Strathmore
UNIVERSITY

INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN STATISTICAL SCIENCES
END OF SEMESTER EXAMINATION
STA 8408: SPATIAL STATISTICS

DATE: Thursday, December 13, 2016 Time: 2 Hours

Instructions

1. This examination consists of **FOUR** questions.
2. Answer **Question ONE (COMPULSORY)** and any other **TWO** questions.

Question 1 (20 Marks)

a) Enumerate the key components of the Cressie and Cassie (1993) **classification of spatial statistics**. Describe each component of this classification, explaining (where necessary) any difference between various components of the classification.

(**6 Marks**)

b) Explain the significance of the following concepts in spatial statistics:
   i) "**Geodesy**"
   ii) "***Tobler's first law of geography***"

(**6 Marks**)

c) Let Y_i denote the observed counts of a disease in region $i = 1, \ldots, n$. Assuming that $Y_i|\theta_i \sim Poisson(\theta_i E_i)$, where E_i denotes the expected counts of the disease in the region and θ_i is the risk of the disease in region i.
   i) Show that the Standardized morbidity Ratio $SMR = \frac{Y_i}{E_i}$ is the Maximum Likelihood estimator of θ_i;
   ii) Explain how Bayesian Spatial Hierarchical models can be used to correct for problems that arise in estimating θ_i in using SMRs.

(**8 Marks**)

Question 2

a) The terms "**Isotropy**" and "**anisotropy**" play an important role in Geostatistics. Explain what each term means.

(**4 Marks**)

b) Describe the key components of a **Variogram** and describe how a plot of the variogram can be used to assess the presence or absence of spatial autocorrelation in empirical data.

(**4 Marks**)

d) Explain what the term "**Kriging**" means in Geostatistics. Hence distinguish between Simple Kriging and Ordinary Kriging.

(**4 Marks**)

e) Given a set of observations $Y(s_1), \ldots, Y(s_n)$ at locations $s_1, \ldots, s_n$ in a study region D, derive an expression for the **Ordinary Kriging estimator** of the observation $Y(s_o)$ at site s_o in the study area using the Lagragian multiplier approach.

(**8 Marks**)

Question 3 (20 Marks)

a) Distinguish between the terms "**clusters**" and "**clustering**." By means of appropriate examples, suggest spatial statistical methods that can be used to detect each of the two phenomena in Areal data sets.

(**4 Marks**)

b) **Spatial error**, **Spatial Lag** and **Spatial Durbin** models are common approaches used to model Areal data. Distinguish between these models, presenting an expression for each model in matrix form and providing a justification for its use.

(**6 Marks**)

c) For the spatial lag model [$\boldsymbol{y} = \boldsymbol{X\beta} + \rho\boldsymbol{W}\boldsymbol{y} + \boldsymbol{\varepsilon}$],

i) Derive an expression for the maximum likelihood estimator of parameters in the model;

(**6 Marks**)

ii) Derive an expression for the variance of the vector of parameter estimates $\boldsymbol{\beta}$.

(**4 Marks**)

Question 4 (20 Marks)

a) Let Y_i denote the observed counts of a disease in region $i = 1, \ldots, n$. Consider the poisson-gamma model, where $Y_i|\theta_i \sim Poisson(\theta_i E_i)$ and where $\theta_i \sim Gamma(\alpha, \beta)$. Prove that the posterior mean of the for this model is a weighted average of the data-based SMR for region i and the prior mean μ.

(**7 Marks**)

b) The ***conditional autoregressive*** (CAR)***model*** and the ***intrinsic autoregressive model*** are widely used as prior distribution for random spatial effects in Bayesian models.
Provide a mathematical description of each of these models, explaining how the two models relate to each other.

(**7 Marks**)

c) The ***Cressie*** Model and the ***Leroux model*** are variants of the CAR model.
Provide a mathematical description of each of these models, explaining how models differ from the CAR model.

(**6 Marks**)