EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

Faculty: Dr. Rajasekharan K. E.

The flipped class room technique is a modern pedagogical tool to enhance the metacognitive abilities of the students. To implement this technique, students are grouped on the basis of learning skills initially. (on the basis of continuous assessment)

Here the conventional class room is just 'flipped' in such a way that the future calluses are planned by the teacher and the resources for the classes (like study materials, videos, web links etc.) are shared in advance. All the students are directed to familiarize the topic and concepts for the next day's class, and discuss their thoughts and understanding of the topic in the class for 40 minutes. They may clarify their doubts with the teacher in the class time. After all the students understood the content of the topic, one or two groups are directed to present their ideas it may go in the form of a discussion among all the students.

I have practiced flipped class room technique in the second semester M A Economics class during the last academic year. (2018-19). The students were grouped as the following.

Group 1

Fathima Suhra, Shahna T, Sheena O, Farhath V K.

Group 2

Najila P K, Safla A, Shahla V P,Rinsila

Group 3

Sameeba Thasnim, Naseeba K, Neethumol, Jishna K

Group 4

Sameela M P, Rahmath I, Mubashir P, Munshida Banu.

Group 5

Shahna P T, Shirin Shahana, Rahna CP, Praseetha .

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

Test concerning the proportion of success of a population

By testing population proportion of success, we mean the testing of the significant difference between population proportion of success and the sample proportion of success.

Now let us familiarise the following notations

P : population proportion of success

p_0 : the assumed value of p (proposed by H_0)

$q_0 = 1 - p_0$

x : the number of successes in the sample

n : sample size

p': $\frac{x}{n}$; the proportion of success in the sample

Suppose we want to test the null hypothesis

$H_0 : p = p_0$ against one of the alternatives

$H_1 : p < p_0$ or

$H_1 : p > p_0$ or

$H_1 : p \neq p_0$, based on a large sample of size n whose proportion of success is p'.

The test statistic is $Z = \frac{p' - p_0}{\sqrt{\frac{p_0 q_0}{n}}}$

For a given significance level , the BCR are respectively,

$\omega : Z < -Z_\alpha$

or $\omega : Z > Z_\alpha$

or $|Z| \geq Z_{\alpha/2}$

Calculate the value of Z and if it lies in the critical region reject H_0 , or otherwise.

Exercise1 : In a survey of 70 business firms, it was found that 45 are planning to expand their capabilities next year. Does the sample information contradict the hypothesis that 70% of the firms are planning to expand next year.

(Hint : $Z = -1.04$)

Exercise 2 : In a die rolling experiment, getting 3 or 6 is identified as a success. Suppose that 9,000 times the die was rolled resulting in 3240 successes. Do you have reasons to believe that the die is an unbiased one?

Test concerning difference of two population proportions

By testing the difference of two population proportions, we are testing the equality of two population proportions. In other words, we are deciding whether the two samples have come from populations having the same proportion of success.

Let us consider the following notations

p_1: $proportion\ of\ success\ of\ the\ first\ population$

p_2: $proportion\ of\ success\ of\ the\ second\ population$

x_1 : $number\ of\ successes\ in\ the\ first\ sample$

x_2: $number\ of\ successes\ in\ the\ second\ sample$

n_1: $size\ of\ firsr\ sample$

n_2: $size\ of\ second\ sample$

$p_1{}'$: $proportion\ of\ success\ in\ the\ first\ sample = \frac{x_1}{n_1}$

$p_2{}'$: $proportion\ of\ success\ in\ \ the\ second\ sample = \frac{x_2}{n_2}$

Suppose we want to test the null hypothesis $H_0: p_1 = p_2$ against one of the alternatives

$H_1 : p_1 < p_2$

or $H_1 : p_1 > p_2$

or $H_1 : p_1 \neq p_2$, based on two independent random samples of sizes n_1 and n_2 with proportions of successes $p_1{}'$ and $p_2{}'$ respectively.

The test statistic is $Z = \frac{p_1{}' - p_2{}'}{\sqrt{p^* q^* \left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$ where $p^* = \frac{n_1 p_1{}' + n_2 p_2{}'}{n_1 + n_2}$ and $q^* = 1 - p^*$

For a given significance level α, the BCR are respectively

$\omega: Z < -Z_\alpha$

or $\omega : Z > Z_\alpha$

or $|Z| \geq Z_{\alpha/2}$

Calculate the value of Z and if it lies in the critical region, reject H_0 or otherwise.

Exercise 1: Before an increase in excise duty on tea, 800 persons out of a sample of 1000 were found to consume tea. After an increase in the duty, 800 people out of 1200used to take tea. Test whether there is significant decrease in the consumption of tea after the increase in the duty. (Hint : $Z = 6.816$)

Exercise 2: In a sample of 600 men from city A, 450were found to be smokers. Out of 900 from city B, 450 were smokers . Do the data indicate that the cities are significantly different with respect to the smoking habit?

Exercise 3 :A machine , in the long run, produces 16 defective items out of every 500 produced. After the machine is repaired, it produced 3 defective items in a batch of 100. Has the machine improved its performance?

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

Test concerning difference of means of two populations

Suppose we want to test the null hypothesis $H_0: \mu_1 - \mu_2 = 0$ against one of the alternatives

$H_1 : \mu_1 - \mu_2 < 0$

or $H_1: \mu_1 - \mu_2 > 0$

or $H_1: \mu_1 - \mu_2 \neq 0$, based on independent random samples of sizes n_1 and n_2 from two populations having the means μ_1 and μ_2 and the known variances ${\sigma_1}^2$ and ${\sigma_2}^2$.

The test statistic is

$$Z = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{{\sigma_1}^2}{n_1} + \frac{{\sigma_2}^2}{n_2}}}$$

For the significance level α, the best critical regions (BCR) are respectively $\omega : Z < -Z_\alpha$

or $\omega : Z > Z_\alpha$

or $\omega : Z \geq Z_{\alpha/2}$

Calculate the value of the statistic Z using the sample information and if it lies in the critical region reject H_0, or otherwise.

Note: When we deal with independent random samples from populations with unknown variances which may not be normal, we can still use the above test with s_1 substituted for σ_1 and s_2 substituted for σ_2 provided n_1 and n_2 are large.

Example

Suppose that 64 students from college A and 81 students from college B had mean heights 68.2" and 67.3" respectively. If the standard deviation for heights of all students is 2.43, is the difference between the two groups significant?

(Hint: Z=2.21, ...=1.96)

Exercise 1.

A random sample of 1000 workers from factory A shows that the mean wages were Rs.47 with a standard deviation of Rs.23. A random sample of 1500 workers from factory B gives a mean wage of Rs.30. Is there any significant difference between their mean level of wages?

Exercise 2.

Given in usual notations,

$$n_1 = 400, \overline{x_1} = 250, s_1 = 40$$

$$n_2 = 400, \overline{x_2} = 220 , s_2 = 55$$

Test whether the two samples have come from populations having the same mean?

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

t – tests

The test of hypothesis based on the Student's t distribution is called t-test. The main applications of t- test are,

1. To test the significance of mean of a small sample from a normal population
2. To test the significance of the difference between the means of two independent samples taken from two normal populations.
3. To test the significance of the difference between the means of two dependent samples taken from a normal population.
4. To test the significance of an observed correlation coefficient.
5. To test the significance of an observed regression coefficient.

t– test for the significance of population mean

To test the mean of a population using Student's t – test, the following assumptions are made.

i) The parent population from which the sample is drawn is normal.
ii) The sample observations are independent and random
iii) The sample size should be small (i.e. $n<30$)
iv) The population standard deviation σ is unknown

By testing the mean of a normal population, we are actually testing the significant difference between sample mean and the hypothetical value of μ proposed by the null hypothesis. In other words, we are testing whether the sample is drawn from the population with the hypothetical mean.

Here we are testing $H_0 : \mu = \mu_0$ against one of the alternatives

$H_1: \mu < \mu_0$

or $H_1: \mu > \mu_0$

or $H_1: \mu \neq \mu_0$

The test statistic is $t = \frac{\bar{x}-\mu_0}{s/\sqrt{n-1}}$ where $= \sqrt{\frac{1}{n}\sum(x_i - \bar{x})^2}$, the sample SD.

For a given significance level , the BCR are respectively,

$\omega : t < -t_\alpha$

or $\omega : t > t_\alpha$

or $\omega: |t| \geq t_{\alpha/2}$

Now calculate the value of t and compare it with the table value of t for $n - 1$ degrees of freedom and a given significance level α.

Exercise 1 : A sample of 10 observations gives a mean equal to 38 and SD 4. Can we conclude that the population mean is 40? (Hint : t = -1.5)

Exercise 2 : A random sample of size 16 has 53 as mean and the sum of squares of the deviations taken from the mean is 150. Can the sample be regarded as arisen from the population with mean 56.

Exercise 3 : A sample of size 8 from a normal population is 6,8,11,5,9,11,10,12. Can such a sample be regarded as drawn from a population with mean 7 at 2% level of significance?

Exercise 4: A personality test was conducted on a random sample of 10 students from a University and the scores obtained were 35,60,55,50.5,44,,41.5,49,53.5,50. Test whether the average “personality test score ”for the University students is 50 at 5% level?

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

Test concerning difference of means of two populations

By testing the equality of two population means, we are deciding whether the two samples are drawn from populations having the same mean.

Assumptions :

i) The two populations from which the samples are drawn follow normal distributions.
ii) The sample observations are independent and random
iii) The sample sizes are small ($n_1 < 30, n_2 < 30$)
iv) The two population variances $\sigma_1{}^2$ and $\sigma_2{}^2$ are equal but unknown

Suppose we want to test the null hypothesis $H_0: \mu_1 - \mu_2 = 0$ against one of the alternatives

$H_1 : \mu_1 - \mu_2 < 0$

or $H_1: \mu_1 - \mu_2 > 0$

or $H_1: \mu_1 - \mu_2 \neq 0$, based on independent random samples of sizes n_1 and n_2 from two populations.

The test statistic is

$t = \frac{\overline{x_1} - \overline{x_2}}{\sqrt{\frac{n_1 s_1{}^2 + n_2 s_2{}^2}{n_1 + n_2 - 2}\left(\frac{1}{n_1} + \frac{1}{n_2}\right)}}$, where s_1 and s_2 are the sample standard deviations of first and second samples respectively.

For the significance level α, the best critical regions (BCR) are respectively $\omega : t < -t_\alpha$

or $\omega : t > t_\alpha$

or $\omega : |t| \geq t_{\alpha/2}$

Calculate the value of the statistic t using the sample information and if it lies in the critical region reject H_0, or otherwise.

Example

The mean life of a sample of 10 electric bulbs was observed to be 1309 hours with a SD of 420 hours. A second sample of 16 bulbs of a different batch showed a mean life of 1205 hours with a SD of 390 hours. Test whether there is significant difference between the means at 5% level. (Hint: t=0.697)

Exercise 1.

A random sample of 16 men from state A had a mean height of 68 inches and sum of squares from the sample mean 132. A random sample of 25 men from state B had the corresponding values 66.5 and 165 respectively. Can the samples be regarded as drawn from populations having the same mean?

Exercise 2.

The following are samples from two independent normal populations. Test the hypothesis that they have the same mean assuming that the variances are equal at 5% level of significance.

Sample 1 : 14,18,12,9,16,24,20,21,19,17

Sample 2: 20,24,18,16,26,25,18

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

Chi square tests

The important applications of chi square test are

i) To test the variance of a normal population
ii) To test the goodness of fit
iii) To test the independence of attributes
iv) To test the homogeneity

Chi square test for population variance

This test is conducted when we want to test if the given normal population has a specified variance, say ${\sigma_0}^2$. Chi square test for variance is generally a right tailed test. Here we are testing

$H_0: \sigma^2={\sigma_0}^2 \; against \; H_1 : \sigma^2 > {\sigma_0}^2$

The test statistic is $\chi^2 = \frac{ns^2}{{\sigma_0}^2}$

For a given significance level $\alpha,$ the BCR is

$\omega : \chi^2 > {\chi_\alpha}^2$, where ${\chi_\alpha}^2$ is obtained by referring the chi square table for n-1 d.f. and a for a given significance level.

Calculate the value of the statistic and if it lies in the critical region, reject the null hypothesis, or otherwise.

Exercise 1 : A manufacturing process is expected to produce goods with a specified weight with variance less than 5unts. A random sample of 10 was found to have variance 6.2 units. Is there reason to suspect that the process variance has increased?

Exercise 2 : A farmer surveyed 4 plots of land and found the following yields. 1269,1271,1263,1265. He believes that his production have a SD of 2. Test at 5% level whether his data is consistent with his supposition.

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

The chi - square test for independence of attributes

Let there be two attributes A and B, 'A' divided into 'm' classes $A_1, A_2, \dots A_m$ and 'B' divided into 'n' classes $B_1, B_2, \dots B_n$. The various cell frequencies can be expressed as in the following table, having 'm' rows and 'n' columns.

B / A	B_1	B_2	.....	B_j	......	B_n	Total
A_1	f_{11}	f_{12}	.	f_{1j}	.	f_{1n}	$f_{1.}$
A_2	f_{21}	f_{22}	.	f_{2j}	.	f_{2n}	$f_{2.}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_i	f_{i1}	f_{i2}	.	f_{ij}	.	f_{in}	$f_{i.}$
. . .	. . .	. . .	. . .	. . .	. . .	. . .	. . .
A_m	f_{m1}	f_{m2}	.	f_{mj}	.	f_{mn}	$f_{m.}$
Total	$f_{.1}$	$f_{.2}$	.	$f_{.j}$	.	$f_{.n}$	$N = f_{..}$

Here

f_{ij} - Frequency of the occurrence of the joint event (A_i, B_j)

$f_{i.}$ - Frequency of occurrence of the event A_i

$f_{.j}$ – frequency of occurrence of the event B_j

Such a table is called m x n contingency table. We test the hypothesis H_0 that the attributes A and B are independent. That is there is no association between A and B.

If H_0 is true, we have,

$P(A_i, B_j) = P(A_i)\, P(B_j)$

i.e., $\frac{f_{ij}}{N} = \frac{f_{i.}}{N} \times \frac{f_{.j}}{N}$

hence, $f_{ij} = \frac{f_{i.}\, f_{.j}}{N}$

The test statistic is given by

$$\chi^2 = \sum_{i=1}^{m} \sum_{j=1}^{n} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$

The BCR is given by $\chi^2 > \chi^2{}_{\alpha}$, where $\chi^2{}_{\alpha}$ is obtained by referring the chi square table for (m-1)(n-1)degrees of freedom. Calculate the value of the statistic and if lies in the critical region, reject the null hypothesis.

Exercise 1 : From the following table on some horticultural data, test the hypothesis that the flower colour is independent of the nature of leaves.

Leaves / Flowers	Flat leaves	Curled leaves
White flowers	99	36
Red flowers	20	5

(Hint : $\chi^2 = 0.249$)

Note: For a 2x2 contingency table, where the cell frequencies are

a	b
c	d

The calculate value of chi square is given by,

$$\chi^2 = \frac{N(ad-bc)^2}{(a+b)(c+d)(a+c)(b+d)}$$, where N=a+b+c+d.

Exercise 2 : Consider the following 2x2 contingency table.

	A_1	A_2
B_1	7	1
B_2	6	8

Apply chi square test at 5% level to test whether the two attributes A and B are independent.

Yates’ correction

Yates’ correction was proposed by F.Yates in 1934.

We have learnt that no cell frequency should be less than 5 for applying chi square test. This is because, when the expected frequencies are less than 5, the chi square table values are not very reliable, especially for 1 degree of freedom.

However if chi square test is applied after considering Yates’ correction, we will get a reduced chi square value. Sometimes, the difference between the chi square value with and without Yates’ correction is so much that it leads to a totally different conclusion.

Consider the 2x2 contingency table

a	b
c	d

With Yates’ correction, the calculated chi square becomes

$$\chi^2 = \frac{N\left(|ad - bc| - \frac{N}{2}\right)^2}{(a+b)(c+d)(a+c)(b+d)}$$

Exercise 4 Consider the previous example and test the independence of attributes with Yates’ correction

Exercise 5 In an experiment on immunization, of human beings, the following results were obtained. Draw your inference on the efficiency of the vaccine at 5% level.

	Died	Survived
Vaccinated	2	10
Not vaccinated	6	4

Exercise 6 A driving school examined the result of 200 candidates who were taking their test for the first time. They found that out of 90 men, 52 passed and out of 110 women 59 passed. Do these data indicate at 1% level of significance that a relationship between gender and the ability to pass the test for the first time?

EMEA COLLEGE OF ARTS AND SCIENCE, KONDOTTI

DEPARTMENT OF STATISTICS

FLIPPING

A REPORT

MA Economics flipped class study material – Testing of Hypothesis.

(Only for private circulation)

Chi square test for goodness of fit

Consider a set possible events 1,2 3,...,n arranged in 'n' classes or cells. Let these events occur with frequencies $O_1, O_2, \ldots O_n$ called 'observed frequencies. Let them be expected to occur with frequencies $E_1, E_2, \ldots E_n$ called the 'expected frequencies. The expected frequencies are calculated on the assumption that the data obeys a certain probability distribution such as binomial, normal etc.

We test the hypothesis that the assumed probability distribution fits good for the given data against the alternative that it is not.

A measure of disagreement existing between the observed and expected frequencies can be found by using the chi square test statistic given by

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i}$$

If $\chi^2 = 0$, the observed frequencies and the expected frequencies will coincide and this shows that there is perfect agreement between theory and observation. When the value computed from the data exceeds limit, we reject H_0. Thus the chi square test for goodness of fit is basically a right tailed test.

The best critical region is $\chi^2 > {\chi_\alpha}^2$, where ${\chi_\alpha}^2$ is obtained by referring the chi square table for n-1 degrees of freedom.

Now calculate the value of the test statistic and if lies in the critical region, reject H_0 or otherwise.

The following points to be considered while conduction this test.

1. If we have to estimate the parameters like ‘p’ (for binomial distribution) or λ (for Poisson distribution), to compute the expected frequencies, then one or more degrees of freedom has to be subtracted for each parameter estimated.
2. If any of the expected cell frequency is less than 5, then we combine (pool) it with the preceding or succeeding cell frequency, so that the resulting frequency is greater than 5. In this case also, one degrees of freedom has to be adjusted.

Exercise 1 : Four coins are tossed 80 times. The distribution of number of heads is given below.

No. of heads :	0	1	2	3	4
Frequency :	4	20	32	18	6

Apply chi square test at 1% level to test whether the coins are unbiased. (Value of statistic=0.73)

Exercise 2 : The demand for refrigerators in a city are found to vary day by day. In a study . the following data were obtained. Test at 5% level of significance whether the demand depends on the day of the week.

Days :	Mon	Tue	Wed	Thur	Fri	Sat
Demand:	115	126	120	110	125	124

Exercise 3: Fit a Poisson distribution for the following data and test for goodness of fit.

X :	0	1	2	3	4	5	6
F :	275	72	30	7	5	2	1

Certified that the data attached herewith is true.

EMEA COLLEGE OF ARTS & SCIENCE
No.
Date
KONDOTTI

PRINCIPAL
E M E A COLLEGE OF ARTS
& SCIENCE, KONDOTTY
MALAPPURAM DT. 673 638