Homework 1 - Stat 534
Due Wednesday, January 18, 2017

Problem 1 requires you to use the R package `spdep`.

1. Our text implies and others state outright that the BB, BW, and WW statistics reveal pretty much the same thing about spatial correlation. The `joincount.mc` function will carry out Monte Carlo tests based on the BB and WW statistics. We do not have an R formula for computing the BW statistic but it is possible to carry out a BW joincount test of spatial autocorrelation (or clustering) using Geary's c.
   (a) Show the relationship between Geary's c and BW.
   (b) Carry out a test based on the BW statistics using `geary.mc`. The data file `atrplx.dat` will be emailed to you at your math department email addresses. The first 2 columns contain the spatial coordinates and the **fourth** column contains the Z values you need. Use the R handout to generate the necessary neighbors and list objects.
   (c) Using the output from `geary.mc` compute BW and $E[BW]$. Do you expect $BW < E[BW]$ or $BW > E[BW]$ in the presence of positive spatial clustering of the plants? Why or why not?
   (d) Reproduce the analysis I presented in class using the *Atriplex* data. Compare the results of the BW test to those of the BB and WW test that were discussed in class. Do these statistics all seem to indicate the same thing about spatial clustering of the plants.
   (e) For grins compute Moran's I and compare that result to those above.

2. Categorize the following examples of spatial data as to their data type:
   (a) Elevations in the foothills of the Allegheny mountains.
   (b) Highest elevation within each state in the United States.
   (c) Concentration of a mineral in soil.
   (d) Plot yields in a uniformity trial.
   (e) Crime statistics giving names of subdivisions where break-ins occurred in the previous year and property loss values.
   (f) Same as previous, but instead of the subdivisions, the individual dwelling, is identified.
   (g) Distribution of oaks and pines in a forest stand.

3. Show that Moran's I is a scale-free statistic, i.e. $Z(\mathbf{s})$ and $\lambda Z(\mathbf{s})$ yield the same value for any constant $\lambda \neq 0$.

4. Let $Y_1, \cdots, Y_n$ be normally distributed with unknown mean μ and known variance σ^2. Let $\text{Cov}\,(Y_i, Y_j) = \sigma^2 \rho$ for $i \neq j$. We will further assume that $\rho > 0$.
   (a) Show that
$$\text{Var}\left(\overline{Y}\right) = \frac{\sigma^2}{n}\left[1 + (n-1)\,\rho\right]$$

(b) Let $n = 10$ and $\rho = 0.26$. Compare and contrast a 95% confidence interval for μ computed using the true standard deviation of $\overline{Y}$ and one computed assuming independence.

(c) Given independence, we know that $\overline{Y}$ is the "best" estimator of μ. One nice property it has is that it is a consistent estimator of the mean. Is $\overline{Y}$ a consistent estimator of the mean given the correlation structure above? Justify your answer.

(d) Recall that effective sample size is a measure of the effect of correlation on inference. An equation for the effective sample size under the equicorrelation model is

$$n^{'} = \frac{n}{1 + (n-1)\rho}.$$

The effective sample size is defined to be the sample size $n^{'}$ of uncorrelated observations that provide the same information (in a sense) as a sample of n correlated observations.

i. Compute the effective sample size when $n = 10, 100,$ and 1000 and $\rho = 0.05, 0.1, 0.25,$ and 0.5.

ii. Find $\lim n^{'}$ as $n \to \infty$.

iii. The effect is extreme here but we would not expect to see this type of correlation structure in a spatial setting. Why not?