**CH5020 Assignment 2**

**Instructions:**

1. Do in groups of 2
2. Submit as word document (Arial font, 11 font size, 1.5 line spacing)
3. Use Matlab 2020 or any other licensed software of your choice; Give Matlab code or other commands as appropriate
4. Consider using live script in Matlab
5. PLEASE do discuss within and outside your group if required but provide original answers- Copying strictly prohibited
6. Pirated software strictly not permitted
7. Due Date: 5th October 2020

**Problem 1**

It is claimed that sports-car owners drive on the average 18,000 miles per year. A consumer firm believes that the average mileage is probably lower. To check, the consumer firm obtained information from 40 randomly selected sports-car owners that resulted in a sample mean of 17,463 miles with a sample standard deviation of 1348 miles. What can we conclude about this claim? Use α = 0.01.

**[7]**

**Problem 2**

A plant is suspected of discharging harmful effluents above the stipulated limit of 200 mg/L into a nearby river. The plant denies this plaint and shows results from sampling of the river carried out by them. These are given in the table below.

However, the Court orders an independent testing agency to sample the effluent concentrations. The results are also tabulated below. The plant lawyer further argues that his client’s results are more accurate as his sample shows less standard deviation. The Court appoints a neutral expert to give his recommendation.

1. State the claims of the Plant and the Neutral Agency
2. What conclusions will be drawn by the expert hypothesis testing? State the hypotheses clearly.
3. What conclusion will be drawn if the Plant tries to confuse the judge by invoking the ≠ alternate hypothesis?
4. What conclusions will be drawn by the expert using 95% confidence interval approaches?

|  |  |  |
| --- | --- | --- |
| **Detail** | **Plant** | **Neutral Agency** |
| Claim | ? | ? |
| Sample size | 3 | 20 |
| Mean concentration (mg/L) | 195 | 205 |
| Sample standard deviation (mg/L) | 4 | 6 |
| Sampling location(s) | points near mixing of river with the sea | random locations near the plant discharge |

**[8]**

**Problem 3**

The electrical resistances of components are measured as they are produced. A sample of six items gives a sample mean of 2.62 ohms and a sample standard deviation of 0.121 ohms. At what observed level of significance is this sample mean significantly different from a population mean of 2.80 ohms? In other words, is there less than 2% probability of getting a sample mean this far away from 2.80 ohms or farther purely by chance when the population mean is 2.80 ohms? Also indicate whether you may get the same conclusion using a 95% confidence interval test.

**[8]**

**Problem 4**

A sample of 15 concrete cylinders was taken randomly during production from a plant. The strength of each specimen was determined, giving a sample standard deviation of 215 kN/m2. Find the 95% confidence interval (with equal probabilities in the two tails) for *standard deviation* of the strengths. Assume the strengths of the concrete cylinders follow a normal distribution.

**[6]**

**Problem 5**

After an Olympiad test is given to all the students in the country, the average is estimated from random samples, with one sample taken from each of the five zones. All these samples have a sample size of 5. The average of each sample was calculated. Let the actual (true) average performance be 35%.

**Part I**

Which of the following is/are TRUE

1. It is highly likely that all the 5 sample means will be identically equal to 35%
2. If a larger sample size had been chosen, the distribution of the sample means around the true mean would have been narrower
3. If a smaller sample size had been chosen, the distribution of the sample means would have been narrower
4. The average of the distribution of the sample means will be 35%

**Part II**

Which of the following is/are TRUE

If the distribution of marks in the Olympiad’s population is **not** normally distributed then

1. The distribution of sample means with sample size equal to 5, will be normal
2. If the sample size had exceeded 30 for all samples, the distribution of the sample means would have been normal
3. Higher the sample size, lower will be the precision of the estimated population mean
4. Higher the sample size, higher will be the precision of the estimated population mean

**[3+3]**

**Problem 6**

From historical data, the yields of power from a nuclear reactor supplied by XYZ Company are normally distributed. This reactor supplied by this company is operated in several plants around the world. The mean daily output of power from a random sample of 6 measurements carried out over different days taken at an Indian plant is 27.33 GW and sample standard deviation is 9 GW.

1. Can the Indian plant accept this yield to be possible if XYZ Company guarantees (or warranties?) an average daily power output of 30 GW from its reactors for a given set of operating conditions?
2. If the same power output and variance are obtained from a sample size of 41, can the observed power output in the Indian plant be still considered to be acceptable?

If you were to make a scientific and unbiased report to the plant management, state your conclusions in both cases clearly.

**[8]**

**Problem 7**

A manufacturer of car batteries guarantees that his batteries will last, on the average, 3 years with a standard deviation of 1 year. If five of these batteries have lifetimes of 1.9, 2.4, 3.0, 3.5 and 4.2 years, is the manufacturer still convinced that his batteries have a standard deviation of 1 year? Assume that the battery lifetime follows a normal distribution.

**[6]**

**Problem 8**

For a chi-squared distribution, find *Χ* 2 such that

1. P(X2>*Χα* 2) =0.99 when ν =4
2. P(X2>*Χα* 2) =0.025 when ν =19
3. P(37.652< X2 < *Χα* 2) = 0.045 when ν =25

**[6]**

**Problem 9**

A normal population with unknown variance has a mean of 20. Is one likely to obtain a random sample of size 9 from this population with a mean of 24 and a standard deviation of 4.1? If not, what conclusion would you draw?

**[4]**

**Problem 10**

In Chennai suburbs, power cuts during summer months (and even otherwise) were quite common. In one such suburb, there was a complete blackout and complaints on the duration of the power cut were quite variable. The electricity board conducted a survey of 25 randomly chosen families and found that the mean duration of the power cut was 12 hours and the sample variance was 4 (hours)2. Construct a 98% CI on the variance assuming the population of families suffering power cuts to be normally distributed (for statistical purposes only).

**[6]**

**Problem 11**

The goodness of fit is used to test whether results may have come from a specified population. So we compare the actual results with those predicted from the hypothesized population. An experiment is conducted by tossing a dye 96 times and recording the values shown on the face in the table below. It is hypothesized that the results belong to the population of face values formed from tossing of a fair die.

The goodness of fit test between observed (oi) and hypothesized (ei) frequencies is determined from the chi-square distribution with k degrees of freedom as follows

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Face Value** | 1 | 2 | 3 | 4 | 5 | 6 |
| Experiment (Oi) | 19 | 15 | 17 | 18 | 14 |  |
| Hypothesized (ei) |  |  |  |  |  |  |

1. Fill in the missing information for face value 6.
2. Fill the hypothesized values (ei) for the different face values in the above table
3. Identify the degrees of freedom k
4. For the proposed population to fit the experimental data well, should the value be high or low? If you conclude that there is indeed goodness of fit, what can you conclude about the fairness of the die?

**[10]**

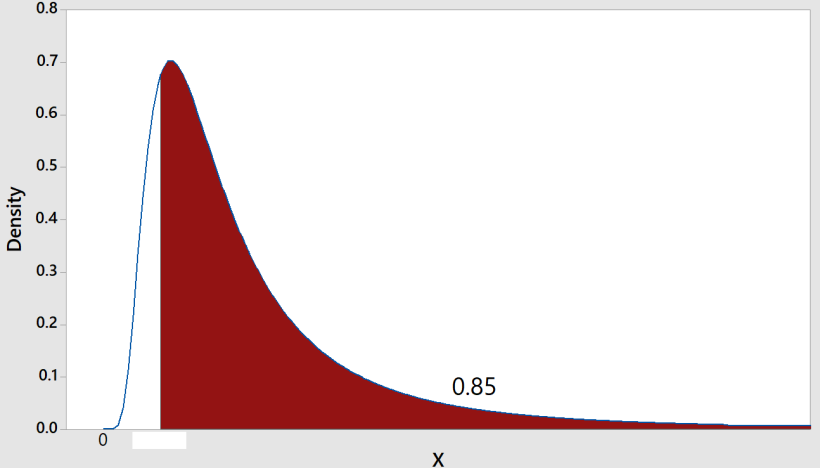
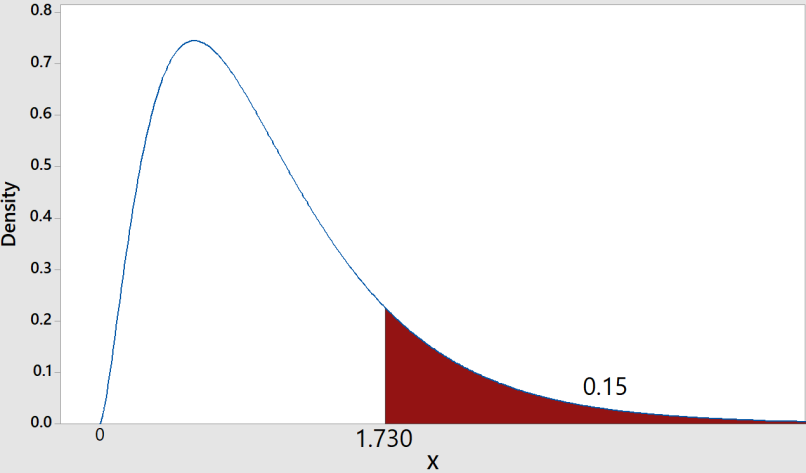
**Problem 12**

Two F-distributions are plotted below(shaded region denotes probability). Fig. 1a has 5 numerator degrees of freedom and unknown νdenominatordegrees of freedom.

**Note:** OnlyParts a-c refer to Fig.1

1. Use the information given in the Fig. 1(**a)** and find ν**.**
2. After interchanging the numerator and denominator degrees of freedom, we get Fig. 1(b). Find the numerical value of **A**.
3. Find the mean and variances for both the distributions.
4. If the numerator and denominator degrees of freedom are *identical*, find f such that

Then what is ?



**A**

**1(a) 1(b)**

**Figure 1.** F-distribution plots for question 12

**[9]**

**Problem 13.**

Answer the following briefly

1. How will demonstrate that your experimental data are affected by only random errors?
2. If a hypothesis is rejected at a particular significance level, can it be accepted at a higher level of significance? When increasing the level of significance, is the confidence interval becoming broader or narrower?
3. Find the degrees of freedom (dof) in T-distribution such that P(Z>2.086404)=P(Tdof>2.5). Why is the value corresponding to the T-distribution higher than the Z value?
4. Using f0.90,4, 6 how will you find f0.10,6,4 ? Demonstrate your procedure.
5. If there are n independent standard normal variables Z1, Z2,...Zn, what will be the distribution formed by

**[3+3+3+3+2]**

**Problem 14.**

Two monitoring stations in a river test concentrations of a pollutant. In the first station, 15 samples had a standard deviation of 3.07 ppm while in the second station, 12 samples had a standard deviation of 0.80. Construct a 98% confidence interval for .

**[4]**

**Problem 15**

A normal distribution of marks has mean 50 and standard deviation 15. Find the following using Matlab

1. Show that P(X>90)=P(X<10)
2. For what value of marks **M** is the probability 0.5?
3. For what value of marks **M** is the probability 0.75?
4. For what value of marks **M** is the probability 0.25?
5. Solve for **v** such that P(Z>**v**) = 0.025
6. Solve for **w** such that P(X<**w**) = 0.5\*P(Z<)

**Problem 16**

Using the T-distribution, what is the degrees of freedom such that

1. P(T>1.25)=0.15 and P(T<1.25)=0.85
2. Find P(T35>0.8)

**[6]**

**Problem 17**

For the following probability values (based on the upper tail) taken from a t-distribution with 20 degrees of freedom, find the corresponding t-values.

1. p=0.5 **b.** p=0.3 **c.** p=0.7 **d.** p= 0.80 **e.** p=0.2

**[5]**

**Problem 18**

A random sample of class marks are taken in an inspection and the values of the random variables are recorded as follows

**X** = [0 15 -5 20 10 9 14]

1. Sample size
2. Sample mean
3. Sample standard deviation
4. Sample variance
5. Construct the two-sided 99% confidence interval for mean μ
6. Construct the one-sided 95% confidence interval for the lower bound on μ
7. Construct the one-sided 98% confidence interval for the upper bound on μ

**[10]**

**Problem 20**

Random samples of size 100 are drawn, with replacement, from two populations P1 and P2 and their means and are computed. If μ1 = 10 σ1 = 2, μ2 = 8 and σ2 = 1, find

1. the probability that the difference between a given pair of sample means is less than 1.5
2. the probability that the difference between a given pair of sample means is greater than 1.75 but less than 2.5.

**[3+4]**

**Problem 21**

A machine is considered to be unsatisfactory if it produces more than8% defectives. It is suspected that the machine is unsatisfactory. A random sample of 120 items produced by the machine contains 14 defectives. Does the sample evidence support the claim that the machine is unsatisfactory? Use α = 0.01.

**[4]**

**Problem 22**

A physician claims that the variance in cholesterol levels of adult men in a certain laboratory is at least 100. A random sample of 25 adult males from this laboratory produced a sample standard deviation of cholesterol levels as 12. Test the physician’s claim at 5% level of significance.

**[4]**

**Problem 23**

Find out how to conduct a t-test on difference of population means involving small samples and a) unknown but equal population variances and b) unknown and unequal population variances. In such cases how do you find the degrees of freedom?

**[6]**

**Problem 24**

The intelligence quotients (IQs) of 17 students from one area of a city showed a sample mean of 106 with a sample standard deviation of 10, whereas the IQs of 14 students from another area chosen independently showed a sample mean of 109 with a sample standard deviation of 7. Is there a significant difference between the IQs of the two groups at α = 0.02? Assume that the population variances are equal.

**[5]**

**Problem 25**

Show how to find the one degree of freedom chi-square distribution value when the upper 100 α percentage point is specified. For example, if the 15% point in the chi-square distribution is required, how will you find it from the normal distribution?

**Problem 26**

An environmental engineer carries out a t-test for 9 samples from a polluted lake and obtains a |t|-value of 2.306. However he has forgotten to take the t-tables and has only the F-tables. How will he estimate the required probability value using the F-tables?

**[5]**

**Problem 27**

Eagle Eye is used in cricket to track the trajectory of the ball. The equipment has been tested rigorously on many overseas cricket pitches over 5 years. After a large number of tests it uses the standard deviation (σ) in bounce of the ball pitched at good length as 50 cm in its tracking calculations. The Eagle Eye tracker is then tested through 5 independent trials in India and the measured standard deviation in the cricket bounce for the ball pitched on good length based on the measurements carried out is 25.74 cm. Can the Eagle Eye be used reliably to track the ball (to give lbw decisions) on Indian pitches?

**[6]**

**Problem 28**

How you will estimate the probability in distributions involving chi-square using the F distribution tables?

**[4]**

**Problem 29**

1. Show that



1. Using f0.90,4, 6 how will you find f0.10,6,4 ? Demonstrate your procedure.

**[4+3]**

**Problem 30**

A sample of 40 alumni from IIT Madras (batch of 2010) working in Indian Organizations is selected to find whether the average annual income may be taken to belong to a population with mean of Rs. 30 lakhs. Population variance is not known. The population may not be assumed to be normally distributed.

1. Explain briefly how will you go about testing the hypothesis that µ = 30 lakhs.
2. If global slowdown of the economy are being considered, what would be the alternate hypothesis?

To compare between IITs, a random sample of 50 alumni from IITX (also batch of 2010) is also taken.

1. How will you go about testing the null hypothesis that the mean salaries of students graduating from both Institutions are the same and the alternate hypothesis that the average salary of IITM is higher than IITX?

State the assumptions that have been made.

**[8]**