Biostatistics and R

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

PART A. Multiple choice questions (answer any 8 out of 10) (8 × 1 Mark)

A1. If 120 dice are thrown, the expected number of them showing an outcome of 5 is
(a) 24
(b) 16
(c) 32
(d) 20

A2. Find the median of the given data set: 5,8,12,17,2,14,6,8, 13, and 7
(a) 8
(b) 7
(c) 5
(d) 14

A3. If K is the Mean of Poisson distribution, then the variance is given by
(a) $\sqrt{K}$
(b) K^2
(c) K
(d) K/2

A4. Which one of the following variables is not categorical?
(a) Gender of a person: male or female
(b) Age of a person.
(c) Choice on a test item: true or false.
(d) Marital status of a person (single, married, divorced)

A5. Which one of these statistics is unaffected by outliers
(a) Mean
(b) Standard deviation
(c) Interquartile range
(d) Range

A6. One use of a regression line is
(a) to determine if any x-values are outliers.
(b) to determine if any y-values are outliers
(c) to determine if X and Yare independent variables
(d) to estimate the change in y for a one-unit change in x.

A7. A chi-square test involves a set of counts called expected counts. What are the expected counts? (a) Hypothetical counts that would occur of the alternative hypothesis were true.
(b) Hypothetical counts that would occur if the null hypothesis were true.
(c) The actual counts that did occur in the observed data.
(d) Counts that would occur in a large data set

A8. A parameter is:
(a) always normally distributed
(b) a population characteristic
(c) a sample characteristics
(d) unknown

A9. The intercept in linear regression represents:
(a) the expected x value when y is zero

(b) the expected y value when x is zero
(c) the strength of the relationship between x and y
(d) a poplation paramter

A10. Which one of the following statistical test is a paramteric test?
(a) Wilcoxon Rank Sum Test
(b) Kruskal-Wallis Test
(c) Chi-square goodness of fit test
(d) ANOVA

PART B **(answer any 4 out of 6)** (4×3 **Marks**)

B1. Define Pearson's correlation coefficient and explain it.

B2. In a certain population an average of 8 new cases of esophageal cancer are diagnosed each year. Find the probability that in a given year at least one case of easophageal cancer is diagnosed.

B3. It is calimed that 15% of ducks in a region are affected by patent schistosome infection. If we randomly sample 10 ducks, what is the probability that

(a) exactly three ducks are found to be infected

(b) no duck is infected

B4. A bird flies a distance of $d = 120 \pm 3\ m$ during a time $20.0 \pm 1.3\ s$. What is the uncertainity in the average speed of the bird?

B5. Suppose the average length of stay for patients of a chronic disease in a hospital is 60 days with a standard deviation of 15. Assuming the length of stay approximately follows a Gaussian distribution, find the probability that a randomly selected patient in the hospital will stay between 30 to 70 days.

B6. Find the value of the Pearson Correlation coefficient for the following data :

Age (X)	:	43, 21, 25, 42, 57, 59
Glucose level (Y)	:	99, 65, 79, 75, 87, 81

PART C **(answer any 3 out of 5)** (5×3 **Marks**)

C1. Bean seeds from supplier A have an 85% germination rate and thosw from supplier B have 70% germination rate. A seed packaging company purchases 40% of their bean seeds from supplier A and the remaining 60% from supplier B and mixes the seed together.

Given that a seed germinates, find the conditional probability that the seed was purchased from supplier B.

C2. Let X and Y be the blood volume (in milliliters) for males who do regular paragliding and the males who do regular sports activities respectively. Seven random observations of X and 8 random observations of Y yielded the following results:

X	:	1612, 1352, 1456, 1222, 1560, 1456, 1924
$Y.$	:	1248, 1092, 1040, 1288, 1248, 910, 1040

Test the hypothesis that the populations means are equal against the two-sided alternate hypothesis. Let $\alpha = 0.05$

C3. Let X and Y denote the heights of blue spruce trees (measured in centimeters) in two large fields. We compare the heights by measuring the heights of two randomly selected sets of trees from each field. The observations are tabulated below:

$$X \; : \; 90.4, 77.2, 75.9, 83.2, 84.0, 90.2, 87.6, 67.4, 77.6, 69.3, 83.3, 72.7$$

$$Y \; : \; 92.7, 78.9, 82.5, 88.6, 95.0, 94.4, 73.1, 88.3, 90.4, 86.5, 84.7, 87.5$$

Using Wilcoxon test, test the hypothesis $H_0 : \; m_X = m_Y$ against an alternative hypothesis $H_1 \; : m_X < m_Y$ to a significance level of $\alpha = 0.05$.

C4. The driver of a diesel powered automobile decided to test the quality of three types of diesel fuel (named A,B abd C) sold in the nearby patrol pumps, based on how many kilometers his vehicle could go per 5 liters of fuel. He collected the following random data by driving his vehicle many times with each brand of diesel:

$$Brand\ A \; : \; 38.7, \;\; 39.2, \;\; 40.1, \;\; 38.9$$
$$Brand - B \; : \; 41.9, \;\; 42.3, \;\; 41.3$$
$$Brand - C \; : \; 40.8, \;\; 41.2, \;\; 39.5, \;\; 38.9, \;\; 40.3$$

Using the method of ANOVA, test the null hypothesis that the mean mileage of the three brands are equal, to a significant level of $\alpha = 0.05$

C5. Chemists use Ion Sensitive Electrodes to measure ionic concentrations of acquous solutions. In order to calibrate this equipment, the output signal in millivolt was measured for known ion concentrations in units of ppm. The data is reproduced here:

concentration (in ppm) :	0.0	50.0	75.0	100.0	150.0	200.0
signal (in mV) :	1.72	2.11	2.36	2.56	3.05	3.42

Calculate a least square regression line between concentration and signal data.