Biostatistics and R Programming Lab

Time: 2 hours **Maximum Marks: 35**

Multiple choice questions (attempt any 5 out of 7) **(5×7 Marks)**

For each question, write an R-script whose name is mentioned below the question.

1. (a) For $x = 12.5$, $y = 22.8$ and $z = 34.6$, compute the following expression in R and print the result:

$$L = \frac{\sqrt{x^3 + y^3 + z^3} * e^{\frac{x}{y}}}{log(\frac{y+z}{x})}$$

(b) A population dynamics model with predation can be written in a dimensionless form as

$$f(x) \ = \ \alpha x(1 - \frac{x}{K}) - \frac{x}{1+x}$$

where α is the growth rate and K is the carrying capacity. Let us choose $K = 20$.

For $\alpha = 1.3$ and $\alpha = 0.3$, plot 2 separete curves of $f(x)$ vs x on the same plot using R.

R-script name : Question-1.r

2. In the attached data file called "test_data.csv", four response variables for a pain index are measured for 100 patients. Write an R script to compute the following:

(a) Read the file into a data frame called "mydat"

(b) create a scatter plot between "pain_index" and "medicine_response".

(c) Compute the Pearson's correlation coefficient between "pain_index" and "medicine_response" and print the value.

(d) Create a subset of "mydat" for "index_value" in the range 50 to 95 and write it into a file called "subset.csv"

R-script name : Question-2.r

3. (a) Plot the Binomial probability density function for $n = 14$ and $p = 0.9$ with the following specifications:

Color : blue
X-axis title = "Number of Trials"
Y-axis title = "Binomial probability for n=14, p=0.9"

(b) Generate 10000 random data points from a Poisson distribution of mean=7. Plot the frequency histogram in the range [0, 14]

R-script name : Question-3.r

4. In an experiment, two sets of observations were made on the growth of pea stem segments. The data for the 2 sets X and Y are given below:

X : 0.8, 1.8, 1.0, 0.1, 0.9, 1.7, 1.0, 1.4, 0.9, 1.2, 0.5
Y : 1.0, 0.8, 1.6, 2.6, 1.3, 1.1, 2.4, 1.8, 2.5, 1.4, 1.9, 2.0, 1.2

Perform an unpaired t-test assuming the equality of the means to test the equality of μ_X and μ_Y and get the p-value.

Use your own R Script for the test, without the R library function "t.test()"

Print your null hypothesis and alternate hypothesis clearly.

At a significance level of 0.05, do you reject or accept the null hypothesis?. Print the conclusion.

R-script name : Question-4.r

5. Two colleges are conducting computational biology courses for last 6 years. Every year, both of them enroll 80 students feach for the course, and few of them drop out in the middle of the course. The data on the number of students who dropped every year is given below for both the colleges:

College-1	6	8	2	4	4	5
College-2	7	10	4	3	5	6

(a) Perform a Wilcoxon Rank Sum test on this data to see whether the mean number of students who dropped out of the course over the years is the same for the two colleges. Use the R library function "wilcox.test()" for this.

(b) Test the same hypothesis by performing a Welsch t-test on the data using the R library function "t.test()".

For both the test, print the results and your conclusions clearly.

R-script name : Question-5.r

6. An experiment was conducted to study the number of days a sick mouse survived after being administered a particular medicine. Four medicines were tried, each on 6 mice. The data is presented below:

Medicine-A :	10	9	8	11	12	10
Medicine-B :	6	7	8	6	5	5
Medicine-C :	8	7	9	10	9	9
Medicine-D :	11	10	12	13	11	12

Using the method of ANOVA, compare the mean effect of different medicines on the survival time of the sick mice. State the null hypothesis, alternate hypothesis and the conclusion clearly. Let $\alpha = 0.05$ for the test. You can use any of the R library functions available for performing ANOVA.

R-script name : Question-6.r

A7. In a laboratory experiment, a quantity Y was measured as a function of an independent quantity X. The data is presented here:

X	1.5	2.5	3.5	4.5	5.5	6.5	7.5
Y	13.1	15.4	22.2	23.4	24.9	32.0	30.9

(a) Fit a straight line of the form $Y = a + bX$ to determine its slope and intercept. You can use R library functions for linear model.

(b) Plot the original data points and the fitted curve on the sample plot with different colors.

(c) Print the expected and the observed Y values for the X values.

R-script name : Question-7.r