STRATHMORE UNIVERSITY
STRATHMORE INSTITUTE OF MATHEMATICAL SCIENCES
MASTER OF SCIENCE IN STATISTICAL SCIENCES
END OF SEMESTER EXAMINATION
STA 8329: MODELLING SOCIAL DATA

Date: 23rd August, 2021. **Time: 3 hours**

Instruction: Answer Question one and any other two questions

Question one [Compulsory]

a) Clearly distinguish the following terms

i. Regression and Causal Inference

ii. Principal Component Analysis and Discriminant Analysis

iii. Cluster Analysis and Mahalanobis distance

(6 marks)

b) Suppose, a latent variable ξ, affects both X and Y. Hence, the true model defines $X = \gamma_1\xi + V$, with Cov $(\xi, V) = 0$, and $Y = \beta X + \gamma_2\xi + \epsilon'$, where Cov $(X, \epsilon') =$ Cov $(\xi, V) = 0$ (We assume that all variables are centered (i.e., deviated from their means), obviating the need for intercept terms). Show that under certain conditions this means either that ξ is a constant for every case, in which case it has no real influence on X or Y, or that ξ has no influence on X, or that ξ has no influence on Y. **(3 marks)**

c) Provide two goals for conducting path analysis **(2 marks)**

d) Describe explicitly a 3-parameter Weibull probability density function as used in life data analysis **(3 marks)**

e) Outline five custom algorithms for machine learning and prediction **(5 marks)**

f) Describe the formula for Proximity Prestige **(5 marks)**

g) Provide clearly 3 uses of Geographical Information System (GIS) in social data models **(6 marks)**

Question two

a) Distinguish the following terms as used in Path Analysis
   i. Structural Equation Modeling and Measurement model
   ii. Latent variable and Observed variable
   iii. Regression path and Loading

(6 marks)

b) Provide three assumptions when undertaking Path Analysis

(3 marks)

c) Elaborate six steps that should be taken when undertaking Path Analysis

(6 marks)

Question three

Consider a bivariate normal distribution

$$(X, Y) \sim Bivariate\ Normal\ (\mu_X; \sigma_X^2;\ \mu_Y; \sigma_Y^2;\ \rho)$$

where ρ is the correlation between X & Y

a) Find the marginal pdf's of X and Y;
b) Prove that X and Y are independent if and only if ρ = 0. (Here ρ is the population correlation coefficient between X and Y.)
c) Find the distribution of $(X + Y)$.
d) Find the conditional pdf of f(x|y), and f(y|x)

(15 marks)

Question four

Describe the following terms explicitly while providing formulas,

a) Conditional Reliability Function
b) Failure Rate Function
c) Mean Life Function (MTTF)
d) Lifetime Distributions
e) Modal Life (or Mode)

(15 marks)

Question five

a) Clearly define a social network model **(2 marks)**
b) Describe the term Walks, Trails, Paths as used in social network **(13 marks)**