Lecture 5

TESTS OF INDEPENDENCE

Introduction

In chapter 4 we saw how to use the χ^2 distribution to look for patterns in **categorical data**.

Introduction

In chapter 4 we saw how to use the χ^2 distribution to look for patterns in **categorical data**.

Instead of comparing **one** categorical variable to a hypothesised probability distribution, we now use the χ^2 distribution to compare **two** sets of categorical variables.

Example: Employment status and gender

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or
- **unemployed**.

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or
- **unemployed**.

Suppose also that we know the graduate's sex:

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or
- **unemployed**.

Suppose also that we know the graduate's sex:

- **male** or

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or
- **unemployed**.

Suppose also that we know the graduate's sex:

- **male** or
- **female**.

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or
- **unemployed**.

Suppose also that we know the graduate's sex:

- **male** or
- **female**.

How can we display such categorical data?

Example: Employment status and gender

Suppose we're interested in the employment status of graduates six months after graduation. Graduates might be:

- in a **permanent** job;
- in a **temporary** job, or
- **unemployed**.

Suppose also that we know the graduate's sex:

- **male** or
- **female**.

How can we display such categorical data?

We've already seen how to construct frequency tables for a **single** categorical variable. But what about **two** categorical variables?

Presenting categorical data

Consider the following table which shows the first few rows of a computer file of data from the previous example:

Presenting categorical data

Consider the following table which shows the first few rows of a computer file of data from the previous example:

Case number	Employment status	Gender
1	2	0
2	0	0
3	1	1
4	0	1
5	1	1
6	0	0
7	0	1
⋮	⋮	⋮

Presenting categorical data

Consider the following table which shows the first few rows of a computer file of data from the previous example:

Case number	Employment status	Gender
1	2	0
2	0	0
3	1	1
4	0	1
5	1	1
6	0	0
7	0	1
⋮	⋮	⋮

We actually have information on **310** graduates. Thus, the above table would have 310 rows!

A more concise way of presenting these data is to consider all combinations of **employment status** and **gender**, and then tabulate the frequencies within these pairings, i.e.

A more concise way of presenting these data is to consider all combinations of **employment status** and **gender**, and then tabulate the frequencies within these pairings, i.e.

Employment status	**Gender**	Frequency
0	0	100
0	1	90
1	0	33
1	1	40
2	0	25
2	1	22

A more concise way of presenting these data is to consider all combinations of **employment status** and **gender**, and then tabulate the frequencies within these pairings, i.e.

Employment status	**Gender**	Frequency
0	0	100
0	1	90
1	0	33
1	1	40
2	0	25
2	1	22

At least we can get this table on a single page!

However, an even more succinct and easily readable way of presenting data from two categorical variables is to use **contingency tables**.

However, an even more succinct and easily readable way of presenting data from two categorical variables is to use **contingency tables**.

- Each **row** in the contingency table corresponds to a single category from one of the categorical variables;

However, an even more succinct and easily readable way of presenting data from two categorical variables is to use **contingency tables**.

- Each **row** in the contingency table corresponds to a single category from one of the categorical variables;
- Each **column** in the contingency table corresponds to a single category from the other categorical variable;

However, an even more succinct and easily readable way of presenting data from two categorical variables is to use **contingency tables**.

- Each **row** in the contingency table corresponds to a single category from one of the categorical variables;
- Each **column** in the contingency table corresponds to a single category from the other categorical variable;
- Contingency tables often show the column totals, row totals and the overall sample size too.

In the previous example, **employment status** has three categories and **gender** has two categories, so we will get a 3×2 contingency table:

In the previous example, **employment status** has three categories and **gender** has two categories, so we will get a 3×2 contingency table:

	Permanent	**Temporary**	**Unemployed**	Total
Male	100	33	25	158
Female	90	40	22	152
Total	190	73	47	310

The χ^2 test for independence

In a χ^2 test for independence, we ask

The χ^2 test for independence

In a χ^2 test for independence, we ask

"Is there any evidence in the data of an association between the two categorical variables?"

The χ^2 test for independence

In a χ^2 test for independence, we ask

"Is there any evidence in the data of an association between the two categorical variables?"

The basic framework for such a hypothesis test is as follows:

The χ^2 test for independence

In a χ^2 test for independence, we ask

"Is there any evidence in the data of an association between the two categorical variables?"

The basic framework for such a hypothesis test is as follows:

1. State the **null hypothesis** (H_0)

The χ^2 test for independence

In a χ^2 test for independence, we ask

"Is there any evidence in the data of an association between the two categorical variables?"

The basic framework for such a hypothesis test is as follows:

1. State the **null hypothesis** (H_0)
   In tests for independence, this is always

   H_0 : There is **no association** between the two variables   or

   H_0 : The two variables are **independent**

The χ^2 test for independence

In a χ^2 test for independence, we ask

"Is there any evidence in the data of an association between the two categorical variables?"

The basic framework for such a hypothesis test is as follows:

1. State the **null hypothesis** (H_0)
   In tests for independence, this is always

   H_0 : There is **no association** between the two variables or

   H_0 : The two variables are **independent**

2. State the **alternative hypothesis** (H_1)

The χ^2 test for independence

In a χ^2 test for independence, we ask

"Is there any evidence in the data of an association between the two categorical variables?"

The basic framework for such a hypothesis test is as follows:

1. State the **null hypothesis** (H_0)
   In tests for independence, this is always

   H_0 : There is **no association** between the two variables or

   H_0 : The two variables are **independent**

2. State the **alternative hypothesis** (H_1)
   This is just the opposite to the null hypothesis, i.e.

   H_1 : There *is* an association between the two variables or

   H_1 : The two variables are *not* independent

3. Calculate the **test statistic**

3. Calculate the **test statistic**
   In tests for independence, this is

$$X^2 = \sum \frac{(O - E)^2}{E},$$

3. Calculate the **test statistic**

In tests for independence, this is

$$X^2 = \sum \frac{(O - E)^2}{E},$$

where O and E are the observed and expected frequencies (respectively).

3. Calculate the **test statistic**

In tests for independence, this is

$$X^2 = \sum \frac{(O - E)^2}{E},$$

where O and E are the observed and expected frequencies (respectively).

The awkward bit is sometimes getting your expected frequencies!

3. Calculate the **test statistic**
   In tests for independence, this is

$$X^2 = \sum \frac{(O - E)^2}{E},$$

   where O and E are the observed and expected frequencies (respectively).

   The awkward bit is sometimes getting your expected frequencies!

   You need to compute these first before you can calculate the test statistic.

4. Find your **p–value**

4. Find your **p–value**
   In tests of independence, we use the chi–squared (χ^2) distribution, with ν degrees of freedom, where

4. Find your **p–value**
In tests of independence, we use the chi–squared (χ^2) distribution, with ν degrees of freedom, where

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

4. Find your **p–value**
In tests of independence, we use the chi–squared (χ^2) distribution, with ν degrees of freedom, where

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

As always in hypothesis tests, we compare our test statistic to the 10%, 5% and 1% critical values to obtain a range for our *p*–value.

4. Find your **p–value**
   In tests of independence, we use the chi–squared (χ^2) distribution, with ν degrees of freedom, where

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

   As always in hypothesis tests, we compare our test statistic to the 10%, 5% and 1% critical values to obtain a range for our *p*–value.

5. Reach a **conclusion**

4. Find your **p–value**
   In tests of independence, we use the chi–squared (χ^2) distribution, with ν degrees of freedom, where

$$\nu = (\text{number of rows} - 1) \times (\text{number of columns} - 1)$$

   As always in hypothesis tests, we compare our test statistic to the 10%, 5% and 1% critical values to obtain a range for our *p*–value.

5. Reach a **conclusion**

The tricky bit is finding the expected frequencies!

Obtaining expected frequencies

If the null hypothesis is true, then we'd expect the two categorical variables to be independent.

Obtaining expected frequencies

If the null hypothesis is true, then we'd expect the two categorical variables to be independent.

- Recall from semester 1 that, for two independent events A and B,

$$\Pr(\text{A and B}) = \Pr(\text{A}) \times \Pr(\text{B})$$

Obtaining expected frequencies

If the null hypothesis is true, then we'd expect the two categorical variables to be independent.

- Recall from semester 1 that, for two independent events A and B,

$$\Pr(\text{A and B}) = \Pr(\text{A}) \times \Pr(\text{B})$$

- For example, for the employment status and gender data, the expected probability of Female (**F**) and Temporary (**T**) is

Obtaining expected frequencies

If the null hypothesis is true, then we'd expect the two categorical variables to be independent.

- Recall from semester 1 that, for two independent events A and B,

$$\text{Pr}(\text{A and B}) = \text{Pr}(\text{A}) \times \text{Pr}(\text{B})$$

- For example, for the employment status and gender data, the expected probability of Female (**F**) and Temporary (**T**) is

$$\text{Pr}(\textbf{F} \text{ and } \textbf{T}) = \text{Pr}(\textbf{F}) \times \text{Pr}(\textbf{T}).$$

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

$$E = \Pr(\textbf{F} \text{ and } \textbf{T}) \times \text{sample size}$$

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

$$\begin{aligned} E &= \Pr(\mathbf{F} \text{ and } \mathbf{T}) \times \text{sample size} \\ &= \Pr(\mathbf{F}) \times \Pr(\mathbf{T}) \times \text{sample size} \end{aligned}$$

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

$$\begin{aligned} E &= \Pr(\mathbf{F} \text{ and } \mathbf{T}) \times \text{sample size} \\ &= \Pr(\mathbf{F}) \times \Pr(\mathbf{T}) \times \text{sample size} \\ &= \frac{\text{row total for } \mathbf{F}}{\text{sample size}} \times \frac{\text{column total for } \mathbf{T}}{\text{sample size}} \times \text{sample size} \end{aligned}$$

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

$$
\begin{aligned}
E &= \Pr(\mathbf{F} \text{ and } \mathbf{T}) \times \text{sample size} \\
&= \Pr(\mathbf{F}) \times \Pr(\mathbf{T}) \times \text{sample size} \\
&= \frac{\text{row total for } \mathbf{F}}{\text{sample size}} \times \frac{\text{column total for } \mathbf{T}}{\text{sample size}} \times \text{sample size} \\
&= \frac{\text{row total for } \mathbf{F} \times \text{column total for } \mathbf{T}}{\text{sample size}}
\end{aligned}
$$

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

$$
\begin{aligned}
E &= \Pr(\mathbf{F} \text{ and } \mathbf{T}) \times \text{sample size} \\
&= \Pr(\mathbf{F}) \times \Pr(\mathbf{T}) \times \text{sample size} \\
&= \frac{\text{row total for } \mathbf{F}}{\text{sample size}} \times \frac{\text{column total for } \mathbf{T}}{\text{sample size}} \times \text{sample size} \\
&= \frac{\text{row total for } \mathbf{F} \times \text{column total for } \mathbf{T}}{\text{sample size}} \\
&= \frac{152 \times 73}{310}
\end{aligned}
$$

Recall from last week that the expected frequency is found by multiplying the expected probability by the sample size; thus, for Female and Temporary, we get

$$
\begin{aligned}
E &= \Pr(\mathbf{F} \text{ and } \mathbf{T}) \times \text{sample size} \\
&= \Pr(\mathbf{F}) \times \Pr(\mathbf{T}) \times \text{sample size} \\
&= \frac{\text{row total for } \mathbf{F}}{\text{sample size}} \times \frac{\text{column total for } \mathbf{T}}{\text{sample size}} \times \text{sample size} \\
&= \frac{\text{row total for } \mathbf{F} \times \text{column total for } \mathbf{T}}{\text{sample size}} \\
&= \frac{152 \times 73}{310} \\
&= 35.794.
\end{aligned}
$$

Generally, the expected frequencies are found as

$$E = \frac{\text{row total} \times \text{column total}}{\text{total sample size}}$$

Example: Employment status and gender

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

Example: Employment status and gender

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

$$H_0 \quad : \quad \text{Employment status and gender are independent}$$

Example: Employment status and gender

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

$$
\begin{aligned}
H_0 &: \text{ Employment status and gender are independent} \\
H_1 &: \text{ Employment status and gender are } \textit{not} \text{ independent}
\end{aligned}
$$

Example: Employment status and gender

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

$$H_0 \; : \; \text{Employment status and gender are independent}$$
$$H_1 \; : \; \text{Employment status and gender are } \textit{not} \text{ independent}$$

Step 3 (calculating the test statistic)

Recall that, for any χ^2 test, the test statistic is

Example: Employment status and gender

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

$$H_0 \quad : \quad \text{Employment status and gender are independent}$$
$$H_1 \quad : \quad \text{Employment status and gender are } \textit{not} \text{ independent}$$

Step 3 (calculating the test statistic)

Recall that, for any χ^2 test, the test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E}.$$

Example: Employment status and gender

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

$$\begin{aligned} H_0 &: \text{Employment status and gender are independent} \\ H_1 &: \text{Employment status and gender are } \textit{not} \text{ independent} \end{aligned}$$

Step 3 (calculating the test statistic)

Recall that, for any χ^2 test, the test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E}.$$

We already have the observed frequencies:

Steps 1 and 2 (hypotheses)

Our null and alternative hypotheses are

$$H_0 \;:\; \text{Employment status and gender are independent}$$
$$H_1 \;:\; \text{Employment status and gender are } \textit{not} \text{ independent}$$

Step 3 (calculating the test statistic)

Recall that, for any χ^2 test, the test statistic is

$$X^2 = \sum \frac{(O-E)^2}{E}.$$

We already have the observed frequencies:

	Permanent	**Temporary**	**Unemployed**	Total
Male	100	33	25	158
Female	90	40	22	152
Total	190	73	47	310

To calculate the test statistic, we also need the **expected** frequencies! Remember that, if the null hypothesis is true, then

To calculate the test statistic, we also need the **expected** frequencies! Remember that, if the null hypothesis is true, then

$$E = \frac{\text{row total} \times \text{column total}}{\text{overall sample size}}.$$

To calculate the test statistic, we also need the **expected** frequencies! Remember that, if the null hypothesis is true, then

$$E = \frac{\text{row total} \times \text{column total}}{\text{overall sample size}}.$$

- For "cell 1" (Male and Permanent), we have

$$\begin{aligned} E_1 &= \frac{158 \times 190}{310} \\ &= \frac{30020}{310} \\ &= 96.839. \end{aligned}$$

- For "cell 2" (Male and Temporary), we have

$$\begin{aligned} E_2 &= \frac{158 \times 73}{310} \\ &= \frac{11534}{310} \\ &= 37.206. \end{aligned}$$

The corresponding frequencies for the other cells are shown in the table below:

	Permanent	**Temporary**	**Unemployed**	Total
Male	96.839	37.206	23.955	158
Female	93.161	35.794	23.045	152
Total	190	73	47	310

So we have

So we have

O	E	$\frac{(O-E)^2}{E}$
100	96.839	0.103
33	37.206	0.475
25	23.955	0.046
90	93.161	0.107
40	35.794	0.494
22	23.045	0.047
		1.272

So we have

O	E	$\frac{(O-E)^2}{E}$
100	96.839	0.103
33	37.206	0.475
25	23.955	0.046
90	93.161	0.107
40	35.794	0.494
22	23.045	0.047
		1.272

and so our test statistic is $X^2 = 1.272$.

Step 4 (finding the p–value)

The degrees of freedom is given by

Step 4 (finding the p–value)

The degrees of freedom is given by

$$\begin{aligned}\nu &= (\text{number of rows} - 1) \times (\text{number of columns} - 1) \\ &= (2-1) \times (3-1) \\ &= 1 \times 2 \\ &= 2.\end{aligned}$$

Step 4 (finding the p–value)

The degrees of freedom is given by

$$\begin{aligned}\nu &= (\text{number of rows} - 1) \times (\text{number of columns} - 1)\\ &= (2-1) \times (3-1)\\ &= 1 \times 2\\ &= 2.\end{aligned}$$

Referring to table 4.1 (in chapter 4), we see that this gives the following critical values:

Significance level	10%	5%	1%
Critical value	4.61	5.99	9.21

Step 4 (finding the p–value)

The degrees of freedom is given by

$$\begin{aligned}\nu &= (\text{number of rows} - 1) \times (\text{number of columns} - 1)\\ &= (2-1)\times(3-1)\\ &= 1\times 2\\ &= 2.\end{aligned}$$

Referring to table 4.1 (in chapter 4), we see that this gives the following critical values:

Significance level	10%	5%	1%
Critical value	4.61	5.99	9.21

Our test statistic $X^2 = 1.272$ lies to the left of the first critical value. Thus, our p–value is **bigger than 10%**.

Step 5 (form your conclusion)

Step 5 (form your conclusion)

- Our p–value is bigger than 10%, and so we have no evidence against the null hypothesis;

Step 5 (form your conclusion)

- Our p–value is bigger than 10%, and so we have no evidence against the null hypothesis;
- we retain H_0;

Step 5 (form your conclusion)

- Our p–value is bigger than 10%, and so we have no evidence against the null hypothesis;
- we retain H_0;
- It appears that the two categorical variables (employment status and gender) *are* independent!

Example (page 48)

In a market research survey 200 people are shown a proposed design for the new Mini Cooper car, and they are asked if they like the design. The responses, broken down by age groups, are shown in the table below.

Example (page 48)

In a market research survey 200 people are shown a proposed design for the new Mini Cooper car, and they are asked if they like the design. The responses, broken down by age groups, are shown in the table below.

	Under 21	21–35	Over 35	Total
Liked design	24	38	70	**132**
Disliked design	35	17	16	**68**
Total	**59**	**55**	**86**	**200**

Example (page 48)

In a market research survey 200 people are shown a proposed design for the new Mini Cooper car, and they are asked if they like the design. The responses, broken down by age groups, are shown in the table below.

	Under 21	21–35	Over 35	Total
Liked design	24	38	70	**132**
Disliked design	35	17	16	**68**
Total	**59**	**55**	**86**	**200**

Is there any evidence to suggest that age is associated with attitude to the proposed design?

Steps 1 and 2 *(hypotheses)*

$$H_0 \quad : \quad \text{There is no association between age and attitude}$$

Steps 1 and 2 *(hypotheses)*

$$H_0 \quad : \quad \text{There is no association between age and attitude}$$

$$H_1 \quad : \quad \text{There } \textbf{is} \text{ an association between age and attitude!}$$

Step 3 *(test statistic)*

Recall that the test statistic is:

$$X^2 \quad = \quad \sum \frac{(O - E)^2}{E}.$$

Step 3 *(test statistic)*

Recall that the test statistic is:

$$X^2 = \sum \frac{(O - E)^2}{E}.$$

We have the O's – these are just the **O**bserved frequencies. We need the "E"'s as well! Remember, these are given by:

Step 3 *(test statistic)*

Recall that the test statistic is:

$$X^2 = \sum \frac{(O - E)^2}{E}.$$

We have the O's – these are just the **O**bserved frequencies. We need the "E"'s as well! Remember, these are given by:

$$E = \frac{\text{row total} \times \text{column total}}{\text{overall sample size}}.$$

Step 3 *(test statistic)*

Recall that the test statistic is:

$$X^2 = \sum \frac{(O-E)^2}{E}.$$

We have the O's – these are just the **O**bserved frequencies. We need the "E"'s as well! Remember, these are given by:

$$E = \frac{\text{row total} \times \text{column total}}{\text{overall sample size}}.$$

For "cell 1" (**Liked design** and **Under 21**), we get:

Step 3 *(test statistic)*

Recall that the test statistic is:

$$X^2 = \sum \frac{(O-E)^2}{E}.$$

We have the O's – these are just the **O**bserved frequencies. We need the "E"'s as well! Remember, these are given by:

$$E = \frac{\text{row total} \times \text{column total}}{\text{overall sample size}}.$$

For "cell 1" (**Liked design** and **Under 21**), we get:

$$E_1 = \frac{\mathbf{132} \times \mathbf{59}}{\mathbf{200}}$$

$$= 38.94.$$

Step 3 *(test statistic)*

For the other cells we get:

Step 3 *(test statistic)*

For the other cells we get:

$$\begin{aligned} E_2 &= \frac{\mathbf{132} \times \mathbf{55}}{\mathbf{200}} \\ &= 36.3. \end{aligned}$$

Step 3 *(test statistic)*

For the other cells we get:

$$E_2 = \frac{\mathbf{132} \times \mathbf{55}}{\mathbf{200}} = 36.3.$$

$$E_3 = \frac{\mathbf{132} \times \mathbf{86}}{\mathbf{200}} = 56.76.$$

Step 3 *(test statistic)*

$$
\begin{aligned}
E_4 &= \frac{\mathbf{68} \times \mathbf{59}}{\mathbf{200}} \\
&= 20.06.
\end{aligned}
$$

Step 3 *(test statistic)*

$$E_4 = \frac{\mathbf{68} \times \mathbf{59}}{\mathbf{200}} = 20.06.$$

$$E_5 = \frac{\mathbf{68} \times \mathbf{55}}{\mathbf{200}} = 18.7.$$

Step 3 *(test statistic)*

$$
\begin{aligned}
E_4 &= \frac{\mathbf{68} \times \mathbf{59}}{\mathbf{200}} \\
&= 20.06. \\
E_5 &= \frac{\mathbf{68} \times \mathbf{55}}{\mathbf{200}} \\
&= 18.7. \\
E_6 &= \frac{\mathbf{68} \times \mathbf{86}}{\mathbf{200}} \\
&= 29.24
\end{aligned}
$$

Step 3 *(test statistic)*

To calculate the test statistic, it helps to draw up a table:

O (Observed frequencies)	E (Expected frequencies)	$\frac{(O-E)^2}{E}$
24	38.94	5.732

Step 3 *(test statistic)*

To calculate the test statistic, it helps to draw up a table:

O (Observed frequencies)	E (Expected frequencies)	$\frac{(O-E)^2}{E}$
24	38.94	5.732
38	36.3	0.080

Step 3 *(test statistic)*

To calculate the test statistic, it helps to draw up a table:

O (Observed frequencies)	E (Expected frequencies)	$\frac{(O-E)^2}{E}$
24	38.94	5.732
38	36.3	0.080
70	56.76	3.088
35	20.06	11.127
17	18.7	0.155
16	29.24	5.995

Step 3 *(test statistic)*

To calculate the test statistic, it helps to draw up a table:

O (Observed frequencies)	E (Expected frequencies)	$\frac{(O-E)^2}{E}$
24	38.94	5.732
38	36.3	0.080
70	56.76	3.088
35	20.06	11.127
17	18.7	0.155
16	29.24	5.995
		26.176

Step 3 *(test statistic)*

To calculate the test statistic, it helps to draw up a table:

O (Observed frequencies)	E (Expected frequencies)	$\frac{(O-E)^2}{E}$
24	38.94	5.732
38	36.3	0.080
70	56.76	3.088
35	20.06	11.127
17	18.7	0.155
16	29.24	5.995
		26.176

Thus, our test statistic is:

$$X^2 = \frac{(O-E)^2}{E}$$

$$= 26.176.$$

Our degrees of freedom to use Table 4.1 is given by

$$
\begin{aligned}
\nu &= (\text{number of rows} - 1) \times (\text{number of columns} - 1) \\
&= (2 - 1) \times (3 - 1) \\
&= 2.
\end{aligned}
$$

Step 4 *(p–value)*

Referring to Table 4.1, we thus get:

p–value	10%	5%	1%
Critical value	**4.61**	**5.99**	**9.21**

Step 4 *(p–value)*

Referring to Table 4.1, we thus get:

p–value	10%	5%	1%
Critical value	**4.61**	**5.99**	**9.21**

Our critical value is $X^2 = 26.176$ which is to the right of the 1% critical value.

Step 4 *(p–value)*

Referring to Table 4.1, we thus get:

p–value	10%	5%	1%
Critical value	**4.61**	**5.99**	**9.21**

Our critical value is $X^2 = 26.176$ which is to the right of the 1% critical value.

Thus, our p–value is **less than 1%**.

Step 5 *(Conclusions)*

- Since p is less than 1%, there is **strong** evidence against H_0

Step 5 *(Conclusions)*

- Since p is less than 1%, there is **strong** evidence against H_0
- Thus, we should **reject** H_0 in favour of H_1

Step 5 *(Conclusions)*

- Since p is less than 1%, there is **strong** evidence against H_0
- Thus, we should **reject** H_0 in favour of H_1
- There is evidence to suggest that age **is** associated with attitude to the proposed design of the new Mini Cooper.