**B.M.S. COLLEGE OF ENGINEERING, BENGALURU-19**

**(***Autonomous Institute, Affiliated to VTU***)**

**DEPARTMENT OF MATHEMATICS**

**THIRD SEMESTER B.E. COURSE (CSE/ISE)**

**Course Title: Statistics and Discrete Mathematics**

**Course Code: 19MA3BSSDM**

# UNIT 4: STATISTICAL INFERENCE

## I. Test of significance for single mean

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $Z = \dfrac{\overline{x} - \mu_O}{\sigma / \sqrt{n}}$ | $H_0 : \mu = \mu_O$ | $H_1 : \mu \neq \mu_O$ | $\lvert z \rvert < z_{\alpha/2}$ |
| | | $H_1 : \mu < \mu_O$ | $z > z_{\alpha}$ |
| | | $H_1 : \mu > \mu_O$ | $z < z_{\alpha}$ |

1. The length of life X of certain computers is approximately normally distributed with mean 800 hours and S.D 40 hours. If a random sample of 30 computers has an average life of 788 hours, test the null hypothesis that μ=800 hours at (a) 5% (b) 1% (c) 10% (d) 15% level of significance
2. Mice with an average lifespan of 32 months will live upto 40 months when fed by a certain nutrious food. If 64 mice fed on this diet have an average lifespan of 38 months and standard deviation of 5.8 months, is there any reason to believe that average lifespan is less than 40 months.
3. A machine runs on an average of 125 hours/year. A random sample of 49 machines has an annual average use of 126.9 hours with standard deviation 8.4 hours. Does this suggest to believe that machines are used on the average more than 125 hours annually at 0.05 level of significance?
4. A company claims that the mean thermal efficiency of diesel engines produced by them is 32.3%. To test this claim, a random sample of 40 engines were examined which showed the mean thermal efficiency of 31.4% and S.D of 1.6%. Can the claim be accepted or not, at 0.01 L.O.S.?
5. It has previously been recorded that the average depth of ocean at a particular region is 67.4 fathoms. Is there reason to believe this at 0.01 L.O.S. if the readings at 40 random locations in that particular region showed a mean of 69.3 with S.D of 5.4 fathoms?
6. A company producing computers states that the mean lifetime of its computers is 1600 hours. Test this claim at 0.01 L.O.S. against the A.H.: μ < 1600 hours if 100 computers produced by this company has mean lifetime of 1570 hours with S.D. of 120 hours.
7. A manufacturer of tyres guarantees that the average lifetime of its tyres is more than 28000 miles. If 40 tyres of this company tested, yields a mean lifetime of 27463 miles with S.D. of 1348 miles, can the guarantee be accepted at 0.01 L.O.S.?

**B.M.S. COLLEGE OF ENGINEERING, BENGALURU-19**
**(*Autonomous Institute, Affiliated to VTU*)**
**DEPARTMENT OF MATHEMATICS**

## II. Test of significance for difference between two means

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $Z = \dfrac{\bar{x} - \mu_O}{\sigma / \sqrt{n}}$ | $H_0 : \mu = \mu_O$ | $H_1 : \mu \neq \mu_O$ | $\lvert z \rvert < z_{\alpha/2}$ |
| | | $H_1 : \mu < \mu_O$ | $z > z_\alpha$ |
| | | $H_1 : \mu > \mu_O$ | $z < z_\alpha$ |
| $Z = \dfrac{(\bar{x}_1 - \bar{x}_2) - \delta}{\sqrt{\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}}}$ | $H_0 : \mu_1 - \mu_2 = \delta$ | $H_1 : \mu_1 - \mu_2 \neq \delta$ | $\lvert z \rvert < z_{\alpha/2}$ |
| | | $H_1 : \mu_1 - \mu_2 > \delta$ | $z < z_\alpha$ |
| | | $H_1 : \mu_1 - \mu_2 < \delta$ | $z > z_\alpha$ |

1. In a random sample of 100 tube lights produced by company A, the mean lifetime (mlt) of tube light is 1190 hours with standard deviation of 90 hours. Also in a random sample of 75 tube lights from company B the mean lifetime is 1230 hours with standard deviation of 120 hours. Is there a difference between the mean lifetimes of the two brands of tube lights at a significance level of (a) 0.05 (b) 0.01?
2. To test the effects a new pesticide on rice production, a farm land was divided into 60 units of equal areas, all portions having identical qualities as to soil, exposure to sunlight etc. The new pesticide is applied to 30 units while old pesticide to the remaining 30. Is there reason to belive that the new pesticide is better than the old pesticide if the mean number of kgs of rice harvested/units using new pesticide (N.P.) is 496.31 with S.D. of 17.18 kgs. Test at a level of significance (a) 0.05 (b) 0.01?
3. A random sample of 40 ‘geyers’ produced by company A have a mean lifetime (mlt) of 647 hours of continuous use with a S.D. of 27 hours, while a sample 40 produced by another company B have mlt of 638 hours with S.D. 31 hours. Does this substantiate the claim of company A that their ‘geyers’ are superior to those produced by company B at (a) 0.05 (b) 0.01 L.O.S.
4. Test at 0.05 L.O.S. a manufacturer’s claim that the mean tensile strength (mts) of a tead A exceeds the mts of thread B by at least 12 kgs. If 50 pieces of each type of thread are tested under similar conditions yielding the following data:

| | Sample Size | Mts (kgs) | S.D. (kgs) |
|---|---|---|---|
| Type A | 50 | 86.7 | 6.28 |
| Type B | 50 | 77.8 | 5.61 |

5. Test the N.H.: $\mu_A - \mu_B = 0$ against the A.H.: $\mu_A - \mu_B \neq 0$ at 0.01 L.O.S. for the following data

| | Sample Size | Mts (kgs) | S.D. (kgs) |
|---|---|---|---|
| Type A | 40 | 247.3 | 15.2 |
| Type B | 30 | 254.1 | 18.7 |

6. If a random sample data show that 42 men earn on the average $\bar{x}_1 = 744.85$ with S.D. $s_1 = 397.7$ while 32 women earn on the average $\bar{x}_2 = 516.78$ with S.D. $s_2 = 162.523$, test at 0.05 level of significance whether the average income for men and women is same or not.

7. A company claims that alloying reduces resistance of electric wire by more than 0.05 ohm. To test this claim samples of standard wire and alloyed wire are tested yielding the following results:

| Type of wire | Sample Size | Mean resistance (ohms) | S.D. (ohms) |
|---|---|---|---|
| Standard | 32 | 0.136 | 0.004 |
| Alloyed | 32 | 0.083 | 0.005 |

## III. Test of Hypothesis: one proportion: large sample

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $Z=\dfrac{x-np_O}{\sqrt{np_Oq_O}}$ | $H_0 : p = p_O$ | $H_1 : p \neq p_O$ | $\lvert z\rvert < z_{\alpha/2}$ |
| | | $H_1 : p < p_O$ | $z > z_\alpha$ |
| | | $H_1 : p > p_O$ | $z < z_\alpha$ |

1. If in a random sample of 600 cars making a right turn at a certain traffic junction 157 drove into the wrong lane, test whether actually 30% of all drivers make this mistake or not at this given junction. Use (a) 0.05 (b) 0.01 L.O.S.
2. Test the claim of a manufacturer that 95% of his ‘stabilizers’ confirm to ISI specifications if out of a random sample of 200 stabilizers produced by this manufacturer 18 were faulty. Use (a) 0.01 (b) 0.05 L.O.S.
3. If in a random sample of 200 persons suffering with ‘headache’ 160 persons got cured by a drug, can we accept the claim of the manufacturer that his drug cures 90% of the sufferers? Use 0.01 L.O.S.
4. A student answers by guess 32 questions correctly in an examination with 50 true or false questions. Are the results significant at (a) 0.05 L.O.S. (b) 0.01 L.O.S.?
5. A hospital claims that at least 40% of the patients admitted are for ‘emergency’ ward. Is there reason to believe this claim if the records shows that only 49 of 150 patients are for ‘emergency’ ward. Use 0.01 L.O.S.

## IV. Test of Hypothesis: Difference between proportions with large sample

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $Z=\dfrac{(\overline{p}_1-\overline{p}_2)-\delta}{\sqrt{\widehat{p}\widehat{q}\left(\dfrac{1}{n_1}+\dfrac{1}{n_2}\right)}}$ where $\hat{p}=\dfrac{x_1+x_2}{n_1+n_2}=\dfrac{n_1p_1+n_2p_2}{n_1+n_2}$, $\hat{q}=1-\hat{p}$ | $H_0 : P_1 - P_2 = \delta$ | $H_1 : P_1 - P_2 \neq \delta$ | $\lvert z\rvert < z_{\alpha/2}$ |
| | | $H_1 : P_1 - P_2 < \delta$ | $z > z_\alpha$ |
| | | $H_1 : P_1 - P_2 > \delta$ | $z < z_\alpha$ |

1. Out of two vending machines at a super bazar', the first machine fails to work 13 times in 250 trials and second machine fails to work 7 times in 250 trials. Test at 0.05 L.O.S. whether the difference between the corresponding sample proportions is significant.
2. If 57 out of 150 patients suffering with certain disease are cured by allopathy and 33 out of 100 patients with same disease are cured by homeopathy, is there reason to believe that allopathy is better than homeopathy at 0.05 L.O.S.
3. A question in a true-false quiz is considered to be smart if it discriminates between intelligent person (IP) and average person (AP). Suppose 205 of 250 IP's and 137 of 250 AP's answer a quiz question correctly. Test at 0.01 L.O.S. whether for the given question, the proportion of correct answers can be expected to be at least 15% higher among IP's than among the AP's.
4. A random sample of 200 parents from urban areas 120, while 240 of 500 parents from rural areas preferred 'private' professional colleges, can we conclude that parents from urban areas prefer 'private' colleges at 0.025 L.O.S.?
5. A study of TV viewers was conducted to find the opinion about the mega serial 'Ramayana'. If 56% of a sample of 300 viewers from south and 48% of 200 viewers from north preferred the serial, test the claim at 0.05 L.O.S. that (a) there is a difference of opinion between south and north (b) 'Ramayana' is preferred in the south.

## V. Small Sample Test Concerning Single Mean: t-Distribution

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $t=\dfrac{\bar{x}-\mu_O}{S/\sqrt{n}}$ here $S^2=\dfrac{\sum(x_i-\bar{x})^2}{n-1}$ OR $t=\dfrac{\bar{x}-\mu_O}{s/\sqrt{n-1}}$ here $s^2=\dfrac{\sum(x_i-\bar{x})^2}{n}$ | $H_0:\mu=\mu_O$ | $H_1:\mu\neq\mu_O$ | $\lvert t\rvert<t_{\alpha/2,n-1}$ |
| | | $H_1:\mu<\mu_O$ | $t>t_\alpha$ |
| | | $H_1:\mu>\mu_O$ | $t<t_\alpha$ |

1. An ambulance service company claims that on an average it takes 20 minutes between a call for an ambulance and the patient's arrival at the hospital. If in 6 calls the time taken (between a call and arrival at hospital) are 27, 18, 26, 15, 20 and 32. Can the company's claim be accepted?
2. Mean lifetime of computers manufactured by a company is 1120 hours with standard deviation of 125 hours. (a) Test the typothesis that mean lifetime of computers has not changed if a sample of 8 computers has a mean life time of 1070 hours (b) Is there decrease in mlt? Use (i) 5% (ii) 1% L.O.S.
3. An auditor claims that he takes on an average 10.5 days to file income tax returns (I.T. returns). Can this claim be accepted if a random sample shows that he took 13, 19, 15, 10, 12, 11, 14, 18 days to file I.T. returns? Use (a) 0.01 (b) 0.05 L.O.S.

4. If 5 pieces of certain ribbon selected at random have mean breaking strength of 169.5 pounds with S.D. of 5.7, do they confirm to the specification mean breaking strength of 180 pounds.
5. In a random sample of 10 bolts produced by a machine the mean length of bolt is 0.53 mm and S.D 0.03 mm. Can we claim from this that the machine is in proper working order if in the past it produced bolts of length 0.50 mm? Use (a) 0.05 (b) 0.01 L.O.S.

## VI. Small Sample Test Concerning Difference Between Two Means: t-Distribution

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $t=\dfrac{(\bar{x}_1-\bar{x}_2)-\delta}{S\sqrt{\dfrac{1}{n_1}+\dfrac{1}{n_2}}}$ <br> (i) $S^2=\dfrac{\sum(x_{1i}-\bar{x}_1)^2+\sum(x_{2i}-\bar{x}_2)^2}{n_1+n_2-2}$ <br> OR <br> (ii) $S^2=\dfrac{n_1s_1^2+n_2s_2^2}{n_1+n_2-2}$ <br> if $s_1^2=\dfrac{\sum(x_{1i}-\bar{x}_1)^2}{n_1}$ and $s_2^2=\dfrac{\sum(x_{2i}-\bar{x}_2)^2}{n_2}$ <br> OR <br> (iii) $S^2=\dfrac{(n_1-1)S_1^2+(n_2-1)S_2^2}{n_1+n_2-2}$ <br> if $S_1^2=\dfrac{\sum(x_{1i}-\bar{x}_1)^2}{n_1-1}$ and $S_2^2=\dfrac{\sum(x_{2i}-\bar{x}_2)^2}{n_2-1}$ | $H_0:\mu_1-\mu_2=\delta$ | $H_1:\mu_1-\mu_2\neq\delta$ | $\lvert t\rvert<t_{\alpha/2,\,n_1+n_2-2}$ |
| | | $H_1:\mu_1-\mu_2>\delta$ | $t<t_\alpha$ |
| | | $H_1:\mu_1-\mu_2<\delta$ | $t>t_\alpha$ |

1. In a mathematics examination 9 students of class A and 6 students of class B obtained the following marks. Test at 1% L.O.S. whether the performance in mathematics is same or not for the two classes A and B. assume that the samples are drawn from normal populations having same variance.

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| A: | 44 | 71 | 63 | 59 | 68 | 46 | 69 | 54 | 48 |
| B: | 52 | 70 | 41 | 62 | 36 | 50 | | | |

2. Out of random sample of 9 mice, suffering with a disease, 5 mice were treated with a new serum while the remaining were not treated. From the time commencement of experiment, the following are the survival times:

| | | | | | |
|---|---|---|---|---|---|
| Treatment | 2.1 | 5.3 | 1.4 | 4.6 | 0.9 |
| NoTreatment | 1.9 | 0.5 | 2.8 | 3.1 | |

Test whether the serum treatment is effective in curing the disease at 5% L.O.S., assuming that the two distributions are normally distributed with equal variances.

3. Random samples of specimens of coal from two mines A and B are drawn and their heat producing capacity (in millions of calories/ton) were measured yielding the following results:

| | | | | | | |
|---|---|---|---|---|---|---|
| Mine A: | 8350 | 8070 | 8340 | 8130 | 8260 | |
| Mine B: | 7900 | 8140 | 7920 | 7840 | 7890 | 7950 |

Is there significant difference between the means of these two samples at 1% L.O.S.?

4. A study is conducted to determine whether the wear of material A exceeds that of B by more than 2 units. If test of 12 pieces of material A yielded a mean wear of 85 units and S.D. of 4 while test of 10 pieces of material B yielded a mean of 81 and S.D. 5, what conclusion can be drawn at 5% L.O.S. Assume that populations are approximately normally distributed with equal variances.

5. To determine whether vegetarian and non-vegetarian diets effects significantly on increase in weight a study was conducted yielding the following data of gain in weight.

| | | | | | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Veg | 34 | 24 | 14 | 32 | 25 | 32 | 30 | 24 | 30 | 31 | 35 | 25 | | | |
| Non-Veg | 22 | 10 | 47 | 31 | 44 | 34 | 22 | 40 | 30 | 32 | 35 | 18 | 21 | 35 | 29 |

Can we claim that the two diets differ pertaining to weight gain, assuming that samples are drawn from normal populations with same variance?

VII. **Paired Sample t-Test**

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $t=\dfrac{\bar{d}-\mu_d}{S_d/\sqrt{n}}$ here $S_d^2=\dfrac{\sum\left(d_i-\bar{d}\right)^2}{n-1}$ OR $t=\dfrac{\bar{d}-\mu_d}{S_d/\sqrt{n-1}}$ here $S_d^2=\dfrac{\sum\left(d_i-\bar{d}\right)^2}{n}$ | $H_0:\mu=\mu_d$ | $H_1:\mu\neq\mu_d$ | $\lvert t\rvert<t_{\alpha/2,n-1}$ |
| | | $H_1:\mu<\mu_d$ | $t>t_{\alpha}$ |
| | | $H_1:\mu>\mu_d$ | $t<t_{\alpha}$ |

1. Use paired sample test at 0.05 L.O.S. to test from the following data whether the differences of means of the weights obtained by two different scales (weighting machines) is significant.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Weight (gms) | Scale: I | 11.23 | 14.36 | 8.33 | 10.50 | 23.42 | 9.15 | 13.47 | 6.47 | 12.40 | 19.38 |
| | Scale: II | 11.27 | 14.41 | 8.35 | 10.52 | 23.41 | 9.17 | 13.52 | 6.46 | 12.45 | 19.35 |

2. The average weekly losses of man-hours due to strikes in an institute before and after a disciplinary program was implemented are as follows: Is there reason to believe that the disciplinary program is effective at 5% L.O.S.?

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| Before | 45 | 73 | 46 | 124 | 33 | 57 | 83 | 34 | 26 | 17 |

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| After | 36 | 60 | 44 | 119 | 35 | 51 | 77 | 29 | 24 | 11 |

3. The pulsality index (P.I.) of 11 patients before and after contracting a disease are given below. Test at 0.05 L.O.S. whether there is a significant increase of the mean of P.I. values.

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Before | 0.4 | 0.45 | 0.44 | 0.54 | 0.48 | 0.62 | 0.48 | 0.60 | 0.45 | 0.46 | 0.35 |
| After | 0.5 | 0.60 | 0.57 | 0.65 | 0.63 | 0.78 | 0.63 | 0.80 | 0.69 | 0.62 | 0.68 |

4. The blood pressure (B.P.) of 5 women before and after intake of a certain drug are given below:

| | | | | | |
|---|---|---|---|---|---|
| Before | 110 | 120 | 125 | 132 | 125 |
| After | 120 | 118 | 125 | 136 | 121 |

Test at 1% L.O.S. whether there is significant change in B.P.

5. Marks obtained in mathematics by 11 students before and after intensive coaching are given below:

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| Before | 24 | 17 | 18 | 20 | 19 | 23 | 16 | 18 | 21 | 20 | 19 |
| After | 24 | 20 | 22 | 20 | 17 | 24 | 20 | 20 | 18 | 19 | 22 |

Test at 5% L.O.S. whether the intensive coaching is useful?

## VIII. Ratio of Variances: F-Distributions

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $F=\frac{S_1^2}{S_2^2}$ if $S_1^2>S_2^2$ <br> OR <br> $F=\frac{S_2^2}{S_1^2}$ if $S_2^2>S_1^2$ <br> here <br> $S_1^2=\frac{n_1s_1^2}{n_1-1}=\frac{\sum(x_{1i}-\bar{x}_1)^2}{n_1-1}$; <br> $S_2^2=\frac{n_2s_2^2}{n_2-1}=\frac{\sum(x_{2i}-\bar{x}_2)^2}{n_2-1}$ | $H_0:\sigma_1^2=\sigma_2^2$ | $H_1:\sigma_1^2\neq\sigma_2^2$ | $F<F_{n_1-1,n_2-1,\alpha}$ <br> OR <br> $F<F_{n_2-1,n_1-1,\alpha}$ |

1. The household net expenditure on health care in south and north India, in two samples of households, expressed as percentage of total income is shown the following table:

| | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|
| South | 15 | 8 | 3.8 | 6.4 | 27.4 | 19 | 35.3 | 13.6 | |
| North | 18.8 | 23.1 | 10.3 | 8 | 18 | 10.2 | 15.2 | 19 | 20.2 |

Test the equality of variances of household's net expenditure on health care in south and north India.

2. Can we conclude that the two population variances are equal for the following data of post graduates passed out from a state and private university.

| | | | | | | |
|---|---|---|---|---|---|---|
| State: | 8350 | 8260 | 8130 | 8340 | 8070 | |
| Private: | 7890 | 8140 | 7900 | 7950 | 7840 | 7920 |

3. Is there reason to believe that the life expected in south and north India is same of not from the following data

| | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| South: | 34 | 39.2 | 46.1 | 48.7 | 49.4 | 45.9 | 55.3 | 42.7 | 43.7 | | |
| North: | 49.7 | 55.4 | 57 | 54.2 | 50.4 | 44.2 | 53.4 | 57.5 | 61.9 | 56.6 | 58.2 |

## IX. Chi-Square Distribution: Goodness of Fit

| Test statistic | Null Hypothesis | Alternative Hypothesis | Accept $H_o$ |
|---|---|---|---|
| $\chi^2 = \dfrac{\sum_{i=1}^{n}(o_i - e_i)^2}{e_i}$ | $H_0$ : There is no significant difference between experimental and theoretical values | $H_1$ : There is significant difference between experimental and theoretical values | $\chi^2 < \chi^2_{n-k,\alpha}$ |

1. Test for goodness of fit of a Poisson distribution at 5% L.O.S. to the following frequency distribution:

| No. of patients arriving/hour: (x) | 0 | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 |
|---|---|---|---|---|---|---|---|---|---|
| Frequency | 52 | 151 | 130 | 102 | 45 | 12 | 5 | 1 | 2 |

2. A sample analysis of examination results of 500 students was made. It was found that 220 students had failed, 170 had secured third class, 90 had secured second class and 20 had secured first class. Do these figures support the general examination result which is in the ration 4:3:2:1 for the respective categories?

*******