****

**GARISSA UNIVERSITY**

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

**SECOND SEMESTER EXAMINATION**

**SCHOOL OF PURE AND APPLIED SCIENCES**

**FOR THE DEGREE OF BACHELOR OF EDUCATION**

**COURSE CODE: STA 318**

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

**EXAMINATION DURATION: 2 HOURS**

**DATE: 17/02/2020 TIME: 09.00-11.00 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. Write down a syntax that creates a sequence that runs from 1.2 to 6.6 increasing by 0.3 at a time stating the output. **[3 marks]**
2. Create a 3 by 4 matrix with entries 1 to 12 and state the output **[5 marks]**
3. Write an R code that computes $3^{2}$, $2^{(-3)}$, $100^{\frac{1}{2}}$. **[3 marks]**
4. State three data types giving examples. **[6 marks]**
5. State two homogeneous data structures and two heterogeneous data structures **[4 marks]**
6. State the logical operators used below and the computer output. **[6 marks]**

|  |
| --- |
| $$3<42$$ |
| $$3>=42$$ |
| $3<4$ & $42>13$ |

1. Write R code to create two vectors as well as a code to test if the vectors are matrices **[3 marks]**

**QUESTION TWO**

1. State three data types giving examples in each case . **[6 marks]**
2. Suppose a grocery store sells “16 ounce” boxes of captain crisp cereal. A random sample of 9 boxes was taken and weighed. The company that makes the Captain Crisp cereal claims that the average weight of a box is at least 16 ounces. Assume the weight of cereal in a box is normally distributed and use a 0.05 level of significance to test the company’s claim.
3. Write down a data frame called Capt\_crisp to store the data on the weights **[4 marks]**
4. Which test statistic is used here? **[2 marks]**
5. Write a syntax to compute the mean and standard deviation by creating variables to store the hypothesized and the sample size. **[4 marks]**
6. The R syntax to compute the test statistic. **[2 marks]**
7. Write the syntax to obtain the p-value of the test. **[2 marks]**

**QUESTION THREE**

Assume that the distributions of X and Y are $N\left(μ\_{1}, σ^{2}\right)$ and $N\left(μ\_{2}, σ^{2}\right)$, respectively. Given the n=6 observations of X (70, 82, 78, 74, 94, 82), and the m=8 observation for Y (64, 72, 60, 76, 72, 80, 84, 68), we would like to test the hypothesis $H\_{0}:μ\_{1}=μ\_{2}$ versus $H\_{1}:μ\_{1}>μ\_{2}$.

1. Write a syntax to store the values of X and Y and compute their respective lengths. **[4 marks]**
2. Write a code to calculate and store their respective means and standard deviations **[4 marks]**
3. Write a code to compute the pooled standard deviation. **[4 marks]**
4. State the test statistics used here and write an R code to compute the test statistic. **[4 marks]**
5. Write a code to compute the p-value **[4 marks]**

**QUESTION FOUR**

1. Consider a random variable X which is $N\left(μ=2, σ^{2}=25\right)$. Write down the R code to compute
2. The pdf at $x=3$. **[2 marks]**
3. The cdf at $x=6$. **[2 marks]**
4. The quantile probability 0.975. **[2 marks]**
5. The random sample of size $n=10$. **[2 marks]**
6. Describe three general naming structure of the relevant R functions in probability and statistics. **[6 marks]**
7. State two ways of creating a heterogeneous data structure with respective syntaxes  **[6 marks]**

**QUESTION FIVE**

Consider the following table of vehicles

|  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Manufacturer | Audi | Honda | Toyota | Volkswagen | Audi | Honda | Toyota | Volkswagen | Honda |
| Model | A4 | Civic | Corolla | Jetta | A4 | Civic | Corolla | Jetta | Civic |
| Displacement | 1.8 | 1.8 | 2.2 | 2.8 | 2.8 | 3.1 | 1.8 | 1.8 | 2 |
| Year | 1999 | 1999 | 2008 | 1999 | 1999 | 2008 | 1999 | 1999 | 2008 |

1. State the data type for each of the factors considered . **[4 marks]**
2. Write down the syntax for computing mean, median, variance, standard deviation, inter-quartile range, minimum, maximum, range and covariance of the displacement  **[10 marks]**
3. Write down the syntax with well labeled axes for computing the histogram and bar-plots for the displacement. **[6 marks]**