****

**GARISSA UNIVERSITY**

**UNIVERSITY EXAMINATION 2019/2020 ACADEMIC YEAR TWO**

**SECOND SEMESTER EXAMINATION**

**SCHOOL OF SCHOOL OF PURE AND APPLIED SCIENCES**

**FOR THE DEGREE OF BACHELOR OF EDUCATION**

**COURSE CODE: STA 224**

**COURSE TITLE: COMPUTING METHODS AND DATA ANALYSIS I**

**EXAMINATION DURATION: 2 HOURS**

**DATE: 18/12/2020 TIME: 03.00-05.00 AM**

**INSTRUCTION TO CANDIDATES**

* **The examination has FIVE (5) questions**
* **Question ONE (1) is COMPULSORY**
* **Choose any other TWO (2) questions from the remaining FOUR (4) questions**
* **Use sketch diagrams to illustrate your answer whenever necessary**
* **Do not carry mobile phones or any other written materials in examination room**
* **Do not write on this paper**

**This paper consists of THREE (3) printed pages *please turn over***

**QUESTION ONE (COMPULSORY)**

1. State the function you can use to get a first look at data. (2 Marks)
2. State the functions that can access the number of rows, number of columns and the dimension of a data frame directly. (3 Marks)
3. Define what we mean by simple linear regression model (4 Marks)
4. Define the coefficient of determination stating the formula. (4 Marks)
5. Describe the functionality of the following commands
6. *lm()* command. (3 Marks)
7. *abline()* command (3 Marks)
8. *predict()* commamd (3 Marks)
9. State three functions that can be used to access stored information in a model

(3 Marks)

1. State the Gauss–Markov theorem. (3 Marks)
2. Explain what we mean by unbiased estimates. (2 Marks)

**QUESTION TWO (20 Marks)**

1. Describe the general naming structure of three relevant R functions (6 Marks)
2. Consider a random variable $X$ which is $N(μ=2, σ^{2}=25)$.
3. Write R syntax that calculates the value of the pdf at $x=3$. (2 Marks)
4. Write R syntax calculates the quantile for probability 0.975. (2 Marks)
5. Write R syntax to generate a random sample of size n = 10. (2 Marks)
6. Define residual as used in simple linear regression (2 Marks)
7. We would like to model the relationship between $X$ and $Y$ using the form $Y=f\left(X\right)+ϵ$. If $Y$ can be taken as the response state three ways in which it can be written. (6 Marks)

**QUESTION THREE (20 Marks)**

1. Given observations $(x\_{i}, y\_{i})$, for $i=1,2, …, n$, derive the values of $β\_{0}$ and $β\_{1}$ which minimize $f\left(β\_{0}¸β\_{1}\right)=\sum\_{i=1}^{n}\left( y\_{i}-(β\_{0}+β\_{1}x\_{i})\right)^{2}$ (6 Marks)
2. Write down R syntax to compute $S\_{xy}$, $S\_{xx}$ and $S\_{yy}$ (6 Marks)
3. Write down R syntax to compute Sum of Squares Total (SST), Sum of Squares Regression (SSR), Sum of Squares Error (SSE) and store the three of them in one vector.(8 Marks)

**QUESTION FOUR (20 Marks)**

1. Derive the maximum likelihood estimates a linear regression model. (10 Marks)
2. We would like to simulate $n=100$ observations from the model $Y\_{i}=β\_{0}+β\_{1}x\_{i}+ ϵ\_{i}$

where $ϵ\_{i}∼N(0, σ^{2})$, $β\_{0}=3$ , $β\_{1}=6$, $σ^{2}=4$.

1. Initialize the parameter required in R (3 Marks)
2. Write R syntax to set a seed for randomization and calculate $S\_{xx}$ (3 Marks)
3. Write R syntax to compute the sampling distributions (4 Marks)

**QUESTION FIVE (20 Marks)**

1. We would like to simulate $n=21$ observations from the model $Y=5-2x+ϵ$.

Let $ϵ∼N(μ=0, σ^{2}=9)$

1. Initialize all the parameters required for simple linear regression model computation in R. (4 Marks)
2. Obtain simulated values of $ϵ\_{i}$ by setting a seed for reproducibility (3 Marks)
3. Generate the $y$ values according the specified functional relationship

(3 Marks)

1. Fit the model to the simulated data estimating coefficients. (3 Marks)
2. Plot the model with well labeled axes and legends. (7 Marks)