Geography 5023
Mid-Term

1. (2 pts) After conducting a linear regression, the estimated coefficient $\hat{\beta}$ is used to test the null hypothesis $H_0 : \beta = 0$, and the p-value $p = .023$ is obtained. Someone claims that there is a 2.3% chance that the null hypothesis is true. What is wrong with this claim, and how might it be "dangerous." How should the p-value be interpreted?

2. Consider the "true" linear regression model

$$Y_i = \beta_0 + \beta_1 X_{1i} + \beta_2 X_{2i} + e_i$$

for individuals $i = 1 \ldots N$. Under classical assumptions, the vector of ordinary regression coefficients $\hat{\boldsymbol{\beta}}$ has a Gaussian distribution with mean vector $\boldsymbol{\beta}$ and variance-covariance matrix $\sigma_e^2(\mathbf{X}'\mathbf{X})^{-1}$.

   (a) (2pts) Why do we care about specifying a distribution for $\hat{\beta}$ when we only have one sample and one vector of regression coefficients?

   (b) (5 pts) What assumptions are needed in order for the vector $\hat{\boldsymbol{\beta}}$ to have the distribution specified above. Specify these assumptions technically and provide the significance of them for practical application.

   (c) (3 pts) Your scientific theory suggests that the unobserved variable Z_i also has an effect on Y_i. When can Z safely be ignored?

   (d) (2 pts) You suspect that there may be outliers present in your data. If so, which assumption(s) does this suggest may be violated?

3. (2 pts) Define a random field **in your own words and without technical jargon**. Give an example of a random field and describe what characteristics make it a random field.

4. (4 pts) Calculate and plot both the semivariogram and correlogram from lags 0 to 3 for the following data.

$$\begin{bmatrix} 0 & 0 & 1 & 1 & 0 & 0 & 0 & 0 & 1 & 1 \end{bmatrix}$$

5. (5 pts) Comment on the claim that statistical inference is fundamentally inappropriate for the analysis of geographical data. On what grounds might such a claim be made, and what are possible responses?