[Academic Script]

Quantitative Techniques for Management

Subject: Business Economics

Course: B.A., 3rd Semester, Undergraduate

Paper No. & Title: Paper – 304 Business Economics

Unit No. & Title: 1(one) Sampling and hypothesis Testing

Lecture No. & Title: 2: Hypothesis – Large Sample Test & Chi square Test

By
Ronak M. Patel
Department of Statistics
SOM-LALIT COLLEGE OF COMMERCE
Ahmedbabd

1. Large sample Tests:

For testing a given hypothesis a random sample is drawn from a population. If the sample size is 30 or more it is generally regarded as a large sample. Here we assume underlying distribution is normal. Even if the underlying distribution is not normal, but sample size is large, by using central limit theorem we can apply large sample tests for such population also.

For large sample tests, we follow the following steps for testing the given hypothesis:

Step 1: Identify the parameter and statistics

Step 2: write null and alternate hypotheses.

Step 3: Under the null hypothesis, calculate the test statistic using the formula

$$|Z| = \frac{|Difference|}{S.E}$$

Step 4: Obtain critical value for the given value of level of significant and compare it with the calculated value. If calculated value ≤ critical value then do not reject the null hypothesis otherwise reject the null hypothesis.

Step 5: write appropriate conclusion.

2. Test for variables:

(i) Test of significance of a mean:

This test is used for testing the mean of the population based on a large sample which is taken from a normal population.

Let us understand with the illustration.

Illustration 1:

It is claimed by the railway authority that a particular train has an average speed of 120 k.m. per hour. During last 100 trips it was found that the average speed was 116 k.m. per hour with standard deviation of 15 k.m. per hour. Is the claim justified?

Solution:

Step 1: Here $n = 100,\ \mu = 120, \bar{x} = 116, s = 15$ and $\alpha = 0.05$.

Step 2: H_o: $\mu = 120$ vs H_1: $\mu < 120$

Step 3: $|Difference| = |\bar{x} - \mu| = |116 - 120| = 4$

$$SE = \frac{s}{\sqrt{n}} = \frac{15}{\sqrt{100}} = 1.5$$

Under H_o the test statistic is

$$|Z| = \frac{|Difference|}{S.E} = \frac{4}{1.5} = 2.67$$

Step 4: Since the given test is one tailed test hence the critical value for 5% level of significance is obtained from the tables of normal distribution and is 1.645

Since $2.67 > 1.645$, the null hypothesis is rejected.

Step 5: At 5% level of significance we reject the null hypothesis hence we conclude that average speed of that train is less than 120 kms per hour.

(ii) Test of significance difference of two means:

This test is used for testing the hypothesis that two large samples are drawn from the normal population with the same mean or there is no significant difference between two means.

Let us consider it with the help of illustration

Illustration 2: A factory produces electric motor in large scale. During the first shift a sample of 150 electric motors shows the average life of 1400 hours with a standard deviation of 120 hours. Another sample during a second shift of 200 motors shows the average life 1200 hours with the standard deviation of 80 hours. Is these difference in two average lives is significant?

Solution:

Step 1: Here $n_1 = 150, n_2 = 200, \overline{x_1} = 1400, \overline{x_2} = 1200, S_1 = 120, S_2 = 80$ and $\alpha = 0.05$.

Step 2: $H_0 : \mu_1 = \mu_2$ vs $H_1 : \mu_1 \neq \mu_2$

Step 3: $|Difference| = |\overline{x_1} - \overline{x_2}| = 200$

$$SE = \sqrt{\frac{S_1^2}{n_1} + \frac{S_2^2}{n_2}} = 11.31$$

Under H_o the test statistic is

$$|Z| = \frac{|Difference|}{S.E} = 17.68$$

Step 4: Since the given test is two tailed test hence the critical value for 5% level of significance is obtained from the tables of normal distribution and is 1.96

Since $17.68 > 1.96$ hence the null hypothesis is rejected.

Step 5: At 5% level of significance we reject the null hypothesis hence we conclude that the average life of electric motors in two shifts differ significantly.

(iii) Test of significance difference of two variances or standard deviations:

This test is used for testing the hypothesis that two large samples are drawn from normal populations with same variance (standard deviations) or there is no significant difference between two variances.

Let us consider the illustration

Illustration 3: Information regarding the marks obtained by boys and girls in an examination are as under:

Gender	Sample size	Mean marks	Standard deviation of marks
Boy	121	83	10
Girl	81	81	12

Is these difference in the standard deviations of the two groups significant?

Solution:

Step 1: Here $n_1 = 121, n_2 = 81, \overline{x_1} = 83, \overline{x_2} = 81, S_1 = 10, S_2 = 12$ and $\alpha = 0.05$.

Step 2: $H_0 : \sigma_1 = \sigma_2$ vs $H_1 : \sigma_1 \neq \sigma_2$

Step 3: $|Difference| = |S_1 - S_2| = 2$

$$SE = \sqrt{\frac{S_1^2}{2n_1} + \frac{S_2^2}{2n_2}} = 1.14$$

Under H_o the test statistic is

$$|Z| = \frac{|Difference|}{S.E} = 1.75$$

Step 4: Since the given test is two tailed test hence the critical value for 5% level of significance is obtained from the tables of normal distribution and is 1.96

Since $1.75 < 1.96$ hence the null hypothesis is not rejected.

Step 5: At 5% level of significance we do not reject the null hypothesis hence we conclude that the variation among the marks of boys and girls is insignificant.

3. Tests for proportion:

(i) Test of significance of proportion:

This test is used for testing the hypothesis that the population proportion has specified value.

Let us consider the illustration

Illustration 4: In a large consignment of commodities, 64 out of 400 are found defective. Test the hypothesis at 1% level of significance that the proportion of the defective items is 20%.

Solution:

Step 1: Here $\mathrm{n = 400,\ x = 64,\ p = \frac{x}{n} = 0.16,\ P = 0.20,\ Q = 1 - P = 0.80}$ *and* $\alpha = 0.05$.

Step 2: $H_0 : P = 0.20$ vs $H_1 : P \neq 0.20$

Step 3: $|Difference| = |p - P| = 0.04$

$$SE = \sqrt{\frac{PQ}{n}} = 0.02$$

Under H_o the test statistic is

$$|Z| = \frac{|Difference|}{S.E} = 2$$

Step 4: Since the given test is two tailed test hence the critical value for 1% level of significance is obtained from the tables of normal distribution and is 2.575

Since $2 < 2.575$ hence the null hypothesis is not rejected.

Step 5: At 1% level of significance we do not reject null hypothesis that the percentage of defective items in the consignment is 20%.

(ii) Test of significance of two proportions:

This test is used for testing the hypothesis that there is no significant difference between the two population proportions.

Illustration 5: In a factory production is to be carried on a machine and it is known that in a batch of 500 articles 16 articles are found defective. After maintenance of that machine, 3 defective articles are found in a batch of 100 articles. Can it be concluded that the performance of machine is improved after maintenance?

Solution:

Step 1: Here $n_1 = 500, n_2 = 100, x_1 = 16, x_2 = 3, p_1 = \frac{x_1}{n_1} = 0.032, p_2 = \frac{x_2}{n_2} = 0.03$

and $\alpha = 0.05$.

Step 2: $H_0 : P_1 = P_2$ vs $H_1 : P_1 > P_2$

Step 3: $|Difference| = |p_1 - p_2| = 0.002$

$$SE = \sqrt{PQ\left(\frac{1}{n_1} + \frac{1}{n2}\right)}$$

$$Where \;\; P = \frac{n_1 p_1 + n_2 p_2}{n_1 + n_2} = \frac{19}{600}, \; Q = \frac{581}{600}$$

$$\therefore SE = 0.0192$$

Under H_o the test statistic is

$$|Z| = \frac{|Difference|}{S.E} = 0.104$$

Step 4: Since the given test is one tailed test hence the critical value for 5% level of significance is obtained from the tables of normal distribution and is 1.645

Since $0.104 < 1.645$ hence the null hypothesis is not rejected.

Step 5: At 5% level of significance we do not reject null hypothesis hence we conclude that the machine is not improved.

Confidence Interval:

As we have discussed earlier, when an estimator is used to predict a single value of parameter then it is called a point estimates. In practice an interval is obtained which may include the value of parameter with a certain degree of confidence. The interval developed by using standard error of the statistic is called confidence interval or fiducial interval.

Confidence interval for the population mean is given as

$$\bar{x} \pm (critical\ value\ for\ given\,\alpha)(SE)$$

Confidence interval for the population proportion is given as

$$p \pm (critical\ value\ for\ given\,\alpha)(SE)$$

For illustration 1 the 95% confidence interval for population mean can be obtained as

$$116 \pm (1.96)(1.5) = (113.06,\ 118.94)$$

For illustration 5 the 99% confidence interval for population proportion can be obtained as under

$$0.16 \pm (2.575)(0.02) = (0.1085, 0.2115)$$

Chi square Distribution χ^2:

A probability distribution of square of standard normal variate is called chi square distribution with one degree of freedom (df). The number of independent terms of a statistic is called degree of freedom. As the number of restrictions increases the degree of freedom decreases. In general, probability distribution of sum of squares of *n* independent standard normal variate is called chi square distribution with *n* degree of freedom. i.e. if $x_1, x_2, \ldots\ldots, x_n$ is a random sample of size *n* drawn from a normal population with mean μ and variance σ^2 then the distribution of statistic $\chi^2 = \sum_{i=1}^{n}\left(\frac{x-\mu}{\sigma}\right)^2$ is called chi square distribution with *n* df. It should be noted that the chi square distribution is a function of its degree of freedom and it is also considered as non parametric test.

Important Properties:

Followings are some important properties of chi square distribution:

i. It is a continuous distribution

ii. It mean is equal to its df and variance is 2(df)

iii. Its skewness is always positive

iv. For large value of sample size it follows normal distribution.

Important Application:

Following are some important applications of chi square distribution.

i. To test the goodness of fit

ii. To test the independency of attributes

iii. To test the significance of variance.

(i) To test goodness of fit:

This test is used to test the hypothesis that there is no significant difference between observed and expected frequencies or to test the hypothesis that the observed frequencies are distributed according to specified probability law.

Let us consider the illustrations to understand the above application

Illustration 6: The information regarding the daily demand of milk bag of a particular dairy at a retail distribution center is given below. Can it be said that the demand of milk bag does not depend on the day of week?

Day	Mon	Tue	Wed	Thu	Fri	Sat	Sun
Demand of milk bags	14	16	8	12	11	9	14

Solution: H_o: demand of milk bag does not depend on the day of week, i.e. the probability of demand of milk bag at any day is same and is $\frac{1}{7}$

Day	Demand (O_i)	$P(x)$	Expected frequency $E_i = N \times P(x)$	$\frac{(O_i - E_i)^2}{E_i}$
Mon	14	$\frac{1}{7}$	12	$\frac{4}{12}$
Tue	16	$\frac{1}{7}$	12	$\frac{16}{12}$
Wed	8	$\frac{1}{7}$	12	$\frac{16}{12}$
Thu	12	$\frac{1}{7}$	12	0

Fri	11	$\frac{1}{7}$	12	$\frac{1}{12}$
Sat	9	$\frac{1}{7}$	12	$\frac{9}{12}$
Sun	14	$\frac{1}{7}$	12	$\frac{4}{12}$
Total	**84**	**1**	**84**	$\frac{50}{12}$

The test statistic is

$$\chi^2 = \sum_{i=1}^{n} \frac{(O_i - E_i)^2}{E_i} = \frac{50}{12} = 4.17$$

At 5% level of significance level of significance and witn $n - 1 = 7 - 1 = 6$ df the critical value from chi square table is 12.59.

Since, 4.17 < 12.59 so the null hypothesis is accepted and hence we conclude that the demand of milk bag does not depend on the day of week.

Illustration 7: In classical random experiment of tossing five coins 320 times the distribution of number of heads is as under:

Number of heads	0	1	2	3	4	5	Total
Observed frequency	8	42	116	90	52	12	320

Can we say that the coins are unbiased?

Solution: H_o: The coins are unbiased. i.e. $p = \frac{1}{2}$ The probability distribution of number of heads is

$$p(x) = {}^nC_x p^x q^{n-x} : x = 0,1,2...5.$$
$$= {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$$

Number of heads	Observed frequency (O_i)	$p(x) = {}^5C_x \left(\frac{1}{2}\right)^x \left(\frac{1}{2}\right)^{5-x}$	Expected frequency $E_i = N \times P(x)$	$\frac{(O_i - E_i)^2}{E_i}$
0	8	${}^5C_0 \left(\frac{1}{2}\right)^0 \left(\frac{1}{2}\right)^{5-0} = \frac{1}{32}$	10	0.40
1	42	${}^5C_1 \left(\frac{1}{2}\right)^1 \left(\frac{1}{2}\right)^{5-1} = \frac{5}{32}$	50	1.28
2	116	${}^5C_2 \left(\frac{1}{2}\right)^2 \left(\frac{1}{2}\right)^{5-2} = \frac{10}{32}$	100	2.56
3	90	${}^5C_3 \left(\frac{1}{2}\right)^3 \left(\frac{1}{2}\right)^{5-3} = \frac{10}{32}$	100	1.00
4	52	${}^5C_4 \left(\frac{1}{2}\right)^4 \left(\frac{1}{2}\right)^{5-4} = \frac{5}{32}$	50	0.08
5	12	${}^5C_5 \left(\frac{1}{2}\right)^5 \left(\frac{1}{2}\right)^{5-5} = \frac{1}{32}$	10	0.40
Total	320	1	320	5.72

The test statistic is $\chi^2 = \sum_i \frac{(O_i - E_i)^2}{E_i} = 5.72$

At 5% level of significance and witn $n - 1 = 6 - 1 = 5$ df the critical value from chi square table is 11.07.

Since, 5.72 < 11.07 so at 5% level of significance and 5 df, the null hypothesis is accepted and hence we conclude that the coins are unbiased.

(ii) Test of independence of attributes:

Illustration 8: The result of last examination of a sample of 100 students is as under:

Gender	First class	Second class	Pass class	Total
Boys	10	28	12	50
Girls	20	22	8	50
Total	30	50	20	100

Can we say that the performance in examination depends upon gender of student?

Solution: H_o: the performance in examination does not depend upon gender of students