University of California, Los Angeles
Department of Statistics

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

Homework 6

EXERCISE 1
Access the Kruger Park data:

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

Load the packages `sp` and `gstat` to construct h-scatterplots with separation distances from 0 to 1 by 0.1.

EXERCISE 2
Access the coal ash data:

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

Answer the following questions:

a. Convert `a` into a `geodata` object using `geoR` and compute the maximum distance for the data.
b. Compute and plot the semivariogram up to distance 10 in the east-west direction. Use both the classical and the robust estimators.
c. Fit a model semivariogram to the empirical semivariograms.
d. Compute the semivariogram cloud for both estimators for direction east-west and then construct the box plot for each cloud.

EXERCISE 3
Access the Maas river data in `R` here:

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

Answer the following questions:

a. Use `gstat` to compute the and plot the sample semivariogram of `log(lead)`.
b. Fit the spherical semivariogram model to the sample semivariogram of part (a) using Cressie's weights and OLS (fit.method=2 or 6).

EXERCISE 4
Access the Wolfcamp data in `R` here:

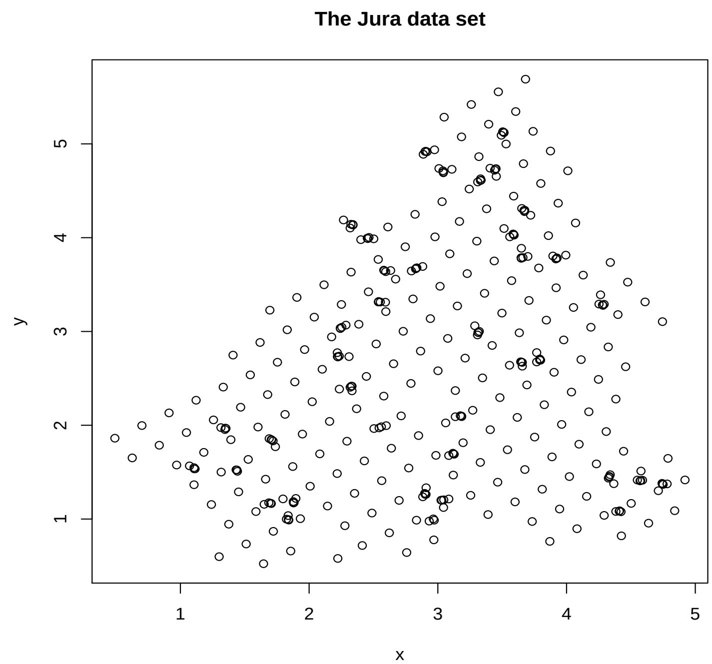
```
a <- read.table("http://www.stat.ucla.edu/~nchristo/statistics_c173_c273/wolfcamp.txt",
header=TRUE)
```

Answer the following questions:

a. Use `gstat` to compute the and plot the sample semivariogram of `level)`. Assume constant mean (use formula = level $\sim$ 1). What do you observe.
b. Remove the trend by using: formula = level $\sim x+y$. Compute the and plot the sample semivariogram using the de-trended data.
c. Fit the spherical semivariogram model to the sample semivariogram of part (b) using fit.method=1,2,6,7).

EXERCISE 5

These Jura data were collected by the Swiss Federal Institute of Technology at Lausanne. See Goovaerts, P. 1997, "Geostatistics for Natural Resources Evaluation", Oxford University Press, New-York, 483 p. for more details. Data were recorded at 359 locations scattered in space (see figure below).



Concentrations of seven heavy metals (cadmium, cobalt, chromium, copper, nickel, lead, and zinc) in the topsoil were measured at each location. The type of land use and rock type was also recorded for each location. The data can be accessed here:

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

```
> names(a)
 [1] "x"       "y"       "Landuse" "Rock"    "Cd"
 [6] "Co"      "Cr"      "Cu"      "Ni"      "Pb"
[11] "Zn"
```

The variables x, y are the coordinates. Landuse and Rock represent type of land use (forest, pasture, meadow, tillage) and rock type (Argovian, Kimmeridgian, Sequanina, Portlandian, and Quaternary). The other variables are concerntrations in ppm of the following chemical elements:

`Cd`: Cadmium
`Co`: Cobalt
`Cr`: Chromium
`Cu`: Copper
`Ni`: Nickel
`Pb`: Lead
`Zn`: Zinc

Compute and plot the semivariogram for each of the variables `Cd`, `Co`, `Cr`, `Cu`, `Ni`, `Pb`, `Zn`, and fit a model semivariogram to them.

EXERCISE 6

Suppose that $Z(s)$ is a second order stationary process with $E[Z(s)] = 0$ and with cubic semivariogram:

$$\gamma(h;\boldsymbol{\theta}) = \begin{cases} 0 & h = 0 \\ c_1\left[7(\frac{h}{\alpha})^2 - 8.75(\frac{h}{\alpha})^3 + 3.5(\frac{h}{\alpha})^5 - 0.75(\frac{h}{\alpha})^7\right] & 0 < h \leq \alpha \\ c_1 & h > \alpha \end{cases}$$

Answer the following questions:

a. Write the covariance function $C(h;\boldsymbol{\theta})$ that corresponds to this cubic semivariogram.

b. Draw (up to $h = 3$ only) this semivariogram and the covariance function on the same plot using `R`. Use $c_1 = 1, \alpha = 3$.

EXERCISE 7

Suppose that the second order stationary process $Z(s)$ can be expressed as a linear combination of three independent random functions as follows: $Z(s) = \mu + 3Y_0(s) - 4Y_1(s) + 2Y_2(s)$, with $E(Y_i(s)) = 0$, $cov(Y_i(s), Y_i(s+h)) = c_i(h)$, and $cov(Y_i(s), Y_j(s+h)) = 0$. Suppose, $Y_0(s)$ is a pure nugget model with $c_0 = 1$, the corresponding model for $Y_1(s)$ is the spherical with $c_1 = 1, \alpha = 1$, and the corresponding model for $Y_2(s)$ is the power with $\omega = 1$. Draw the three basic semivariogram models on the same graph and the semivariogram model of the combination of the three models on another plot.

EXERCISE 8

Use your own data that you chose for the final project. Begin with non-spatial data analysis by constructing histograms, box plots and computing the descriptive statistics and the variance covariance and correlation matrix, etc. Compute the semivariogram for each of the variables in the data set and fit a theoretical model to them.