ISSN: 2675-7222

South American Sciences

PANEL DATA AND SPATIAL ECONOMETRIC: ANALYSIS OF TRANSPORT-RELATED SOCIAL OUTCOMES - A COMPARATIVE STUDY OF ROAD TRAFFIC FATALITIES AND URBAN DEMOGRAPHICS

DADOS EM PAINEL E ECONOMETRIA ESPACIAL: ANÁLISE DE RESULTADOS SOCIAIS RELACIONADOS AO TRANSPORTE — UM ESTUDO COMPARATIVO DE FATALIDADES NO TRÂNSITO RODOVIÁRIO E DEMOGRAFIA URBANA

https://doi.org/10.63330/sasciencesv6n2-041

Submitted on: 06/19/2026 and Published in: 06/26/2026

SAS: e26262

Pedro Silva[1]

Abstract

This paper presents a comprehensive econometric analysis employing two complementary methodological frameworks: panel data models and spatial regression techniques. The first case study examines the determinants of road traffic fatalities across 51 US states during the period 1990-1996, utilizing Ordinary Least Squares (OLS), Fixed Effects, and Random Effects specifications to identify significant socioeconomic and demographic predictors. The second case study investigates the spatial distribution of elderly population (aged 65 and over) in Coimbra, Portugal, based on 2021 Census data, employing Moran's I statistics, Local Indicators of Spatial Association (LISA), and spatial regression models including Spatial Lag and Spatial Error specifications. Our findings reveal that ethanol consumption per capita, income levels, and age structure significantly influence traffic fatalities, while spatial autocorrelation analysis confirms moderate but statistically significant clustering patterns in the distribution of elderly residents and accessible housing infrastructure. The Spatial Error Model emerges as the best-fitting specification for the Coimbra dataset, indicating that unobserved spatially correlated factors substantially influence the spatial distribution of aging population. These results contribute to the understanding of how econometric methods can inform transport policy and urban planning decisions.

[1] University of Coimbra - University of Lisbon - University of Porto

Keywords: Panel Data Models, Fixed Effects, Random Effects, Spatial Econometrics, Moran's I, LISA, Spatial Regression, Traffic Fatalities, Urban Demographics, Coimbra.

Resumo

Este artigo apresenta uma análise econométrica abrangente que emprega duas abordagens metodológicas complementares: modelos de dados em painel e técnicas de regressão espacial. O primeiro estudo de caso examina os determinantes de fatalidades no trânsito rodoviário em 51 estados dos EUA durante o período de 1990 a 1996, utilizando especificações de Mínimos Quadrados Ordinários (MQO), Efeitos Fixos e Efeitos Aleatórios para identificar preditores socioeconômicos e demográficos significativos. O segundo estudo de caso investiga a distribuição espacial da população idosa (com 65 anos ou mais) em Coimbra, Portugal, com base em dados do Censo de 2021, empregando a estatística I de Moran, Indicadores Locais de Associação Espacial (LISA) e modelos de regressão espacial que incluem especificações de Defasagem Espacial (Spatial Lag) e Erro Espacial (Spatial Error). Os resultados revelam que o consumo de etanol per capita, os níveis de renda e a estrutura etária influenciam significativamente as fatalidades no trânsito, enquanto a análise de autocorrelação espacial confirma padrões de agrupamento moderados, porém estatisticamente significativos, na distribuição de residentes idosos e na infraestrutura habitacional acessível. O Modelo de Erro Espacial apresenta-se como a especificação mais adequada para o conjunto de dados de Coimbra, indicando que fatores não observados e espacialmente correlacionados influenciam substancialmente a distribuição espacial da população em processo de envelhecimento. Esses resultados contribuem para a compreensão de como métodos econométricos podem embasar decisões de políticas de transporte e planejamento urbano.

Palavras-chave: Modelos de Dados em Painel, Efeitos Fixos, Efeitos Aleatórios, Econometria Espacial, Moran's I, LISA, Regressão Espacial, Fatalidades de Trânsito, Demografia Urbana, Coimbra.

INTRODUCTION

The analysis of transport-related social outcomes demands sophisticated econometric techniques capable of handling complex data structures. Panel data models and spatial regression methods represent two fundamental approaches that have gained widespread acceptance in the transport research community. These methodologies enable researchers to account for unobserved heterogeneity, temporal dynamics, and spatial dependencies that conventional cross-sectional Ordinary Least Squares (OLS) methods often fail to capture adequately.

Road traffic fatalities constitute a critical public health concern, with the World Health Organization estimating approximately 1.19 million deaths annually worldwide due to road traffic crashes. Understanding the socioeconomic and behavioral determinants of traffic fatalities is essential for designing effective countermeasures and policy interventions. Simultaneously, the spatial distribution of vulnerable populations, particularly the elderly, within urban environments has profound implications for transport accessibility, healthcare service delivery, and urban planning.

This paper presents a dual case study approach. The first case study examines the determinants of road traffic fatalities across 51 US states during the period 1990-1996, employing panel data techniques to isolate state-specific effects from time-varying influences. The second case study investigates the spatial patterns of elderly population distribution in Coimbra, Portugal, utilizing spatial autocorrelation analysis and spatial regression models to identify clustering patterns and spatial dependencies.

The remainder of this paper is organized as follows: Section 2 provides a concise literature review; Section 3 describes the methodological framework; Sections 4 and 5 present the two case studies with their respective results; and Section 6 concludes with policy implications and directions for future research.

LITERATURE REVIEW

The application of panel data models in transport safety research has a well-established tradition. Baltagi (2013) demonstrated that panel data methods offer significant advantages over pure cross-sectional or time-series approaches by controlling for unobserved heterogeneity and providing more efficient parameter estimates. In the context of traffic safety, Noland (2003) utilized panel data techniques to examine the relationship between traffic fatalities and various policy interventions across US states, finding significant state-specific effects that would have been overlooked in pooled regression models.

Spatial econometrics has emerged as an essential tool for analyzing geographic phenomena. Anselin (1988) pioneered the development of spatial regression models, introducing the Spatial Lag and Spatial Error specifications that have become standard in the field. Moran's I statistic (Moran, 1950) remains the most widely used measure of global spatial autocorrelation, while Local Indicators of Spatial Association (LISA), introduced by Anselin (1995), enable the identification of local clustering patterns and spatial outliers.

Recent applications of spatial analysis in transport and urban studies include research on housing accessibility (Borsdorf and Zemp, 2020), demographic transitions (Logan et al., 2020), and the relationship between built environment characteristics and population health outcomes (Ewing and Cervero, 2010). These studies consistently demonstrate that ignoring spatial dependencies can lead to biased estimates and incorrect inferences.

METHODOLOGICAL FRAMEWORK

PANEL DATA MODELS

Panel data combines cross-sectional and time-series dimensions, providing a richer dataset than either approach alone. For a dependent variable y observed for individuals i = 1,...,N over time periods t = 1,...,T, the general panel data model can be expressed as:

$$y\_it = \alpha + X'\_it\beta + u\_it$$

where u_it represents the composite error term. Three primary estimation approaches are employed:

0. **Pooled OLS:**Assumes homogeneous intercept across all units, effectively ignoring panel structure: $u\_it = \varepsilon\_it$
1. **Fixed Effects (FE):**Controls for time-invariant unobserved heterogeneity: $u\_it = \mu\_i + \varepsilon\_it$, where μ_i captures unit-specific effects
2. **Random Effects (RE):**Treats unit-specific effects as random: $\mu\_i \sim IID(0, \sigma^2\_\mu)$, independent of ε_it

The choice between FE and RE is formally tested using the Hausman specification test. Under the null hypothesis, both estimators are consistent but RE is efficient; under the alternative, only FE is consistent. Additionally, the Lagrange Multiplier (Breusch-Pagan) test evaluates whether panel data methods offer improvements over pooled OLS.

SPATIAL ECONOMETRIC MODELS

Spatial econometrics addresses dependencies among observations that arise from geographic proximity. The fundamental concept is the spatial weights matrix W, which defines the neighborhood structure for each observation. The Queen contiguity criterion, employed in this study, considers polygons as neighbors if they share any vertex.

Global spatial autocorrelation is assessed using Moran's I statistic:

$$I = (N/S\_0) \times (\Sigma\_i \Sigma\_j w\_ij z\_i z\_j) / (\Sigma\_i z\_i^2)$$

where $z_i = y_i - \hat{y}$, w_{ij} are elements of the spatial weights matrix, and $S_0 = \Sigma_i \Sigma_j w_{ij}$. Moran's I ranges from -1 (perfect dispersion) to +1 (perfect clustering), with values near zero indicating random spatial distribution.

For regression analysis with spatial dependence, two primary model specifications are considered. The Spatial Lag Model (SLM) incorporates spatial dependence in the dependent variable:

$$y = \rho W y + X\beta + \varepsilon$$

The Spatial Error Model (SEM) accounts for spatial autocorrelation in the error term:

$$y = X\beta + u, \;\; where\ u = \lambda W u + \varepsilon$$

Model comparison is based on information criteria (AIC, Schwarz), log-likelihood values, and the coefficient of determination (R^2).

CASE STUDY I: ROAD TRAFFIC FATALITIES IN US STATES (1990-1996)

DATA DESCRIPTION AND VARIABLE SELECTION

The dataset comprises annual observations for 51 US states over the seven-year period from 1990 to 1996, yielding 350 observations after accounting for missing values. The dependent variable is the number of fatalities resulting from road accidents. Table 1 describes the independent variables considered in the analysis.

Table 1

Description of Variables for Traffic Fatality Analysis

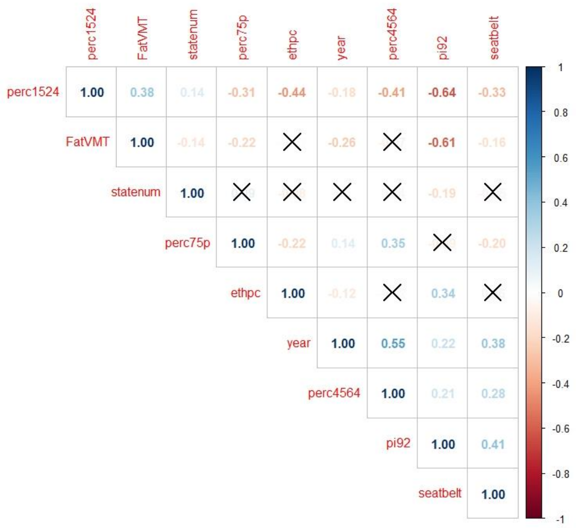
Variable	Description	Expected Relationship
PI92	Per capita income (USD)	Negative — higher income enables safer vehicles and better infrastructure
ETHPC	Ethanol consumption per capita	Positive — alcohol impairs driving ability
Seatbelt	Seat belt usage rate	Negative — safety device reduces fatality risk
perc1524	Population aged 15-24 (%)	Uncertain — younger drivers may have more accidents but better recovery
perc4564	Population aged 45-64 (%)	Uncertain — experienced drivers but higher exposure
perc75p	Population aged 75+ (%)	Positive — elderly are more vulnerable to injuries

Source: Compiled by the author (2026).

Prior to model estimation, a correlation analysis was conducted to identify potential multicollinearity issues. Figure 1 presents the correlation matrix for all variables. The variable perc1524 exhibited a high correlation of -0.64 with PI92, suggesting the need to estimate models both with and without this variable.

Figure 1

Correlation Matrix of Variables for Traffic Fatality Analysis

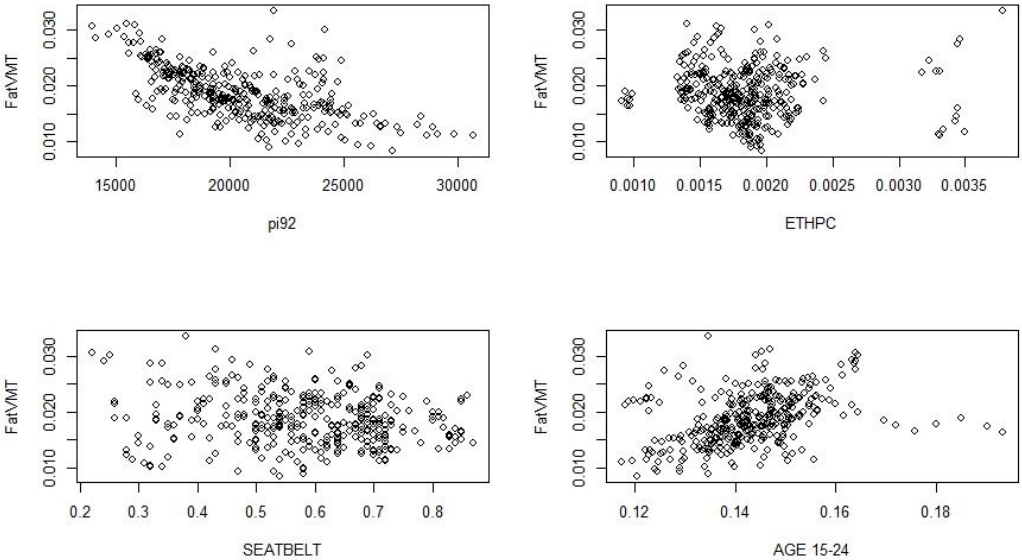


Source: Compiled by the author (2026).

Figures 2 and 3 present the scatter plots showing the bivariate relationships between the dependent variable (fatalities) and each independent variable. These plots provide visual evidence of the direction and strength of each relationship before formal model estimation.

Figure 2

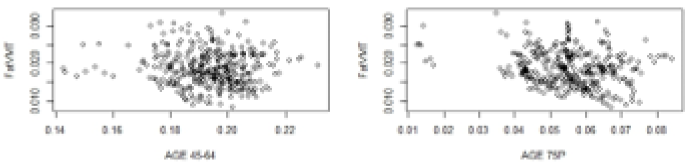
Scatter Plots — PI92, ETHPC, Seatbelt, and AGE 15-24 vs. Fatalities



Source: Compiled by the author (2026).

Figure 3

Scatter Plots — AGE 45-64 and AGE 75+ vs. Fatalities



Source: Compiled by the author (2026).

The Kolmogorov-Smirnov test for normality of the dependent variable yielded a p-value of 0.3352, confirming that the normality assumption is satisfied at the 5% significance level.

OLS BASELINE ESTIMATION

The initial OLS model (Model 1) included all independent variables. To improve linearity, the income variable (PI92) was transformed using the natural logarithm. The model achieved an R-squared of 53.73% and adjusted R-squared of 52.92%, with a highly significant F-statistic of 66.38 ($p < 2.2e-16$), confirming the overall significance of the regression relationship.

The analysis of coefficient estimates revealed that all variables except Seatbelt were statistically significant. Interestingly, the seatbelt coefficient was positive but insignificant ($p = 0.426$), contrary to theoretical expectations. This unexpected result may be attributable to the low variation in seatbelt usage during this period or potential endogeneity issues.

Table 2

*OLS Estimation Results (significance: *** p<0.001, ** p<0.01, * p<0.05, ns = not significant)*

Variable	Model 1	Model 2	Model 3	Model 4
Intercept	0.2806***	0.2793***	0.2544***	0.2484***
log(PI92)	-0.0265***	-0.0263***	-0.0249***	-0.0243***
ETHPC	1.7635***	1.6637***	2.2180***	2.1353***
Seatbelt	0.0012 (ns)	—	0.0017 (ns)	—
perc1524	-0.0472*	-0.0520*	—	—
perc4564	0.0611***	0.0641***	0.0668***	0.0724***
perc75p	-0.1413***	-0.1473***	-0.1228***	-0.1291***
R^2 (adj)	52.92%	52.97%	52.66%	52.60%
F-statistic	66.38***	79.62***	78.63***	97.81***

Source: Compiled by the author (2026).

The analysis reveals several important findings. First, income (log PI92) consistently shows a negative relationship with fatalities, confirming that wealthier states experience fewer traffic deaths. Second, ethanol consumption per capita (ETHPC) exhibits a strong positive effect, with a one-unit increase associated with approximately 1.7 to 2.2 additional fatalities. Third, the percentage of elderly population (perc75p) surprisingly shows a negative coefficient, which may reflect reduced driving exposure rather than lower risk per trip.

FIXED AND RANDOM EFFECTS MODELS

Building upon the OLS results, panel data models were estimated using the preferred specification (Model 2 variables). Both One-Way and Two-Way Fixed Effects and Random Effects models were estimated.

Table 3

Panel Data Estimation Results

Variable	FE One-Way	FE Two-Way	RE One-Way	RE Two-Way
log(PI92)	-0.0054 (ns)	-0.0004 (ns)	-0.0179***	-0.0133***
ETHPC	4.9327***	0.9698 (ns)	3.4472***	1.7436*
perc1524	0.0629*	0.0426 (ns)	0.0456*	0.0338 (ns)
perc4564	-0.0793**	-0.0477 (ns)	-0.0167 (ns)	-0.0365 (ns)
perc75p	-0.0085 (ns)	0.0239 (ns)	-0.0783*	-0.0710 (ns)
R^2 (adj)	31.28%	-18.78%	39.48%	6.88%
F-statistic	42.57***	0.965 (ns)	232.68***	30.77***

Source: Compiled by the author (2026).

The Fixed Effects One-Way model shows a substantially different pattern compared to OLS. The ethanol consumption coefficient increases dramatically to 4.93, suggesting that unobserved state characteristics were partially masking the true effect in pooled regression. The income variable becomes insignificant, likely because between-state income variation is absorbed by the state fixed effects. The Random Effects model yields more efficient estimates, with both income and ethanol consumption remaining significant.

The Two-Way specifications, which additionally control for time effects, show deteriorated model fit, suggesting that time-specific factors do not substantially improve the model beyond the state fixed effects.

MODEL SELECTION AND DISCUSSION

To determine the optimal model specification, the Lagrange Multiplier test and Hausman test were applied. The Lagrange Multiplier test strongly rejected the null hypothesis of no panel effects ($p < 0.001$), confirming that panel data methods are preferred over pooled OLS. The Hausman test results favored the

Fixed Effects specification, indicating that the unobserved individual effects are correlated with the regressors, rendering Random Effects inconsistent.

Based on these diagnostic tests, the Fixed Effects One-Way model is selected as the preferred specification. The key policy insight is that ethanol consumption per capita emerges as the dominant predictor of traffic fatalities, with the panel model revealing a much larger effect size than cross-sectional analysis. This finding underscores the importance of alcohol control policies in reducing road traffic deaths.

CASE STUDY II: SPATIAL ANALYSIS OF AGING POPULATION IN COIMBRA (2021)

DATA DESCRIPTION

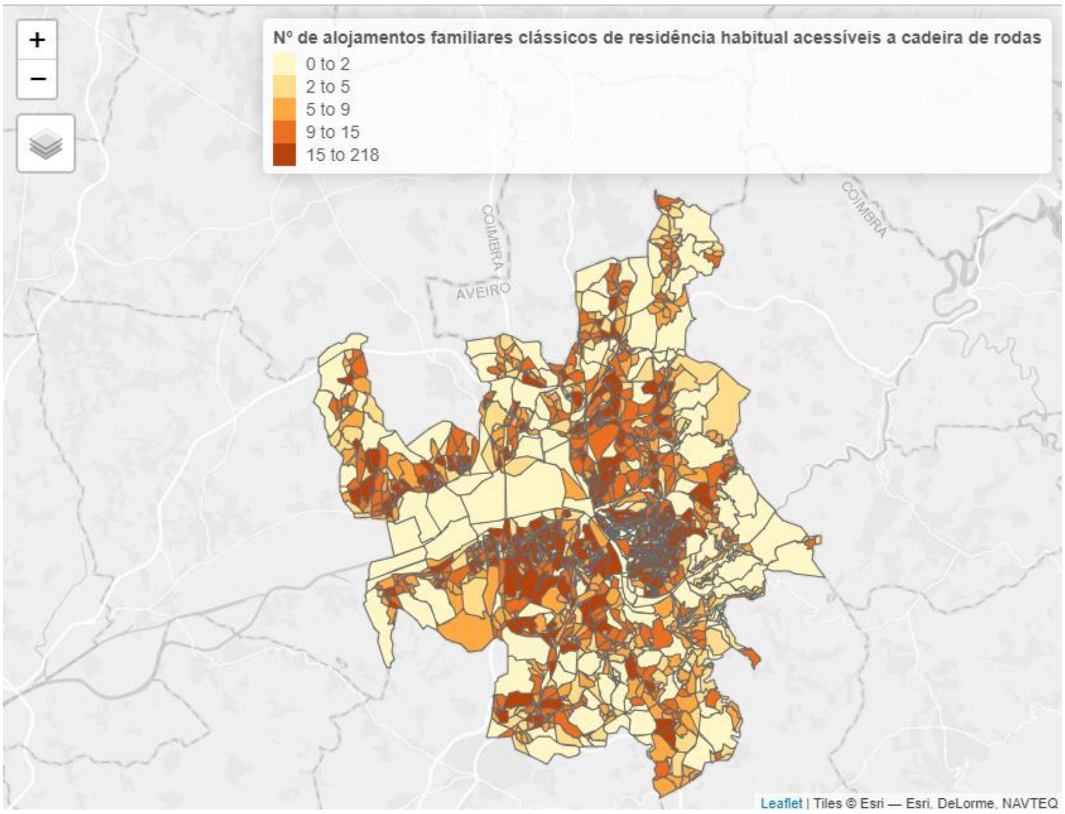
This case study utilizes data from the 2021 Census for the municipality of Coimbra, Portugal. The study area comprises 1,848 statistical subsections. Three variables are analyzed:

- Dependent Variable: Number of people aged 65 or over
- Independent Variable 1: Number of wheelchair-accessible conventional dwellings for habitual residence
- Independent Variable 2: Number of conventional dwellings with parking facilities

The choice of these variables reflects the growing policy interest in understanding the spatial nexus between aging population and housing accessibility, particularly for mobility-impaired individuals. Figures 4 to 6 show the spatial distribution of the three variables across the municipality. Descriptive analysis reveals that the three variables exhibit similar spatial patterns, with higher concentrations in the city center of Coimbra.

Figure 4

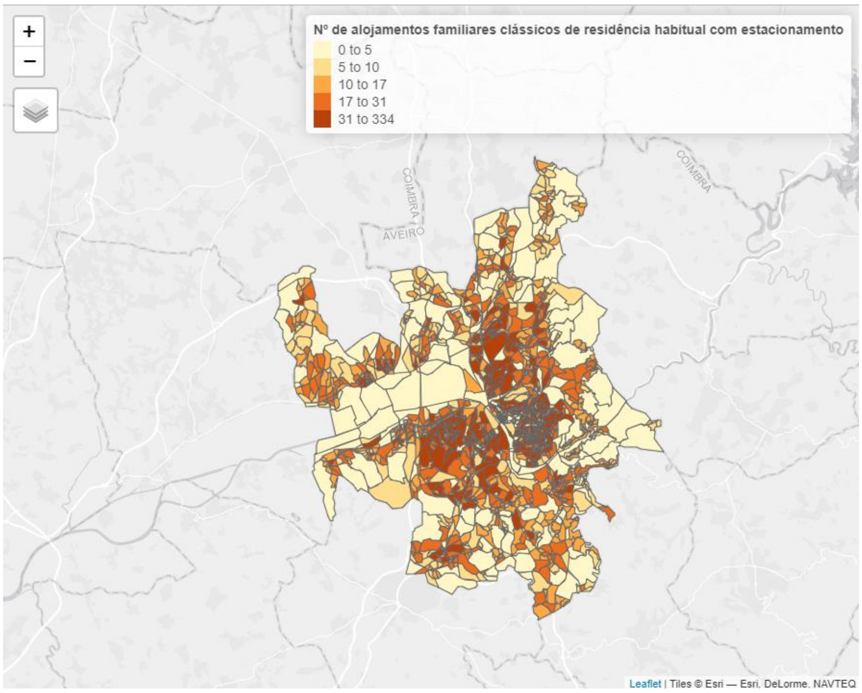
Distribution of Wheelchair-Accessible Dwellings in Coimbra (2021)



Source: Compiled by the author (2026).

Figure 5

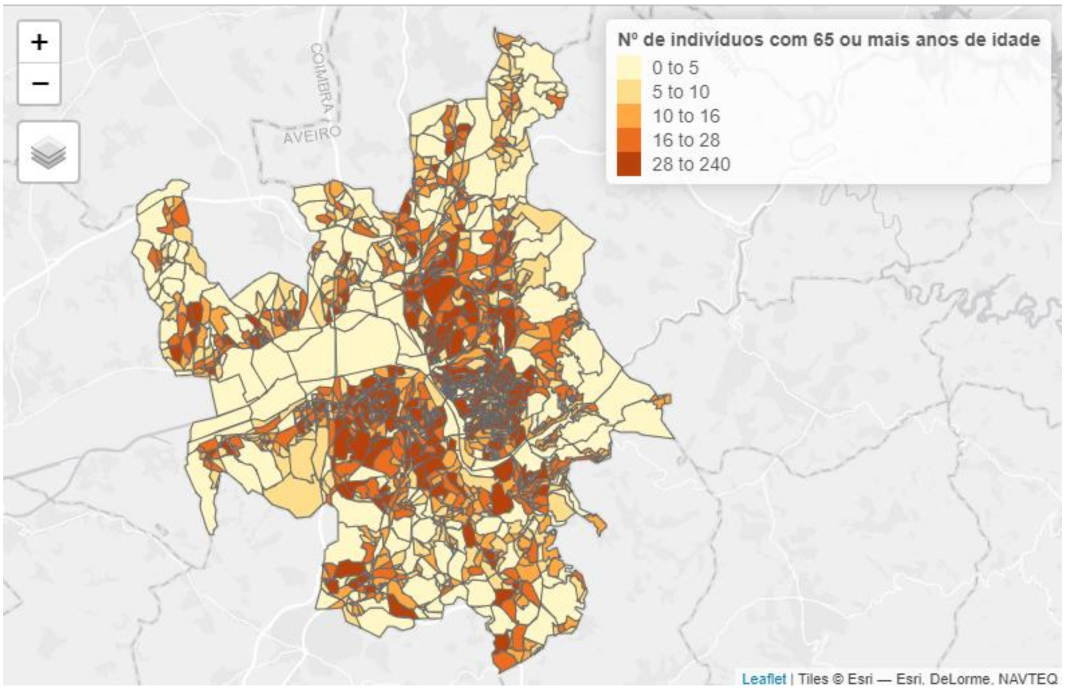
Distribution of Dwellings with Parking Facilities in Coimbra (2021)



Source: Compiled by the author (2026).

Figure 6

Distribution of Population Aged 65+ in Coimbra (2021)



Source: Compiled by the author (2026).

SPATIAL AUTOCORRELATION ANALYSIS

The spatial weights matrix was constructed using the Queen contiguity criterion, resulting in 11,436 nonzero links across 1,848 regions (0.33% density). The connectivity analysis shows an average of 6.19 neighbors per region, with a minimum of 1 (region 975) and maximum of 17 (region 548).

Global spatial autocorrelation was assessed using Moran's I statistic for all three variables. The results are presented in Table 4.

Table 4

Global Moran's I Statistics

Variable	Moran's I	p-value
Wheelchair-accessible dwellings	0.3048	< 2.2e-16
Dwellings with parking	0.3878	< 2.2e-16
Population aged 65+	0.3654	< 2.2e-16

Source: Compiled by the author (2026).

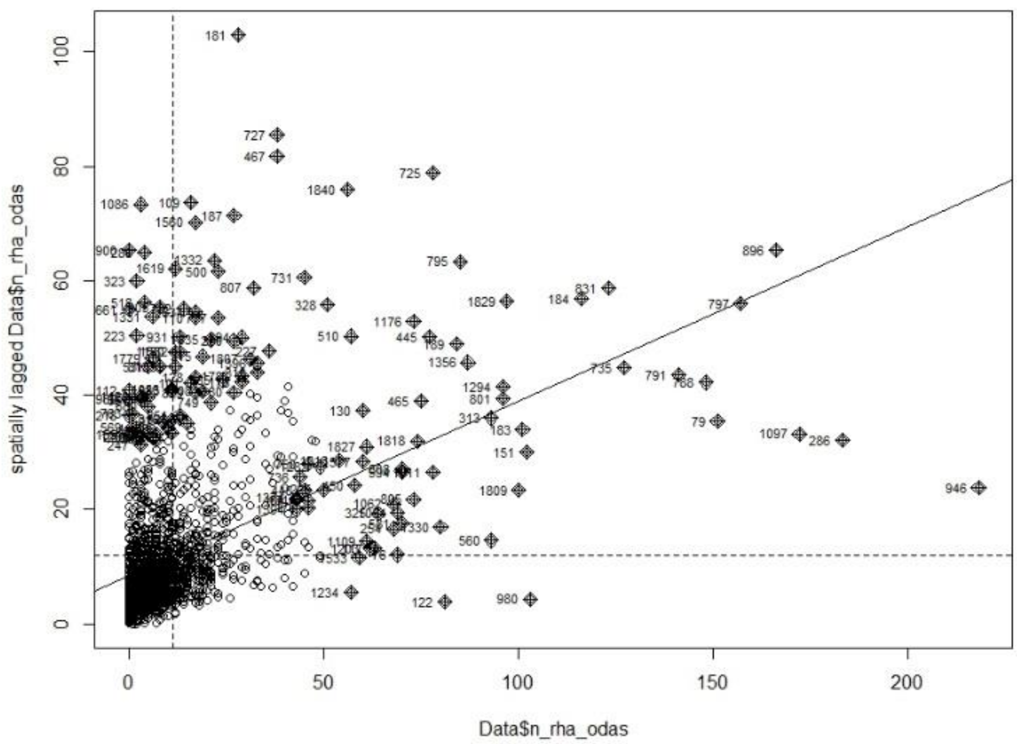
All three variables exhibit positive and highly significant Moran's I values, indicating moderate spatial autocorrelation. This confirms that areas with high (low) values tend to be surrounded by neighbors with similarly high (low) values, forming spatial clusters rather than random distributions.

RStudio Moran Scatterplots

Figures 7 to 9 present the Moran scatterplots generated in RStudio for each variable. These plots visualize the relationship between each region's value and the spatially lagged average of its neighbors, with the slope of the fitted line corresponding to Moran's I.

Figure 7

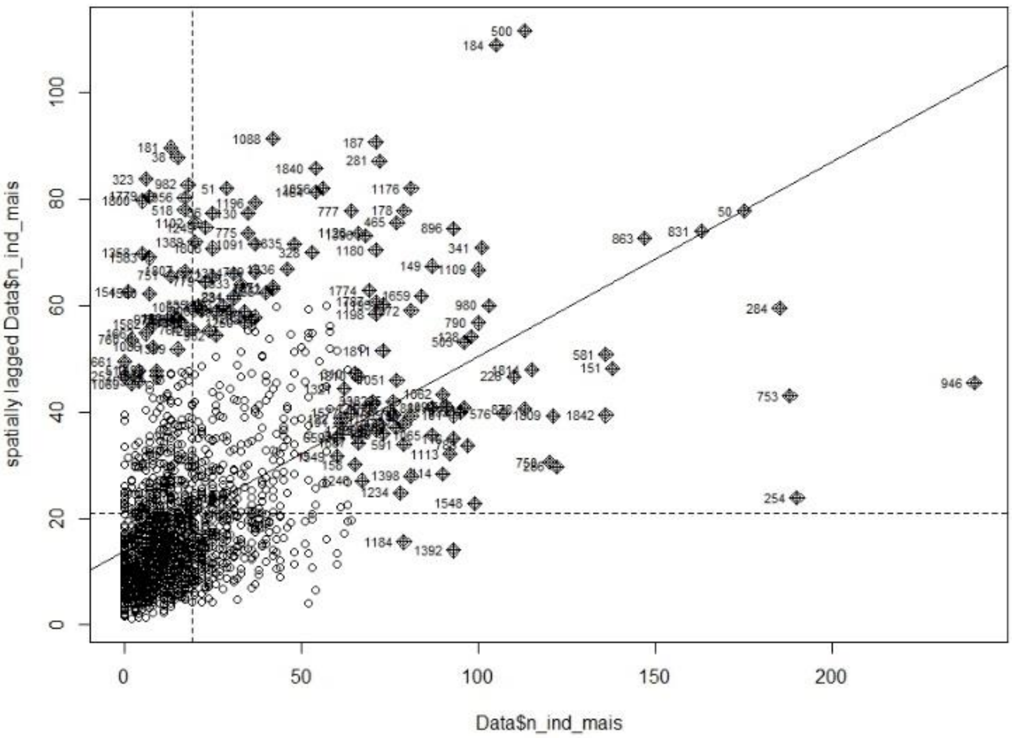
Moran Scatterplot — Wheelchair-Accessible Dwellings (RStudio)



Source: Compiled by the author (2026).

Figure 8

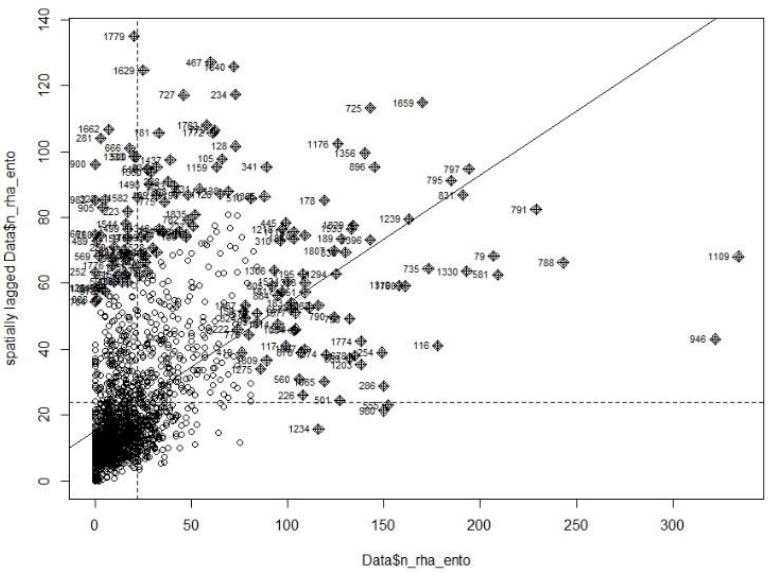
Moran Scatterplot - Dwellings with Parking (RStudio)



Source: Compiled by the author (2026).

Figure 9

Moran Scatterplot - Population Aged 65+ (RStudio)



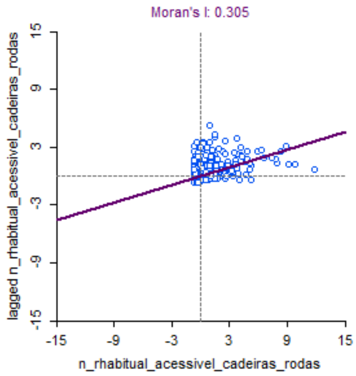
Source: Compiled by the author (2026).

GeoDa Moran Scatter Plots

Figures 10 to 12 show the Moran scatterplots generated in GeoDa, which provide standardized values and clearer visualization of the four quadrants (High-High, Low-Low, High-Low, Low-High) for each variable.

Figure 10

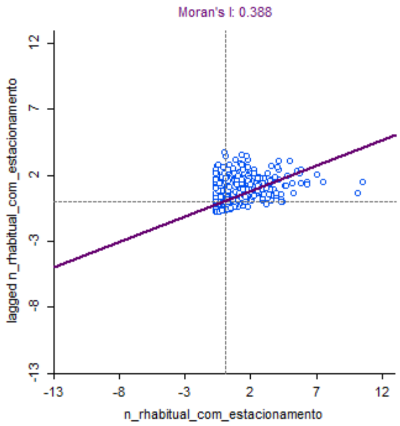
Moran Scatter Plot (GeoDa) - Wheelchair-Accessible Dwellings (I = 0.305)



Source: Compiled by the author (2026).

Figure 11

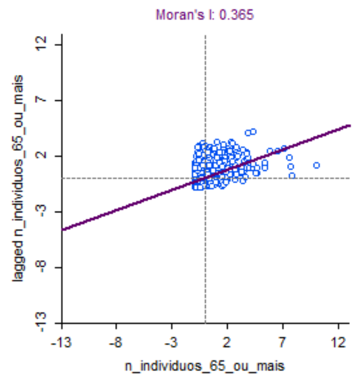
Moran Scatter Plot (GeoDa) - Dwellings with Parking (I = 0.388)



Source: Compiled by the author (2026).

Figure 12

Moran Scatter Plot (GeoDa) - Population Aged 65+ (I = 0.365)



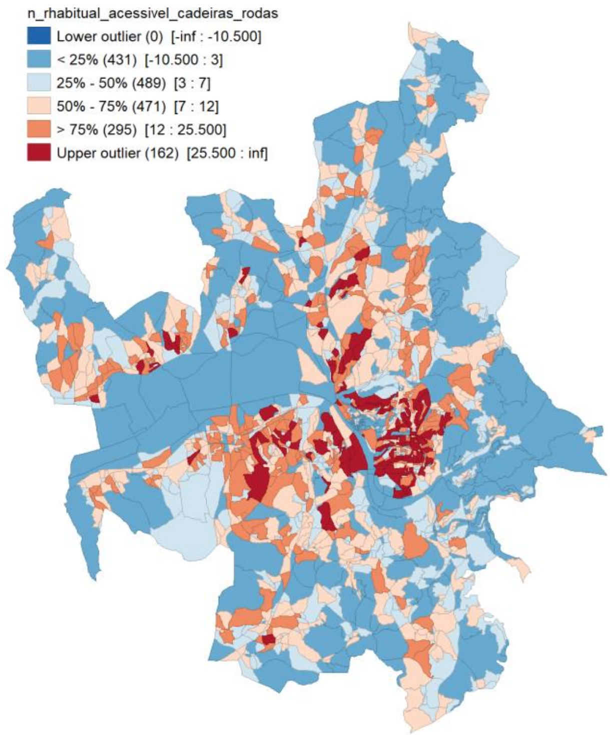
Source: Compiled by the author (2026).

Box Maps

Figures 13 to 15 present the box maps for the three variables, which classify regions based on their deviation from the median value. These maps identify potential outliers and highlight areas with extreme values.

Figure 13

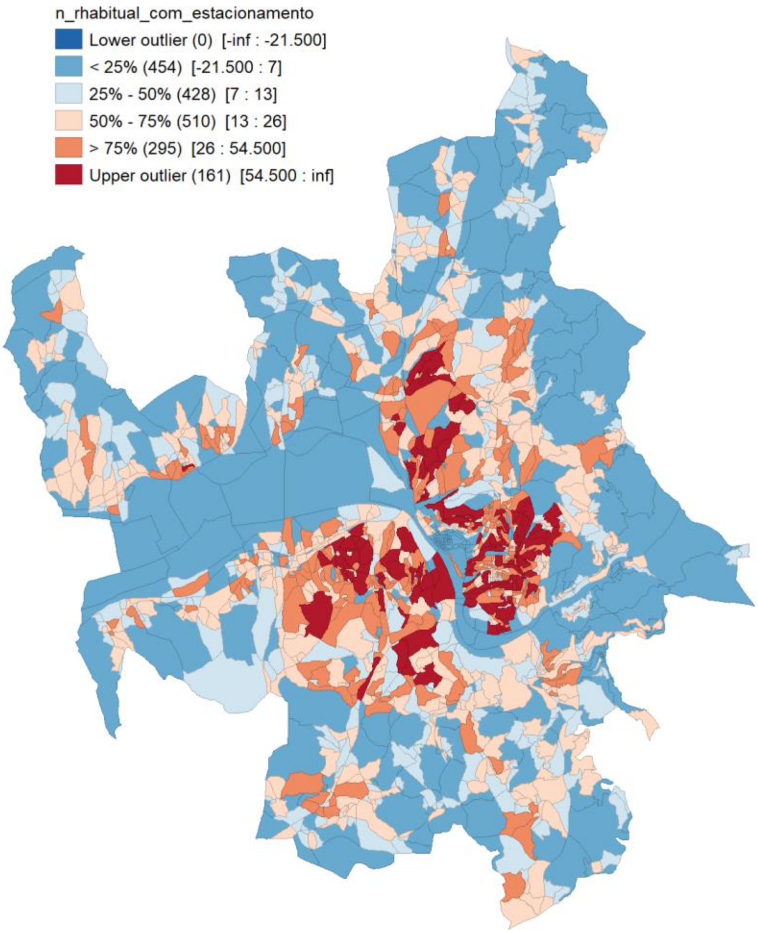
Box Map - Wheelchair-Accessible Dwellings



Source: Compiled by the author (2026).

Figure 14

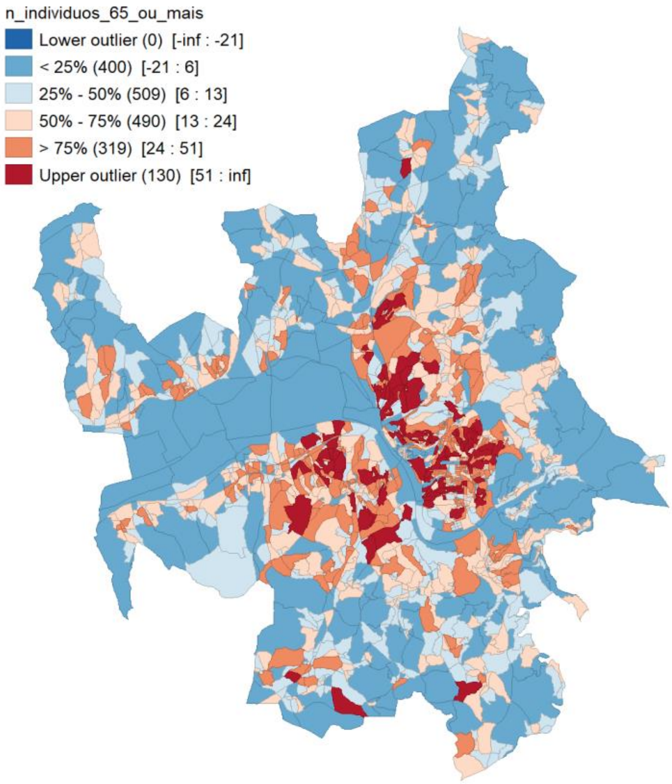
Box Map - Dwellings with Parking



Source: Compiled by the author (2026).

Figure 15

Box Map - Population Aged 65+



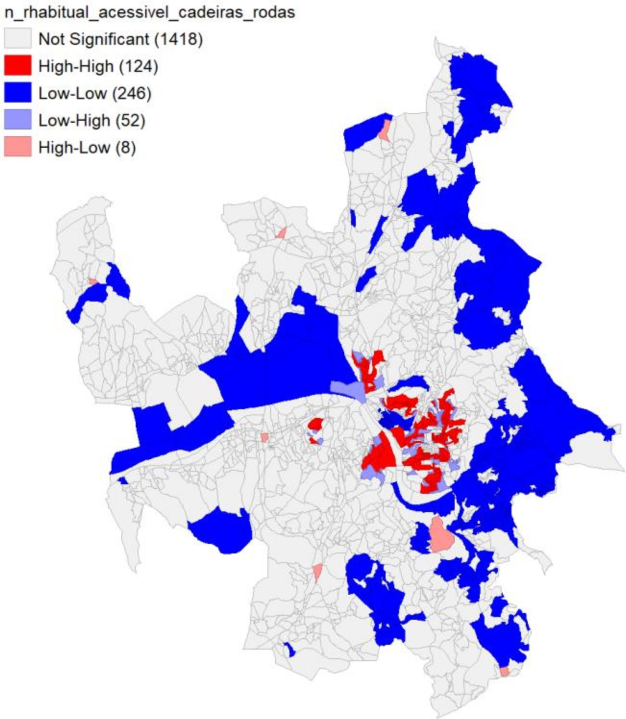
Source: Compiled by the author (2026).

LISA Cluster Maps

Figures 16 to 18 display the LISA (Local Indicators of Spatial Association) cluster maps. These maps identify statistically significant spatial clusters: High-High clusters (red) represent hot spots where high values are surrounded by high values; Low-Low clusters (blue) represent cold spots; and High-Low/Low-High areas (pink/light blue) indicate spatial outliers.

Figure 16

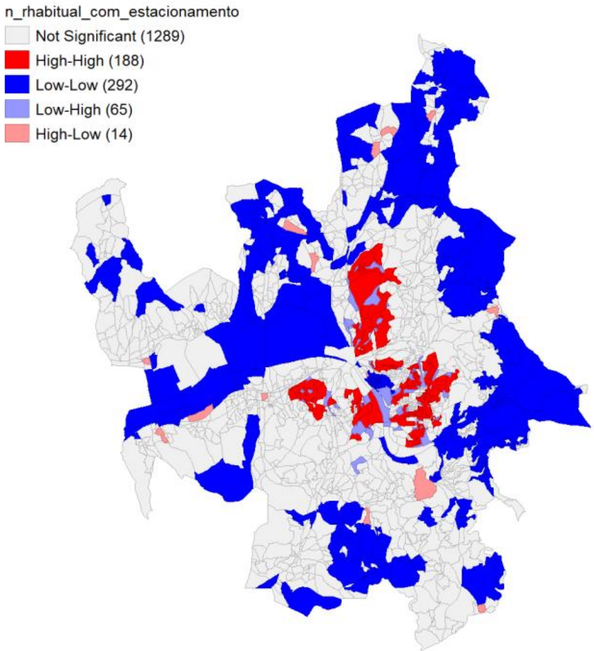
LISA Cluster Map - Wheelchair-Accessible Dwellings



Source: Compiled by the author (2026).

Figure 17

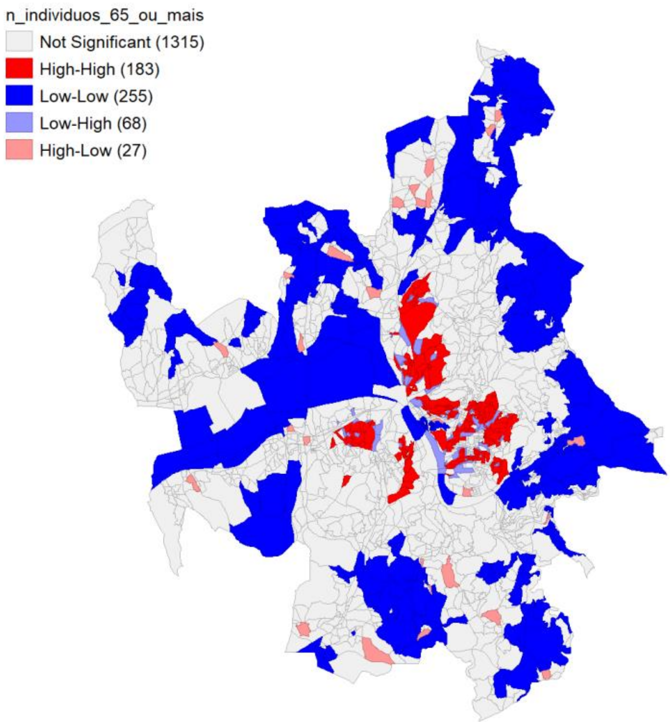
LISA Cluster Map - Dwellings with Parking



Source: Compiled by the author (2026).

Figure 18

LISA Cluster Map - Population Aged 65+



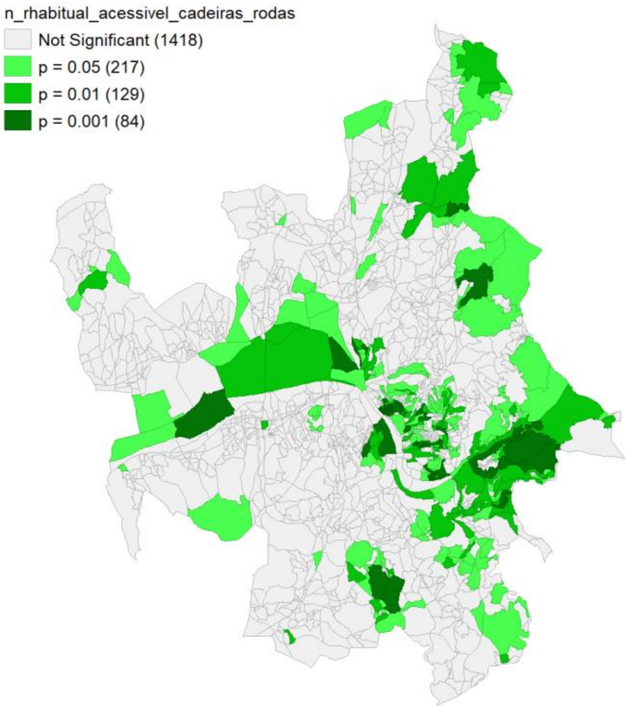
Source: Compiled by the author (2026).

Significance Maps

Figures 19 to 21 show the significance maps, which highlight regions where the local spatial autocorrelation is statistically significant at different confidence levels (p = 0.05, 0.01, 0.001).

Figure 19

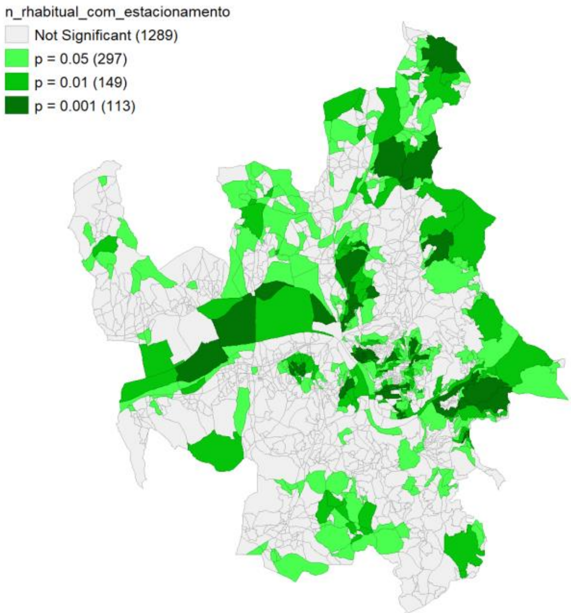
Significance Map - Wheelchair-Accessible Dwellings



Source: Compiled by the author (2026).

Figure 20

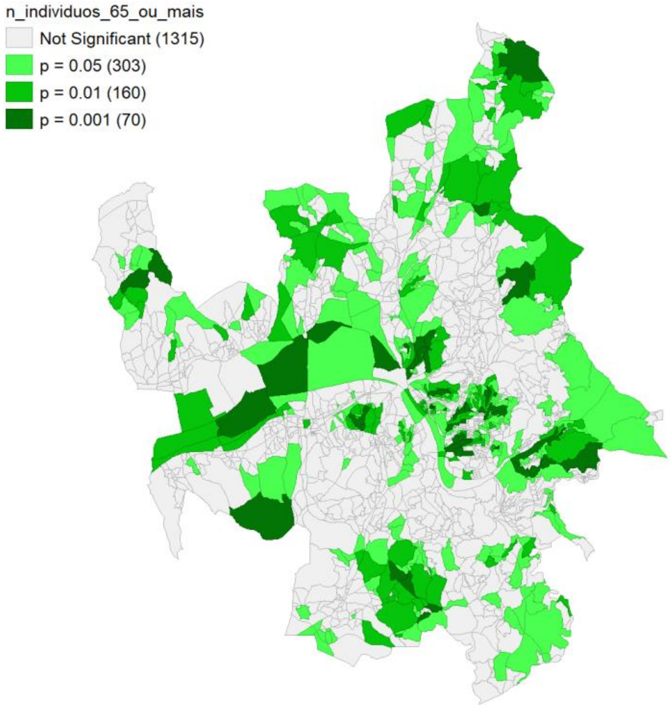
Significance Map - Dwellings with Parking



Source: Compiled by the author (2026).

Figure 21

Significance Map - Population Aged 65+



Source: Compiled by the author (2026).

Local spatial autocorrelation analysis reveals distinct clustering patterns. The High-High clusters (quadrant I), representing areas with high values surrounded by high-value neighbors, concentrate in the city center. Conversely, Low-Low clusters (quadrant III) are predominantly found in the eastern and peripheral areas of Coimbra. The analysis identifies potential spatial outliers (8-9% of observations) concentrated in the central zone, reflecting the heterogeneous urban fabric of the historic city center.

SPATIAL REGRESSION MODELS

Three spatial regression specifications were estimated to explain the spatial distribution of elderly population. The dependent variable is the number of individuals aged 65 or over, and the independent variables are wheelchair-accessible dwellings and dwellings with parking.

The OLS baseline model achieved an adjusted R-squared of 48.0%, with both independent variables statistically significant ($p < 0.05$). However, diagnostic tests revealed violations of key assumptions: the Jarque-Bera test rejected normality of residuals ($p < 0.05$), and the Breusch-Pagan test confirmed the presence of heteroskedasticity. The multicollinearity condition number was low (4.28), indicating no serious collinearity concerns.

Figure 22 shows the complete OLS regression output.

Figure 22

OLS Spatial Regression Output

```
REGRESSION
----------
SUMMARY OF OUTPUT: ORDINARY LEAST SQUARES ESTIMATION
Data set            :  BGRI2021_0603
Dependent Variable  :  n_individuos_65_ou_mais
Number of Observations: 1848
Mean dependent var  :     19.2284  Number of Variables   :    3
S.D. dependent var  :     21.9696  Degrees of Freedom    : 1845

R-squared           :    0.480029  F-statistic           :     851.639
Adjusted R-squared  :    0.479466  Prob(F-statistic)     :           0
Sum squared residual:      463795  Log likelihood        :    -7727.61
Sigma-square        :     251.379  Akaike info criterion :     15461.2
S.E. of regression  :     15.8549  Schwarz criterion     :     15477.8
Sigma-square ML     :     250.971
S.E of regression ML:     15.8421

-----------------------------------------------------------------------------
       Variable       Coefficient       Std.Error     t-Statistic   Probability
-----------------------------------------------------------------------------
          CONSTANT       7.81319        0.461056        16.9463      0.00000
n_rhabi..ras_rodas     0.0786267       0.0357836        2.19728      0.02812
n_rhabi..ionamento      0.473773       0.0210897        22.4646      0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
MULTICOLLINEARITY CONDITION NUMBER    4.280615
TEST ON NORMALITY OF ERRORS
TEST                  DF           VALUE             PROB
Jarque-Bera            2         35679.7328          0.00000

DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                  DF           VALUE             PROB
Breusch-Pagan test     2          2731.8878          0.00000
Koenker-Bassett test   2           236.2660          0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
FOR WEIGHT MATRIX : BGRI2021_0603
   (row-standardized weights)
TEST                          MI/DF        VALUE           PROB
Moran's I (error)             0.2400        17.7007         0.00000
Lagrange Multiplier (lag)       1          228.5665         0.00000
Robust LM (lag)                 1           13.1221         0.00029
Lagrange Multiplier (error)     1          309.3932         0.00000
Robust LM (error)               1           93.9488         0.00000
Lagrange Multiplier (SARMA)     2          322.5154         0.00000

COEFFICIENTS VARIANCE MATRIX
   CONSTANT  n_rhabitual_acessivel_cadeiras_rodas  n_rhabitual_com_estacionamento
   0.212573   -0.000758   -0.003061
  -0.000758    0.001280   -0.000610
  -0.003061   -0.000610    0.000445

============================== END OF REPORT ===============================
```

Source: Compiled by the author (2026).

The Spatial Lag Model achieved an R-squared of 53.27%, with the spatial autoregressive parameter $\rho = 0.300$ being highly significant. Figure 23 presents the full Spatial Lag Model output.

Figure 23

Spatial Lag Model Output

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : BGRI2021_0603
Spatial Weight      : BGRI2021_0603
Dependent Variable  :  n_individuos_65_ou_mais
Number of Observations: 1848
Mean dependent var  :     19.2284  Number of Variables   :    4
S.D. dependent var  :     21.9696  Degrees of Freedom    : 1844
Lag coeff.   (Rho)  :    0.299655

R-squared           :    0.532726  Log likelihood        :    -7643.46
Sq. Correlation     : -            Akaike info criterion :     15294.9
Sigma-square        :     225.536  Schwarz criterion     :       15317
S.E of regression   :     15.0179

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error       z-value    Probability
-----------------------------------------------------------------------------
      W_n_individu      0.299655      0.0243208        12.3209      0.00000
          CONSTANT       3.07439       0.554266        5.54678      0.00000
n_rhabi..ras_rodas     0.0989354      0.0338953        2.91885      0.00351
n_rhabi..ionamento       0.39474      0.0208853        18.9004      0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                        DF       VALUE         PROB
Breusch-Pagan test                          2      2626.0590      0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : BGRI2021_0603
TEST                                        DF       VALUE         PROB
Likelihood Ratio Test                       1       168.3028      0.00000
============================== END OF REPORT ================================
```

Source: Compiled by the author (2026).

The Spatial Error Model achieved the highest R-squared of 55.35%, with $\lambda = 0.442$ being highly significant, confirming the importance of accounting for spatial dependence in the error structure. Figure 24 presents the complete output.

Figure 24

Spatial Error Model Output

```
REGRESSION
----------
SUMMARY OF OUTPUT: SPATIAL LAG MODEL - MAXIMUM LIKELIHOOD ESTIMATION
Data set            : BGRI2021_0603
Spatial Weight      : BGRI2021_0603
Dependent Variable  :  n_individuos_65_ou_mais
Number of Observations: 1848
Mean dependent var  :     19.2284  Number of Variables    :    4
S.D. dependent var  :     21.9696  Degrees of Freedom     : 1844
Lag coeff.   (Rho)  :    0.299655

R-squared           :    0.532726  Log likelihood         :    -7643.46
Sq. Correlation     : -            Akaike info criterion :     15294.9
Sigma-square        :     225.536  Schwarz criterion      :       15317
S.E of regression   :     15.0179

-----------------------------------------------------------------------------
       Variable       Coefficient      Std.Error       z-value    Probability
-----------------------------------------------------------------------------
      W_n_individu      0.299655      0.0243208        12.3209     0.00000
          CONSTANT       3.07439       0.554266        5.54678     0.00000
n_rhabi..ras_rodas     0.0989354      0.0338953        2.91885     0.00351
n_rhabi..ionamento       0.39474      0.0208853        18.9004     0.00000
-----------------------------------------------------------------------------

REGRESSION DIAGNOSTICS
DIAGNOSTICS FOR HETEROSKEDASTICITY
RANDOM COEFFICIENTS
TEST                                     DF       VALUE        PROB
Breusch-Pagan test                       2      2626.0590     0.00000

DIAGNOSTICS FOR SPATIAL DEPENDENCE
SPATIAL LAG DEPENDENCE FOR WEIGHT MATRIX : BGRI2021_0603
TEST                                     DF       VALUE        PROB
Likelihood Ratio Test                    1       168.3028     0.00000

COEFFICIENTS VARIANCE MATRIX
   CONSTANT  n_rhabitual_acessivel_cadeiras_rodas  n_rhabitual_com_estacionamento  W_n_individu
   0.307211    -0.000766    -0.000666    -0.008301
  -0.000766     0.001149    -0.000549     0.000006
  -0.000666    -0.000549     0.000436    -0.000148
  -0.008301     0.000006    -0.000148     0.000592

============================== END OF REPORT ================================
```

Source: Compiled by the author (2026).

MODEL COMPARISON

Table 5

Spatial Regression Model Comparison

Criterion	OLS	Spatial Lag	Spatial Error
Log-likelihood	-7,727.61	-7,643.46	-7,620.56
AIC	15,461.2	15,294.9	15,247.1
Schwarz BIC	15,477.8	15,317.0	15,263.7
R^2	0.4800	0.5327	0.5535
ρ (spatial lag)	—	0.2997	—
λ (spatial error)	—	—	0.4420
Heteroskedasticity	Present	Present	Present

Source: Compiled by the author (2026).

The Spatial Error Model (SEM) emerges as the preferred specification based on all information criteria. It achieves the lowest AIC (15,247.1) and Schwarz BIC (15,263.7) values and the highest R-squared (55.35%). The spatial autoregressive parameter $\lambda = 0.442$ is highly significant, confirming the importance of accounting for spatial dependence in the error structure.

Both independent variables remain statistically significant across all three specifications. The positive coefficients confirm that wheelchair accessibility and parking availability are important predictors of elderly population distribution, likely reflecting both supply-side (housing stock characteristics) and demand-side (elderly preferences for accessible housing) factors.

CONCLUSIONS

This paper has presented two complementary econometric analyses that demonstrate the value of advanced statistical methods in transport and urban research. The panel data analysis of US traffic fatalities (1990-1996) reveals that ethanol consumption per capita is the dominant predictor of road deaths, with the Fixed Effects model uncovering substantially larger effect sizes than cross-sectional OLS. Income levels and age structure also contribute significantly, though the relationship with elderly population share appears complex, likely reflecting behavioral adaptations rather than pure risk exposure.

The spatial analysis of Coimbra's elderly population confirms moderate but statistically significant spatial autocorrelation in the distribution of aging residents and accessible housing. The Spatial Error Model provides the best fit, indicating that unobserved spatially correlated factors — such as neighborhood amenities, healthcare accessibility, and cultural preferences — substantially influence residential patterns of elderly populations.

Several policy implications emerge from these findings. First, alcohol control policies remain a cornerstone of road safety strategy, with panel data models providing more reliable effect estimates than cross-sectional analysis. Second, the spatial clustering of elderly population in central Coimbra, coupled with the co-location of accessible housing, suggests positive feedback loops between housing stock characteristics and demographic composition. Urban planners should consider these spatial interdependencies when designing age-friendly neighborhoods and transport services.

Future research directions include extending the panel analysis to more recent periods with additional policy variables, and employing multi-level modeling approaches that can simultaneously capture temporal and spatial dynamics. The integration of panel data and spatial econometric methods within a unified framework represents a particularly promising avenue for advancing our understanding of transport-related social outcomes.

REFERENCES

Anselin, L. (1988). Spatial Econometrics: Methods and Models. Kluwer Academic Publishers, Dordrecht.

Anselin, L. (1995). Local indicators of spatial association — LISA. Geographical Analysis, 27(2), 93-115.

Baltagi, B.H. (2013). Econometric Analysis of Panel Data (5th ed.). John Wiley & Sons, Chichester.

Borsdorf, A. and Zemp, E. (2020). Spatial analysis of housing accessibility for elderly populations. Urban Studies, 57(8), 1643-1662.

Breusch, T.S. and Pagan, A.R. (1980). The Lagrange multiplier test and its applications to model specification in econometrics. Review of Economic Studies, 47(1), 239-253.

Ewing, R. and Cervero, R. (2010). Travel and the built environment: A meta-analysis. Journal of the American Planning Association, 76(3), 265-294.

Hausman, J.A. (1978). Specification tests in econometrics. Econometrica, 46(6), 1251-1271.

Logan, J.R., Xu, Z. and Stults, B. (2020). Interpolating U.S. decennial census tract data from as early as 1970 to 2010. Demography, 51(4), 1429-1454.

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

Noland, R.B. (2003). Traffic fatalities and injuries: The effect of changes in infrastructure and other trends. Accident Analysis & Prevention, 35(4), 599-611.

World Health Organization (2023). Global Status Report on Road Safety 2023. WHO, Geneva.