Birla Institute of Technology & Science, Pilani

Second Semester 2019-2020, MATH F113 (Probability & Statistics)

Assignment

Submission Deadline: May 11, 2020 **Max. Marks: 30**

Note:

1. The assignment contains 11 questions. Questions Q1 to Q10 carry two marks each, while Q11 carries 10 marks.
2. Please attempt all questions including their subparts. Provide a complete solution for each question. Just writing the final answer will receive zero credit.
3. For Q1 to Q10, prepare a handwritten solution. For Q11, in addition to the handwritten solution, please attach a screenshot of your analysis based on the "Data Analysis Toolpak" in MS-Excel. Finally, please submit the scanned copy of the assignment solution in a pdf format.
4. Write your name and ID on every page of the assignment solution.
5. For submission, please use the following naming convention of the scanned pdf file: Tut Sec_ID_Name. For example, T19_2019A1PS0714P_Punya Gupta.
6. Detailed instructions regarding the submission process will be shared to you shortly.

1. The joint density function of X and Y is given by

$$f(x,y)=\begin{cases} a\left(3xy^2+bx^5\right); & 0\le x,y\le 1 \\ 0 & ;\text{otherwise} \end{cases}$$

Find the real constants a and b such that X and Y become independent.

2. Let Z be a standard normal random variable and $Y = a + bZ + cZ^2$, where a, b and c are real constants. Compute $\rho(Y,Z)$.

3. Find the distribution of $X=\sum_{i=1}^{n} X_i$, where X_i's are i.i.d. geometric random variables with parameter p.

4. One observed value of a binomial random variable X with $p=\frac{1}{3}$ is $x=2$. With this information, find the maximum likelihood estimate for the parameter n.

5. If the margin of error for the population mean for 99% confidence level is 39, what was the population standard deviation if the sample size selected was 50?

6. Researchers are concerned about the impact of browsing social networking sites while they are enrolled in classes. They would like to know if students are spending too much on these sites and therefore are spending less time on their classes than they should. For this, the researchers need to find out, on an average, how many hours in a week students are spending on social networking sites. From previous studies, it is known that the population standard deviation of the associated random variable is about 5 hours (per week). A survey of 200 students provides a sample mean of 7.10 hours spent (per week) on social networking sites. What is a 95% confidence interval of population mean based on this sample?

7. In cosmology, the Hubble constant H_0 measures the expansion rate of the Universe, and the H_0 measurements from different sources are assumed Gaussian. A sample of 49 measurements of H_0 from different sources suggest that any two galaxies separated by 1 Mpc distance in the Universe are moving away from each other

with a velocity 70 km/s and standard deviation of 5 km/s/Mpc, and hence the observed value of H_0 is 70±5 km/s/Mpc. Find a 95% confidence interval for the true value of H_0.

8. The average adult male height in a certain country is 170 cm. A group of researchers suspect that the adult male height at the southernmost city of the country might have a different average height due to the influence of extreme weather conditions. To test, they have chosen a random sample of nine adult males from the city and obtained the following observed values (in cm):

   176.2 157.9 160.1 180.9 165.1 167.2 162.9 155.7 166.2

   Assuming normality of the population, carry out a hypothesis test at $\alpha = 0.05$.

9. A sample of 9 days over the past six months showed that a showroom had the following number of air conditioner sales: 22, 25, 20, 18, 15, 22, 24, 19 and 26. If the number of sales per day is normally distributed, would an analysis of these sample data reject the hypothesis that the variance in the number of sales per day is equal to 10? Use $\alpha = 0.1$ to draw conclusion.

10. In studying the effect of air quality on a lake, the experimenter takes observations on the *ph*-level of the water and the air quality as measured on an air quality index. The index goes from 0 to 100 with larger numbers representing high pollution. These data are provided below.

ph-level (x)	4.5	4.1	4.8	4.0	5.0	6.0	3.5	4.9	3.2	6.1
air quality (y)	40	50	30	60	20	10	70	30	85	15

    Calculate point estimates of the slope and intercept of the linear regression line. Hence, provide a point estimate of the standard deviation σ.

11. To answer the questions below, please use MS-Excel's "Data Analysis Toolpak" (https://support.office.com/en-us/article/load-the-analysis-toolpak-in-excel-6a63e598-cd6d-42e3-9317-6b40ba1a66b4). A screenshot of your analysis needs to be uploaded as a part of submission.

    Consider the following data obtained from a chemical process where the yield of the process is thought to be related to the reaction temperature.

Observation Number	Temperature (x_i)	Yield (y_i)	Observation Number	Temperature (x_i)	Yield (y_i)
1	50	122	16	80	182
2	53	118	17	82	180
3	54	128	18	85	183
4	55	121	19	87	188
5	56	125	20	90	200
6	59	136	21	93	194
7	62	144	22	94	206
8	65	142	23	95	207
9	67	149	24	97	210
10	71	161	25	100	219
11	72	167			
12	74	168			
13	75	162			
14	76	171			
15	79	175			

    (i) Use the "Descriptive Statistics" tool to compute sample mean, sample median and sample standard deviation for the temperature and yield data.

    (ii) Use the "Regression" analysis tool to find the equation of the estimated linear regression line, SSE, SST, SSR and the coefficient of determination. Provide a point estimate of the population correlation coefficient.