well. CR is divided into 14 regions, namely Karlovarský, Plzeňský, Ústecký, Liberecký, Středočeský, Prague, Jihočeský, Královéhradecký, Pardubický, Vysočina, Olomoucký, Jihomoravský, Zlínský and Moravskoslezský as describes (ČSÚ, 2017).

We are able to measure the strength of association between and within variables, reveal the structure of the common factors, classify the regions according to similarity and finally arrange the regions by descending order from the lowest mortality due to these diseases into the highest by using multivariate statistical methods such as correlation analysis, factor analysis, cluster analysis and multidimensional comparative analysis. For example, multivariate statistical methods are described in following publications (Hebák et. al., 2007a; Hebák et. al.; 2007b, Kubanová, 2008; Řezanková et. al., 2009; Stankovičová & Vojtková, 2007). On the other hand, these methods are used within these articles, namely (Jindrová & Kopecká, 2017; Kopecká & Jindrová, 2017; Pacáková & Papoušková, 2016; Pacáková et. al., 2016; Petr et. al., 2010).

2. **Methods**

2.1 **Data**

All data associated with mortality caused by oncological and cardiovascular diseases are obtained from the database of National Health Information System CR (NZIS CR) as demonstrates (ÚZIS ČR, 2016). Each variable related to mortality is drawn for each region for the entire 14-year period (1994-2015) and they are converted to the European standard ASR (E) 2013. ASR (E) is called age-standardized...
rate for Europe 2013. It means that this rate is created by using European Standard Population (ESP) representing European population with a standard age structure. This population age structure was updated in 2013. ASR (E) 2013 is calculated per 100,000 population. For example, this rate is often used in case of mortality which represents the proportion of deaths from a particular population during certain period. For details, see (National archives, 2016).

First, it is necessarily to compile a data matrix where the rows of this matrix are represented by 14 regions (objects) and the columns are presented by 12 quantitative indicators of mortality caused by oncological and cardiovascular diseases (variables). The variables that are intended for using multivariate statistical methods include: malignant neoplasms \( (X_1) \), malignant neoplasms of lymphoid hematopoietic and related tissue \( (X_2) \), malignant melanoma of the skin \( (X_3) \), malignant neoplasm of the cervix uteri \( (X_4) \), malignant neoplasm of the larynx, trachea, bronchus and lung \( (X_5) \), malignant neoplasm of the prostate \( (X_6) \), malignant neoplasm of the female breast \( (X_7) \), malignant neoplasm of the colon and rectum \( (X_8) \), malignant neoplasm of the stomach \( (X_9) \), heart diseases, including cerebrovascular \( (X_{10}) \), ischemic heart diseases \( (X_{11}) \), cerebrovascular diseases \( (X_{12}) \). All mortality variables are expressed in ASR (E) as demonstrates (ÚZIS ČR, 2016).

2.2 Correlation analysis and Kayser-Meyer-Olkin index

Non-correlated variables belong to the important assumption for using cluster analysis. The Pearson and the Spearman correlation coefficients are the correlation coefficients which measure strength of associations between variables, see (Hebák et. al., 2007a, Kopecká & Jindrová, 2017; Kubanová, 2008, Řezanková et. al., 2009). The Kayser-Meyer-Olkin index (KMO) is another way of measuring associations. This index measures associations within the group of variables. The Measure of sampling adequacy rate (MSA) is simplified and analogous rate of the KMO index. When the MSA rate points to considerable associations among variables, it is a good result for application of component analysis before using cluster analysis, see (Hebák et. al., 2007b, Kopecká & Jindrová, 2017; Stankovičová & Vojtková, 2007).

2.3 Factor analysis

First, appropriate number of factors have to be determined by using eigenvalues at the beginning of factor analysis. The main aim of factor analysis is to determine \( m \) variables by using a smaller number of common factors. For details, see (Hebák et. al., 2007b, Jindrová & Kopecká, 2017; Stankovičová & Vojtková, 2007). Relationship between variances of original and derived factor is significant result of factor analysis. Rotation of factors is the important part of factor analysis. Varimax rotation belongs to the most used rotations of factors. Next, correlations are expressed by factor loadings between the selected variables and the factors. Factor loadings point out what percentage of the variance the factor explains within the selected variables. However, factor scores are the final results of factor analysis. They assign the values of the rotated factor scores to the each object. For details, see (Pacákoveá & Papoušková, 2016).

2.4 Cluster analysis

Cluster analysis is used to classifying the objects into groups so that the objects are the most different among the group and the most similar inside the group. Cluster analysis is based on the input data matrix. The rows of this matrix are presented by \( n \)
objects and the columns are presented by $m$ selected variables. Euclidean distance is the most common distance between the objects and Ward’s method belongs to the most important and the most used hierarchical methods as describe (Hebák et. al., 2007b, Kopecká & Jindrová, 2017; Petr et. al., 2010, Řezanková et. al., 2009).

2.5 Multidimensional comparative analysis

Multidimensional comparative analysis is applied because of comparing objects which are evaluated by several variables. It is useful for arranging objects by descending order from the best into the worst. First of all, the type of each variable should be defined. The reason is that the “great” values of some variables influence analysis positively (stimulants) and “small” values of some variables influence analysis negatively (destimulants). Standardization of stimulants is expressed by formula (1) and standardization of destimulants is expressed by formula (2), see (Pacáková & Papoušková, 2016, Pacáková et. al., 2016, Stankovičová & Vojtková, 2007):

$$b_{ij} = \frac{x_{ij}}{x_{\text{max},j}} \cdot 100,$$  

(1)

$$b_{ij} = \frac{x_{\text{min},j}}{x_{ij}} \cdot 100,$$  

(2)

where $x_{ij}$ is value of $i$-th object on $j$-th variable, $x_{\text{max},j}$ is maximal value of $j$-th variable and $x_{\text{min},j}$ is minimal value of $j$-th variable. Finally, the score for each country is average of $b_{ij}$.

3. Problem solving

All results which are presented by employing methods mentioned above are constructed by programs EXCEL and STATISTICA in this chapter. First, associations between and among variables are determined. Then structure of common factors and similarity of the regions are found out and finally, the regions are arranged by descending order from the lowest mortality into the highest.

3.1 Results of correlation analysis and Kayser-Meyer-Olkin index

Overall the MSA rate acquiring value 0.43 points to poor association among the variables which are mentioned above, which is reason for absence of component analysis before using cluster analysis mainly because of inappropriateness of original variables for component analysis and loss of information from the variables. However, the significant associations exist between some pairs of the variables according to the Spearman correlation coefficient. For example, the considerable associations are between variables namely $X_1$ and $X_4, X_5, X_6, X_7, X_8$ (0.85, 0.96, 0.54, 0.70, 0.93) and next between $X_5$ and $X_4, X_8$ (0.92, 0.81). But on the other hand, a very poor association is between $X_1$ and $X_{10}$ (0.36).

3.2 Results of factor analysis

First of all, the eigenvalues are determined of the correlation matrix in Fig. 3.
Fig. 1: Scree plot presenting eigenvalues

Fig. 1 shows that the top four factors are suitable for using because values of eigenvalue numbers achieve values higher than 1 and explain together more than 87% of the original variables variability.

Now, the groups of variables are determined by employing four factors through the factor loadings after Varimax rotation in Tab. 1, which is the main aim of factor analysis.

**Tab. 1: Factor loadings after the Varimax rotation**

<table>
<thead>
<tr>
<th>Variables</th>
<th>Factor 1</th>
<th>Factor 2</th>
<th>Factor 3</th>
<th>Factor 4</th>
</tr>
</thead>
<tbody>
<tr>
<td>$X_1$</td>
<td>0.96</td>
<td>0.04</td>
<td>0.16</td>
<td>0.19</td>
</tr>
<tr>
<td>$X_2$</td>
<td>0.23</td>
<td>-0.53</td>
<td>0.62</td>
<td>0.21</td>
</tr>
<tr>
<td>$X_3$</td>
<td>-0.07</td>
<td>-0.04</td>
<td>0.80</td>
<td>-0.37</td>
</tr>
<tr>
<td>$X_4$</td>
<td>0.92</td>
<td>0.20</td>
<td>-0.15</td>
<td>-0.01</td>
</tr>
<tr>
<td>$X_5$</td>
<td>0.97</td>
<td>0.05</td>
<td>0.01</td>
<td>0.14</td>
</tr>
<tr>
<td>$X_6$</td>
<td>0.42</td>
<td>0.33</td>
<td>0.64</td>
<td>0.25</td>
</tr>
<tr>
<td>$X_7$</td>
<td>0.71</td>
<td>-0.48</td>
<td>0.32</td>
<td>0.07</td>
</tr>
<tr>
<td>$X_8$</td>
<td>0.84</td>
<td>0.12</td>
<td>0.29</td>
<td>0.24</td>
</tr>
<tr>
<td>$X_9$</td>
<td>0.41</td>
<td>0.83</td>
<td>-0.09</td>
<td>-0.14</td>
</tr>
<tr>
<td>$X_{10}$</td>
<td>0.25</td>
<td>0.71</td>
<td>-0.21</td>
<td>0.58</td>
</tr>
<tr>
<td>$X_{11}$</td>
<td>-0.12</td>
<td>0.90</td>
<td>0.18</td>
<td>0.23</td>
</tr>
<tr>
<td>$X_{12}$</td>
<td>0.20</td>
<td>0.06</td>
<td>-0.03</td>
<td>0.92</td>
</tr>
</tbody>
</table>

Each of factors can be interpreted according to factor loadings after Varimax rotation. The factor 1 is presented by variables which provide information on mortality caused by oncological diseases, namely malignant neoplasm of the cervix uteri, malignant neoplasm of the larynx, trachea, bronchus and lung, malignant neoplasm of the female breast and malignant neoplasm of the colon and rectum. The factor 2 is described primarily by malignant neoplasm of the stomach and hearth diseases, concretely ischemic heart diseases. After that, the factor 3 is the next factor which...
presents mortality caused by oncological diseases, namely malignant neoplasms of lymphoid hematopoietic and related tissue, malignant melanoma of the skin and malignant neoplasm of the prostate. Finally, the last factor 4 is described by cerebrovascular diseases.

The significant positive correlations exist between the original variables and the factors after Varimax rotation. It means that the high values of factor scores of individual factors indicate the high mortality caused by these diseases.

The factor 2, factor 3 and factor 4 which are described above are always presented with the factor 1 within Fig. 2, Fig. 3 and Fig. 4. The positions of the regions, associations between factors and outliers of the regions can be revealed according to these four factors through these figures.

**Fig. 2: Location of the regions according to factor 1 and factor 2**

![Fig. 2: Location of the regions according to factor 1 and factor 2](source: own)

**Fig. 3: Location of the regions according to factor 1 and factor 3**

![Fig. 3: Location of the regions according to factor 1 and factor 3](source: own)
The state of mortality can be detected in individual regions based on these three figures above because the regions can be divided into 4 quadrants. The regions are situated according to factor 1 which is represented by the most frequently occurring oncological diseases and factor 2 which is described by malignant neoplasm of the stomach and hearth diseases, concretely ischemic heart diseases in Fig. 2. The worst situation is in Ústecký region. On the other hand, Královéhradecký region belongs to the regions with the lowest mortality due to these factors. The first quadrant is created by Ústecký, Karlovarský, Liberecký and Moravskoslezský regions where the situation in mortality is bad according to factor 1 and factor 2. The second quadrant contains regions such as Zlínský and Olomoucký where the situation is bad according to factor 2 and good according to factor 1. Prague, Královéhradecký, Pardubický, Vysočina, and Jihomoravský regions creates the third quadrant where the situation in mortality is the best according to the both factors. The last quadrant is created by Plzeňský, Jihočeský and Středočeský regions where the situation is worse according to factor 1 and better according to factor 2.

Next, the regions are situated according to factor 1 and factor 3 which provides information on mortality caused by less frequent oncological diseases in Fig. 3. Plzeňský and Liberecký regions belong to the regions with the highest mortality and Pardubický and Zlínský regions belong to the regions with the best situation in mortality in this case. The first quadrant includes Ústecký, Karlovarský, Plzeňský, Jihočeský and Liberecký regions where the situation in mortality is bad according to factor 1 and factor 3. The second quadrant contains regions such as Prague and Vysočina where the situation is worse according to factor 3 and better according to factor 1. Good situation in mortality is within Královéhradecký, Pardubický, Jihomoravský, Olomoucký and Zlínský regions according these factors. The last quadrant is created by Středočeský and Moravskoslezský regions.

Finally, according to factor 1 and factor 4 which represents mortality caused by cerebrovascular diseases Ústecký, Karlovarský, Plzeňský and Středočeský regions belong to the worst regions and Prague, Královéhradecký, Pardubický, Olomoucký, Jihomoravský and Zlínský regions belong to the best in Fig. 4.
However, the outliers include Ústecký and Karlovarský regions in all three figures. These regions belong to the regions with worse situation of mortality caused by oncological and cardiovascular diseases in all cases mentioned above.

In term of associations between these four factors the significant correlation 0.73 is detected only between factor 1 and factor 4, which can be observed in Fig. 4.

3.3 Results of cluster analysis

Hierarchical Ward’s method is applied by using normalized dataset within cluster analysis. Ward’s method classifies the regions according to similarity. The important clusters are revealed in Fig. 5.

Fig. 5: Dendrogram representing similarities within regions

![Dendrogram](source: own)

Fig. 5 shows significant three clusters at a distance which equals 6. First cluster contains regions such as Moravskoslezský, Zlínský, Pardubický, Jihomoravský, Královéhradecký, Olomoucký, Vysočina and Středočeský. These eight regions belong to the regions with lower mortality caused by malignant neoplasm of the cervix uteri, malignant neoplasm of the larynx, trachea, bronchus and lung, malignant neoplasm of the female breast and malignant neoplasm of the colon and rectum according to factor analysis. Next cluster includes only two regions, namely Ústecký and Karlovarský regions which represent regions with higher mortality caused by all diseases mentioned above. Finally, regions such as Liberecký, Plzeňský, Jihočeský and Prague create the last cluster. These regions belong to the regions where the situation in mortality is average within results of factor analysis. However, these four regions are worse in mortality caused by malignant neoplasms of lymphoid hematopoietic and related tissue, malignant melanoma of the skin and malignant neoplasm of the prostate. But situation in mortality is better in Prague than in these three remaining regions within the last cluster.

3.4 Results of multidimensional comparative analysis

Multidimensional comparative analysis is useful for determining the order. It means that the regions are arranged by descending order from the lowest mortality caused by oncological and cardiovascular diseases to the highest. The score 1 \( (X_1-\bar{X}) \) describing
mortality caused by oncological diseases, score 2 (X10-X12) describing mortality from cardiovascular diseases and finally, score 3 (X7-X12) indicating mortality caused by both of them are displayed in Tab. 2.

**Tab. 2: Ranking of the regions according to score 1, score 2 and score 3**

<table>
<thead>
<tr>
<th>Rank</th>
<th>Region</th>
<th>Score 1</th>
<th>Region</th>
<th>Score 2</th>
<th>Region</th>
<th>Score 3</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Pardubický</td>
<td>93.41</td>
<td>Prague</td>
<td>100.00</td>
<td>Pardubický</td>
<td>91.72</td>
</tr>
<tr>
<td>2</td>
<td>Zlínský</td>
<td>92.50</td>
<td>Karlovarský</td>
<td>89.74</td>
<td>Prague</td>
<td>89.75</td>
</tr>
<tr>
<td>3</td>
<td>Vysočina</td>
<td>90.27</td>
<td>Královéhrad.</td>
<td>87.61</td>
<td>Zlínský</td>
<td>89.35</td>
</tr>
<tr>
<td>4</td>
<td>Jihomoravský</td>
<td>89.35</td>
<td>Pardubický</td>
<td>86.62</td>
<td>Královéhrad.</td>
<td>88.75</td>
</tr>
<tr>
<td>5</td>
<td>Královéhrad.</td>
<td>89.13</td>
<td>Jihočeský</td>
<td>86.20</td>
<td>Jihomoravský</td>
<td>88.44</td>
</tr>
<tr>
<td>6</td>
<td>Olomoucký</td>
<td>86.72</td>
<td>Jihomoravský</td>
<td>85.73</td>
<td>Vysočina</td>
<td>87.38</td>
</tr>
<tr>
<td>7</td>
<td>Prague</td>
<td>86.33</td>
<td>Olomoucký</td>
<td>84.38</td>
<td>Olomoucký</td>
<td>86.14</td>
</tr>
<tr>
<td>8</td>
<td>Středočeský</td>
<td>84.99</td>
<td>Plzeňský</td>
<td>83.70</td>
<td>Středočeský</td>
<td>83.56</td>
</tr>
<tr>
<td>9</td>
<td>Jihočeský</td>
<td>82.24</td>
<td>Moravskoslo.</td>
<td>83.56</td>
<td>Jihočeský</td>
<td>83.23</td>
</tr>
<tr>
<td>10</td>
<td>Moravskoslo.</td>
<td>82.19</td>
<td>Liberecký</td>
<td>83.52</td>
<td>Moravskoslo.</td>
<td>82.53</td>
</tr>
<tr>
<td>11</td>
<td>Plzeňský</td>
<td>80.12</td>
<td>Zlínský</td>
<td>79.88</td>
<td>Plzeňský</td>
<td>81.01</td>
</tr>
<tr>
<td>12</td>
<td>Liberecký</td>
<td>79.32</td>
<td>Středočeský</td>
<td>79.26</td>
<td>Liberecký</td>
<td>80.37</td>
</tr>
<tr>
<td>13</td>
<td>Karlovarský</td>
<td>75.06</td>
<td>Vysočina</td>
<td>78.69</td>
<td>Karlovarský</td>
<td>78.73</td>
</tr>
<tr>
<td>14</td>
<td>Ústecký</td>
<td>74.69</td>
<td>Ústecký</td>
<td>69.72</td>
<td>Ústecký</td>
<td>73.45</td>
</tr>
</tbody>
</table>

Source: own

The ranking of the regions is changing in all three cases according to Tab. 2. Pardubický region belongs to the regions with the lowest mortality. On the other hand, Ústecký region shows the highest mortality caused by these diseases. A poor association exists between score 1 and score 2 by using the Spearman correlation coefficient which takes the value 0.12. But on the other hand, a strong association is found between score 1 and score 2, namely 0.92. It is possible to see that the ranking of the regions is similar in case of score 1 and score 3 and different in case of score 2 and score 3.

**Discussion**

The data which carry information on mortality caused by critical illnesses in the regions within the CR were obtained based on database of NZIS CR. This information was obtained by employing multivariate statistical methods and using programs EXCEL and STATISTICA.

First, the Spearman correlation coefficients together with MSA rate revealed several significant associations, mainly between mortalities from oncological diseases and a poor association among the entire group of the selected variables.

Next, the four important factors and their structure were found out by using factor analysis. The first of them factor 1 is described by the most frequently occurring deaths caused by oncological diseases. The factor 2 represents deaths due to malignant neoplasm of the stomach and hearth diseases, concretely ischemic heart diseases which can be caused by stress. Factor 3 is determined by deaths because of less frequently occurring oncological diseases. The last factor 4 expresses deaths which are caused
by cerebrovascular diseases. It can be said that Ústecký and Karlovarský regions belong to the most affected regions but on the other hand, Pardubický region belongs among the least affected regions in the CR based on these factors.

After that, three significant clusters were detected based on cluster analysis. The first cluster contains Moravskoslezský, Zlínský, Pardubický, Jihomoravský, Královéhradecký, Olomoucký, Vysočina and Středočeský regions which belong to the regions with the best situation in mortality according to factor analysis. Ústecký and Karlovarský regions create the second cluster which includes the regions with the worst situation in mortality and the last cluster includes Liberecký, Plzeňský, Jihočeský and Prague regions which are considered as average in case of mortality. Finally, the results of cluster analysis are very similar to the results of multidimensional comparative analysis in case of mortality from oncological diseases but different in case of mortality from cardiovascular diseases.

**Conclusion**

Malignant neoplasms belong to the diseases with the high mortality within CR. On the other hand, cardiovascular diseases are the most dangerous diseases at all.

The main aim of this article was to provide and present the results of multivariate statistical methods and to provide an overview of mortality caused by oncological and cardiovascular diseases by selected indicators of mortality within different regions in the CR.

The results of multivariate statistical methods can be useful for health sector to take appropriate measures against these diseases and insurance companies to improve offer of the critical illnesses insurance within the CR. This overview of mortality brings information on significant differences in mortality situation within individual regions in CR. These differences explain one of the causes of socio-economic inequalities within regions of CR and they can lead to worsening of the situation, especially in the most affected regions.

**Acknowledgement**

This work was supported by SGS University of Pardubice, FEA, No. SGS_2017_022, “Ekonomický a sociální rozvoj v soukromém a veřejném sektoru”.

**References**


**Contact Address**

**Ing. Lucie Kopecká**

University of Pardubice, Faculty of Economics and Administration, Institute of Mathematics and Quantitative Methods
Studentská, 95, 532 10, Pardubice, Czech Republic
E-mail: Lucie.Kopecka1@student.upce.cz
Phone number: +420466036017

Received: 18. 04. 2017, reviewed: 20. 02. 2018
Approved for publication: 27. 06. 2018