One-Sample Tests

SESSION 9

Structure

9.1 Introduction

Objectives

9.2 Problem Description

9.3 Z-Test for Population Mean

9.4 t-Test for Population Mean

9.5 Z-Test for Population Proportion

9.6 Chi-Square Test for Population Variance

9.1 INTRODUCTION

In Block 3 of MST-004 (Statistical Inference), you have learnt the general procedure of testing a hypothesis about the population parameter(s). In the same block, you have studied one-sample parametric tests such as Z-test, t-test and chi-square test and learnt which of these tests should be applied in a given situation.

In this lab session, you will learn how to apply the Z-test, t-test and chi-square test for one-sample using MS Excel 2007. In the next lab session, you will study how to apply the two-sample tests using MS Excel 2007.

Prerequisite

- Lab Sessions 6 and 7 of MSTL-001 (Basic Statistics Lab).
- Block 3 of MST-004 (Statistical Inference).

A parametric test is a test which is based on the fundamental assumption that the form of parent population is known.

A non-parametric test is a test that does not make any assumption regarding the form of the parent population.

Objectives

After performing the activities of this session, you should be able to:

- prepare the spreadsheet in MS Excel 2007;
- apply the Z-test for population mean;
- apply the t-test for population mean;
- apply the Z-test for population proportion; and
- apply the chi-square test for population variance.

9.2 PROBLEM DESCRIPTION

In this lab session, we state three problems to illustrate the applications of different kinds of one-sample parametric tests:

1. We consider a problem related to test the claim of a company manufacturing motorcycles. The company claims that its motorcycles give an average mileage of 60 km/liter. For testing the claim of the company, an analyst randomly selects 24 motorcycles of that company and records their mileage. The data so obtained are given in Table 1.

Table 1: Mileage of motorcycles

S. No.	Mileage (km/liter)	S. No.	Mileage (km/liter)
1	2587	13	2001
2	2789	14	2197
3	2376	15	2231
4	2225	16	2412
5	2400	17	2187
6	2580	18	2221
7	2452	19	2366
8	2241	20	2262
9	2500	21	2357
10	2400	22	2348
11	2457	23	2150
12	2245	24	2014

Assuming that the mileage of the motorcycles is normally distributed:

- Formulate the null and alternative hypotheses.
- Use a suitable test for testing the claim of the company at 5% level of significance when the standard deviation of mileage of the motorcycles is

  i) known to be 6 km/litre and

  ii) unknown.

2. Suppose a researcher wants to test the effectiveness of a particular injection on a disease. He/she takes a sample of 100 patients suffering from that disease and gives the injection. He/she records the data by taking **1** if the disease is cured and **0** if the disease is not cured. The data so recorded are given in Table 2. Test the hypothesis that the proportion of the patients whose disease is cured is more than 80% at 1% level of significance.

Table 2: Category-wise data of 100 patients

Patient Number	Effect (cured or not cured)	Patient Number	Effect (cured or not cured)	Patient Number	Effect (cured or not cured)
1	1	35	0	69	0
2	1	36	0	70	1
3	1	37	0	71	0
4	0	38	0	72	0
5	0	39	1	73	1
6	1	40	0	74	1
7	1	41	1	75	1
8	0	42	0	76	0
9	0	43	0	77	0
10	1	44	0	78	0
11	0	45	1	79	1
12	0	46	1	80	1
13	0	47	0	81	0
14	0	48	0	82	1
15	1	49	1	83	1
16	1	50	1	84	0

Patient Number	Effect (cured or not cured)	Patient Number	Effect (cured or not cured)	Patient Number	Effect (cured or not cured)
17	0	51	1	85	0
18	0	52	0	86	0
19	1	53	0	87	1
20	1	54	1	88	0
21	0	55	0	89	0
22	0	56	0	90	1
23	1	57	1	91	0
24	0	58	0	92	1
25	0	59	0	93	1
26	1	60	0	94	1
27	1	61	1	95	1
28	1	62	0	96	0
29	0	63	1	97	1
30	1	64	0	98	0
31	0	65	0	99	0
32	1	66	0	100	1
33	1	67	0		
34	0	68	1		

3. A fruit juice manufacturing company uses automatic machines to fill 500 ml juice bottles. The company assumes that the variance of the volume of the juice bottles is 4.0 sq ml as a standard. To test this, a quality control inspector of the company collected 20 juice bottles at random from the process and measured the volume of each sampled bottle. She obtained the data given in Table 3.

Table 3: Volume of fruit juice in the sampled bottles

Sample Number	Volume of Juice per bottle (in ml)	Sample Number	Volume of Juice per bottle (in ml)
1	497.32	14	499.56
2	504.76	15	502.24
3	499.24	16	501.76
4	499.26	17	500.65
5	502.32	18	501.12
6	502.12	19	501.00
7	499.34	20	497.50
8	499.38		
9	501.26		
10	498.60		
11	502.44		
12	501.26		
13	497.32		

Assuming that the volume of the juice bottle is normally distributed, is there enough evidence that the volume of juice in the bottles has variance less than the given standard at 5% level of significance?

9.3 Z-TEST FOR POPULATION MEAN

In Unit 10 of MST-004, you have learnt that the Z-test is used for testing the population mean when the population standard deviation (σ) is known and the population under study is normal or non-normal for the large sample. But when sample size is small, we apply the Z-test only when the population under study is normal.

The testing procedure of the Z-test has been described in Unit 10. We briefly mention the main steps as follows:

A test for testing the null hypothesis is said to be a two-tailed test if the alternative hypothesis is two-tailed. If the alternative hypothesis is one-tailed then the test is said to be one-tailed test.

Step 1: We first formulate the null hypothesis (H_0) and alternative hypothesis (H_1). If μ_0 is the value of the population mean which has to be tested, we can take the null and alternative hypotheses as follows:

$$H_0: \mu = \mu_0 \text{ and } H_1: \mu \neq \mu_0 \qquad [\text{for two-tailed test}]$$

or

$$\left.\begin{array}{l} H_0: \mu \leq \mu_0 \text{ and } H_1: \mu > \mu_0 \\ H_0: \mu \geq \mu_0 \text{ and } H_1: \mu < \mu_0 \end{array}\right\} \qquad [\text{for one-tailed test}]$$

Step 2: We calculate the value of the test statistic using the formula given below:

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}} \qquad \text{... (1)}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – the sample mean, and

σ – the population standard deviation.

Let Z_{cal} be the calculated value of the test statistic Z.

Step 3: We obtain the critical (cut-off or tabulated) value(s) of the test statistic Z corresponding to the given level of significance (α).

Step 4: We take the decision about the null hypothesis as follows:

i) Using the critical region approach

In this approach, we compare the calculated value of the test statistic (Z_{cal}) with the critical value(s) obtained in Step 2 and Step 3, respectively. Since critical value(s) depends upon the nature of the test, i.e., whether it is one-tailed or two-tailed, the following cases arise:

In case of two-tailed test:

$$H_0: \mu = \mu_0 \text{ and } H_1: \mu \neq \mu_0$$

Since the test is two-tailed, there exist two critical values, one lying on the left tail and the other on the right tail. Let $-z_{\alpha/2}$ and $z_{\alpha/2}$ be the critical values on the left tail and the right tail, respectively, at α% level of significance (see Fig. 9.1). If $Z_{cal} \geq z_{\alpha/2}$ or $Z_{cal} \leq -z_{\alpha/2}$, it means that Z_{cal} lies in the rejection region and we reject H_0. If $-z_{\alpha/2} < Z_{cal} < z_{\alpha/2}$,

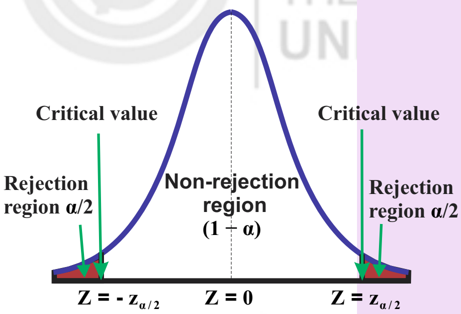


Fig. 9.1

it means that Z_{cal} lies in the non-rejection region and we do not reject H_0.

In case of right-tailed test:

$H_0 : \mu \le \mu_0$ and $H_1 : \mu > \mu_0$

Let z_α be the critical value at α% level of significance which lies on the right tail (see Fig.9.2). If $Z_{cal} \ge z_\alpha$, it means that Z_{cal} lies in the rejection region and we reject H_0. If $Z_{cal} < z_\alpha$, it means that Z_{cal} lies in the non-rejection region and we do not reject H_0.

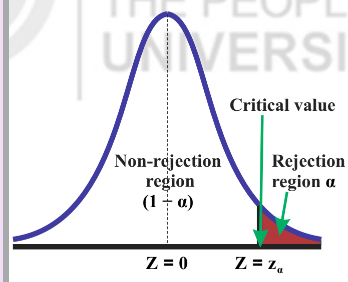


Fig. 9.2

In case of left-tailed test:

$H_0 : \mu \ge \mu_0$ and $H_1 : \mu < \mu_0$

Let $-z_\alpha$ be the critical value at α% level of significance which lies on the left tail in this case (see Fig 9.3). If $Z_{cal} \le -z_\alpha$, it means that Z_{cal} lies in the rejection region and we reject H_0. If $Z_{cal} > -z_\alpha$, it means that Z_{cal} lies in the non-rejection region and we do not reject H_0.

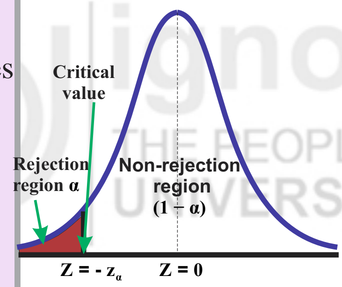


Fig. 9.3

ii) Using p-value approach

We calculate the p-value using the following formulae as required:

$$\text{p-value} = 2P\left[Z \ge |Z_{cal}|\right] \quad [\text{for two-tailed test}] \quad \dots (2)$$

$$\text{p-value} = P[Z \ge Z_{cal}] \quad [\text{for right-tailed test}] \quad \dots (3)$$

$$\text{p-value} = P[Z \le Z_{cal}] \quad [\text{for left-tailed test}] \quad \dots (4)$$

We compare the calculated p-value with the given level of significance (α). If p-value is less than or equal to α, we reject the null hypothesis and if it is greater than α, we do not reject the null hypothesis.

Step 5: Conclusion

If null hypothesis is rejected, we conclude that the sample provides us sufficient evidence against the null hypothesis at α% level of significance.

If null hypothesis is not rejected, we conclude that the sample does not provide us sufficient evidence against the null hypothesis at α% level of significance.

Steps in Excel

In the given Problem 1, the sample size n = 24 (< 30) is small, the standard deviation of the mileage of the motorcycles (population standard deviation) is known to be 6 km/liter and the mileage of the motorcycles is normally distributed. So we use the Z-test for population mean.

For applying the Z-test, we first have to set up the null and alternative hypotheses. Here we have to test the claim of the motorcycle manufacturer that the motorcycles give an average mileage of 60 km/litre. So we can take the null and alternative hypotheses as follows:

$H_0 : \mu = \mu_0 = 60$ (claim)

and $H_1 : \mu \ne \mu_0 = 60$ [two-tailed test]

The following steps are used to apply the one-sample Z-test in Excel 2007:

Step 1: We enter the data (given in Table 1) in Excel 2007 spreadsheet. We start by entering the heading of the data in Row 1 in Excel sheet and the data itself from Row 2. The entries will go up to Row 25 and Column B. For the given data, the spreadsheet will look as shown in Fig. 9.4.

	A	B
1	S.No.	Mileage (km/litre)
2	1	56
3	2	62
4	3	54
5	4	63
6	5	65
7	6	59
8	7	57
9	8	60
10	9	66
11	10	59

Fig. 9.4: Partial screenshot of the spreadsheet for the given data.

Step 2: We calculate the average (mean) mileage of the motorcycles in the sample, i.e., we calculate the sample mean of the mileage of all 24 motorcycles. For this, we use Cell B26. We compute the sample mean $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ as explained under the heading 'Steps in Excel' of Sec. 6.3 of Lab Session 6. That is, we select Cell B26 and click on the ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Average***. Then we get the sample mean in Cell B26 as shown in Fig. 9.5.

B26 | f_x =AVERAGE(B2:B25)

	A	B	C	D
19	18	50		
20	19	55		
21	20	58		
22	21	60		
23	22	54		
24	23	57		
25	24	55		
26	$\bar{X}$	57.8333		
27				

Fig. 9.5

Step 3: We now type the values of μ_0, σ, n and α in Cells B27, B28, B29 and B30, respectively (see Fig. 9.6).

I36 | f_x

	A	B	C	D
25	24	55		
26	$\bar{X}$	57.8333		
27	μ_0	60		
28	σ	6		
29	n	24		
30	α	0.05		
31				

Fig. 9.6

Step 4: We compute the value of $\alpha/2$ (since the test is two tailed) in Cell B31 by typing "=B30/2" and pressing ***Enter*** (Figs. 9.7a and 9.7b).

NORMSDIST | f_x =B30/2

	A	B	C
29	n	24	
30	α	0.05	
31	α/2	=B30/2	
32			

(a)

B31 | f_x =B30/2

	A	B	C
29	n	24	
30	α	0.05	
31	α/2	0.0250	
32			

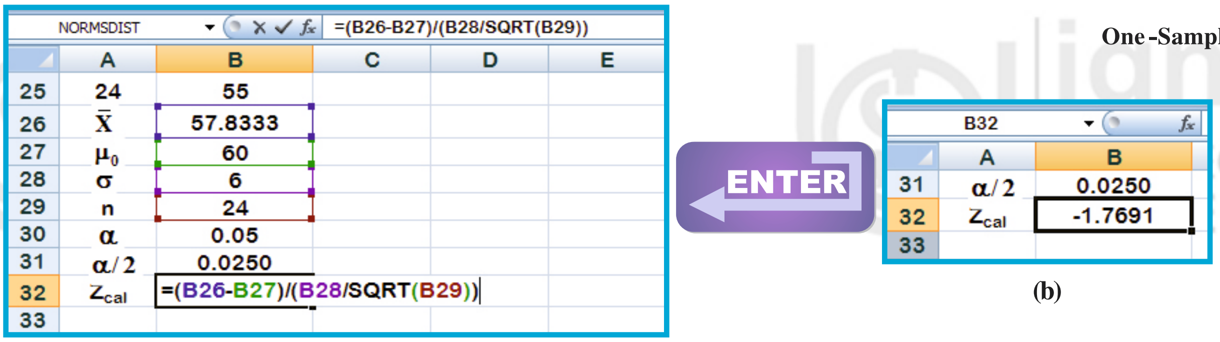
(b)

Fig. 9.7

The test statistic for the Z-test is

$$Z = \frac{\bar{X} - \mu_0}{\sigma/\sqrt{n}}.$$

Step 5: We now compute the value of the test statistic Z using equation (1). Here we shall use Cell B32 for putting the value of Z_{cal}. Since the values of $\bar{X}$, μ_0, σ and n are given in Cells B26, B27, B28 and B29, respectively (see Fig. 9.6), we type "=(B26-B27)/(B28/(Sqrt(B29))" in Cell B32 and press ***Enter***. Then we get the value of Z_{cal} in Cell B32 (Fig. 9.8).



(a) (b)

Fig. 9.8

Step 6: Decision using the critical region approach

We now obtain the critical (tabulated) values with the help of Excel as follows:

1. We select Cell B33,
2. Then we click on ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Normsinv*** function as shown in Fig. 9.9.

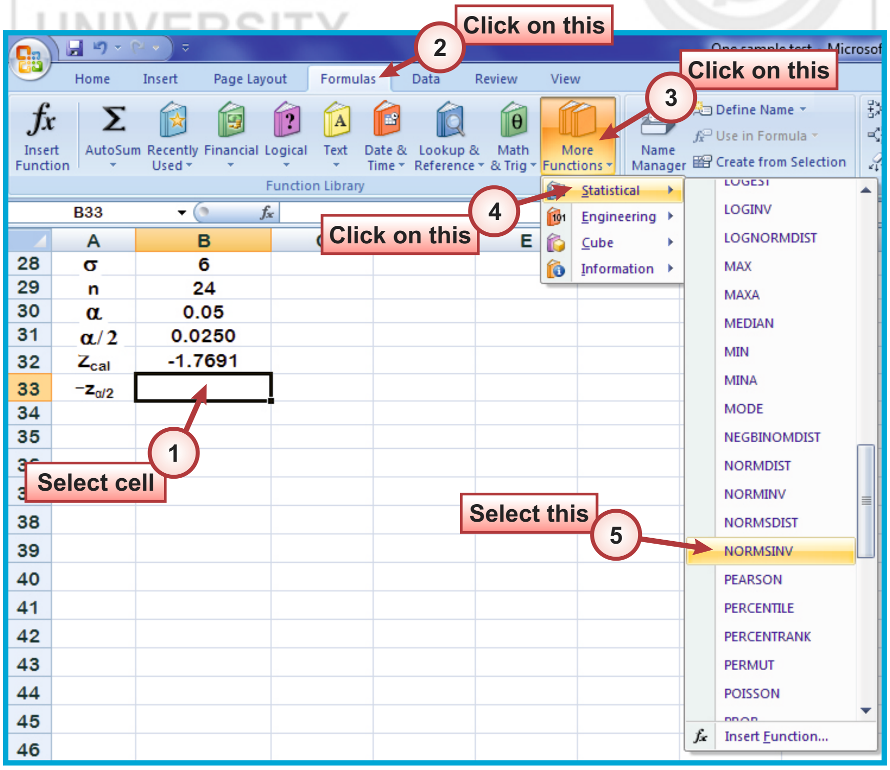
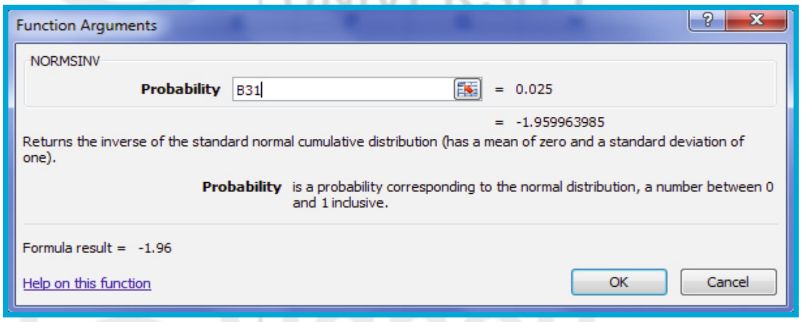


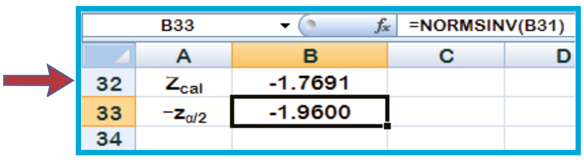
Fig. 9.9

3. A new dialog box is opened as shown in Fig. 9.10a. This dialog box requires the ***probability***, i.e., the level of significance (for two tailed tests, it is $\alpha/2$) which is calculated in Cell B31. So we select Cell B31 and click on ***OK***. This function gives the value shown in Fig. 9.10b.

The function ***Normsinv*** of Excel calculates the probability of the normal distribution on the left-tail, i.e., $P[X < a]$.



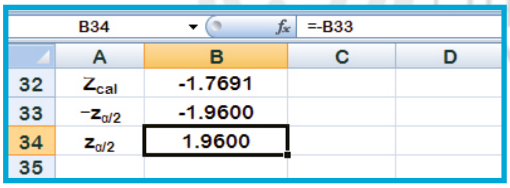
(a)



(b)

Fig. 9.10

4. Since the test is two-tailed, there are two critical values. We now obtain the value of $z_{\alpha/2}$ in Cell B34 by typing "=-B33" and pressing ***Enter***. We get the output shown in Fig. 9.11.



Fig· 9·11

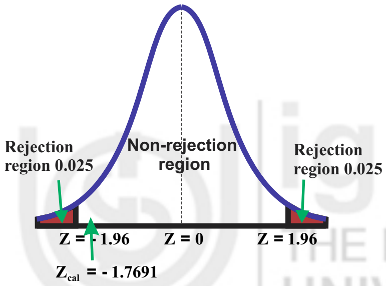


Fig. 9.12

Conclusion

We compare the calculated and critical values of the test statistic obtained in Steps 5 and 6. Since $-z_{\alpha/2}(=-1.96) < Z_{cal}(=-1.7691) < z_{\alpha/2}(=1.96)$, it means that Z_{cal} lies in the non-rejection region as shown in Fig. 9.12. We do not reject the null hypothesis. Since our claim is the null hypothesis, we support the claim. Hence, we conclude that the sample does not provide us sufficient evidence against the claim. So we may assume that the average mileage of the motorcycles of the company is 60 km/liter at 5% level of significance.

Step 7: Decision using the p-value approach

Since the test is two tailed,

p-value = $2P\left[Z \geq \left|z_{cal}\right|\right] = 2P\left[Z \geq \left|-1.7691\right|\right]$. We calculate the p-value as follows:

1. We first calculate $\left|z_{cal}\right| = \left|-1.7691\right|$ in Cell B35 by typing "=Abs(B32)" and pressing ***Enter***. We get the output shown in Fig. 9.13.

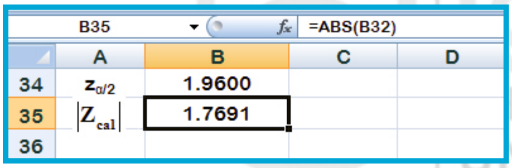


Fig. 9.13

2. Since Excel directly calculates $P\left[Z \leq a\right]$, we first calculate $P\left[Z \leq \left|Z_{cal}\right|\right]$ by selecting ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Normsdist*** function as shown in Fig. 9.14.

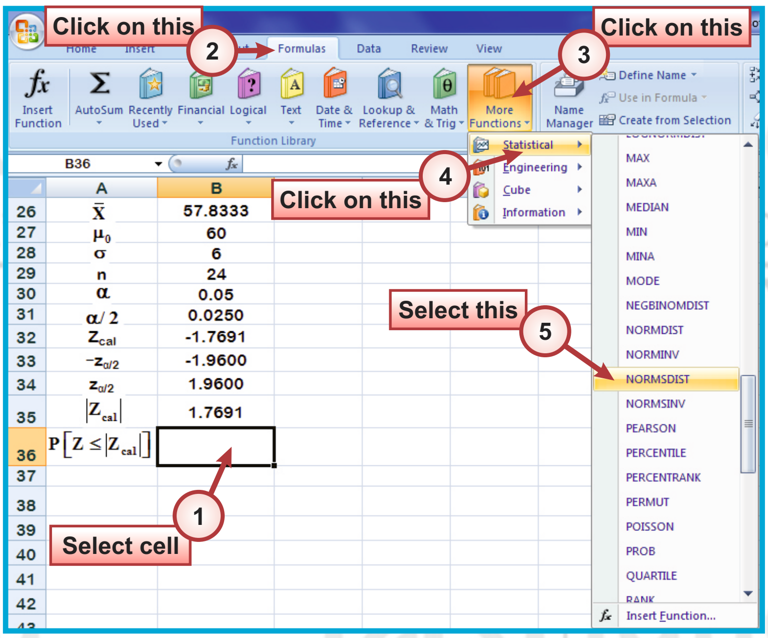
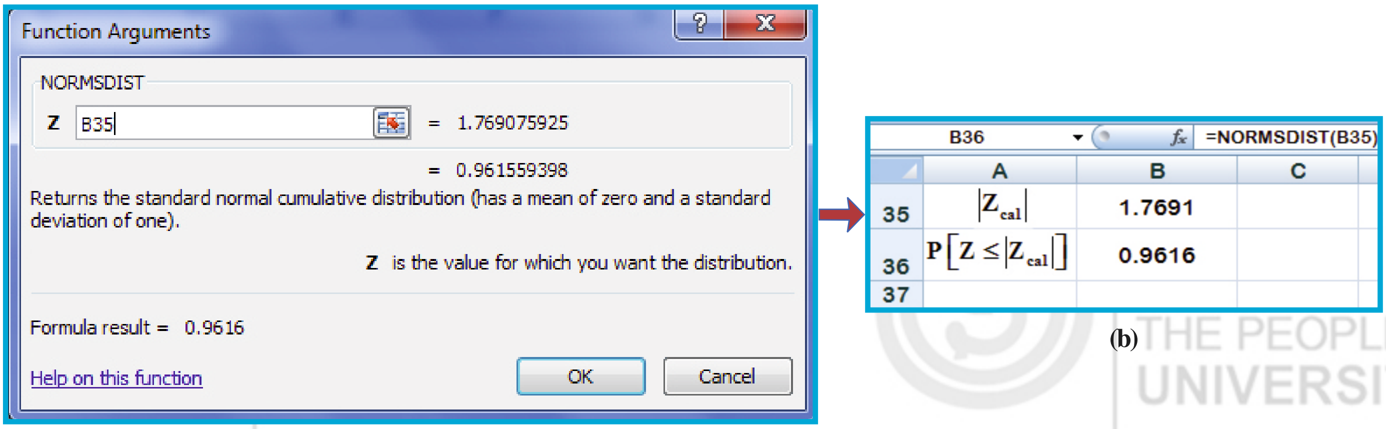


Fig. 9.14

3. We get a new dialog box as shown in Fig. 9.15a. This dialog box requires the value of ***Z*** (i.e., the value up to which we want to calculate the probability). In our problem, $Z = |Z_{cal}|$ which is calculated in Cell B35. So we select Cell B35 in this box and click on ***OK*** (see Fig. 9.15a). This function gives the value of $P[Z \le |Z_{cal}|]$ in Cell B36 as shown in Fig. 9.15b.



(a)

(b)

Fig. 9.15

4. We now calculate the p-value = $2P[Z \ge |Z_{cal}|]$ in Cell B37 by typing "=2*(1-B36)" and pressing ***Enter***. We get the p-value as shown in Fig. 9.16.

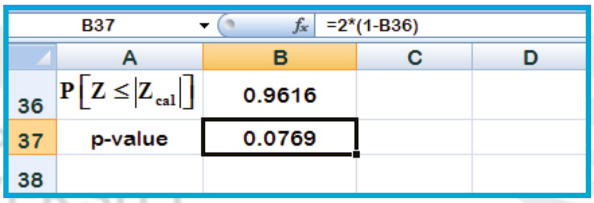


Fig. 9.16

Conclusion

We now take the decision about the null hypothesis on the basis of the p-value approach. We compare the calculated p-value with the level of significance α. Since the p-value (= 0.0769) > α (= 0.05), we do not reject the null hypothesis. Hence, we conclude that the sample does not provide us sufficient evidence against the claim. So we may assume that the average mileage of the motorcycles of the company is 60 km/liter.

9.4 t-TEST FOR POPULATION MEAN

In Unit 11 of MST-004, you have learnt that the t-test is used for testing the population mean when the population standard deviation (σ) is unknown and the population under study is normal.

The testing procedure of this test has been described in Unit 11. We briefly mention the main steps as follows:

Step 1: We first formulate the null hypothesis (H_0) and alternative hypothesis (H_1) in the same way as discussed in Step 1 of the procedure of the Z-test in Sec. 9.3.

Step 2: We calculate the value of the test statistic Z using the formula given ahead:

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}} \qquad \text{... (5)}$$

where $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – the sample mean,

μ_0 – the value of the population mean which has to be tested, and

$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2}$ – the sample standard deviation.

Let t_{cal} be the calculated value of the test statistic t.

Step 3: We obtain the critical (cut-off or tabulated) value(s) of the test statistic t corresponding to the given level of significance (α).

Step 4: We take the decision about the null hypothesis as follows:

i) Using the critical region approach

We compare the calculated value of the test statistic (t_{cal}) with the critical value(s) of the test statistic obtained in Step 2 and Step 3, respectively. Since critical value(s) depends upon the nature of the test (that is, whether it is one-tailed test or two-tailed test), the following cases arise:

In case of two-tailed test:

$H_0: \mu = \mu_0$ and $H_1: \mu \neq \mu_0$

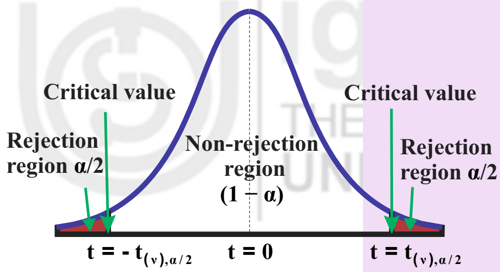


Fig. 9.17

Since the test is two-tailed, there exist two critical values, one on the left tail and the other on the right tail. Let $-t_{(v),\alpha/2}$ and $t_{(v),\alpha/2}$ be the two critical values on the left tail and right tail, respectively, at α% level of significance (see Fig. 9.17) in this case. If $t_{cal} \geq t_{(v),\alpha/2}$ or $t_{cal} \leq -t_{(v),\alpha/2}$, it means that t_{cal} lies in the rejection region and we reject H_0. If $-t_{(v),\alpha/2} < t_{cal} < t_{(v),\alpha/2}$, it means that t_{cal} lies in the non-rejection region and we do not reject H_0.

In case of right-tailed test:

$H_0: \mu \leq \mu_0$ and $H_1: \mu > \mu_0$

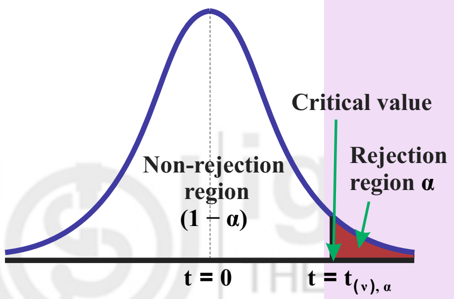


Fig. 9.18

Let $t_{(v),\alpha}$ be the critical values at α% level of significance which lies on the right tail in this case (see Fig. 9.18). If $t_{cal} \geq t_{(v),\alpha}$, it means that t_{cal} lies in the rejection region and we reject H_0. If $t_{cal} < t_{(v),\alpha}$, it means that t_{cal} lies in the non-rejection region and we do not reject H_0.

In case of left-tailed test:

$H_0: \mu \geq \mu_0$ and $H_1: \mu < \mu_0$

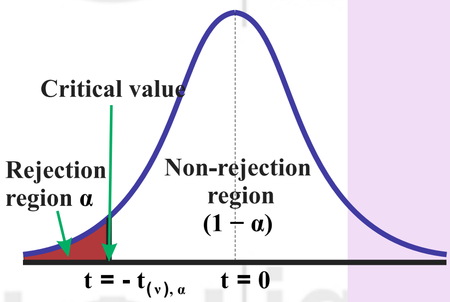


Fig. 9.19

Let $-t_{(v),\alpha}$ be the critical value at α% level of significance which lies on the left tail in this case (see Fig. 9.19). If $t_{cal} \leq -t_{(v),\alpha}$, it means that t_{cal} lies in the rejection region and we reject H_0. If $t_{cal} > -t_{(v),\alpha}$, it means that t_{cal} lies in the non-rejection region and we do not reject H_0.

ii) Using the p-value approach:

We calculate the p-value using the following formulae as required:

$$\text{p-value} = 2P\left[t \geq |t_{cal}|\right] \quad \text{[for two-tailed test]} \quad \ldots (6)$$

$$\text{p-value} = P[t \geq t_{cal}] \quad \text{[for right-tailed test]} \quad \ldots (7)$$

$$\text{p-value} = P[t \leq t_{cal}] \quad \text{[for left-tailed test]} \quad \ldots (8)$$

We compare the calculated p-value with the given level of significance (α). If the p-value is less than or equal to α, we reject the null hypothesis and if it is greater than α, we do not reject the null hypothesis.

Step 5: Conclusion

We draw the conclusion about the null hypothesis as discussed in Step 5 of the procedure of the Z-test in Sec. 9.3.

Steps in Excel

In Problem 1(ii), the population standard deviation is unknown and the mileage of the motorcycles is normally distributed. So we can use the t test for the population mean. The following steps are used to apply the t-test in Excel 2007:

Step 1: We repeat Steps 1 and 2 under the heading 'Steps in Excel' of Sec. 9.3 to enter the data and calculate the sample mean. We now calculate the sample standard deviation (S) using the formula

$$S = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2} \text{ as follows:}$$

1. We select Cell B27.
2. We click on ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Stdev*** function as shown in Fig. 9.20.

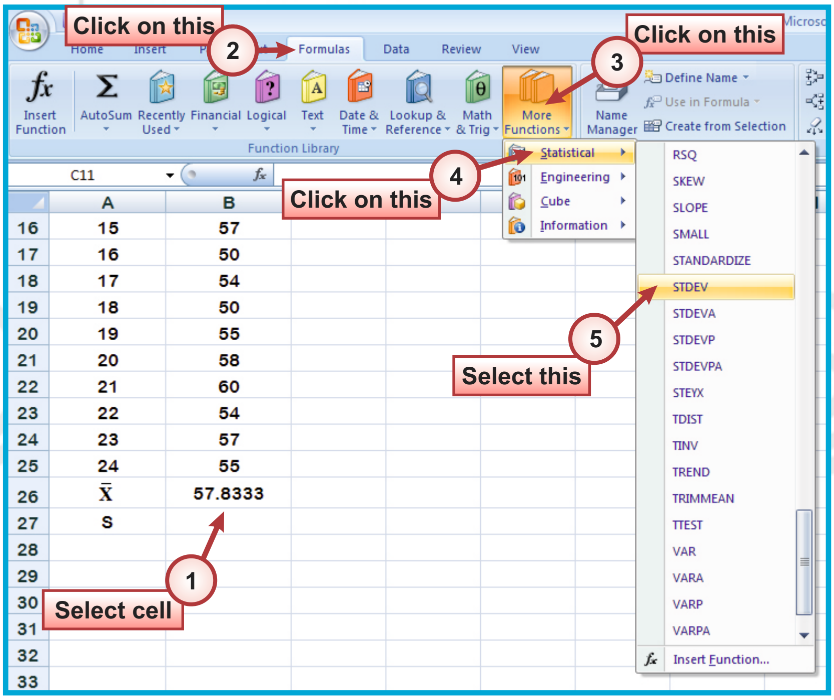
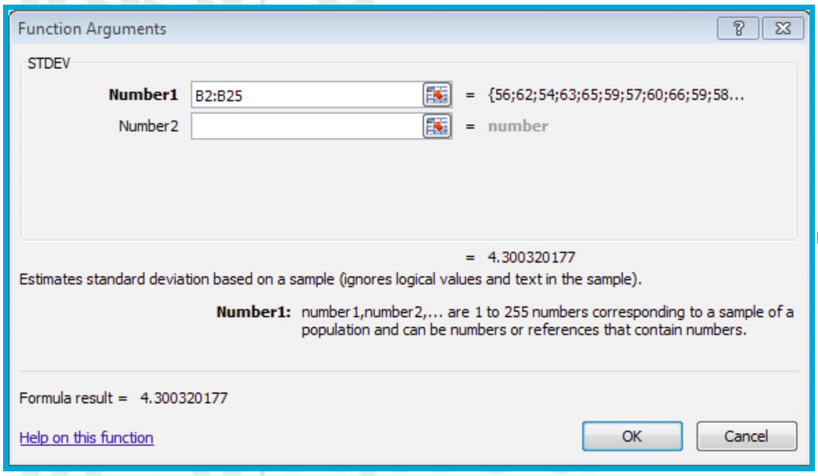


Fig. 9.20

3. We get a new dialog box. This dialog box requires the data for which we want to calculate the standard deviation(S) under ***Value 1***.

So we select Cells B2:B25 in this box and then click on ***OK*** (see Fig. 9.21a). We get the value of S as shown in Fig. 9.21b.



(a)

B27 f_x =STDEVA(B2:B25)

	A	B	C
25	24	55	
26	$\bar{X}$	57.8333	
27	S	4.3003	
28			

(b)

Fig. 9.21

The test statistic

$$t = \frac{\bar{X} - \mu_0}{S/\sqrt{n}}.$$

Step 2: We now type the values of μ_0, n, df (n – 1 = 23) and α in Cells B28, B29, B30 and B31, respectively (Fig. 9.22).

	A	B	C	D
26	$\bar{X}$	57.8333		
27	S	4.3003		
28	μ_0	60		
29	n	24		
30	df	23		
31	α	0.05		
32				

Fig. 9.22

Step 3: We compute the value of the test statistic t using equation (5). Here we shall use Cell B32 for putting the value of the test statistic. Since the values of $\bar{X}$, μ_0, S and n are given in Cells B26, B27, B28 and B29, respectively (see Fig. 9.22), we type "=(B26 - B28)/(B27/(Sqrt(B29))" in Cell B32, and press ***Enter*** (see Fig. 9.23a). Then we get the value of t_{cal} in Cell B32 as shown in Fig. 9.23b.

STDEVA f_x =(B26-B27)/(B28/SQRT(B29))

	A	B	C	D
25	24	55		
26	$\bar{X}$	57.8333		
27	S	4.3003		
28	μ_0	60		
29	n	24		
30	df	23		
31	α	0.05		
32	t_{cal}	=(B26-B27)/(B28/SQRT(B29))		
33				

(a)

	A	B	C
31	α	0.05	
32	t_{cal}	4.3710	
33			

(b)

Fig. 9.23

Step 4: Decision using the critical region approach

We now obtain the critical (tabulated) values with the help of Excel as follows:

1. We select Cell B33,
2. Then we click on ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Tinv*** function as shown in Fig. 9.24.

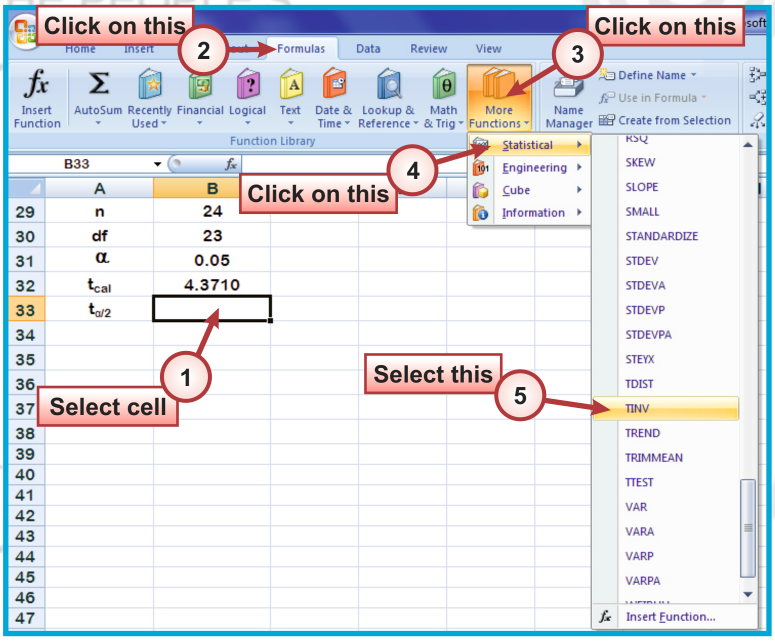
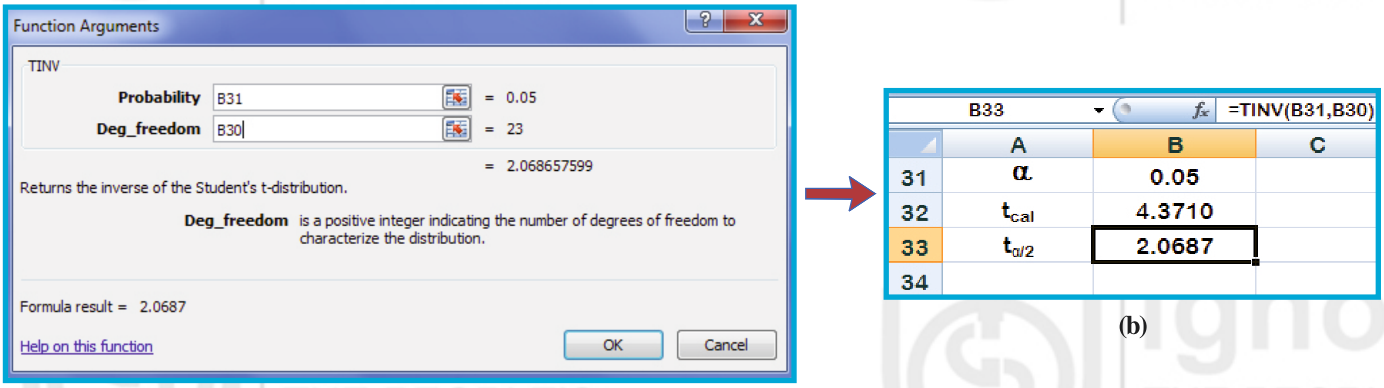


Fig. 9.24

3. We get a new dialog box (Fig. 9.25a). The dialog box requires the value of ***probability*** (level of significance) and ***Deg_ freedom***. The function ***Tinv*** of Excel calculates only right tail critical value, for two tailed t-test. So we consider $\alpha = 0.05$ (for two tailed test), which is put in Cell B30. We select Cell B31 in ***Probability*** and Cell B30 in ***Deg_ freedom*** and click on ***OK*** (see Fig. 9.25a). This function gives the value of $t_{\alpha/2}$ as shown Fig. 9.25b.



(a)

Fig. 9.25

4. We now obtain the value of $-t_{\alpha/2}$ in Cell B34 by typing "=-B33", (see Fig. 9.26a) and pressing ***Enter***. We get the output shown in Fig. 9.26b.

(a) (b)

Fig. 9.26

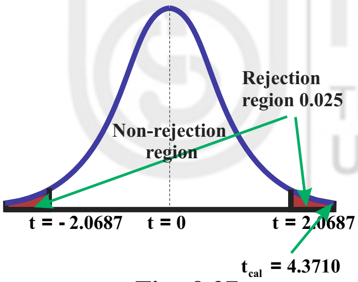


Fig. 9.27

Conclusion

We now take the decision about the null hypothesis. We compare the calculated value of the test statistic (t_{cal}) with the critical values obtained in Step 3 and Step 4, respectively. Since t_{cal} (= 4.3710) > $-t_{\alpha/2}$ (= –2.0687) and $t_{\alpha/2}$ (= 2.0687), it means that it lies in the rejection region (Fig. 9.27). So we reject the null hypothesis. Since the null hypothesis is the claim, we reject the claim. Hence, we conclude that the sample provides us sufficient evidence against the claim. So we assume that the average mileage of the motorcycles of the company is not 60 km/litre at 5% level of significance.

Step 5: Decision using the p-value approach

Since the test is two-tailed, p-value $= 2P\left[t \geq |t_{cal}|\right]$

We first calculate $P\left[t \geq t_{cal}\right] = P\left[t \geq 4.3710\right]$ as follows:

1. We select Cell B35.
2. Then we click on ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Tdist*** function as shown in Fig. 9.28.

Tdist function of Excel calculates $P\left[t \geq a\right]$.

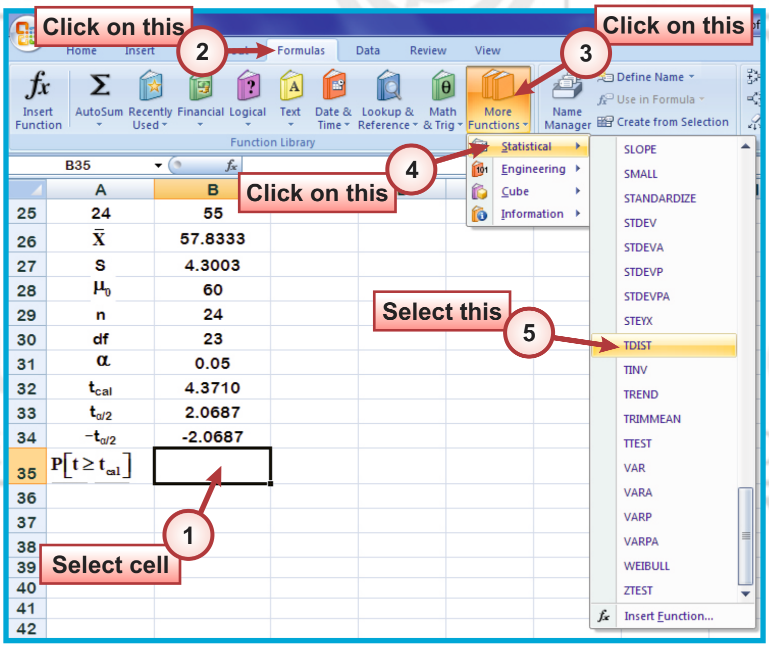
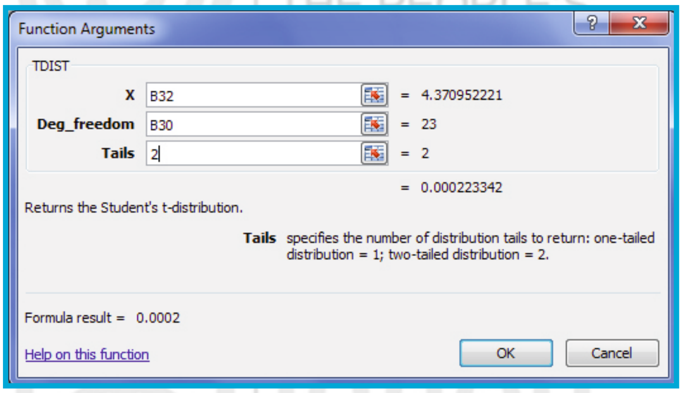
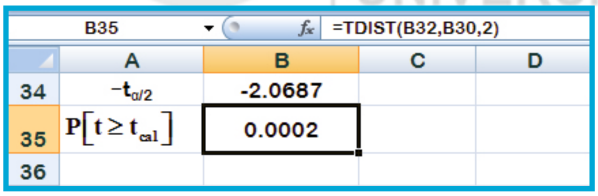


Fig. 9.28

3. We get a new dialog box. The dialog box requires the value of ***X*** (t_{cal}), ***Deg_freedom*** (df) and ***Tails*** (two-tailed). The value of t_{cal} is calculated in Cell B32 and df is given in Cell B30. So we select Cell B32 in ***X*** and Cell B30 in ***Deg_freedom*** and type 2 for two-tailed test in ***Tails*** and then click on ***OK*** (Fig. 9.29a). This function gives the value of $P[t \geq t_{cal}]$ as shown in Fig. 9.29b.



(a)



(b)

Fig. 9.29

4. We calculate the p-value = $2P\left[t \geq |t_{cal}|\right] = 2P\left[t \geq 4.3710\right]$ in Cell B36 by typing "=2*(1-B36)", and pressing ***Enter***. We get the p-value as shown in Fig. 9.30.

B36 — f_x =2*B35

	A	B	C	D
35	$P\left[t \geq t_{cal}\right]$	0.0002		
36	p-value	0.0004		
37				

Fig. 9.30

Conclusion

We now take the decision about the null hypothesis on the basis of the p-value. We compare the calculated p-value with the level of significance α. Since the p-value (= 0.0004) is less than α (= 0.05), we reject the null hypothesis. Hence, we conclude that the sample provides us sufficient evidence against the claim. So we assume that the average mileage of the motorcycles of the company is not 60 km/liter at 5% level of significance.

9.5 Z-TEST FOR POPULATION PROPORTION

In Unit 10 of MST-004, you have learnt that sometimes data are in the form of counts or data are classified into two categories or groups according to an attribute or a characteristic. For example, the people living in a colony may be classified into two groups (male and female) with respect to the characteristic sex, the lot of articles may be classified as defective and non-defective, etc. In such situations, the data are available in dichotomous or binary outcomes, which is a special case of nominal scale. Data categorised into two mutually exclusive and exhaustive classes are generally known as success and failure outcomes. For example, the characteristic sex can be measured as success if male and failure if female or vice versa. So in such situations, we use proportion instead of mean. And for testing the claim about the population proportion, we use the Z-test for population proportion.

In Unit 10 of MST-004, you have learnt the procedure of the Z-test for testing population proportion. We briefly mention the main steps as follows:

Step 1: We first formulate the null hypothesis (H_0) and alternative hypothesis (H_1). If P_0 is the specified value of the population proportion (P) which we wish to test, we can take the null and alternative hypotheses as follows:

$$H_0 : P = P_0 \text{ and } H_1 : P \neq P_0 \qquad [\text{for two-tailed test}]$$

or

$$\left.\begin{array}{l} H_0 : P \leq P_0 \text{ and } H_1 : P > P_0 \\ H_0 : P \geq P_0 \text{ and } H_1 : P < P_0 \end{array}\right\} \qquad [\text{for one-tailed test}]$$

Step 2: We calculate the value of the test statistic Z using the formula given below:

$$Z = \frac{p - P_0}{\sqrt{\frac{P_0 Q_0}{n}}} \qquad \dots (9)$$

where n – the sample size, and

$p = \frac{X}{n}$ – the sample proportion, X denotes the number of observations or elements that possess a certain attribute out of n observations of the sample, and

$Q_0 = 1 - P_0$.

Step 3: The other steps are the same as discussed in the procedure for the Z-test in Sec. 9.3.

Steps in Excel

We apply the Z-test for population proportion for Problem 2. We have to set up the null hypothesis (H_0) and alternative hypothesis (H_1) first. Here we have to test the hypothesis that the proportion of the patients cured the disease is more than 80%. So we can take the null and alternative hypotheses as follows:

$$H_0 : P \leq P_0 = 0.80 \text{ and } H_1 : P > P_0 = 0.80 \text{ (claim)} \quad [\text{right-tailed test}]$$

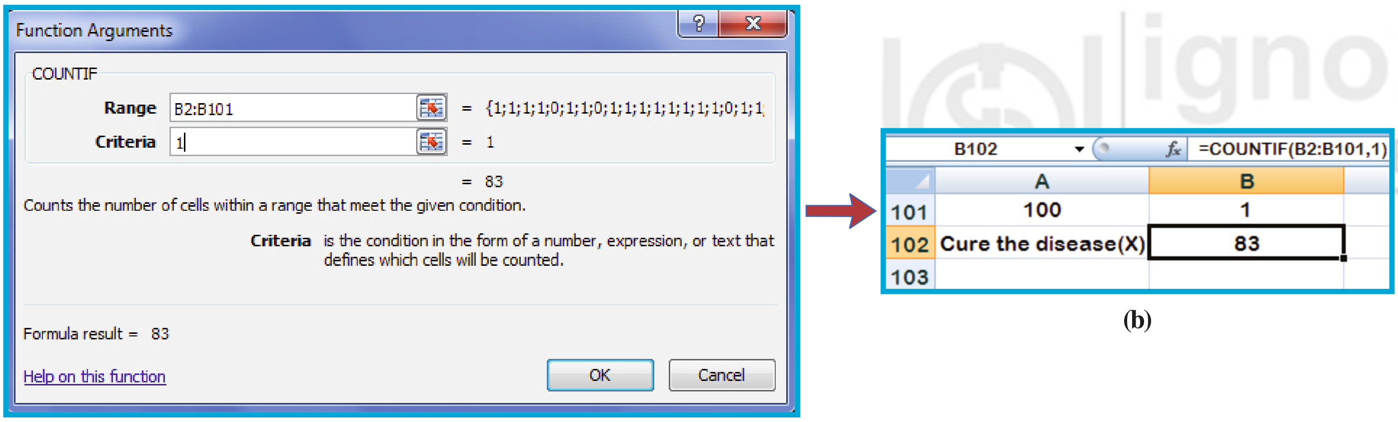
The following steps are used to apply the Z-test for population proportion in Excel 2007:

Step 1: We enter the data (given in Table 2) in Excel sheet as shown in Fig. 9.31.

	A	B	C
1	Person No.	Effect (cure or non-cure)	
2	1	1	
3	2	1	
4	3	1	
5	4	1	
6	5	0	
7	6	1	
8	7	1	
9	8	0	
10	9	1	
11	10	1	

Fig. 9.31: Partial screenshot of the spreadsheet for the given data.

Step 2: Here the attribute is that the disease is cured. So we first count the number of patients whose disease is cured, in Cell B102, as explained under the heading 'Steps in Excel' of Sec. 2.3 of Lab Session 2 (Fig. 9.32a). We get the output shown in Fig. 9.32b.



(a)

(b)

Fig. 9.32

Step 3: We type the values of n, P_0 and α in Cells B103, B104 and B105, respectively, as shown in Fig. 9.33.

	A	B	C
102	Cure the disease(X)	83	
103	n	100	
104	P_0	0.80	
105	α	0.01	
106			

Fig. 9.33

Step 4: We calculate the value of Q_0 (= 1– P_0) in Cell B106 by typing "=1-B104" (see Fig. 9.34a) and the value of p (= X/n) in Cell B107 by typing "=B102/B103" (Fig. 9.34b).

B106 f_x =1-B104

	A	B
105	α	0.01
106	Q_0	0.20
107		

(a)

B107 f_x =B102/B103

	A	B
106	Q_0	0.20
107	p	0.83
108		

(b)

Fig. 9.34

Step 5: We now calculate the value of the test statistic Z using equation (9). Here we shall use Cell B108 for putting the value of the test statistic. The formula for the test statistic Z is $Z = \dfrac{p - P_0}{\sqrt{\dfrac{P_0 Q_0}{n}}}$. Since the values of p, P_0, Q_0 and n are given in Cells B107, B104, B106 and B103, respectively, we type "=(B107-B104)/Sqrt(B104*B106/B103)" in Cell B108, and press ***Enter*** (see Fig. 9.35a). Then we get the calculated value of the test statistic Z in Cell B108 (Fig. 9.35b).

COUNTIF f_x =(B107-B104)/SQRT(B104*B106/B103)

	A	B	C	D
102	Cure the disease(X)	83		
103	n	100		
104	P_0	0.80		
105	α	0.01		
106	Q_0	0.20		
107	p	0.83		
108	Z_{cal}	=(B107-B104)/SQRT(B104*B106/B103)		
109				

(a)

B108 f_x =(B107-B104)/SQRT(B104*B106/B103)

	A	B	C	D
107	p	0.83		
108	Z_{cal}	0.75		
109				

(b)

Fig. 9.35

Step 6: Decision using the critical region approach

Since the test is one tailed, we compute the critical (tabulated) value by taking α = 0.01 in place of α/2 in Cell B109 as explained in Step 6 under the heading 'Steps in Excel' of Sec. 9.3. The output is shown in Fig. 9.36.

B109 | f_x =NORMSINV(B105)

	A	B	C
108	Z_{cal}	0.75	
109	$-z_{\alpha}$	-2.3263	
110			

Fig. 9.36

Step 7: Since Excel calculates the left-tailed critical value ($-z_{\alpha}$), we compute the right-tailed critical value in Cell B110 by typing "=-B109" and pressing ***Enter***. We get the output shown in Fig. 9.37.

B110 | f_x =-B109

	A	B	C
109	$-z_{\alpha}$	-2.3263	
110	z_{α}	2.3263	
111			

Fig. 9.37

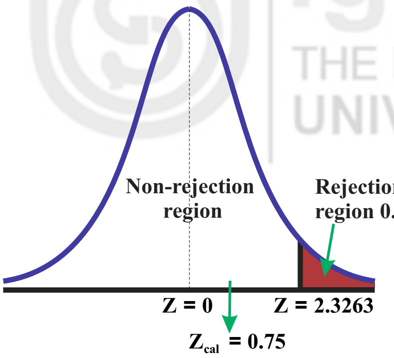


Fig. 9.38

Conclusion

Since the calculated value of the test statistic Z (= 0.75) is less than the critical value z_{α} (= 2.3263), it means that z_{α} lies in the non-rejection region as shown in Fig. 9.38. We do not reject the null hypothesis. Since the alternative hypothesis is the claim, we reject the claim. Hence, we conclude that the sample provides us sufficient evidence against the claim. So we may assume that the proportion of the patients cured the disease is not more than 80% at 1% level of significance.

Step 8: Decision using the p-value approach

Since the test is right-tailed, p-value = $P[Z \geq Z_{cal}]$. Since Excel gives $P[Z \leq Z_{cal}]$, we first calculate $P[Z \leq Z_{cal}]$ and then $P[Z \geq Z_{cal}]$ as explained in Step 7 under the heading 'Steps in Excel' of Sec. 9.3. We get the outputs shown in Figs. 9.39a and 9.39b.

B111 | f_x =NORMSDIST(B108)

	A	B	C
110	z_{α}	2.3263	
111	$P[Z \leq Z_{cal}]$	0.7734	
112			

(a)

B112 | f_x =1-B111

	A	B	C
111	$P[Z \leq Z_{cal}]$	0.7734	
112	$P[Z \geq Z_{cal}]$	0.2266	
113			

(b)

Fig. 9.39

Conclusion

We now compare the calculated p-value with the level of significance α. Since the p-value (= 0.2266) is greater than α(= 0.01), we do not reject the null hypothesis.

9.6 CHI-SQUARE TEST FOR POPULATION VARIANCE

You have learnt about the chi-square test (χ^2-test) for population variance or standard deviation in Unit 12 of MST-004. This test is used when we are interested in testing the hypothesis about the variability of a characteristic under study when the characteristic under study follows the normal distribution. For example, an analyst may be interested in testing the claim of an ambulance agency that the standard deviation of the duration of serving time is less than

5 minutes or the claim of a cigarette manufacturer that the variance of nicotine content in the cigarettes is 0.62, and so on.

The procedure for this test is described in Unit 12. We briefly mention the main steps as follows:

Step 1: We first formulate the null hypothesis (H_0) and alternative hypothesis (H_1). If σ_0^2 is the specified value of the population variance (σ^2) which we wish to test, we can take the null and alternative hypotheses as follows:

$$H_0 : \sigma^2 = \sigma_0^2 \text{ and } H_1 : \sigma^2 \neq \sigma_0^2 \qquad [\text{for two-tailed test}]$$

or

$$\left.\begin{array}{l} H_0 : \sigma^2 \leq \sigma_0^2 \text{ and } H_1 : \sigma^2 > \sigma_0^2 \\ H_0 : \sigma^2 \geq \sigma_0^2 \text{ and } H_1 : \sigma^2 < \sigma_0^2 \end{array}\right\} \quad [\text{for one-tailed test}]$$

Step 2: We calculate the value of the test statistic χ^2 using the formula given below:

$$\chi^2 = \frac{\sum\left(X_i - \bar{X}\right)^2}{\sigma_0^2} = \frac{(n-1)S^2}{\sigma_0^2} \qquad \dots (10)$$

where $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}\left(X_i - \bar{X}\right)^2$ – the sample variance, and

$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$ – the sample mean.

Let χ^2_{cal} be the calculated value of the test statistic χ^2.

Step 3: We obtain the critical (cut-off or tabulated) value(s) of the test statistic χ^2 corresponding to the given level of significance (α).

Step 4: We take the decision about the null hypothesis as follows:

i) Using the critical region approach

We compare the calculated value of the test statistic (χ^2_{cal}) with the critical value(s) obtained in Step 2 and Step 3, respectively. Since critical value(s) depends upon the nature of the test (that is, whether it is one-tailed test or two-tailed), the following cases arise:

In case of two-tailed test:

$H_0 : \sigma^2 = \sigma_0^2$ and $H_1 : \sigma^2 \neq \sigma_0^2$

Let $\chi^2_{(\nu),(1-\alpha/2)}$ and $\chi^2_{(\nu),\alpha/2}$ be the two critical values at the left-tail and right-tail, respectively, on pre-fixed level of significance α in this case (see Fig. 9.40). If $\chi^2_{cal} \geq \chi^2_{(\nu),\alpha/2}$ or $\chi^2_{cal} \leq \chi^2_{(\nu),(1-\alpha/2)}$, it means that χ^2_{cal} lies in rejection region and we reject H_0. If $\chi^2_{(\nu),(1-\alpha/2)} < \chi^2_{cal} < \chi^2_{(\nu),\alpha/2}$, it means that χ^2_{cal} lies in the non-rejection region and we do not reject H_0.

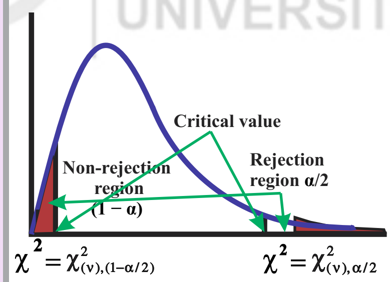


Fig. 9.40

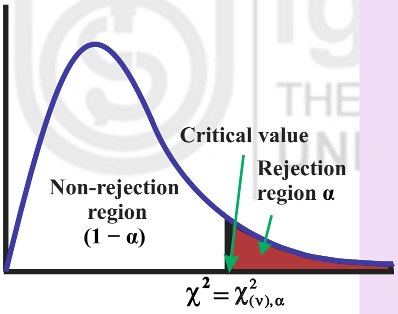


Fig. 9.41

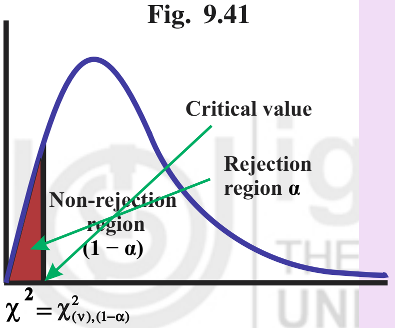


Fig. 9.42

In case of right-tailed test:

$H_0 : \sigma^2 \le \sigma_0^2$ and $H_1 : \sigma^2 > \sigma_0^2$

Let $\chi^2_{(\nu),\alpha}$ be the critical value at α% level of significance which lies in the right tail in this case (see Fig. 9.41). If $\chi^2_{cal} \ge \chi^2_{(\nu),\alpha}$, it means that χ^2_{cal} lies in the rejection region and we reject H_0. If $\chi^2_{cal} < \chi^2_{(\nu),\alpha}$, it means that χ^2_{cal} lies in the non-rejection region and we do not reject H_0.

In case of left-tailed test:

$H_0 : \sigma^2 \ge \sigma_0^2$ and $H_1 : \sigma^2 < \sigma_0^2$

Let $\chi^2_{(\nu),(1-\alpha)}$ be the critical value at α% level of significance which lies in the left tail in this case (see Fig. 9.42). If $\chi^2_{cal} \le \chi^2_{(\nu),(1-\alpha)}$, it means that χ^2_{cal} lies in the rejection region and we reject H_0. If $\chi^2_{cal} > \chi^2_{(\nu),(1-\alpha)}$, it means that χ^2_{cal} lies in the non-rejection region and we do not reject H_0.

ii) Using the p-value value approach

We calculate the p-value using the following formulae as required:

$$\text{p-value} = 2P\left[\chi^2 \ge \chi^2_{cal}\right] \qquad [\text{for two-tailed test}] \qquad \dots (11)$$

$$\text{p-value} = P\left[\chi^2 \ge \chi^2_{cal}\right] \qquad [\text{for right-tailed test}] \qquad \dots (12)$$

$$\text{p-value} = P\left[\chi^2 \le \chi^2_{cal}\right] \qquad [\text{for left-tailed test}] \qquad \dots (13)$$

We compare the calculated p-value with the given level of significance (α). If p-value is less than or equal to α, we reject the null hypothesis and if it is greater than α, we do not reject the null hypothesis.

Step 5: Conclusion

We draw the conclusion as discussed in Step 5 of the procedure for the Z-test in Sec. 9.3.

Steps in Excel

In Problem 3, the volume of the juice bottles is normally distributed and we have to test the claim about the population variance. So we shall use the χ^2-test.

For applying the χ^2-test, we have to set up the null hypothesis (H_0) and the alternative hypothesis (H_1). Here we have to test the hypothesis that the volume of the juice bottles has variance (σ^2) less than the given standard 4.0. So we can take the null and alternative hypotheses as follows:

$$H_0 : \sigma^2 \ge 4.0 \text{ and } H_1 : \sigma^2 < 4.0 \text{ (claim)} \qquad [\text{left-tailedtest}]$$

The following steps are used to apply the χ^2-test for testing the population variance in Excel 2007:

Step 1: We enter the data (given in Table 3) in Excel sheet as shown in Fig. 9.43.

	A	B	C
1	Sample No.	Volume of Juice per bottle (in ml)	
2	1	497.32	
3	2	504.76	
4	3	499.24	
5	4	499.26	
6	5	502.32	
7	6	502.12	
8	7	499.34	
9	8	499.38	
10	9	501.26	
11	10	498.6	

Fig. 9.43: Partial screenshot of the spreadsheet for the given data.

Step 2: We calculate the sample variance (S^2) using the formula $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$. Here we shall use Cell B22 for putting the value of S^2. To calculate the value of S^2, we click on ***Formulas*** → ***More Functions*** → ***Statistical*** → ***Var*** function and repeat the procedure as explained in Step 1 under the heading 'Steps in Excel' of Sec 9.4 for calculating the sample variance. We get the output shown in Fig. 9.44.

B22 | =VAR(B2:B21)

	A	B	C
21	20	497.5	
22	s^2	3.8736	
23			

Fig. 9.44

Step 3: We type the values of σ_0^2, n, df (n – 1 = 19) and α in Cells B23, B24, B25 and B26, respectively, as shown in Fig. 9.45.

	A	B	C
22	s^2	3.8736	
23	σ_0^2	4.0	
24	n	20	
25	df	19	
26	α	0.05	
27			

Fig. 9.45

Step 4: We compute the value of (1 – α) since the test is left-tailed and Excel calculates right-tail value, in Cell B27 by typing "=1-B26" and pressing ***Enter***. We get the output shown in Fig. 9.46.

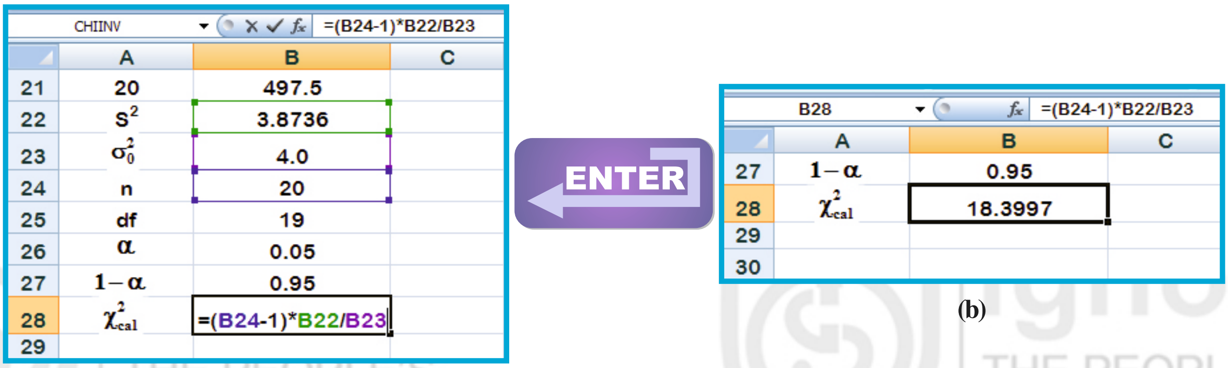
B27 | =1-B26

	A	B	C
26	α	0.05	
27	1–α	0.95	
28			

Fig. 9.46

Step 5: We compute the value of the test statistic χ^2 using equation (10). Here we shall use Cell B28 for putting the value of the test statistic. The formula for calculating the value of test statistic χ^2 is

$\chi^2 = \frac{(n-1)S^2}{\sigma_0^2}$. Since the values of n, S^2 and σ_0^2 are given in Cells B24, B22 and B23, respectively (see Fig. 9.45), we type "=(B24-1)*B22/B23" in Cell B28 and press ***Enter*** (see Fig. 9.47a). Then we get the value of the χ^2_{cal} in Cell B28 (Fig.9.47b).



(a) (b)

Fig. 9.47

Step 6: Decision using the critical region approach

We now obtain the critical (tabulated) value with the help of Excel as follows:

1. We select Cell B29,
2. Then we click on ***Formulas***→***More Functions***→***Statistical*** → ***Chinv*** function as shown in Fig. 9.48.

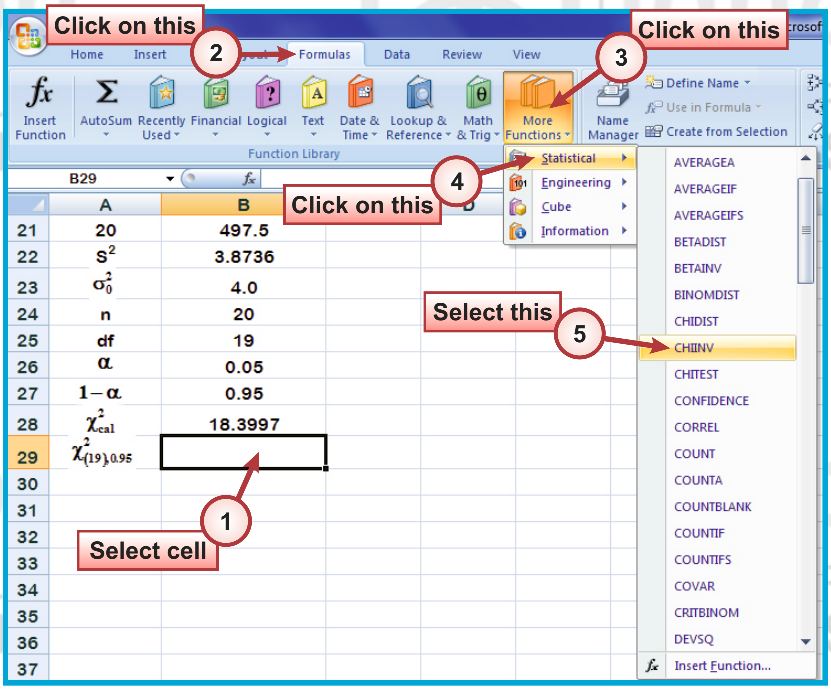
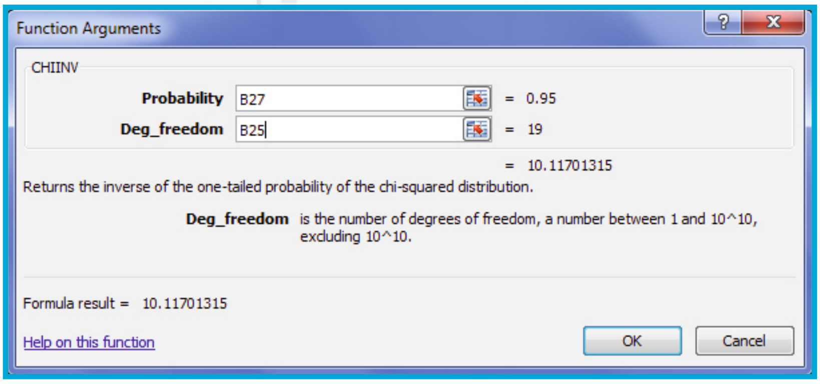
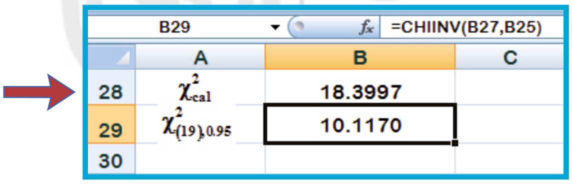


Fig. 9.48

3. We get a new dialog box. This dialog box requires the value of ***probability*** (level of significance α) and ***Deg_freedom***. Since Excel gives the critical value for the right-tailed χ^2-test and we have to find the critical value for the left tail, we consider $1 - \alpha = 0.95$ which is given in Cell B27. So we select Cell B27 in ***Probability*** and Cell B25 in ***Deg_freedom*** and then click on ***OK*** (Fig. 9.49a). This function gives the critical value $\chi^2_{(19),0.95}$ (Fig. 9.49b).



(a)



(b)

Fig. 9.49

Conclusion

To take a decision about the null hypothesis, we compare the calculated value of the test statistics χ^2_{cal} with the critical value $\chi^2_{(19),0.95}$ obtained in Steps 5 and 6, respectively. Since χ^2_{cal} (= 18.3997) is greater than $\chi^2_{(19),0.95}$ (= 10.1170), it means that the calculated value lies in the non-rejection region as shown in Fig. 9.50. So we do not reject the null hypothesis. Since our claim is under the alternative hypothesis, we reject the claim. Hence, we conclude that the sample provides us sufficient evidence against the claim. So we may assume that the volume of juice bottles does not have variance less than the given standard 4.0 at 5% level of significance.

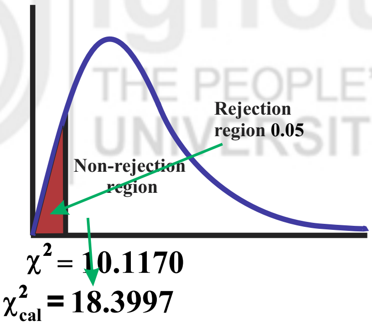


Fig. 9.50

Step 7: Decision according to the p-value

Since our problem is left-tailed, the p-value = $P\left[\chi^2 \le \chi^2_{cal}\right]$. But Excel gives $P\left[\chi^2 \ge \chi^2_{cal}\right]$. So we calculate $P\left[\chi^2 \ge \chi^2_{cal}\right]$ first and then the p-value as follows:

1. We select Cell B30.
2. Then we click on ***Formulas*** → ***More Functions*** → ***Statistical***→ ***Chidist*** function as shown in Fig. 9.51.

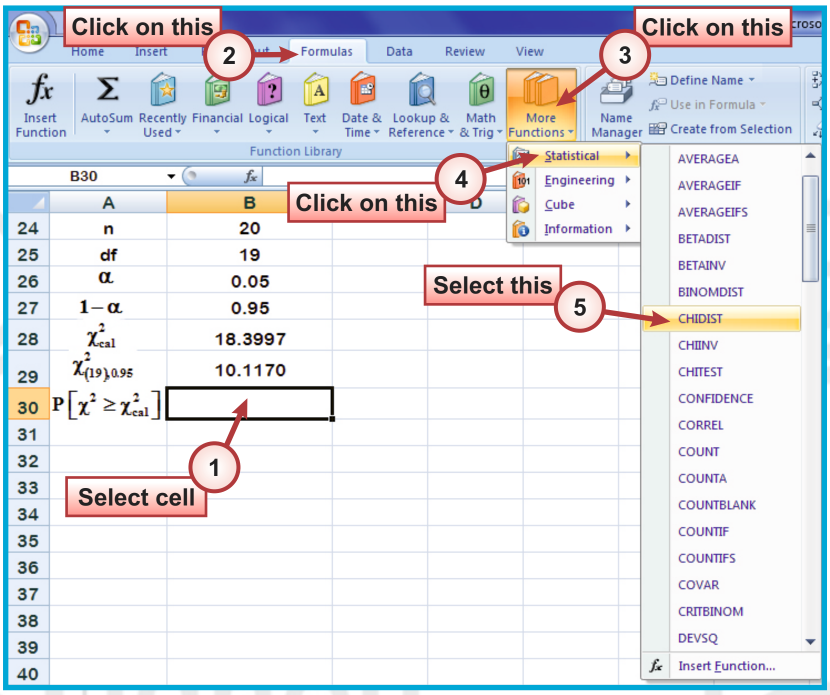
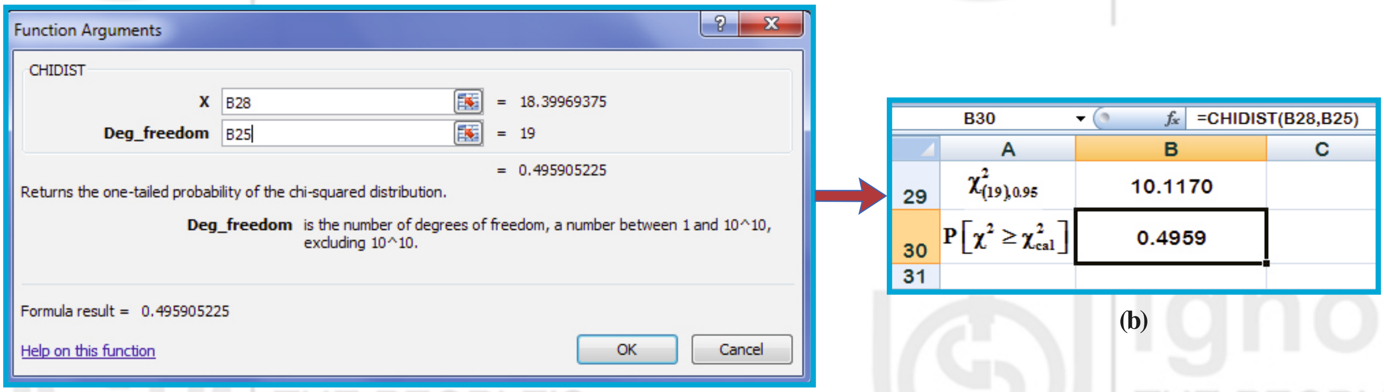


Fig. 9.51

3. A new dialog box is opened. The dialog box requires the value of ***X*** (χ^2_{cal}) and ***Deg_ freedom*** (df). The value of χ^2_{cal} is calculated in Cell B28 and df is given in Cell B25. So we select Cell B28 in ***X*** and Cell B25 in ***Deg_ freedom*** and then click on ***OK*** (Figs. 9.52a). We get the output shown in Fig. 9.52b.



(a) (b)

Fig. 9.52

4. We now calculate the p-value $= P\left[\chi^2 \leq \chi^2_{cal}\right]$ in Cell B31 by typing "=1–B30" in Cell B31 and then pressing ***Enter***. We get the p-value shown in Fig. 9.53.

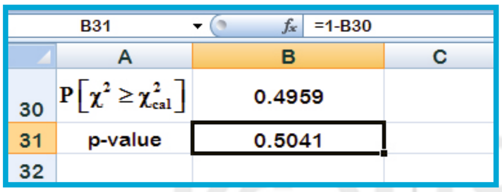


Fig. 9.53

Conclusion

We now compare the calculated p-value with the level of significance $\alpha = 0.05$. Since the p-value (= 0.5041) is greater than α (= 0.05), we do not reject the null hypothesis.

You should now apply these tests in MS Excel to other problems for practice.

Activity

Apply the suitable test on the data with the help of MS Excel 2007 for the following exercises and interpret the results:

A1) Examples 1, 2, 7 and 8 given in Unit 10 of MST-004.

A2) Exercises E4, E5, E10 and E11 given in Unit 10 of MST-004.

A3) Examples 1 and 2 given in Unit 11 of MST-004.

A4) Exercises E2 and E3 given in Unit 11 of MST-004.

A5) Examples 1 and 2 given in Unit 12 of MST-004.

A6) Exercises E1 and E2 given in Unit 12 of MST-004.

Match the results with the manual computation done in Units 10, 11 and 12 of MST-004.

Continuous Assessment 9

1. A tyre manufacturer claims that the average life of a particular category of tyre manufactured by him is 20000 km when used under normal driving conditions. To test the claim, 26 tyres are selected and their lives are measured. The data are given in the following table.

Table 4: Life of tyre (in km)

S. No.	Life of Tyre (in km)	S. No.	Life of Tyre (in km)
1	18500	14	20000
2	18000	15	20500
3	24000	16	16000
4	19000	17	25000
5	17500	18	22000
6	18000	19	23500
7	22000	20	18500
8	21000	21	20000
9	20500	22	19500
10	21000	23	21500
11	24000	24	24000
12	18500	25	26000
13	22000	26	24500

Assuming that the life of the tyres are normally distributed:

- Formulate the null and alternative hypotheses.
- Use the suitable test for testing the claim of the company at 1% level of significance when the standard deviation of the lives of the tyres is

  i) known to be 3000 km; and

  ii) unknown.

2. A national survey of employees found that 75% of them felt that work stress had a negative impact on their personal lives. To check this statement, an analyst took a sample of 120 employees. She recorded the data by recording **Yes** for the employees who said that the claim was true and **No** for those who said that the claim was false. The data are given in Table 5.

Table 5: Data of 120 employees

Employee Number	Answer	Employee Number	Answer	Employee Number	Answer
1	Yes	41	No	81	Yes
2	Yes	42	Yes	82	Yes
3	Yes	43	Yes	83	Yes
4	Yes	44	Yes	84	Yes
5	Yes	45	No	85	Yes
6	Yes	46	Yes	86	Yes
7	No	47	Yes	87	Yes
8	Yes	48	Yes	88	Yes
9	Yes	49	Yes	89	Yes
10	Yes	50	Yes	90	Yes
11	Yes	51	No	91	No
12	No	52	Yes	92	Yes

Employee Number	Answer	Employee Number	Answer	Employee Number	Answer
13	No	53	Yes	93	No
14	No	54	Yes	94	Yes
15	Yes	55	Yes	95	Yes
16	Yes	56	No	96	Yes
17	Yes	57	No	97	No
18	Yes	58	Yes	98	No
19	Yes	59	Yes	99	No
20	Yes	60	Yes	100	Yes
21	No	61	Yes	101	Yes
22	No	62	Yes	102	Yes
23	Yes	63	Yes	103	Yes
24	Yes	64	Yes	104	Yes
25	Yes	65	Yes	105	Yes
26	Yes	66	Yes	106	Yes
27	Yes	67	Yes	107	Yes
28	Yes	68	No	108	Yes
29	Yes	69	No	109	No
30	Yes	70	No	110	No
31	Yes	71	Yes	111	Yes
32	Yes	72	Yes	112	Yes
33	Yes	73	Yes	113	Yes
34	Yes	74	Yes	114	Yes
35	No	75	Yes	115	No
36	Yes	76	Yes	116	Yes
37	Yes	77	Yes	117	Yes
38	Yes	78	Yes	118	Yes
39	Yes	79	No	119	Yes
40	Yes	80	No	120	Yes

Formulate the null and alternative hypotheses and test the claim at 5% level of significance.

3. Suppose a researcher studies the variation in the upper blood pressure of male workers in corporate sectors. She randomly selects 28 male workers and records their upper blood pressure in the following table:

Table 6: Upper blood pressures of sampled workers

Worker Number	Upper Blood Pressure (in mm Hg)	Worker Number	Upper Blood Pressure (in mm Hg)
1	131	15	124
2	120	16	130
3	145	17	142
4	124	18	150
5	180	19	126
6	162	20	100
7	150	21	110
8	100	22	130
9	120	23	110
10	124	24	150
11	152	25	170
12	140	26	132
13	160	27	135
14	155	28	150

The researcher wishes to test the hypothesis that the standard deviation in the upper blood pressure of the workers is greater than 30 mm Hg.

Assuming that the upper blood pressure is normally distributed:

- Formulate the null and alternative hypotheses.
- Use a suitable test for testing the hypothesis at 1% level of significance.

Home Work: Do It Yourself

1) Follow the steps explained in Secs. 9.3, 9.4, 9.5 and 9.6 to apply the tests on the data of Tables 1, 2 and 3. Take the final screenshots and keep them in your record book.

2) Develop the spreadsheets for the exercises of "Continuous Assessment 9" as explained in this lab session. Take a screenshot of the final spreadsheet.

3) **Do not forget** to keep all screenshots in your record book as these will contribute to your continuous assessment in the Laboratory.