GWANDA STATE UNIVERSITY

FACULTY OF ENGINEERING AND THE ENVIRONMENT

DEPARTMENTS OF GEOMATICS AND SURVEYING

PROBABILITY THEORY & STATISTICS

EGS 1111

This examination paper consists of 5 pages

Date:	**June 2023**
Total Marks:	**100**
Time:	**2:30 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

- Graph Paper
- Calculator
- Statistical Tables

SECTION A: Answer ALL questions [40].

A1. Define the following terms

(a) Population [2]

(b) Geostatistics [2]

(c) Spatial Data [2]

A2. The co-ordinate from a satellite gives the following model, the length and width, denoted as X and Y, respectively, of each map evaluated. Let
A denote the event of $48 < X < 52$
B denote the event of $9 < Y < 11$
Plot and shade the regions represented by the following events in a xy-plane

(a) $A \cap B$ [2]

(b) $A' \cup B'$ [3]

(c) i If these events were mutually exclusive, how successful would this mapping be? [1]

ii Would the satellite process produce maps with $X = 50$ centimeters and $Y = 10$ centimeters? [2]

A3. (a) Given the following dataset $\{8, 1, -2, 3, -3, 5\}$

(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) Determine the value c so that each of the following functions can serve as a probability distribution of the discrete random variable X:
$f(x) = c(x^2 + 4)$, 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) In a $NiCd$ battery, a fully charged cell is composed of nickelic hydroxide. Nickel is an element that has multiple oxidation states. Assume the following proportions of the states:

Nikel Charge	Propotions Found
0	0.17
+2	.35
+3	0.33
+4	0.15

Determine the mean and variance of the nickel charge. [4]

(b) Suppose that the number of customers who enter a store in an hour is a Poisson random variable, and suppose that $P(X = 0) = 0.05$.
Determine the mean and variance of X. [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. (a) What is the other name given to Chi-Square Experiment? [2]

(b) State the mean for Chi-Square Experiment. [2]

(c) Outline the similarities and differences between a Chi-Square distribution and the Binomial distribution? [4]

(d) In 2021 the GIS research team at Gwanda State University performed a study the opinions of ArcGis use by students. The percentage distribution of their responses is shown in the following table.

Opinion	Percentage
A: GIS mapping is something that is just at its infancy, and its use will grow exponentially over the next few years	45
B: GIS mapping is something that most student and the staff members will use, but it will not move more into the mainstream	21
C: GIS mapping is already over, and it is time to find the next best thing	17
D: I do not know enough about GIS mapping to have an opinion	17

Recently (in 2022) 800 randomly selected students were asked the same question. The following table lists the number of students in this sample who gave each response.

Opinion	Frequency
A: GIS mapping is something that is just at its infancy, and its use will grow exponentially over the next few years	374
B: GIS mapping is something that most student and the staff members will use, but it will not move more into the mainstream	183
C: GIS mapping is already over, and it is time to find the next best thing	127
D: I do not know enough about GIS mapping to have an opinion	116

Test at the 2.5% level of significance whether the current distribution of opinions is different from that for 2022. [12]

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) The derivation of Least Squares Method approach involves the minimization of function:

$$\Phi(a,b) = \sum_{i=1}^{n} (a + bx_i - y_i)^2$$

(i) How do we obtain a and b to minimize $\Phi(a,b)$? [2]

(ii) Determine the unknown a and b. [6]

(b) The dataset for unknown location can be used to contract a simple linear regression model to express drain current y (in milliamperes)as a function of ground-to-source voltage x (in volts). The data are as follows:

y	x	y	x
.734	1.1	1.50	1.6
.886	1.2	1.66	1.7
1.04	1.3	1.81	1.8
1.19	1.4	1.97	1.9
1.35	1.5	2.12	2.0

(i) State regression model. [4]

(ii) Fit a simple linear regression model to this data [8]

B9. State, compare and contrast between two main hypothesis in geostatistics. [20]

B10. (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
$\gamma(h)$									

[12]

(ii) Plot the variogram using the data in (i). [5]

(iii) Comment on the theoretical variogram model that can be fitted from the information obtained in (i) and (ii) above. [3]

END OF QUESTION PAPER