****

**JARAMOGI OGINGA ODINGA UNIVERSITY OF SCIENCE AND TECHNOLOGY**

**SCHOOL OF MATHEMATICS AND ACTURIAL SCIENCE**

**UNIVERSITY EXAMINATION FOR DEGREE OF BACHELOR OF SCIENCE ACTURIAL**

**3rd YEAR 2nd SEMESTER 2016/2017 ACADEMIC YEAR**

**MAIN REGULAR**

**COURSE CODE: SAS 312**

**COURSE TITLE: STATISTICAL COMPUTING II**

EXAM VENUE: STREAM: (Bsc. **Actuarial Science with IT)**

DATE: EXAM SESSION:

TIME: 2.00 HOURS

**Instructions:**

1. **Answer questions one and any other two only.**
2. **Candidates are advised not to write on the question paper.**
3. **Candidates must hand in their answer booklets to the invigilator while in the examination room.**

**QUESTION ONE (30 MARKS)**

(a) Explain the following giving examples in each case

1. Nominal and ordinal scales of measurements (2 marks)
2. Numeric and character data types (2 marks)

(b) List four advantages of R programming software (4 marks)

(c) In R, how do you import data in the following file formats

1. Csv (1 mark)
2. Stata (1 mark)
3. SPSS (1 mark)

(d) Provide R-codes used to generate random numbers from the following distributions

1. Normal (1 mark)
2. Exponential (1 mark)
3. Gamma (1 mark)
4. Poisson (1 mark)
5. Binomial (1 mark)

(e) List any six functions that R provides (6 marks)

(f) Explain how R commands are written (1 mark)

(g) How can one save and or write data in csv and stata file formats with R (2 marks)

(h) Illustrate four data structures in R that are used to perform statistical analysis and create graphs

(4 mark)

(i)In R, how missing values are represented (3 mark)

**QUESTION TWO (20 Marks)**

The following data represent the body-breadth (x, in cm) and body-weight (y, in cm) of 14 randomly selected sea fishes.

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| x | 0.5 | 0.6 | 0.8 | 0.4 | 0.5 | 0.7 | 1.0 | 1.0 | 0.6 | 0.7 | 1.5 | 0.5 | 0.5 | 0.6 |
| y | 10 | 15 | 25 | 12 | 15 | 14 | 25 | 28 | 18 | 20 | 40 | 18 | 15 | 20 |

1. Draw a scatter diagram providing an R-code (3 marks)
2. At $α$=0.05, examine whether body-weight and body-breadth of fishes are significantly correlated (5 marks)
3. Provide an R-code to obtain spearman and Pearson correlation coefficients (2 marks)
4. Fit a simple linear regression of y on x extracting regression slope and intercept (8 marks)
5. Write an R-code that gives a summary and ANOVA of the fit in (iv) (2 marks)

**QUESTION THREE (20 Marks)**

(a)The following data are the random sample of observations recorded from an industry producing juice per hour in different days

xi : 50,55,62,67,45,68,70,62,73,64,75,55,50,68,64,60,66,60,56,59,60,60,63,67,66,68,70,65,54,55,70,66,67

1. Plot a boxplot to explore the above data (5 marks)
2. Under the assumption xi $\~ N(μ, σ^{2})$, test the significance of

H0: $μ=60$ against H1: $μ\ne μ\_{0}$ at $α=0.05$ (5 marks)

1. Provide R-codes (i) and (ii) (2 marks)

(b)If the number of boys and girls who are regular in their exercises is distributed as given below, it is required to test whether there is statistical difference in the exercise habits between boys and girls.

|  |  |  |  |
| --- | --- | --- | --- |
| Habit | Boys | Girls | Total |
| Exercise regularly | 2 | 8 | 10 |
| Do not exercise regularly | 10 | 4 | 14 |
| Total | 12 | 12 | 24 |

1. Use an appropriate Chi-square test to assess whether exercise regularly is associated with sex at 95% confidence level. (5 marks)
2. Provide an R code that can generate the results in (i). (3 marks)

**QUESTION FOUR (20 Marks)**

(a)A sample of 200 tribal people is selected and the people are classified according to different sub-tribes. The number of people in sub-tribes is shown below

|  |  |  |
| --- | --- | --- |
| Number of people | People in Sub-tribes | Total |
| 1 | 2 | 3 | 4 | 5 | 200 |
| 20 | 45 | 18 | 36 | 81 |

(i)At $α=0.05$,test whether the proportions in different sub-tribes are homogeneous. Show all your workings. (8 marks)

(ii) Write an R-code for doing the analysis in (i) above. (2 marks)

(b)A researcher in the field of Botany has collected some leaves from a forest in 4 occasions. The collected leaves are classified into 2 classes according to the species A and B.

|  |  |  |
| --- | --- | --- |
| Species | Occasion | Total |
| 1 | 2 | 3 | 4 |
| A | 5 | 12 | 8 | 10 | 35 |
| B | 30 | 36 | 22 | 45 | 133 |
| Total | 35 | 48 | 30 | 55 | 168 |

(i)Test whether the proportions of species A and B in each occasion same at $α=0.05$ (8 marks)

(ii)Write an R-code for doing the analysis in (i) above. (2 marks)

(c)

**QUESTION FIVE (20 Marks)**

(a)Suppose a binomial probability mass function is given by

$$f\left(x\right)=\left\{\begin{array}{c}\left(\genfrac{}{}{0pt}{}{n}{k}\right)p^{x}(1-p)^{n-x};x=0,1,2, \cdots , n\\0; elsewhere\end{array}\right.$$

(i)Show that E(x)=np and var(x)=npq (8 marks)

(ii)Give an R-code that would be used to generate 100 random numbers when p=0.68 and n=120. (2 marks)

(b)Show that for a Poisson distribution, E(x)=var(x)=$λ$ (8 marks)

(ii) Give an R-code that would be used to generate 100 random numbers when $λ$ = 0.24 (2 marks)