University of California, Los Angeles
Department of Statistics

Statistics C173/C273 **Instructor: Nicolas Christou**

Homework 3

EXERCISE 1
We discussed on Friday, 01/21 different methods for testing for autocorrelation. One of the methods using spatiotemporal data is to transform the data into binary data. Please create a simple example to test the hypothesis of spatial and temporal clustering using a contingency table.

EXERCISE 2
Also on Friday, 01/21 we started discussing testing for autocorrelation for lattice data, and we introduced the idea of BB and BW joins. The formula for the number of BB and BW joins are given as follows:
$BB = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} Z(s_i)Z(s_j)$ and
$BW = \frac{1}{2}\sum_{i=1}^{n}\sum_{j=1}^{n} w_{ij} [Z(s_i) - Z(s_j)]^2$.
Please provide an example using a regular lattice (similar to the one from lecture) to verify that BB and BW provides the number of joins.

EXERCISE 3
Consider the county adjacency file downloaded from
`https://www.census.gov/geographies/reference-files/2010/geo/county-adjacency.html`. You can read the data in R using:

```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/
county_adjacency.txt", sep="\t", fill=FALSE, strip.white=TRUE)[,c(1,3)]
```

Answer the following questions:

a. Construct the adjacency matrix for the counties of California, $w_{ii} = 0, w_{ij} = 1$, etc., as we discussed in class.

b. Use the `maps` package to plot the map of California with the county lines. Add the seat of each county on the map. You can access the county seats here:

   ```
   a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/
   ca_seats_coord.txt", header=TRUE)
   ```

   Compute the distance matrix for the California county seats.

EXERCISE 4
Show that the Moran's I statistic is not affected if the data are multiplied by a constant λ. For example, using $Z(s_i), i = 1, \ldots, n$ or $\lambda Z(s_i), i = 1, \ldots, n$ will give the same I statistic.

EXERCISE 5
Consider the 2×3 regular lattice with observations $Z(s_1) = 5, Z(s_2) = -3, Z(s_3) = -6, Z(s_4) = 2, Z(s_5) = 4, Z(s_6) = -2$ as in the following layout:

	Column 1	Column 2	Column 3
Row 1	$Z(s_1) = 5$	$Z(s_2) = -3$	$Z(s_3) = -6$
Row 2	$Z(s_4) = 2$	$Z(s_5) = 4$	$Z(s_6) = -2$

For the adjacency matrix assume that the lattices are connected in the east-west and south-north directions,, i.e. there is connectivity east or west of the lattice and north or south of the lattice. For example, lattice 2 is adjacent to lattices 1 to west direction, 3 to east direction, and 5 to south direction.

Answer the following questions:

a. Find the mean and variance of the Moran's I statistic using the permutation test by enumerating all the 6! permutations of the lattice. Construct the histogram of the 6! values of I.

b. Under normality, the mean of I is given by $E(I) = -\frac{1}{n-1}$ and
$EI^2 = \frac{1}{(n-1)(n+1)S_0^2}\left[n^2 S_1 - nS_2 + 3S_0^2\right]$.
Note:
Note:
$S_0 = \sum_{i=1}^n \sum_{j=1}^n w_{ij}$,
$S_1 = \frac{1}{2}\sum_{i=1}^n \sum_{j=1}^n (w_{ij} + w_{ji})^2$,
$S_2 = \sum_{i=1}^n \left(\sum_{j=1}^n w_{ij} + \sum_{j=1}^n w_{ji}\right)^2$,

Compute the empirical p-value using the permutation test under the hypothesis of no spatial autocorrelation (use the data from (a)) and compare it with the p-value based on normal approximation using the formulas in (b). Note: Using EI and EI^2 compute the variance of I and then compute $Z = \frac{I-E[I]}{\sqrt{\mathrm{var}[I]}}$.