LAJOS BÁLINT

Spatial Gender Differences in Life Expectancy at Birth

The study presents the major micro-regional characteristics of life expectancy at birth between 2005 and 2009 according to the classification in force at present. The spatial pattern of life expectancy at birth by gender was described with the help of global and local autocorrelation indicators. Due to the strong relation between the regional life expectancy of men and women, the applied non-spatial cluster analysis served as a more exact methodological frame for typology. The sole purpose of the study was to describe the spatial structure of life expectancy at birth with descriptive tools.

Life expectancy at birth goes far beyond its demographic content. A number of researchers emphasize that life expectancy can be considered a stable imprint of social processes, and is – among macro-level indicators – the most important component of life quality (Sen 1998, Wilkinson 1992, Bobak and Marmot 1996, Mazumdar 2001, Dasgupta 2000). As opposed to other indicators of life quality and different composite indicators, which are often difficult to interpret, life expectancy in itself has an unambiguous content irrespective of spatial and temporal context. To live longer and to remain healthy as long as possible is one of the individual and collective aims that are naturally obvious for everybody.

In Hungary, a large number of publications dealt with the topic of regional mortality (Daróczi 1997). Until the establishment of the structure of micro-regions, researches focused mainly on county-level analyses and the differences between the districts of the capital city (Daróczi 1997, Józan and Forster 1999). Demographic and spatial epidemiological researches at a more nuanced regional scale gained ground owing to the establishment of the structure of micro-regions as well as to the more and more refined methodological tools. In the Hungarian scientific literature numerous micro-regional mortality analyses were prepared (Hablicsek 2004, Klinger 2003, 2006a, 2006b). The focus of the analyses of demographers was first of all on the compilation of regional mortality tables and the descriptive characterization of differences. Within the current framework, the emphasis is definitely put on spatiality. This study outlines the present differences in life expectancy with the help of spatial cross-sectional data.

Data and method

The micro-regional abridged mortality tables were calculated with the Chiang method separately for men and women (Chiang 1984). The detailed description of the method is well documented in numerous publications of the Hungarian scientific literature (Hablicsek 2003, Daróczi 2004). The abridged mortality tables <1, 1–4, 5–9, …, 90+ are built up from the data of age groups. The mortality and mid-year population data come from the DEMOgráfiai program of the Hungarian Central Statistical Office which creates tables. In the interest of authenticity, micro-regional life expectancies at birth contain the

aggregate data of five years. The selected regional scale is the micro-regional classification in force at present[1], which divides the country into 174 disjunct regions in a compact way. The population sizes of these regions are quite different. The division of the population of the capital into districts seems a possible solution, by which the number of observations is naturally increasing, while population size differences between them are decreasing. However, changing the spatial scale leads to well-known consequences. In case of life expectancy at birth, the mean of the "sample" increases, and the results of the autocorrelation tests, especially of the local tests change. Regions with the best life expectancy are practically only the districts of Budapest and some micro-regions of the agglomeration. The main argument in favour of the unchanged traditional micro-regional scale is that the divide between the districts of the capital is nowadays much less striking than the differences between the capital and the peripherical regions in the rural areas.

Table 1

Descriptive statistics

Denomination	Men	Women
Minimum	63.54	74.54
Maximum	73.09	80.51
Range	9.55	5.97
1st quartile	67.39	76.50
Median	68.58	77.10
3rd quartile	70.02	78.17
Interquartile range	2.63	1.67
Mean	68.67	77.27
Dispersion	1.80	1.22
Variance	3.23	1.50
Skegness	–0.04	0.17
Kurtosis	–0.27	–0.56
r	0.83[a)]	
N=	174	174

a) $p < 0{,}001$.

General features

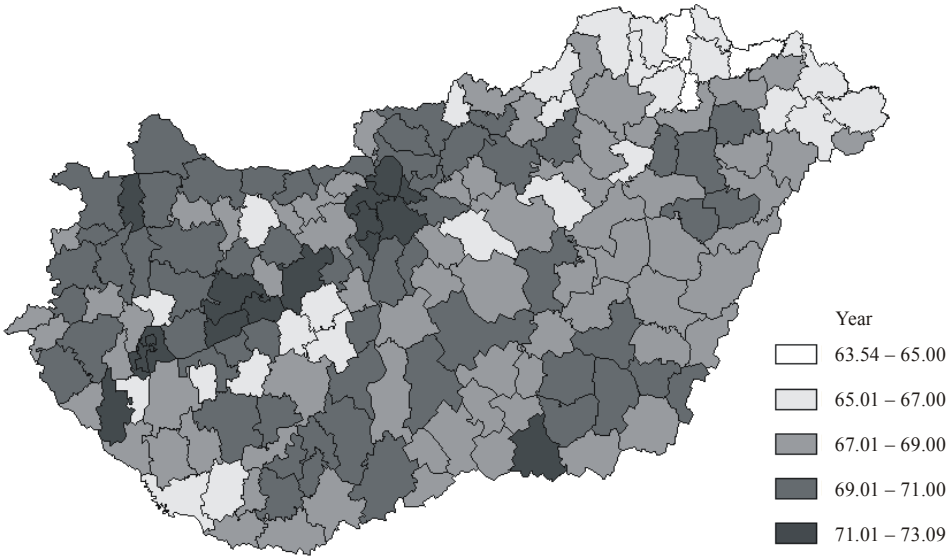
Life expectancy at birth for *men* by micro-region varied between 63.5 and 73.1 years in the observed period. The range of nearly ten years can be considered a significant difference. Micro-regions with the lowest life expectancy have wider, contiguous areas mainly in the north-eastern part of the country (typically in Borsod-Abaúj-Zemplén county) and in the region of Southern Transdanubia. Life expectancy is low in the micro-regions in the southern parts of Fejér county (Aba, Enying and Sárbogárd) as well. Furthermore, there are isolated areas with high mortality in the region Central Transdanubia. The location of areas with the most unfavourable life expectancy is similar to an hourglass rotated towards the south-west and north-east, while in the case of areas with more favourable life expectancy it is more difficult to observe such abstraction.

1 The number of micro-regions changed to 175 after the manuscript had been handed in (*the editor*).

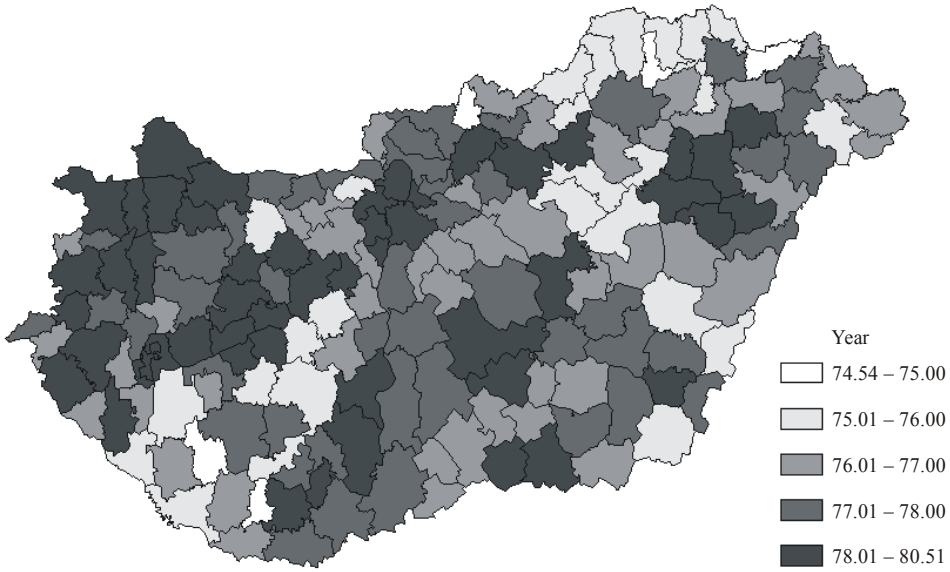
Figure 1–2

Life expectancy at birth by micro-region between 2005 and 2009

1. Men



2. Women



The areas with the highest life expectancy are in the capital and its agglomeration, as well as on the northern shore of Lake Balaton. The majority of micro-regions in Western Transdanubia have lower life expectancies in each region, first of all in the more urbanized areas.

The spatial differences in life expectancy of women by micro-region are much smaller (Table 1), as confirmed by the range (6 years) and the interquartile range as well (1.7 years as opposed to 2.4 years in case of men). However, the difference of six years between the areas with the highest and the lowest life expectancy is also considerable. The areas in the most favourable situation are – similarly as in the case of men – in Budapest and its agglomeration, in the region Western Transdanubia and on the northern shore of Lake Balaton (Figures 1–2), but, differently from men, there are such areas in the different parts of the Great Plain (e.g. in Hajdúság) as well. In the eastern part of the country women have higher life expectancy in areas centred in large cities and in their surroundings (Szeged, Debrecen, Eger, Békéscsaba).

The spatial pattern of life expectancy for men and women does not considerably differ from each other, insofar as visual impression may have some relevance. The definite correlation between the two genders is shown by the high value of the Pearson linear correlation coefficient as well ($r=0.83$, $p<0.001$).

Spatial autocorrelation

There are numerous application opportunities of spatial autocorrelation, of which I use the aspect of the explorative spatial data analysis (ESDA). I try to find an answer to the question of how far the values observed in certain locations are similar to, or different from, those of the neighbours. The present spatial structure of life expectancy is described by determining the degree of clustering of life expectancies and by presenting the location of clusters. The differences between autocorrelation tests applied for irregular polygons can be essentially attributed to the different interpretations of similarity (Waller and Gotway 2004). In the following, I present three often applied autocorrelation tests, and apply them to micro-regional life expectancies.

Moran I

The Moran *I* (Moran 1950, Cliff and Ord 1981) is built up similarly to the Pearson product-moment correlation. Moran himself used the concept of spatial correlation. The difference is the correction of spatiality by the weight matrix (W). The coefficient shows the relation of the variable with itself, i.e. with the same variable of the neighbouring locations and not the strength of the linear relation between two variables. Therefore, the Moran coefficient is an univariant and, due to the inherent feature of spatiality, multi-directional indicator. Its formula is written in the following well-known way:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

where $S_0 = \sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}$. If the weight matrix is row-standardized, i.e. the sum of the rows is one, i.e. each existing neighbour is normed by the number of neighbours, the sum of S_0 is equal to the number of observations *(n)*. The equation is reduced:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n}(y_i - \bar{y})(y_j - \bar{y})}{\sum_{i=1}^{n}(y_i - \bar{y})^2},$$

where y_i is the value of the *y* variable in the *i-th* location, y_j is the value of the y variable in the *j-th* location, $\bar{y}$ is the mean of observations. The expected value of the Moran coefficient is $E(I) = -1/(n-1)$; according to the null hypothesis, observations are independent from each other, i.e. there is no autocorrelation. The value range of the indicator is approximately between –1 and 1. The Moran value is sensitive to the population differences in the relevant observations (Oden 1995) and the outliers, and its size depends on the selected weight matrix as well (Waller and Gotway 2004). Normality and random permutation tests, as well as saddle point approximation are available for testing the hypothesis (Bivand 2009). The value of Moran *I* of course does not change, but the significance level may alter. The normality assumption is often not true for the probability distribution of test statistics, so it is examined usually by one of the above mentioned processes, whether the observations are random or clustered.

Gear's c coefficient

The null hypothesis of Moran *I* is based on the covariance structure of the examined spatial variable $[(y_i - \bar{y})(y_j - \bar{y})]$. The assumption is that the values of locations do not differ consistently from the mean of observations. The null hypothesis of Geary c is based on another interpretation of independence, and states that the neighbouring spatial elements do not differ from each other (Geary 1954). According to the conclusion of the hypothesis, there is no consistent spatial pattern in the differences of neighbours, which are great in some places and smaller in others. Geary c can be computed by the following formula:

$$c = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} W_{ij}(y_i - y)^2}{2W_{ij}\sum_{i=1}^{n}(y_i - \bar{y})^2}.$$

The symbols are the same as those used in the foregoing. The expected value of Geary c in the case of spatial independence is: $E(c)=1$. The connection between Geary c and Moran *I* is negative. As Moran is the best invariant autocorrelation test, Geary is rarely applied. The less than one indicator refers to positive autocorrelation, i.e. to the similarity (small difference) of the neighbouring observations. In the case of extreme positive autocorrelation the value of c is 0. Any value, higher than one, refers to aconsiderable difference between certain locations i.e., to negative autocorrelation. The extreme value of negative autocorrelation is 2. Since the indicator is based on the squared

difference between the values of the neighbouring locations, outliers have a considerable effect on estimating autocorrelation (Fortin–Dale 2005).

Getis–Ord General G statistics

The general or global G indicator (Getis–Ord 1992) is based on the concept of the distance approach of neighbourhood, but it does not preclude the possibility of determining the weight matrix on the basis of topological relation. In the case of global G observations are generally, but not exclusively identified with a single point i.e., their centroids, whose Descartes coordinates are known. For aggregate demographic data applying the population centroid instead of geometrical procedures would be more reasonable which, in the case of micro-regions, would be equal to the coordinates of the centroid of the most populated settlement (the centre of the micro-region). Statistics examine the values of pairs of x_i and x_j points of located in d distance from each other. The formula of the indicator is:

$$G(d) = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}(d)x_i x_j}{\sum_{i=1}^{n}\sum_{j=1}^{n} x_i x_j}, \quad j \neq i$$

The global G identifies spatial relation with the product of the point pair values. In the case of a random pattern the expected value is: *E[G(d)]=W/[n(n–1)]*. If the value of G is higher than the value expected in the case of independence, then the spatial pattern is characterized by the concentration of high values. For low values we can state that the spatial pattern is much rather dominated by the low value pairs. In addition to the normality test, the significance value of global G is determined with the help of the Z value.

It is worth making some comments on the practice of the widely used G statistics published in 1992 (Getis–Ord 1992).[2] Its limit is that it can be applied only for positive values in their own (not transformed) units of measurement. Non-linear transformation considerably influences the value of test statistics. Accepting the null hypothesis, i.e. that observations are independent, does not necessarily mean a random pattern. High and low clusters may be equally present in the spatial structure. Consequently, as opposed to other global autocorrelation indicators, G statistics reveal which type of clustering is present in the observed spatial structure (Aldstadt 2010).

Results of global life expectancy test

In the case of topological weight matrices the relation was always based on first order queen contiguity. In line with symbols used in scientific literature, I applied B (binary) and W-type (row-standardized) weights, and created the binary weight matrix on the basis of the nearest 5, 10, 15 neighbours. In order to define the significance level of

2 In 1995, the authors elaborated a more flexible test, which can be applied not only for variables with positive sign but also for binary weights (Ord–Getis 1995).

autocorrelation, I used a randomization test for the Moran and the Geary values, and assumed normality for Getis statistics. The computations were made with the help of an R 13.0 program (R Development Core Team 2011).

It can be stated in general that the spatial pattern of life expectancy at birth for men shows slightly more considerable autocorrelation based on both the Moran *I* and the Geary c indicator than the same tests for women (Table 2). So, the spatial similarity of their life expectancy in the different micro-regions is also more significant. Except for the first order binary neighbourhood of women, the global G value proved to be significant in case of each weight matrix. Low values hardly differing from zero indicate that in the micro-regional pattern of life expectancy in the Hungarian spatial structure the more definite role of low values is characteristic, although it is not obvious at all based only on the impressions provided by the map.

Table 2

Results of autocorrelation tests with different weight matrices

Type of weight matrix	Moran *I*	Geary c	Global G
		Men	
B	0.371***	0.631***	0.031*
W	0.394***	0.613***	–
KNN5	0.398***	0.631***	0.029**
KNN10	0.327***	0.691***	0.058***
KNN15	0.270***	0.749***	0.087***
		Women	
B	0.237***	0.717***	0.031
W	0.264***	0.716***	–
KNN5	0.233***	0.770***	0.029***
KNN10	0.199***	0.804***	0.058***
KNN15	0.159***	0.838***	0.087***

* $p<0.05$, ** $p<0.01$, *** $p<0.001$.

Correlogram

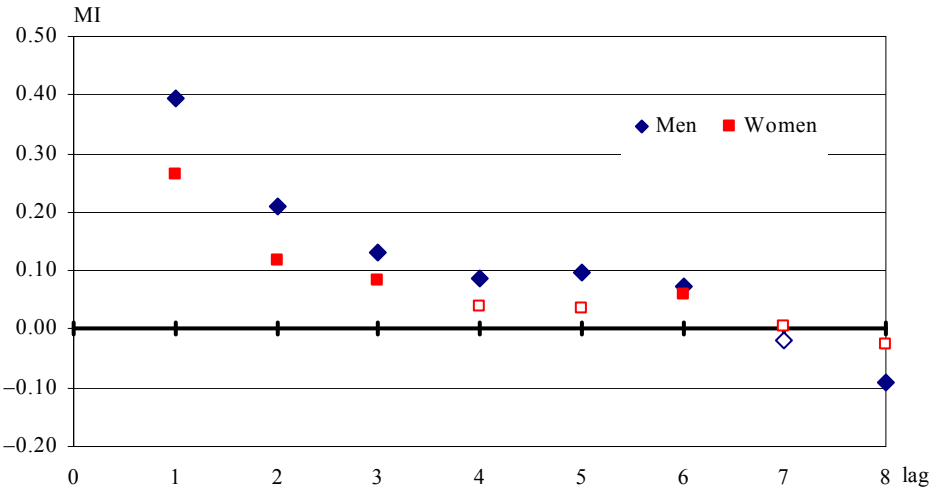
Characterizing spatial structure is a complex task, which involves the size, intensity and direction (isotropic or anisotropic) of spatial processes. One of the tools of characterizing the pattern is the spatial correlogram i.e, the representation of Moran *I* values as the function of distance (D=1, 2, …, d) or the different orders of neighbourhood (K=1, 2,…, k). The correlogram gives a clear visualization of the permanence of spatial dependence. The coefficients in the correlogram are usually marked by specific symbols depending on their significance; the empty sign refers to the absence of significance, while the signs of significant coefficients (α=0.05) are filled (Fortin–Dale 2005). According to common experience, autocorrelation is the largest in the case of first order or nearest neighbourhood, and dependence is gradually diminishing by the increase of distance or the higher order of neighbourhood. For those who are familiar with geostatistics, the

similarity between the correlogram and the semivariogram is easily recognizable. The correlogram can be practically considered the inverse of the semivariogram.[3]

I took into account 1–8 lags of the 174 observations, i.e. the eighth possible neighbour (Figure 3). It is clear to see in the diagram that, on the one hand, the autocorrelation of men is always more significant than that of women, and, on the other hand, its intensity is strongly present up to the sixth order of neighbours, while in the case of women it practically disappears after the fourth order . At the same time, a significant autocorrelation not differing from that of men occurs again at the sixth order of neighbourhood.

Figure 3

Moran I correlogram



Note: The empty formation shows the insignificant spatial effect (the absence of spatial effect).

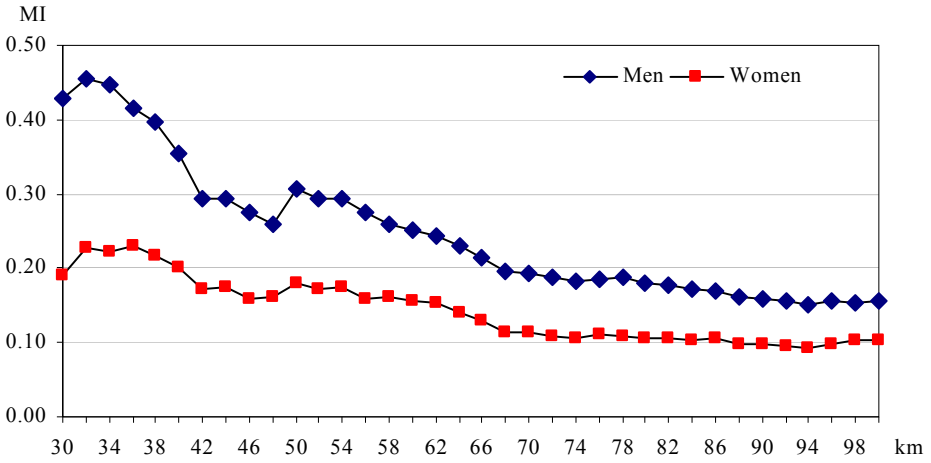
Compared to the correlogram which is based on the different orders of neighbourhood, distance-type approaches are also accepted, yet their substantive content is not at all unambiguous due to the nature of the data, the irregularity of spatial elements, i.e., the differences between distances. I defined the distance based on the Euclidean distance between the centroids of the micro-regions in an interval between 30 and 100 kilometres, for every two kilometres (Figure 4). The minimum 30 kilometre distance ensures that each observation (micro-region) has at least one neighbour. In the distance-based approach, the maximum value of Moran *I* was realized at 32 kilometres (MI = 0.447) for men and between 32 and 36 kilometres (MI = 0.22) for women. The highest value has an outstanding role, since the spatial effect is the largest here, and further additive effects are not realizable any more. After this, the value of the Moran coefficient is practically continuously decreasing but, in the case of neighbours within a

3 On the senivariogram, the distances between the observed points are on the x axis and the variances are on the y axis. In the case of similarity between two near points, their variance is less different, but by the increase of the distance, the differences are increasing as well. The fitting of the proper function shows the operation of the spatial structure.

circle of 100-kilometre radius, representing a size of a region in Hungary, a considerable and strongly significant autocorrelation can be observed.

Figure 4

Distance-based Moran I correlogram



Indicators or local autocorrelation

As opposed to the global autocorrelation tests, local approaches reveal the local features of spatial structures and describe the differences and similarities within the observed spatial structure by giving an exact answer to the questions 'where?' and 'which are those?'. While global indices inform on the extent of autocorrelation in one single indicator, their local variants evaluate each observation one by one. The study by Anselin, (1995) published in the mid-1990s, was a significant leap in the field of autocorrelation of polygon-type spatial data. He marked the family of local indicators with the acronym LISA (Local Indicators of Spatial Association). Each LISA indicator is proportional to its global equivalent. In other words, global indicators can be decomposed into their local components.

Local Moran

Among the indicators of the LISA family the most widely spread is the local Moran connected to Anselin. ,It can be defined for the i-th location as follows:

$$I_i = z_i \sum_j w_{ij} z_j \ ,$$

where z_i and zj are the differences from the mean, $(y_i - \bar{y})$ and $(y_j - \bar{y})$, w_{ij} is the local weight matrix. Due to the summation of values of j observations, only the neighbouring values are involved, i.e. $j \in J_i$. In the interest of simpler interpretation, w_{ij} is usually row-standardized, and – not necessarily, but in the general practice – w_{ii}=0. The expected value of local Moran is $E(I_i)$=–w_i/(n–1). Significant local clusters may be defined also

along with the assumption of normality and randomization. The interpretation of local Moran is the same as the equivalent types of quadrants of the Moran scatter diagram. Four significant outputs are possible. There are high-high and low-low clusters, where the fix locations and their surroundings have similar values. Furthermore, there are also low-high and high-low clusters. Significant clusters may also be defined by assuming normality on the basis of its known momentums. An especially wide-spread solution is the random permutation test recommended by Anselin and the saddle point approximation of Tiefelsdorf (LLoyd 2011).

Getis–Ord local G

The Getis–Ord-statistics (Getis–Ord 1992, 1995) is less widely spread in the Hungarian practice, so its more detailed description, including its momentums, seems to be justified. The local version of the Getis–Ord-statistics can be defined as follows:

$$G_i(d) = \frac{\sum_{j=1}^{n} w_{ij}(d)x_j}{\sum_{j=1}^{n} x_j}, \qquad i \neq j,$$

where w_{ij} is the symmetrical spatial weight matrix. The value of the matrix elements is 1 if they are located within distance *(d)* defined by us and it is 0 in all other cases. According to this type of indicator, the given location is not neighbouring with itself, and the value of the weight matrix is 0. This distinguishes it from the indicator $G_i^*(d)$, where $i = j$. The $G_i(d)$ statistics actually identifies the strength of spatial association with the concentration of weighted spatial points. In the case of clustering of the above average values within a given distance the value of G_i will be high, while in the case of concentration of low values it will be low.
The expected value of the indicator can be defined as follows:

$$E(G_k) = \frac{W_i}{(n-1)}, \qquad \text{where } W_i = \sum_{j=1}^{1} w_{ij}(d), \quad i \neq j.$$

For the variance of the indicator we have to define:

$$Y_{i1} = \frac{\sum_{j=1}^{n} x_j}{n-1} \qquad \text{and} \qquad Y_{i2} = \frac{\sum_{j=1}^{n} x_j^2}{n-1} - Y_{i1}^2, \qquad i \neq j.$$

The variance is:

$$Var(G_i) = \frac{W_i(n-1-W_i)}{(n-1)^2(n-2)}\left(\frac{Y_{i2}}{Y_{i1}^2}\right).$$

The difference between the observed and the expected value of the G_i-statistics gives an answer as to whether the clustering of the high or the low value of the variable is characteristic in the surroundings of the location. The standardized Z value, $G_i(Z)$ belonging to G_i can be written as follows:

$$Z_i = \frac{G_i(d) - E[G_i(d)]}{\sqrt{Var(G_i(d))}}$$

It may be easy to fefining significance under the above conditions,, since the critical value of $G_i(Z)$ is 1.96 on a confidence level of 95%. When defining local G_i value, the key issue is to define the optimum *d* distance. According to the proposal of Getis (1995), defining the distance with the maximum extent of autocorrelation is the most obvious.

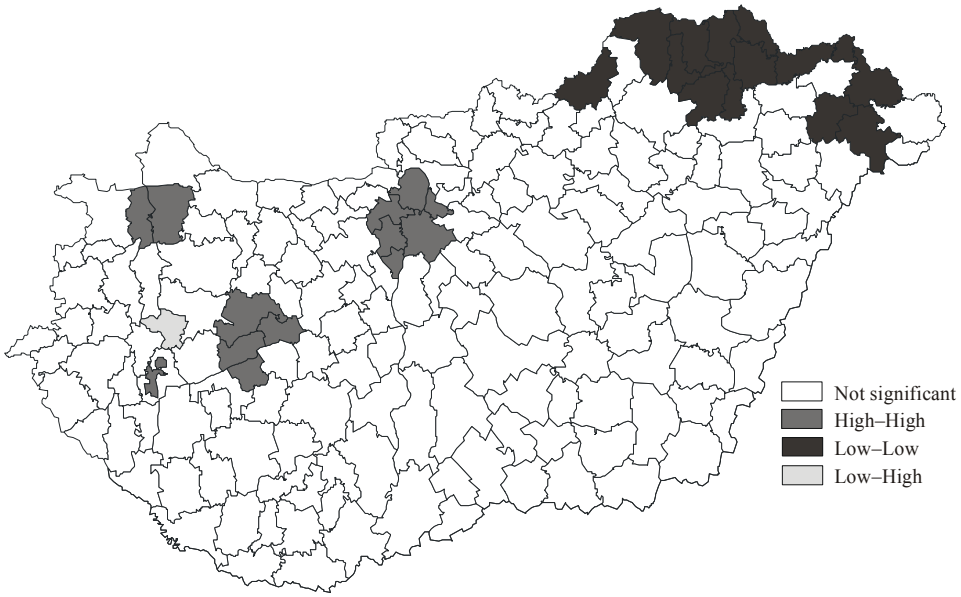
Local autocorrelation of micro-regional life expectancies

I examined the local features with the local Moran- (Anselin 1995) and Getis–Ord- (1992) statistics.Frst-order binary weight matrix based on queen contiguity was applied in both tests. Instead of mapping the local values of autocorrelation, I only wish to point out the significant clusters. In both tests the significance was defined in the same way, along with the assumption of normality, the limitations of which were indicated previously.

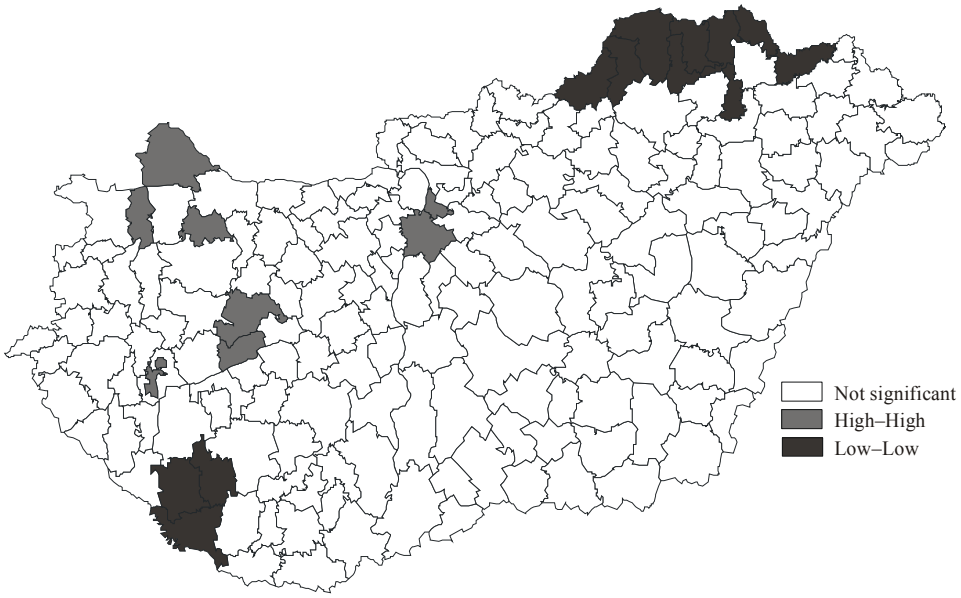
Figure 5–8

Local autocorrelation of micro-regional life expectancies
Local Moran I_i

5. Men

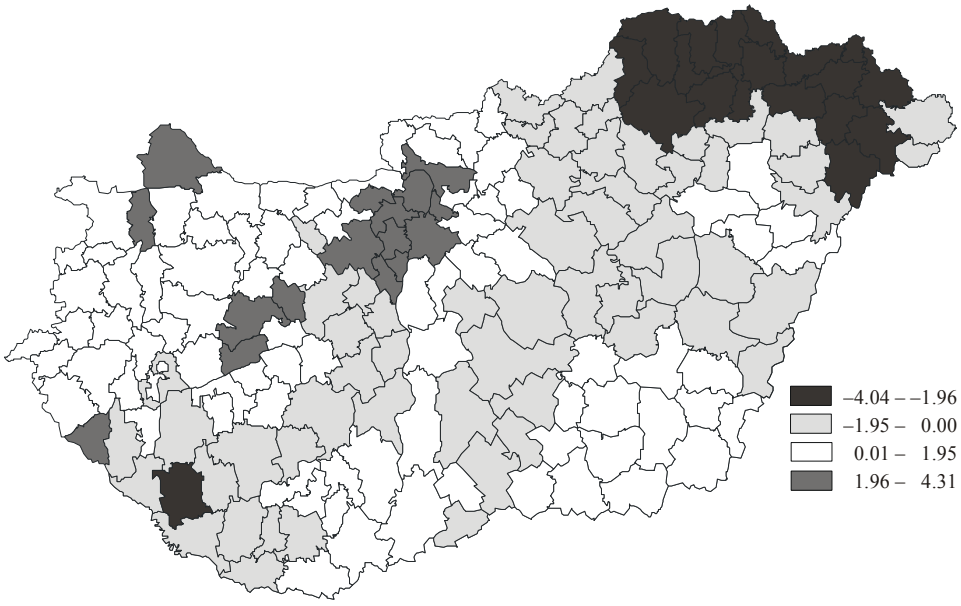


6. Women

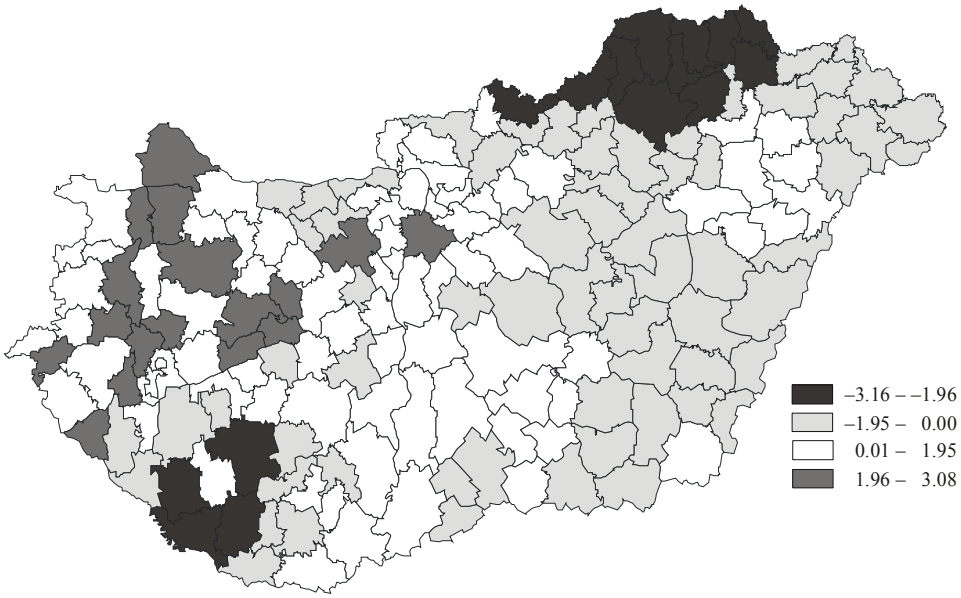


Local $G_i(Z)$

7. Men



8. Women



On the whole, the two tests led to very similar results both for men and women. The cluster of men is in the north-eastern part of the country. The cluster according to local G turned out to be somewhat more extensive than that according to the local version of Moran. It is interesting that, on the basis of the local Moran, there was no significantly low cluster in Southern Transdanubia, while in case of the Getis test, a cluster centre emerged (Nagyatád micro-region). On the basis of clustering of high values, four-four clusters can be distinguished, differing practically only in size. The areas with high life expectancy for men are mostly in the neighbourhood of the capital city, in the surroundings of Balatonfüred and Veszprém (the northern areas of Lake Balaton) and in Győr-Moson-Sopron county. According to local Moran, the Hévíz micro-region surrounded by the Keszthely micro-region, while according to Getis–Ord, the Letenye micro-region were identified as a high cluster.

The areas with low life expectancy are in the northern edge of the country in case of women as well, however, this cluster proved to be much narrower and somewhat shorter than that for men. Both local tests showed a cluster of more considerable extent in the micro-regions of Somogy in Southern Transdanubia (Nagyatád, Barcs, Kadarkút). It should be noted that according to $G_{i,}$ the Kaposvár micro-region was also a part of the low cluster. Except for Budapest, areas with longer lifetime cluster compactly as well. The cluster around Budapest is not at all as extensive as that for men, and only the centre of the cluster, i.e. Budapest proved to be significant. It seems that the effect of higher living standards is different for the two genders in the central region of the country,. While in the case of men a diffuse effect in the areas around Budapest is clearly visible, in the case of women this seems to be limited. The high clusters are in the micro-regions of Western Transdanubia. Compared to the former difference, the difference between the

two testsis considerable. The number of clusters shown by local Moran is considerably lowerthan the number of those shown by local G.

Gender differences

The difference between life expectancies of men and women became obvious after the 1st world war, and since that time life expectancy of women has been higher than that of men in each more developed country (Nathanson 1984). The more favourable mortality of women can be observed in each age group; however, differences diminish in older age (Read and Gorman 2011). In older age, not only is the difference in mortality probabilities moderate, , the structure of causes does not show any significant difference, either (Cockerham 2004). The extent of gender differences considerably changed over time (Annandale 2010). In addition to the rlevant time lag, the fluctuation of differences can be partly explained by lifestyle factors and the consequent changes in the structure of causes of death (Cutler et al. 2006). Nowadays, the difference is much smaller in countries with low mortality than the peak values in the late 1960s and early 1970s. In the most developed countries differences between 4 and 7 years can be observed in general. So, for example, in Austria, where mortality is on Western European level, life expectancy at birth was 77.7 years for men and 83.3 years for women in 2009 (a difference of 5.6 years). Among the Nordic countries a smaller, 4–5 years difference is characteristic of Sweden (79.3, 83.4) and Norway (78.5, 83.4), while the difference is more considerable, 6.8 years in Finland (76.7, 83.5). Among the Western European countriesgender differences are strongly significant in France (77.8, 84.8) (HFA–DB).

In the Central and Eastern European region differences are incomparably larger: in the former Soviet region differences much above ten years can be considered usual (Cockerham 1999). In Russia the difference between men and women is exactly 12 years (62.7, 74.7) (HMD), and the situation is the same in the Ukraine as well (62.3, 74.1) (HFA–DB).

In Hungary the difference between life expectancies of men and women was 8 years (70.0, 77.9) in 2010, which can be considered a medium between Eastern European countries with extremely high and Western European countries with low mortality. Gender differences decreased in the past one and a half decades, but the lag is considerable compared even to the higher differences in the Western European.

To a certain extent, gender differences can be explained by biological (genetic, hormonal) features. The temporal fluctuation of macro-level differences is a reasonable argument for the limits to the biological explanation. In addition to biological reasons, various social, behaviour and psychological factors have a determinant role. Behaviour differences refer to the different lifestyle of men and women. Apart from some exceptions, health damaging habits (smoking, alcohol consumption) occur more often among men, their eating habits are less healthier than those of women, high-risk activity forms (injuries, accidents) occur more regularly among them and more men fall victims of fatal violent crimes too. The majority of victims of completed suicides are also men. It is also known that the participation of men in screenings and healthcare is lower than that of women.

In the spatial examination of gender differences two more important empirical trends can be defined. One of them clearly focuses on the analysis of the spatial structure, while the other is looking for an answer to the reasons for the differences. Preston (1976) intended to explain the inequality on the basis of the different (economic, social, psychological) dimensions of modernization. He stressed the different effects of the various factors on the structure of causes of death. The Swedish study examining the connection between gender differences and health status concluded a negative correlation between the health status and the different segments of equality, i.e. the equality in gender roles generated an increasing inequality in health, which was just the opposite of the most important finding of inequality literature (Backhans 2007).

The spatial connection between healthy life expectancies of men and women was examined by Groenewegen and his colleagues (Groenewegen et al. 2003). He found a relatively strong connection between the regional values for men and for women. The regional analysis of the British health authorities examining the data of the first half of the 1990s found more significant differences in the case of men than in the case of women. Gender differences were more significant in deprived areas and more balanced in areas of higher social status (Relaigh and Kiri 1997). When analysing 373 regional units in Poland, Malczewski (2009) obtained entirely different results. Life expectancies of men and women correlated only to a medium degree, and the spatial patterns of the two genders considerably differed from each other. A Scottish study examined the regional data of alcohol-related mortality by gender (Emslie and Mitchell 2009). The authors found a strong connection between the mortalities of the two genders. According to their finding, the same social factors are in the background of alcohol-related mortality in case of both genders.

Experiences in Hungary

According to Hungarian micro-regional data, the median and mean value of the differences between men and women is around 8.5 years, which is somewhat higher than macro-level differences. The lowest difference, 6.35 years was observed in the Szentendre micro-region, where life expectancy at birth was 73.04 years (95%, KI: 73.44–73.64) for men and 79.39 years (95%, KI: 78.82–79.96) for women. In Budapest the difference was almost the same as in Szentendre (6.46 years). At the same time, especially considering that these are areas with the most equal values , it is worth mentioning that difference of more than six years is is greater than the difference of four years which is not rare in Western Europe. The largest regional gender difference was observed in the Abaúj–Hegyköz micro-region (12.16 years), where life expectancy of men was 63.54 years (KI: 62.11–4.97), while that of women was 75.7 (KI: 74.02–77.37) years.

Table 3

Descriptive statistics of differences in life expectancies of men and women

Denomination	Difference
Minimum	6.35
Maximum	12.16
Range	5.81
1st quartile	7.87
Median	8.52
3rd quartile	9.28
Interquartile range	1.41
Mean	8.59
N=	174

The map illustrating the differences (Figure 9) and the high level of the correlation coefficient mentioned earlier imply that genders are good predictors to each other. The differences are more moderate in those regions where life expectancies are higher, while absolute disadvantages mean relative disadvantages between genders as well. I applied the cluster analysis for the typisation of the regional features of gender differences.. Clusters were formed on the basis of only two vectors, the difference in life expectancy between genders and the life expectancy of women. Before clustering, both variables were standardized (z-score). Among the possible solutions of partitioning, I used agglomerative hierarchical and non-hierarchical methods (PAM, k-means). In the hierarchical method the identity of the single groups was examined by single and complete linkage clustering and also with the centroid, group mean and sum of squares methods (Ward). (For a detailed description of the methods see Kauffmann and Rousseeuw 2005, Füstös et al. 2004.) The most adequate method was the k-means cluster (R 'cclust' package). The distance matrix between observations was based on the most general Euclidean metric:

$$d(i,j) = \sqrt{(x_{i1} - x_{jl})^2 + (x_{i2} - x_{j2})^2} \ ,$$

Defining the optimum number of clusters is the most difficult task in clustering procedures. Numerous tools, which are not necessarily in line with each other, are available to carry out the most appropriate clustering (for evaluating the different cluster algorithms see first of all Gan et al. 2007, Borcard et al. 2011). In the present case, I applied the silhouette coefficient (SC) of Rousseeuw (Rousseeuw 2005). The indicator takes into account the dissimilarity of observations belonging to the cluster *a*, and examines the same in relation with cluster *b* being the closest to cluster *a.*

$$s(i) = \frac{b(i) - a(i)}{\max\{a(i), b(i)\}} .$$

The highest s(i) mean coefficient value define the number of clusters. The value of the indicator is $-1 \le s(i) \le 1$. The definite advantage of the theory is that the silhouette value of each cluster member can be visualized, so clusters whose members do not fit properly to their groups can be seen well. The mean of the coefficients by cluster

indicates the adequacy of clustering; in the case of unequal number of groups SC is not recommended.

Due to the coefficient two clusters had to be defined. The value of the average silhouette coefficient was 0.53, which showed a stable cluster scheme. In case of partitioning with a higher number of clusters, the value of SC was always below 0.4 (the value of SC was 0.36 in case of three and 0.23 in case of four clusters), indicating the weakness of the structure, i.e. the artificial nature of partitioning. For validating clustersthe variance analysis (ANOVA) was used separately for both variables. The normal distribution of residuals was verified by the Shapiro–Wilk-test, while the homogeneity of variances by the Bartlett-test. Both tests met the requirements specified for the variance analysis. The analysis confirmed that life expectancies at birth significantly differ in the function of the defined factor (cluster members).

Figure 9

Gender differences in life expectancies at birth

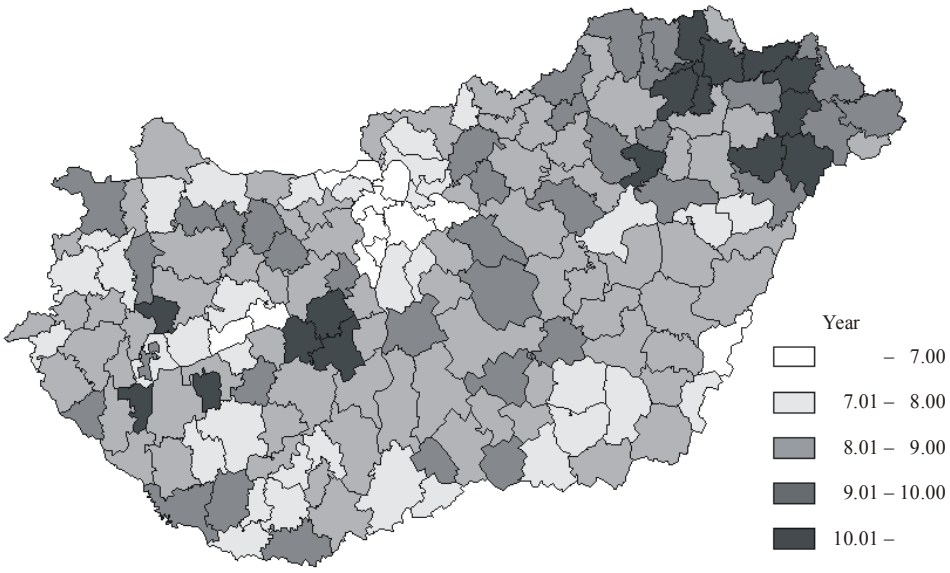
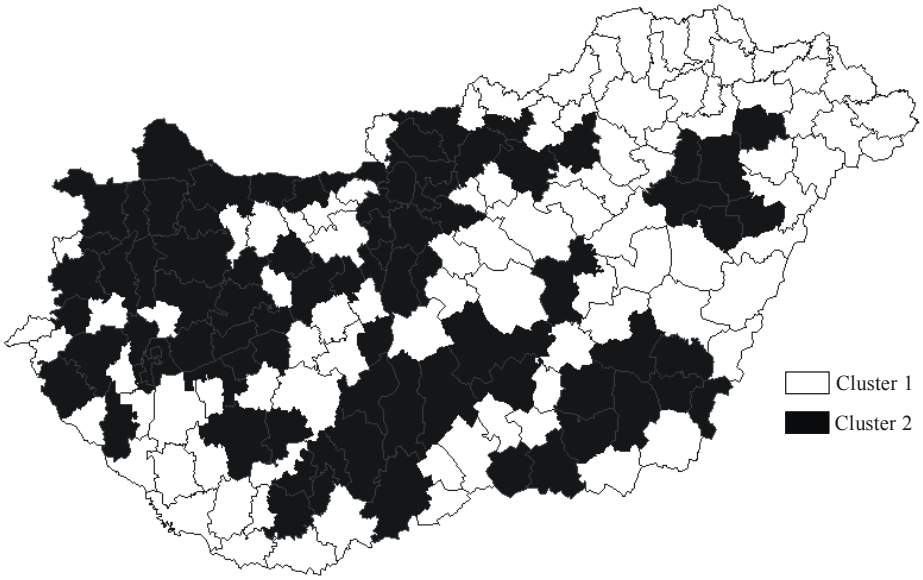


Figure 10

Result of k-mean clustering



According to the results the country can be divided into two types on the basis of regional gender differences, which are of course differences of degree. According to present data, the expected gender features are consistent in space. Where women can expect longer lifetime on the basis of the present level of mortality, men have similar life chances as well. So, my most important finding is that according to present data, inconsistent clusters cannot be demonstrated in the spatial structure in Hungary with the applied method. Life expectancies by gender strongly correlate with each other.

Summing up briefly the results of the cluster analysis, it can be stated that the first group is made up of micro-regions (n=93) with lower life expectancies for both genders. The mean lifetime of men was 67.3, while that of women was 76.4 years, and the mean difference amounted to 9.1 years. In the second cluster (n=81), the mean life expectancies were 70.3 and 78.3 years, respectively, and the difference was somewhat smaller. The spatial distribution does not correspond either to the distribution by east–west or to the one by centre–periphery. In the western part of the country those areas are in majority where differences are somewhat smaller on a national scale, and life chances are better for both genders. Such areas can be found in the regions Southern Great Plain, Northern Great Plain and Southern Transdanubia; moreover, they covers larger territories. The second-type cluster is dominant mainly in Southern Great Plain. The concentrated occurrence of observations belonging to the first cluster can be seen also in Southern Transdanubia, Northern Great Plain and mainly in Northern Hungary, however, they occur more sporadically in any other Hungarian region as well.

Conclusions

In the study I examined the spatial pattern of life expectancies at birth. The life expectancies were estimated with own computations on the basis of the method of Chiang (1984). Of the data analysis techniques the global and the local clustering provided a proper framework for describing spatial features and revealing local patterns. Local clusters were defined with the binary weight matrix and by assuming normality in case of both genders. The location of clusters proved to be similar by gender as well, while smaller differences occurred in the extent of the cluster.

Due to problems deriving from spatial delimitation and the different size of the observed sub-regions, regional inequalities in Hungary cannot be unambiguously compared to the similar data of other countries; the differences reveal significant regional disparity. Compared to the results of Western European researches, the differences in Hungary proved to be definitely more considerable. In the entire analysis presentation of the features by genders played a determinant role. It was not surprising that regional differences were more robust in the case of men. Gender differences were more than six years even in the most balanced micro-regions, which only just approach the mean in Western Europe, while areas with larger differences show similarities with the Eastern European, former Soviet region. We could observe a very strong correlation between life expectancies of men and women. The cluster analysis revealed only differences of degree, i.e. we cannot report any cluster or region that would refer to behaviours inconsistent by gender. In Hungary life expectancies by gender are clearly effective macro-level predictors to each other.

REFERENCES

Aldstadt, J. (2010): Spatial Clustering. In: Fischer, M. M. – Getis, A. (eds.): *Handbook of Applied Spatial Analysis*. Software Tools, Methods and Applications, Springer-Verlag, Berlin–Heidelberg.

Annandale, E. (2010): Health Status and Gender. In: Cockerham W. C. (ed.): *The New Blackwell Companion to Medical Sociology*. Wiley–Blackwell.

Anselin, L. (1995): Local Indicators of Spatial Association – Lisa. *Geographical Analysis* 27(2): 93–115.

Backhans, Mona C. – Lundberg, M.l – Mansdotter, A. (2007): Does increased gender equality lead to convergence of health outcomes for men and women? A study of Swedish municipalities. *Social Science and Medicine* 64 (9): 1892–903.

Bivand, R. (2009): „The Problem of Spatial Autocorrelation" forty years on. http://cran.r-project.org/web/packages/ spdep/vignettes/CO69.pdf (downloaded: May 2012).

Borcard, D. – Gillet, F. – Legendre, P. (2011): *Numerical Ecology with R. USE R!* Springer New York, Dordrecht, London, Heidelberg.

Chiang, C. L. (1984): *The Life Table and its Applications*. Robert E. Kriger Publishing Company Malaber, Florida.

Cockerham, W. C. (1999): *Health and Social Change in Russia and Eastern Europe*. Routledge, New York and London.

Cockerham, W. C. (2004): *Medical Sociology*. Ninth Edition, Pearson Prentice Hall, Upper Saddle River, New Jersey.

Cutler, D. – Deaton, A. – Lleras-Muney, A. (2006): The determinants of Mortality. *Journal of Economic Perspectives* 20(3): 97–120.

Daróczi Etelka – Hablicsek László (2008): A halandóság területi és időbeli különbségei az életkor függvényében. *Demográfia* 51 (1): 7–50.

Daróczi Etelka (1997): *A halandóság területi eltérései Magyarországon 1959/60–1992*. KSH Népességtudományi Intézet Kutatási Jelentések, 60., Budapest.

Daróczi Etelka (2004): Területi és társadalmi különbségek a középkorú férfiak és nők halandóságában Magyarországon 2001 körül. In: Daróczi Etelka – Kovács Katalin (szerk.): *Halálozási viszonyok az ezredfordulón: társadalmi és földrajzi választóvonalak*. KSH, Népességtudományi Kutatóintézet. Kutatási Jelentések, 77., Budapest.

Dasgupta, P. (2000): Valuation and Evaluation: Measuring the Quality of Life and Evaluating Policy. Reseurces for the Future, Discussion Paper 00–24. http://www.rff.org.

Emslie, C. – Mitchell, R. (2009): Are there gender differences in the geography of alcohol-related mortality in Scotland? An ecological study. BMC Public Health, 9 (58).

Fortin, M. J. – Dale, M. R. T. (2005): *Spatial Analysis – A Guide for Ecologists*. Cambridge University Press, Cambridge.

Füstös László – Kovács Erzsébet – Meszéna György – Simonné Mosolygó Nóra (2004): *Alakfelismerés (Sokváltozós statisztikai módszerek)*. Új Mandátum Könyvkiadó, Budapest.

Gan, G. – Ma, C. – Wu, J. (2007): *Data Clustering: Theory, Algorithms, and Applications*. ASA-SIAM Series on Statistics and Applied Probability, SIAM, Philadelphia, ASA, Alexandria, VA.

Geary, R. C. (1954): The Contiguity Ratio and Statistical Mapping. *The Incorporated Statistician* 5. (3): 115–127+129–146.

Getis, A. – Ord, J. K. (1992): The analysis of spatial association by use of distance statistics. *Geographical Analysis* 24 (3): 189–206.

Getis, A. (1995): Spatial Filtering in a Regression Framework: Experiments on Regional Inequality, Government Expenditures, and Urban Crime. In: Anselin, L. – Florax, R. (eds.): *New Directions in Spatial Econometrics* Springer, Berlin.

Getis, A. (2010): Spatial Autocorrelation. In: Fischer, M. M. – Getis, A. (eds.): *Handbook of Applied Spatial Analysis*. *Software Tools. Methods and Applications* Springer–Verlag, Berlin–Heidelberg.

Groenewegen, P. P. – Westert, G. P. – Boshuizen, H. C. (2003): Regional differences in healthy life expectancy in the Netherlands. *Public Health* 117 (6): 424–429.

Hablicsek László (2003): Térségi halandósági különbségek a középkorúak körében. In: Daróczi Etelka (szerk.) (2003): *Kettős szorításban. A középgenerációk élete és egészsége*. KSH Hivatal Népességtudományi Kutatóintézet, Kutatási Jelentések, 74., Budapest .

HFA-DB: European Health for All Database, http://data.euro.who.int/hfadb/.

HMD: Human Mortality Database, http://www.mortality.org/.

Józan, P. – Forster, D. P. (1999): Social inequalities and health: an ecological study of mortality in Budapest, 1980–83 and 1990–93. *British Medical Journal* 318 (7188): 914–915.

Kauffmann, L. – Rousseeuw, P. J. (2005): *Finding Groups in Data. An Introduction to Cluster Analysis*. John Wiley & Sons, Hoboken, New Jersey.

Klinger András (2003): A kistérségek halandósági különbségei. *Demográfia* 46 (1): 9.

Klinger András (2006a): Újabb adatok a vidéki kistérségek és a budapesti kerületek halandósági különbségeiről (I.). *Demográfia* 49 (2–3): 197–231.

Klinger András (2006b): Újabb adatok a vidéki kistérségek és a budapesti kerületek halandósági különbségeiről (II.). *Demográfia* 49 (4): 342–365.

LLoyd, C. D. (2011): *Local Models for Spatial Analysis*. 2nd. Ed. CRC Press, Taylor & Francis Group, Boca Raton, London, Routledge.

Malczewski, J. (2009): Exploring spatial autocorrelation of life expectancy in Poland with global and local statistics. *Geojournal* 75(1): 79–92.

Mazumdar, K. (2001): Improvements in life expectancy: 1960–1995 an exploratory analysis. *Social Indicators Research* 55 (3): 303–328.

Moran, P. A. P. (1950): Notes on continuous stochastic phenomena. *Biometrika* 37 (1-2): 17–23.

Nathanson, C. A. (1984): Sex differentials in Mortality. *Annual Review of Sociology* 10: 191–213.

Oden, N. (1995): Adjusting Moran's I for population density. *Statistics in Medicine* 14(1):17–26.

Ord, J. K. – Getis, Arthur (1995): Local Spatial Autocorrelation Statistics: Distributional Issues and an Application. *Geographical Analysis* 27 (4): 286–306.

Preston, Samual H. (1976): *Mortality Patterns in National Populations with Special Reference to Recorded Causes of Death*. Academic, New York.

Read, Jen'nan Ghazal – Borman, Bridget K. (2011): Gender and Health Revisited. In: Pescolido, B. A. –Martin, J. K. –McLeod, J. D. –Rogers, A. (eds): *Handbook of the Sociology of Health, Illness, and Healing*. A Blueprint for the 21st Century. Springer, New York, Dordrecht, Heidelberg, London.

Relaigh, Veena Soni – Kiri, Victor A. (1997): Life expectancy in England: variations and trends by gender, health authority, and level of deprivation. *Journal of Epidemiology and Community Health* 51(6): 649–658.

Sen, Amartya (1998): Mortality as an Indicator of Economic Success and Failure. *The Economic Journal* 108 (446): 1–25.

Waller, Lance A. – Gotway, Carol A. (2004): *Applied Spatial Statistics for Public Health Data*. John Wiley & Sons, New York.

Wilkinson, R. G. (1992): Income distribution and life expectancy. *British Medical Journal,* 304 (6820): 165–168.

Keywords: life expectancy at micro-regional level, global autocorrelation, local autocorrelation, k-means cluster, spatial demography.

Abstract

This article examines the spatial gender differences of life expectancy at birth in Hungary using the most recent data available. The estimation of life expectancies is based on modified version of Chiang's method which was calculated by the author. We experienced much larger differences amongst men than women. Our empirical result corresponds to the conventional experience. The range regarding men was 9.5 years and 6 years for women. Both values refer to enormous large spatial disparities. Nonetheless the strength of relationship between genders, based on a simple correlation coefficient was very high (r=0,83). It implies that genders are good predictors to one another. The exploration of spatial pattern of life expectancy is based on well-known global and local spatial autocorrelation tests. It has been found that the global spatial autocorrelations are somewhat larger for men than for women. However the differences are not remarkable. Not surprisingly the local patterns were very similar as well. Minor differences were found regarding the extent of local clusters. Areas with high life expectancies can be found in the central region, especially in the capital and its sorrounding zones, and Western Transdanubia region. The areas with the lowest life expectancy lie in South Western and North-Eastern regions of the country.

Finally we attempted to partition areas according to women's life expectancy and gender differences. We used k-means cluster method and silhouette coefficient (SC) to classify proper cluster structure. Because of the above mentioned strong relationship we could define only two clusters. SC value was more than 0.5. We experienced that, where the life expectancy was high or relatively high the gender differences were significantly smaller.