UNIT 7 APPLICATIONS OF CHI-SQUARE IN PROBLEMS WITH CATEGORICAL DATA

Structure **Page Nos.**

7.1 INTRODUCTION

In this block, you have already studied several problems of testing of hypotheses. The tests that you have studied so far relate to problems where the sample data have been obtained from a continuous distribution, for example, the normal distribution. In practice, however, one often obtains data in which the sampled "observations" are classified into classes according to one or more attributes. For example, a sample of flowers can be classified according to their colour — some of the flowers in the sample could be white, the others could be purple. Again, suppose it is claimed that a vaccine controls a disease. To 'verify' the truth of this claim, a sample of N individuals is taken and these individuals can be classified according to two attributes — inoculated or not inoculated, and affected or not affected by the disease.

When the sampled data are classified according to one or more attributes, we say that we have a set of **categorical data**. How do we tackle the inference problems arising out of categorical data? In this unit we shall discuss the use of one of the most widely used tests, the chi-square test, in this context.

To start with, in Sec.7.2, we shall consider the use of the chi-square test in "goodness-of-fit" problems. Then, in Sec.7.3, we shall see how the chi-square test can help us compare two features of a population to see if there is any relationship between them or not. In other words, we test to see if the features occur independent of each other or not.

While studying this unit, please keep comparing the situations in this unit and Units 5 and 6 to really understand the difference in the questions being asked and answered.

Objectives

After studying this unit, you should be able to

- define categorical data;
- identify inference problems associated with categorical data;
- use the chi-square test for solving some inference problems arising in categorical data.

7.2 GOODNESS-OF-FIT

Let us begin by trying to solve Ms.Dalta's problem. She is the marketing director of a company that sells four types of steel almirahs. As part of her duties, she has to make sure that there is no loss of sales due to less stock availability. So far she has been ordering new cupboards assuming that the demand for all four types is the same.

Recently, however, the stock inventories have become more difficult to control. Therefore, Ms.Dalta feels that she should check whether her hypothesis of uniform demand is valid or not.

Can you apply any of the methods you have studied so far for helping Ms.Dalta? There is no parameter that she is estimating and no assumption regarding the distribution of the population. So Ms.Dalta needs to look for some new tools. What she needs to do is to test the hypothesis :

H_0 is the null hypothesis, and H_A is the alternative hypothesis.

H_0 : The demand is uniform for all four types of almirahs

against

H_A : The demand is not uniform for all four types of almirahs.

For doing this, she selects a sample of 80 almirahs sold over the past few months. Ms.Dalta assumes that the demand is uniform. So the probability of an almirah of Type i being bought is the same, for i = 1, 2, 3, 4. If we denote this probability by p_i, then $p_1 = p_2 = p_3 = p_4 = \frac{1}{4}$. So, if the demand is uniform, she can expect $80\left(\frac{1}{4}\right) = 20$ almirahs of each type to be sold. But the observed sales of each type are 23, 19, 18 and 20, respectively. Her problem is to see how well her hypothesis of uniform demand fits the observed sales. In other words, how can this data set be used for testing H_0?

More generally, suppose a sample of n individuals are classified into k classes. Suppose the number of individuals falling in the ith class is O_i (i = 1, ..., k). The problem of "**goodness-of-fit**" consists in testing the hypothesis, H_0, that the probability of an individual falling in the ith class is p_i (i = 1, ..., k), where $\sum_{i=1}^{k} p_i = 1$. In other words, the hypothesis H_0 to be tested is that the number of individuals (in a sample of size n) falling in the ith class is np_i (i = 1, ..., k). This is to be tested against the hypothesis H_A, that H_0 is not true.

Note that, $\chi^2_{\alpha,k-1}$ is denoted by 'χ^2_α **with k-1 degrees of freedom' in Unit 4.**

So, in this general situation, the "expected" number of individuals falling in the ith class is $E_i = np_i$ (i = 1, ..., k). Note that these "expected" numbers are known to us because to start with we assume H_0 and calculate them. That is, we assume that the probability of an individual falling in the ith class is p_i (i = 1, ..., k). Based on these numbers $O_1, O_2, \ldots, O_k, E_1, E_2, \ldots, E_k$, there is a way of testing the validity of H_0. Let us see what this method is.

U is also called the **sample χ^2 value**, or the **observed value of χ^2** for the data.

Let $U = \sum_{i=1}^{k} \frac{(O_i - E_i)^2}{E_i}$. This statistic U, under some mild conditions, is known to have an **approximate** chi-square distribution with **k – 1 degrees of freedom**, where k is the number of classes. If we want to test the hypothesis H_0 at the α level of significance, then we need to find $\chi^2_{\alpha,k-1}$, from the standard χ^2 distribution tables (given at the end of this block). If $U > \chi^2_{\alpha,k-1}$, we reject H_0. Otherwise we do not reject H_0.

To see how this test works, let us consider Ms.Dalta's data, presented in Table 1.

Table 1

Type of almirah	Observed sales (O_i)	Expected sales (E_i) $= np_i = 80 \times \left(\frac{1}{4}\right)$
I	23	20
II	19	20
III	18	20
IV	20	20

Here $U = \frac{(O_1 - E_1)^2}{E_1} + \frac{(O_2 - E_2)^2}{E_2} + \frac{(O_3 - E_3)^2}{E_3} + \frac{(O_4 - E_4)^2}{E_4}$

$$= \frac{(23-20)^2}{20} + \frac{(19-20)^2}{20} + \frac{(18-20)^2}{20} + \frac{(20-20)^2}{20}$$

$$= \frac{9+1+4+0}{20} = 0.7$$

Here k = 4. If Ms.Dalta wants to test H_0 at a 5% level of significance, $\alpha = 0.05$. Now, $\chi^2_{0.05,3} = 7.815$. Since $U < \chi^2_{0.05,3}$, Ms.Dalta does not reject H_0. In other words, Ms.Dalta concludes that the demand for the four types of almirahs is uniform.

Another example may help you to see how this test works.

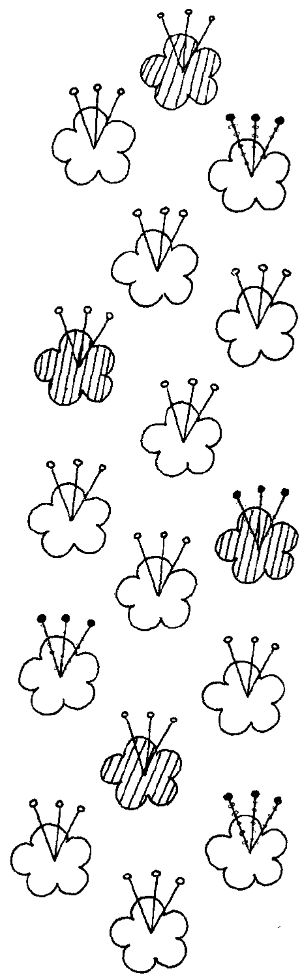


Fig.1

Example 1 (Experiment on the breeding of flowers of a certain species) :
Jaswant is interested in breeding flowers of a certain species. The experimental breeding can result in four possible types of flowers :

(a) magenta flowers with a green stigma (MG),
(b) magenta flowers with a red stigma (MR),
(c) red flowers with a green stigma (RG),
(d) red flowers with a red stigma(RR).

According to the well-known Mendel's law, these four kinds of flowers should come out in the ratio 9 : 3 : 3 : 1. Jaswant found that under her experiment, out of 160 flowers that bloomed, the number of flowers with types MG, MR, RG and RR were 84, 35, 28 and 13, respectively. She wants to find out whether these data are compatible with Mendel's law or not.

If they are compatible, then the probabilities of each of these types of blooming are $p_1 = \frac{9}{16}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{3}{16}$, and $p_4 = \frac{1}{16}$. So Jaswant wants to test the hypothesis

H_0 : The distribution of the flower types is multinomial with
$p_1 = \frac{9}{16}$, $p_2 = \frac{3}{16}$, $p_3 = \frac{3}{16}$, $p_4 = \frac{1}{16}$.
against
H_A : H_0 is not true, that is, the distribution is not multinomial with the specified probabilities.

See the appendix to the unit for a brief introduction to the multinomial distribution.

Jaswant's data can be presented as shown in Table 2.

Table 2

Flower type	Observed number O_i	Expected number E_i (= np_i)
MG	84	90
MR	35	30
RG	28	30
RR	13	10
Total (n)	160	160

Here k = 4, and

$$U = \sum_{i=1}^{4} \frac{(O_i - E_i)^2}{E_i}$$

$$= \frac{(84-90)^2}{90} + \frac{(35-30)^2}{30} + \frac{(28-30)^2}{30} + \frac{(13-10)^2}{10}$$

$$= 2.27$$

Jaswant needs to compare this value with the appropriate critical χ^2-value. She takes the significance level of the test as $\alpha = 0.05$. Also, in this case, since the number of classes is 4, the degrees of freedom are $4 - 1 = 3$. So, she finds $\chi^2_{0.05,3}$, which is 7.81. Since $U = 2.27 < 7.81 = \chi^2_{0.05,3}$, she does not reject H_0. Thus, Jaswant concludes that her data is compatible with Mendel's law.

* * *

In the two situations above, the hypothetical probabilities $p_1, p_2, \ldots$ were known to us from before because of the type of assumption H_0 was. However, in some problems, these probabilities may have to be estimated from the data itself. The following example illustrates this.

Example 2 : A consultant was employed by a city council to study the pattern of bus arrivals and departures at a very busy interstate bus terminus. Since many

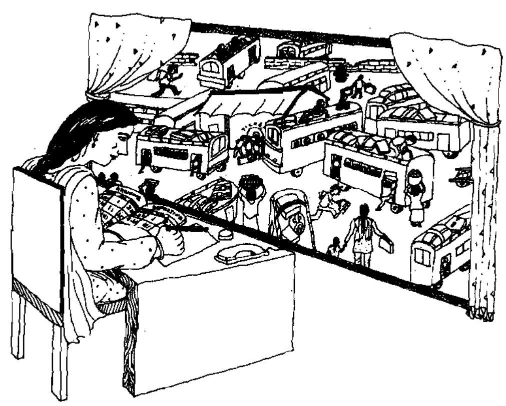


Fig.2

arrival processes fit the Poisson distribution, she decided to test the following hypothesis :

H_0 : The arrivals are distributed as a Poisson random variable,
against
H_A : The arrivals are not Poisson distributed.

She sampled the number of arrivals in 200 minutes. Then she grouped the arrivals into k = 6 categories, and noted her observations, as shown in Column 2 of Table 3 below.

However, since the parameter of the Poisson distribution is unspecified in the hypothesis, the consultant needed to estimate this from the data itself. For this she first computed the sample mean as

$$\bar{x} = \frac{(1\times 23)+(2\times 45)+\cdots+(5\times 41)}{200}$$
$$= 2.96$$

So, she estimated the parameter of the Poisson distribution as $\hat{\lambda} = 2.96$.

With this value of $\hat{\lambda}$, she computed the Poisson probabilities for the different classes from the tables (which are also provided at the end of this block). These are shown in Column 3 of the table below.

Table 3 : Arrivals at ISBT

No. of arrivals	Observed frequencies O_i	Prob. according to Poisson dist. p_i	Expected frequencies $E_i(=np_i)$
0	10	0.0524	10.48
1	23	0.1545	30.90
2	45	0.2277	45.54
3	49	0.2238	44.76
4	32	0.1651	33.02
5 or more	41	0.1765	35.30
Total	200	1.0000	200.00

According to her data,

$$U = \frac{(10-10.48)^2}{10.48} + \frac{(23-30.90)^2}{30.9} + \ldots + \frac{(41-35.3)^2}{35.3}$$
$$= 0.022 + 2.02 + 0.006 + 0.402 + 0.032 + 0.92$$
$$= 3.402$$

Here k = 6 but one parameter has been estimated. So, the degrees of freedom associated with the chi-square distribution is $(k-1)-1 = k-2 = 4$. The critical value of chi-square at 4 degrees of freedom and 1 percent level of significance is 13.27. Since 3.402 < 13.27, the consultant did not reject the null hypothesis. In other words, she was in a position to conclude that the arrivals and departures at the bus terminus were Poisson distributed.

To find p_i for $\lambda = 2.96$, we take the average of the values in the columns corresponding to 2.9 and 3.0, respectively. Thus, $p_1 = \frac{0.055+0.0498}{2} = 0.0524$.

* * *

Let us now look at a problem involving normal distribution. While solving it, the following very important point about applying the χ^2-test will show up.

Remark 1 : If, corresponding to a category, say j, the expected value E_j is small, i.e., less than 5, then the chi-square approximation for the distribution of U will not be good. So, if the condition $E_i \geq 5$ is not satisfied for all i, then we should combine the category j with $E_j < 5$ with its adjacent categories $j+1, j+2, \ldots, j+r$, where $E_j + E_{j+1} + \ldots + E_{j+r} \geq 5$ but $E_j + E_{j+1} + \ldots + E_{j+r-1} < 5$. The number of classes, accordingly, gets reduced by r.

This remark will become more clear as you study the solution of Problem 1.

Problem 1 : A chemical company wants to know if its sales of a liquid chemical are normally distributed. This information will help them in planning and controlling the inventory. The sales record for a random sample of 200 days is given in Table 4.

We assume that the upper limit of a class shows that quantities less than that limit are in the class. So, for example, 35.5 will be included in the third class interval, not the second one.

Table 4

Sales (in 1000 litres)	Number of days
Less than 34.0	0
34.0-35.5	13
35.5-37.0	20
37.0-38.5	35
38.5-40.0	43
40.0-41.5	51
41.5-43.0	27
43.0-44.5	10
44.5-46.0	1
46.0 or more	0
Total	200

At the 5% level of significance, test the hypothesis that the company's sales are normally distributed.

Solution : Let us start by clearly stating our hypotheses.

H_0 : The company's sales are normally distributed.
against
H_A : The company's sales are **not** normally distributed.

Now, we assume for just now that H_0 is valid. By methods known to us, we can calculate the sample mean and sample standard deviation $\bar{x}$ and s_x. You can check that these are :

$\bar{x}$ = 40,000 litres, s_x= 2.5 thousand litres.

Now, we need to find the expected frequencies E_i corresponding to each O_i. You know that $E_i = 200 \times p_i$, where p_i is the probability for each class in Table 4, computed under the assumption of normal distribution.

So, let us expand Table 4 to include all the class probabilities (Column 3), the expected frequencies (Column 4) and the corresponding values of $\frac{(O_i - E_i)^2}{E_i}$ (Column 5).

To get the first entry in Column 3, we compute $z = \frac{(x-\mu)}{\sigma}$ for x = 34. As you know, μ and σ are estimated by $\bar{x}$ and s_x, respectively. So, $z = \frac{(34-40)}{2.5} = -2.4$.

Now, from the table of normal probabilities in the Block Appendix, you know that $P[-2.4 \le Z \le 0] = P[0 \le Z \le 2.4] = 0.4918$.
So, the probability we want is
$p_1 = 0.5 - P[-2.4 \le Z \le 0] = 0.5 - 0.4918 = 0.0082$.
Therefore, $E_1 = 200\ (0.0082) = 1.64$.

Similarly, you can compute the other expected frequencies and complete the 4th column of Table 5. You may wonder about the brackets in Columns 2, 3 and 4 of the table. This is because, as we have mentioned in Remark 1, **the χ^2 goodness-of-fit test is a good approximation only if the E_i are not very small.** This is why we have grouped the first two classes and the last two classes in Table 5.

To fill in the fifth column of Table 5, we treat the bracketed classes as a single class. So, $\frac{(O_1 - E_1)^2}{E_1} = \frac{(13-7.18)^2}{7.18} = 4.7176$. You can similarly calculate the other entries of Column 5 in the table below.

Table 5

Sales (in 1000 litres)	Observed frequency (O_i)	Class probability (p_i)	Expected frequency (E_i)	$\frac{(O_i - E_i)^2}{E_i}$
less than 34.0	0	0.0082	1.64	
34.0 – 35.5	13	0.0277	5.54 (combined: 7.18)	4.7176
35.5 – 37.0	20	0.0792	15.84	1.0925
37.0 – 38.5	35	0.1592	31.84	0.3136
38.5 – 40.0	43	0.2257	45.14	0.1015
40.0 – 41.5	51	0.2257	45.14	0.7607
41.5 – 43.0	27	0.1592	31.84	0.7357
43.0 – 44.5	10	0.0792	15.84	2.1531
44.5 – 46.0	1	0.0277	5.54	
greater than 46.0	0	0.0082	1.64 (combined: 7.18)	5.3193

Now, summing up the entries in the last column of Table 5, we get $U = 15.194$.

Next, to see whether we accept or reject H_0, we look up the value of χ^2 at the 5% level of significance and for the appropriate number of degrees of freedom. Note that, though we started with the data categorised into 10 classes, we needed to group two sets of 2 frequencies each. So, for purposes of the χ^2 test we now have 8 classes. Also, we have estimated two parameters, μ and σ. Therefore, the degrees of freedom are $(8 - 1) - 2 = 5$.

So, from the χ^2 table, we find $\chi^2_{0.05,5} = 11.07$.

Since $U > \chi^2_{0.05,5}$, we must reject H_0. That is, the normal distribution is not a good fit to the data.

* * *

Now try the following exercises.

E1) In Table 6 below you find the distribution of the heights for 100 college students. Estimate the mean and the standard deviation of the distribution. Check whether the sample is drawn from a normally distributed population at 5% level of significance.

Table 6

Class (cm)	Number of students (O_i)
Less than 161	4
161 - 164	11
164 - 167	16
167 - 170	19
170 - 173	25
173 - 176	18
176 - 179	4
179 - 182	2
182 or more	1
Total	100

E2) Test whether the observed frequencies, as given below, in 4 phenotypic classes AB, Ab, aB, ab are in agreement with the expected ratio 9: 3: 3: 1.

Class	AB	Ab	aB	ab
Frequency	102	25	28	5

E3) A die is rolled 1200 times with the following results :

No. that comes up	1	2	3	4	5	6
Frequency	205	279	217	257	133	109

Test if the die is unbiased.

In all the situations so far, the problem was related to data that were classified according to one attribute. Now let us see how the χ^2 test can be used to infer about situations in which the data are classified according to two or more attributes.

7.3 TEST OF INDEPENDENCE

Fig.3 :" Don't worry! You take this medicine, and you won't have any more colds in future."

In this section we shall look at inference problems like the following one.

Dr.Surya had recently developed a serum that she thought might be effective in preventing colds. But, she needed to verify its efficacy. For this purpose she carried out an experiment.

One thousand individuals were classified into two groups of the same size. The serum was administered to the members of the first group only. The number of individuals in each group who caught a cold zero times, or once, or more than once during some period after the treatment was noted. The data are shown in the following table having 2 rows and 3 columns.

Table 7 : Table showing the effect of serum

Category	The number catching a cold			Total
	Zero times	Once	More than once	
Group given serum	252	145	103	500
Untreated group	224	136	140	500
Total	476	281	243	1000

Dr.Surya's problem was to examine whether or not this serum is effective in preventing a cold. In other words, she wants to know if a person can catch a cold one or more times whether s/he has taken the serum or not. We can reword this as: is the treatment by the serum **independent** of the number of times of catching a cold?

So, Dr.Surya formulated the following null and alternative hypotheses to be tested:

H_0 : There is no interdependence between the serum treatment and the number of times of getting a cold.

H_A : H_0 is not true, i.e., the serum has some effect on preventing colds.

To test H_0 against H_A she planned to use the χ^2 test at the 5% significance level. As you know, to do so she needed to calculate the expected frequencies corresponding to each of the 6 entries in the 2×3 table, Table 7, assuming H_0, i..e., the independence of the number of times one gets a cold and of taking serum treatment.

Let us see how she obtained E_{11}. For this, she used the fact that out of the 1000 people, 476 had no cold. So, out of the 500 in the treatment group,

$\frac{476}{1000} \times 500 = 238$ were expected to not have any cold. Note that this is

$$\frac{\text{(sum of the first row entries)} \times \text{(sum of first column entries)}}{\text{(total sample size)}}$$

Similarly, she calculated the other expected frequencies :

$E_{12} = 500 \times \frac{281}{1000} = 140.5, E_{13} = 121.5,\ E_{21} = 238,\ E_{22} = 140.5,\ E_{23} = 121.5.$

Then , Surya calculated the sample statistic U as

$$U = \frac{(252-238)^2}{238} + \frac{(145-140.5)^2}{140.5} + \frac{(103-121.5)^2}{121.5} + \frac{(224-238)^2}{238} + \frac{(136-140.5)^2}{140.5} + \frac{(140-121.5)^2}{121.5} = 7.57$$

She took the significance level of the test as $\alpha = 0.05$. Also, in this case the number of degrees of freedom was $(2-1)(3-1) = 2$. So, comparing the value of U with $\chi^2_{0.05,2} = 5.99$, she found that $U > \chi^2_{0.05,2}$

For an $m \times n$ table, the number of degrees of freedom is $(m-1)(n-1)$.

So, she rejected H_0, and concluded that the serum has some effect in preventing colds.

Let us look closely at the steps Dr. Surya went through for testing the independence of two features of the population under study.

Step 1 : She stated the hypothesis regarding the independence of two features of the sample.

Step 2 : As in the case of the goodness-of-fit tests, she noted the frequencies — how many of each type of person (treated or untreated) had which kind of feature (the number of times they catch a cold). These frequencies were written in a table, called a **contingency table**.

In this case, the contingency table had 2 rows and 3 columns, because corresponding to each of the two groups of people there were 3 possibilities about the cold they did or did not catch. In brief, we say that the table was a 2×3 contingency table.

Step 3 : Corresponding to each of the 6 cells of the contingency table, Dr.Surya calculated the expected frequency. She did this as follows :
E_{ij} = expected frequency for ith row and jth column

$$= \frac{\text{(sum of entries of ith row)(sum of entries of jth column)}}{\text{(total sample size)}}$$

where i = 1, 2 and j = 1, 2, 3.

Step 4 : Then the sample χ^2, U, was calculated by

$$U = \sum_{i=1}^{2}\sum_{j=1}^{3}\frac{(O_{ij}-E_{ij})^2}{E_{ij}},$$ where O_{ij} was the entry in the ith row and jth column.

Note that, more generally, if she had had an $m \times n$ contingency table, the value would be

$$U = \sum_{i=1}^{m}\sum_{j=1}^{n}\frac{(O_{ij} - E_{ij})^2}{E_{ij}}$$, where m and n are natural numbers.

Step 5 : She compared this value with the value of $\chi^2_{\alpha,d}$, where α is the level of significance and d = (m − 1) (n− 1) is the number of degrees of freedom. Then, as you saw in Sec. 7.2, if $U < \chi^2_{\alpha,d}$, H_0 is accepted. Otherwise H_0 is rejected.

Another example may help you to clarify your understanding regarding the process of testing for independence.

Example 4 : The Glorious Watch Company wants to find out if there is any relationship between the income of a person and the importance she attaches to the price of a brand name. Mr.Zafar, the Chief of the Marketing Division, wants to test the hypothesis

H_0 : Income of a person and importance to her of price attached are independent.
against
H_A : H_0 is not true.

Zafar does a survey among the customers. To analyse his results, he groups them into 3 income levels, and asks them to mark the level of importance they give on a 3-point scale — great, moderate or low. He noted the results in a contingency table (see Table 8). In this table, you will also find the expected frequency corresponding to each observed frequency written alongside. As you know, these will be calculated as follows :

$$E_{11} = \frac{\text{(sum of entries of first row) (sum of entries of first column)}}{\text{(total sample size)}}$$

$$= \frac{170 \times 187}{500} = 63.58.$$

All the other E_{ij}s are calculated in the same way.

Table 8

Feature 1 (Importance Level)	Feature 2 (Income)						Total
	Low		Middle		High		
	O_{i1}	E_{i1}	O_{i2}	E_{i2}	O_{i3}	E_{i3}	
Great	79	63.58	58	61.2	33	45.22	170
Moderate	48	59.09	65	56.88	45	42.03	158
Low	60	64.33	57	61.92	55	45.75	172
Total	187		180		133		500

So, the sample χ^2 value that Zafar calculated was

$$U = \frac{(79-63.58)^2}{63.58} + \frac{(58-61.2)^2}{61.2} + \ldots + \frac{(57-61.92)^2}{61.92} + \frac{(55-45.75)^2}{45.75}$$

$$= 3.74 + 0.167 + 3.302 + 2.081 + 1.159 + 0.21 + 0.291 + 0.391 + 1.87$$

$$= 13.211$$

Then Zafar compared this with the value of χ^2 for (3−1) (3−1) = 4 degrees of freedom and at the 2% level of significance, which is $\chi^2_{0.02,4} = 11.668$.

He found $U > \chi^2_{0.02,4}$, which made him decide that he should reject H_0. In other words, Zafar is 98% certain that the level of income of a person is related to the importance she gives to the price of the brand of watches.

* * *

In the example above, it is interesting to note that if Zafar had chosen to be 99% certain, then $\chi^2_{0.01,4} = 13.277 > U$. So that, he would not have rejected H_0. What does this tell us about statistical analyses? Think about it.

And now here are some problems for you to solve.

E4) The data in the following table give mortality rates among vaccinated and non-vaccinated patients. Test if the vaccine has any effect in curing the disease.

Categories	Living	Dead	Total
Vaccinated	320	125	445
Non-vaccinated	98	230	328
Total	418	355	773

E5) Do the following data on sociability of soldiers recruited in cities and villages suggest that city soldiers are more sociable than village soldiers?

Place \ Sociability	Sociable	Non-sociable
City	13	6
Village	7	14

E6) A group of 1650 school children were classified according to their performance in school tests and family economic level. Test if there is any association between these two attributes.

Economic level \ Performance	Very Good	Good	Average	Poor	Total
Very Rich	4	7	16	25	52
Rich	13	37	79	73	202
Average	105	372	298	175	950
Poor	36	213	75	123	446
Total	157	629	468	396	1650

E7) In an experiment to study whether smoking affects health, the following data were collected. Test the hypothesis that smoking does not affect health.

	Light smoking	Moderate smoking	Heavy smoking
Health affected	16	29	35
Health not affected	36	23	17

In this section you have seen situations in which the population is tested to see if two or more common features of the population are related or not. This is as far as we intend to discuss the use of χ^2 for analysing categorical data. Let us end with a brief look at what we have covered in this unit.

7.4 SUMMARY

In this unit we have started with a look at data presented in the form of frequencies falling in different categories or classes. Based on such data we have undertaken different tests of hypotheses using the chi-squared distribution. We have considered two types of tests :

1) **Test of goodness-of-fit** : The hypotheses are given by
H_0 : The data fit a given distribution ; against
H_A : H_0 is not true, i.e., the data do not fit that distribution.

For testing whether H_0 is acceptable, we consider the observed and expected frequencies of the various categories in the data.

Suppose there are k categories with O_i as the observed frequency and E_i as the expected frequency of the ith category. Then the sample χ^2 value is

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

If H_0 were acceptable, then this value should be less than $\chi^2_{\alpha, k-s-1}$ with 100α % significance level, where s is the number of parameters estimated in finding the expected frequencies.

So, if $U < \chi^2_{\alpha, k-s-1}$, then H_0 is not rejected. Otherwise, H_0 is rejected.

2) **Test of independence** : Suppose a population can be classified into r categories on the basis of feature A, and into c categories on the basis of feature B. The hypotheses are given by :

H_0 : There is no interdependence between the features A and B
H_A : H_0 is not true, that is, A has an effect on B.

The data is presented in the form of an $r \times c$ contingency table. Let n_{ij} be the frequency in the ith row and jth column and let

$$n_{i0} = \sum_j n_{ij}, \quad n_{0j} = \sum_i n_{ij}, \quad n = \sum_i \sum_j n_{ij}$$

If the two classification criteria are mutually independent, the expected value E_{ij} for the ith row and jth column is given by

$$E_{ij} = \frac{n_{i0} \times n_{0j}}{n}$$

Then, the sample χ^2 value, $U = \sum_{i=1}^{r}\sum_{j=1}^{c} \frac{(O_{ij} - E_{ij})^2}{E_{ij}}$.

If this value is less than $\chi^2_{\alpha,(r-1)(c-1)}$, then H_0 is acceptable at the α level of significance.

And now you may like to check whether you have achieved the objectives of the unit listed in Sec.7.1. Also, while doing the exercises in this unit, you may have had some doubts. If so, please go through the following section also.

7.5 SOLUTIONS/ANSWERS

E1) Here $\bar{x} = 170$, $s^2 = 36$ and $n = 100$.

H_0 : The sample is drawn from a population with normal distribution $N(170.0, 6^2)$.
H_A : H_0 is not true.

In order to solve this problem by the same method as in Example 1, we consider the classes in Table 6 corresponding to categories of a multinomial distribution. Let O_i be the observed value for the ith category. Then, what is the expected value for the ith category in this case? Since the population distribution is completely specified as $N(170, 6^2)$ under the null hypothesis H_0, we can obtain the probability p_i with which the height of a student chosen randomly falls into the ith category. The expected value for the ith category is obtained by $E_i = np_i$. To compute the values p_i, the boundary points of the classes should be standardised by the population means and the standard deviation so as to make use of the table for a standard normal distribution. The standardised boundary points are given below in Table 11.

Table 11

Boundary Points of Class (x_i)	161	164	167	170	173	176	179	182
Standardised Boundary Points (z_i)	–1.5	–1.0	–0.5	0.0	0.5	1.0	1.5	2.0

Here, $z_i = \dfrac{x_i - 170.0}{6.0}$ The p_i and E_i values can be obtained as follows.

$p_1 = P[-\infty < Z < -1.5] = 1 - P[-\infty < Z < 1.5] = 0.0668$; $E_1 = 100 \times 0.0668 = 6.68$

$p_2 = P[-1.5 \leq Z < -1.0] = 0.0919$; $E_2 = 9.19$, and so on.

The values are all given in Table 12 below.

Table 12

Class (cm)	Number of students (O_i)	Probabilities (p_i)	Expected values E_i ($=np_i$)
Less than 161	4	0.0668	6.68
161 - 164	11	0.0919	9.19
164 - 167	16	0.1498	14.98
167 - 170	19	0.1915	19.15
170 - 173	25	0.1915	19.15
173 - 176	18	0.1498	14.98
176 - 179	4	0.0919	9.19
179 - 182	2	0.0440	4.40 } 6.68
182 or more	1	0.0228	2.28
Total	100	1.0000	100

Let us now test H_0 against H_A at the 5% significance level.

Now, from Table 12,

$$U = \frac{(4-6.68)^2}{6.68} + \frac{(11-9.19)^2}{9.19} + \frac{(16-14.98)^2}{14.98} + \frac{(19-19.15)^2}{19.15} + \frac{(25-19.15)^2}{19.15} + \frac{(18-14.98)^2}{14.98} + \frac{(4-9.19)^2}{9.19} + \frac{(3-6.68)^2}{6.68}$$
$$= 8.86$$

Taking the significance level of the test $\alpha = 0.05$, we have $\chi^2_{0.05,5} = 11.07$.

The degrees of freedom $8 - 3 = 5$, because the number of categories, after combining the last two categories is 8, and the number of parameters estimated is 2.

Since $U = 8.86 < 11.07 = \chi^2_{0.05,5}$, we conclude that there is good agreement between the observed frequencies and the fitted values. So H_0 is accepted.

E2) If the data are compatible with the given ratios, the expected frequencies are : AB: 90, Ab: 30, aB: 30, ab: 10.

The value of U is 5.07. This is less than $\chi^2_{0.05,3} = 7.815$. Hence, we accept the null hypothesis that the given data are in agreement with the expected ratios.

E3) Here H_0 : the expected frequency is 200 in each class.
H_A : H_0 is not true.

Therefore, $U = 112.87 > \chi^2_{0.05,5} = 11.070$. Hence, we conclude on the basis of the given data that we reject H_0. So the die is not unbiased.

E4) The hypothesis here is:

H_0 : There is no effect of the vaccine on mortality.
against
H_A : H_0 is not true.

The expected frequencies E_{ij} are given in the table below.

	Living	Dead	Total
Vaccinated	241	204	445
Non-Vaccinated	177	151	328
Total	418	355	773

The observed value of χ^2 is $U = 133.08$.
The number of degrees of freedom $= (2 - 1)(2 - 1) = 1$.
$\chi^2_{0.05,1} = 3.84 < U$.
Hence, we conclude that we cannot accept H_0. So, on the basis of the given data, we conclude that the vaccine has a definite effect on the mortality rate.

E5) H_0 : There is no interdependence between place and sociability level.
H_A : H_0 is not true.

The table of expected frequencies is

Place \ Sociability	Social	Non-social	Total
City	9.5	9.5	19
Village	10.5	10.5	21
Total	20	20	40

So, $U = 12.25\left(\frac{2}{9.5} + \frac{2}{10.5}\right) = 4.9.$

The number of degrees of freedom = 1.

$\chi^2_{0.05,\,1} = 3.84 < U.$

Therefore, we reject H_0. So, the data suggests that the place a soldier comes from affects her/his sociability level.

E6) The expected frequencies are given below :

Performance / Economic level	Very Good	Good	Average	Poor
Very Rich	4.95	19.82	14.75	12.48
Rich	9.22	77.00	57.29	48.48
Average	90.39	362.15	269.45	228
Poor	42.44	170.02	126.50	107.04

The value of the sample χ^2 is $U = 127.61 > 25.0 = \chi^2_{0.05,\,15}$. Hence, the hypothesis of independence between the categories is rejected.

E7) Under the assumption that smoking does not affect health, the expected frequencies are given below.

	Light smoking	Moderate smoking	Heavy smoking
Health affected	26.67	26.67	26.67
Health not affected	25.33	25.33	25.33

The observed value of χ^2 is $U = 14.52 > \chi^2_{0.05,\,2} = 5.991$. Hence, it is concluded on the basis of the given data that smoking affects health.

APPENDIX : MULTINOMIAL DISTRIBUTION

This distribution is an extension of the binomial distribution that you studied in Unit 3. It shows up in the following situation :

There is an experiment which consists of n identical trials, which are independent. Each trial can have k possible outcomes. Suppose the probability of each of these outcomes is $p_1, p_2, \ldots, p_k$, with $p_1 + p_2 \ldots + p_k = 1$. These probabilities remain the same from trial to trial.

Mathematically, this situation is represented by considering k random variables $X_1, \ldots, X_k$ with probabilities $p_1, \ldots, p_k$ that $X_1 = x_1, \ldots, X_k = x_k$, respectively, where $\sum_{i=1}^{k} x_i = n$, $\sum_{i=1}^{k} p_i = 1$, $p_i \neq 0 \,\forall i = 1, \cdots, k$. If the random vector $(X_1, \ldots, X_k)$ is multinomially distributed, the $P[X_1 = x_1, \cdots, X_k = x_k] = \frac{n!}{x_1! \cdots x_k!} p_1^{x_1} \cdots p_k^{x_k}$.

So, if we are testing if a certain population is multinomially distributed, we will test

H_0 : The population of size n is multinomially distributed with probabilities $p_1, p_2, \ldots, p_k$ (known to us);

against

H_A : The population is not multinomially distributed.

As in all the other cases discussed in the unit, if the np_i are not very small, then the test statistic $U = \sum_{i=1}^{k} \frac{(O_i - np_i)^2}{np_i}$ has approximately a chi-squared distribution with $(k-1)$ degrees of freedom. The approximation is usually good for $E_i = np_i \geq 5$.

APPENDIX – 1

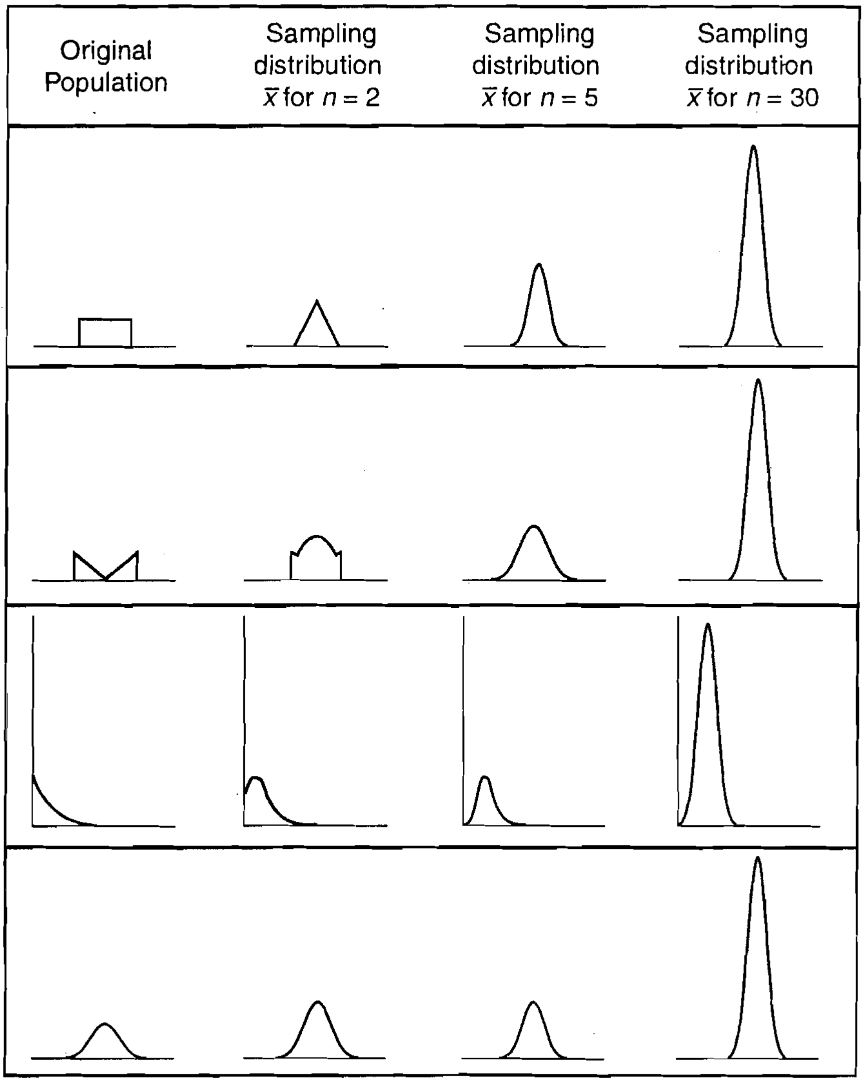


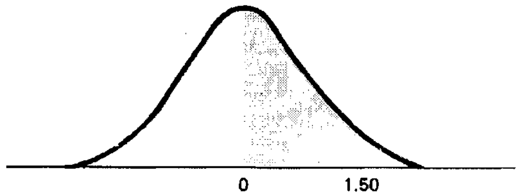
Fig. 1 Sampling distribution of $\bar{x}$ for different populations and different sample sizes.

APPENDIX-2

TABLE 1

AREAS UNDER THE STANDARD NORMAL CURVE

This table shows the area between zero (the mean of a standard normal variable) and z. For example, if $z = 1.50$, this is the shaded area shown below which equals .4332.

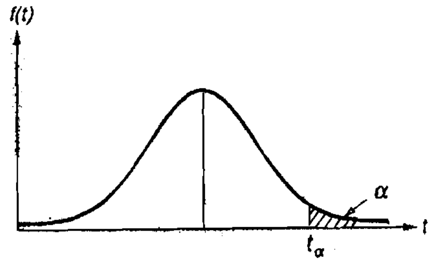


z	.00	.01	.02	.03	.04	.05	.06	.07	.08	.09
0.0	.0000	.0040	.0080	.0120	.0160	.0199	.0239	.0279	.0319	.0359
0.1	.0398	.0438	.0478	.0517	.0557	.0596	.0636	.0675	.0714	.0753
0.2	.0793	.0832	.0871	.0910	.0948	.0987	.1026	.1064	.1103	.1141
0.3	.1179	.1217	.1255	.1293	.1331	.1368	.1406	.1443	.1480	.1517
0.4	.1554	.1591	.1628	.1664	.1700	.1736	.1772	.1808	.1844	.1879
0.5	.1915	.1950	.1985	.2019	.2054	.2088	.2123	.2157	.2190	.2224
0.6	.2257	.2291	.2324	.2357	.2389	.2422	.2454	.2486	.2517	.2549
0.7	.2580	.2611	.2642	.2673	.2704	.2734	.2764	.2794	.2823	.2852
0.8	.2881	.2910	.2939	.2967	.2995	.3023	.3051	.3078	.3106	.3133
0.9	.3159	.3186	.3212	.3238	.3264	.3289	.3315	.3340	.3365	.3389
1.0	.3413	.3438	.3461	.3485	.3508	.3531	.3554	.3577	.3599	.3621
1.1	.3643	.3665	.3686	.3708	.3729	.3749	.3770	.3790	.3810	.3830
1.2	.3849	.3869	.3888	.3907	.3925	.3944	.3962	.3980	.3997	.4015
1.3	.4032	.4049	.4066	.4082	.4099	.4115	.4131	.4147	.4162	.4177
1.4	.4192	.4207	.4222	.4236	.4251	.4265	.4279	.4292	.4306	.4319
1.5	.4332	.4345	.4357	.4370	.4382	.4394	.4406	.4418	.4429	.4441
1.6	.4452	.4463	.4474	.4484	.4495	.4505	.4515	.4525	.4535	.4545
1.7	.4554	.4564	.4573	.4582	.4591	.4599	.4608	.4616	.4625	.4633
1.8	.4641	.4649	.4656	.4664	.4671	.4678	.4686	.4693	.4699	.4706
1.9	.4713	.4719	.4726	.4732	.4738	.4744	.4750	.4756	.4761	.4767
2.0	.4772	.4778	.4783	.4788	.4793	.4798	.4803	.4808	.4812	.4817
2.1	.4821	.4826	.4830	.4834	.4838	.4842	.4846	.4850	.4854	.4857
2.2	.4861	.4864	.4868	.4871	.4875	.4878	.4881	.4884	.4887	.4890
2.3	.4893	.4896	.4898	.4901	.4904	.4906	.4909	.4911	.4913	.4916
2.4	.4918	.4920	.4922	.4925	.4927	.4929	.4931	.4932	.4934	.4936
2.5	.4938	.4940	.4941	.4943	.4945	.4946	.4948	.4949	.4951	.4952
2.6	.4953	.4955	.4956	.4957	.4959	.4960	.4961	.4962	.4963	.4964
2.7	.4965	.4966	.4967	.4968	.4969	.4970	.4971	.4972	.4973	.4974
2.8	.4974	.4975	.4976	.4977	.4977	.4978	.4979	.4979	.4980	.4981
2.9	.4981	.4982	.4982	.4983	.4984	.4984	.4985	.4985	.4986	.4986
3.0	.4987	.4987	.4987	.4988	.4988	.4989	.4989	.4989	.4990	.4990

Source: This table is adapted from National Bureau of Standards, *Tables of Normal Probability Functions*, Applied Mathematics Series 23, U.S. Department of Commerce, 1953.

TABLE-2

t-distribution



ν	$t_{.100}$	$t_{.050}$	$t_{.025}$	$t_{.010}$	$t_{.005}$	$t_{.001}$	$t_{.0005}$
1	3.078	6.314	12.706	31.821	63.657	318.31	636.62
2	1.886	2.920	4.303	6.965	9.925	22.326	31.598
3	1.638	2.353	3.182	4.541	5.841	10.213	12.924
4	1.533	2.132	2.776	3.747	4.604	7.173	8.610
5	1.476	2.015	2.571	3.365	4.032	5.893	6.869
6	1.440	1.943	2.447	3.143	3.707	5.208	5.959
7	1.415	1.895	2.365	2.998	3.499	4.785	5.408
8	1.397	1.860	2.306	2.896	3.355	4.501	5.041
9	1.383	1.833	2.262	2.821	3.250	4.297	4.781
10	1.372	1.812	2.228	2.764	3.169	4.144	4.587
11	1.363	1.796	2.201	2.718	3.106	4.025	4.437
12	1.356	1.782	2.179	2.681	3.055	3.930	4.318
13	1.350	1.771	2.160	2.650	3.012	3.852	4.221
14	1.345	1.761	2.145	2.624	2.977	3.787	4.140
15	1.341	1.753	2.131	2.602	2.947	3.733	4.073
16	1.337	1.746	2.120	2.583	2.921	3.686	4.015
17	1.333	1.740	2.110	2.567	2.898	3.646	3.965
18	1.330	1.734	2.101	2.552	2.878	3.610	3.922
19	1.328	1.729	2.093	2.539	2.861	3.579	3.883
20	1.325	1.725	2.086	2.528	2.845	3.552	3.850
21	1.323	1.721	2.080	2.518	2.831	3.527	3.819
22	1.321	1.717	2.074	2.508	2.819	3.505	3.792
23	1.319	1.714	2.069	2.500	2.807	3.485	3.767
24	1.318	1.711	2.064	2.492	2.797	3.467	3.745
25	1.316	1.708	2.060	2.485	2.787	3.450	3.725
26	1.315	1.706	2.056	2.479	2.779	3.435	3.707
27	1.314	1.703	2.052	2.473	2.771	3.421	3.690
28	1.313	1.701	2.048	2.467	2.763	3.408	3.674
29	1.311	1.699	2.045	2.462	2.756	3.396	3.659
30	1.310	1.697	2.042	2.457	2.750	3.385	3.646
40	1.303	1.684	2.021	2.423	2.704	3.307	3.551
60	1.296	1.671	2.000	2.390	2.660	3.232	3.460
120	1.289	1.658	1.980	2.358	2.617	3.160	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.090	3.291

TABLE-3
CHI-SQUARED DISTRIBUTION

ν	$\alpha = .995$	$\alpha = .990$	$\alpha = .975$	$\alpha = .950$	$\alpha = .900$	$\alpha = .750$	$\alpha = .500$	$\alpha = .250$	$\alpha = .100$	$\alpha = .050$	$\alpha = .025$	$\alpha = .010$	$\alpha = .005$
1	0.0000393	0.000157	0.000982	0.00393	0.01579	0.10153	0.45494	1.32330	2.70554	3.84146	5.02389	6.63490	7.87944
2	0.01003	0.02010	0.05064	0.10259	0.21072	0.57536	1.38629	2.77259	4.60517	5.99146	7.37776	9.21034	10.59663
3	0.07172	0.11483	0.21580	0.35185	0.58437	1.21253	2.36597	4.10834	6.25139	7.81473	9.34840	11.34487	12.83816
4	0.20699	0.29711	0.48442	0.71072	1.06362	1.92256	3.35669	5.38527	7.77944	9.48773	11.14329	13.27670	14.86026
5	0.41174	0.55430	0.83121	1.14548	1.61031	2.67460	4.35146	6.62568	9.23636	11.07050	12.83250	15.08627	16.74960
6	0.67573	0.87209	1.23734	1.63538	2.20413	3.45460	5.34812	7.84080	10.64464	12.59159	14.44938	16.81189	18.54758
7	0.98926	1.23904	1.68987	2.16735	2.83311	4.25485	6.34581	9.03715	12.01704	14.06714	16.01276	18.47531	20.27774
8	1.34441	1.64650	2.17973	2.73264	3.48954	5.07064	7.34412	10.21885	13.36157	15.50731	17.53455	20.09024	21.95495
9	1.73493	2.08790	2.70039	3.32511	4.16816	5.89883	8.34283	11.38875	14.68366	16.91898	19.02277	21.66599	23.58935
10	2.15586	2.55821	3.24697	3.94030	4.86518	6.73720	9.34182	12.54886	15.98718	18.30704	20.48318	23.20925	25.18818
11	2.60322	3.05348	3.81575	4.57481	5.57778	7.58414	10.34100	13.70069	17.27501	19.67514	21.92005	24.72497	26.75685
12	3.07382	3.57057	4.40379	5.22603	6.30380	8.43842	11.34032	14.84540	18.54935	21.02607	23.33666	26.21697	28.29952
13	3.56503	4.10692	5.00875	5.89186	7.04150	9.29907	12.33976	15.98391	19.81193	22.36203	24.73560	27.68825	29.81947
14	4.07467	4.66043	5.62873	6.57063	7.78953	10.16531	13.33927	17.11693	21.06414	23.68479	26.11895	29.14124	31.31935
15	4.60092	5.22935	6.26214	7.26094	8.54676	11.03654	14.33886	18.24509	22.30713	24.99579	27.48839	30.57791	32.80132
16	5.14221	5.81221	6.90766	7.96165	9.31224	11.91222	15.33850	19.36886	23.54183	26.29623	28.84535	31.99993	34.26719
17	5.69722	6.40776	7.56419	8.67176	10.08519	12.79193	16.33818	20.48868	24.76904	27.58711	30.19101	33.40866	35.71847
18	6.26480	7.01491	8.23075	9.39046	10.86494	13.67529	17.33790	21.60489	25.98942	28.86930	31.52638	34.80531	37.15645
19	6.84397	7.63273	8.90652	10.11701	11.65091	14.56200	18.33765	22.71781	27.20357	30.14353	32.85233	36.19087	38.58226
20	7.43384	8.26040	9.59078	10.85081	12.44261	15.45177	19.33743	23.82769	28.41198	31.41043	34.16961	37.56623	39.99685
21	8.03365	8.89720	10.28290	11.59131	13.23960	16.34438	20.33723	24.93478	29.61509	32.67057	35.47888	38.93217	41.40106
22	8.64272	9.54249	10.98232	12.33801	14.04149	17.23962	21.33704	26.03927	30.81328	33.92444	36.78071	40.28936	42.79565
23	9.26042	10.19572	11.68855	13.09051	14.84796	18.13730	22.33688	27.14134	32.00690	35.17246	38.07563	41.63840	44.18128
24	9.88623	10.85636	12.40115	13.84843	15.65868	19.03725	23.33673	28.24115	33.19624	36.41503	39.36408	42.97982	45.55851
25	10.51965	11.52398	13.11972	14.61141	16.47341	19.93934	24.33659	29.33885	34.38159	37.65248	40.64647	44.31410	46.92789
26	11.16024	12.19815	13.84390	15.37916	17.29188	20.84343	25.33646	30.43457	35.56317	38.88514	41.92317	45.64168	48.28988
27	11.80759	12.87850	14.57338	16.15140	18.11390	21.74940	26.33634	31.52841	36.74122	40.11327	43.19451	46.96294	49.64492
28	12.46134	13.56471	15.30786	16.92788	18.93924	22.65716	27.33623	32.62049	37.91592	41.33714	44.46079	48.27824	50.99338
29	13.12115	14.25645	16.04707	17.70837	19.76774	23.56659	28.33613	33.71091	39.08747	42.55697	45.72229	49.58788	52.33562
30	13.78672	14.95346	16.79077	18.49266	20.59923	24.47761	29.33603	34.79974	40.25602	43.77297	46.97924	50.89218	53.67196

TABLE-4
F-distribution

$F_{0.05}$

ν_2 = Degrees of freedom for denominator	ν_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	161	200	216	225	230	234	237	239	241	242	244	246	248	249	250	251	252	253	254
2	18.50	19.00	19.20	19.20	19.30	19.30	19.40	19.40	19.40	19.40	19.40	19.40	19.40	19.50	19.50	19.50	19.50	19.50	19.50
3	10.10	9.55	9.28	9.12	9.01	8.94	8.89	8.85	8.81	8.79	8.74	8.70	8.66	8.64	8.62	8.59	8.57	8.55	8.53
4	7.71	6.94	6.59	6.39	6.26	6.16	6.09	6.04	6.00	5.96	5.91	5.86	5.80	5.77	5.75	5.72	5.69	5.66	5.63
5	6.61	5.79	5.41	5.19	5.05	4.95	4.88	4.82	4.77	4.74	4.68	4.62	4.56	4.53	4.50	4.46	4.43	4.40	4.37
6	5.99	5.14	4.76	4.53	4.39	4.28	4.21	4.15	4.10	4.06	4.00	3.94	3.87	3.84	3.81	3.77	3.74	3.70	3.67
7	5.59	4.74	4.35	4.12	3.97	3.87	3.79	3.73	3.68	3.64	3.57	3.51	3.44	3.41	3.38	3.34	3.30	3.27	3.23
8	5.32	4.46	4.07	3.84	3.69	3.58	3.50	3.44	3.39	3.35	3.28	3.22	3.15	3.12	3.08	3.04	3.01	2.97	2.93
9	5.12	4.26	3.86	3.63	3.48	3.37	3.29	3.23	3.18	3.14	3.07	3.01	2.94	2.90	2.86	2.83	2.79	2.75	2.71
10	4.96	4.10	3.71	3.48	3.33	3.22	3.14	3.07	3.02	2.98	2.91	2.85	2.77	2.74	2.70	2.66	2.62	2.58	2.54
11	4.84	3.98	3.59	3.36	3.20	3.09	3.01	2.95	2.90	2.85	2.79	2.72	2.65	2.61	2.57	2.53	2.49	2.45	2.40
12	4.75	3.89	3.49	3.26	3.11	3.00	2.91	2.85	2.80	2.75	2.69	2.62	2.54	2.51	2.47	2.38	2.38	2.30	2.30
13	4.67	3.81	3.41	3.18	3.03	2.92	2.83	2.77	2.71	2.67	2.60	2.53	2.46	2.42	2.38	2.34	2.30	2.25	2.21
14	4.60	3.74	3.34	3.11	2.96	2.85	2.76	2.70	2.65	2.60	2.53	2.46	2.39	2.35	2.31	2.27	2.22	2.18	2.13
15	4.54	3.68	3.29	3.06	2.90	2.79	2.71	2.64	2.59	2.54	2.48	2.40	2.33	2.29	2.25	2.20	2.16	2.11	2.07
16	4.49	3.63	3.24	3.01	2.85	2.74	2.66	2.59	2.54	2.49	2.42	2.35	2.28	2.24	2.19	2.15	2.11	2.06	2.01
17	3.45	3.59	3.20	2.96	2.81	2.70	2.61	2.55	2.49	2.45	2.38	2.31	2.23	2.19	2.15	2.10	2.06	2.01	1.96
18	4.41	3.55	3.16	2.93	2.77	2.66	2.58	2.51	2.46	2.41	2.34	2.27	2.19	2.15	2.11	2.06	2.02	1.97	1.93
19	4.38	3.52	3.13	2.90	2.74	2.63	2.54	2.48	2.42	2.38	2.31	2.23	2.16	2.11	2.07	2.03	1.98	1.93	1.88
20	4.35	3.49	3.10	2.87	2.71	2.60	2.51	2.45	2.39	2.35	2.28	2.20	2.12	2.08	2.04	1.99	1.95	1.90	1.84
21	4.32	3.47	3.07	2.84	2.68	2.57	2.49	2.42	2.37	2.32	2.25	2.18	2.10	2.05	2.01	1.96	1.92	1.87	1.81
22	4.30	3.44	3.05	2.82	2.66	2.55	2.46	2.40	2.34	2.30	2.23	2.15	2.07	2.03	1.98	1.94	1.89	1.84	1.78
23	4.28	3.42	3.03	2.80	2.64	2.53	2.44	2.37	2.32	2.27	2.20	2.13	2.05	2.01	1.96	1.91	1.86	1.81	1.76
24	4.26	3.40	3.01	2.78	2.62	2.51	2.42	2.36	2.30	2.25	2.18	2.11	2.03	1.98	1.94	1.89	1.84	1.79	1.73
25	4.24	3.39	2.99	2.76	2.60	2.49	2.40	2.34	2.28	2.24	2.16	2.09	2.01	1.96	1.92	1.87	1.82	1.77	1.71
30	4.17	3.32	2.92	2.69	2.53	2.42	2.33	2.27	2.21	2.16	2.09	2.01	1.93	1.89	1.84	1.79	1.74	1.68	1.62
40	4.08	3.23	2.84	2.61	2.45	2.34	2.25	2.18	2.12	2.08	2.00	1.92	1.84	1.79	1.74	1.69	1.64	1.58	1.51
60	4.00	3.15	2.76	2.53	2.37	2.25	2.17	2.10	2.04	1.99	1.92	1.84	1.75	1.70	1.65	1.59	1.53	1.47	1.39
120	3.92	3.07	2.68	2.45	2.29	2.18	2.09	2.02	1.96	1.91	1.83	1.75	1.66	1.61	1.55	1.50	1.43	1.35	1.25
∞	3.84	3.00	2.60	2.37	2.21	2.10	2.01	1.94	1.88	1.83	1.75	1.67	1.57	1.52	1.46	1.39	1.32	1.22	1.00

TABLE-5
F-distribution

$F_{0.01}$

ν_2 = Degrees of freedom for denominator	ν_1 = Degrees of freedom for numerator																		
	1	2	3	4	5	6	7	8	9	10	12	15	20	24	30	40	60	120	∞
1	4,052	5,000	5,403	5,625	5,764	5,859	5,928	5,982	6,023	6,056	6,106	6,157	6,209	6,235	6,261	6,287	6,313	6,339	6,366
2	98.50	99.00	99.20	99.20	99.30	99.30	99.40	99.40	99.40	99.40	99.40	99.40	99.40	99.50	99.50	99.50	99.50	99.50	99.50
3	34.10	30.80	29.50	28.70	28.20	27.90	27.70	27.50	27.30	27.20	27.10	26.90	26.70	26.60	26.50	26.40	26.30	26.20	26.10
4	21.20	18.00	16.70	16.00	15.50	15.20	15.00	14.80	14.70	14.50	14.40	14.20	14.00	13.90	13.80	13.70	13.70	13.60	13.50
5	16.30	13.30	12.10	11.40	11.00	10.70	10.50	10.30	10.20	10.10	9.89	9.72	9.55	9.47	9.38	9.29	9.20	9.11	9.02
6	13.70	10.90	9.78	9.15	8.75	8.47	8.26	8.10	7.98	7.87	7.72	7.56	7.40	7.31	7.23	7.14	7.06	6.97	6.88
7	12.20	9.55	8.45	7.85	7.46	7.19	6.99	6.84	6.72	6.62	6.47	6.31	6.16	6.07	5.99	5.91	5.82	5.74	5.65
8	11.30	8.65	7.59	7.01	6.63	6.37	6.18	6.03	5.91	5.81	5.67	5.52	5.36	5.28	5.20	5.12	5.03	4.95	4.83
9	10.60	8.02	6.99	6.42	6.06	5.80	5.61	5.47	5.35	5.26	5.11	4.96	4.81	4.73	4.65	4.57	4.48	4.40	4.31
10	10.00	7.56	6.55	5.99	5.64	5.39	5.20	5.06	4.94	4.85	4.71	4.56	4.41	4.33	4.25	4.17	4.08	4.00	3.91
11	9.65	7.21	6.22	5.67	5.32	5.07	4.89	4.74	4.63	4.54	4.40	4.25	4.10	4.02	3.94	3.86	3.78	3.69	3.60
12	9.33	6.93	5.95	5.41	5.06	4.82	4.64	4.50	4.39	4.30	4.16	4.01	3.86	3.78	3.70	3.62	3.54	3.45	3.36
13	9.07	6.70	5.74	5.21	4.86	4.62	4.44	4.30	4.19	4.10	3.96	3.82	3.66	3.59	3.51	3.43	3.34	3.25	3.17
14	8.86	6.51	5.56	5.04	4.70	4.46	4.28	4.14	4.03	3.94	3.80	3.66	3.51	3.43	3.35	3.27	3.18	3.09	3.00
15	8.68	6.36	5.42	4.89	4.56	4.32	4.14	4.00	3.89	3.80	3.67	3.52	3.37	3.29	3.21	3.13	3.96	2.96	2.87
16	8.53	6.23	5.29	4.77	4.44	4.20	4.03	3.89	3.78	3.69	3.55	3.41	3.26	3.18	3.10	3.02	2.93	2.84	2.75
17	8.40	6.11	5.19	4.67	4.34	4.10	3.93	3.79	3.68	3.59	3.46	3.31	3.16	3.08	3.00	2.92	2.83	2.75	2.65
18	8.29	6.01	5.09	4.58	4.25	4.01	3.84	3.71	3.60	3.51	3.37	3.23	3.08	3.00	2.92	2.84	2.75	2.66	2.57
19	8.19	5.93	5.01	4.50	4.17	3.94	3.77	3.63	3.52	3.43	3.30	3.15	3.00	2.92	2.84	2.76	2.67	2.58	2.49
20	8.10	5.85	4.94	4.43	4.10	3.87	3.70	3.56	3.46	3.37	3.23	3.09	2.94	2.86	2.78	2.69	2.61	2.52	2.42
21	8.02	5.78	4.87	4.37	4.04	3.81	3.64	3.51	3.40	3.31	3.17	3.03	2.88	2.80	2.72	2.64	2.55	2.46	2.36
22	7.95	5.72	4.82	4.31	3.99	3.76	3.59	3.45	3.35	3.26	3.12	2.98	2.83	2.75	2.67	2.58	2.50	2.40	2.31
23	7.88	5.66	4.76	4.26	3.94	3.71	3.54	3.41	3.30	3.21	3.07	2.93	2.78	2.70	2.62	2.54	2.45	2.35	2.26
24	7.82	5.61	4.72	4.22	3.90	3.67	3.50	3.36	3.26	3.17	3.03	2.89	2.74	2.66	2.58	2.49	2.40	2.31	2.21
25	7.77	5.57	4.68	4.18	3.86	3.63	3.46	3.32	3.22	3.13	2.99	2.85	2.70	2.62	2.53	2.45	2.36	2.27	2.17
30	7.56	5.39	4.51	4.02	3.70	3.47	3.30	3.17	3.07	2.98	2.84	2.70	2.55	2.47	2.39	2.30	2.21	2.11	2.01
40	7.31	5.18	4.31	3.83	3.51	3.29	3.12	2.99	2.89	2.80	2.66	2.52	2.37	2.29	2.20	2.11	2.02	1.92	1.80
60	7.08	4.98	4.13	3.65	3.34	3.12	2.95	2.82	2.72	2.63	2.50	2.35	2.20	2.12	2.03	1.94	1.84	1.73	1.60
120	6.85	4.79	3.95	3.48	3.17	2.96	2.79	2.66	2.56	2.47	2.34	2.19	2.03	1.95	1.86	1.76	1.66	1.53	1.38
∞	6.63	4.61	3.78	3.32	3.02	2.80	2.64	2.51	2.41	2.32	2.18	2.04	1.88	1.79	1.70	1.59	1.47	1.32	1.00