GWANDA STATE UNIVERSITY

FACULTY OF ENGINEERING AND THE ENVIRONMENT

DEPARTMENTS OF GEOMATICS AND SURVEYING

PROBABILITY THEORY & STATISTICS

EGS 1111

This examination paper consists of 6 pages

Date:	**March/April 2024**
Total Marks:	**100**
Time:	**3:00 hours**
Examiner's Name:	**Mr. M. Ndlovu**

INSTRUCTIONS

Candidates should attempt **ALL** questions from **Section A** (40 marks) and **ANY THREE** questions from **Section B** (60 marks).

ADDITIONAL MATERIALS

- Calculator
- Statistical Tables

SECTION A: Answer ALL questions [40].

A1. Define the following terms

(a) Map [2]

(b) Mapping [2]

(c) Big Data [2]

A2. Suppose a satellite provides coordinates for the length and width of maps evaluated, denoted as X and Y respectively. Let, A denote the event where the length X falls between 48 and 52 centimeters, i.e., $48 < X < 52$. B denote the event where the width Y falls between 9 and 11 centimeters, i.e., $9 < Y < 11$.

(a) On a xy-plane, plot and shade the regions represented by the following events:

(i) $A \cap B$ (intersection of A and B). [2]

(ii) $A' \cup B'$ (union of the complement of A and the complement of B). [3]

Additionally:

(b) If these events were mutually exclusive, discuss the effectiveness of this mapping. [1]

(c) Evaluate whether the satellite process would produce maps with a length of $X = 50$ centimeters and a width of $Y = 10$ centimeters. [2]

A3. (a) Given the following dataset $\{-2, -2, -2, -3, -3, -3\}$

(i) Write down the formulas for both arithmetic mean and geometric mean. [4]

(ii) Using the given dataset compute the difference between the arithmetic mean and geometric mean. [4]

(b) Find the value of c such that each of the following functions can serve as a probability distribution of the discrete random variable X:

$$f(x) = \left(c^2 - \frac{79}{12}\right) x, \quad \text{for } x = 0, 1, 2, 3.$$

[5]

(c) Marketing estimates that a new instrument for the analysis of soil samples will be very successful, moderately successful, or unsuccessful with probabilities 0.3, 0.6, and 0.1, respectively. The yearly revenue associated with a very successful, moderately successful, or unsuccessful product is \$10 million, \$5 million, and \$1 million, respectively. Let the random variable X denote the yearly revenue of the product. Determine the probability distribution of X. [4]

A4. (a) Consider a survey area where soil samples are collected for analysis. Each soil sample contains various concentrations of different elements. Let's focus on the presence of gold in these soil samples. Assume the following proportions of gold oxidation states found in the soil samples:

Gold Charge	**Propotions Found**
0	0.25
+2	.15
+3	0.40
+4	0.20

Calculate the mean and variance of the gold charge across the soil samples. [4]

(b) Imagine a scenario where the number of rainfall events occurring in a particular region within an hour follows a Poisson distribution. Given that the probability of having no rainfall events in an hour is 0.05, calculate the average number of rainfall events and the variability (variance) in the number of rainfall events within this region. [5]

SECTION B: Answer ANY three questions [60].

B5. (a) The design of a communication system considered the following questions

(i) How many three digit phone prefixes that are used to represent a particular geographic are (such as area code) can be created from the digits 0 through 9? [2]

(ii) As in part (i), how many three-digit phone prefixes are possible that do not start with 0 or 1, but contain 0 or 1 as the middle digits? [2]

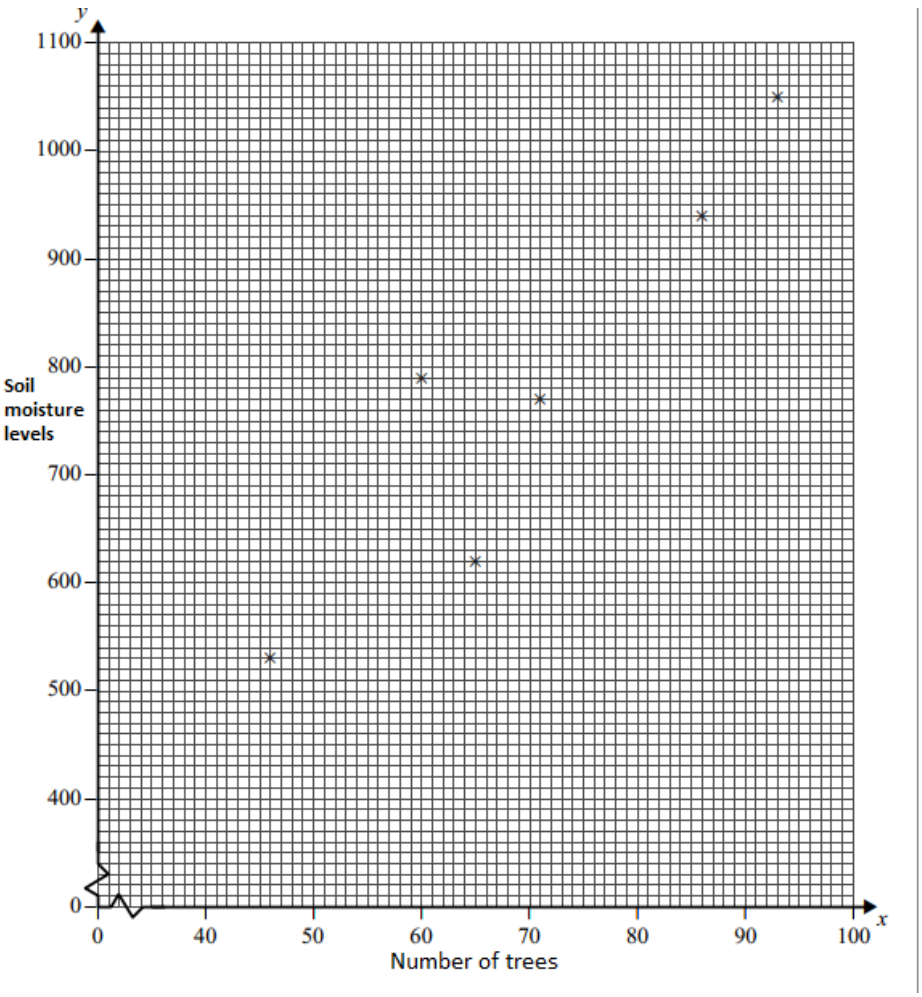
(iii) How many three-digit phone prefixes are possible in which no digit appears more than once in each prefix? [2]

(b) A major west coast city provides one of the most comprehensive emergency medical services in the world. Operating in a multiple hospital system with approximately 20 mobile medical units, the service goal is to respond to medical emergencies with a mean time of 12 minutes or less and standard deviation of 3.2 minutes.The director of medical services wants to formulate a hypothesis test with a .05 level of significance, to determine whether or not the service goal of 12 minutes or less is being achieved. [12]

B6. Suppose we have collected data on the number of trees (x) and the soil moisture levels (y) in a forested area during 10 randomly selected days:

x No of Tress	86	60	65	46	71	93	56	81	75	57
y Moisture Levels	940	790	620	530	770	1050	690	780	860	550

(a) The first 6 pairs of data values in this table are plotted on the scatter diagram shown below. Plot the final 4 pairs of data values on the scatter diagram. [4]



(b) (i) Define the terms in a linear regression model and explain their significance in spatial analysis. [5]

(ii) Compute the equation of the least squares regression line in the form $y = a + bx$ and plot the line on the scatter diagram. [10]

(iii) Interpret the value of b in the context of the relationship between tree density and soil moisture levels. [2]

(iv) Discuss why the value of a may not have a practical interpretation in this geostatistical context. [2]

(c) Estimate, to the nearest 10, the expected soil moisture level when the number of trees is 50. [2]

B7. (a) Complete the following general ANOVA Table for (**a** - **g**):

Source of Variance	Sum of Square	Degree of freedom	**a**	F-value
Treatments = A	SSA	$k-1$	**f**	$\frac{MSA}{MSE}$
Blocks	SSB	$j-1$	$\frac{SSB}{j-1}$	**g**
b	**c**	**d**	**e**	
Total	SST	N-1		

[7]

(b) First year collage students were randomly assigned to three groups to experiment with three different methods of teaching Geostatistics. At the end of the semester, the same test was given to all 15 students. The table gives the scores of students in the three groups

Method I	Method II	Method III
48	55	
73	85	68
51	70	95
65	69	74
	90	67

(i) At the 10% significance level, will you reject the null hypothesis that the three teaching methods in geostatistics class are all the same? [10]

(ii) Hence, compute and intepret the p-value for the F-value in (i). [3]

B8. (a) Suppose all the data point are along the same line and there are equally spaced

u	1	2	3	4	5	6	7	8	9	10
$Z(u)$	41.2	40.2	39.7	39.2	40.1	38.3	39.1	40.0	41.1	40.3

with the variogram given by:

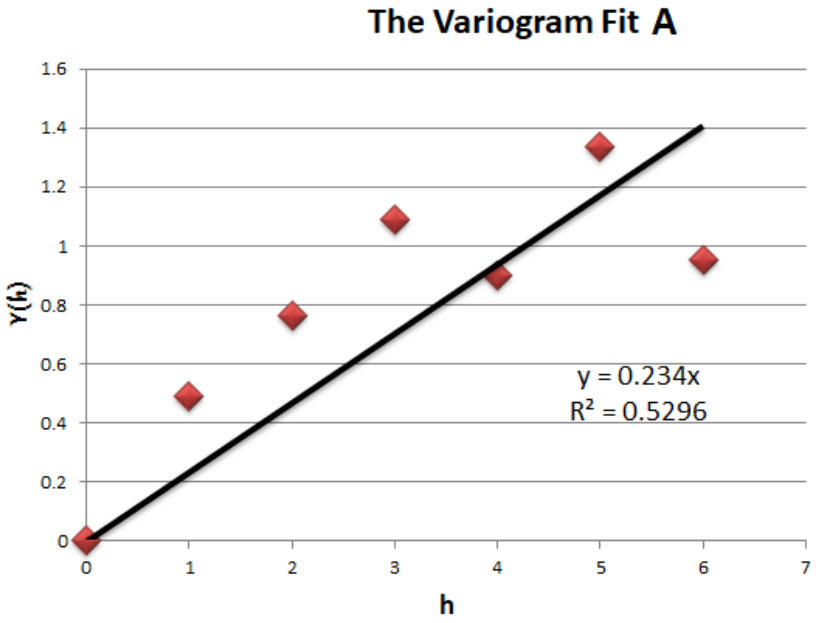
$$\gamma(h) = \frac{1}{2N(h)} \sum_{u_i - u_j = h} \big(Z(u_i) - Z(u_j)\big)^2$$

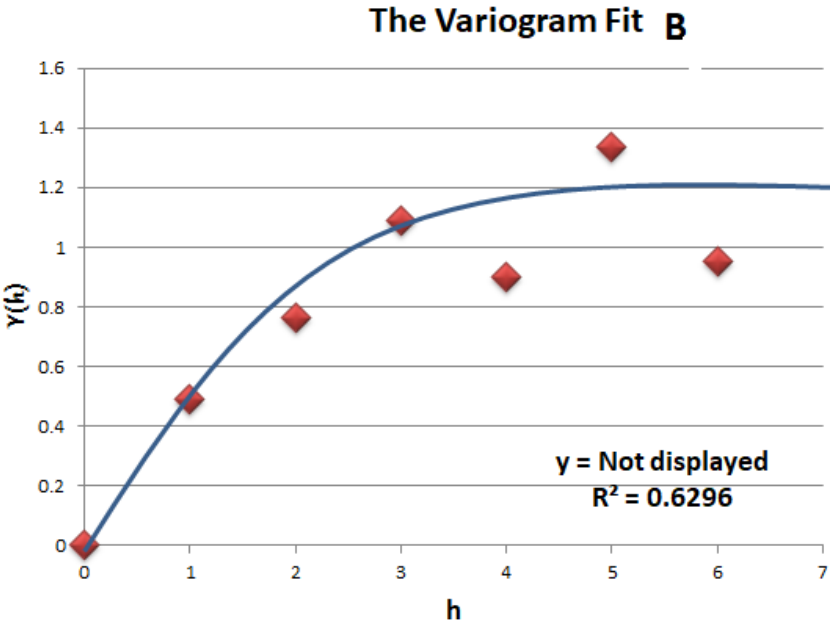
(i) Complete the following table giving your value to 2 decimal places:

h	1	2	3	4	5	6	7	8	9
$\gamma(h)$									

[9]

(ii) Two plots for the variograms were obtained from the table in (i)





Using terms from geostatistics compare and contrast these two diagrams (variogram fit A with variogram fit B). [5]

(iii) Comment on the theoretical variogram model that can be fitted from the information obtained in (i) and (ii) above. [4]

(iv) State which plot contains a more accurate fitted model? [2]

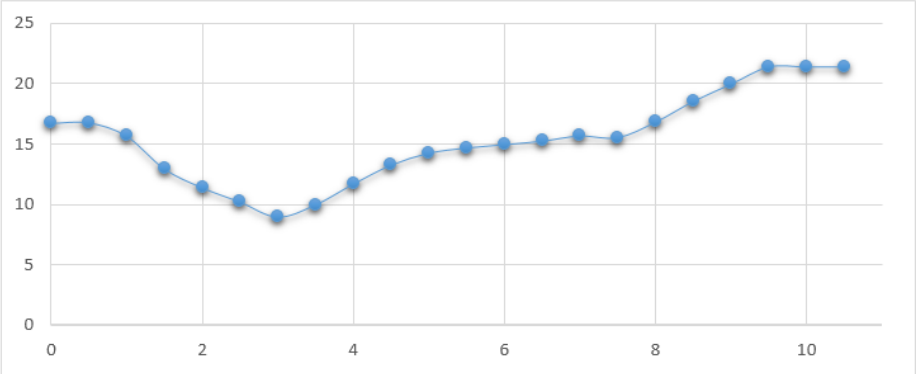
B9. In this study we are going to apply curve fitting technique and technological software to elaborate the application of mathematics in real world sceneries around Gwanda State University. Curve fitting is the process of constructing a curve, or mathematical function, that has the best fit to a series of data points, possibly subject to constraints. Comment on the following stages of this study presentated by these graphs.

(a) Stage 1: Data collection [3]

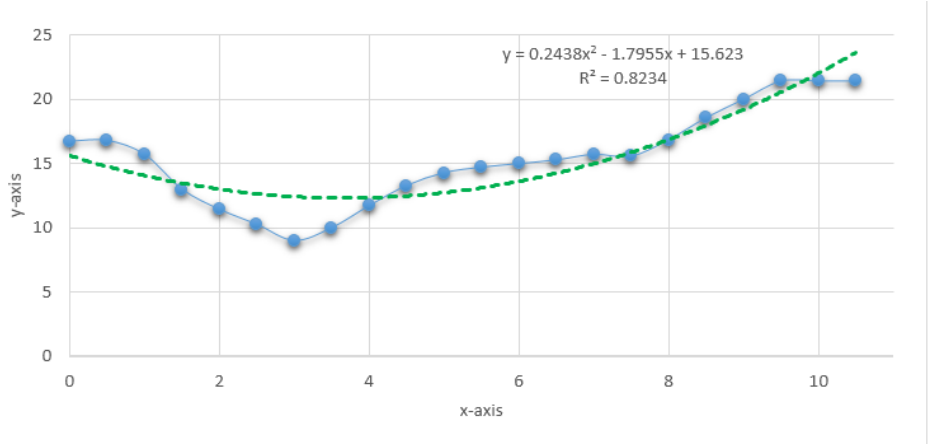
(b) Stage 2: Data cleaning or data transformation [3]

x	0	0.5	1	1.5	2	2.5	3	3.5	4	4.5	5	5.5	6	6.5	7	7.5	8	8.5	9	9.5	10	10.5
y	16.75	16.8	15.71	12.99	11.42	10.25	8.99	9.99	11.71	13.28	14.28	14.71	14.99	15.28	15.71	15.57	16.85	18.57	20	21.43	21.43	21.43
y top down	3.25	3.2	4.29	7.01	8.58	9.75	11.01	10.01	8.29	6.72	5.72	5.29	5.01	4.72	4.29	4.43	3.15	1.43	0	-1.43	-1.43	-1.43
y down up	16.75	16.8	15.71	12.99	11.42	10.25	8.99	9.99	11.71	13.28	14.28	14.71	14.99	15.28	15.71	15.57	16.85	18.57	20	21.43	21.43	21.43

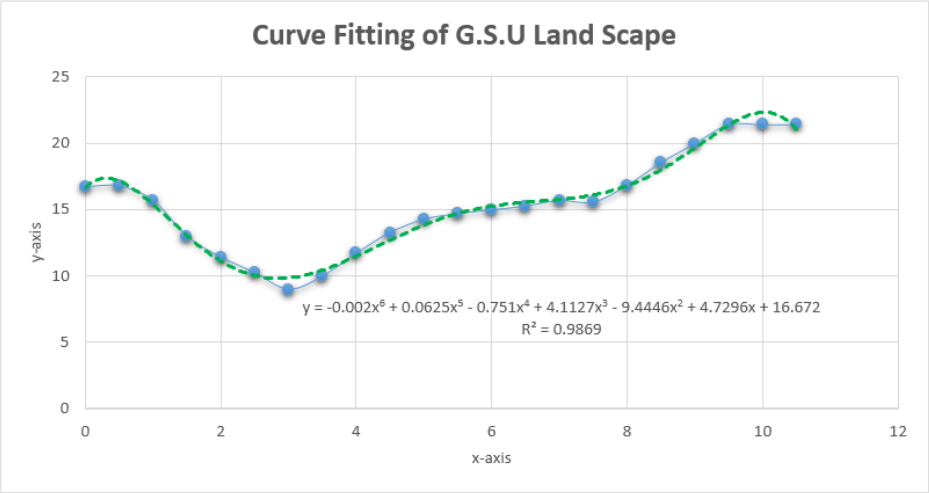
(c) Stage 3: Scatter Plot [2]



(d) Stage 4: Curve fitting (i) [3]



(e) Stage 5: Curve fitting (ii) [3]



(f) Stage 6: Stage Compare different fitted curves and conclusion [4]

END OF QUESTION PAPER