**CHAPTER FOUR**

**CHI – SQUARE (Χ2) DISTRIBUTION**

**Objectives:**

After completing this chapter, you should be able to:

* Understand the nature of a Chi - Square distribution.
* Identify the areas of applications for a Chi - Square distribution.
* Test the equality of more than two population proportions.
* Conduct goodness-of-fit tests for the three distributions (normal, binomial, and poisson).
* Test the independence between two variables.
  1. **Introduction**

In the previous chapter – Hypothesis Testing – we have attempted to determine whether a hypothesis that assumes some value for a population parameter is reasonable or not and whether to accept or reject the assumption. In this chapter, while our concern is still testing a hypothesis made about a population, with information gathered from a sample, our focus will be only on Chi - Square test – a right tailed test. A right tailed test will reject the null hypothesis if the sample statistic is significantly higher than the hypothesized population parameter.

**4.2 The Chi - Square Distribution**

The mathematical expression for the Chi- Square distribution contains only one parameter, the number of degrees of freedom, ν. There is a particular chi - square distribution for each particular number of degrees of freedom. For example, the distribution of Z2 (the square of the standard normal variable) is the chi - square distribution with ν = 1 degree of freedom. The random variable of the distributions is denoted by χ2 (χ is the Greek letter Chi – which is read as “ki”). The χ2 distribution is a continuous distribution and since it is a probability distribution, the total area under each curve is 1.

The variable χ2 cannot be negative; so chi - square curves do not extend to the left of zero having a positive skew-ness as they extend indefinitely in the positive direction. When ν exceeds 2, χ2 curves have one mode but as ν increases the skew-ness becomes less apparent and in fact, when ν is very large, the chi - square distribution is almost the same as a normal distribution having a mean equal to ν and a standard deviation equal to 

but in applications, ν is not large enough to permit us to use normal curve probabilities as approximations of χ2 probabilities. The significance levels α in this chapter will be right - tail areas of chi - square distributions. The symbol χ2 α, ν means the value of chi - square such that the distribution with ν degrees of freedom has a right tail area of α.

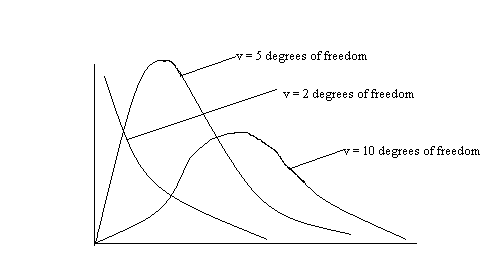


Figure 4-1 Chi-square distribution

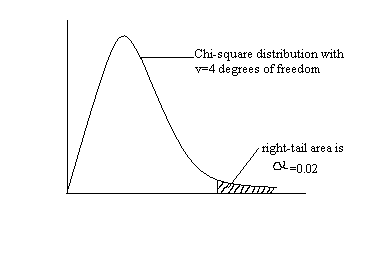


Figure 4-2 The meaning of χ2 α, ν

Example: The value of χ2 0.02, 4 = 11.668

The values of χ2 α, ν are given in χ2 table.

Exercise: What is the value of χ2 0.05, 3?

Answer: 7.815

**4.3 Areas of Application**

The Chi - square distribution is used in number of statistical tests including:

* Test for independence between two variables
* Goodness-of-fit tests
* Test for equality of several proportions
  + 1. **The χ2 test for independence**

This involves a contingency table analysis. A contingency table is one that consists of count data (data obtained by counting the frequency of an occurrence given a certain category from a simple random sample) arranged in r – rows and c – columns. The actual sample counts are called observed frequencies, and are denoted by fo.

The expected and observed frequencies, fe and fo are used to compute a sample statistic for testing the hypothesis that the row and column categories are independent. The underlying idea is that the observed frequencies should be close to the frequencies that would be expected if the categories are independent. Large differences will lead us to reject the hypothesis of independence. The statistic that is used for the test is called the sample χ2. It is computed as:

**Sample χ2 =** 

and the degree of freedom, ν (read as ‘nu’) is: **ν = (r – 1) (c – 1)**

The formula shows that the larger the squared differences are relative to their respective expected frequencies, the larger will be the value of the sample χ2. Therefore, large values of the sample χ2 will lead to rejecting the independence hypothesis.

The expected frequencies corresponding to a single observed frequency must not be too small i.e. fe ≥ 5. If fe < 5 then we shall combine adjacent rows (or columns) in the contingency table to get fe values of at least 5 before computing the sample χ2, also ν will be computed after combining rows or columns (if any combining exists).

The steps to be followed in the contingency table test are illustrated along the following example.

**Example:** The manager of Wabi-Shebele Hotel has collected opinions on the quality of their service from a random sample of customers. The customers visited the Hotel’s branches in three regions, namely Addis Ababa, Nazareth and Awasa – Langanoo, and rated the services on a scale of 1 (best) to 4. The sample data is given in Table 7.1 below. The manager wants to know whether quality ratings are or are not independent of the respective regions where the Hotel’s branches are found. Perform test for independence at 5% level.

**Table 4.1**

**Customer quality ratings of service by region**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Quality**  **Rating** | **Hotel's Branch** | | | **Row**  **Total** |
| **Addis Ababa** | **Nazareth** | **Langano** |
| 1  2  3  4 | 15  7  11  3 | 10  13  12  8 | 6  12  8  15 | 31  32  31  26 |
| Column Total | 36 | 43 | 41 | 120 - **Grand total** |

**Solution:**

###### Steps

**1. State the hypotheses** *Ho: Quality rating is independent of Hotel's Branch*

*Ha: Quality rating is not independent of Hotel's Branch*

**2. State the decision** *Table 7.1 has r = 4 rows and c = 3 columns; it is a 4 by 3*

**rule** *contingency table. The number of degrees of freedom is:*

*ν = (r - 1) (c - 1) = (4 - 1) (3 - 1)*

*= 6*

*with significance level α = 0.05, we find from the Table X*

*in the appendix that χ20.05 , 6 = 12.592*

**3. Compute the** *Sample χ2 =* 

**Sample χ2**

*fe = (row total) x (column total)*

*grand total*

*for the cell in the first row, first column,*

*fe = 31 x 36 = 9.3*

*120*

*for the cell in the first row, 2nd column, and 3rd column,*

*fe = 31 x 43 = 11.1 fe = 31 x 41 = 10.6*

*120 120*

*or for the last i.e. 3rd column we can find fe by adding the previous expected frequencies and subtracting it from the row total*

*i.e. fe = 31 - (9.3 + 11.1) = 10.6*

Table 4.2 shows the observed and expected frequencies for each cell and from these frequencies we compute the value of the test statistic – sample χ2 – as shown in Table 7.3.

**Table 4.2**

**Service quality ratings observed and expected frequencies**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| **Quality**  **Ratings** | **Hotel’s Branch** | | | | | |
| **Addis Ababa** | | **Nazareth** | | **Langano** | |
| **fo** | **fe** | **fo** | **fe** | **fo** | **fe** |
| 1 | 15 | 9.3 | 10 | 11.1 | 6 | 10.6 |
| 2 | 7 | 9.6 | 13 | 11.5 | 12 | 10.9 |
| 3 | 11 | 9.3 | 12 | 11.1 | 8 | 10.6 |
| 4 | 3 | 7.8 | 8 | 9.3 | 15 | 8.9 |

**Table 4.3**

**Calculating the sample χ2 value**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| ***fo*** | ***fe*** | ***fo - fe*** | ***(fo - fe)2*** |  |
| **15**  **10**  **6**  **7**  **13**  **12**  **11**  **12**  **8**  **3**  **8**  **15** | 9.3  11.1  10.6  9.6  11.5  10.9  9.3  11.1  10.6  7.8  9.3  8.9 | 5.7  -1.1  -4.6  -2.6  1.5  1.1  1.7  0.9  -2.6  -4.8  -1.3  6.1 | 32.49  1.21  21.16  6.76  2.25  1.21  2.89  0.81  6.76  23.04  1.69  37.21 | 3.4935  0.1090  1.9962  0.7042  0.1957  0.1110  0.3108  0.0730  0.6377  2.9538  0.1817  4.1809  Sample χ2 = 14.9475 |

*Sample χ2 = 14.9475*

**4. Accept or reject Ho** *Because the value of the sample test statistic,*

***14.9475****, exceeds the value* ***12.592*** *in the decision rule, we reject Ho.*

The test result means that customer quality rating of services is not independent of the Hotel’s branch where the service took place. The sample provides convincing evidence that service quality rating depends on the Hotel’s Branch.

* + 1. **Testing the Equality of more-than-two Population Proportions**

A test for equality of two population proportions was described in previous chapter . The hypotheses there were: Ho: P1 = P2; and

Ha: P1 ≠ P2

The hypotheses of the test for equal proportions can be stated either in terms of equal row proportions or in terms of equal column proportions. We will use row proportions Pr (that is, cell proportions to column totals) in our hypotheses. Thus, the hypotheses will be: Ho: The proportions Pr in any row are equal.

Ha: The Pr in at least one row is not equal.

**Example:** Table 4.5 contains counts for a random sample of n = 200 workers. We want to test the hypothesis that the population proportions of satisfactory workers in education levels 1, 2, and 3, are equal at a 5% significance level.

### Table 4.5

Performance rating of workers by educational level

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Supervisor**  **rating** | **Educational level** | | | Row total |
| **Elementary (1)** | **Junior (2)** | **Secondary (3)** |
| Satisfactory | 12 | 63 | 65 | 140 |
| Not-satisfactory | 8 | 17 | 35 | 60 |
| Column total | 20 | 80 | 100 | 200 - grand total |

# Solution

**1. Hypotheses** Ho: the cell proportions Pr in any row are equal

Ha: The cell proportions Pr in at least one row are now equal.

ν = (r - 1) (c - 1) = (2 - 1) (3 - 1)

= 2

with α = 0.05 and ν = 2, χ20.05 , 2 = 5.991

Hence the decision rule is

**2. Decision Rule** Reject Ho if sample χ2 > 5.991

Next we must compute the expected frequencies as we did in section

7.3.1. The results are as shown in Table 7.6.

**Table 4.6**

**Performance rating observed and expected frequencies**

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| **Supervisor Rating** | **Educational Level** | | | | | | **Row**  **Total** |
| **Elementary** | | **Junior high school** | | **Secondary high school** | |
| **fo** | **fe** | **fo** | **fe** | **fo** | **fe** |
| Satisfactory | 12 | 14 | 63 | 56 | 65 | 70 | 140 |
| Not-satisfactory | 8 | 6 | 17 | 24 | 35 | 30 | 60 |
| Column Total | 20 | | 80 | | 100 | | 200 |

Then compute Sample χ2 as shown below:

**Table 4.7 Performance rating χ2 calculation**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **fo** | **fe** | **fo - fe** | **(fo - fe)2** |  |
| 12  63  65  8  17  35 | 14  56  70  6  24  30 | -2  7  -5  2  -7  5 | 4  49  25  4  49  25 | 0.2857  0.8750  0.3571  0.6667  2.0417  0.8333  χ2 = 5.0595 |

**3. Sample χ2**

The sample χ2 value i.e. 5.0595 does not exceed the value 5.991, in the decision rule, so we accept Ho.

**4. Accept or reject:** Accept Ho.

Accepting Ho means that the proportion of satisfactory rated workers is the same for all three educational levels.

**Check Point 4.2**

The quality-control manager of Vita Company examined a random sample of parts made during the three shifts that the company operates. The manager classified the parts as good or defective as shown in Table 7.8. Perform, at the 5 percent level, a test of the hypothesis that equal proportions of defective parts are made by the three shifts.

### Table 4.8

|  |  |  |  |
| --- | --- | --- | --- |
|  | Shift | | |
| **Day** | **Middle** | **Night** |
| Number of good parts | 427 | 273 | 240 |
| Number of defective parts | 23 | 27 | 10 |

**Answer:** The proportions of defective parts made by the shifts are not all equal. The sample χ2, 7.19, is greater than 5.991.

* + 1. **Goodness-of-fit tests**

Goodness-of-fit tests use sample data as a basis for accepting or rejecting assumptions about a population’s distribution. The assumptions are stated as:

The null hypothesis: Ho: The population distribution is the specified one

Ha: The population distribution is not the specified one

A goodness-of-fit test is a hypothesis test that is performed by first computing the frequencies that would be expected if Ho is true, then the differences between the frequencies observed in the sample fo and the expected frequencies fe are used to calculate:

**Sample χ2**= 

Finally, the sample χ2 is compared with the appropriate χ2 α, ν to decide whether Ho should be accepted or rejected.

Goodness-of-fit tests differ from independence tests both in the methods used to compute expected frequencies and in the rule for determining the number of degrees of freedom. In a goodness-of-fit test, the method for calculating the expected frequencies depends on the population assumptions that are made; and the number of degrees of freedom in a goodness-of-fit test is:

ν = ne – 1 – g

Where: ne = number of fe values used in computing the sample χ2

g = number of population parameters (e.g. μ, δ) estimated from the sample.

**a. Goodness-of-fit: Binomial Distribution**

A particular Binomial distribution is specified by the values of two parameters, n and p, where: n = sample size (or number of trials)

p = probability of success in a trial

From Binomial Probability tables, we can calculate the probability of 0, 1, 2 … successes in n trials. Each of these probabilities, when multiplied by the sample size n, is an expected frequency for the number of successes in n trials.

**Example:** Mr. X, a sales representative for Moon Paper Company has five accounts to visit per day. It is suggested that sales by Mr. X may be described by the binomial distribution with the probability of selling each account being 0.4. Given the following frequency distribution of Mr. X’s number of sales per day, can we conclude that the data do in fact follow the binomial distribution with n = 5 and p = 0.4? Use the 0.05 significance level.

### Number of Sales per day Frequency of No. of sales

* + 1. 10
    2. 41
    3. 60
    4. 20
    5. 6
    6. 3

140

**Solution:**

We first need to find the expected numbers of frequency of sales with n = 5 and p = 0.4 from the binomial distribution table with p = 0.4. Thus the expected frequencies are calculated as:

### Table 4.9

|  |  |  |
| --- | --- | --- |
| **No. of sales per day** | **fo** | **fe** |
| 0  1  2  3  4  5 4 or more | 10  41  60  20  6  3 9  140 | 0.0778 x 140 = 10.892  0.2592 x 140 = 36.288  0.3456 x 140 = 48.384  0.2304 x 140 = 32.256  0.0768 x 140 = 10.752  0.0102 x 140 = 1.428 12.18 |

Since we previously said that fe ≥ 5, thus we combine the frequencies for 4 and 5 number of sales per day as shown in Table 9 above and proceed with the calculations for χ2.

### Table 4.10

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **No. of sales per day** | **fo** | fe | **(fo - fe)2** | **(fo - fe)2**  **fe** |
| 0  1  2  3  4 or more | 10  41  60  20  9 | 10.892  36.288  48.384  32.256  12.18 | 0.7957  22.2029  134.9315  150.2095  10.1124 | 0.0731  0.6119  2.7888  4.6568  0.8302  χ2 = 8.9608 |

We did not state the decision rule at the beginning because, when goodness-of-fit tests are performed, it is often not possible to know how many degrees of freedom there will be until after the expected frequencies are computed; that’s because it may be necessary to combine some frequencies before the sample χ2 is computed. So now that we have the information needed to determine the number of degrees of freedom, let’s start at the beginning:

The hypotheses are:

1. **Ho:** The number of sales per day by Mr. X follows a binomial distribution with n = 5, and p = 0.4

**Ha:**  The number of sales per day by Mr. X does not follow a binomial distribution with n = 5, and p = 0.4

The number of degree of freedom is

ν = ne – 1 – g

= 5 – 1 – 0 = 4

with α = 0.05 and ν = 4, we find (from the table at the back of the module) χ2 0.05, 4 = 9.49

Hence, the decision rule is

1. Reject Ho if sample χ2 > 9.49.
2. Sample χ2 = 8.9608

As 8.9608 does not exceed 9.49

1. Accept Ho.

Thus the sample evidence supports the hypothesis that the number of sales per day by Mr. X follows a binomial distribution with n = 5 and p = 0.4.

**Check Point 4.3**

A manufacturer packages drinking glasses in boxes of 50. All glasses from a sample of 100 boxes were examined, and the number of defective glasses in each box was recorded. The sample data are given in Table 7.11:

a. How much glasses were examined?

b. How many defective glasses were found?

c. Compute the sample proportion defective.

d. Do the sample data support the null hypothesis that the numbers of defective glasses in box are binomially distributed? (Perform a goodness-of-fit test at the 5 percent level).

### Table 4.11

|  |  |
| --- | --- |
| **Number of defectives in a box** | **Number of boxes** |
| 0  1  2  3  4  5 | 69  22  4  1  3  1 |

**Answer:** **a.** 5000 **b.** 50 **c.** 0.01

1. Accept the hypothesis that the distribution is a binomial distribution. The sample χ2,3.59 is less than 3.841.

**b. Goodness-of-fit: Normal Distribution**

The test of a normal fit is similar to the binomial fit test in the foregoing; only here the expected frequencies are determined using normal probabilities. (Table at the back)

**Example:** For inventory planning and control purposes, a certain chemical company wants to know if its sales of a liquid chemical are normally distributed. Sales for a random sample of 200 days are given in Table 7.12. The sample mean and sample standard deviation calculated from the 200 sample daily sales numbers are:

 = 40 thousand gallons

Sx = 2.5 thousand gallons

At a 5 percent level, perform a test of the hypothesis that sales are normally distributed. (The values to be used for the parameters μ and δ are the sample estimates. That will cost us 2 degrees of freedom when we compute the value of ν.)

**Table 4.12**

**Sales for 200 days**

|  |  |
| --- | --- |
| **Sales**  **(in thousands of gallons)** | **No. of days**  **fo** |
| Less than 34.0  34.0 and under 35.5  35.5 " " 37.0  37.0 " " 38.5  38.5 " " 40.0  40.0 " " 41.5  41.5 " " 43.0  43.0 " " 44.5  44.5 " " 46.0  46.0 or more | 0  13  20  35  43  51  27  10  1  0  200 |

**Solution:**

To compute the expected frequency for the “less than 34.0” class, we first find the probability for this class, and then multiply this probability by the sample size, 200.

at x = 34, we compute Z = x – μ

δ

= 34 – 40

2.5

= -2.4

From table at the back of the module;

P (0 to -2.4) = p (0 to 2.4) = 0.4918

The tail area probability we want is

0.5 – p (0 to -2.4) = 0.5 – 0.4918 = 0.0082

Since the sample includes 200 numbers, the expected frequency for the “less than 34.0” class is.

fe: 0.0082 (200) = 1.64

The expected frequency for the “34.0 and under 35.5” class is:

34 - 40 = -2.4

2.5 P(0 to 2.4) = 0.4918

35.5 - 40 = -1.8 P(0 to 1.8) = 0.4641

2.5 0.0277

fe = 0.0277 x 200 = 5.54

**c. Goodness-of-fit: Poisson distribution**

The Poisson process formula that you have came across in chapter 3 provides the probability of the number of "arrivals" in an interval of time. To remind you the formula is,

**P(x) = e-λt (λt)x**

**x!**

Where x = number of arrivals in t units of time

λ = average arrival rate per unit of time

t = number of units of time

In this section we will let t = 1 unit of time, so λt = λ (1) = λ. Then the formula becomes.

**P(x) = e-λ (λ)x**

**x!**

Where x = number of arrivals in 1 unit of time

λ = average arrival rate per unit of time

The formula (the Poisson distribution table) is used to determine expected frequencies in a test of the hypothesis that a distribution is a Poisson distribution with a stated value of the parameter λ.

**Example:** When a beer bottle-filling machine breaks a bottle, the machine must be shut down while the broken glass is removed. The production manager at Harar Brewery has been using a Poisson distribution with λ = 3 shutdowns per day, on the average, to determine the probabilities of 0,1,2,3… shutdowns in a day. The manager has tabulated the number of shutdowns per day in a random sample of 120 operating days, as shown in Table 4.14. We want to test, at the 5 percent level, the hypothesis that the number of shutdowns in a day has a Poisson distribution with λt = λ = 3.

### Table 4.14

Number of shutdowns in a day

|  |  |
| --- | --- |
| **No. of shutdowns in a day**  **X** | **Number of days**  **fo** |
| 0  1  2  3  4  5  6 or more | 3  20  29  22  23  10  13 |

**Solution:**

The formula for the probability of shutdowns in a day is:

P(x) = e-λ (λ)x  = e-3 (3)x

x! x!

To compute the expected number of days when there will be x = 0 shutdowns, we first compute

P(0) = e-3 (3)0

0!

= e-3 (1)= 1

= 0.0498

Then we multiply P(0) by the number of days in the sample, 120, to obtain:

fe(0) = 0.0498 (120) = 5.976

As the expected frequency of x = 0 shutdowns. Poisson probabilities can be computed on a hand calculator, or you can be found in the tables.

The probability of a or more arrivals is 1 minus the probability that the number of arrivals is 0, or 1, or 2, or…or (a-1). That is:

P (a or more) = 

In our example

P (6 or more) = 

Thus P(6 or more) = 1 - P(0) - P(1) - ….. - P (15)

= 1 - (0.0498) - (0.1494) - … - (0.1008)

= 0.0840

Therefore, P (6 or more) = 0.0840 (120) = 10.080

**Summary**

In this chapter, we analyzed discrepancies between frequencies to that we actually observe in a random sample and frequency fe that we would expect to occur if a stated null hypothesis Ho is true.

* The sample test statistic used was

**Sample χ2**= 

* The distribution of this test static is closely approximated by the mathematical chi-square distribution with parameter ν provided each fe value is at least five.
* The χ2 distribution has a single parameter ν - degree of freedom.
* Large values of χ2 indicate significant differences between the fo and the frequency that would be expected if Ho is true.
* A contingency table is a set of observed sample frequencies fo arranged in the cells of a table that has r-rows and c-columns.
* Tests of assumptions made about a population are called goodness-of-fit tests.
* The hypotheses in a contingency table for independence are

Ho: The row and column categories are independent

Ha: The row and column categories are not independent

* In a test for the equality of population proportions, Pr denotes the cell proportions to column total in a row of a contingency table. The hypotheses are:

Ho: The Pr (proportions) in any row are equal.

Ha: The Pr in at least one row are not equal.

* A test is made by obtaining a random sample of observed frequencies fo. Then the probability distribution specified in Ho is used to compute expected frequencies fe.
* The chi-square distribution could be applied to:
  + - * Test the equality of several proportions
      * Goodness-of-fit tests
      * Test the independence between two variables

**Exercise**: For the brewery example, use the formula to compute the expected number of days when there will be 4 shutdowns.

**Answer**: 20.16 (0.1680 x 120)

**Table 4.15**

Calculations of expected frequencies and χ2 for testing a Poisson fit.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **No. of shutdowns in a day** | **No. of days**  **fo** | **P(x)** | **120 times P(X)**  **fe** | **fo -fe** | **(fo - fe)2**  **fe** |
| 0  1  2  3  4  5  6 or more | 3  20  29  22  23  10  13 | 0.0498  0.1494  0.2240  0.2240  0.1680  0.1008  0.0840 | 5.976  17.928  26.880  26.880  20.160  12.096  10.080 | -2.976  2.072  2.120  -4.880  2.840  -2.096  2.920 | 1.482  0.239  0.167  0.886  0.400  0.363  0.846  χ2 = 4.383 |

Now we have calculated all fe and find the sample χ2, let's summarize our findings. We started with;

1. Ho: the number of shutdowns per day is a Poisson distribution with λ = 3 per day

Ha: the number of shutdowns per day is not a Poisson distribution with λ = 3

In the table, we computed χ2 using ne = 7 expected frequencies; the parameter λ was not estimated from the sample rather it is given in the Ho. Consequently,

ν = ne - 1 - g = 7 - 1 - 0

= 6

with α = 0.05, we find χ20.05,6 = 12.592. Hence, the decision rule is

1. Reject Ho if sample χ2 > 12.592

Next, from Table 7.15

1. Sample χ2 = 4.383

The sample χ2 does not exceed the 12.592 in the decision rule, so we

1. Accept Ho,

Accepting Ho means we conclude that shutdowns have a Poisson distribution with an average of λ = 3 shutdowns per day.

**Check Point 4.5**

The number of patients to arrive at an emergency room each day is recorded in the accompanying table for 100 days. The average number of emergency room patients is approximately 2 per day. Do the data support the hypothesis that the number of emergency room patients follows a Poisson distribution? Do a 1% test.

No. of patient Frequency

* 1. 15
  2. 17
  3. 21
  4. 20
  5. 19

5 or more 8

**Answer:** Sample χ2 = 17.93

Degree of freedom = 5

Critical Value = 15.086

Decision = reject Ho

**Check Point 4.1**

In investigating whether there is a relationship between the qualification test scores of persons who have gone through a certain job-training program and their subsequent performance on the job, 400 samples were taken and the results are as depicted in Table 7.4 below.

### Table 4.4

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Qualification test scores** | Performance | | | RowTotal |
| Poor | Fair | Good |
| Below Average | 67 | 64 | 25 | 156 |
| Average | 42 | 76 | 56 | 174 |
| Above Average | 10 | 23 | 37 | 70 |
| Column total | 119 | 163 | 118 | 400 - **Grand total** |

At a 0.01 level of significance, test whether the on-the-job performance of persons who have gone through the training program is independent of their qualification test score.

**Answer:** Since χ2 = 40.89 exceeds 13.277, reject Ho. That is, we conclude that there is a relationship between qualification test score and on-the-job performance.

**Check Point 4.4**

In a hospital, 150 patients were checked for their cholesterol level. Do the data collected support the hypothesis that the cholesterol levels follow a normal distribution with a population mean of 200 and a population standard deviation of 11?

|  |  |
| --- | --- |
| **Cholesterol level** | **Frequency** |
| 180 or less | 10 |
| 180 – 200 | 67 |
| 200 – 220 | 65 |
| 220 or more | 8 |

Answer: Sample χ2 =6.56,

Degree of freedom = 3,

Critical Value = 7.815,

Decision - do not reject Ho.

**Exercise:** Compute the expected frequency for the "35.5 and under 37.0" class.

**Answer:** 15.84

The complete list of expected frequencies, computed as just illustrated, is given in the fourth column of table 4.13.

### Table 4.13

Calculation of expected frequencies and the sample χ2 for a normal goodness-of-fit test

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Sales Class** | **fo** | **Sales class probability** | **200 times probability,**  **fe** | **fo - fe** | **(fo - fe)2**  **fe** |
| Less than 34.0  34.0 and under 35.5  35.5 " " 37.0  37.0 " " 38.5  38.5 " " 40.0  40.0 " " 41.5  41.5 " " 43.0  43.0 " " 44.5  44.5 " " 46.0  46.0 or more | 0  13 **13**  20  35  43  51  27  10  1  0 **1** | 0.0082  0.0277  0.0792  0.1592  0.2257  0.2257  0.1592  0.0792  0.0277  0.0082 | 1.64  5.54 **7.18**  15.84  31.84  45.14  45.14  31.84  15.84  5.54  1.64 **7.18** | **5.82**  4.16  3.16  -2.14  5.86  -4.84  -5.84  **-6.18** | **4.7176**  1.0925  0.3136  0.1015  0.7607  0.7357  2.1531  **5.3193**  χ2 = 15.1940 |

After combining frequencies so that fe values are at least 5, the sample χ2 is computed to be 15.1940. Now let's summarize the test.

The hypotheses are:

1. Ho: the distribution is normally distributed

Ha: the distribution is not normally distributed

The sample χ2 was computed using ne = 8 expected frequencies. Two parameters, μ and δ, were estimated from the sample. Hence, the number of degrees of freedom is

ν = ne - 1 - g

ν = 8 - 1 - 2 = 5 degrees of freedom

with α = 0.05, χ20.05,5 = 11.070 (Table at the back) so the decision rule is,

1. Reject Ho if sample χ2 > 11.070.

From table 7.13,

1. Sample χ2 = 15.1940

The sample χ2 exceeds the value 11.070 in the decision rule, so we

1. Reject Ho.

Our conclusion is that daily sales are not normally distributed.

***Exercises***

1.A production supervisor is interested in knowing if the number of breakdowns on four machines is independent of the shift using the machines. Test this hypothesis based on the following sample information;

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Shift | A | B | C | D | Total |
| Morning | 15 | 10 | 18 | 12 | 55 |
| Evening | 12 | 8 | 15 | 10 | 45 |
| Total | 27 | 18 | 33 | 22 | 100 |

2.An automobile manufacturing firm is bringing out a new model. In order to map out its advertising campaign, it wants to determine whether the model appeal depends on age group or not. The firm takes a random sample from persons attending a preview of the new model and obtained the results summarized below:

***Age groups***

Persons who under 20 20-40 40-50 50 and over *Total*

Liked the car 146 78 48 28 **300**

Disliked the car 54 52 32 62 **200**

*Total* **200** **130 80 90 500**

Test whether the model appeal and age groups are independent.

3.A total of 200 students were registered for the introductory course in statistics, in four sections. Section A had 50 students, section B had 60 students, section C had 45 students and section D also had 45 students. A common exam was administered to all the four sections. Each section was thought by a different professor. The following table represents the number of students who passed or failed in each section.

Section A B C D Total

Passed 45 47 38 42 **172**

Failed 5 13 7 3 **28**

Total **50 60 45 45 200**

Tests the null hypothesis that the proportion students failed by the four professors are equal at 5% level of significance.

4.A survey of 320 families with 5 children each revealed the following distribution of boys: *Number of boys*  *Number of families*

0 12

1 40

2 100

3 110

4 56

* 1. 14

At 5% level of significance, can we conclude that male and female births are equally probable.(Hint: Use binomial distribution to find the expected number families having 0,1,2,3,4 or 5 children.)

5.Three candidates A,B and C are running for mayor of the city. A random sample of registered voters is selected from each of the three areas of Manhattan. Those in the sample showed their preferences for each of the candidate as follows:

**Candidate**

**Area A B C**

Uptown 18 42 10

Midtown 20 35 15

Downtown 15 15 10

At =0.01,test if the voter preference is related to the area of the city.