

Approccio geo-statistico all'analisi territoriale: l'autocorrelazione spaziale

Francesco G. Truglia
truglia@istat.it

1.5. Struttura spaziale dei dati: contiguità

Prima legge della geografia di Tobler secondo la quale “tutto è correlato con tutto, ma le cose vicine sono più correlate delle cose lontane”

Problema del vicinato - Ordinamento delle unità

Serie storiche? *Tempo unidirezionali*. Dati ordinati in base alla freccia del tempo.

Serie temporali? *Spazio multidirezionale*.

Matrice di contiguità

-Scelta criterio vicinato: Torre, Regina, Distanze euclidee, blocchi, ecc.

-Indipendentemente dal criterio di vicinato i dati sono ricodificati in forma binaria C o standardizzata W.

$$C = \begin{bmatrix} 0 & c_{12} & \dots & c_{1n} \\ c_{21} & 0 & \dots & c_{2n} \\ \dots & \dots & \dots & \dots \\ c_{n1} & c_{n2} & \dots & 0 \end{bmatrix}$$

$c_{ij} = 1$ se le subaree o zone i e j hanno un lato o un vertice in comune

$c_{ij} = 0$ negli altri casi

$$w_{ij} = \frac{c_{ij}}{\sum_j c_{ij}}$$

$$\sum_{i=1}^n \sum_{j=1}^n c_{ij} = \sum_{i=1}^n \sum_{j=1}^n w_{ij} = n(n-1)$$

$$W = \begin{bmatrix} 0 & w_{12} & \dots & w_{1n} \\ w_{21} & 0 & \dots & w_{2n} \\ \dots & \dots & \dots & \dots \\ w_{n1} & w_{n2} & \dots & 0 \end{bmatrix}$$

$0 < w_{ij} \leq 1$ se le subaree o zone i e j hanno un lato o un vertice in comune

$w_{ij} = 0$ negli altri casi

Struttura spaziale dei dati: contiguità

#1° Ordine

```
V <- poly2nb(STR.ZU, row.names=STR.ZU$ZUR,queen=F)#statistiche vicinato
```

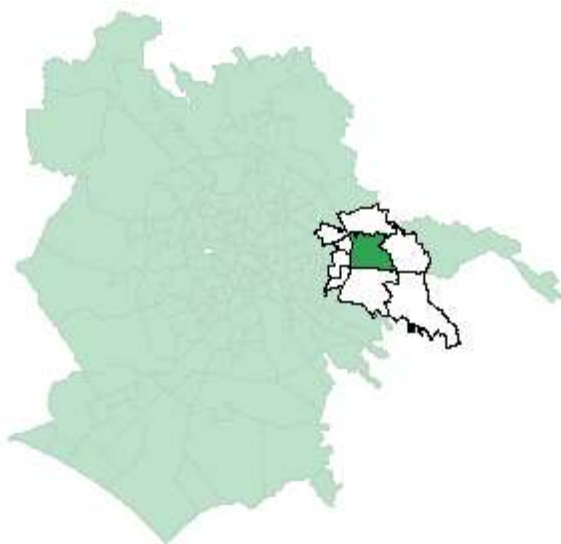
```
Wm <- nb2mat(V, style='W',zero.policy =T)#Matrice W STYLE = w-> STANDARD; B NOSTANDARD
```

```
W <- nb2listw(V, style='W',zero.policy =T)##Elenco vicini
```

#2° Ordine

```
lags2 <- nblag(V, 2)
```

```
W2 <- nb2listw(lags2[[2]], style="W", zero.policy=TRUE)
```



summary(W)

Neighbour list object:

Number of regions: 154

Number of nonzero links: 798

Percentage nonzero weights: 3.364817

Average number of links: 5.181818

Link number distribution:

1	2	3	4	5	6	7	8	9	10
1	5	15	28	42	40	13	4	3	3

1 least connected region:

12n with 1 link

3 most connected regions:

11x 12a 19a with 10 links

nb_w\$neighbours[1:3]

```
[[1]]  
[1] 2 6 132 140 142  
[[2]]  
[1] 1 3 5 132 152  
[[3]]  
[1] 2 4 5 8 11 19 152
```

nb_w\$weights[1:3]

```
[[1]]  
[1] 0.2 0.2 0.2 0.2 0.2  
[[2]]  
[1] 0.2 0.2 0.2 0.2 0.2  
[[3]]  
[1] 0.142 0.142 0.142 0.14 0.142 0.142 0.142
```

Weights style: W

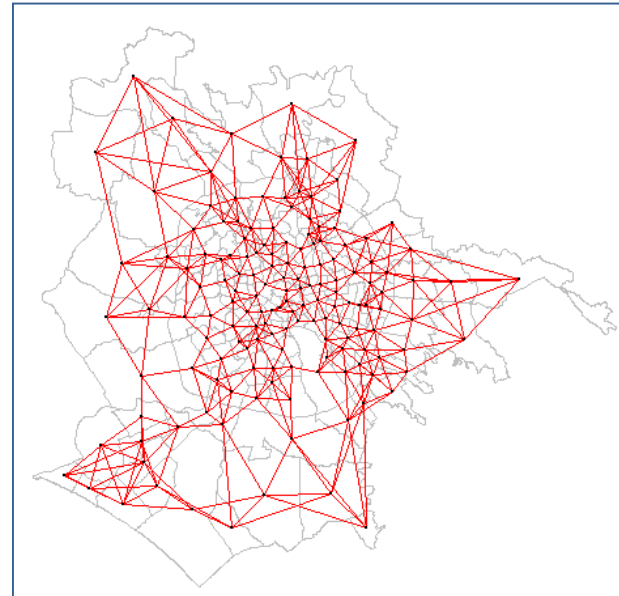
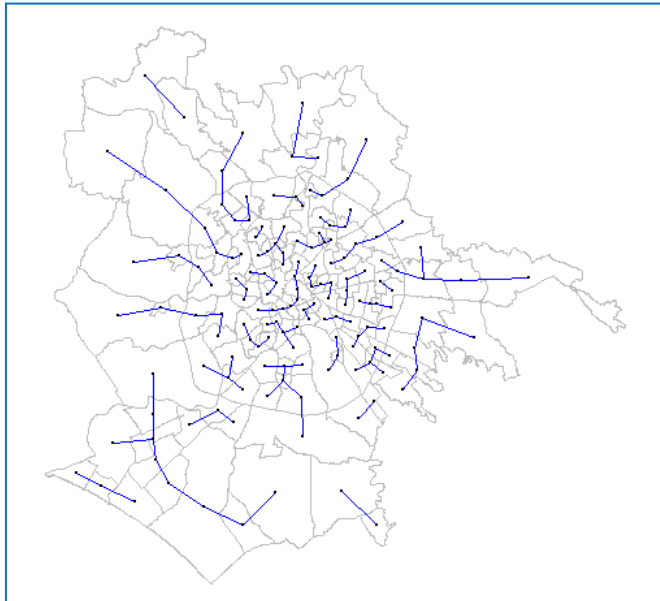
Weights constants summary:

	n	nn	S0	S1	S2
W	154	23716	154	63.36271	636.4623

Struttura spaziale dei dati: contiguità

```
IDs <- sort(STR.ZU$ZUR)
coord <- coordinates(STR.ZU)
plot(STR.ZU, border="grey")
W_knn1 <- knn2nb(knearneigh(coord, k=1), row.names=IDs)
W_knn1_mat <- nb2listw(W_knn1) # statistiche vicinato  $S = n^{\circ}$  vicini all'i-esimo punto
plot(W_knn1_mat, coords=coord, pch=19, cex=0.1, col="blue", add=T) # sovrappone centroidi - poligoni
title("k=1 (Baricentri)")
```

```
plot(STR.ZU, border="grey")
W_knn5 <- knn2nb(knearneigh(coord, k=5), row.names=IDs) ## i primi più vicini di ordine 1
W_knn5_mat <- nb2listw(W_knn5) # statistiche vicinato  $S = n^{\circ}$  vicini all'i-esimo punto
plot(W_knn5_mat, coords=coord, pch=19, cex=0.1, col="red", add=T) # sovrappone centroidi ai poligoni
title("k=5 (Baricentri)")
```



4. Struttura spaziale dei dati geo-statistici

- 4.1. Autocorrelazione spaziale: I di Moran
- 4.2. Relazione tra I-Moran e r-Pearson
- 4.3. Scomposizione I di Moran
- 4.4. Scomposizione territoriale: Local Moran
- 4.5. Scomposizione territoriale: Local Moran
- 4.6. I di Moran Bivariato
- 4.7. L di LEE globale e locale

3.1. Autocorrelazione spaziale: I di Moran

Statistiche per la misura dell'autocorrelazione spaziale

Moran (1948), Geary (1950), Getis ed Ord (1992), Whittle (1954).

Questi indici, tranne l'ultimo, sono disponibili sia per l'analisi dell'autocorrelazione globale che locale (*LISA- Local Indicator of Spatial Association*).

$$I = \frac{N}{S_0} \frac{z' Cz}{z' z} \quad I = \frac{z' Wz}{z' z} \quad I = \frac{N}{S_0} \frac{\sum_i^N \sum_j^N (x_i - M)(x_j - M) w_{ij}}{\sum_i^N (x_i - M)^2}$$

Dove:

- z è il vettore composto da N osservazioni della variabile x_i espressa in scarti dalla media ($z = x_i - M$);
- $\mathbf{C}$ o $\mathbf{W}$ matrice di contiguità binaria o standardizzata;
- S_0 è uno scalare :il numero delle unità contigue.

Autocorrelazione negativa $I < -1/N-1$

Autocorrelazione nulla $I = -1/N-1$

Autocorrelazione positiva $I > -1/N-1$

$$\min \lambda_W < I < \max \lambda_W$$

```
moran.range <- function(lw) {  
  wmat <- listw2mat(lw)  
  return(range(eigen((wmat+t(wmat))/2)$values))  
}  
moran.range(W)
```

3.1. Autocorrelazione spaziale: I di Moran

$$Z(I) = \frac{I - E(I)}{S_{E(I)}} \quad E(I) = -\frac{1}{N-1}$$

La deviazione standard viene calcolata in maniera differente a seconda se si assume che:

- $Z(I)$. segua una normale standard, $S_{E(I)}$ è data da (Ebdon, 1988):

$$S_{E(I)} = \sqrt{\frac{N^2 \sum_{ij} w_{ij}^2 + 3(\sum_{ij} w_{ij})^2 - N \sum_i (\sum_j w_{ij})^2}{(N^2 - 1)(\sum_{ij} w_{ij})^2}}$$

La distribuzione spaziale degli eventi sia equiprobabile in ogni punto del dominio spaziale

$$S_{E(I)} = \sqrt{\frac{N((S_1(N^2 - 3N + 3) - NS_2 + 3S_0^2) - \frac{K(S_1(N^2 - N) - 2NS_2 + 6S_0^2)}{(N-1)(N-2)(N-3)S_0^2} - \left(\frac{1}{N-1}\right)^2)}{(N-1)(N-2)(N-3)S_0^2}}$$

$$S_0 = \sum_{i=1}^N \sum_{j=1}^N w_{ij} \quad S_2 = \sum_{i=1}^N (\sum_{j=1}^N w_{ij} + \sum_{j=1}^N w_{ji})^2 \quad S_1 = \frac{1}{2} \sum_{i=1}^N \sum_{j=1}^N (w_{ij} + w_{ji})^2$$

Weights style: W

Weights constants summary:

n	nn	S0	S1	S2
W 154	23716	154	63.36271	636.4623

3.1. Autocorrelazione spaziale: I di Moran

```
lx<-moran.test(x, listw=W, zero.policy=T)#1°ordine
l2x<-moran.test(x, listw=W2, zero.policy=T)#2°ordine
l3x<-moran.test(x, listw=W3, zero.policy=T)#3°ordine
moran.plot(x, W,zero.policy=F,xlab="X", ylab="WX",
           labels=T,main="Asia19",pch=16,cex=0.2, quiet=F)#Labels=punti con alta influenza
```

Moran I test under randomisation

data: x

weights: W

Moran I statistic standard deviate = 9.3362, p-value < 2.2e-16

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.468826346	-0.006535948	0.002592466

data: x

weights: W2

Moran I statistic standard deviate = 8.7688, p-value < 2.2e-16

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.290081524	-0.006535948	0.001144223

data: x

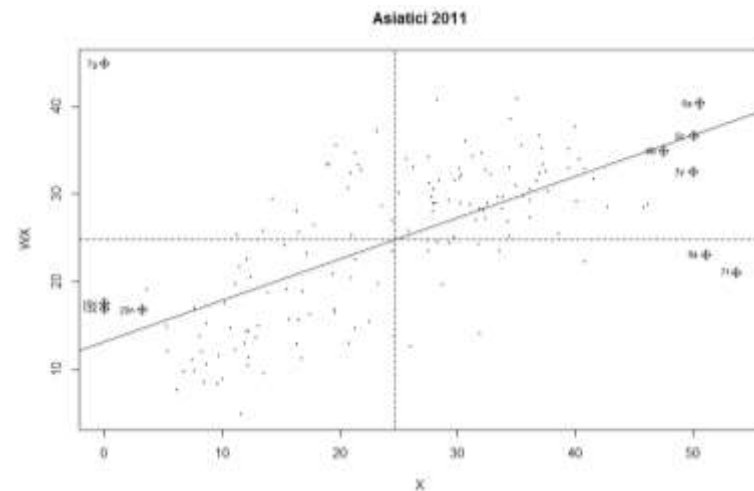
weights: W3

Moran I statistic standard deviate = 3.2175, p-value = 0.00064

alternative hypothesis: greater

sample estimates:

Moran I statistic	Expectation	Variance
0.0808391853	-0.0065359477	0.0007374496



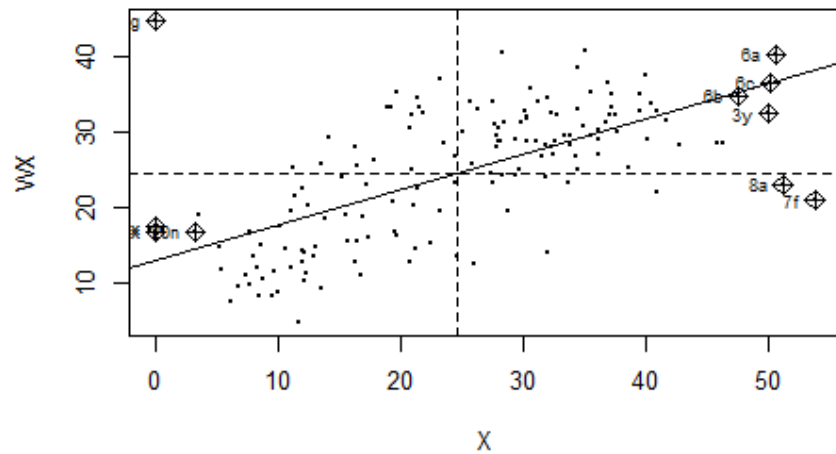
Potentially influential observations of

lm(formula = wx ~ x) :

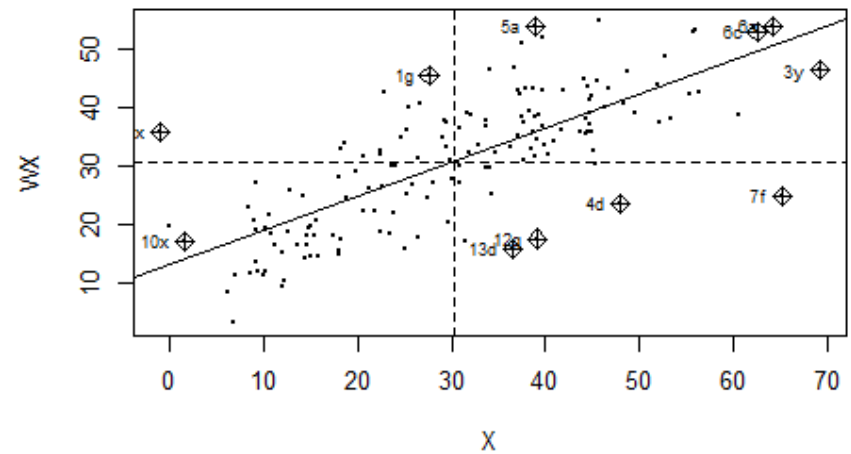
dfb.1_	dfb.x	dfbt	cov.r	cook.d	hat
10x	0.12	-0.11	0.12	1.04_*	0.01 0.03
13x	0.10	-0.09	0.10	1.04_*	0.01 0.03
20n	0.05	-0.05	0.05	1.04_*	0.00 0.03
3y	0.07	-0.11	-0.12	1.05_*	0.01 0.04
6a	-0.06	0.09	0.10	1.05_*	0.01 0.04
6b	0.01	-0.01	-0.02	1.04_*	0.00 0.03
6c	0.00	0.00	0.00	1.05_*	0.00 0.04
7f	0.39	-0.54	-0.58_*	0.96	0.16 0.04_*
7g	0.97	-0.87	0.97_*	0.75_*	0.40 0.03
8a	0.27	-0.39	-0.43_*	0.99	0.09 0.04

3.1. Autocorrelazione spaziale: I di Moran

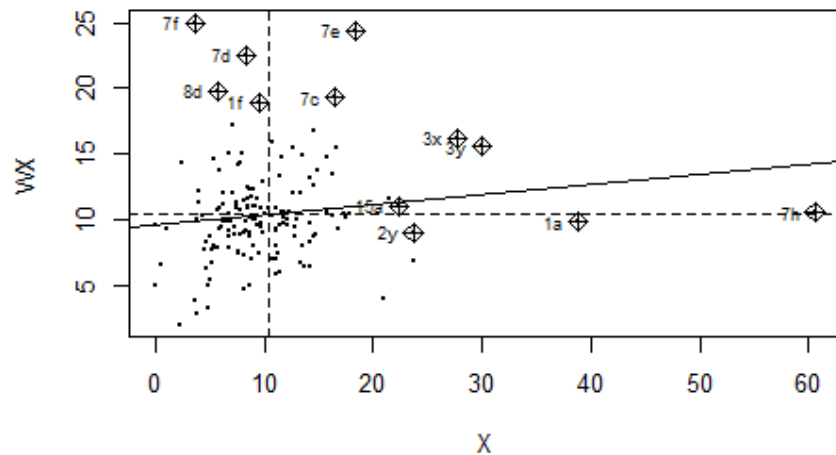
Asiatici 2011



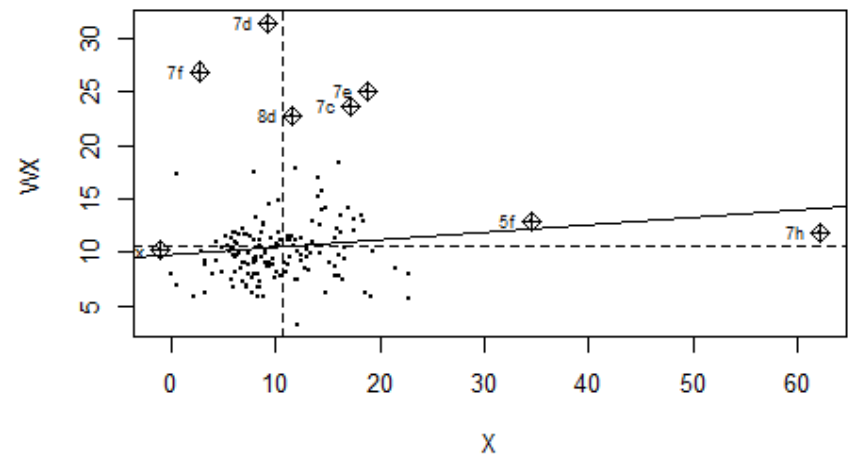
Asiatici 2019



Africani 2011



Africani 2019



3.2. Relazione tra I-Moran e r-Pearson

$$z_j^* = \sum_i (x_i - M) w_{ij} = \sum_i w_{ij} z_{ij}$$

$$M(z) \cong M(z^*)$$

$$\text{Var}(z) \geq \text{Var}(z^*)$$

#Spatiality lagged variable

`x<-STR.ZU$As11`

`Wx<-lag.listw(W,x,zero.policy=T)`

$z = I z^* \rightarrow I$ coefficiente di regressione

$$r = \frac{\text{Cov}(z, z^*)}{\sqrt{\text{Var}(z)} \sqrt{\text{Var}(z^*)}}; I = \frac{\text{Cov}(z, z^*)}{\sqrt{\text{Var}(z)} \sqrt{\text{Var}(z)}}$$

r e I stesso numeratore ; denominatore di $r < I$

$$I \leq \sqrt{\frac{\text{Var}(z^*)}{\text{Var}(z)}}$$

$$I = r \sqrt{\frac{\text{Var}(z^*)}{\text{Var}(z)}}$$

3.4. Scomposizione I di Moran

Varianza totale

Varianza unità contigue

Varianza tra unità non connesse

$$\text{Var}(z) = \text{Var}(z^*) + \text{Var}(\#z)$$

$$I = \frac{\text{cov}(zz^*)}{\text{Var}(z)} \cdot \frac{\text{Var}(z^*)}{\text{Var}(z^*)} = \frac{\text{cov}(zz^*)}{\text{Var}(z^*)} \cdot \frac{\text{Var}(z^*)}{\text{Var}(z)} = I^* \cdot RC$$

- $I^* \rightarrow$ registra il segno dell'autocorrelazione spaziale
- $RC \rightarrow$ fornisce la misura dell'autocorrelazione spaziale

$RC < 1 \rightarrow \text{Var}(z^*) < \text{Var}(z)$. La struttura spaziale del vicinato riduce la **variabilità intrinseca**. In relazione alla variabile esaminata, le zone contigue sono più simili rispetto a tutte le altre aree.

Il contrario se **$RC > 1 \rightarrow \text{Var}(z^*) > \text{Var}(z)$** . La struttura spaziale del vicinato non modifica la variabilità del fenomeno.

I_z e I_{z^*} non sono simmetrici $\rightarrow$ non sono indici di correlazione

$$\frac{\text{cov}(zz^*)}{\text{Var}(z)} \neq \frac{\text{cov}(zz^*)}{\text{Var}(z^*)}$$

```
var(x)
var(Wx)
cov(x,Wx)
RC<-var(Wx)/var(x)
I<-cov(x,Wx)/var(x)
Iw<-cov(x,Wx)/var(Wx)
Iw*RC# =I
```

3.4.Scomposizione territoriale: Local Moran

L'indice di Moran fornisce una misura sintetica dell'autocorrelazione globale.

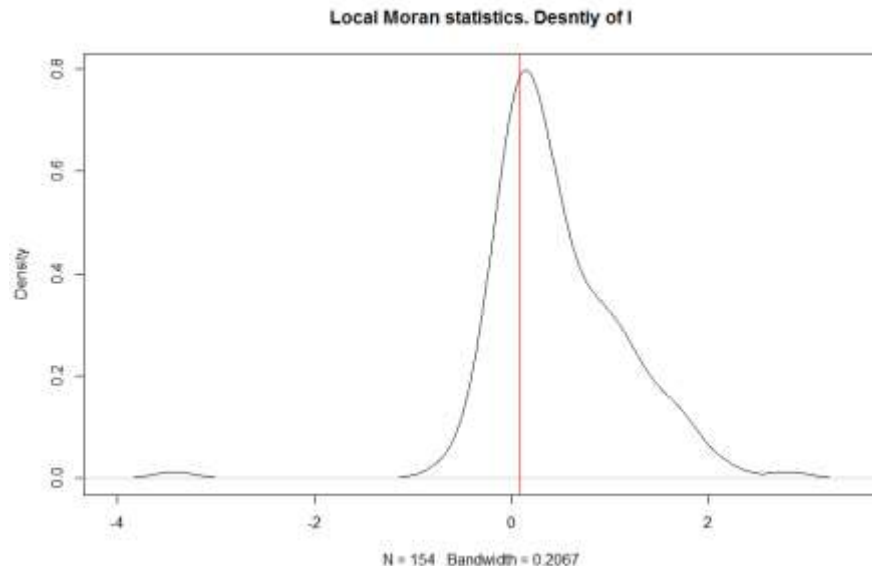
È plausibile, tuttavia, che non tutte le unità abbiano lo stesso peso sulla determinazione di questa statistica.

Per identificare le unità che contribuiscono in modo significativo alla formazione di I è possibile utilizzare l'indice di autocorrelazione locale $LISA$.

$$L_i = \frac{Z_i}{\sum_i Z_i^2} \sum_j w_{ij} Z_j \quad \rightarrow \quad I = \sum_i L_i$$

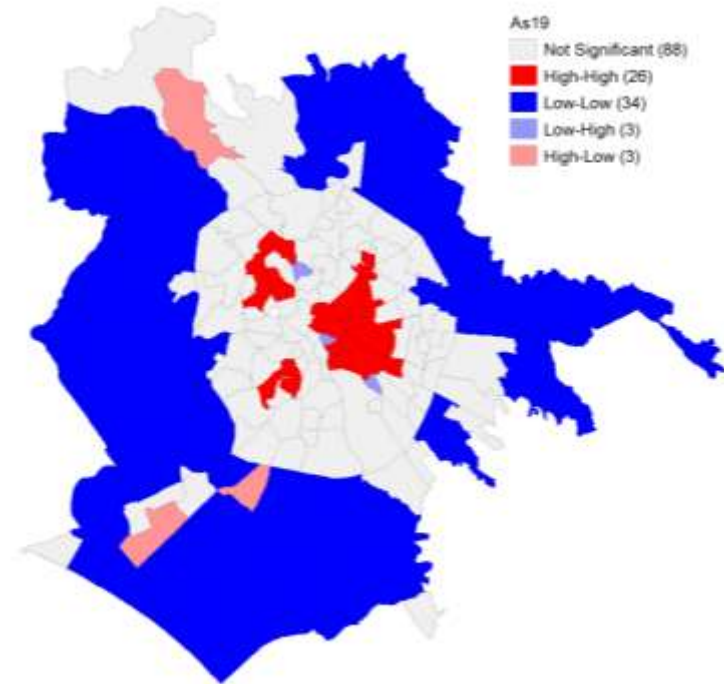
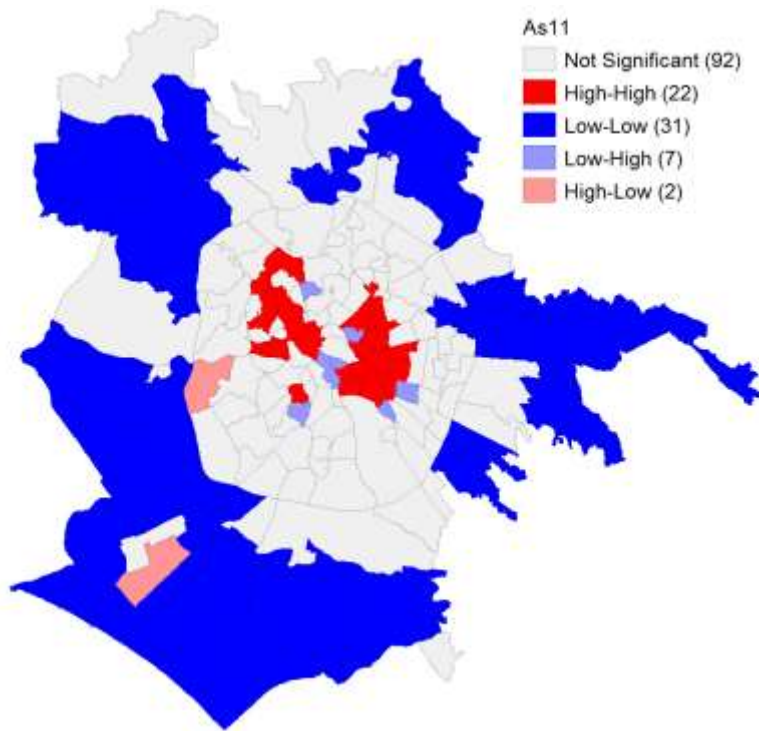
```
LISA.x<-localmoran(x, W,zero.policy=T,alternative = "greater")
```

```
plot(density(LISA.x[,1]),main="Local Moran statistics. Desntiy of I")  
abline(v=0.08, col="red")# v=Ix
```



	li	E.li	Var.li	Z.li	Pr(z > 0)
10a	0.82338	-0.00653	0.19309	1.8886	2.946927e-02
10b	0.48333	-0.00653	0.19309	1.1148	1.324673e-01
10c	0.13223	-0.00653	0.13607	0.3761	3.533879e-01

3.5.Scomposizione territoriale: Local Moran



```
STR.ZU <- st_as_sf(STR.ZU) %>%  
  mutate(quad_sig = ifelse(y > 0 & Wy > 0 & LISA.y[,5] <= 0.05, "high-high",  
    ifelse(y <= 0 & Wy <= 0 & LISA.y[,5] <= 0.05, "low-low",  
    ifelse(y > 0 & Wy <= 0 & LISA.y[,5] <= 0.05, "high-low",  
    ifelse(y <= 0 & Wy > 0 & LISA.y[,5] <= 0.05, "low-high",  
    "non-significant")))))
```

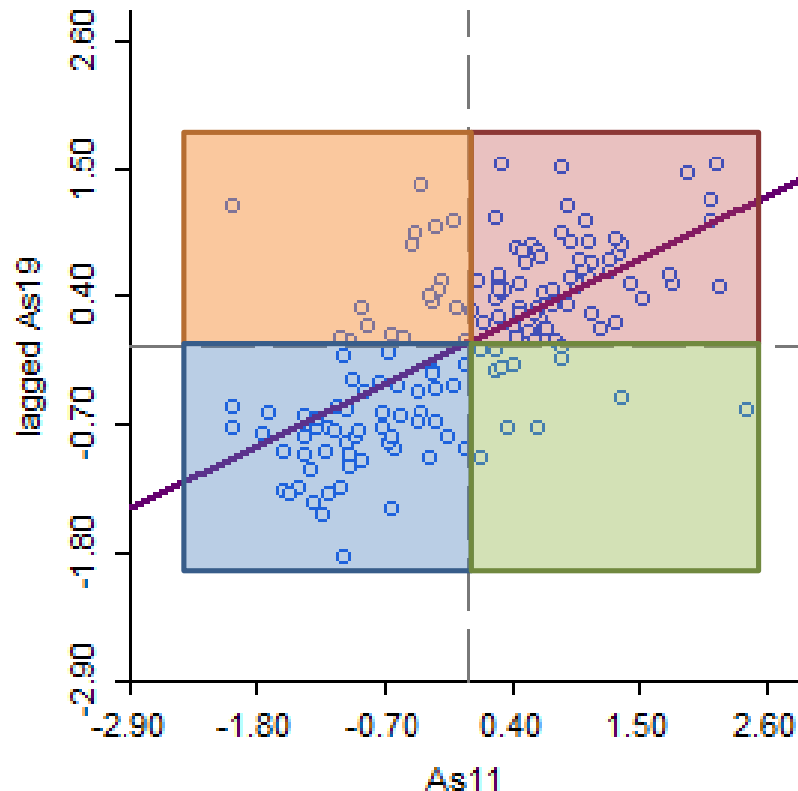
```
qtm(STR.ZU, fill="quad_sig", fill.title="LISA-Asiatici 19")
```

3.6.I di Moran Bivariato

La versione bivariata dell'indice di Moran (cfr. Wartenberg, 1985) registra l'intensità della relazione **tra due variabili diverse** o con **la stessa variabile ritardata nel tempo** (autocorrelazione spazio-temporale) delle quali una spazialmente ritardata e, per matrici **W** si può scrivere nel seguente modo:

$$I_b = \frac{Z'_x W Z_y}{Z'_x Z_x} \quad I_{t-k,t} = \frac{Z'_{t-k} W Z_t}{Z'_{t-k} Z_{t-k}}$$

Moran's I: 0.495



NB: Non c'è una funzione R per il calcolo di questo indice .
C'è una specifica funzione in **GeoDa** .

(Anselin, Syabri e Smirnov,2002)

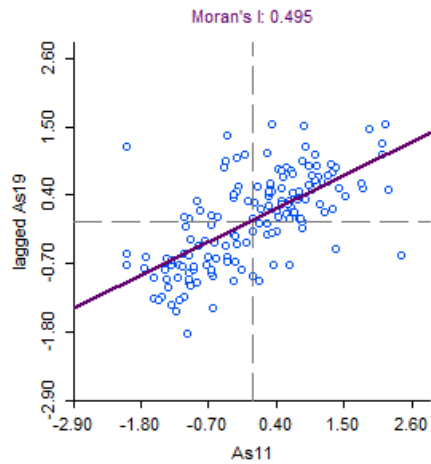
Zu con presenza **Alta** di Asiatici nel 2011 e nel 2019
“ **territori a forte attrazione** ”

Zu con presenza di Asiatici **Bassa** nel 2011 e **Alta** 2019
“ **territori di nuovo insediamento** ”

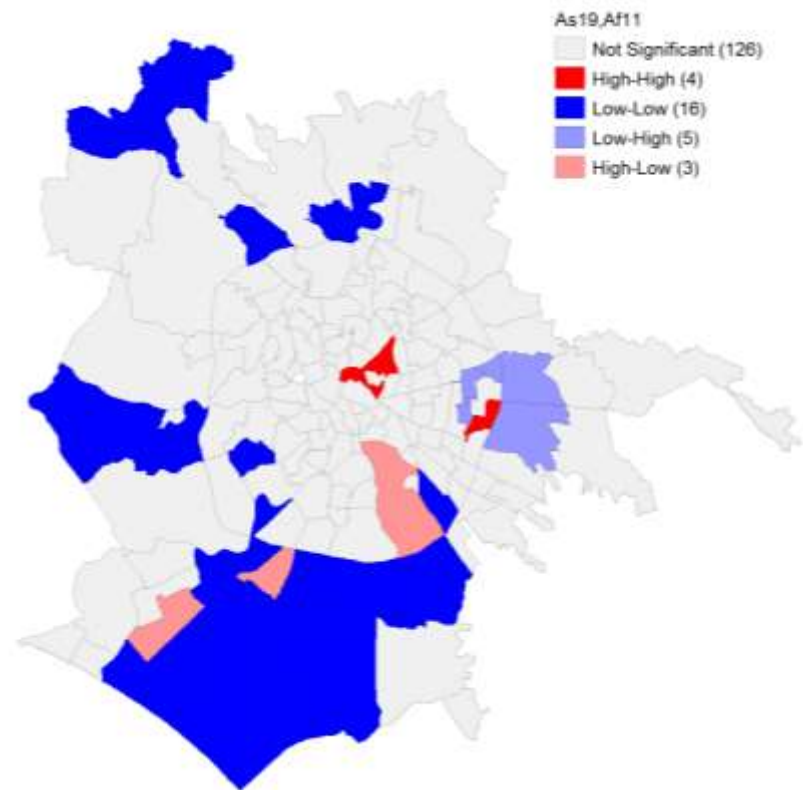
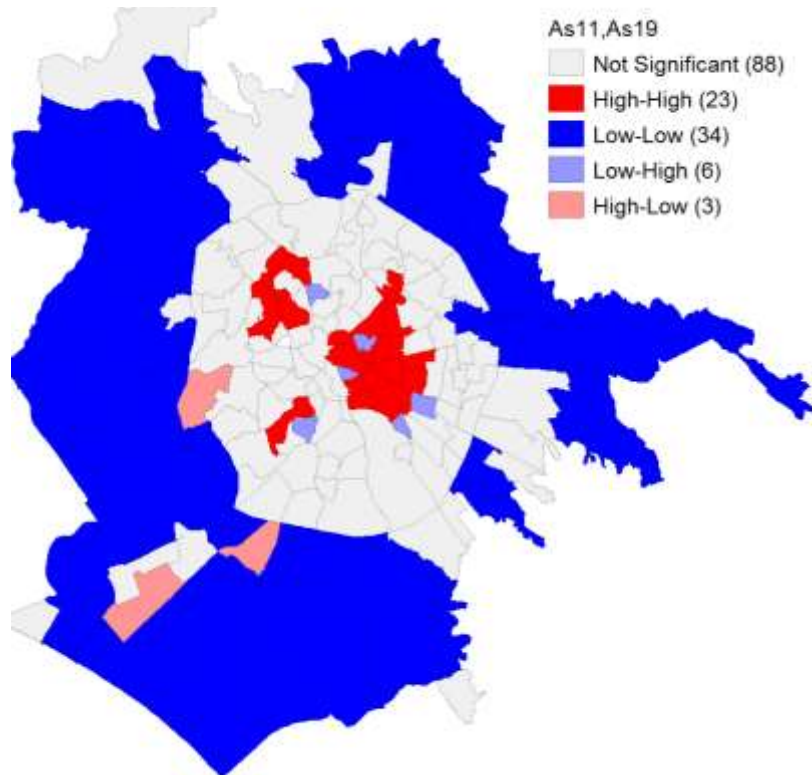
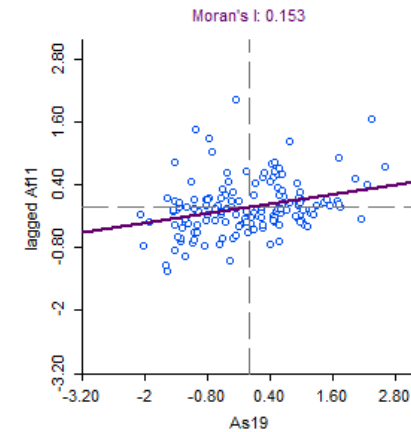
Zu con presenza di Asiatici **Bassa** nel 2011 e nel 2019
“ **territori a bassa attrazione** ”

Zu con presenza di Asiatici **Alta** nel 2011 e **Bassa** 2019
“ **territori a bassa attrazione** ”

3.6.I di Moran Bivariato



$$L_i = \frac{Z_i}{\sum Z_i^2} \sum_j w_{ij} u_j$$



3.7.L di LEE globale e locale

Criticità del Moran-bivariato: non tiene conto della correlazione tra le due variabili **nella stessa** unità areale ***I_b***. Si potrebbe sovrastimare l'autocorrelazione spaziale che potrebbe invece essere dovuta principalmente o anche alla correlazione tra le due variabili nell'*i*-esima unità.

Lee (2001) fornisce un'alternativa all'indice *I_b* tenendo conto proprio della correlazione tra le due variabili nella medesima unità territoriale

$$L_{zu^*} = \frac{\text{Cov}(z^*, u^*)}{\sqrt{\text{Var}(z)} \sqrt{\text{Var}(u)}} = \sqrt{\frac{\text{Var}(z^*)}{\text{Var}(z)} \frac{\text{Var}(u^*)}{\text{Var}(u)}} r_{z^*u^*}$$
$$r_{z^*u^*} = \frac{\text{Cov}(z^*, u^*)}{\sqrt{\text{Var}(z^*)} \sqrt{\text{Var}(u^*)}}$$

$$E(L) = \frac{\text{tr}(W^T W)}{N-1} r_{zu}$$

```
lee.test(x, y, W, zero.policy=TRUE, alternative="less", na.action=na.omit)
```

$$IL_{zu^*} = \frac{N(z^*, u^*)}{\sqrt{\text{Var}(z)} \sqrt{\text{Var}(u)}}$$

```
Lxy <- lee(x, y, W, length(x), zero.policy=TRUE)
```

```
$localL
```

```
[1] 0.4778735900 0.1484630725 0.0802392750 0.0051015530 0.0285288195 0.0897475105  
[7] 0.9074014313 1.3089842416 0.8867839844 1.1600136372 0.5215443032 0.1981826162  
[13] 1.1036577926 0.3501365645 0.2519642386 0.1892556637 0.2172667429 0.1351776863  
[19] 0.0131240350 0.4078133303 0.0612037879 0.0279133543 0.0006632198 -0.0018986064  
[25] -0.0119344110 0.2617379913 0.7502988931 0.4771051054 2.0054396268 1.2775996204  
[31] 3.0062089331 1.5771852237 0.1542583396 0.6354878592 0.8074864192 0.5486499887
```

3.7.L di LEE globale e locale

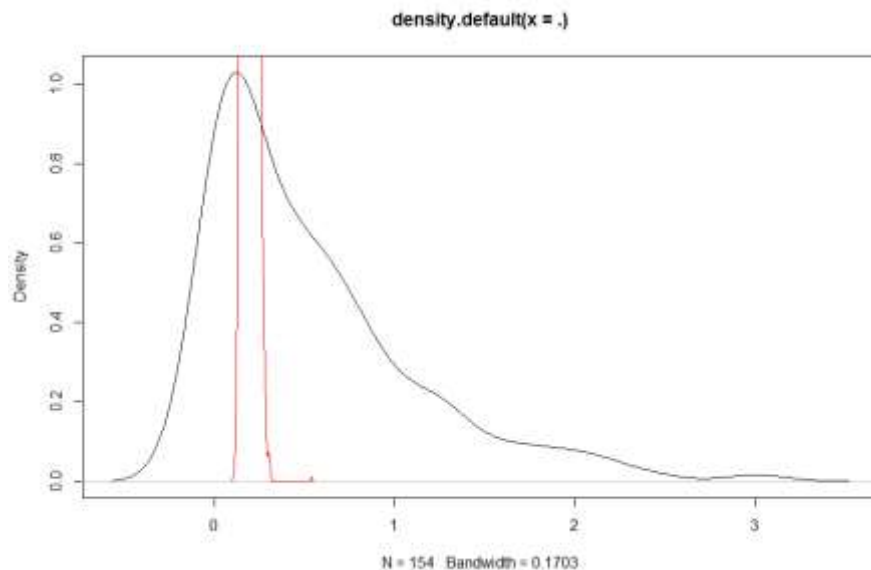
k permutations to estimate pseudo p-values

```
LMCxy <- lee.mc(x, y, nsim=10000, W, zero.policy=TRUE, alternative="less")
```

quick plot of local L

```
Lxy[[2]] %>% density() %>% plot() # Lee's local L statistic (Local)
```

```
LMCxy[[7]] %>% density() %>% lines(col="red") # plot values simulated 10k times
```

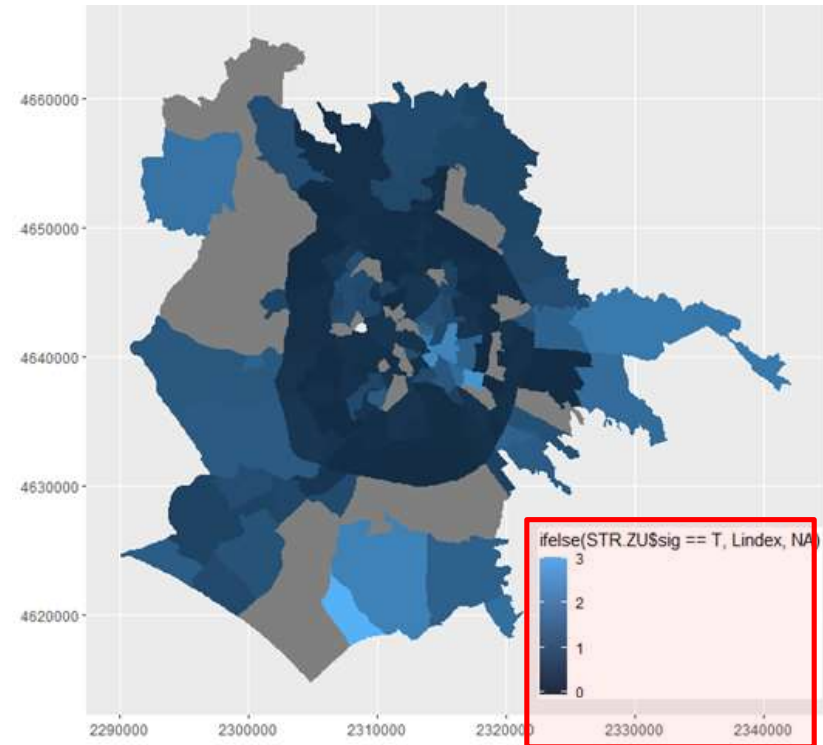
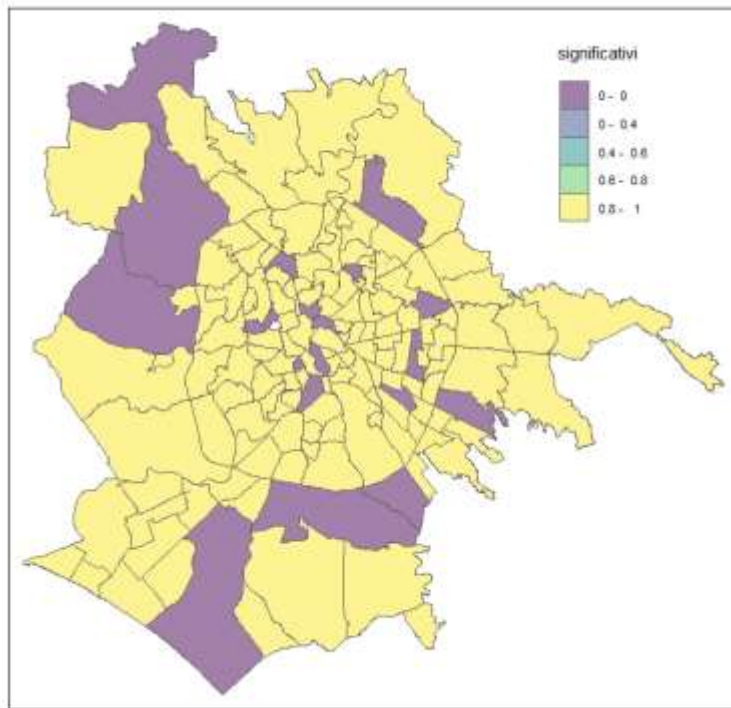


get confidence interval of 95% (mean +- 2 standard deviations)

```
two_sd_above <- mean(LMCxy[[7]]) + 2 * sd(LMCxy[[7]])
```

```
two_sd_below <- mean(LMCxy[[7]]) - 2 * sd(LMCxy[[7]])
```

3.7.L di LEE globale e locale



Valori min e max
di Local Lee

convert spatial object to sf class for easier/faster use

```
STR_sf <- st_as_sf(STR.ZU)
```

add L index values to map object

```
STR_sf$Lindex <- Lxy[[2]]
```

Preparing for plotting

identify significant local results

```
STR.ZU$sig <- if_else( STR_sf$Lindex < 2*two_sd_below, 1, if_else( STR_sf$Lindex > 2*two_sd_above, 1, 0))
```

Map of Local L index but only the significant results

```
ggplot() + geom_sf(data=STR_sf, aes(fill=ifelse( STR.ZU$sig==T, Lindex, NA)), color=NA)
```