Prof. Dr. Reinhold Kosfeld
University of Kassel

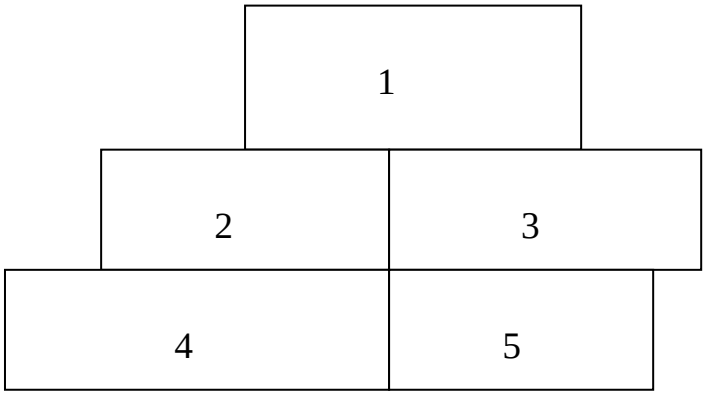
Spatial Econometrics

Assignments with Solutions

Part I – Assignments 1- 10

General information on a system of regions and data

Five regions are arranged in the following form:



Distances between regional centres (in km):

Region	1	2	3	4	5
1		8	10	20	22
2			16	12	14
3				20	12
4					10

Data on geo-referenced variables:

Region	u	v
1	6	3
2	8	3
3	8	2
4	11	1
5	12	1

u: Unemployment rate
v: Vacancy rate

Beveridge curve:
Relation between the unemployment rate (u) and vacancy rate (v)

Assignment 1 (Connectivity)

The arrangement of five regions is given along with the distances between regional centres (see “General information on a system of regions and data”).

a) Specify the contiguity matrix **W*** and the corresponding standardized weight matrix **W** for the given irregular arrangement!

- Contiguity matrix **W*** (Neighbourhood matrix):

$$\mathbf{W}^* = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix}$$

- Standardized weight matrix **W**:

Row-standardization: $w_{ij} = w^*_{ij} / \sum_{j=1}^{n} w^*_{ij}$

for example: 2nd element of 1st row (w_{12})

$$w_{12} = w^*_{12} / \sum_{j=1}^{n} w^*_{1j} = 1 / (0+1+1+0+0) = 1/2 = 0.5$$

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

b) Which spatial weight matrix **W*** is implied by the gravity model of spatial interaction?

Gravity of spatial interaction:

Use of power function $w_{ij}^{*} = d_{ij}^{-\alpha}$ with $\alpha = 2$

$$\mathbf{W}^{*} = \begin{bmatrix} 0 & 8^{-2} & 10^{-2} & 20^{-2} & 22^{-2} \\ 8^{-2} & 0 & 16^{-2} & 12^{-2} & 14^{-2} \\ 10^{-2} & 16^{-2} & 0 & 20^{-2} & 12^{-2} \\ 20^{-2} & 12^{-2} & 20^{-2} & 0 & 10^{-2} \\ 22^{-2} & 14^{-2} & 12^{-2} & 10^{-2} & 0 \end{bmatrix}$$

$$= \begin{bmatrix} 0 & 0.0156 & 0.0100 & 0.0025 & 0.0021 \\ 0.0156 & 0 & 0.0039 & 0.0069 & 0.0051 \\ 0.0100 & 0.0039 & 0 & 0.0025 & 0.0069 \\ 0.0025 & 0.0069 & 0.0025 & 0 & 0.0100 \\ 0.0021 & 0.0051 & 0.0069 & 0.0100 & 0 \end{bmatrix}$$

c) Determine the spatial weight matrix **W*** on the basis of the negative exponential function with ß calculated at $\gamma = 0.5$!

Average distance between regional centres:

$$\bar{d} = \frac{2}{n(n-1)} \sum_{i=1}^{n} \sum_{j>i} d_{ij} = \frac{2}{5 \cdot 4} \sum_{i=1}^{5} \sum_{j>i} d_{ij}$$

$$= \frac{1}{10}(8 + 10 + 20 + 22 + 16 + 12 + 14 + 20 + 12 + 10) = \frac{1}{10} 144 = 14.4$$

Distance decay parameter ß for $\gamma = 0.5$:

$$\beta = -\frac{\ln (1-\gamma)}{\bar{d}} = -\frac{\ln (1-0.5)}{\bar{d}} = -\frac{\ln 0.5}{14.4} = -\frac{-0.693}{14.4} = 0.048$$

Negative exponential function:

$$w_{ij}^{*} = e^{-\beta \cdot d_{ij}} = e^{-0.048 \cdot d_{ij}}$$

for example: 2nd element of 1st row (w^*_{12})

$$w_{12}^{*} = e^{-0.048 \cdot d12} = e^{-0.048 \cdot 8} = e^{-0.384} = 0.6811$$

$$W^* = \begin{bmatrix} 0 & 0.6811 & 0.6188 & 0.3829 & 0.3478 \\ 0.6811 & 0 & 0.4639 & 0.5621 & 0.5107 \\ 0.6188 & 0.4639 & 0 & 0.3829 & 0.5621 \\ 0.3829 & 0.5621 & 0.3829 & 0 & 0.6188 \\ 0.3478 & 0.5107 & 0.5621 & 0.6188 & 0 \end{bmatrix}$$

d) Compare the standardized weight matrices arsing from the distance measures used in parts b) and c)!

Row-standardization: $w_{ij} = w^*_{ij} / \sum_{j=1}^{n} w^*_{ij}$

● Standardized weight matrix based on the power function (part b))

e.g. $w_{12} = w^*_{12} / \sum_{j=1}^{5} w^*_{1j}$ =0.0156 / (0 + 0.0156 + 0.0100 + 0.0025 + 0.0021)

= 0.0156 / 0.0302 = 0.5166

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5166 & 0.3311 & 0.0828 & 0.0695 \\ 0.4952 & 0 & 0.1238 & 0.2190 & 0.1619 \\ 0.4292 & 0.1674 & 0 & 0.1073 & 0.2961 \\ 0.1142 & 0.3151 & 0.1142 & 0 & 0.4566 \\ 0.0871 & 0.2116 & 0.2863 & 0.4149 & 0 \end{bmatrix}$$

● Standardized weight matrix based on the negative exponential function (part c))

e.g. $w_{12} = w^*_{12} / \sum_{j=1}^{5} w^*_{1j}$ = 0.6811 / (0 + 0.6811 + 0.6188 + 0.3829 + 0.3478)

= 0.6811 / 2.0306 = 0.3354

$$\mathbf{W} = \begin{bmatrix} 0 & 0.3354 & 0.3047 & 0.1886 & 0.1713 \\ 0.3071 & 0 & 0.2092 & 0.2534 & 0.2303 \\ 0.3052 & 0.2288 & 0 & 0.1888 & 0.2772 \\ 0.1967 & 0.2887 & 0.1967 & 0 & 0.3179 \\ 0.1705 & 0.2504 & 0.2756 & 0.3034 & 0 \end{bmatrix}$$

<u>Comparison:</u>
The spatial weights calculated with the negative exponential function are much more alike compared to those calculated with the power function. In general, nearer (remote) regions have lower (larger) weights with the negative exponential function than with the power function.

Assignment 2 (Spatial Lag)

The data on the vacancy rate of five regions

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

are given along with the unstandardized and standardized contiguity matrices:

$$\mathbf{W}^* = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

a) Determine the first-order spatial lag of the vacancy rate on the basis of the

- unstandardized contiguity matrix **W***,
- standardized contiguity matrix **W**!

How can the spatial lags be interpreted?

● First-order spatial lag on the basis of the unstandardized contiguity matrix:

$$\mathrm{L} \cdot \mathbf{v} = \mathbf{W}^* \cdot \mathbf{v} = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0\cdot 3+1\cdot 3+1\cdot 2+0\cdot 1+0\cdot 1 \\ 1\cdot 3+0\cdot 3+1\cdot 2+1\cdot 1+0\cdot 1 \\ 1\cdot 3+1\cdot 3+0\cdot 2+0\cdot 1+1\cdot 1 \\ 0\cdot 3+1\cdot 3+0\cdot 2+0\cdot 1+1\cdot 1 \\ 0\cdot 3+0\cdot 3+1\cdot 2+1\cdot 1+0\cdot 1 \end{bmatrix} = \begin{bmatrix} 5 \\ 6 \\ 7 \\ 4 \\ 3 \end{bmatrix}$$

The interpretation of the spatial lag on the basis of the unstandardized contiguity matrix is not straightforward as it renders for each region the sums of the vacancy rates in the surrounding regions.

● First-order spatial lag on the basis of the standardized contiguity matrix:

$$L \cdot \mathbf{v} = \mathbf{W} \cdot \mathbf{v}$$

$$= \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \cdot 3 + 0.5 \cdot 2 \\ (1/3) \cdot 3 + (1/3) \cdot 2 + (1/3) \cdot 1 \\ (1/3) \cdot 3 + (1/3) \cdot 3 + (1/3) \cdot 1 \\ 0.5 \cdot 3 + 0.5 \cdot 1 \\ 0.5 \cdot 2 + 0.5 \cdot 1 \end{bmatrix} = \begin{bmatrix} 2.5 \\ 2 \\ 2\,1/3 \\ 2 \\ 1.5 \end{bmatrix}$$

The first-order spatial lag $L \cdot \mathbf{v}$ renders the mean values of the vacancy rates in the immediate neighbouring regions.

b) What is the second-order spatial lag of the vacancy rate based on the standardized contiguity matrix?

Second-order spatial lag: $L^2 \cdot \mathbf{v} = \mathbf{W}_2 \cdot \mathbf{v}$

Second-order unstandardized neighbourhood matrix:

$$\mathbf{W}_2^* = \begin{bmatrix} 0 & 0 & 0 & 1 & 1 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \end{bmatrix}$$

Second-order standardized neighbourhood matrix:

$$\mathbf{W}_2 = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{bmatrix}$$

Second-order spatial lag of the vacancy rate:

$$L^2 \cdot \mathbf{v} = \mathbf{W}_2 \cdot \mathbf{v} = \begin{bmatrix} 0 & 0 & 0 & 0.5 & 0.5 \\ 0 & 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0.5 & 0 & 0.5 & 0 & 0 \\ 0.5 & 0.5 & 0 & 0 & 0 \end{bmatrix} \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} 0.5 \cdot 1 + 0.5 \cdot 1 \\ 1 \cdot 1 \\ 1 \cdot 1 \\ 0.5 \cdot 3 + 0.5 \cdot 2 \\ 0.5 \cdot 3 + 0.5 \cdot 3 \end{bmatrix} = \begin{bmatrix} 1 \\ 1 \\ 1 \\ 2.5 \\ 3 \end{bmatrix}$$

The second-order spatial lag $L^2 \cdot \mathbf{v}$ renders the average vacancy rate in the second-order neighbouring regions i.e. in the regions around the immediate neighbours.

Assignment 3 (Moran coefficient I)

Given are the data on the unemployment and vacancy rate of five regions

Region	1	2	3	4	5
Unemploy-ment rate (u)	6	8	8	11	12
Vacancy rate (v)	3	3	2	1	1

and the unstandardized and standardized contiguity matrices:

$$\mathbf{W}^* = \begin{bmatrix} 0 & 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 1 & 0 \\ 1 & 1 & 0 & 0 & 1 \\ 0 & 1 & 0 & 0 & 1 \\ 0 & 0 & 1 & 1 & 0 \end{bmatrix} \text{ and } \mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

a) Compute Moran's I for the unemployment rate on the basis of the

- unstandardized contiguity matrix **W***,
- standardized contiuity matrix **W**

with the sum formula and interpret it!

● Moran's I on the basis of the unstandardized contiguity matrix:

$$I = \frac{n}{S_0} \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij}^* \cdot (x_i - \bar{x}) \cdot (x_j - \bar{x})}{\sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Arithmetic mean of unemployment rate:

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}(6+8+8+11+12) = \frac{1}{5}45 = 9$$

Table: Cross-products $(x_i - \bar{x})(x_j - \bar{x})$

Region	1	2	3	4	5
1	$(6-9)^2=9$	(6-9)(8-9)=3	(6-9)(8-9)=3	(6-9)(11-9)=-6	(6-9)(12-9)=-9
2	(8-9)(6-9)=3	$(8-9)^2=1$	(8-9)(8-9)=1	(8-9)(11-9)=-2	(8-9)(12-9)=-3
3	(8-9)(6-9)=3	(8-9)(8-9)=1	$(8-9)^2=1$	(8-9)(11-9)=-2	(8-9)(12-9)=-3
4	(11-9)(6-9)=-6	(11-9)(8-9)=-2	(11-9)(8-9)=-2	$(11-9)^2=4$	(11-9)(12-9)=6
5	(12-9)(6-9)=-9	(12-9)(8-9)=-3	(12-9)(8-9)=-3	(12-9)(11-9)=6	$(12-9)^2=9$

Table: Weighted cross-products $w_{ij}^{*}(x_i - \bar{x})(x_j - \bar{x})$

Region	1	2	3	4	5
1	0·9 = 0	1·3 = 3	1·3 = 3	0·(-6) = 0	0·(-9) = 0
2	1·3 = 3	0·1 = 0	1·1 = 1	1·(-2) = -2	0·(-3) = 0
3	1·3 = 3	1·1 = 1	0·1 = 0	0·(-2) = 0	1·(-3) = -3
4	0·(-6) = 0	1·(-2) = -2	0·(-2) = 0	0·4 = 0	1·6 = 6
5	0·(-9) = 0	0·(-3) = 0	1·(-3) = -3	1·6 = 6	0·9 = 0
ΣΣ	Sum of weighted cross-products				16

Table: Sum of squared deviations $(x_i - \bar{x})^2$

Region i	x_i	$x_i - \bar{x}$	$(x_i - \bar{x})^2$
1	6	6 – 9 = -3	9
2	8	8 – 9 = -1	1
3	8	8 – 9 = -1	1
4	11	11 – 9 = 2	4
5	12	12 – 9 = 3	9
Σ	Sum of squared deviations		24

Moran's I (unstandardized contiguity matrix): n=5, $S_0 = 12$

$$I = \frac{5}{12} \cdot \frac{16}{24} = 0{,}2778$$

● Moran's I on the basis of the standardized contiguity matrix:

$$I = \frac{\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot (x_i - \bar{x}) \cdot (x_j - \bar{x})}{\sum_{i=1}^{n} (x_i - \bar{x})^2}$$

Denominator: Sum of squared deviations $(x_i - \bar{x}^2)$

$\sum_{i=1}^{n=5} (x_i - \bar{x}^2) = 24$ (see the unstandardized case above)

Cross-products $(x_i - \bar{x})(x_j - \bar{x})$ see the unstandardized case above

Table: Weighted cross-products

Region	1	2	3	4	5
1	0·9 = 0	(1/2)·3 = 1.5	(1/2)·3 = 1.5	0·(-6) = 0	0·(-9) = 0
2	(1/3)·3 = 1	0·1 = 0	(1/3)·1 = 1/3	(1/3)·(-2) = -2/3	0·(-3) = 0
3	(1/3)·3 = 1	(1/3)·1 = 1/3	0·1 = 0	0·(-2) = 0	(1/3)·(-3) = -1
4	0·(-6) = 0	(1/2)·(-2) = -1	0·(-2) = 0	0·4 = 0	(1/2)·6 = 3
5	0·(-9) = 0	0·(-3) = 0	(1/2)·(-3) = -1.5	(1/2)·6 = 3	0·9 = 0
ΣΣ	Sum of weighted cross-products				7.5

Moran's I (standardized contiguity matrix):

$$I = \frac{7.5}{24} = 0.3125$$

b) Compute Moran's I for the vacancy rate on the basis of the

- unstandardized contiguity matrix **W***,
- standardized contiuity matrix **W**

with the matrix formula and interpret it!

● Moran's I on the basis of the unstandardized contiguity matrix:

$$I = \frac{n}{S_0} \cdot \frac{(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{W}^* (\mathbf{x} - \bar{\mathbf{x}})}{(\mathbf{x} - \bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}})}$$

Attribute vector (vector of vacancy rates): **x** (= **v**) = [3 3 2 2 1 1]'

Arithmetic mean of vacancy rate: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}(3+3+2+1+1) = \frac{1}{5}10 = 2$

- Numerator (quadratic form):

$$\mathbf{x}_i - \bar{\mathbf{x}} = [1 \quad 1 \quad 0 \quad -1 \quad -1]'$$

$$(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{W}^* (\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix}1 & 1 & 0 & -1 & -1\end{bmatrix} \begin{bmatrix}0 & 1 & 1 & 0 & 0\\ 1 & 0 & 1 & 1 & 0\\ 1 & 1 & 0 & 0 & 1\\ 0 & 1 & 0 & 0 & 1\\ 0 & 0 & 1 & 1 & 0\end{bmatrix} \begin{bmatrix}1\\1\\0\\-1\\-1\end{bmatrix}$$

$$= \begin{bmatrix}1 & 1 & 0 & -1 & -1\end{bmatrix} \begin{bmatrix}1\\0\\1\\0\\-1\end{bmatrix} = 2$$

- Denominator (scalar product):

$$\mathbf{x}_i - \bar{\mathbf{x}} = [1 \quad 1 \quad 0 \quad -1 \quad -1]'$$

$$(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix}1 & 1 & 0 & -1 & -1\end{bmatrix} \begin{bmatrix}1\\1\\0\\-1\\-1\end{bmatrix} = 4$$

Moran's I (unstandardized contiguity matrix): n = 5, $S_0 = 12$

$$I = \frac{5}{12} \cdot \frac{2}{4} = 0.2083$$

● Moran's I on the basis of the standardized contiguity matrix:

$$I = \frac{(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{W} (\mathbf{x} - \bar{\mathbf{x}})}{(\mathbf{x} - \bar{\mathbf{x}})' (\mathbf{x} - \bar{\mathbf{x}})}$$

- Numerator (quadratic form):

$$(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{W} (\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0.5 \\ 0 \\ 1/3 \\ 0 \\ -0.5 \end{bmatrix} = 1 \cdot 0.5 + 1 \cdot 0 + 0 \cdot (1/3) + (-1) \cdot 0 + (-1) \cdot (-0.5) = 1$$

- Denominator (scalar product):

$$(\mathbf{x}_i - \bar{\mathbf{x}})' (\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix} = 4$$

Moran's I (standardized contiguity matrix):

$$I = \frac{1}{4} = 0.25$$

Assignment 4 (Moran coefficient II)

Given are the data on the vacancy rate of five regions:

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

a) Compute Moran's I for the vacancy rate on the basis of the unstandardized distance-based weights matrix **W***,

$$\mathbf{W}^* = \begin{bmatrix} 0 & 0.0156 & 0.0100 & 0.0025 & 0.0021 \\ 0.0156 & 0 & 0.0039 & 0.0069 & 0.0051 \\ 0.0100 & 0.0039 & 0 & 0.0025 & 0.0069 \\ 0.0025 & 0.0069 & 0.0025 & 0 & 0.0100 \\ 0.0021 & 0.0051 & 0.0069 & 0.0100 & 0 \end{bmatrix},$$

with the matrix formula!

● Moran's I on the basis of the unstandardized distance matrix:

$$I = \frac{n}{S_0} \cdot \frac{(\mathbf{x} - \bar{\mathbf{x}})' \mathbf{W}^* (\mathbf{x} - \bar{\mathbf{x}})}{(\mathbf{x} - \bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}})}$$

Attribute vector (vector of vacancy rates): **x** (= **v**) = [3 3 2 1 1]'

Arithmetic mean of vacancy rate: $\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}(3+3+2+1+1) = \frac{1}{5}10 = 2$

- Numerator (quadratic form):

$$\mathbf{x}_i - \bar{\mathbf{x}} = [1 \quad 1 \quad 0 \quad -1 \quad -1]'$$

$$(\mathbf{x}_i - \bar{\mathbf{x}})' \mathbf{W}^* (\mathbf{x}_i - \bar{\mathbf{x}}) = [1 \quad 1 \quad 0 \quad -1 \quad -1] \begin{bmatrix} 0 & 0.0156 & 0.0100 & 0.0025 & 0.0021 \\ 0.0156 & 0 & 0.0039 & 0.0069 & 0.0051 \\ 0.0100 & 0.0039 & 0 & 0.0025 & 0.0069 \\ 0.0025 & 0.0069 & 0.0025 & 0 & 0.0100 \\ 0.0021 & 0.0051 & 0.0069 & 0.0100 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0.0110 \\ 0.0036 \\ 0.0045 \\ -0.0006 \\ -0.0028 \end{bmatrix} = 0.0180$$

- Denominator (scalar product):

$$\mathbf{x}_i - \bar{\mathbf{x}} = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix}'$$

$$(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix} = 4$$

Moran's I (unstandardized distance matrix): n = 5, S_0 = 0.1310

$$I = \frac{5}{0.1310} \cdot \frac{0.0180}{4} = 0.1718$$

b) Compute Moran's I for the vacancy rate on the basis of the standardized distance-based weights matrix **W**,

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5166 & 0.3311 & 0.0828 & 0.0695 \\ 0.4952 & 0 & 0.1238 & 0.2190 & 0.1619 \\ 0.4292 & 0.1674 & 0 & 0.1073 & 0.2961 \\ 0.1142 & 0.3151 & 0.1142 & 0 & 0.4566 \\ 0.0871 & 0.2116 & 0.2863 & 0.4149 & 0 \end{bmatrix}$$

with the matrix formula!

● Moran's I on the basis of the standardized distance matrix:

$$I = \frac{(\mathbf{x} - \bar{\mathbf{x}})'\mathbf{W}(\mathbf{x} - \bar{\mathbf{x}})}{(\mathbf{x} - \bar{\mathbf{x}})'(\mathbf{x} - \bar{\mathbf{x}})}$$

- Numerator (quadratic form):

$$(\mathbf{x}_i - \bar{\mathbf{x}})'\mathbf{W}(\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0 & 0.5166 & 0.3311 & 0.0828 & 0.0695 \\ 0.4952 & 0 & 0.1238 & 0.2190 & 0.1619 \\ 0.4292 & 0.1674 & 0 & 0.1073 & 0.2961 \\ 0.1142 & 0.3151 & 0.1142 & 0 & 0.4566 \\ 0.0871 & 0.2116 & 0.2863 & 0.4149 & 0 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix}$$

$$= \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 0.3643 \\ 0.1143 \\ 0.1932 \\ -0.0273 \\ -0.1162 \end{bmatrix} = 0.6221$$

- Denominator (scalar product):

$$(\mathbf{x}_i - \bar{\mathbf{x}})'(\mathbf{x}_i - \bar{\mathbf{x}}) = \begin{bmatrix} 1 & 1 & 0 & -1 & -1 \end{bmatrix} \begin{bmatrix} 1 \\ 1 \\ 0 \\ -1 \\ -1 \end{bmatrix} = 4$$

Moran's I (standardized distance matrix):

$$I = \frac{0.6221}{4} = 0.1555$$

Assignment 5 (Moran's I test under normality)

Given are the data on the vacancy rate of five regions

Region	1	2	3	4	5
vacancy rate (v)	3	3	2	1	1

and the standardized contiguity matrix:

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

Is the Moran coefficient under normality of the vacancy rate significant at the 5% level (right-sided test)?

● Moran's I for vacancy rate (with standardized contiguity matrix): I = 0.25

Null hypothesis: H_0: No spatial autocorrelation (spatial randomness)
Alternative hypothesis: Spatial autocorrelation (spatial dependence)

Test statistic: $$Z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} \overset{a}{\sim} N(0,1)$$

Expected value: $$E(I) = -\frac{1}{n-1} = -\frac{1}{5-1} = -0.25$$

Variance (for normal approx.): $$Var(I) = \frac{n^2 \cdot S_1 - n \cdot S_2 + 3 \cdot S_0^2}{(n^2 - 1) \cdot S_0^2} - [E(I)]^2$$

$$S_0 = \sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} = 5$$

$$S_1 = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n}(w_{ij}+w_{ji})^2$$

$$= [(w_{12}+w_{21})^2 + (w_{13}+w_{31})^2 + (w_{21}+w_{12})^2 + (w_{23}+w_{32})^2$$
$$+ (w_{24}+w_{42})^2 + (w_{31}+w_{13})^2 + (w_{32}+w_{23})^2 + (w_{35}+w_{53})^2$$
$$+ (w_{42}+w_{24})^2 + (w_{45}+w_{54})^2 + (w_{53}+w_{35})^2 + (w_{54}+w_{45})^2$$
$$= (1/2+1/3)^2 + (1/2+1/3)^2 + (1/3+1/2)^2 + (1/3+1/3)^2$$
$$+ (1/3+1/2)^2 + (1/3+1/2)^2 + (1/3+1/3)^2 + (1/3+1/2)^2$$
$$+ (1/2+1/3)^2 + (1/2+1/2)^2 + (1/2+1/3)^2 + (1/2+1/2)^2]/2$$
$$= [\frac{25}{36}+\frac{25}{36}+\frac{25}{36}+\frac{4}{9}+\frac{25}{36}+\frac{25}{36}+\frac{4}{9}+\frac{25}{36}+\frac{25}{36}+1+\frac{25}{36}+1]/2 = \frac{304}{36\cdot 2} = 4.2222$$

$$S_2 = \sum_{i=1}^{n}\left(\sum_{j=1}^{n} w_{ij} + \sum_{j=1}^{n} w_{ji}\right)^2 = \sum_{i=1}^{n}(w_{i\bullet}+w_{\bullet i})^2$$

$$= (w_{1\bullet}+w_{\bullet 1})^2 + (w_{2\bullet}+w_{\bullet 2})^2 + (w_{3\bullet}+w_{\bullet 3})^2 + (w_{4\bullet}+w_{\bullet 4})^2$$
$$+ (w_{5\bullet}+w_{\bullet 5})^2 = (1+2/3)^2 + (1+4/3)^2 + (1+4/3)^2 + (1+5/6)^2$$
$$+ (1+5/6)^2 = \frac{25}{9}+\frac{49}{9}+\frac{49}{9}+\frac{121}{36}+\frac{121}{36} = \frac{734}{36} = \frac{367}{18} = 20.3889$$

$$\mathrm{Var}(I) = \frac{n^2\cdot S_1 - n\cdot S_2 + 3\cdot S_0^2}{(n^2-1)\cdot S_0^2} - [E(I)]^2$$

$$= \frac{5^2\cdot 4.2222 - 5\cdot 20.3889 + 3\cdot 5^2}{(5^2-1)\cdot 5^2} - \left(-\frac{1}{4}\right)^2 = \frac{78.6105}{600} - \frac{1}{16}$$
$$= 0{,}1310 - 0{,}0625 = 0.0685$$

Test statistic (z-score): $z(I) = \frac{I - E(I)}{\sqrt{Var(I)}} = \frac{0.25 - (-0.25)}{\sqrt{0.0685}} = \frac{0.5}{0.2617} = 1.9106$

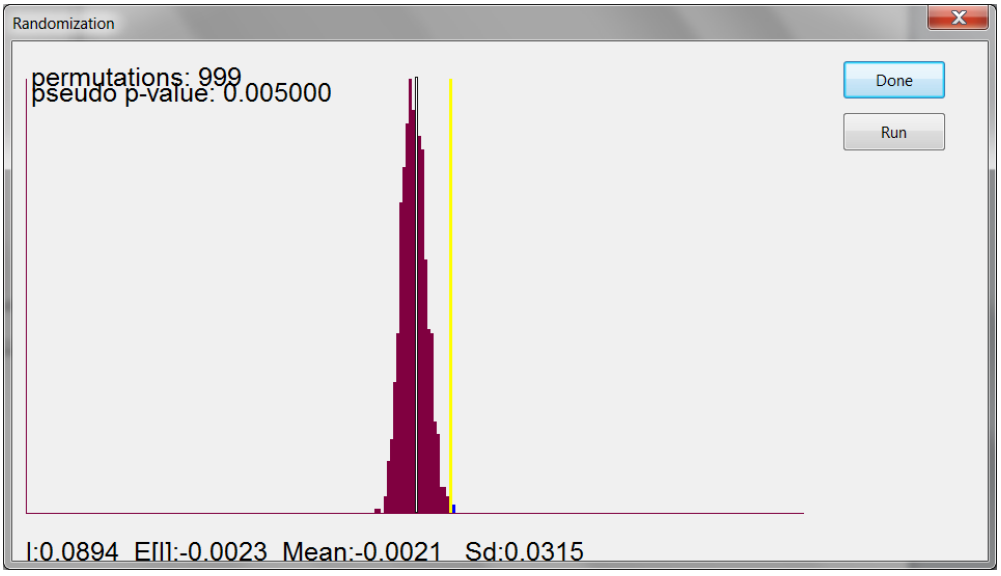
Critical value (α=0,05, right-sided test): $z_{0.95}$ = 1,6449

Testing decision:
z(I) = 1.9106 > $z_{0.95}$ = 1,6449 => Reject H_0

Interpretation:
Significant positive spatial autocorrelation of the vacancy rate

Assignment 6 (Moran's I permutation test)

Using GeoDa, the Moran's I test under randomization for the share of employees in manufacture of electrical and optical instruments across German districts (n=439) yields the following results:



a) Explain the idea of the permutation test!

The permutation test is a numerical approach to testing for significance of a statistic. It is used in GeoDa for testing the significance of Moran's I in order to improve the results of the approximate normal test and gaining insight in its sampling distribution under spatial randomness.

In each replication the observed values of variable are randomly assigned to the regions. In this way random map patterns of the spatial distribution of a variable are obtained. For each random pattern, the Moran coefficient is computed. The observed value of Moran's I is compared to simulated sample distribution.

The observed Moran's I value has a low probability to stem from a spatial random distribution of the variable, if it is found in the tails of the sample distribution. In particular the null hypothesis of spatial randomness has to rejected, if the pseudo p-value of Moran's I is lower than the significance level set by the user.

With an increasing number of replications the approximation of the generated sample distribution sample is improved. The choice of final digits of 99 for the number of permutations in GeoDa is motivated by the computation of the pseudo p-value (see part c)).

b) Interprete the statistics I, Mean and Sd and the parameter E[I]!

• I = 0.0894 and E[I] = -0. 0023

I is the Moran statistic for the share of employees in the manufacture of electrical and optical instruments. E[I] is the expected value of Moran's I for the study area i.e. the 439 German districts:

$$E(I) = -\frac{1}{n-1} = -\frac{1}{438} = -0.0023.$$

Because of

$$I = 0.0894 > E(I) = -0.0023,$$

Moran's I points to possible positive spatial autocorrelation. However, it can only be concluded from a significance test whether the measured spatial dependence is a characteristic feature of the variable in the population or due to sampling errors.

• Mean = -0.0021 and E[I] = -0. 0023

The "Mean" gives the average of Moran's I for the simulated distributions. This estimator for the expected value E(I) is the mean of 999 I values values from the randomized samples. that is a parameter of the sample distribution of Moran's I. The difference

$$(\text{Mean} = -0.0021) - (E(I) = -0.0023 = 0.0002$$

is attributed to the sample error.

• Std = 0.0315

The statistic Std is the standard error of Moran's I computed from its simulated distribution.

c) Test the null hypothesis of spatial randomness at a significance level of 5%! How can the pseudo p-value be interpreted?

The null hypothesis of spatial randomness can be tested by comparing the nominal significance level α of 0.05 with the pseudo p-value of 0.005:

pseudo p-value = 0.005 $<$ α = 0.05 $\Rightarrow$ H_0 rejected.

According to this testing outcome, the alternative hypothesis

H_1: positive spatial dependence (for one-sided test)

or

H_1: spatial dependence (for two-sided test)

is accepted.

In the case of the first alternative hypothesis, the pseudo p-value reflects the fraction of randomized samples where the Moran's I statistic is larger or equal than the observed value for the study region:

$$\text{pseudo p} - \text{value} = \frac{M+1}{S+1}.$$

M: # permutations with I $\geq$ observed I, S: # replications

On this account, the pseudo p-value can be interpreted as the approximate actual significance level of the test.

Assignment 7 (Moran scatterplot)

Given are the data on the vacancy rate of five regions,

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

and the standardized spatial weight matrix

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}.$$

a) Determine the regression line of Lv on v!

Attribute vector: $\mathbf{x} = \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix}$

$$\mathrm{L}\mathbf{x} = \mathbf{W}\mathbf{x} = \begin{bmatrix} 0 & 1/2 & 1/2 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 1/2 & 0 & 0 & 1/2 \\ 0 & 0 & 1/2 & 1/2 & 0 \end{bmatrix} \cdot \begin{bmatrix} 3 \\ 3 \\ 2 \\ 1 \\ 1 \end{bmatrix} = \begin{bmatrix} (3+2)/2 \\ (3+2+1)/3 \\ (3+3+1)/3 \\ (3+1)/2 \\ (2+1)/2 \end{bmatrix} = \begin{bmatrix} 2\frac{1}{2} \\ 2 \\ 2\frac{1}{3} \\ 2 \\ 1\frac{1}{2} \end{bmatrix}$$

● Regression: L**x** on **x**

Regression line: $\hat{Lx_i} = a + b \cdot x_i$

Ordinary least squares (OLS) estimators for a and b:

i	x_i	Lx_i	$x_i \cdot Lxi$	x_i^2
1	3	2 1/2	7 1/2	9
2	3	2	6	9
3	2	2 1/3	4 2/3	4
4	1	2	2	1
5	1	1 1/2	1 1/2	1
Σ	10	10 1/3	21 2/3	24

$$\hat{b} = \frac{n\sum x_i \cdot Lx_i - \sum x_i \sum Lx_i}{n\sum x_i^2 - (\sum x_i)^2} = \frac{5 \cdot 21\frac{2}{3} - 10 \cdot 10\frac{1}{3}}{5 \cdot 24 - 10^2} = \frac{108\frac{1}{3} - 103\frac{1}{3}}{120 - 100} = \frac{5}{20} = \frac{1}{4} = 0.25$$

(slope = Moran's I)

$$\hat{a} = \frac{\sum Lx_i}{n} - \hat{b} \cdot \frac{\sum x_i}{n} = \frac{10\frac{1}{3}}{5} - 0.25 \cdot \frac{10}{5} = 2.067 - 0.5 = 1.567$$

b) Draw the Moran scatterplot for the vacancy rate!

● Regression values $\hat{Lx}_i$

$\hat{Lx}_1 = 1{,}567 + 0{,}25 \cdot 3 = 2{,}317$ $\quad$ $\hat{Lx}_2 = 1{,}567 + 0{,}25 \cdot 3 = 2{,}317$

$\hat{Lx}_3 = 1{,}567 + 0{,}25 \cdot 2 = 2{,}067$ $\quad$ $\hat{Lx}_4 = 1{,}567 + 0{,}25 \cdot 1 = 1{,}817$

$\hat{Lx}_5 = 1{,}567 + 0{,}25 \cdot 1 = 1{,}817$

● Standardization of the vacancy rate and form its lagged values:

Sum of x-values and sum of squared x-values see part a)

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \frac{1}{5} \cdot 10 = 2$$

$$s_x^2 = \frac{1}{n}\sum_{i=1}^{n} x_i^2 - \bar{x}^2 = \frac{1}{5} \cdot 24 - 2^2 = 4.8 - 4 = 0.8 \;\rightarrow\; s_x = \sqrt{0.8} = 0.894$$

Standardized values (z- values) of the x-values:

$$z_{xi} = \frac{x_i - \bar{x}}{s_x}$$

Hint: For standardization we calculate the variance of X with the factor 1/n, while GeoDa uses the factor 1/(n-1).

z_{xi}	Lagged values of z_{xi}: Lz_{xi}
z_{x1} = (3-2)/0.894 = 1.119	Lz_{x1} =0.5·1.119 + 0.5·0 = 0.560
z_{x2} = (3-2)/0.894 = 1.119	Lz_{x2} =1.119/3 + 0/3 – 1.119/3 = 0
z_{x3} = (2-2)/0.894 = 0	Lz_{x3} = 1.119/3 + 1.119/3 – 1.119/3 = 0.373
z_{x4} = (1-2)/0.894 = -1.119	Lz_{x4} = 0.5·1.119 – 0.5·0.119 = 0
z_{x5} = (1-2)/0.894 = -1.119	Lz_{x5} = 0.5·0 - 0.5·1.119 = -0.560

● Regression line for the standardized values (Moran scatterplot):

• Lowest z_{xi} value (regions 4 and 5):

$z_{x5} = z_{x5} = -1.119$

$\hat{Lx}_5 = \hat{Lx}_4 = 1.817$

Standardized $\hat{Lx}_5 = \hat{Lx}_4$ value:

$$\hat{z}_{Lx_4} = \frac{\hat{Lx}_4 - \bar{x}}{s_x} = \frac{1.817 - 2}{0.894} = \frac{-0.183}{0.894} = -0.205$$

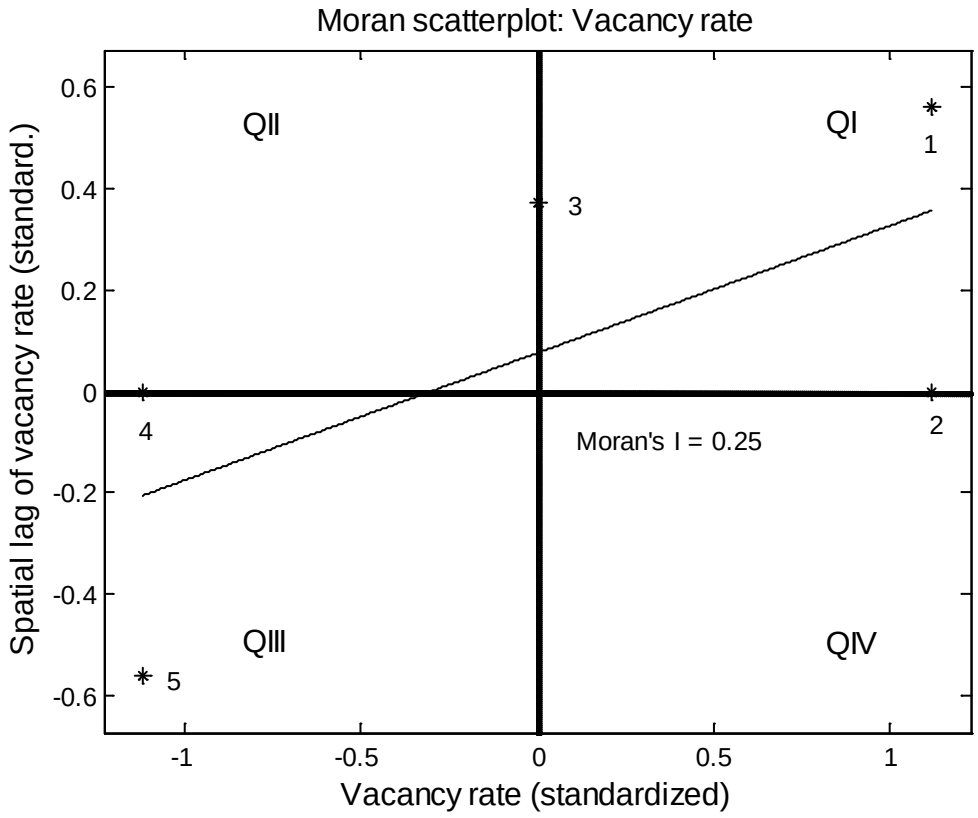
• Largest z_{xi} value (regions 1 and 2):

$z_{x1} = z_{x2} = 1.119$

$\hat{Lx}_1 = \hat{Lx}_2 = 2.317$

Standardized $\hat{Lx}_5 = \hat{Lx}_4$ value:

$$\hat{z}_{Lx_1} = \frac{\hat{Lx}_1 - \bar{x}}{s_x} = \frac{2.317 - 2}{0.894} = \frac{0.317}{0.894} = 0.355$$



c) Interpret the Moran scatterplot with respect to local characteristics!

Table: Types of local spatial association

		Spatially lagged vacancy rate (LX)	
		high	low
Vacancy rate (X)	high	Quadrant I: HH Region 1	Quadrant IV: HL -
	low	Quadrant II: LH -	Quadrant III: LL Regions 5

Region 2: Border between quadrants I and IV
Region 3: Border between quadrants I and II
Region 4: Border between quadrants II and III

Positive spatial association: Regions 1 and 5 (spatial clusters)
No spatial association: Region 2, 3 and 4
No spatial outliers

Assignment 8 (Regression outliers)

Identity outliers in the regression of Lv on v (see Assignment 7) on the basis of the

- 2s rule for the residuals,
- hat matrix and leverage points,
- Cook's distance!

● 2s rule for the residuals

OLS residuals e_i and squared residuals e_i^2 of the regression of **Wx** on **x**

i	x_i	Lx_i	$\widehat{Lx}_i$	e_i	e_i^2
1	3	2.5	2.3167	0.1833	0.0336
2	3	2	2.3167	-0.3167	0.1003
3	2	2.3333	2.0667	0.2666	0.0711
4	1	2	1.8167	0.1833	0.0336
5	1	1.5	1.8167	-0.3167	0.1003
Σ	10	10.3333	10.3335	-0.0002≈0	0.3389

e_i: OLS residual of the regression of **Wx** on **x**

s^2: Unbiased estimate of the error variance: $s^2 = \sum_{i=1}^{n} e_i^2 /(n-k)$

$$s^2 = \frac{1}{5-2}\sum_{i=1}^{n} e_i^2 = \frac{1}{3}\cdot 0.3389 = 0.1130$$

Standard deviation: $s = \sqrt{0.1130} = 0.3362$

2s interval: [-2·s; 2·s] = [-2·0.3362; 2·0.3362] = [-0.6724; 0.6724]

No residuals outside the 2s interval → no outliers in the OLS residuals

● Hat matrix and leverage points

Observation matrix:

$$\mathbf{X} = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \qquad \mathbf{X}'\mathbf{X} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 2 & 1 & 1 \end{bmatrix} \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} = \begin{bmatrix} 5 & 10 \\ 10 & 24 \end{bmatrix}$$

$$(\mathbf{X}'\mathbf{X})^{-1} = \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix}$$

Hat matrix:

$$\mathbf{H} = \mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}' = \begin{bmatrix} 1 & 3 \\ 1 & 3 \\ 1 & 2 \\ 1 & 1 \\ 1 & 1 \end{bmatrix} \begin{bmatrix} 1.2 & -0.5 \\ -0.5 & 0.25 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 2 & 1 & 1 \end{bmatrix}$$

$$= \begin{bmatrix} -0.3 & 0.25 \\ -0.3 & 0.25 \\ 0.2 & 0 \\ 0.7 & -0.25 \\ 0.7 & -0.25 \end{bmatrix} \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 3 & 3 & 3 & 3 & 3 \end{bmatrix} = \begin{bmatrix} 0.45 & 0.45 & 0.2 & -0.05 & -0.05 \\ 0.45 & 0.45 & 0.2 & -0.05 & -0.05 \\ 0.2 & 0.2 & 0.2 & 0.2 & 0.2 \\ -0.05 & -0.05 & 0.2 & 0.45 & 0.45 \\ -0.05 & -0.05 & 0.2 & 0.45 & 0.45 \end{bmatrix}$$

Diagonal elements of the „hat“ matrix **H** (cut-off value: 4/(n=5)=0,8):
h_1=0.45, h_2=0.45, h_3=0.2, h_4=0.45, h_5=0.45

Interpretation:
$h_i < 0.8$ for all i → no leverage points

● Cook's distance

$$D_i = \frac{e_i^2 \cdot h_i}{k \cdot s^2 \cdot (1 - h_i)^2}$$

k: number of explanatory variables (inc. intercept)

s^2: Unbiased estimate of the error variance $\quad s^2 = \sum_{i=1}^{n} e_i^2 / (n - k)$

Cut-off value: $D_i > 1$ (influential observation)

$$s^2 = \sum_{i=1}^{n} e_i^2 / (n - k) = 0.3389 / (5 - 2) = 0.1130$$

$$D_1 = \frac{e_1^2 \cdot h_1}{k \cdot s^2 \cdot (1 - h_1)^2} = \frac{0{,}0336 \cdot 0{,}45}{2 \cdot 0{,}1130 \cdot (1 - 0{,}45)^2} = \frac{0{,}01512}{0{,}068365} = 0{,}2212$$

$$D_2 = \frac{e_2^2 \cdot h_2}{k \cdot s^2 \cdot (1 - h_2)^2} = \frac{0.1003 \cdot 0{,}45}{2 \cdot 0.1130 \cdot (1 - 0{,}45)^2} = \frac{0.045135}{0.068365} = 0.6602$$

$$D_3 = \frac{e_3^2 \cdot h_3}{k \cdot s^2 \cdot (1 - h_3)^2} = \frac{0.0711 \cdot 0.2}{2 \cdot 0.1130 \cdot (1 - 0.2)^2} = \frac{0.01422}{0.14464} = 0.0983$$

$$D_4 = \frac{e_4^2 \cdot h_4}{k \cdot s^2 \cdot (1 - h_4)^2} = \frac{0{,}0336 \cdot 0.45}{2 \cdot 0{,}1130 \cdot (1 - 0.45)^2} = \frac{0{,}01512}{0{,}068365} = 0.2212$$

$$D_5 = \frac{e_5^2 \cdot h_5}{k \cdot s^2 \cdot (1 - h_5)^2} = \frac{0.1003 \cdot 0.45}{2 \cdot 0.1130 \cdot (1 - 0.45)^2} = \frac{0.045135}{0.068365} = 0.6602$$

Interpretation:
$D_i < 1$ for all i → no influential observations

Assignment 9 (Geary's c)

Compute Geary's c for the vacancy rate

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

on the basis of the standardized contiguity matrix

$$\mathbf{W} = \begin{bmatrix} 0 & 0.5 & 0.5 & 0 & 0 \\ 1/3 & 0 & 1/3 & 1/3 & 0 \\ 1/3 & 1/3 & 0 & 0 & 1/3 \\ 0 & 0.5 & 0 & 0 & 0.5 \\ 0 & 0 & 0.5 & 0.5 & 0 \end{bmatrix}$$

and interpret it!

Geary's c:
$$C = \frac{(n-1)\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} \cdot (x_i - x_j)^2}{2 \cdot n \cdot \sum_{i=1}^{n}(x_i - \bar{x})^2}$$

Table: Weighted squared differences $w_{ij} \cdot (x_i - x_j)^2$

Region	1	2	3	4	5
1	0	$(1/2)(3-3)^2 = 0$	$(1/2)(3-2)^2 = 1/2$	0	0
2	$(1/3)(3-3)^2 = 0$	0	$(1/3)(3-2)^2 = 1/3$	$(1/3)(3-1)^2 = 1\ 1/3$	0
3	$(1/3)(2-3)^2 = 1/3$	$(1/3)(2-3)^2 = 1/3$	0	0	$(1/3)(2-1)^2 = 1/3$
4	0	$(1/2)(1-3)^2 = 2$	0	0	$(1/2)(1-1)^2 = 0$
5	0	0	$(1/2)(1-2)^2 = 1/2$	$(1/2)\cdot(1-1)^2 = 0$	0
ΣΣ	Sum of weighted squared differences				5 2/3

$$\bar{x} = \frac{1}{5}\sum_{i=1}^{5} x_i = \frac{1}{5}\cdot(3+3+2+1+1) = \frac{1}{5}\cdot 10 = 2$$

$$\sum_{i=1}^{5}(x_i - \bar{x})^2 = \sum_{i=1}^{5}(x_i - 2)^2 = (3-2)^2 + (3-2)^2 + (2-2)^2 + (1-2)^2 + (1-2)^2 =$$
$$= 1+1+0+1+1 = 4$$

Geary‘s c (n=5): $C = \frac{(5-1)\cdot 5\frac{2}{3}}{2\cdot 5\cdot 4} = \frac{68/3}{40} = 0.5667$

Interpretation:
$0 \leq C=0.5667 < 1$ → positive spatial autocorrelation

Assignment 10 (Getis-Ord G)

The arrangement of five regions is given along with the distances between regional centres (see “General information on a system of regions and data”).

a) What value does Getis-Ord G statistic take for the vacancy rate

Region	1	2	3	4	5
Vacancy rate (v)	3	3	2	1	1

with a critical distance of $\bar{d}$?

Getis-Ord G statistic: $$G(d) = \frac{\sum_{i=1}^{n}\sum_{j\neq i} w_{ij}(d)\cdot x_i \cdot x_j}{\sum_{i=1}^{n}\sum_{j\neq i} x_i \cdot x_j}$$

Distance matrix **D**:

$$\mathbf{D} = \begin{bmatrix} 0 & 8 & 10 & 20 & 22 \\ 8 & 0 & 16 & 12 & 14 \\ 10 & 16 & 0 & 20 & 12 \\ 20 & 12 & 20 & 0 & 10 \\ 22 & 14 & 12 & 10 & 0 \end{bmatrix}$$

$$\bar{d} = \frac{2}{n(n-1)}\sum_{i=1}^{n}\sum_{j>i} d_{ij} = \frac{2}{5(5-1)}\sum_{i=1}^{5}\sum_{j>i} d_{ij}$$

$$= \frac{1}{10}(8+10+20+22+16+12+14+20+12+10) = \frac{144}{10} = 14.4$$

Spatial weight matrix **W**(d) corresponding with $d=\bar{d}=14.4$:

$$\mathbf{W}(d=14.4) = \begin{bmatrix} 1 & 1 & 1 & 0 & 0 \\ 1 & 1 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 & 1 \\ 0 & 1 & 0 & 1 & 1 \\ 0 & 1 & 1 & 1 & 1 \end{bmatrix}$$

Attribute vector **x**: $\mathbf{x} = [3 \;\; 3 \;\; 2 \;\; 1 \;\; 1]'$

Calculation of the denominator of G:

Region	1	2	3	4	5
1	-	3·3=9	3·2=6	3·1=3	3·1=3
2	3·3=9	-	3·2=6	3·1=3	3·1=3
3	2·3=6	2·3=6	-	2·1=2	2·1=2
4	1·3=3	1·3=3	1·2=2	-	1·1=1
5	1·3=3	1·3=3	1·2=2	1·1=1	-
ΣΣ j≠i	Sum of products $x_i \cdot x_j$, j≠i				76

Calculation of the numerator of G:

Region	1	2	3	4	5
1	-	3·3=9	3·2=6	0	0
2	3·3=9	-	0	3·1=3	3·1=3
3	2·3=6	0	-	0	2·1=2
4	0	1·3=3	0	-	1·1=1
5	0	1·3=3	1·2=2	1·1=1	-
ΣΣ j≠i	Sum of weighted products $w_{ij}(d=14.4) \cdot x_i \cdot x_j$, j≠i				48

Getis-Ord G statistic with $d = \bar{d} = 14.4$::

$$G(d = 14.4) = \frac{\sum_{i=1}^{n} \sum_{j \neq i} w_{ij}(d) \cdot x_i \cdot x_j}{\sum_{i=1}^{n} \sum_{j \neq i} x_i \cdot x_j} = \frac{48}{76} = 0.6316$$

b) Interpret the value of the G statistic for the vacancy rate!

Expected value of G(d):

$$E[G(d)] = \frac{W}{n(n-1)} \qquad \text{with} \qquad W = \sum_{i=1}^{n} \sum_{j \neq i} w_{ij}(d)$$

Calculation of W:

Region	1	2	3	4	5
1	-	1	1	0	0
2	1	-	0	1	1
3	1	0	-	0	1
4	0	1	0	-	1
5	0	1	1	1	-
$\Sigma\Sigma$ $j \neq i$	Sum of w_{ij}, $j \neq i$				12

$$E[G(d)] = \frac{W}{n(n-1)} = \frac{12}{5 \cdot (5-1)} = \frac{12}{20} = 0.6$$

If there is no overall concentration of high vacancy data, Getis-Ord G statistic is expected to take a value of 0.6. The expected value results when the vacancy rates are randomly distributed over space. The G value of 0.6316 with the critical distance d=14.4 for the given arrangement of regions indicates only a small tendency of concentration of the vacancy rate.