Chapter 7

The χ^2 hypothesis test for contingency tables and goodness-of-fit

Contents

7.1 Introduction

In this final chapter, we will study the χ^2 test in three hypothesis testing problems, two (very similar) problems related to 'contingency table' data, and one concerned with testing whether data can be modelled with a specific choice of probability distribution. It is worth repeating what we said in Chapter 6 (section 6.9):

For any hypothesis testing method

- the basic approach is always the same;
- all that changes is that we use a different test statistic, which will have different distribution under H_0;
- in this module, you will not be expected to construct suitable test statistics for different situations, but you should know how to *recognise* a suitable test statistic: increasing (absolute) values should become less likely under H_0, and more likely under H_A.

Make sure you have understood Chapter 6 before reading on!

Motivating example

Below are (real!) data for two successive years of exam results for this module: numbers of students (at their first attempt) with result in the six categories: fail, pass, 3rd, 2.2, 2.1 and 1st.

year	fail	pass	3rd	2.2	2.1	1st
2015	29	13	11	21	76	27
2016	15	19	11	47	68	29

We call this sort of table a **contingency table**: a table giving **counts** of how many times a particular combination of 'row' and 'column' (year and exam result) has occurred.

Suppose we wish to know whether the results were 'noticeably' different between the two years; perhaps changes were made to the module from one year to the next, and we want to know if these affected the exam results. There is always bound to be *some* variation from one year to the next, so how can we tell if any differences are 'statistically significant' or not?

7.2 Plotting contingency table data in R

Before doing any formal analysis, we should always try to plot the data if possible; a plot is usually the best way to present the data in any case, and it can often tell us what results to expect from our formal analysis; if we make a mistake in the analysis, we may be able to spot it from the plot.

In R, we can produce a 'grouped' bar chart. Because the total numbers of students were different in the two years, we will convert raw counts to percentages

```
options(digits = 2)
# First create a matrix of the data, the same shape as the table
counts <- matrix(c(29, 15, 13, 19, 11, 11, 21, 47, 76, 68, 27, 29),
                 nrow = 2,
                 ncol = 6)
counts
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,]   29   13   11   21   76   27
## [2,]   15   19   11   47   68   29
```

```
# Create an 'empty' matrix of percentages, which we fill row-by-row
percentages <- matrix(0, nrow = 2, ncol = 6)
percentages[1, ] <- 100 * counts[1, ] / sum(counts[1, ])
percentages[2, ] <- 100 * counts[2, ] / sum(counts[2, ])
percentages
```

```
##      [,1] [,2] [,3] [,4] [,5] [,6]
## [1,] 16.4  7.3  6.2   12   43   15
## [2,]  7.9 10.1  5.8   25   36   15
```

Now we can produce a bar chart. Note that we call this a bar chart and not a histogram, because the width of the bars does not represent anything.

```
barplot(percentages, beside = TRUE,
        names.arg = c("fail", "pass", "3rd", "2.2", "2.1", "1st"),
        legend = c("2015", "2016"),
        xlab = "classification",
        ylab = "percentage of students in classification",
        args.legend = list(x="topleft"))
```

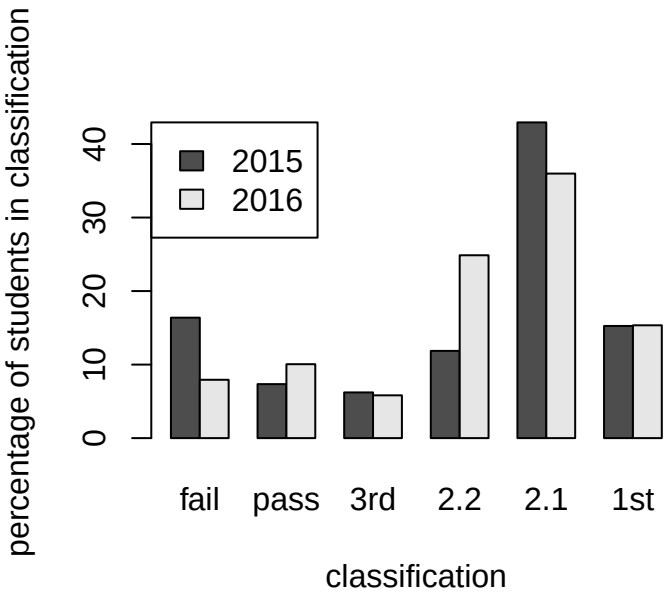

Figure 7.1: Comparing exam results in 2015 and 2016. This suggests that there was a change, with a smaller proportion of fails in 2016.

The plot suggests there was a change between 2015 and 2016, with a higher proportion of fails and lower proportion of 2.2s in 2015. We might expect any formal hypothesis test to conclude that there is a difference between the two years.

Aside: do not use pie charts!

You may be tempted to plot such data in a pie chart. In most situations, this will be a bad choice! Bar charts will often display the data more clearly, as heights are easier to compare than volumes, and *comparing* two pie charts is awkward. Compare Figures 7.1 and 7.2: which one do you think is easier to read?

```
pie(percentages[1,],
    labels = c("fail", "pass", "3rd", "2.2", "2.1", "1st"),
    main = "2015")
pie(percentages[2,],
    labels = c("fail", "pass", "3rd", "2.2", "2.1", "1st"),
    main = "2016")
```

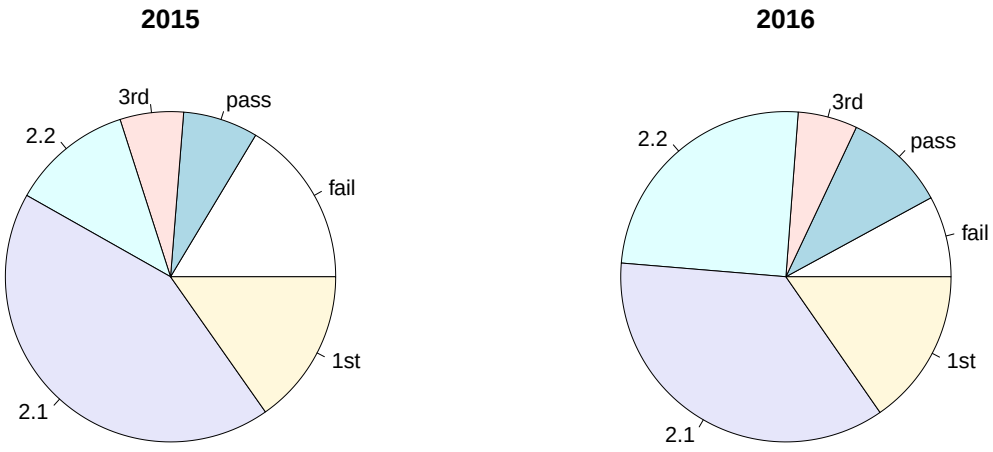

Figure 7.2: An attempt to compare exam results for two years using a pie chart. Do not use pie charts! Usually, the information will be displayed more clearly in a bar chart. Compare this with Figure 7.1.

And if, for some reason, you *have* to produce a pie chart, **never** draw a '3d' pie chart. These look awful, and you will impress no-one!

7.3 χ^2 test I: testing for 'row homogeneity'

We will now introduce a formal hypothesis test that can be used to analyse the data.

7.3.1 Notation and hypotheses

We first need some notation. Suppose that in 2015, the probabilities of a student achieving each of the six outcomes were $\theta_{1,1}, \ldots, \theta_{1,6}$, so that, for example, the probability of a pass in 2015 was $\theta_{1,2}$. Now define the corresponding probabilities for 2016 as $\theta_{2,1}, \ldots, \theta_{2,6}$, so to summarise, the probabilities are given by

year	fail	pass	3rd	2.2	2.1	1st
2015	$\theta_{1,1}$	$\theta_{1,2}$	$\theta_{1,3}$	$\theta_{1,4}$	$\theta_{1,5}$	$\theta_{1,6}$
2016	$\theta_{2,1}$	$\theta_{2,2}$	$\theta_{2,3}$	$\theta_{2,4}$	$\theta_{2,5}$	$\theta_{2,6}$

We can think of the table as an observation of two multinomial random variables, with distributions $multinom(177; \theta_{1,1}, \ldots, \theta_{1,6})$ in row 1, and distribution $multinom(189; \theta_{2,1}, \ldots, \theta_{2,6})$

If there was no change between the two years, we should have $\theta_{1,i} = \theta_{2,i}$ for $i = 1, \ldots, 6$, so we can write the hypotheses as

$$H_0 \quad : \quad \theta_{1,i} = \theta_{2,i} \quad \text{for } i = 1, \ldots, 6, \tag{7.1}$$

$$H_A \quad : \quad \theta_{1,i} \neq \theta_{2,i} \quad \text{for at least one } i. \tag{7.2}$$

In other words, in the null hypothesis, we say that the probability distributions of grades for the two rows are the same: we are testing for **row homogeneity**.

7.3.2 Calculating expected counts under the null hypothesis

As with all hypothesis tests, we will now assume the null hypothesis is true. In a χ^2 test, we consider what counts we would expect to get if the null hypothesis were true. Under the null hypothesis, we will re-write the table of probabilities as

year	fail	pass	3rd	2.2	2.1	1st
2015	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6
2016	θ_1	θ_2	θ_3	θ_4	θ_5	θ_6

so that in any column, the probabilities in the two rows are the same. What would the expected counts the null hypothesis? In each row, this would be the total number of students taking the exam in that year, multiplied by the corresponding probability:

year	fail	pass	3rd	2.2	2.1	1st
2015	$177\times\theta_1$	$177\times\theta_2$	$177\times\theta_3$	$177\times\theta_4$	$177\times\theta_5$	$177\times\theta_6$
2016	$189\times\theta_1$	$189\times\theta_2$	$189\times\theta_3$	$189\times\theta_4$	$189\times\theta_5$	$189\times\theta_6$

In the null hypothesis, we haven't specified particular values for $\theta_1, \ldots, \theta_6$, only that they didn't change from one year to the next, so we will now estimate them. To estimate θ_1, we had 366 students take the exam in total, and 44 students fail, so we will estimate θ_1 by

$$\hat{\theta}_1 = \frac{44}{366}. \tag{7.3}$$

Similarly, to estimate θ_2, we note that 32 out of 366 students got a 'pass' mark, so we estimate θ_2 by $\hat{\theta}_2 = \frac{32}{366}$. The expected counts are therefore

year	fail	pass	3rd	2.2	2.1	1st
2015	$177\times\frac{44}{366}$	$177\times\frac{32}{366}$	$177\times\frac{22}{366}$	$177\times\frac{68}{366}$	$177\times\frac{144}{366}$	$177\times\frac{58}{366}$
2016	$189\times\frac{44}{366}$	$189\times\frac{32}{366}$	$189\times\frac{22}{366}$	$189\times\frac{68}{366}$	$189\times\frac{144}{366}$	$189\times\frac{58}{366}$

Note that to get the expected count in row i, column j, we have calculated

$$\frac{\text{total in row } i\times \text{ total in column } j}{\text{grand total}}.$$

7.3.3 The test statistic

Define $O_{i,j}$ to be the observed count in row i, column j of the table, and $E_{i,j}$ to be the corresponding expected count. For example, we have $O_{2,4} = 47$ and $E_{2,4} = \frac{189\times 68}{366} = 35.11$. We use the test statistic

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j} - E_{i,j})^2}{E_{i,j}}, \tag{7.4}$$

where r denotes the number of rows in the table (2), and c denotes the number of columns (6).

- Recall that for any test statistic, we want a 'large' value to make us think H_0 is false. If H_0 is false, there should be some 'large' differences between what we observe ($O_{i,j}$) and what we would expect to observe if H_0 were true ($E_{i,j}$): these large differences will make χ^2 large.
- This test statistic cannot be negative: when calculating a critical region, we consider the right-hand tail only of its distribution under H_0: values in the left hand tail would be *consistent* with H_0.

The observed test statistic is

$$\chi^2_{obs} = \frac{(29-21.2787)^2}{21.2787} + \frac{(13-15.4754)^2}{15.4754} + \ldots + \frac{(29-29.9508)^2}{29.9508} \quad (7.5)$$

$$= 15.66 \quad (7.6)$$

7.3.4 The distribution of the test statistic under H_0

If H_0 is true, then approximately,

$$\chi^2 \sim \chi^2_\nu. \quad (7.7)$$

We do not give a formal justification here, but note that

- the test statistic is a sum of the squares of $(O_{i,j} - E_{i,j})/\sqrt{E_{i,j}}$: the χ^2_ν distribution results from summing the squares of ν independent standard normal random variables;
- the test statistic cannot be negative, so the normal or t distribution would be a bad approximation here: the χ^2 distribution will be more suitable.

7.3.5 Calculating the degrees of freedom parameter ν

Informally, we can think of the degrees of freedom as the number of 'pieces of information' we have to use in our hypothesis test to compare the two distributions. We have 12 observations in total, but there are two constraints:

- The first row total is fixed at 177. Given the counts 29, 13, 11, 21, and 76 in row 1, first 5 columns, we know that the count in row 1, column 6 must be $177-29-13-11-21-76$.
- The second row total is fixed at 189, and so a similar constraint applies.

This leave only 10 'free' observations. But then we have also had to estimate $\theta_1, \ldots, \theta_5$ (with θ_6 determined by $1-\theta_1-\theta_2-\ldots-\theta_6$), so we think of this as 'using up' another 5 observations. The total degrees of freedom is therefore $12-2-5=5$. Note that we can write this as

$$\nu = (r-1)(c-1) \quad (7.8)$$

where r is the number of rows and c is the number of columns in the table.

7.3.6 Calculating the critical region

For a (Neyman-Pearson) test of size 0.05, we want the 95th percentile of the χ^2_5 distribution, which we obtain using R.

```
qchisq(0.95, 5)

## [1] 11.07
```

Repeating what we said previously, it doesn't make sense to use the 2.5th and 97.5th percentiles here. The test statistic cannot be negative, and a small value close to 0 simply means that the observed values are similar to the expected values under H_0: this wouldn't suggest H_0 is false.

Since the observed test statistic lies inside the critical region, we reject H_0 at the 5% level of significance, and state that we do have evidence of a difference in distributions of grades between the two years.

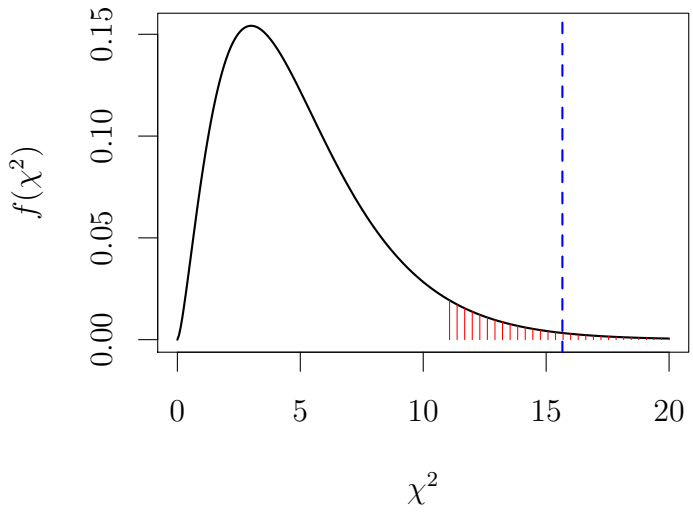

Figure 7.3: The solid line shows the approximate distribution of the random test statistic under H_0. The dashed line shows the value of the observed test statistic. The shaded area indicates the 5% critical region. If H_0 were true, the probability of the random test statistic lying in this region would be exactly 0.05. As the observed test statistic does lie in this region, we do reject H_0 at the 5% level of significance.

7.3.7 Calculating the p-value

To calculate the p-value, we want the probability, under H_0 that the test statistic will be at least as large as the one we observed:

$$p := P(\chi^2 \geq \chi^2_{obs}), \tag{7.9}$$

which we obtain from R as follows:

```
1 - pchisq(15.66, 5)

## [1] 0.007885
```

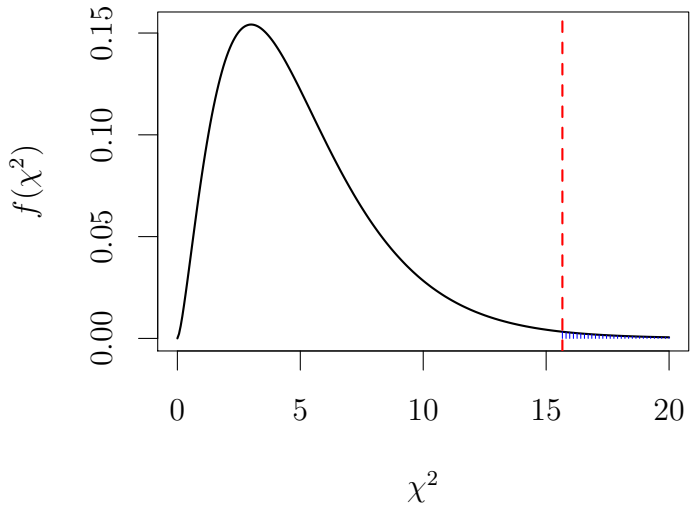

Figure 7.4: The solid line shows the distribution of the random test statistic under H_0. The dashed line shows the value of the observed test statistic. The p-value is the probability of observing a value this 'extreme', and is indicated by the shaded (which is only just visible).

7.4 χ^2 test II: testing for independence

The calculations in this test are identical to those in the previous test for row homogeneity. All that changes is that we write the hypotheses in a different way.

A slightly different scenario is when we want to test for independence 'between rows and columns' of a contingency table. A (famous) example is as follows. The following table shows data for graduate school admissions to UC Berkeley in 1973.

	admitted	rejected
males	3738	4704
females	1494	2827

Hence 44% of male applicants were successful, and only 35% of female applicants were successful. Was this evidence of gender discrimination?

7.4.1 Notation and hypotheses

We formulate this as test for independence between rows and columns. We think of each applicant as being counted in one of the four cells, with probabilities

	admitted	rejected
males	$\theta_{1,1}$	$\theta_{1,2}$
females	$\theta_{2,1}$	$\theta_{2,2}$

so that, for example, $\theta_{2,1}$ is the probability that an applicant will be both female and admitted. Note that here, we think of the data as arising from a *single multinomial* distribution: $multinom(12763; \theta_{1,1}, \theta_{1,2}, \theta_{2,1}, \theta_{2,2})$ (rather than one per row, as in the test for row homogeneity).

We define $\theta_{1,\bullet}$ to be the probability of a male applicant, so that

$$\theta_{1,\bullet} = \theta_{1,1} + \theta_{1,2}, \tag{7.10}$$

and $\theta_{\bullet,1}$ to be the probability an applicant is admitted, so that

$$\theta_{\bullet,1} = \theta_{1,1} + \theta_{2,1}. \tag{7.11}$$

Similarly, we define $\theta_{2,\bullet}$ and $\theta_{\bullet,2}$ respectively as the probability of a female applicant and the probability an applicant is rejected.

We now write the hypotheses as

$$H_0 \ : \ \theta_{i,j} = \theta_{i,\bullet} \times \theta_{\bullet,j} \quad \text{for all } i, j, \tag{7.12}$$

$$H_A \ : \ \theta_{i,j} \neq \theta_{i,\bullet} \times \theta_{\bullet,j} \quad \text{for at least one } i, j. \tag{7.13}$$

The null hypothesis states, for example,

$$P(\text{male and accepted}) = P(\text{male})P(\text{accepted}), \tag{7.14}$$

hence the two events are independent. This implies

$$P(\text{accepted} \mid \text{male}) = P(\text{accepted}), \tag{7.15}$$

hence knowledge that the applicant is male does not change the probability that the applicant would be accepted.

7.4.2 Calculating expected counts under the null hypothesis

The expected count in cell i, j would be $12763 \times \theta_{i,j}$, and under H_0, this would be $12763 \times \theta_{i,\bullet}\theta_{\bullet,j}$. We estimate $\theta_{1,\bullet}$, the probability of a male applicant, by

$$\hat{\theta}_{1,\bullet} = \frac{3738 + 4704}{12763} = \frac{8442}{12763},$$

and we estimate $\theta_{\bullet,1}$, the probability of being admitted, by

$$\hat{\theta}_{\bullet,1} = \frac{3738 + 1494}{12763} = \frac{5232}{12763}$$

Hence under H_0, the expected count in row 1, column 1 is

$$12763 \times \hat{\theta}_{1,\bullet} \times \hat{\theta}_{\bullet,1} = \frac{8442 \times 5232}{12763}.$$

We must have $\theta_{2,\bullet} = 1 - \theta_{1,\bullet}$ and $\theta_{\bullet,2} = 1 - \theta_{\bullet,1}$, so the estimates of $\theta_{2,\bullet}$ and $\theta_{\bullet,2}$ are determined automatically by the estimates of $\theta_{1,\bullet}$ and $\theta_{\bullet,1}$. The four expected counts are therefore

	admitted	rejected
males	$\frac{8442\times5232}{12763}$	$\frac{8442\times7531}{12763}$
females	$\frac{4321\times5232}{12763}$	$\frac{4321\times7531}{12763}$

Note that again, to get the expected count in row i, column j, we have calculated

$$\frac{\text{total in row } i\times \text{ total in column } j}{\text{grand total}},$$

the same calculation we did in the test for row homoegeneity.

7.4.3 The test statistic

We use the same test statistic as before:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{(O_{i,j}-E_{i,j})^2}{E_{i,j}}. \tag{7.16}$$

The observed test statistic is

$$\begin{aligned}\chi^2_{obs} &= \frac{(3738-3460.671)^2}{3460.671}+\frac{(4704-4981.329)^2}{4981.329}+\\ &\quad \frac{(1494-1771.329)^2}{1771.329}+\frac{(2827-2549.671)^2}{2549.671} &(7.17)\\ &= 111.25 &(7.18)\end{aligned}$$

7.4.4 The distribution of the test statistic under H_0 and calculating the degrees of freedom

If H_0 is true, then approximately,

$$\chi^2 \sim \chi^2_\nu. \tag{7.19}$$

To calculate the degrees of freedom, we have four cells in the table, but one constraint: the observations must sum to the grand total of 12763. We have also estimated two parameters: $\theta_{1,\bullet}$ and $\theta_{\bullet,1}$ so this leaves a total of one degree of freedom. Again, we have

$$\nu = (r-1)(c-1), \tag{7.20}$$

the same as we had in the test for row homogeneity.

7.4.5 Calculating the critical region

For a (Neyman-Pearson) test of size 0.05, we want the 95th percentile of the χ^2_1 distribution, which we obtain using R.

```
qchisq(0.95, 1)

## [1] 3.841
```

Since the observed test statistic lies inside the critical region, we reject H_0 at the 5% level of significance, and state that we do have evidence of a difference in acceptance probability for males and female applicants.

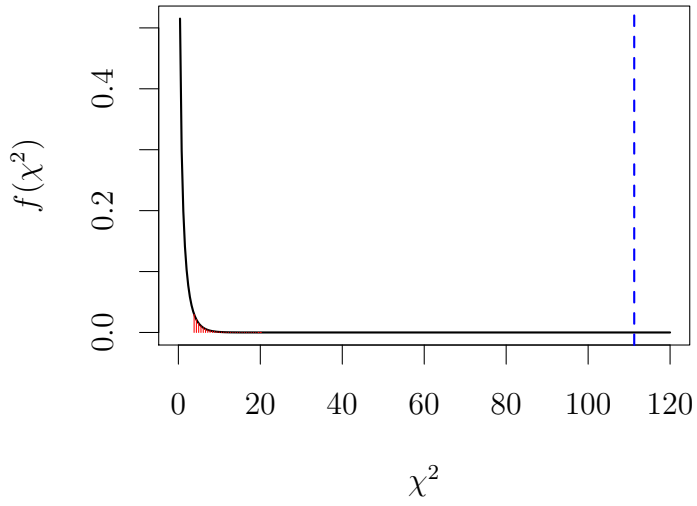

Figure 7.5: The solid line shows the approximate distribution of the random test statistic under H_0. The dashed line shows the value of the observed test statistic. The shaded area indicates the 5% critical region. If H_0 were true, the probability of the random test statistic lying in this region would be exactly 0.05. As the observed test statistic does lie in this region (it is *far* above the critical value), we do reject H_0 at the 5% level of significance.

7.4.6 Calculating the p-value

To calculate the p-value, we want the probability, under H_0 that the test statistic will be at least as large as the one we observed:

$$p := P(\chi^2 \geq \chi^2_{obs}), \tag{7.21}$$

which we obtain from R as follows:

```
1 - pchisq(111.25, 1)

## [1] 0
```

This is so small, R has reported a value of 0.

7.4.7 Simpson's paradox

In the admissions data, we had 44% of male applicants being admitted, and only 35% of female applicants being admitted, with this difference (strongly) statistically significant. But all is not quite as it first seems! Below, data are given for the six largest departments

Department	Male applicants	Percent admitted	Female applicants	Percent admitted
A	825	62%	108	82%
B	560	63%	25	68%
C	325	37%	593	34%
D	417	33%	375	35%
E	191	28%	393	24%
F	373	6%	341	7%

Observe how the percentage of successful female applicants is similar to or even *higher* than the percentage of successful male applicants. Looking at department A, one might even conclude that, if anything, there is gender discrimination *against* male applicants. This is an example of **Simpson's** paradox, where an association between two variables is 'reversed' after the introduction of a third variable:

- from the original table, we thought female applicants were *less* likely to be accepted
- after the breakdown by department, we now think female applicants were *more* likely to be accepted

The resolution of this paradox comes from noting that:

1. some departments were harder to get in to than others: they had lower acceptance rates;
2. female applicants were more likely to apply to these 'more competitive' departments.

This can explain why overall, the success rate for females was lower.

7.5 χ^2 test III: testing for goodness-of-fit

We may wish to know whether some data can be modelled by a particular distribution or not, and we can use the same χ^2 test as before to do this. The only modifications will be in how the expected counts are calculated, and the corresponding degrees of freedom parameter.

Motivating example

In a long sequence of observations on the behaviour of a laboratory animal, it is hypothesized that the numbers of actions categorised as 'grooming' in a day should follow a Poisson distribution. Actual numbers of actions, over 60 days, are as follows.

```
##  [1] 1 1 3 4 0 4 5 3 5 0 0 0 5 1 4 6 2 3 6 2 3 3 2 0 0 1 0 0 2 5 3 1 2 3 1
## [36] 2 1 0 3 3 3 5 3 1 0 3 3 0 2 0 2 3 4 0 3 2 1 2 3
```

We group the data in the following table (e.g., on 9 days, the number of actions was 1):

Number of actions	0	1	2	3	4	5	6	7 or more
Frequency	13	9	10	16	5	5	2	0

Are these data consistent with a Poisson distribution?

7.5.1 Notation and hypotheses

Before collecting the data, denote the number of action on the 60 days by the random variables $X_1, \ldots, X_{60}$. Our null hypothesis is

$$H_0 : X_1, \ldots, X_{60} \overset{iid}{\sim} Poisson(\lambda), \tag{7.22}$$

with the alternative that the data have some other distribution. Note that we have *not* specified a *value* of λ: the null hypothesis is simply that the data have a Poisson distribution (we don't care what the rate parameter λ is). Denote the 60 observed values by $x_1, \ldots, x_{60}$. From the previous table, we know that 13 of the values in $x_1, \ldots, x_{60}$ are equal to zero, 9 of the values are equal to 1 and so on.

7.5.2 Calculating expected counts under the null hypothesis

We first need an estimate of the Poisson rate parameter λ: then we can consider what the expected counts would be. We have $\lambda = E(X_i)$: the population mean is λ, so we can estimate it by the sample mean $\bar{x}$. Within $x_1, \ldots, x_{60}$, we have 13 lots of 0, 9 lots of 1, 10 lots of 2 and so on. Hence we estimate λ by

$$\bar{x} = \frac{9 + (10 \times 2) + (16 \times 3) + (5 \times 4) + (5 \times 5) + (2 \times 6)}{60} = 2.233 \text{ (to 3 d.p.)}. \tag{7.23}$$

The probabilities of the different possible values are then obtained by calculating

$$P(X = x) = \frac{e^{-2.233}(2.233)^x}{x!}$$

for $x = 0, \ldots, 6$ and for "7 or more", we calculate

$$1 - \sum_{x=0}^{6} \frac{e^{-2.233}(2.233)^x}{x!}$$

In R we do

```
dpois(0:6, 2.233)

## [1] 0.107 0.239 0.267 0.199 0.111 0.050 0.018

1-ppois(6, 2.233)

## [1] 0.0081
```

The expected count in each cell will be 60× the corresponding probability:

Number of actions	0	1	2	3	4	5	6	7 or more
Observed frequency	13	9	10	16	5	5	2	0
Expected frequency	6.4	14.4	16.0	11.9	6.6	3.0	1.1	0.5

In any χ^2 test, we use a rule-of-thumb that the expected count any cell of the table should be at least 5, otherwise the approximate distribution of the test statistic under H_0 may not be very accurate. Cells with smaller expected counts should be merged.

We merge the cells for "Number of actions" of 4 and above, as if the original table were

Number of actions	0	1	2	3	4 or more
Observed frequency (O_j)	13	9	10	16	12

so that the corresponding expected frequencies are

Number of actions	0	1	2	3	4 or more
Expected frequency (E_j)	6.4	14.4	16.0	11.9	11.2

7.5.3 The test statistic

We use the same test statistic as before, but for a table of counts with just one row:

$$\chi^2 = \sum_{j=1}^{5} \frac{(O_j - E_j)^2}{E_j}. \tag{7.24}$$

The observed test statistic is

$$\begin{aligned} \chi^2_{obs} &= \frac{(13-6.4)^2}{6.4} + \frac{(9-14.4)^2}{14.4} + \\ &\quad \frac{(10-16)^2}{16} + \frac{(16-11.9)^2}{11.9} + \frac{(12-11.2)^2}{11.2} && (7.25) \\ &= 12.4. && (7.26) \end{aligned}$$

7.5.4 The distribution of the test statistic under H_0 and calculating the degrees of freedom

If H_0 is true, then approximately,

$$\chi^2 \sim \chi^2_\nu. \tag{7.27}$$

To calculate the degrees of freedom, we have five cells in the table (after merging cells), but one constraint: the observations must sum to 60. We have also estimated one parameter: λ so this leaves a total of three degrees of freedom.

7.5.5 Calculating the critical region and p-value

For a (Neyman-Pearson) test of size 0.05, we want the 95th percentile of the χ^2_3 distribution, which we obtain using R.

```
qchisq(0.95, 3)
```

```
## [1] 7.8
```

Since the observed test statistic lies inside the critical region, we reject H_0 at the 5% level of significance, and state that we do have evidence that the data does not follow a Poisson distribution.

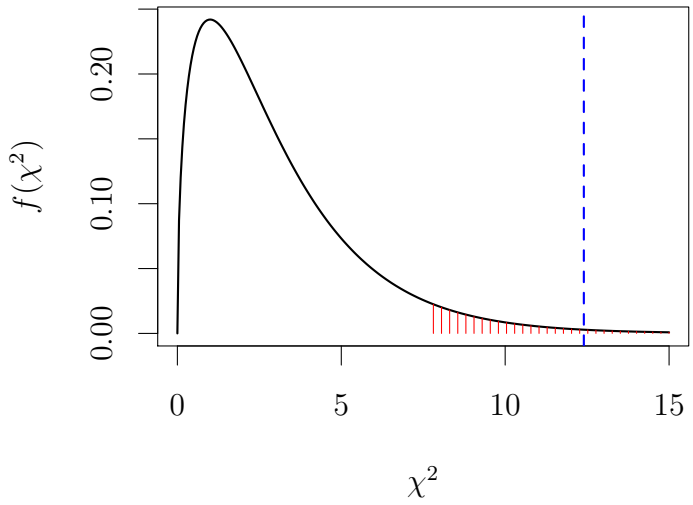

Figure 7.6: The solid line shows the approximate distribution of the random test statistic under H_0. The dashed line shows the value of the observed test statistic. The shaded area indicates the 5% critical region. If H_0 were true, the probability of the random test statistic lying in this region would be exactly 0.05. As the observed test statistic does lie in this region, we do reject H_0 at the 5% level of significance.

Calculating the p-value

To calculate the p-value, we want the probability, under H_0 that the test statistic will be at least as large as the one we observed:

$$p := P(\chi^2 \geq \chi^2_{obs}), \tag{7.28}$$

which we obtain from R as follows:

```
1 - pchisq(12.4, 3)

## [1] 0.0061
```

This is small (less than 0.01), indicating strong evidence against the null hypothesis.

It is clear from comparing observed and expected numbers that there are more zeroes in the data than would be expected from a Poisson distribution with the appropriate mean, and that the rest of the data are rather higher than expected.