Question 1. A researcher is concerned about the level of knowledge possessed by university students regarding United States history. Students completed a high school senior level standardized U.S. history exam. Major for students was also recorded. Data in terms of percent correct is recorded below for 32 students. Compute the appropriate test for the data provided below.

Education	Business/Management	Behavioral/Social Science	Fine Arts
62	72	42	80
81	49	52	57
75	53	31	87
50	68	70	54
67	39	22	28
40	69	61	29
46	40	68	52
36	15	76	45

What is your computed answer?
What would be the null hypothesis in this study?
What would be the alternate hypothesis?
What probability level did you choose and why?
What were your degrees of freedom?
Is there a significant difference between the four testing conditions?
Interpret your answer.

Hints: Use ANOVA

Question-2 In a gene expression data, we have total 100 genes which have higher expression in a disease X compared to normal population. In total there were 15000 genes expressed in both samples (diseased and normal). There are 3 functions F1, F2 and F3 for which we want to know if they are enriched in diseased cases. Among 15000 expressed genes 40 genes have F1 function while 50 genes are of F2 and 60 genes are of F3. In the disease-upregulated subset of 100 genes, 15 have F1 function, 20 have F2 and 12 have F3. What is the significance (P-value) of enrichment of the functions F1, F2, F3 in the upregulated gene set (of 100).

Question-3. Apply Fischer exact test systematically on contigency table without using functions in R or any tool

	Math	science
Fail	2	4
Pass	6	7

Find if the fraction of pass in math is correlated to fraction of pass in science.

Question-4. you are playing Ludo with your friend. Your friend gets higher number of sixes in a dice which you know is unbiased. You decide to test wether he is cheating and you start counting how many times he get sixes in a set of 5 throws. You got following observation in total 40 such set..

Number of sixes in a set (5 throws)	number of sets (observed)
0	4 sets
1	6 sets
2	7 sets
3	5 sets
4	10 sets
5	8 sets

What will you conclude ? Try two methods.

Question-5: You are playing Ludo with your friend. However your friend says that he will throw his own dice about which you have suspicion. In 170 throws you observe following counts

count (for 1) = 20, count(for 2) = 20, count (for 3) = 20
Count (for 4) = 40, count(for 5) = 40 , count (for 6) = 30

(a) Is his dice unbiased assuming alpha=0.05 for P-value ?

Then your friend starts checking your dice and gets following counts in 95 throws of your dice.

count (for 1) = 10, count(for 2) = 5, count (for 3) = 10
Count (for 4) = 20, count(for 5) = 30 , count (for 6) = 20

(b) He claims that your dice is similar to his dice. Is it true ? What is the significance of your conclusion ?

You go to another friend (B) to make judgment, However that friend(B) does not have patience and throws each dice less number of times and observes following for two dice

Your Dice: for 1:2, for2: 2, for 3: 3 , for 4: 4, for 5: 4, for 6:3
Friends Dice: for 1:1, for 2:2, for 3: 2 , for 4: 3, for 5: 3, for 6:3

c) what would his conclusion and the significance of his conclusion.

question-6. In the Assignment2-Data file uploaded on backpack. We have tow matrices. Kindly use goodness of fit to test if they fit Gaussian, Binomial or Poisson distribution. What is the P-value in each of fit.
Hint: try dividing the data in to 10 or 13 category and find observed and expected frequency in each category

question-7.Suppose in IIITD we following number students in different branches
CSE : 300
ECE: 200
CB: 30

In following courses we have enrollment as such

SC : CSE: 30, ECE: 20, CB: 10
GT : CSE: 50, ECE: 20, CB: 5
AI : CSE: 100, ECE: 30, CB:8

what is the likelihood of the combinations in all three cases (subject)
In what is the P-value of enrichment of CB in each subject.