# Spatial Autocorrelation

*Stamatis Kalogirou*

*2015-04-15*

## Introduction

Moran's I is one of the oldest statistics used to examine spatial autocorrelation. This global statistic was first proposed by Moran (1948, 1950). Later, Cliff and Ord (1973, 1981) present a comprehensive work on spatial autocorrelation and suggested a formula to calculate the $I$ which is now used in most textbooks and software: $I = \frac{n}{W} \frac{\sum_{i=1}^{n} \sum_{j=1}^{n} w_{ij} z_i z_j}{\sum_{i=1}^{n} z_i^2}$ where $n$ is number of observations, $W$ is the sum of the weights $w_{ij}$ for all pairs in the system, $z_i = x_i - \bar{x}$ where $x$ is the value of the variable at location $i$ and $\bar{x}$ the mean value of the variable in question (Eq. 5.2 Kalogirou, 2003).The implementation here allows only nearest neighbour weighting schemes. Resampling and randomization null hypotheses have been tested following the discussion of Goodchild (1986, pp. 24-26).

## Exploring the data

The lctools package has some built in data to allow for practising the various spatial analysis techniques. One of the datasets is GR.Municipalities. The latter is a SpatialPolygonsDataFrame that refers to Municipalities of Greece. The descriptive data for each municipality include demographic and economic variables the source of which are the Population Census of 2001 and the General Secretariat for Information Systems, respectively.

```
library(lctools)
data(GR.Municipalities)
names(GR.Municipalities@data)
```

```
## [1] "OBJECTID"   "X"          "Y"          "Name"       "CodeELSTAT"
## [6] "PopM01"     "PopF01"     "PopTot01"   "UnemrM01"   "UnemrF01"
## [11] "UnemrT01"   "PrSect01"   "Foreig01"   "Income01"
```

To learn more about the above data set, try `help(GR.Municipalities)`. To calculate the global Moran's I statistic one can make use of the function `MoransI` with three arguments: the coordinates of the observations, the number of nearest neighbours and the variable for which the statistic will be calculated for. The coordinates refer to the geometric centro-ids of the municipalities in Greece, the number of nearest neighbours is set to 6 and the variable to be analysed refers to the mean annual recorded income in 2001 (in Euros).

```
Coords<-cbind(GR.Municipalities@data$X, GR.Municipalities@data$Y)
bw<-6
mI<-moransI(Coords,bw,GR.Municipalities@data$Income01)
moran.table<-matrix(data=NA,nrow=1,ncol=6)
colnames(moran.table) <- c("Moran's I", "Expected I", "Z resampling", "P-value resampling",
                    "Z randomization", "P-value randomization")
moran.table[1,1]<-mI$Morans.I
moran.table[1,2]<-mI$Expected.I
moran.table[1,3]<-mI$z.resampling
moran.table[1,4]<-mI$p.value.resampling
moran.table[1,5]<-mI$z.randomization
moran.table[1,6]<-mI$p.value.randomization
```

| Moran's I | Expected I | Z resampling | P-value resampling | Z randomization | P-value randomization |
|---|---|---|---|---|---|
| 0.6544124 | -0.0030864 | 22.27952 | 0 | 22.40946 | 0 |

The results suggest a strong positive spatial autocorrelation that is statistically significant using either the randomization or resampling hypotheses (Cliff and Ord, 1973; 1981; Goodchild, 1986). In order to examine the sensitivity of the above results, one could try different bandwidth sizes (i.e. number of nearest neighbours).

```
bw<-c(3, 4, 6, 9, 12, 18, 24)

moran<-matrix(data=NA,nrow=7,ncol=7)
colnames(moran) <- c("ID", "k", "Moran's I", "Z resampling", "P-value resampling",
                     "Z randomization", "P-value randomization")
counter<-1

for(b in bw){
    moranI<-moransI(Coords,b,GR.Municipalities@data$Income01)
    moran[counter,1]<-counter
    moran[counter,2]<-b
    moran[counter,3]<-moranI$Morans.I
    moran[counter,4]<-moranI$z.resampling
    moran[counter,5]<-moranI$p.value.resampling
    moran[counter,6]<-moranI$z.randomization
    moran[counter,7]<-moranI$p.value.randomization
    counter<-counter+1
}
```

The results are presented in the table below. The Moran's I for the mean annual recorded income is above 0.5 in all cases suggesting a strong positive spatial autocorrelation.

| ID | k | Moran's I | Z resampling | P-value resampling | Z randomization | P-value randomization |
|---|---|---|---|---|---|---|
| 1 | 3 | 0.6823 | 16.4559 | 0 | 16.5519 | 0 |
| 2 | 4 | 0.6797 | 18.8569 | 0 | 18.9669 | 0 |
| 3 | 6 | 0.6544 | 22.2795 | 0 | 22.4095 | 0 |
| 4 | 9 | 0.6422 | 26.7714 | 0 | 26.9274 | 0 |
| 5 | 12 | 0.6206 | 29.9977 | 0 | 30.1725 | 0 |
| 6 | 18 | 0.5802 | 34.8110 | 0 | 35.0136 | 0 |
| 7 | 24 | 0.5523 | 38.8086 | 0 | 39.0342 | 0 |

```
plot(moran[,2], moran[,3], main="Global Moran's I", sub="", xlab="# of neighbours",
     ylab="Moran's I")
```

# Global Moran's I
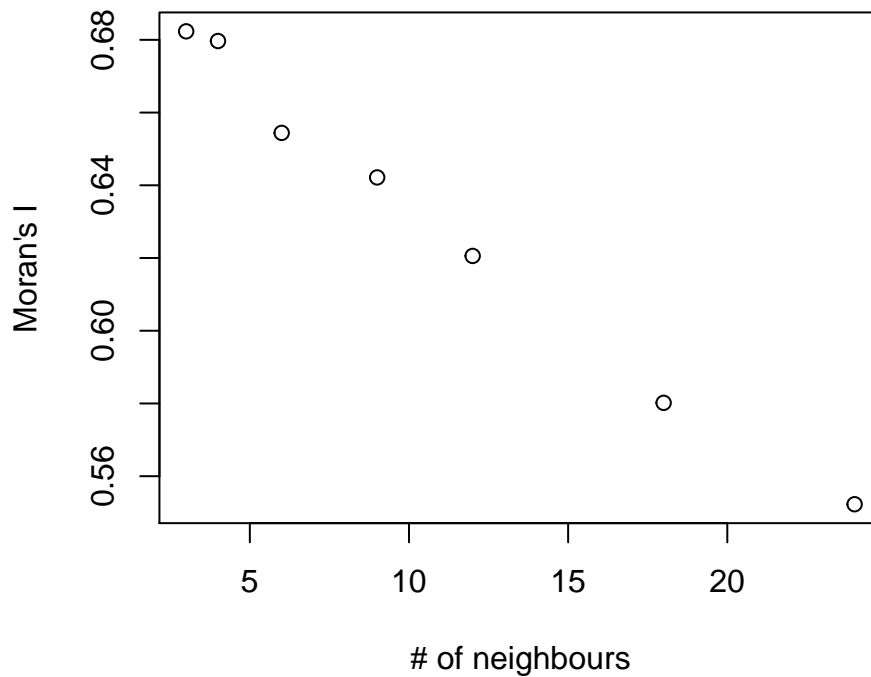


## Local Moran's I

The next step is the calculation of local Moran $I_i$ using the command `l.moransI` as follows:

```
l.moran<-l.moransI(Coords,6,GR.Municipalities@data$Income01)
```

The object `l.moran` generated includes a plethora of results that can be plotted or mapped. The code below creates the classic Moran's I scatter plot.

```
xmin<-round(ifelse(abs(min(l.moran[,7])) > abs(min(l.moran[,8])), abs(min(l.moran[,7])),
                  abs(min(l.moran[,8]))))
xmax<-round(ifelse(abs(max(l.moran[,7])) > abs(max(l.moran[,8])), abs(max(l.moran[,7])),
                  abs(max(l.moran[,8]))))
xmax<-ifelse(xmin>xmax,xmin,xmax)+1
ymax<-xmax
xmin<- -xmax
ymin<- -ymax
reg1 <- lm(l.moran[,8]~l.moran[,7])
```

```
plot(l.moran[,7], l.moran[,8], main="Moran's I scatterplot", sub="", xlab="Income",
     ylab="lagged Income", xlim=c(xmin, xmax), ylim=c(ymin, ymax))
abline(h=0)
abline(v=0)
abline(reg1, col="red")
```

## Moran's I scatterplot
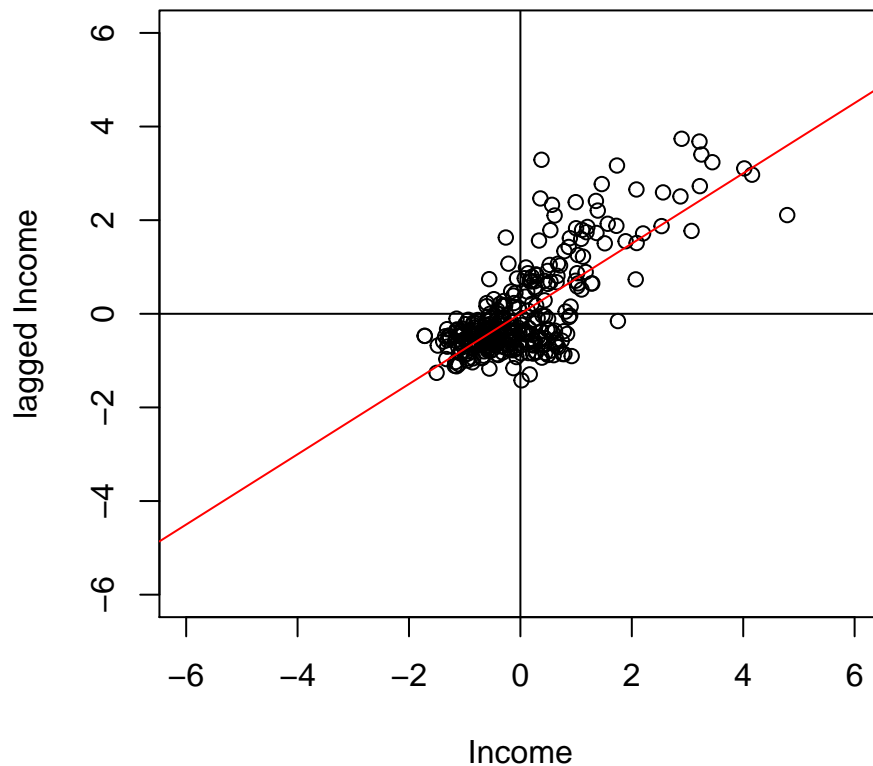
### References

Cliff, A.D., and Ord, J.K., 1973, Spatial autocorrelation (London: Pion).

Cliff, A.D., and Ord, J.K., 1981, Spatial processes: models and applications (London: Pion).

Goodchild, M. F., 1986, Spatial Autocorrelation. Catmog 47, Geo Books.

Kalogirou, S., 2003, The Statistical Analysis And Modelling Of Internal Migration Flows Within England And Wales, PhD Thesis, School of Geography, Politics and Sociology, University of Newcastle upon Tyne, UK. URL: https://theses.ncl.ac.uk/dspace/handle/10443/204

Moran, P.A.P., 1948, The interpretation of statistical maps, Journal of the Royal Statistics Society, Series B (Methodological), 10, 2, pp. 243 - 251.

Moran, P.A.P., 1950, Notes on continuous stochastic phenomena, Biometrika, 37, pp. 17 - 23.