



TUTORIAL-III

Statistical Learning

(Part-B)



OBJECTIVES

1. Consider a hypothesis H_0 where $\phi_0 = 5$ against H_1 where $\phi_1 > 5$. The test is?
 - a) Right tailed ✓
 - b) Left tailed
 - c) Two tailed
 - d) Cross tailed

OBJECTIVES

2. Type 1 error occurs when?

- a) We reject H_0 if it is True ✓
- b) We reject H_0 if it is False
- c) We accept H_0 if it is True
- d) We accept H_0 if it is False

OBJECTIVES

3. The probability of Type 1 error is referred as?

a) $1-\alpha$

b) β

c) α ✓

d) $1-\beta$

OBJECTIVES

4. What is the assumption made for performing the hypothesis test with T-distribution?
 - a) the distribution is non-symmetric
 - b) the distribution has more than one modal class
 - c) the distribution has a constant variance
 - d) the distribution follows a normal distribution ✓

OBJECTIVES

5. If a hypothesis is rejected at 0.6 Level of Significance then _____

- a) it will be rejected at any level
- b) it must be rejected at 0.5 level
- c) it may be rejected at 0.5 level ✓
- d) it cannot be rejected at 0.5 level

OBJECTIVES

6. In a two tailed test when a Null Hypothesis is rejected for a True Alternative Hypothesis then it has _____
- a) Type 1 error
 - b) Type 2 error
 - c) No error ✓
 - d) Many errors

OBJECTIVES

7. Match the items of List - II with the items of List - I and select the code of correct matching

List - I		List - II	
(a)	Chi-square Test	(i)	Testing the significance of the differences of the mean values among more than two sample groups.
(b)	ANOVA (F-test)	(ii)	Testing the goodness of fit of a distribution
(c)	Z – test	(iii)	Testing the significance of the difference of the mean values between two large sized samples
(d)	t – test	(iv)	Testing the significance of the difference of the mean values between two small sized samples when population standard deviation is not available

- a) (a) - (i), (b) - (ii), (c) - (iii), (d) - (iv)
b) (a) - (iv), (b) - (iii), (b) - (ii), (d) - (i)
c) (a) - (i), (b) - (ii), (c) - (iv), (d) - (iii)
d) (a) - (ii), (b) - (i), (c) - (iii), (d) - (iv) ✓

OBJECTIVES

8. In a hypothesis test, what does the p value signify?


- a) *smallest level of significance for rejection of Null Hypothesis* ✓
- b) *largest level of significance for rejection of Null Hypothesis*
- c) *smallest level of significance for acceptance of Null Hypothesis*
- d) *smallest level of significance for acceptance of Null Hypothesis*

OBJECTIVES

9. Arrange the following steps in the process of hypothesis testing in proper sequence:

- A. Select the level of significance
- B. Setup null and alternative hypothesis
- C. Establish the decision rule
- D. Performance computations
- E. Select test statistics
- F. Draw conclusion

Choose the correct answer from the options given below

- a) A, B, C, D, E, F
- b) A, B, E, D, C, F
- c) B, A, C, D, E, F
- d) B, A, E, C, D, F 

OBJECTIVES

10. The choice of one-tailed test and two-tailed test depends upon

- a) Null hypothesis
- b) Alternative hypothesis ✓
- c) Composite hypothesis
- d) None of these

NUMERICALS

1. A drug company is testing a drug intended to increase heart rate. A sample of 100 yielded a mean increase of 1.4 beats per minute, with a standard deviation known to be 3.6. Since the company wants to avoid marketing an ineffective drug, it proposes a 0.001 significance level. Should it market the drug?

ANS:

NUMERICALS

2. The Environmental Protection Department requires a mill to aerate its effluent so that the mean dissolved oxygen (DO) level is above 6.0mg/L. To maintain the compliance, the Department collects air samples at 12 randomly selected dates. The data collected is given below.

5.85	6.28	6.50	6.21
5.94	6.12	6.65	6.14
6.34	6.19	6.29	6.40

The Department requires a strong evidence that the mean DO is high. It is required to test it with significance level, $\alpha = 1\%$. Answer the following questions. Assume that the population data follows normal distribution.

- What should be the null and alternate hypothesis?
- Which test statistics: z-, t- or χ^2 -test is applicable in this case?
- According to your test statistics, calculate the critical value.
- From the sample data, calculate the test value.
- Decide if the null hypothesis is to be accepted or rejected.

ANS:-

(a) The null hypothesis, H_0 and alternate hypothesis, H_1 are given below.

$$H_0: \mu = 6.0; H_1: \mu > 6.0$$

(b) For the given data, population mean is to be infer and population variance is not known. Hence, it comes under t-test.

(c) With degree of freedom = 11 and $\alpha = 0.01$, from the t-distribution table, the critical t-value, $t_c = 2.7181$ for one tailed test.

(d) From the sample data, $\bar{x} = 6.2425$, $S^2 = 0.04957$ and $n = 12$. Hence $t = \frac{\bar{x} - \mu}{s/\sqrt{n}} = 3.773$

(e) Since $t > t_c$, the null hypothesis is rejected. This means that mill does not violate the Department's compliance.

NUMERICALS

3. It is claimed that an automobile is driven on the average more than 20,000 kilometres per year. To test the claim, a random sample of 25 automobile owners is asked to keep a record of the kilometres they travel. The random sample showed an average of 23500 kilometres and a standard deviation of 3900 kilometres. It is planned to test the above with parametric based hypothesis testing. Assume 5% confidence level.
- a) What should be the null and alternate hypothesis?
 - b) Which test statistics: z-, t- or χ^2 -test is applicable in this case?
 - c) According to your test statistics, calculate the critical value.
 - d) From the sample data, calculate the test value.
 - e) Decide if the null hypothesis is to be accepted or rejected.

Ans:-

- a) Null hypothesis $H_0 : \mu = 20,000$ Alternate hypothesis $H_1: \mu > 20,000$
- b) This hypothesis comes under the case of t-test.
- c) The critical region from the statistical table is: $t = 1.711$
- d) The test value from the sample statistics is given below:
$$t = \frac{X - \mu}{s/\sqrt{n}} = \frac{23500 - 20000}{3900/\sqrt{25}} = 4.48$$
- b) Decision from the hypothesis testing is: H_0 is rejected that means an automobile is driven on the average more than 20,000 kilometres per year.

NUMERICALS

4. A survey was conducted among 500 students who are studying either in “government funded collages” (GVT) or “privately funded colleges” (PVT). The objective of the survey to see the choice of “classroom based learning” (C) over the “Internet based learning” (I).

C oll eg es	Learning		
	C	I	
G V T	75	125	200
P V T	60	240	300
	135	365	500

It is proposed to apply the χ^2 -test to verify if there is exist any association between “colleges” and “learning”.

- Decide the null and alternate hypotheses in this case. Justify your answer.
- Calculate the χ^2 –value from the sample data.
- Test the hypothesis with 5% confidence level.

ANS:-

- The hypothesis of the χ^2 -test is given below.

H_0 : There is no association between the attributes College and Learning

H_1 : There is an association between the attributes College and Learning

$$b) \quad \chi^2 = \frac{(75-54)^2}{54} + \frac{(125-146)^2}{146} + \frac{(60-81)^2}{81} + \frac{(240-219)^2}{219}$$

$$= 8.16 + 3.02 + 5.44 + 2.01 = 18.63$$

- Critical value is = 3.841

As $|\chi^2| > 3.841$

Reject the null hypothesis that means class room based learning is not equal to internet based learning.

C o l l e g e s	Learning		
	C	I	
G V T	75 (54)	125 (146)	200
P V T	60 (81)	240 (219)	300
	135	365	500

NUMERICALS

5. The accompanying data on breakdown voltage of electrically stressed circuits was read from a normal probability plot that appeared in the article “Damage of Flexible Printed Wiring Boards Associated with Lightning-Induced Voltage Surges” (IEEE Transactions on Components, Hybrids, and Manuf. Tech., 1985: 214–220). The straightness of the plot gave strong support to the assumption that breakdown voltage is approximately normally distributed.

1470	1510	1690	1740	1900
2000	2030	2100	2190	2200
2290	2380	2390	2480	2500
2580	2700			

- a) $P(\bar{X} \geq 2200)$
b) $P(\bar{X} \leq 2000)$
c) $P(1900 \leq \bar{X} \leq 2100)$
d) Confidence Interval of sample mean \bar{X} at 95%

ANS:-