*Education Inequality Within the European Union: A Spatial Statistics Approach*

Andrea Furková, University of Economics in Bratislava, Slovakia

The Asian Conference on Education 2023
Official Conference Proceedings

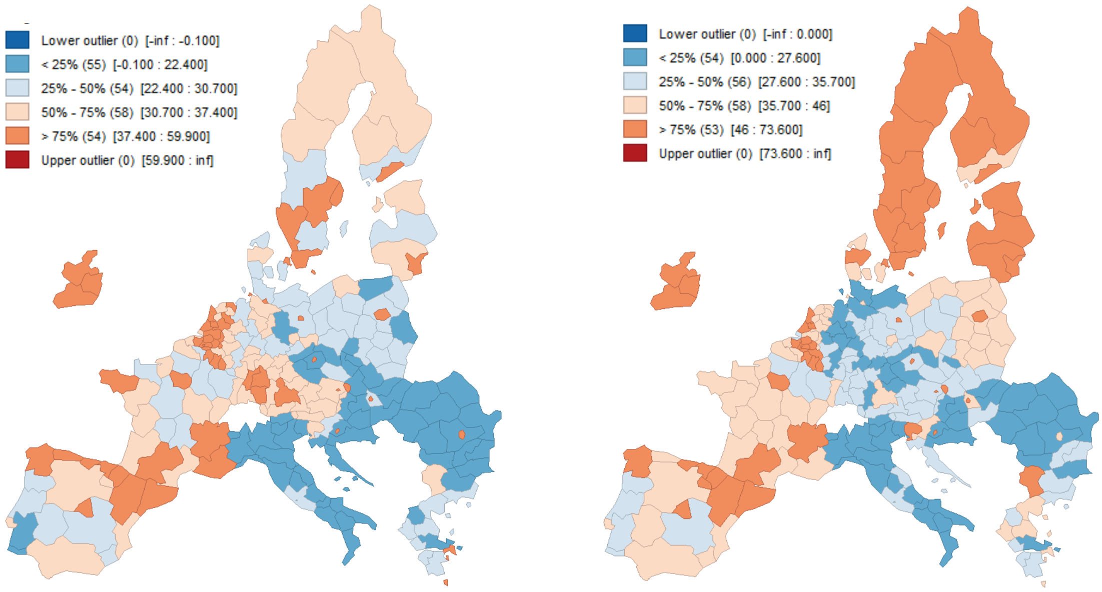
**Abstract**
This study aims to investigate the spatial distribution of educational inequality within the European Union (EU) using a spatial statistics approach. Educational inequality remains a significant challenge for policymakers seeking to ensure equal opportunities for all EU citizens. By examining the spatial dimension of this issue, the study contributes to a deeper understanding of the geographical patterns of educational disparities across EU member states. The research employs spatial statistical techniques, including Exploratory Spatial Data Analysis (ESDA) to analyse data on educational indicators such as educational attainment levels and other relevant factors. These indicators are obtained from Eurostat for the most recent years available. The analysis will be carried out in the context of EU regions and we will focus on tertiary education. The findings reveal significant spatial variations in higher educational inequality across EU member states. ESDA techniques help identify clusters of regions with pronounced disparities in access to higher education, providing valuable insights for targeted policy interventions. Spatial autocorrelation analysis quantifies the extent of spatial dependence, highlighting areas where similar levels of higher educational inequality are clustered. Despite efforts to promote equal opportunities, disparities in access to higher education persist across member states. By analysing the spatial dimension of this issue, the study contributes to a comprehensive understanding of the geographical patterns of higher educational inequality in the EU.

Keywords: Tertiary Educational Attainment Level, Educational Inequality Within the European Union, Spatial Autocorrelation

## Introduction

Education is a fundamental right and a crucial driver of economic growth and social well-being. Within the European Union (EU), the pursuit of equal educational opportunities for all citizens has been a central tenet of regional policy. The European Union has made significant strides in promoting education as a means of fostering social cohesion and economic development. In the context of the EU, The European Education Area strategic framework (Council of the European Union, 2021) was put in place to structure collaboration between EU Member States and key stakeholders to achieve their collective vision. As part of this collaboration, several goals related to education were set. For example, by 2030 at least 45% of 25-34-year-olds should have a higher education qualification. Despite the attention paid to the issue of education in the EU, educational inequality remains a persistent challenge, with disparities in educational attainment posing barriers to social mobility and economic development. Numerous studies have explored educational inequality within the EU, with a particular emphasis on gender disparities, socioeconomic factors, and policy interventions (e.g. Muszynska & Wedrowska, 2023; Palmisano et al., 2022). However, the spatial dimension of these disparities, along with their localized geographical patterns, has received comparatively less attention. Exploratory Spatial Data Analysis (ESDA) techniques offer valuable insights into the localized nature of educational inequality. This paper addresses the issue of gender-based disparities in tertiary educational attainment within the EU by employing Local Moran and Local Geary statistics. By focusing on local spatial patterns, we aim to gain a deeper understanding of the geographical distribution of educational disparities across EU member states.

Figure 1 (two box maps – Male and Female) illustrates spatial distribution of population by tertiary educational attainment level across the European regions separately for Male and Female. In addition, to point out the gender-based disparities in tertiary educational attainment, two boxplots were constructed (Male and Female). Local spatial patterns evident from Figure 1 indicate gender-based educational disparities. Significant differences in tertiary education between men and women are also confirmed by the comparison of boxplots presented in Figure 2.




(a) (b)

Figure 1: Box maps – population by tertiary educational attainment level (in %): Male (a) and Female (b)

Source: author's elaboration in GeoDa

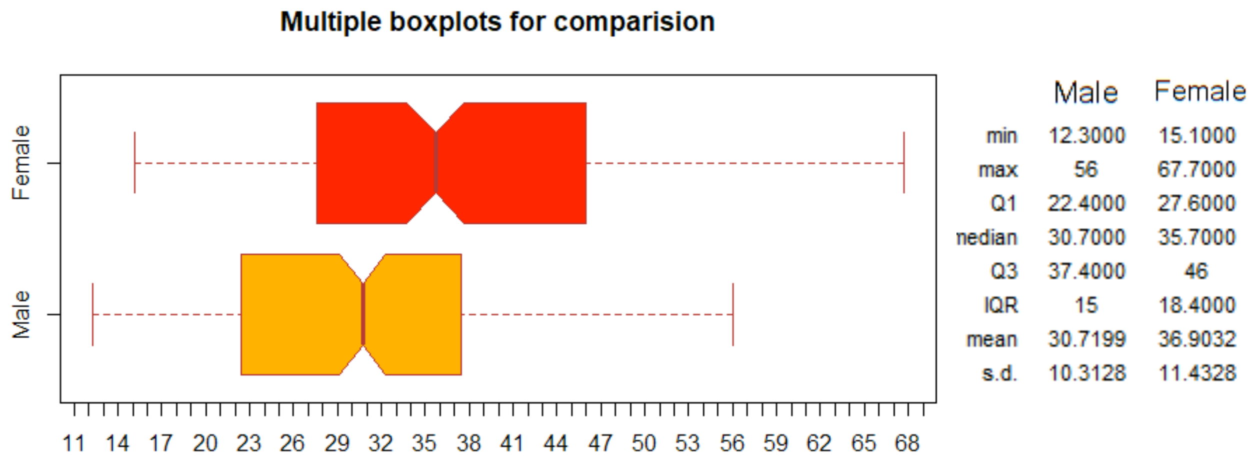



Figure 2: Multiple boxplots for comparison – population by tertiary educational attainment level (in %): Male and Female.

Notes: Q1 – Fist quartile; Q3 – Third quartile; IQR – Interquartile range; s.d. – Standard deviation

Source: author's elaboration in RStudio

## 2. Research Methods

In this section, we will briefly introduce the primary methodological framework underpinning our empirical analysis. Our empirical analysis relies on selected tools from Exploratory Spatial Data Analysis (ESDA). ESDA tools facilitate the assessment of spatial connections among observations, or spatial units (such as regions or countries). Spatial association, also known as spatial autocorrelation or spatial dependence, occurs when spatial units are not independent across the geographic area. This implies that neighbouring spatial units are linked in some way. For a more in-depth exploration of this concept, we can refer to works like Getis (2010) or Anselin & Rey (2014).

ESDA encompasses various techniques that help us describe and visualize spatial distributions, identify unusual locations or spatial outliers, and uncover patterns of spatial association, clusters, or hotspots. To detect spatial autocorrelation, we can employ global and local indicators of spatial association, including well-known statistics like Moran's *I*, Getis-Ord statistics, and Geary's C statistic. These statistics are employed to examine the overall spatial autocorrelation of the variable of interest, essentially testing for general spatial trends across the entire area. Conversely, the local versions of these statistics allow for a more detailed analysis of local spatial patterns. In this paper, we specifically consider local versions of Moran's *I* and Geary's *C* statistics to measure spatial associations.

A local version of Moran's *I* statistic has been proposed by Anselin (1995) to further analyse local spatial patterns. In this case, particular location *i* is fixed. The local Moran's $I_i$ statistic for the location *i* is defined as (Feldkircher, 2006):

$$I_i = \frac{\left(x_i - \bar{x}\right)}{\frac{1}{N}\sum_{k=1}^{N}\left(x_k - \bar{x}\right)^2} \sum_{j=1}^{N} w_{ij}\left(x_j - \bar{x}\right) \tag{1}$$

where $x_i$ represents the underlying variable for region *i*, $\bar{x}$ represents the mean of the variable, *N* is the number of regions in the data set and $w_{ij}$ are the elements of spatial weight matrix $\mathbf{W}$ of dimension $N \times N$ (for more details see, e.g., Getis, 2010 or Anselin & Rey, 2014). Each location (region) has an associated test statistic and spatial pattern can be visualised by cluster map. This graphical tool enables to detect which of the spatial unit has a statistically significant relationship with its neighbours, and show the type of relationship (*high-high* and *low-low* – positive spatial associations or *high-low*, *low-high* – negative spatial associations).

Next, we briefly discuss a local Geary statistic. As in its global counterpart (for more details see Anselin, 2019b), the focus is on squared differences, or, rather dissimilarity than similarity. Small values of the statistic suggest positive spatial autocorrelation (see Getis, 2010), whereas large values suggest negative spatial autocorrelation. The local Geary statistic takes on the following form:

$$G_i = \sum_{j} w_{ij}\left(x_i - x_j\right)^2 \tag{2}$$

where all variables were defined before.

Statistical inference can be based on a conditional permutation procedure and is interpreted in the same way as for, e.g., local Moran statistic or Getis Ord statistic. However, the interpretation of significant locations in terms of the type of association is not as straightforward for the local Geary as it is for the local Moran statistic. Closer examination (see formula (2)) reveals that this statistic consists of a weighted sum of the squared distance in attribute space for the geographical neighbours of observation *i*. The attribute similarity is not a cross-product, and thus has no direct correspondence with the slope in a scatter plot (Anselin, 2019a; Anselin, 2019b).

## 3. Empirical Results

The paper uses a set of data from the Eurostat regional statistical database (Eurostat, 2023) to perform spatial education inequality analysis. Database contains 221 European regions at NUTS 2 level (NUTS - Nomenclature of territorial units for statistics). Figures 3 and 4 provide an overview of the study area. These figures (left sides) show real spatial distributions for population by tertiary educational attainment level (male and female) across the EU regions.

It is clear that the levels of tertiary education attainment among men and women are not evenly distributed, but the level of education probably tends to be spatially correlated. Figures 3 and 4 illustrate the difference between the true – likely spatially autocorrelated distribution (left) and the simulated random distribution (right) for the population by levels of tertiary education attainment. The well-known Moran's *I*[1] test can be considered a quick check for spatial autocorrelation. If the observations are randomly distributed in space, there should be no particular relationship between the indicator population by level of tertiary education attainment and its spatial lag. This is the case of a simulated random distribution (see Figures 3 and 4 - right sides). The corresponding values of the Moran *I* statistic are shown in Table 1. Conversely, the observations have a particular spatial structure if the corresponding values of the Moran *I* statistic are statistically significant. It is therefore clear that the geographical location of the region and the characteristics of the neighborhood play a significant role in the analysis of educational inequalities within the EU.

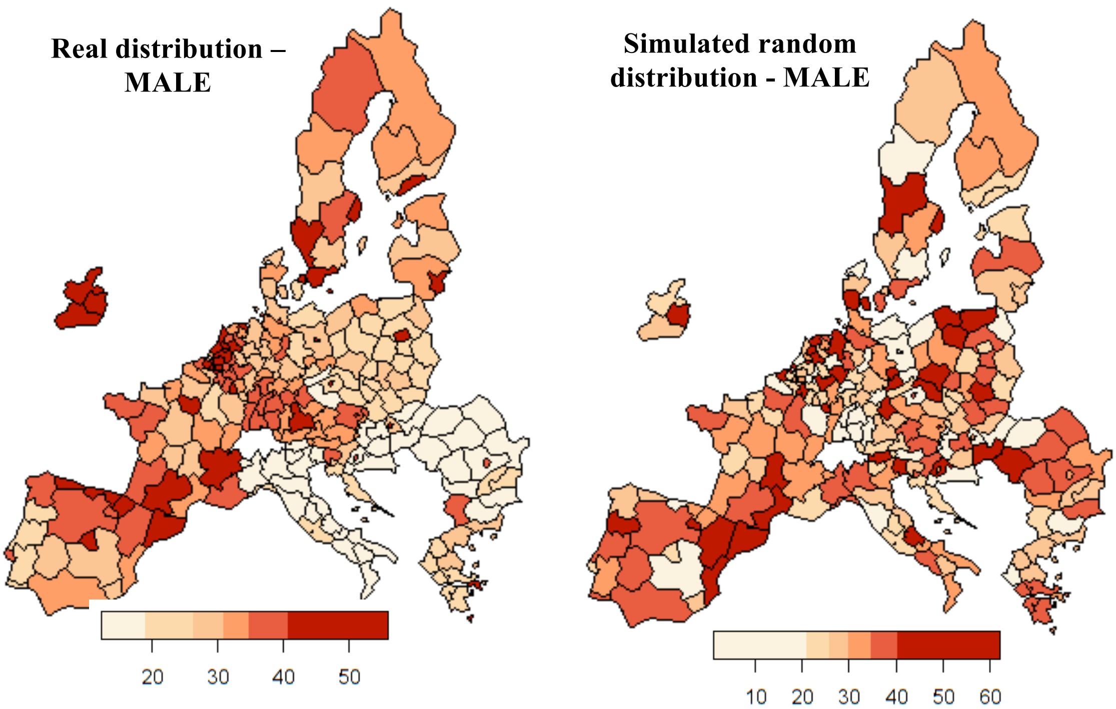



Figure 3: Illustration of the gap between real - spatially autocorrelated distribution (left) and simulated random distribution (right) for population by tertiary educational attainment level (in %) - Male
Source: author's elaboration in RStudio

[1] For calculation of Moran´*I* statistics, spatial weighting matrix of queen contiguity scheme was used. This form of matrix is used in all parts of our spatial analysis (for more details see, e.g., Anselin & Rey 2014).

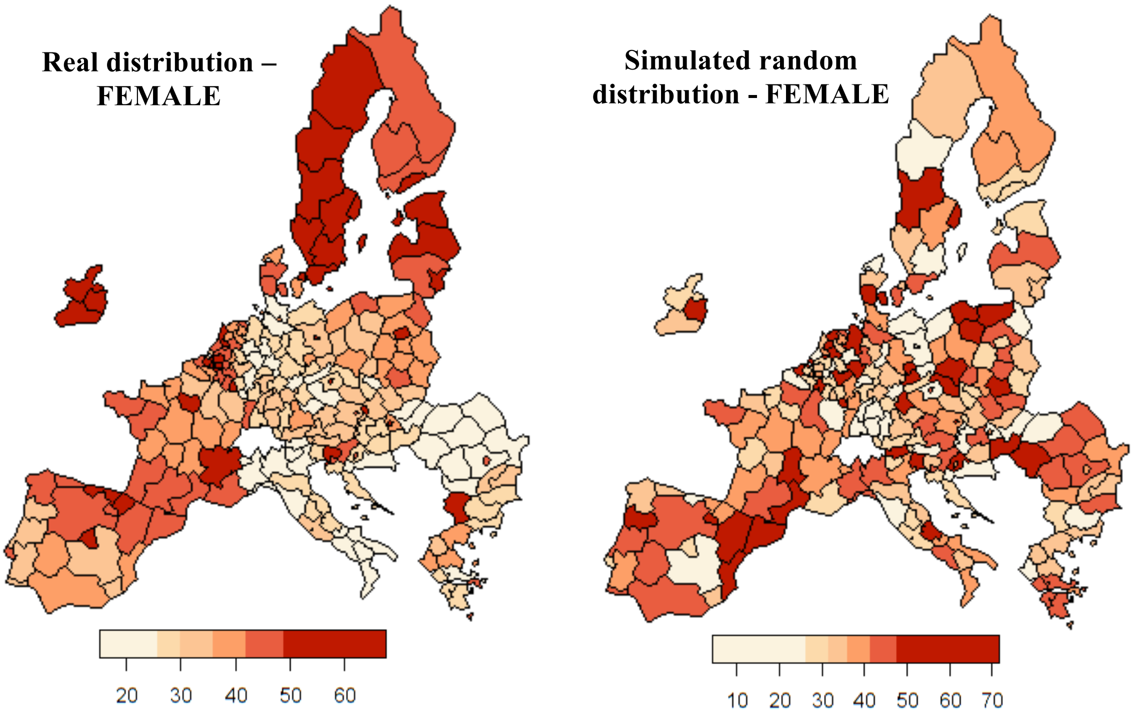


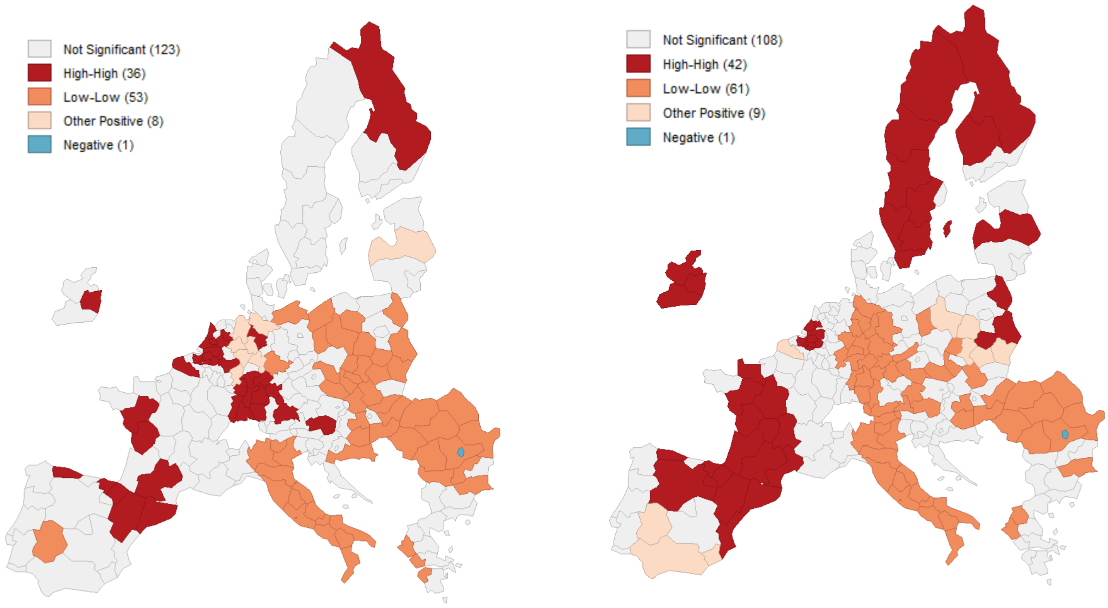
Figure 4: Illustration of the gap between real - spatially autocorrelated distribution (left) and simulated random distribution (right) for population by tertiary educational attainment level (in %) - Female
Source: author's elaboration in RStudio

| | Population by tertiary educational attainment level (in %) - MALE | | Population by tertiary educational attainment level (in %) - FEMALE | | Population by tertiary educational attainment level (in %) - TOTAL | |
|---|---|---|---|---|---|---|
| | Real distribution | Simulated random distribution | Real distribution | Simulated random distribution | Real distribution | Simulated random distribution |
| Moran´s Index | 0.4843*** | -0.0003 | 0.4280*** | -0.0003 | 0.4965*** | -0.0374 |
| Pseudo *p*-value | 0.0010 | 0.4650 | 0.0010 | 0.4540 | 0.0010 | 0.2510 |

Table 1: Moran's *I* index of population by tertiary educational attainment level (in %) – Male, Female and Total
Note: Symbol *** indicates statistical significance at 1% level of significance

Geary cluster maps (constructed based on the local Geary statistic defined by formula [2]) provide more evidence about indicated unequal distribution and spatial clustering of tertiary education levels within the EU. Based on the Figure 5 we identify significant locations – regions with positive spatial autocorrelation. A significant local Geary statistic that is less than its expected value under the null hypothesis of spatial randomness suggests a clustering of similar values (small differences imply similarity). For those observations, the association high-high or low-low can be detected. Based on the indicator population by tertiary educational attainment level - Male, 36 high-high and 53 low-low local Geary clusters were identified (see Figure 5[a]). Similar spatial pattern can be seen from Figure 5 (b) where the results for the indicator population by tertiary educational attainment level – Female are depicted (42 high-high and 61 low-low local Geary clusters). The high-high locations (so-called hot spots) are mainly regions of Finland, Sweden, France, Ireland and Spain. These regions are regions where high values of tertiary education levels are clustered. Low-low values (so-called cold spots locations) are mainly concentrated in regions of Italy and most of the Eastern European regions.

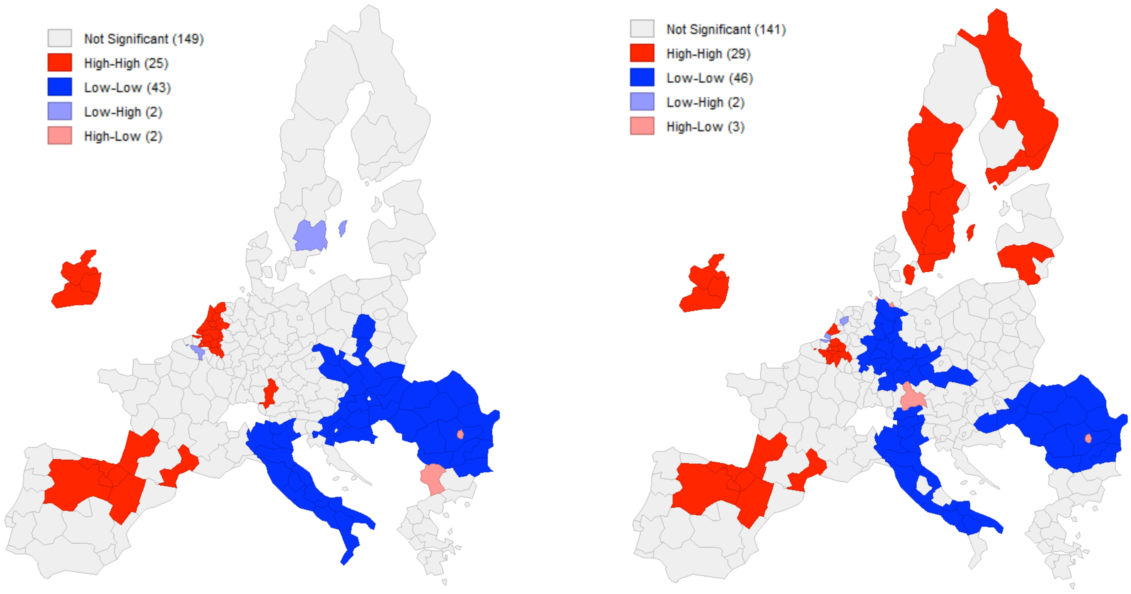
As the calculation of local Geary statistic is based on the squared difference (see formula [2]), there may be observations for which a classification to high-high or low-low clusters is not possible. This is because the squared difference can cross the mean (expected value). These locations are referred as other positive spatial autocorrelation (see Figure 5). As for negative spatial autocorrelation (large values imply dissimilarity), it is not possible to assess whether the association is between high-low or low-high outliers, since the squaring of the differences removes the sign (Anselin, 2019b). In this analysis, there is one region with this type of association for Male and Female.




(a) (b)

Figure 5: Local Geary clusters: population by tertiary educational attainment level (in %): Male (a) and Female (b)
Source: author's elaboration in GeoDa

In contrast to the local Geary statistic presented so far, the local Moran *I* statistic allows to assess whether the association is high-low or low-high in the case of negative spatial autocorrelation. We can see these associations on the Figure 6 (Lisa cluster maps) and they are calculated based on the formula given in (1) . Hot spot as well as cold spot localities detected on the basis of Moran's *I* statistics are in significant agreement with the results from Geary's statistics. The results thus indicate that inequalities in tertiary education within the EU regions are influenced by the spatial distribution of the regions, both with regard to the education of men and women.




(a) (b)

Figure 6: Lisa cluster maps: population by tertiary educational attainment level (in %): Male (a) and Female (b)
Source: author's elaboration in GeoDa

**Conclusion**

The aim of this study was to examine the spatial distribution of educational inequality within the EU using a spatial statistics approach. The research uses spatial statistical techniques such as Geary and Moran statistics belonging to Exploratory spatial data analysis. The analysis was carried out in the context of the EU regions and focused on tertiary education for men and women. The findings reveal significant spatial differences in higher education inequality in the EU member states. By analyzing the spatial dimension of this issue, the study contributes to a comprehensive understanding of the geographic patterns of higher education inequality in the EU, and the results can be used in the creation of regional policies.

**Acknowledgements**

This work was supported by the Grant Agency of the Slovak Republic VEGA 1/0047/23 "The importance of spatial spillover effects in the context of the EU's greener and carbon-free Europe priority".

## References


Anselin, L. (1995). Local Indicators of Spatial Association – LISA. *Geographical Analysis*, *27*(2).

Anselin, L. (2019a). A Local Indicator of Multivariate Spatial Association: Extending Geary's c. *Geographical Analysis*, *51*(2), 133-150.

Anselin, L. (2019b). *GeoDa An Introduction to Spatial Data Analysis*. Retrieved November 9, 2021 from https://geodacenter.github.io/workbook/6a_local_auto/lab6a.html#local-geary.

Anselin, L., & Rey, S.J. (2014). *Modern spatial econometrics in practice: a guide to GeoDa, GeoDaSpace and PySAL*. Chicago, USA: GeoDa Press LLC.

Council of the European Union (2021). *Council Resolution on a strategic framework for European cooperation in education and training towards the European Education Area and beyond (2021-2030) 2021/C 66/01.* Retrieved June 20, 2023 from https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:32021G0226(01)

Eurostat (2023). Statistics by theme. Retrieved June 20, 2023 from http://ec.europa.eu/eurostat/data/browse-statistics-by-theme (10.06.2023).

Feldkircher, M. (2006). *Regional Convergence within the EU-25: A Spatial Econometric Analysis*. Retrieved January 5, 2018 from https://www.oenb.at/Publikationen/Volkswirtschaft/Workshopbaende/2006/Workshop-No.-09.html

Getis, A. (2010). Spatial autocorrelation. In: Fischer, M.M. & Getis, A. *Handbook of Applied Spatial Analysis. Software Tools, Methods and Applications* (pp. 255-278). Berlin, Heidelberg: Springer-Verlag.

Muszynska, J., & Wedrowska, E. (2023). Does Education Affect Income Inequality? A *Comparative Review of Fourteen European Countries. Economy of Regions*, 19(2), 397- 409.

Palmisano, F., & Biagi, F., & Peragine, V. (2022). Inequality of Opportunity in Tertiary Education: Evidence from Europe. *Research in Higher Education*, *63*(4), 1-52.



**Contact email:** andrea.furkova@euba.sk