National University
of Computer & Emerging Sciences
Islamabad

Probability and Statistics (Sections: Cyber-Sec-M, DS-N, Artificial Intelligence – J, K)

Semester Project

Course Instructor: Muhammad Usman Ashraf

Total Marks: 8 Absolutes

Project Submission Guidelines:

- **Title page contain each group members name and roll number.**
- **You must highlight the contribution from each group in completing the project.**
- This is a group project.
- Each group must complete this project.
- A group can have maximum of 3 students.
- If a student wishes to work individually, he / she can submit individual project.
- Plagiarized work (from internet or fellow students) will result in zero marks.
- Deadline for complete project submission via email @ **for sections AI (J), AI (K):**

  i180721@nu.edu.pk and **for sections DS (N), cyber security (M):** i180646@nu.edu.pk

  (one pdf file and one program file) is Thursday 17 June 2021 latest by 9 pm.
- Name of your project files must be as per following format: **Section_ID1_ID2_ID3**. (e.g. A_123456_654321_987654)
- Deadline for complete project submission in hard copy is Thursday 17 June 2021 latest by 3.30 Pm.
- 3 absolute Marks will be deducted in case of missing hard copy of the project.

- Late submissions will not be considered.

Question No # 01:

According to the Union of Concerned Scientists (www.ucsusa.org), as of November 2012, there were 502 low Earth orbit (LEO) and 432 geosynchronous orbit (GEO) satellites in space. Each satellite is owned by an entity in either the government, military, commercial, or civil sector. A breakdown of the number of satellites in orbit for each sector is displayed in the accompanying table. Use this information to construct pair of graphs (i. Rectangles, Pie Chart) that compare the ownership sectors of LEO and GEO satellites in orbit. What observations do you have about the data?

Ownership Sector	**LEO Satellites**	**GEO Satellites**
Government	229	59
Military	109	91
Commercial	118	281
Civil	46	1

Note: All the graph in this should be construct using R-language

Question No# 02:

a) A number of particular articles has been classified according to their weights. After drying for two weeks the same articles have again been weighted and similarly classified. It is known that the median weight in the first weighing was 20·83 oz. while in the second weighing it was 17·35 oz. Some frequencies *a* and *b* in the first weighing and *x* and *y* in the second are missing. Find out the values of the missing frequencies.

	Frequencies	
Class	**1st weighing**	**2nd weighing**
0-4	a	x
5-9	b	Y
10-14	11	40
15-19	52	50
20-24	75	30

25-29	22	28

b) Obtain the variance of $Y = 2X_1 + 3X_2 - 4X_3$ where X_1, X_2 and X_3 are three random variables with means given as 3, 4, and mean($X3$) = last digit of your roll number along with variances 10, 20, 30 respectively, and co-variances between them are zero. Find the mean and variance of Y

Question No# 03:

In the manufacturing of a certain scientific instrument great importance is attached to the life of a particular critical component. This component is obtained in bulk from two sources, A and B, and in the course of inspection, the lives of 1000 of the components from each source are determined. The following frequency tables are obtained:

1998						2003					
43	76	51	14	0	10	10	11	14	20	15	6
20	0	5	17	67	25	17	0	5	19	127	4
38	0	56	8	0	9	31	5	88	1	1	16
14	5	37	14	95	20	14	19	20	9	138	22
23	12	33	0	3	45	13	10	20	20	20	12

1. Find Median and two quartiles (Lower and Upper Quartile) for each group.
2. Find mean and Standard deviation for each source and compare them.
3. Which source do you think providing better quality of components and why? **Note:** answer this part by considering results of mean and standard deviation.
4. Calculate absolute and relative measure of skewness for each group and comment on life of the components for each source.
5. Also find Minimum and Maximum for each group and Construct Box plot

Note: Every calculation in this question must be carried out using R-language.

Question No# 04:

I. Find the 23th, 48th ,64th and 82th Percentile and 3rd, 6th and 9th Decile of the following data:
56, 65, 65, 67, 72, 73, 75, 77, 77, 78, 78, 78, 80, 80, 80, 82, 83, 85, 85, 86, 87, 87, 87, 88, 90, 92, 93, 95, 98

II. What will be the value of Mean, Standard Deviation and Variance if you:

a. Add sum of last three digits of your roll number (e.g. your roll number is 19I-1234 than sum of last three digits of your roll number will be "2+3+4 = 9") to each value.

b. Multiple sum of last three digits of your roll number (e.g. your roll number is 19I-1234 than sum of last three digits of your roll number will be "2+3+4 = 9") to each value.

Question No# 05:

I. Take a suitable data set (related to Your Field) to construct the component Multiple bar chart through R-lang.

II. Take a suitable data set (related to Your Field) to construct the component Histogram through R-language.

III. Take a suitable data set (related to Your Field) to construct the component Ogive through R-language.

IV. Take a suitable data set (related to Your Field) to construct the component Scatter plot through R-language.

V. Take a suitable data set (related to Your Field) to construct the component Line Chart through R-language.

Note: Also write the link of dataset in pdf.

Question No# 06

a) In arithmetic, a computer program generates random questions that students must answer within a fixed time frame. The probability of replying correctly to the first question is 0.95. When a question is answered correctly, the next generated question becomes more difficult, and the probability of a correct answer being given is reduced by 0.2. Every time a question is answered incorrectly, the next question is of the same standard, and the probability of being given correctly remains unchanged. The third question is answered correctly, what is the chance that the second question will be answered correctly.

b) A programmer is taking a two-hours-time-limit makeup examination. Suppose the probability that the programmer will finish the exam at most y hours is y/3, for all $0 \leq y \leq 2$. Given that the student is still working after 1.75 hours, what is the conditional probability that the full time is used?

c) A multiple choice test has 7 questions with 3 wrong choices and 1 correct choice each. How many ways are there to answer the test? What is the probability that two papers have the same answers?

Question No# 07

The proportion of people who respond to a certain mail-order solicitation is a continuous random variable that has the density function

$$f(x)=\begin{cases} ax-bx^2 & \text{for } 0\le x\le 2. \\ 0 & \text{elsewhere} \end{cases}$$

Observations on X indicate that the mean is sum of the last two digits of your roll number.

i. Find the values of a and b.

ii. If two independent observations are made on X what is the probability that at least one of them is less than 0.5.

Question No# 08

As described in the book's Introduction, Benford's Law arises in a variety of situations as a model for the first digit of a number:

$$p(x) = P(\text{lst digit is } x) = \log_{10}\left(\frac{x+1}{x}\right),\ x = 1,2,\ldots,9$$

(a) Without computing individual probabilities from this formula, show that it specifies a legitimate Probability mass function.

(b) Compute the individual probabilities and compare to the distribution where 1, 2. . . 9 are equally likely.

(c) Calculate the measure of skewness and interpret your result.

(d) Calculate the mode and interquartile range of the probability distribution.

Question No# 09

a) Processor made by a certain company are considered to be defective with an independent probability 0.01. The company provides the processor in 10 pack. It offers a money-back guarantee that 1 of the 10 processor is faulty at most. What percentage will the company substitute parcels sold? What is expected number of packages sold must the company replace?

b) A modem transmits over an error-prone channel, so it repeats every "0" or "1" bit transmission five times. We call each such group of five bits a "code word." The channel changes an input bit to its complement with probability p = 1/10 and it does so independently of its treatment of other input bits. The modem receiver takes a majority vote of the five received bits to estimate the input signal. Find the probability that the receiver makes the wrong decision.

Question No# 10

Take one example of Binomial Distribution, hypergeometric Distribution and Poisson distribution and Find.

i. Construct the complete probability distribution.
ii. Construct the Cumulative probability distribution.
iii. Find the prob. of different events.
iv. Draw probability mass function
v. Cumulative probability function.
vi. Discuss the different behavior of the line graph when n will be changed (only in case of disc. Prob. Dis.).
vii. Also generate 50 random numbers from each distribution.
viii. Find the First Second and Third Quartiles of each distribution under study.
ix. Calculate the 50th, 70th, 95th percentiles.
x. Calculate the second decile and D6

Note: Clearly specified the complete procedure to execute each part in R