

UNIVERSITY EXAMINATIONS

SECOND SEMESTER 2023/2024 ACADEMIC YEAR

**SECOND YEAR EXAMINATION FOR THE DEGREE OF
BACHELOR OF SCIENCE (STATISTICS & ECONOMICS)**

STAT 214: STATISTICAL COMPUTER PROGRAMMING

STREAM: R

TIME: 2 HRS

DAY: THURSDAY[11.30AM-1.30PM] DATE: 18/04/2024

THIS QUESTION PAPER CONSISTS OF FOUR (4) PAGES

PLEASE DO NOT OPEN UNTIL THE INVIGILATOR SAYS SO

Instructions

Answer question **ONE** and any other **TWO** questions.

QUESTION ONE [30 Marks]

- a) Define the following terms as used in computer programming [3 Marks]
 - i) Statistical computer programming.
 - ii) R scripts
 - iii) Syntax
- b) Identify any three features that high level programming languages possess [3 Marks]
- c) Write down the valuation results of the following R expressions [8 Marks]
 - i) $1:8-1:2$
 - ii) $A <- \text{matrix}(c(2,4,8,9,11,13,15,16,17,20,22,26), \text{nrow} = 3, \text{ncol} = 4, \text{byrow} = \text{TRUE})$
 - iii) $B <- \text{array}(1:24, c(3, 4, 2))$
 - iv) $\text{rep}(\text{seq}(4,16, \text{by}=4),3)$
- d) Perform conversion of the following number systems [4 Marks]
 - 11010_B to decimal and
 - 40_D to hexadecimal
- e) State any three important features for data structures? [3 Marks]
- f) Explain the following two R codes with respect to how they generate random variates [4 Marks]
 - $x <- \text{pnorm}(4)$
 - $y <- \text{rnorm}(20,10,2)$
- g) Consider the following table of random variates and fill in the missing entries (?) [5 Marks]

Distribution	R code	Parameters
Binomial	?	size,prob
Poisson	rpois	?
?	rnorm	?
Chi-square	?	df,ncp

QUESTION TWO [20 Marks]

- a) What do you understand by the term data frame? Describe its basic structure [4 Marks]
- b) Consider the following table and filling in the missing entries (#) [4 Marks]

Data type	Description	Example
Numeric	#	#
#	Categorical variable	#
- c) In a follow up study on breast cancer in country A to evaluate whether patients in public hospitals had better survival gave mean lifetime (ml) as $ml \pm \frac{1.99*sd \text{ for age}}{\sqrt{\text{No. of iterations}}}$
Write program in R that computes 95% confidence interval ($\alpha = 0.01$) [4 Marks]



- d) Suppose that in a study to determine if there exists any relationship between length and gestational age of low birth weight infants in a population, the findings were as shown in the table below.

Gestational age (weeks)	24	26	28	30	32
Length (centimeters)	26	30	34	38	42

With brief and appropriate explanations, write down R codes that would;

- i) Fit a simple linear regression model [4 Marks]
 ii) Determine SSE using for loop. [4 Marks]

QUESTION THREE [20 Marks]

- a) What is meant by power of a test? [3 Marks]
 b) Write down a program that access the three variables below, compute their average and returns the mean value. [5 Marks]

is_disabled_w10, FRR_w10 and is_disabled_w8

- c) Write a program in R that generates ten variates given $y_i = \beta_0 + \beta_1 x_i + e_i$ where $x_i \sim N(9, 0.2)$, $e_i \sim N(0, 0.5)$, $\beta_0 = 4.6$, $\beta_1 = 2.8$ $y_i = \beta_0 + \beta_1 x_i + e_i$ [5 Marks]

- d) Write a function in R for determining the standard deviation of any given set of numbers from

first principles where, $sd = \sqrt{\frac{\sum_{i=1}^n (x_i - \bar{x})^2}{n}}$, hence determine the standard deviation of (5,4,-9,6,9,8,5,3,6,9) [7 Marks]

QUESTION FOUR [20 Marks]

- a) State any **four** properties of good random number generators. [4 Marks]
 b) Generate six random numbers using each of the following pseudo-random number generators methods i) Midsquare method, given $z_0=7082$ [5 Marks]
 ii) Linear congruential method, given $z_0=7$, $a=5$, $c=3$ and $m=16$. [6 Marks]
 c) Give and explain R codes that will generate five uniform pseudo-random numbers on the interval [0, 1], and ten uniform such numbers on the interval [-3, -1] [5 Marks]

QUESTION FIVE [20 Marks]

- a) Explain briefly the concepts of Computer simulation and Monte Carlo Simulation. [4 Marks]
 b) Describe a scenario where Monte Carlo simulation has often been used. [3 Marks]
 c) Suppose in a University campus where students are contracted by academic departments to type documents, the rate of production (number of pages) is generated as shown by the discrete random variables table below. These pages costs \$6.0 each and variable cost for producing each is \$2.5 whereas spoiled pages must be disposed at the cost of \$0.10 per page.

Further that;

Total production cost as produced*unit production cost.

If more pages are produced than demanded, the number of units spoiled = produce - demand. otherwise, more units are spoiled.

Disposal costs are determined as $\text{UnitDispCost} * \text{if}(\text{produced} > \text{demand}, \text{produced} - \text{demand}, 0)$

Profit is determined as revenue - total variables cost- totalDispCost.

Rate of production	1500	2000	2500
Probability	0.20	0.30	0.50

On the basis of above information write down R program that determines maximum profit.

[13 Marks]

