Biostatistics and R Programming Lab

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

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) Create a sequence of number from 100 to 120. To each number in this sequence, add a string "_product" using "paste" command. (eg, 100_product, 101_product,....,120_product) and print these strings.

(2 marks)

(b) For $A = 14.0$, $B = 22.5$ and $C = 34.6$, compute the quantity

$$Q = \frac{\sqrt{A^3 + B^3}}{(1 + \frac{A}{B})^2}$$

(3 marks)

(c) Plot the function $Y = \frac{10\ X}{5 + X}$ in the X range [0,20] with the following specifications:

(1) Join the points with a red colored line
(2) X axis title : "Concentration in micromoles/L"
(3) Y axis title : "Reaction rate in micromoles/L/min"

(2 marks)

R-script name : Question-1.r

2. The attached file "crop_table.csv" has the yearly data on the yield of various crops in million metric tons over 12 consequtive years starting from 1950. Read the filw into a R data frame to perform the following tasks:

(a) Plot the yield of Wheat from 1950 to 2012 with points and lines

(b) Create a scatter plot of Cereal Vs Wheat yield with points and lines.

Plots (a) and (b) should be produced on the same canvas by splitting canvas into two portions.

(4 marks)

(c) Create a new column named "Total_production" by adding the yields of Rice, Wheat and Cereals and print this column on the screen.

(3 marks)

R-script name : Question-2.r

3. . Write a R-scipt for performing the following tasks:

(a) Plot a Poisson distribution $P(x,\ \mu = 7)$ in the x range 0 to 14. The title of the plot should be "Poisson distribution for mean = 7".

(2 marks)

(b) Generate 10000 points randomly drawn from a t-distribution with 6 degrees of freedom. Plot the histogram of these data points

(2 marks)

(c) For this data, find the number of data points more than 2 standard deviations above the mean value and print this number.

(3 marks)

R-script name : Question-3.r

4. The following are the systolic blood pressure (mmHg) of 14 patients undergoing drug therapy for hypertension:

X = {183, 152, 178, 157, 194, 163, 144, 114, 178, 152, 118, 158, 172, 138 }

Assuming the systolic blood pressure folows a Gaussian distribution, write an R script to perform an appropriate test to conclude, based on this data, whether the population mean is less than 165. Print the hypothesis and results clearly. [7 marks]

R-script name : Question-4.r

5. In a clinical trial, the placebo and treatment groups consisting of 12 patients each were subjected to a test. The results are presented here:

Placebo : 3.1,5.2,5.3,4.7,5.4,5.7,3.8,6.2,6.9,5.5,4.1,7.8

Treatment : 4.9,6.9,7.1,4.9,4.5,6.1,6.4,6.2,6.3,7.4,5.4,4.4

(a) Assuming that the two populations follow Gaussian distributions with unequal variance, perform a two sample Welsch t-test using the appropriate R library function for comparing the population mean of these two groups. Use $\alpha = 0.05$ for accepting or rejecting null hypothesis.

(3 marks)

(b) On the same data, perform a Wilcoxon Rank Sum test using the appropriate R library function for comparing the population mean of these two groups.

(4 marks)

Print the null and alternative hypothesis along with the results of both the tests.

R-script name : Question-5.r

6. Let BC1, BC2, BC3 and BC4 be the blood cholesterol level of women belonging to 4 different age groups. Assume that the distributions of BC_i to be $N(\mu_i, \sigma)$, i=1,2,3,4.

The data sets are given here:

BC1 :	221	213	202	183	185	197	162
BC2 :	271	192	189	209	227	236	142
WC3 :	262	193	224	201	161	178	265
WC4 :	192	253	248	278	232	267	289

Using an appropriate function from R library, perform an ANOVA test on this data to see if a significant level of differences exists between the mean cholesterol levels of these 4 age groups. Print your null hypothesis and the conclusion along with the results clearly.

(7 marks)

R-script name : Question-6.r

A7. A biological quantity Q is known to vary linearly with time T. The quantity was measured every hour and the data is presented here.

T	(hours)	0	1	2	3	4	5	6	7	8
Q	(mg/liter)	0.2	1.8	4.3	5.7	8.3	9.7	12.4	13.6	16.3

Make a least square fit to a straight line of the form $Q = a + bT$ to determine the slope b and intercept a. Plot the data points and the fitted straight line in the same plot.

(7 marks)

R-script name : Question-7.r