Unit I – Testing of Hypothesis

* Sampling distributions
* Statistical hypothesis
* Tests for single mean and Difference of means (large and small samples)
* Tests for single variance and equality of variances
* Chi square test for goodness of fit
* Independence of attributes

19MAT202 – Statistics and Numerical methods

Unit – I

Testing of Hypothesis

Basic Definitions:

Population:

A population is used to refer any collection of individual it may be finite or infinite.

Sample:

A sample is a small portion selected (or drawn) from the population and the process of drawing a sample from a population is called sampling.

Sample size:

The no. of individual in a selected sample is called the sample size.

Parameter and Statistics:

Any statistical method computed from population data is known as parameter and Any statistical method computed from sample data is known as statistics.

Notations:

measure	Population	Sample
size	N	n
mean	μ	$\bar{x}$
Standard deviation	σ	s
Proportion	P	p'
Variance	σ^2	s^2

Sampling Distribution :

The various value of Statistics so obtained may be arrange as a frequency distribution which is known as sampling distributions.

Standard Error :

The standard deviation of sampling distribution of a statistic is known as its standard error.

Null hypothesis :

Null hypothesis is the hypothesis which is tested for possible rejection under the assumption that it is true and is denoted by H_0.

Alternative Hypothesis :

A hypothesis that is complementary to null hypothesis is called Alternative hypothesis and is denoted by H_1.

A procedure for deciding whether to accept or reject the null hypothesis is called the test of hypothesis.

Level of significance :

It is the prob. level below which the null hypothesis is rejected. Generally 5% and 1% level of significance are used.

Critical region :

The critical region of a test of statistical hypothesis is that region which leads to the rejection of null hypothesis, H_0. Those region which lead to the acceptance of H_0 is called critical region.

Error in Sampling:

Errors are Type I, Type II errors.

Type I error: Reject H_0 when it is true

Type II error: Accept H_0 when it is false

$P(\text{Type I error}) = \alpha$ & $P(\text{Type II error}) = \beta$

One tail & Two tail test:

If μ_0 is population parameter and μ is the sample statistics, then the null hypothesis is given by

$H_0 : \mu = \mu_0$

Alternative Hypothesis is given by,

$H_1 : \mu \neq \mu_0$ (Two tailed)

$H_1 : \mu > \mu_0$ (Right tailed - One tail)

$H_1 : \mu < \mu_0$ (Left tailed - One tail)

Procedure for Testing a Hypothesis:

1. Formulate H_0 and H_1.
2. Choose the level of significance α
3. Compute the test statistic, using the data available.
4. Pick out the critical value from the tabulation
5. Conclusion: Compare the computed value of the test statistic with the critical value at the given level of significance.

Critical Values (or) Significant values:

The sample values of the statistics beyond which the null hypo. will be rejected are called C.Vs.

Nature of Test	LOS 1%	5%	10%
Two tailed test	2.58	1.96	1.645
One tailed test	2.33	1.645	1.28 (Right)
	-2.33	-1.645	-1.28 (Left)

Scanned with CamScanner

Sampling

- Large Samples ($n \geq 30$)
  - Test for single mean
  - Test for Difference of means
  - Test for Single proportion
  - Test for Difference of Proportions
- Small Samples ($n < 30$)
  - t Test
    - Single mean
    - Difference of mean
  - F Test
  - χ^2 Test

Test of Significance of large Sample :

Test for single mean :

Null hypo. : $H_0 : \mu = \mu_0$

Test Statistics : $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$ (or) $z = \dfrac{\bar{x} - \mu}{s/\sqrt{n}}$

---

1. A sample of 900 members is found to have a mean of 3.4 cm & S.D. 2.61 cm. Is the sample from a large population of mean 3.25 cm & S.D. 2.61 cms. If the population is normal & its mean is unknown, find 95 % confident limits

Soln.

Given $n = 900$, $\bar{x} = 3.4$

$\mu = 3.25$, $\sigma = 2.61$

Step 1:

$H_0 : \mu = 3.25$

$H_1 : \mu \neq 3.25$ (Two tailed test)

Step 2:

Level of Significance $\alpha = 5\% = 0.05$

Step 3:

Test Statistic : $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{3.4 - 3.25}{2.61/\sqrt{900}}$$

$$= 1.724$$

Step 4: critical value at 5% is $z_\alpha = 1.96$

Step 5: conclusion: since $|z| = 1.724 < 1.96 = |z_\alpha|$

$\therefore$ H_0 is accepted at 5% LOS.

$\therefore$ The sample is taken from population whose mean 3.25 cm.

$$\mu = \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$= 3.4 \pm 1.96 \times \frac{2.61}{\sqrt{900}}$$

$$= 3.4 \pm 0.17$$

$$= 3.4 + 0.17,\ 3.4 - 0.17$$

$$= 3.23,\ 3.57$$

i.e., $3.23 < \mu < 3.57$

---

2. The mean height of college students in a city are normally distributed with SD 6 cms. A sample of 100 students has mean height of 158 cms. Test the hypothesis that the mean height of college students in the city 160 cms. Also obtain 99% confidence limit for the true mean.

Soln. :

Given $n=100$, $\bar{x}=158$, $\mu=160$ and $\sigma=6$

Step 1 :

$H_0 : \mu = 160$

$H_1 : \mu \neq 160$ (Two tailed test)

Step 2 :

Level of significance $\alpha = 1\% = 0.01$

Step 3 :

Test Statistic : $$z = \frac{\bar{x}-\mu}{\sigma/\sqrt{n}}$$

$$= \frac{158-160}{6/\sqrt{100}}$$

$$z = 3.33$$

Step 4 :

Critical value at 1% is $z_\alpha = 2.58$

Step 5 :

Conclusion :

$$|z| = 3.33 > 2.58 = |z_\alpha|$$

∴ H_0 is rejected at 1% of LOS.

∴ The mean height of the college students in the city is 160 cms, which is not true.

Confidence limits :

$$\mu = \bar{x} \pm z_\alpha \frac{\sigma}{\sqrt{n}}$$

$$= 156.452, 159.548$$

ie, $156.452 \nless \mu = 160 \nless 159.548$

Here 160 does not lies in the interval

3. A random sample of 200 employees at a large corporation showed their average to be 42.8 yrs with a SD of 6.89 yrs. Test the hypo $H_0: \mu = 40$ and $H_1: \mu > 40$ at $\alpha = 0.01$ level of significance.

Soln.

Given $n = 200$, $\bar{x} = 42.8$, $\mu = 40$, $\sigma = 6.89$

Step 1:

$H_0: \mu = 40$

$H_1: \mu > 40$ (Right – one tail test)

Step 2:

LOS $\alpha = 1\% = 0.01$

Step 3:

Test statistic: $z = \dfrac{\bar{x} - \mu}{\sigma/\sqrt{n}}$

$$= \frac{42.8 - 40}{6.89/\sqrt{200}}$$

$$= 5.747$$

Step 4:

Critical value at 1% is $z_\alpha = 2.33$ (One tail - Right)

Step 5:

Conclusion:

$|z| = 5.747 > 2.33 = |z_\alpha|$

$\therefore$ H_0 is rejected at 1% of LOS.

$\therefore$ $H_1: \mu > 40$ is accepted.

Hw The mean height of college students in a city are normally distributed with SD 6cms. A sample of 100 students has mean height 158

Test for Difference of Two means:

Null hypothesis: $H_0 : \mu_1 = \mu_2$

Test statistic. $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}}, \quad \sigma_1 = \sigma_2 = \sigma$$

$$= \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$(\text{or})\ z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}}}$$

---

1. The means of two simple large samples of 1000 and 2000 members are 67.5 inches and 68 inches resp. Can the samples be regarded as drawn from the same population of SD of 2.5 inches? Test at 5% level of significance.

Soln.

Given $n_1 = 1000$, $\bar{x}_1 = 67.5$

$n_2 = 2000$, $\bar{x}_2 = 68$, $\sigma = 2.5$

Step 1:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 \neq \mu_2$ (Two tailed test)

Step 2:

Level of significance $\alpha = 5\% = 0.05$

Step 3:

Test statistics $$z = \frac{\bar{x}_1 - \bar{x}_2}{\sigma\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

$$= \frac{67.8-68}{2.5\sqrt{\frac{1}{1000}+\frac{1}{2000}}}$$

$$= -5.164$$

Step 4:

Critical value at 5% is $Z_\alpha = 1.96$

Step 5:

Conclusion:

$$|Z| = 5.164 > 1.96 = |Z_\alpha|$$

$\therefore$ H_0 is rejected at 5% LOS.

$\therefore$ The Samples cannot be regarded as drawn from the Same population of S.D 2.5 inches.

---

2. A simple sample of height of 6400 English men has a mean of 170cm & SD of 6.4 cm, while a simple sample of heights of 1600 Americans has a mean of 172 cm and a SD of 6.3 cm. Do the data indicate that Americans are the average taller than the English men?

Soln.

Given $n_1 = 6400$, $\bar{x}_1 = 170$, $s_1 = 6.4$

$n_2 = 1600$, $\bar{x}_2 = 172$, $s_2 = 6.3$

Step 1:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$ (One tail- Right)

$\mu_1 \rightarrow$ Englishmen

$\mu_2 \rightarrow$ Americans

Step 2:

Level of significance $\alpha = 5\% = 0.05$

Step 3:

$$z = \frac{\bar{x}_1 - \bar{x}_2}{\sqrt{\frac{s_1^2}{n_1}+\frac{s_2^2}{n_2}}} = \frac{170-172}{\sqrt{\frac{(6.4)^2}{6400}+\frac{(6.3)^2}{1600}}} = -11.32$$

Step 4:

Critical value at 5% is $z_\alpha = 1.645$

Step 5:

Conclusion :

$$|z| = 11.32 > 1.645 = |z_\alpha|$$

$\therefore$ H_0 is rejected

Hence H_1 is accepted.

$\therefore$ Americans are taller than the Englishmen.

3]. A simple sample of height of 6400 Sailors has a mean of 67.85 inches and S.D. of 2.56 inches while a simple sample of heights of 1600 soldiers has a mean of 68.55 inches and S.D. of 2.52 inches. Do the data indicate that soldiers are on the average taller than sailors? Use 5% LOS.

Soln.

Given Sailors: $n_1 = 6400$, $\bar{x}_1 = 67.85$, $s_1 = 2.5$

Soldiers: $n_2 = 1600$, $\bar{x}_2 = 68.55$, $s_2 = 2.$

Step 1:

$H_0 : \mu_1 = \mu_2$

$H_1 : \mu_1 < \mu_2$ (One tailed test – Left)

Step 2:

LOS at 5%

Step 3:

Test Statistic : $z = \dfrac{\bar{x}_1 - \bar{x}_2}{\sqrt{\dfrac{s_1^2}{n_1} + \dfrac{s_2^2}{n_2}}}$

$$z = \frac{67.85 - 68.55}{\sqrt{\dfrac{(2.56)^2}{6400} + \dfrac{(2.52)^2}{1600}}}$$

$z = -9.91$

Step 4 :

critical value at 5%, $z_\alpha = 1.645$

Step 5 :

Conclusion :

$|z| = 9.91 > 1.645 = |z_\alpha|$

$\therefore$ H_0 is rejected at 5% of LOS.

Test for Single Proportion :

If x is the no. of person or items possessing a certain attribute in a sample of n items or person, the sample proportion $p' = \frac{x}{n}$

Null Hypothesis : $H_0 : P = P_0$ $P \to$ population proportion

$H_1 : P \neq P_0$ $p' \to$ Sample proportion

Test statistics :

$$z = \frac{p' - P}{\sqrt{\frac{PQ}{n}}} \quad \text{where } Q = 1 - P$$

Q. A coin is tossed 400 times and its turn up head 216 times. Discussed whether the coin may regarded as unbiased one.

Soln.

Given $n = 400$, $p' = \frac{x}{n} = \frac{216}{400} = 0.5156$

and $P = \frac{1}{2} \Rightarrow Q = 1 - P = \frac{1}{2}$

Step 1:

$H_0 : P = \frac{1}{2}$

$H_1 : P \neq \frac{1}{2}$

Step 2 :

LOS $\alpha = 5\%$

Step 3 :

$$Z = \frac{p' - p}{\sqrt{\frac{pq}{n}}}$$

$$= \frac{0.5156 - 0.5}{\sqrt{\frac{(0.5)(0.5)}{400}}}$$

$$= 1.6$$

Step 4 :

Critical value at 5%, $Z_\alpha = 1.96$

Step 5 :

Conclusion :

$$|Z| = 1.6 < 1.96 = |Z_\alpha|$$

$\therefore H_0$ is accepted.

Hence the coin is unbiased one.

2]. 40 people were attacked by a disease & only 36 survived. Will you reject the hypothesis that the survival rate, if attacked by this disease, is 85% in favour of the hypothesis that it is more at 5% los

Soln.

Given $n = 40$, $P = 85\% = 0.85$, $p' = \frac{36}{40} = 0.9$

$q = 1 - P = 1 - 0.85$
$= 0.15$

Step 1 :

$H_0 : P = 0.85$ (People survived after attack)

$H_1 : P > 0.85$ (one tail)

Step 2 :

$\alpha = 5\%$

Step 3:

$$z = \frac{p' - p}{\sqrt{\frac{pq}{n}}}$$

$$= \frac{0.9 - 0.85}{\sqrt{\frac{0.85(0.15)}{40}}} = \frac{0.05}{0.056} = 0.886$$

Step 4:

critical value at 5% is 1.645

Step 5:

conclusion:

$|z| = 0.886 < 1.645 = |z_\alpha|$

$\therefore$ H_0 is accepted.

Hence 85% of people were affected by cholera.

3] In a sample of 1000 people in Karnataka 540 are rice eaters. Can we assume that both rice & wheat are equally popular in this state at 1% LOS?

Ans: $z = 2.532 \Rightarrow H_0$ is accepted

conclusion: Both rice & wheat are equally popular

Test for Difference of proportions:

$$z = \frac{p_1' - p_2'}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} \quad \text{where } p_1' = \frac{x_1}{n}$$

$$p_2' = \frac{x_2}{n}$$

and $$p = \frac{p_1' n_1 + p_2' n_2}{n_1 + n_2}$$

$$q = 1 - p$$

Scanned with CamScanner

J. In a large city A, 20 % of a random sample of 900 school boys had def eye-sight. In another large city B, 18.5 % of a random sample of 1600 school boys had the same defect. Is the difference b/w the 2 proportions significant?

Soln.

Given $n_1 = 900$, $p_1' = 0.20$
$n_2 = 1600$, $p_2' = 0.185$

Step 1:

$H_0 : p_1 = p_2$

$H_1 : p_1 \neq p_2$ (Two tailed test)

Step 2:

LOS $\alpha = 5\%$

Step 3:

$$z = \frac{p_1' - p_2'}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$$

$$= 0.918$$

$$p = \frac{n_1 p_1' + n_2 p_2'}{n_1 + n_2}$$

$$= \frac{900(0.20) + 1600(0.185)}{900 + 1600}$$

$p = 0.1904$

$q = 0.8096$

Step 4:

Critical value at 5 % at $Z_\alpha = 1.96$

Step 5:

Conclusion:

$$|Z| = 0.918 < 1.96 = |Z_\alpha|$$

∴ H_0 is accepted.

∴ The difference b/w the 2 proportions are not significant.

2]. Before an increase in excise duty on tea, 800 persons out of sample of 1000 persons were found to be tea drinkers. After an increase in duty, 800 people were tea drinkers in a sample of 1200 people. Using standard error of proportion state whether there is a significant decrease in the consumption of tea after the increase in excise duty.

Soln.

Given $n_1 = 1000, \quad p_1' = \frac{800}{1000} = 0.8$

$n_2 = 1200, \quad p_2' = \frac{800}{1200} = 0.667$

Now, $p = \frac{n_1 p_1' + n_2 p_2'}{n_1 + n_2}$

$= \frac{1000(0.8) + 1200(0.667)}{1000 + 1200}$

$= \frac{1600.4}{2200}$.

$p = 0.727$

$q = 1 - p = 0.273$

Step 1:

$H_0 : P_1 = P_2$

$H_1 : P_1 > P_2$ (Right tailed test)

Step 2:

LOS, $\alpha = 5\%$

Step 3:

$$z = \frac{p_1' - p_2'}{\sqrt{pq\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}} = \frac{0.8 - 0.667}{\sqrt{(0.727)(0.273)\left(\frac{1}{1000} + \frac{1}{1200}\right)}}$$

$$= \frac{0.8 - 0.667}{0.019} = 7$$

Step 4:
Critical value at 5% is, $Z_\alpha = 1.645$

Step 5:
Conclusion :

$$|z| = 7 > 1.645 = |Z_\alpha|$$

$\therefore H_0$ is rejected.

ie, There is a difference in the consumption of tea before and after the increase in excise duty.

3]. Random Samples of 400 men and 600 women were asked whether they would like to have a flyover near their residence. 200 men & 325 women were in favour of the proposal. Test the hypothesis that proportions of men & women in favour of the proposal are same, at 5% level.

Soln.

Given $n_1 = 400$, $p_1' = \frac{200}{400} = 0.5$

$n_2 = 600$, $p_2' = \frac{325}{600} = 0.541$

Step 1:
$H_0 : P_1 = P_2$
$H_1 : P_1 \neq P_2$ [Two tailed test]

Step 2:
LOS, $\alpha = 5\%$

Step 3:

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

$$= \frac{0.5 - 0.541}{\sqrt{0.525(0.475)\left(\frac{1}{400} + \frac{1}{600}\right)}}$$

$$= 1.34$$

$$P = \frac{n_1 P_1' + n_2 P_2'}{n_1 + n_2}$$

$$= \frac{400(0.5) + 600(0.541)}{400 + 600}$$

$P = 0.525$

$q = 1 - P = 0.475$

Scanned with CamScanner

Step 4:

critical value at 5% is $Z_\alpha = 1.96$

Conclusion:

$$|Z| = 1.34 < 1.96 = |Z_\alpha|$$

$\therefore$ H_0 is accepted.

$\therefore$ proportions of the men and women in favour of the proposal are same.