University of California, Los Angeles
Department of Statistics

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

Homework 4

EXERCISE 1
In homework 3 we constructed the adjacency matrix for the counties of California. Please find an attribute of interest at the county level of California. This will be the response variable $\mathbf{Z}$ vector. Now find few predictors that you think might be associated with $\mathbf{Z}$. You can use for example the coordinates x and y (longitude and latitude) for each county seat, but also you should include other predictors. Note: You can use data for other states or countries. Answer the following questions:

a. Assume there is an intercept in the model $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$. Construct and compute the following: $\mathbf{X}, \mathbf{X}'\mathbf{X}, \hat{\boldsymbol{\beta}}, \mathbf{H}, \hat{\mathbf{Y}}, \mathbf{e}$

b. Compute the Moran's I test statistic using the residuals and its expected value $E[I]$.

EXERCISE 2
Access the soil data:

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

Answer the following questions:

a. Convert `a` into a `geodata` object using `geoR`.

b. Compute and plot the empirical semivariogram up to a maximum distance of 1800.

c. Fit by eye a spherical theoretical semivariogram and add the plot on the empirical variogram.

EXERCISE 3
Access the jura data:

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

Answer the following questions:

a. Create a data frame using the first two columns (coordinates) and variable `Cd`. Convert this data frame into a `geodata` object using `geoR` and compute the maximum distance in your data.

b. Compute and plot the empirical semivariogram up to a maximum distance of 3.

c. Fit by eye a spherical theoretical semivariogram and add the plot on the empirical semivariogram.

d. Indicator semivariograms: Transform your data into 0, 1 data based on some threshold. Form example

```
I <- ifelse(a$Cd >= 1, 1,0)
```

Your variable now is `I`. Construct a data frame using the first two columns (coordinates) and this indicator variable `I`. Compute and plot the empirical semivariogram up to a maximum distance of 5.5, fit by eye a gaussian theoretical semivariogram, and add the plot on the empirical variogram. Do the same using maximum distance of 2.

EXERCISE 4
Consider the Moran's test statistic I for spatial autocorrelation. If we use the model $\mathbf{Z} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$ then the test statistic is given by $I = \frac{n}{S_0}\frac{\mathbf{e}'\mathbf{W}\mathbf{e}}{\mathbf{e}'\mathbf{e}}$. Using this formula, show that when the model is $\mathbf{Z} = \mu\mathbf{1} + \boldsymbol{\epsilon}$ we get $E(I) = -\frac{1}{n-1}$. Find and expression for $E(I)$ when $k = 1$ predictor (simple regression, $y = \beta_0 + \beta_1 x_i + \epsilon_i$).

EXERCISE 5
Consider the spatial locations for the ozone monitoring stations in California. You can access them here:

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

Answer the following questions:

a. Using the Cholesky decomposition method and assuming multivariate normal data with mean $\mu = 0.10$ and variance covariance matrix $\boldsymbol{\Sigma}$, simulate manually the following isotropic process: Exponential model with $c_0 = 0, c_1 = 0.05, \alpha = 2.8$.

b. Compute the sample semivariogrm and fit the exponential model to it. On the same graph plot the theoretical model used in (a).