STAT 696, Spatial Statistics Midterm Exam, Spring 2011

Due 5:30PM, Thursday March 17

This assignment is a take-home midterm exam. You **may not** collaborate with *any* other person (whether in the class or not). You **may** use any reading material (class notes, books, etc.) you wish. Professor Bailey will answer questions.

Please follow the lab report directions for Homework, i.e. include commands and output you used to answer the questions.

3 Problems 25, 25, 50: 100 total points.

1. The ordinary kriging equations were derived in class. This problem will derive the more commonly given equations.

(a) Using the form of an inverse partitioned matrix (found in Searle, *Linear Models*, 1971, p. 27) or other places, show that the weights for ordinary kriging are given by:

$$\boldsymbol{w}_1 = \begin{bmatrix} w_1 \\ \vdots \\ w_n \end{bmatrix} = \boldsymbol{C}_1^{-1}\boldsymbol{c}_0 - \frac{\boldsymbol{C}_1^{-1}\mathbf{1}\mathbf{1}'\boldsymbol{C}_1^{-1}\boldsymbol{c}_0}{\mathbf{1}'\boldsymbol{C}_1^{-1}\mathbf{1}} + \frac{\boldsymbol{C}_1^{-1}\mathbf{1}}{\mathbf{1}'\boldsymbol{C}_1^{-1}\mathbf{1}}$$

Note: The partitioned matrix is the form:

$$\begin{bmatrix} \boldsymbol{A} & \boldsymbol{B} \\ \boldsymbol{B}' & \boldsymbol{D} \end{bmatrix}$$

(b) Let $\hat{m}_{GLS}$ be the generalized least squares estimate of the mean given by:

$$\hat{m}_{GLS} = (\mathbf{1}'\boldsymbol{C}_1^{-1}\mathbf{1})^{-1}\boldsymbol{C}_1^{-1}\boldsymbol{z}_1$$

where $\boldsymbol{z}_1' = (Z_1, Z_2, \ldots, Z_n)$ is the vector of response variables.

Show that the predicted value of Z at $\boldsymbol{s}_0$ which is given by:

$$\hat{Z}_0 = \boldsymbol{w}_1'\boldsymbol{z}_1$$

can be written in the form:

$$\hat{Z}_0 = \hat{m}_{GLS} + \boldsymbol{c}_0'\boldsymbol{C}_1^{-1}(\boldsymbol{z}_1 - \hat{m}_{GLS}\mathbf{1})$$

2. We will consider the kriging example with six data points: $\boldsymbol{s}_1 = (0,0), \boldsymbol{s}_2 = (1,1), \boldsymbol{s}_3 = (2,3), \boldsymbol{s}_4 = (3,2), \boldsymbol{s}_5 = (4,1), \boldsymbol{s}_6 = (3,0)$. And the point that we which to predict Z is $\boldsymbol{s}_0 = (2,1)$.

Take $\gamma(||\boldsymbol{h}||) = 1 - \exp(-||\boldsymbol{h}||/2)$.

(a) Using the `krige` function to obtain the kriging variance at $\boldsymbol{s}_0 = (2,1)$. Note: we do not need Z values to obtain kriging variances, but we do need then to use the `krige` function. You can just assign all Z values a 0, so that you can use the function.

(b) To see the effect that kriging locations may have on the ordinary kriging predictor and the kriging variance, repeat (a) for each of four locations labeled as A=(1,0), B=(2,2), C=(3,1), and D=(2,0). Comment on why you think the kriging variances at these four locations are ordered as they are.

(c) Suppose that you which to minimize the kriging variance at $\boldsymbol{s}_{\mathbf{0}} = (2,1)$ and we have sufficient resources to take an observation at any one of the four sites in part (b). Find the kriging variance at $\boldsymbol{s}_{\mathbf{0}} = (2,1)$ corresponding to the addition of each of the sites A, B, C, and D. What is the best additional site? Explain.

3. **Walker Lake Data**. We will perform a spatial analysis of a dataset derived from a digital elevation model from the western United States of the Walker Lake area in Nevada. We will consider a subsample of the original dataset with 470 sites. We will use the variable V, the concentration of a pollutant (in ppm). The data is available in the gstat package using the R command `data(walker)`.

(a) Summarize the distribution of 470 V values with a histogram and any other informative summary statistics. Does the data appear to be normal? Explain. Make a scatter plot of (X,Y) so that you can see the locations where the data has been collected. What do you notice about the sampling scheme of the data. Make a bubble plot of the data and describe any spatial trends and/or patterns that you observe.

(b) Model any trend in the data with a linear trend function. Include a diagnostic and summary plot of the best fitting linear trend function.

(c) Make a plot of the sample semivariogram of the data assuming isotropy. Fit an omnidirectional variogram to the data. Give the estimates of the best fitting variogram. How well does your semivariogram fit the sample semivariogram?

(d) Make plots of the semivariogram in each of 4 (or more) directions. Comment on the nature of each sample semivariogram (i.e. are there nugget effects, etc.). Is there any evidence of anisotropy for these data?

(e) Use the `krige` function and your fitted variogram model from part (c) to obtain kriging predictions and kriging variances for the original data values. Make a histogram of the residuals of the kriging predictions and original data values. What do you conclude? Make a bubble plot of the kriging predictions and kriging variances. Give a summary of the kriging variances. Where are the largest and smallest kriging variances located on the grid? Note: if you use more than just an overall mean in your formula statement, you will be using universal kriging, which is fine for this problem.

(f) Instead of predicting at the original data locations as in part (e), predict at three new sample points (120,180), (40,45), and (215,280). Compute a 95% kriging prediction interval for each of these three points. Compare the three intervals. Can you explain what you observe?