# 18MAB301T - Probability and Statistics

## Unit – 3
## Testing of Hypothesis

Dr.R.Varadharajan
Assistant Professor (S.G)
Department of Mathematics
SRM Institute of Science and Technology
Kattankulathur - 603203

# *Testing of hypotheses*

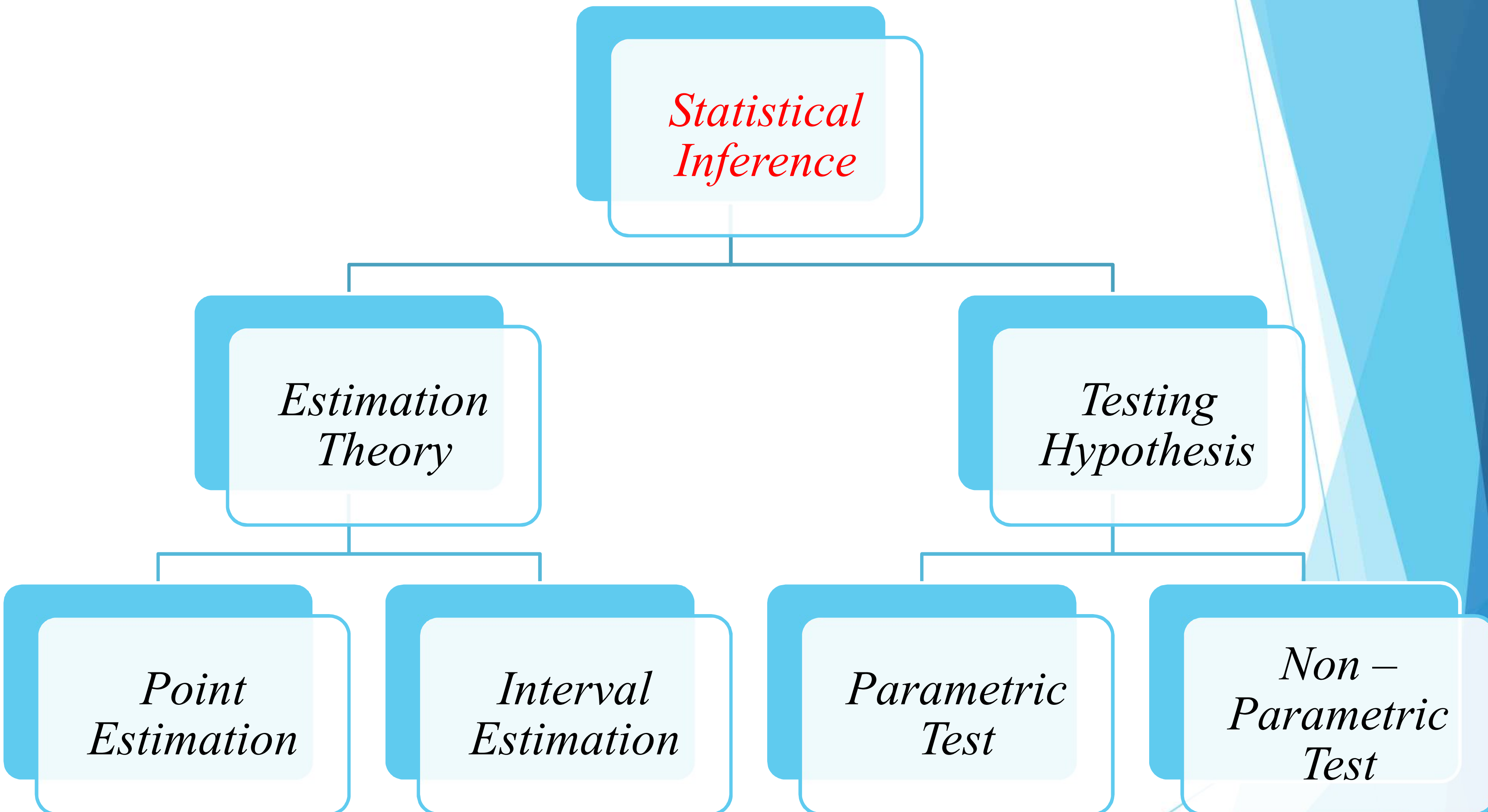
***Learning Objectives:***

***To understand the concept of testing hypothesis.***

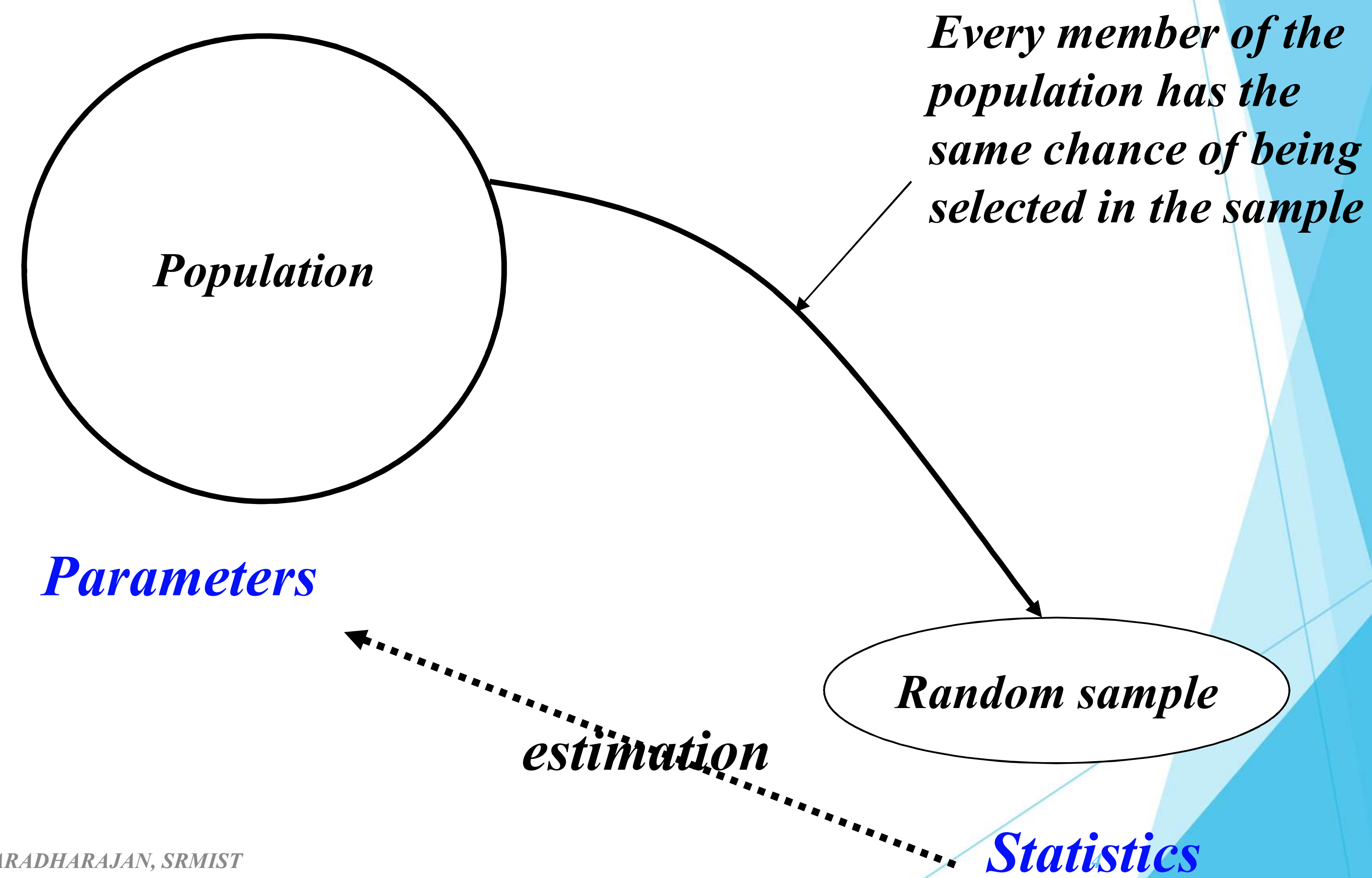
***To distinguish the null and alternative hypotheses.***

***To interpret type I and II errors***

***To understand and solve the testing hypothesis problem***


Statistical Inference
Estimation Theory
Testing Hypothesis
Point Estimation
Interval Estimation
Parametric Test
Non – Parametric Test

# *Statistical estimation*



# *What is HYPOTHESIS?*

- *A hypothesis (plural: hypotheses) is a prediction about the outcome of a scientific investigation. Like all predictions, hypotheses are based on a person's observations and previous knowledge or experience.*
- *Hypothesis can also be defined as Something taken to be true for the purpose of argument or investigation; an assumption.*
- *Some of the Synonyms for Hypothesis are Supposition, Theory, Reason and Guess.*

# *Why do you need a hypothesis?*

- *A good hypothesis will help you to focus your investigation.*

- *As you progress through your investigation you might notice that more and more information comes out.*

- *Your hypothesis will ensure that you stay on course in your investigation.*

# *Steps in Testing of hypotheses*

*Step1: Set up the hypotheses*

*Step2: Select the level of significance*

*Step3: Select the suitable test criterion*

*Step4: Doing Computation*

*Step5: Making Decision*

# *Step1: Set up the hypothesis*

- ***The first step in the process is to set up the decision making process.***
- ***This involves identifying the null and alternative hypotheses***
- ***We will be writing two types of hypotheses:***
  1. ***Null Hypothesis (H0 or HN)***
  2. ***Alternative Hypothesis (H1 or HA)***

# *Step2: Select the level of significance*

- ***In the second step, we should select the suitable Level of Significance to our given problem . Some times this may be given in the question.***

- ***The probability of a false rejection of the null hypothesis in a statistical test, it is also called significance level.***

- ***The customary used  level of significance are 5%***

# *Step3: Select the suitable test criterion*

- ***In the this step, we have to select the suitable test criterion to our given problem .***

- ***For example if we have the small sample information, we can choose the t – test to test null hypothesis against the alternative hypothesis.***

# *Step4: Doing Computation*

- ***Based on the selected test criterion, we can calculate the necessary information's.***

- ***For example if we have the small sample information, we can choose the t – test to test null hypothesis against the alternative hypothesis.***

***That is*** $$|t| = \left| \frac{\bar{X} - \mu}{\sqrt{s^2 / n - 1}} \right|$$

***Based on the above statistics, we can calculate*** $s^2 and \bar{X}$

# *Step5: Making Decision*

- ***This is the final step in the testing hypothesis procedure, we can write the conclusion Based on the calculated value and the table value.***

- ***If the calculated value is less than the respective distribution table value, we can accept the null hypothesis.***

- ***Otherwise we can reject the null hypothesis. That is, If the calculated value is greater than the respective distribution table value, we can reject the null hypothesis.***

# *Type I and Type II Errors*

| | Population Condition | |
|---|---|---|
| Conclusion | $H_0$ True | $H_0$ False |
| Accept $H_0$ | Correct Decision | Type II Error |
| Reject $H_0$ | Type I Error | Correct Decision |

**False Negative**

**False Positive**

# *Test of Significance*

1. ***Test of significance for Attributes***
   - ***a. Test for Single Proportion***
   - ***b. Test for the Difference between two population proportions test***

2. ***Test of significance for Variables***
   - ***a. Small sample test***
     - ***(i) Test for Single Mean***
     - ***(ii) Test for the Difference between two means***
     - ***(iii) Paired t - test***
   - ***b. Large sample test***
     - ***(i) Test for Single Mean***
     - ***(ii) Test for the Difference between two means***

3. ***F – Test***

# *Test of Significance for Attributes*

## *(a). Test for Single Proportion*

*To test the null hypothesis if the population proportion can be a given value, we take the null hypothesis*

*Ho=P=P0 (given)*

*When n is large (greater than 30 for practical problem) we know*

$$|Z| = \left| \frac{p - P_0}{\sqrt{PQ/n}} \right|$$

*If* $|Z| < Z_\alpha$ *, we accept the null hypothesis, where* $Z_\alpha$ *is the table value of normal distribution at* $\alpha\%$ *level of significance.*

***Problem 1:***

***8 oranges were found to be bad in a sample of 100 drawn at random from a lot of oranges. Can we conclude that the lot contains 10% bad oranges.***

***Solution:***

***Given that***

***Sample size = n=100***

***No. of bad oranges x = 8***

***Sample proportion***

$$p = \frac{x}{n} = \frac{8}{100} = 0.08$$

***To test if the lot contains 10% bad oranges, we take the null hypothesis***

***Ho=P= 0.10 (given)***

***The level of significance is 5%.***

***The test statistic to be used when n is large (greater than 30 for practical problem) we know***

$$|Z| = \left|\frac{p - P_0}{\sqrt{PQ/n}}\right| = \left|\frac{0.08 - 0.10}{\sqrt{(0.10)(1-0.10)/100}}\right| = 0.667$$

***Inference:***

***Since the calculated value is less than the normal distribution table value (0.667<1.96). so we accept the null hypothesis. It may be conclude that the lot may contain 10% bad oranges.***

***Contd..***

***(b). Test for the Difference between two population proportions***

***To test the null hypothesis if two population proportion can be a given value, we take the null hypothesis***

***Ho=P1=P2=P (say)***

***When n1 and n2 are large (greater than 30 for practical problem) we know***

$$|Z| = \left| \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right|$$

***If*** $|Z| < Z_\alpha$ ***, we accept the null hypothesis, where*** $Z_\alpha$ ***is the table value of normal distribution at*** $\alpha\%$ ***level of significance.***

***Problem 2:***

***20 persons were identified as a smokers in a sample of 80 drawn from a village. Similarly 60 persons founded to be a smokers in a sample of 120 drawn at random from another village. Test whether the ratio of smokers in both villages are equal.***

***Solution:***

***We are given, No. of smokers in sample 1:x1=20, sample size n1 = 80***

***sample proportion for 1st sample =*** $p_1 = \frac{x_1}{n_1} = \frac{20}{80} = 0.25$

***No. of smokers in sample 2:x2=60, sample size n2 = 120***

***sample proportion for 1st sample =*** $p_2 = \frac{x_2}{n_2} = \frac{60}{120} = 0.5$

***we take the null hypothesis***

***Ho=P1=P2 (ie) The ratio of smokers in both villages are equal.***

***The level significance is 5% .***

***Contd..***

***The test statistic to be used when n is large (greater than 30 for practical problem) we know***

$$|Z| = \left| \frac{p_1 - p_2}{\sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| = \left| \frac{0.25 - 0.5}{\sqrt{(0.4)(0.6)\left(\frac{1}{80} + \frac{1}{120}\right)}} \right| = 3.535$$

***Where,***

$$P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{80(0.25) + 120(0.5)}{80 + 120} = 0.4$$

$$Q = 1 - P = 1 - 0.4 = 0.6$$

***The normal distribution table for 5% L.S is 1.967***

***Inference:***

***Since the calculated value is greater than the normal distribution table value 1.967 at 5% L.S. So we reject the null hypothesis. It may be conclude that the ratio of smokers in both villages are not equal.***

# *Test of Significance for Variables (<30)*

## *(a). Test for Single mean when population variance is unknown.*

***To test the null hypothesis if the population mean can be a given value, we take the null hypothesis***

$$H_0 = \mu = \mu_0$$

***The test statistic to be used is***

$$|t| = \left| \frac{\bar{x} - \mu_0}{\sqrt{s^2 / n - 1}} \right| \sim t$$ ***distribution with (n-1) degrees of freedom***

***If*** $|t| < t_\alpha$ ***, we accept the null hypothesis, where*** $t_\alpha$ ***is the table value of t – distribution at*** $\alpha\%$ ***level of significance.***

***Problem 3:***

***If 40, 42, 50, 60, 45, 40, 55, 58, 62, 60 are the sample observations drawn from a normal population. Test if the population mean can be equal to 50.***

***Solution:***

***Given that***

***Sample size = n=10***

***Population mean*** $\mu = 50$

***To test if the population mean can be a given value, we take the null hypothesis***

$$H_0 = \mu = \mu_0 = 50$$

***The level of significance is 5%.***

***The test statistic to be used when n is small***

$$|t| = \left| \frac{\bar{x} - \mu_0}{\sqrt{s^2 / n - 1}} \right| \sim t$$ ***- distribution with (n-1) degrees of freedom***

***Contd…***

| | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| ***x:*** | ***40*** | ***42*** | ***50*** | ***60*** | ***45*** | ***40*** | ***55*** | ***58*** | ***62*** | ***60*** |
| $x^2$ **:** | ***1600*** | ***1764*** | ***2500*** | ***3600*** | ***2025*** | ***1600*** | ***3055*** | ***3364*** | ***3844*** | ***3600*** |

$$\sum x = 512; \sum x^2 = 29622; n = 10$$

$$\bar{x} = \frac{\sum x}{n} = \frac{512}{10} = 51.2$$

$$s^2 = \frac{\sum x^2}{n} - \left(\frac{\sum x}{n}\right)^2 = \frac{29622}{10} - (51.2)^2 = 70.76$$

$$|t| = \left|\frac{\bar{x} - \mu_0}{\sqrt{s^2 / n - 1}}\right| = \left|\frac{51.2 - 50}{\sqrt{70.76/10 - 1}}\right| = 0.427$$

***The 5% t distribution table value is 2.262.***

***Inference:***

***Since the calculated value is less than the t distribution table value at 5% L.S with n-1=10-1=9 d.f. So we accept the Ho. It may be conclude that the population mean is equal to 50.***

***Contd…***

***(b). Test for the difference between two means when population variance are unknown.***

***To test the null hypothesis if the population means are equal, we take the null hypothesis***

***The test statistic to be used is***

$$H_0 = \mu_1 = \mu_2$$

$$|t| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \sim t$$ ***distribution with (n1+n2-2) degrees of freedom***

***If*** $|t| < t_\alpha$ ***, we accept the null hypothesis, where*** $t_\alpha$ ***is the table value of t – distribution at*** $\alpha\%$ ***level of significance.***

***Problem 4:***

***Test whether the following samples could have come from the normal population with equal mean.***

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| ***Sample 1:30*** | ***20*** | ***25*** | ***35*** | ***40*** | ***20*** | ***35*** | ***30*** |
| ***Sample2:50*** | ***20*** | ***20*** | ***40*** | ***55*** | ***30*** | ***25*** | ***25*** |
| ***60*** | ***20*** | | | | | | |

***Solution:***

***Given that*** $n_1 = 8$ ***and*** $n_2 = 10$

***We take the null hypothesis,*** $H_0 = \mu_1 = \mu_2$

***since the sample sizes are small and population variance are unknown then the test statistic to be used is***

$$|t| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \right| \sim t$$

***- distribution with (n1+n2-2) degrees of freedom***

| $x_1$ | $x_2$ | $x_1^2$ | $x_2^2$ |
|---|---|---|---|
| 30 | 50 | 900 | 2500 |
| 20 | 20 | 400 | 400 |
| 25 | 20 | 625 | 400 |
| 35 | 40 | 1225 | 1600 |
| 40 | 55 | 1600 | 3025 |
| 20 | 30 | 400 | 900 |
| 35 | 25 | 1225 | 625 |
| 30 | 25 | 900 | 625 |
| | 60 | | 3600 |
| | 20 | | 400 |
| 235 | 345 | 7275 | 14075 |

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = 235/8 = 29.37$$

$$s_1^2 = \frac{\sum x_1^2}{n_1} - \left(\frac{\sum x_1}{n_1}\right)^2 = \frac{7275}{8} - (29.37)^2 = 46.77$$

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = 345/10 = 34.5$$

$$s_2^2 = \frac{\sum x_2^2}{n_2} - \left(\frac{\sum x_2}{n_2}\right)^2 = \frac{14075}{10} - (34.5)^2 = 217.25$$

$$S^2 = \frac{n_1 s_1^2 + n_2 s_2^2}{n_1 + n_2 - 2} = \frac{8(46.77) + 10(217.25}{8 + 10 - 2} = 159.166$$

$$|t| = \left|\frac{\bar{x}_1 - \bar{x}_2}{\sqrt{s^2\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}\right| = \left|\frac{29.37 - 34.5}{\sqrt{159.166\left(\frac{1}{8} + \frac{1}{10}\right)}}\right| = 0.85$$

***The t – distribution table value at 5% L.S with (n1+n2-2)=(8+10-2)=16 d.f. is 1.746***

***Inference:***

***Since the calculated value is less than the t – distribution table value at 5% L.S with (n1+n2-2)=(8+10-2)=16 d.f. So we accept the Ho.***

## *(c). Test for Single mean (Paired t – test)*

***To test the null hypothesis if the population means are equal when population variance are unknown and sample sizes are small and equal also the sample observations are related, we take the null hypothesis***

$$H_0 = \mu_1 = \mu_2$$

***The test statistic to be used is***

$$|t| = \left|\frac{\bar{d}}{\sqrt{s^2 / n - 1}}\right| \sim t$$ ***distribution with (n-1) degrees of freedom***

***If*** $|t| < t_\alpha$ ***, we accept the null hypothesis, where*** $t_\alpha$ ***is the table value of t – distribution at*** $\alpha\%$ ***level of significance.***

***Note:*** $d_i = (x_{1i} - x_{2i}); i = 1,2,3 \ldots n$

$$\bar{d} = \frac{\sum d}{n} \qquad s^2 = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2$$

***Problem 5:***

***To verify whether a course in accounting improved performance, a similar test was given to 12 participants both before and after the course. The original marks recorded in alphabetical order of the participants- were 44, 40, 61, 52, 32, 44, 70, 41, 67, 72, 53, and 72. After the course, the marks were in the same order, 53, 38, 69, 57, 46, 39, 73, 48, 73, 74, 60, and 78. Was the course useful?***

***Solution:***

***Let us take the hypothesis that there is no difference in the marks obtained before and after the course. i.e the course has not been useful.***

***The level of significance is 5%.***

***Then the test statistic to be used is*** $|t| = \left| \dfrac{\bar{d}}{\sqrt{s^2/n-1}} \right| \sim t$ ***- distribution with (n-1) d.f.***

***Where,***

$$d_i = (x_{1i} - x_{2i}); i = 1,2,3 \ldots n$$

$$\bar{d} = \frac{\sum d}{n}$$

$$s^2 = \frac{\sum d^2}{n} - \left( \frac{\sum d}{n} \right)^2$$

| Participants | Before $x_1$ | After $x_2$ | $d = x_1 - x_2$ | $d^2$ |
|---|---|---|---|---|
| A | 44 | 53 | 9 | 81 |
| B | 40 | 38 | -2 | 4 |
| C | 61 | 69 | 8 | 64 |
| D | 52 | 57 | 5 | 25 |
| E | 32 | 46 | 14 | 196 |
| F | 44 | 39 | -5 | 25 |
| G | 70 | 73 | 3 | 9 |
| H | 41 | 48 | 7 | 49 |
| I | 67 | 73 | 6 | 36 |
| J | 72 | 74 | 2 | 4 |
| K | 53 | 60 | 7 | 49 |
| L | 72 | 78 | 6 | 36 |
| Total | | | 60 | 578 |

$$\bar{d} = \frac{\sum d}{n} = 60/12 = 5$$

$$s^2 = \frac{\sum d^2}{n} - \left(\frac{\sum d}{n}\right)^2 = \frac{578}{12} - (5)^2 = 23.1667$$

$$|t| = \left|\frac{\bar{d}}{\sqrt{s^2/n-1}}\right| = \left|\frac{5}{\sqrt{23.1667/12-1}}\right| = 3.443$$

***The t – distribution table value at 5% L.S with (n-1)=(12-1)=11 d.f. is 2.201***

***Inference:***

***Since the calculated value is greater than the t – distribution table value at 5% L.S with (n-1)=(12-1)=11 d.f. So we reject the Ho. Hence the course has been useful.***

# *Test of Significance for Variables (>30)*

## *(a). Test for Single mean when population variance is known.*

***To test the null hypothesis if the population mean can be a given value when population variance is known, we take the null hypothesis***

$$H_0 = \mu = \mu_0$$

***The test statistic to be used is***

$$|Z| = \left| \frac{\bar{x} - \mu_0}{\sqrt{\sigma^2 / n}} \right| \sim Z(0,1)$$

***If*** $|Z| < Z_\alpha$ ***, we accept the null hypothesis, where*** $Z_\alpha$ ***is the table value of normal distribution at*** $\alpha\%$ ***level of significance.***

***Problem 7:***

***The mean height obtained from a random sample of size 100 is 64 inches. The standard deviation of the distribution of height of the population is known to be 3 inches. Test the statement that the mean height of the population is 67 inches at 5% L.S.***

***Solution:***

***Given that***

***Population mean*** $\mu = 67$, ***Population SD*** $\sigma = 3$

***Sample size = n=100, Sample mean*** $\bar{x} = 64$

***To test if the population mean can be a given value, we take the null hypothesis***

$$H_0 = \mu = \mu_0 = 67$$

***The level of significance is 5%.***

***The test statistic to be used when n is small***

$$|Z| = \left|\frac{\bar{x} - \mu_0}{\sqrt{\sigma^2 / n}}\right| \sim Z(0,1), \textit{ then } |Z| = \left|\frac{\bar{x} - \mu_0}{\sqrt{\sigma^2 / n}}\right| = \left|\frac{64 - 67}{\sqrt{9/100}}\right| = 10$$

***Inference:***

***Since the calculated value is greater than the normal distribution table value, so we reject the Ho. Hence the mean height of the population could not be 67 inches.***

***Contd…***

***(b). Test for the difference between two means when population variance are known.***

***To test the null hypothesis if the population means are equal when the population variances are known, we take the null hypothesis***

***The test statistic to be used is***

$$H_0 = \mu_1 = \mu_2$$

$$|Z| = \left| \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left( \frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2} \right)}} \right| \sim N(0,1)$$

***If $|Z| < Z_\alpha$ , we accept the null hypothesis, where $Z_\alpha$ is the table value of normal distribution at $\alpha\%$ level of significance.***

**Problem 8:**

**Intelligence test on two groups of boys and girls gave the following results:**

| | Mean | S.D | n |
|---|---|---|---|
| Girls | 75 | 15 | 150 |
| Boys | 70 | 20 | 250 |

**Is there a significant difference in the mean scores obtained by boys and girls?**

**Solution:**

**Let us take the hypothesis that there is no significant difference in the mean scores obtained by boys and girls.**

$$H_0 = \mu_1 = \mu_2$$

**The level of significance is 1%.**

**then the test statistic to be used is** $|Z| = \left| \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\left(\dfrac{\sigma_1^2}{n_1} + \dfrac{\sigma_2^2}{n_2}\right)}} \right| \sim N(0,1)$

$$|Z| = \left| \frac{75-70}{\sqrt{\dfrac{(15)^2}{150} + \dfrac{(20)^2}{250}}} \right| = 2.84$$

**Inference:**

**Since the calculated value is greater than the normal distribution table value2.58, so we reject the Ho. Hence there seems to be a significant difference in the mean scores obtained by boys and girls.**

***Contd…***

***Test for the ratio of two variances when the population means are unknown.***

***To test whether two independent samples x and y have been drawn from normal population with same variance (or) two independent estimates of the population variance are homogeneous, we take the null hypothesis***

$$H_0 = \sigma_1^{\ 2} = \sigma_2^{\ 2} = \sigma^2$$

***Then the test statistic to be used is***

$$F = \frac{S_1^2}{S_2^2} \sim$$ ***F distribution with (n1-1, n2-1) d.f.***

***Reject Ho, If*** $F < F_\alpha$

***Where*** $F_\alpha$ ***is the table value of normal distribution at*** $\alpha\%$ ***level of significance.***

***Note: Here*** $S_1^2 = \dfrac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1}, S_2^2 = \dfrac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1},$ ***and the numerator always having the greater variance.***

***Problem 9:***

***Two random samples drawn from two normal populations are***

***Sample 1:*** 55 54 52 53 56 58 52 50 51 49

***Sample2:*** 108 107 105 105 106 107 104 103 104 101 105

***obtain the estimates of the variance of the population that have the same variance.***

***Solution:***

***Let us take the null hypothesis*** $H_0 = \sigma_1^2 = \sigma_2^2$ ***then the test statistic to be used is***

$$F = \frac{S_1^2}{S_2^2} \sim F \text{ distribution with } (n1-1,\ n2-1) \text{ d.f.}$$

***Where ,*** $S_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1}, S_2^2 = \frac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1}$, ***and the numerator always having the greater variance.***

| $x_1$ | $x_2$ | $(x_1 - \bar{x}_1)$ | $(x_1 - \bar{x}_1)^2$ | $(x_2 - \bar{x}_2)$ | $(x_2 - \bar{x}_2)^2$ |
|---|---|---|---|---|---|
| 55 | 108 | 2 | 4 | 3 | 9 |
| 54 | 107 | 1 | 1 | 2 | 4 |
| 52 | 105 | -1 | 1 | 0 | 0 |
| 53 | 105 | 0 | 0 | 0 | 0 |
| 56 | 106 | 3 | 9 | 1 | 1 |
| 58 | 107 | 5 | 25 | 2 | 4 |
| 52 | 104 | -1 | 1 | -1 | 1 |
| 50 | 103 | -3 | 9 | -2 | 2 |
| 51 | 104 | -2 | 4 | -3 | 9 |
| 49 | 101 | 4 | 16 | -4 | 16 |
| | 105 | | | 0 | 0 |
| 530 | 1155 | | 70 | | 40 |

$$\bar{x}_1 = \frac{\sum x_1}{n_1} = 530 / 10 = 53$$

$$s_1^2 = \frac{\sum (x_1 - \bar{x}_1)^2}{n_1 - 1} = \frac{70}{10 - 1} = 7.77$$

$$\bar{x}_2 = \frac{\sum x_2}{n_2} = 1155 / 11 = 105$$

$$s_2^2 = \frac{\sum (x_2 - \bar{x}_2)^2}{n_2 - 1} = \frac{40}{11 - 1} = 4$$

$$F = \frac{S_1^2}{S_2^2} = \frac{7.77}{4} = 1.95$$

***The F – distribution table value at 5% L.S with (n1-1, n2-1)=(10-1, 11-1)=(9,10)d.f. is 3.03.***

***Inference:***

***Since the calculated value is less than the F – distribution table value at 5% L.S. so we accept the Ho. i.e population variance are equal.***

## *Chi – Square Test:*

- ***Chi – square test is applied in statistics to test the goodness of fit to verify the distribution of observed data with assumed theoretical distribution.***
- ***It is a measure to study the divergence of actual and expected frequency .***
- ***Thus the chi-square test describes the discrepancy between theory and observation.***

## *Assumptions:*

1. ***All the observation must be independent.***
2. ***All the events must be mutually exclusive.***
3. ***There must be large observations.***
4. ***For comparison purposes, the data must be in original units.***

## *Chi – Square Test for Goodness of fit:*

- ***Through the test we can find out the deviation between the observed value and expected value.***
- ***Here we are not concern with the parameters but concern with the form of distribution.***
- ***Karl Pearson has developed a method to test the difference between the theoretical value and the observed value.***
- ***The test is done by comparing the computed value with the table value of the chi-square distribution for the described d.f.***
- ***A Greek letter $\chi^2$ is used to describe the magnitude of the difference between the fact and theory.***

***The chi – square may be defined as***

$$\chi^2 = \sum \frac{(O - E)^2}{E} \sim \chi^2 \text{ distribution with (n-1) d.f.}$$

***Where, O = Observed frequencies***

***E = Expected frequencies***

***Problem 11:***

***Four coins were tossed 160 times and the following results were obtained***

| | | | | | |
|---|---|---|---|---|---|
| ***No. of heads:*** | ***0*** | ***1*** | ***2*** | ***3*** | ***4*** |
| ***Observed frequency:*** | ***17*** | ***52*** | ***54*** | ***31*** | ***6*** |

***Under the assumption that coins the coins are balanced. Find the expected frequencies of getting (0, 1, 2, 3 or 4) heads and test the goodness of fit.***

***Solution:***

***Let us take the null hypothesis the coins are unbiased. So p=1/2, q=1/2 and n=4***

$$p(X = x) = P(x) = {}^{n}C_{x}\, p^{x} q^{n-x}$$

$$P(x) = {}^{4}C_{x}\, (1/2)^{x} (1/2)^{4-x}$$

| No. of heads | $P(x)={}^4C_x(1/2)^x(1/2)^{4-x}$ | $N.P(x)$ | EX .F | OB .F | $(O-E)^2$ | $\chi^2=\sum\frac{(O-E)^2}{E}$ |
|---|---|---|---|---|---|---|
| 0 | $P(0)={}^4C_0(1/2)^0(1/2)^{4-0} = 0.06$ | 160*0.0625= 9.6 | 10 | 17 | 49 | 4.9 |
| 1 | $P(1)={}^4C_1(1/2)^1(1/2)^{4-1} = 0.24$ | 160*0.24=39. 4 | 40 | 52 | 144 | 3.6 |
| 2 | $P(2)={}^4C_2(1/2)^2(1/2)^{4-2} = 0.36$ | 160*0.36=59. 6 | 60 | 54 | 36 | 0.6 |
| 3 | $P(3)={}^4C_3(1/2)^3(1/2)^{4-3} = 0.24$ | 160*0.36=39. 4 | 40 | 31 | 81 | 2.025 |
| 4 | $P(4)={}^4C_4(1/2)^4(1/2)^{4-4} = 0.0625$ | 160*0.0625= 9.6 | 10 | 6 | 16 | 1.6 |
| | | | | | | 12.725 |

***The table value for chi – square distribution at 5% L.S with (r-1=5-1=4) d. f. is 9.48.***

***Inference:***

***since the calculated chi – square value is greater than the chi – square table value, so we reject the Ho. Hence the B.D does not provide the good fit to the given data.***

## *Chi – Square Test for independent of two attributes:*

*To test the independent of the attributes A and B, we take the null hypothesis, the attributes are independent, the test statistic to be used is*

$$\chi^2 = \sum \frac{(O-E)^2}{E} \sim \chi^2 \textit{distribution with (m-1)(n-1) d.f.}$$

*Where, O = Observed frequencies*

*E = Expected frequencies*

*E= Row total * Column total / Net total*

*Reject Ho, If* $\chi^2 < \chi_\alpha^2$

*Where* $\chi_\alpha^2$ *is the table value of chi – square distribution at alpha % L.S.*

## *Problem 12:*

*The response of boys and girls to a particular questions are given below*

| | Yes | No |
|---|---|---|
| Boys | 62 | 34 |
| Girls | 56 | 26 |

## *Solution:*

***1. Null Hypothesis:***

***The boys and girls differ significantly in their response.***

| | Yes | No | Total |
|---|---|---|---|
| Boys | 62 | 34 | 96 |
| Girls | 56 | 26 | 82 |
| Total | 118 | 60 | 178 |

| O | E = RT*CT / NT | $(O-E)^2$ | $\chi^2 = \sum \frac{(O-E)^2}{E}$ |
|---|---|---|---|
| 62 | 63.64 | 2.6896 | 0.0422 |
| 34 | 32.35 | 2.7225 | 0.0841 |
| 56 | 54.35 | 2.7225 | 0.0500 |
| 26 | 27.64 | 2.6896 | 0.0973 |
| | | | 0.2736 |

***The table value for chi – square distribution at 5% L.S with (r-1)(c-1)=1*1=1) d. f. is 3.841.***

***Inference:***

***since the calculated chi – square value is less than the chi – square table value, so we accept the Ho. Hence, we conclude that boys and girls are differ. significantly in their response.***

# THANK YOU