Department of Mathematics
MTL 601 (Probability and Statistics)
Tutorial Sheet No. 8

1. There has been a great deal of controversy in recent years over the possible dangers of living near a high-level electromagnetic field (EMF). After hearing many anecdotal tales of large increase among children living near EMF, one researcher decided to study the possible dangers. In order to do his study, he followed following steps: (a) studied maps to find the locations of electric power lines, (b) used these maps to select a fairly large community that was located in a high-level EMF area. He interviews people in the local schools, hospitals, and public health facilities in order to discover the number of children who had been affected by any type of cancer in the previous 3 years. He found that there had been 32 such cases. According to government public health committee, the average number of cases of childhood cancer over a 3-year period in such a community was 16.2, with a standard deviation of 4.7. Is the discovery of 32 cases of childhood cancers significantly large enough, in comparison with the average number of 16.2, for the researcher to conclude that there is some special factor in the community being studied that increases the chance for children to contract cancer? Or is it possible that there is nothing special about the community and that the greater number of cancers is solely due to chance?

2. Let $Y_1 < Y_2 < \cdots < Y_n$ be the order statistics of a random sample of size 10 from a distribution with the following PDF
$$f(x;\theta) = \frac{1}{2}e^{-|x-\theta|}, \quad -\infty < x < \infty$$
for all real θ. Find the likelihood ratio test λ for testing $H_0: \ \theta = \theta_0$ against the alternative $H_1: \ \theta \neq \theta_0$.

3. Let $X_1, X_2, \ldots, X_n$ and $Y_1, Y_2, \ldots, Y_n$ be independent random samples from the two normal distributions $N(0, \theta_1)$ and $N(0, \theta_2)$.
   (a) Find the likelihood ratio test λ for testing the composite hypothesis $H_0: \ \theta_1 = \theta_2$ against the composite alternative hypothesis $H_1: \ \theta_1 \neq \theta_2$.
   (b) The test statistic λ is a function of which F statistic that would actually be used in this test.

4. Let $X_1, X_2, \ldots, X_{50}$ denote a random sample of size 50 from a normal distribution $N(\theta, 100)$. Find a uniformly most powerful critical region of size $\alpha = 0.10$ for testing $H_0: \ \theta = 50$ against $H_1: \ \theta > 50$.

5. Consider a queueing system that describes the number of telephone ongoing calls in a particular telephone exchange. The mean time of a queueing system is required to be at least 180 s. Past experience indicates that the standard deviation of the talk time is 5 s. Consider a sample of 10 customers who reported the following talk time

   $210, 195, 191, 202, 152, 70, 105, 175, 120, 150.$

   Would you conclude at the 5% level of significance that the system is unacceptable? What about at the 10% level of significance.

6. Let $X_1, X_2, \ldots, X_n$ be a random sample from a distribution with the following PDF
$$f(x;\theta) = \begin{cases} \theta x^{\theta-1}, & 0 < x < \infty \\ 0 & \text{otherwise} \end{cases}$$
where $\theta > 0$. Find a sufficient statistics for θ and show that a uniformly most powerful test of $H_0: \ \theta = 6$ against $H_1: \ \theta < 6$ is based on this statistic.

7. If $X_1, X_2, \ldots, X_n$ is a random sample from a beta distribution with parameters $\alpha = \beta = \theta > 0$, find a best critical region for testing $H_0: \ \theta = 1$ against $H_1: \ \theta = 2$.

8. Let $X_1, X_2, \ldots, X_n$ denote a random sample of size 20 from a Poisson distribution with mean θ. Show that the critical region C defined by $\displaystyle\sum_{i=1}^{20} x_i \geq 4$.

9. Let X be a discrete type random variable with PMF

$$P(x;\theta)=\begin{cases}\theta^x(1-\theta)^{1-x}, & x=0,1\\ 0 & \text{otherwise}\end{cases}.$$

We test the simple hypothesis $H_0:\ \theta=\frac{1}{4}$ against the alternative composite hypothesis $H_1:\ \theta<\frac{1}{4}$ by taking a random sample of size 10 and rejecting H_0 if and only if the observed values $x_1, x_2, \ldots, x_{10}$ of the sample observations are such that $\sum_{i=1}^{10} x_i < 1$. Find the power of this test.

10. In a certain chemical process, it is very important that a particular solution that is to be used as a reactant has a pH of exactly 8.20. A method for determining pH that is available for solutions of this type is known to give measurements that are normally distributed with a mean equal to the actual pH and with a standard deviation of .02. Suppose ten independent measurements yielded the following pH values: 8.18, 8.17, 8.16, 8.15, 8.17, 8.21, 8.22, 8.16, 8.19, 8.18.

    1. What conclusion can be drawn at the $\alpha = 0.10$ level of significance?
    2. What about at the $\alpha = 0.05$ level of significance?

11. An automobile manufacturer claims that the average mileage of its new two-wheeler will be at least 40 km. To verify this claim 15 test runs were conducted independently under identical conditions and the mileage recorded (in km) as: 39.1, 40.2, 38.8, 40.5, 42, 45.8, 39, 41, 46.8, 43.2, 43, 38.5, 42.1, 44, 36. Test the claim of the manufacturer at $\alpha = 0.05$ level of significance.

12. The life of certain electrical equipment is normally distributed. A random sample of lives of twelve such equipments has a standard deviation of 1.3 years. Test the hypothesis that the standard deviation is more than 1.2 years at 10% level of significance.

13. Random samples of the yields from the usage of two different brands of fertilizers produced the following results: $n_1 = 10, \overline{X} = 90.13, s_1^2 = 4.02; n_2 = 10, \overline{Y} = 92.70, s_2^2 = 3.98$.Test at 1 and 5% level of significance whether the difference between the two sample means is significant.

Table 1: Data for Problem 10.14

15	7	7	15	11	9
20	12	17	12	18	18
25	14	18	18	19	19
30	19	25	22	19	23
35	7	10	11	15	11

14. Consider the strength of a synthetic fiber that is possibly affected by the percentage of cotton in the fiber. Five levels of this percentage are considered with five observations at each level. The data are shown in Table 1. Use the F-test, with $\alpha = 0.05$ to see if there are differences in the breaking strength due to the percentages of cotton used.Please check the edits made in Table captions (Tables 5.27–5.34, 5.36–5.40, 5.42), and correct if necessary.

15. It is desired to determine whether there is less variability in the marks of probability and statistics course by IITD students than in that by IITB students. If independent random samples of size 10 of the two IIT's yield $s_1 = 0.025$ and $s_2 = 0.045$, test the hypothesis at the 0.05 level of significance.

16. Two analysts A and B each make +ve determinations of percent of iron content in a batch of prepared ore from a certain deposit. The sample variances for A and B turned out to be 0.4322 and 0.5006, respectively. Can we say that analyst A is more accurate than B at 5% level of significance?

17. Elongation measurements are made of ten pieces on steel, five of which are treated with method A (aluminum only), and the remaining five are method B (aluminum plus calcium). The results obtained are given in Table 2. Test the hypothesis that

    1. $\sigma_A^2 = \sigma_B^2$.
    2. $\mu_B - \mu_A = 10\%$.

    at 2% level of significance by choosing approximate alternatives.

Table 2: Data for Problem 10.17

Method A	78	29	25	23	30
Method B	34	27	30	26	23

18. Suppose the weekly number of accidents over a 60-week period in Delhi is given in Table 3. Test the hypothesis that the number of accidents in a week has a Poisson distribution. Assume $\alpha = 0.05$.

Table 3: Data for Problem 10.18

1	0	0	1	3	4	0	2	1	4
2	2	0	0	5	2	1	3	0	1
1	8	0	2	0	1	9	3	3	5
1	3	2	0	7	0	0	0	1	3
3	3	1	6	3	0	1	2	1	2
1	1	0	0	2	1	3	0	0	2

19. A study was investigated to see if Southern California earthquakes of at least moderate size (having values of at least 4.4 on the Richter Scale) are more likely to occur on certain days of the week then on others. The catalogs yielded the following data on 1100 earthquakes given in Table 4. Test at the 5% level of significance, the hypothesis that an earthquake is equally likely to occur on any of the 7 days of the week.

Table 4: Data for Problem 10.19

Day	Sun	Mon	Tue	Wed	Thur	Fri	Sat
Number of Earthquakes (f_i)	156	144	170	158	172	148	152

20. A builder claims that a particular brand water heaters are installed in 70% of all homes being constructed today in the city of Delhi, India. Would you agree with this claim if a random survey of new homes in this city shows that 9 out of 20 had water heater installed? Use a 0.10 level of significance.

21. The proportions of blood types O, A, B and AB in the general population of a particular country are known to be in the ratio 49:38:9:4, respectively. A research team, investigating a small isolated community in the country, obtained the frequencies of blood type given in Table 5. Test the hypothesis that the proportions in this community do not differ significantly from those in the general population. Test at the 5% level of significance.

Table 5: Data for Problem 10.21

Blood type	O	A	B	AB
Frequency (f_i)	87	59	20	4

Table 6: Data for Problem 10.22

4	3	3	1	2	3	4	6	5	6
2	4	1	3	4	5	3	4	3	4
3	3	4	5	4	5	6	4	5	1
6	3	6	2	4	6	4	6	3	5
6	3	6	2	4	6	4	6	3	2
5	4	6	3	3	3	5	3	1	4

22. Consider the data of Table 6 that correspond to 60 rolls of a die. Test the hypothesis that the die is fair ($P_i = \frac{1}{6},\ i = 1, \ldots, 6$), at 0.5% level of significance.

23. A sample of 300 cars having cellular phones and one of 400 cars without phones are tracked for 1 year. Table 7 gives the number of cars involved in accidents over that year. Use the above to test the hypothesis that having a cellular phone in your car and being involved in an accident are independent. Use the 5% level of significance.

Table 7: Data for Problem 10.23

	Accident	No accident
Cellular phone	22	278
No phone	26	374

24. A randomly chosen group of 20,000 nonsmokers and one of 10,000 smokers were followed over a 10-year period. The data of Table 8 relate the numbers of them that developed lung cancer during the period. Test the hypothesis that smoking and lung cancer are independent. Use the 1% level of significance.

Table 8: Data for Problem 10.24

	Smokers	Nonsmokers
Lung cancer	62	14
No lung cancer	9938	19986

25. A politician claims that she will receive at least 60% o the votes in an upcoming election. The results of a simple random sample of 100 voters showed that 58 of those sampled would vote for her. Test the politician's claim at the 5% level of significance.

26. Use the 10% level of significance to perform a hypothesis test to see if there is any evidence of a difference between the Channel A viewing area and Channel B viewing area in the proportion of residents who viewed a news telecast by both the channels. A simple random sample of 175 residents in the Channel A viewing area and 225 residents in the Channel B viewing area is selected. Each resident in the sample is asked whether or not he/she viewed the news telecast. In the Channel A telecast, 49 residents viewed the telecast, while 81 residents viewed the Channel B telecast.

27. Can it be concluded from the following sample data of Table 9 that the proportion of employees favouring a new pension plan is not the same for three different agencies. Use $\alpha = 0.05$.Please provide caption for Table 5.35, as they are mandatory.

28. In a study of the effect of two treatments on the survival of patients with a certain disease, each of the 156 patients was equally likely to be given either one of the two treatments. The result of the above was that 39 of the 72 patients given the first treatment survived and 44 of the 84 patients given the second treatment survived. Test the null hypothesis that the two treatments are equally effective at $\alpha = 0.05$ level of significance.

Table 9: Data for Problem 27

	Agency 1	Agency 2	Agency 3
For the pension plan	67	84	109
Against the pension plan	33	66	41
Total	100	150	150

29. Three kinds of lubricants are being prepared by a new process. Each lubricant is tested on a number of machines, and the result is then classified as acceptable or nonacceptable. The data in the Table 10 represent the outcome of such an experiment. Test the hypothesis that the probability p of a lubricant resulting in an acceptable outcome is the same for all three lubricants. Test at the 5% level of significance.

Table 10: Data for Problem 29

	Lubricant 1	Lubricant 2	Lubricant 3
Acceptable	144	152	140
Not acceptable	56	48	60
Total	200	200	200

30. Twenty-five men between the ages of 25 and 30, who were participating in a well-known heart study carried out in New Delhi, were randomly selected. Of these, 11 were smokers, and 14 were not. The data given in Table 11 refer to readings of their systolic blood pressure. Use the data of Table 11 to test the hypothesis that the mean blood pressures of smokers and nonsmokers are the same at 5% level of significance.

Table 11: Data for Problem 30

Smokers	124	134	136	125	133	127	135	131	133	125	118			
Nonsmokers	130	122	128	129	118	122	116	127	135	120	122	120	115	123

31. Polychlorinated biphenyls (PCB), used in the production of large electrical transformers and capacitors, are extremely hazardous when released into the environment. Two methods have been suggested to monitor the levels of PCB in fish near a large plant. It is believed that each method will result in a normal random variable that depends on the method. Tests the hypothesis at the $\alpha = 0.10$ level of significance that both methods have the same variance, if a given fish is checked eight times by each method with the data (in parts per million) recorded given in Table 12.

Table 12: Data for Problem 31

Method 1	6.2	5.8	5.7	6.3	5.9	6.1	6.2
Method 2	6.3	5.7	5.9	6.4	5.8	6.2	6.3

32. The average weight of 25 new born babies is 2.78 kgms with a standard deviation of 0.72 kgms. Construct a 99% confidence interval for the true mean weight of such babies under the assumption that the population follows normal distribution.

33. Can it be concluded from the following sample data that the proportion of employees favoring a new pension plan is not the same for three different government agencies. Use $\alpha = .05$

	Agency I	Agency II	Agency III
For the pension plan	67	84	109
Against the pension plan	33	66	41

34. An economist estimates that the average Indian household saves 15% of its income. In a random sample of 64 households, the average saving is found to be 14%, and the standard deviation is 7%. Assume that population is normally distributed. Do we have enough evidence to refute the economists claim? Use a 0.05 level of significance level.

35. Suppose that a public opinion poll is surveyed with random sample of 1000 voters for 15th Lok Sabha Election 2009. Respondents were classified by gender (male or female) and by voting preference (Congress, BJP or CPI-M). Results are shown in the contingency table below.

	Congress	BJP	CPI-M
Male	200	150	50
Female	250	300	50

Can it be concluded from the above sample data that there is no relationship between gender and voting preference. Use a 0.05 level of significance level.

36. A vote is to be taken among the students of MAL 250 course in two sections to determine whether a quiz should be conducted. To determine if there is a significant difference in the proportion of two section voters favouring the proposal, a poll is taken. If 24 out of 40 students in section I favour the proposal and 24 out of 50 students in section II favour it, would you agree that the proportion of section I student voters favouring the proposal is higher than the proportion of section II student voters? Use an $\alpha = 0.05$ level of significance.

37. An oil company claims that the sulfur content of its diesel fuel is at most 0.15 percent. To check this claim, the sulfur contents of 40 randomly chosen samples were determined; the resulting sample mean, and sample standard deviation was 0.162 and 0.040, respectively. Using the five percent level of significance, can we conclude that the company's claims are invalid?

38. Historical data indicate that 4% of the components produced at a certain manufacturing facility are defective. A particularly acrimonious labor dispute has recently been concluded, and management is curious about whether it will result in any change in this figure of 4%. If a random sample of 500 items indicated 16 defectives, is this significant evidence, at the 5% level of significance, to conclude that a change has occurred.

39. An auto rental firm is using 15 identical motors that are adjusted to run at fixed speeds to test three different brands of gasoline. Each brand of gasoline is assigned to exactly five of the motors. Each motor runs on ten gallons of gasoline until it is out of fuel. Table 13 gives the total mileage obtained by the different motors. Test the hypothesis that the average mileage obtained is not affected by the type of gas used. Use the 5% level of significance.

Table 13: Data for Problem 39

Gas 1	Gas 2	Gas 3
220	244	252
251	235	272
226	232	250
246	242	238
260	225	256

40. To examine the effects of pets and friends in stressful situations, researchers recruited 45 people to participate in an experiment and data are shown in Table 14. Fifteen of the subjects were randomly assigned to each of the 3 groups to perform a stressful task alone (Control Group), with a good friend present, or with their dog present. Each subject mean heart rate during the task was recorded. Using ANOVA method, test the appropriate hypothesis at the $\alpha = 0.05$ level to decide if the mean heart rate differs between the groups.

Table 14: Data for Problem 40

	n	Mean	SD
Control	15	82.52	9.24
Pets	15	73.48	9.97
Friends	15	91.325	8.34

41. Suppose that to compare the food quality of three different hostel students on the basis of the weight on 15 students as shown in Table 15.
The means of these three samples are 68, 80, and 77. We want to know whether the differences among them are significant or whether they can be attributed to chance.

Table 15: Data for Problem 41

Narmada	72	58	74	66	70
Tapti	76	85	82	80	77
Kaveri	77	81	71	76	80

42. A fisheries researcher wishes to conclude that there is a difference in the mean weights of three species of fish (A,B,C) caught in a large lake. The data are shown in Table 16. Using ANOVA method, test the hypothesis at $\alpha = 0.05$ level.References "Cassella and Berger (2002), Freund and Miller (2004), Freund et al. (2010), Hoel et al. (1971), Medhi (1992), Meyer (1970), Panik (2012), Rao (1973), Ross (2014) and Shao (2003)" are given in list but not cited in text. Please cite in text or delete from list.

Table 16: Data for Problem 42

A	B	C
1.5	1.5	6
4	1	4.5
4.5	4.5	4.5
3	2	5.5