CHAPTER 9

Measuring associations between two categorical variables

Conceptual metaphors and tests of independence

What you will learn from this chapter:

This chapter focuses on associations between two categorical variables. You will learn how to measure the association strength using odds ratios, Cramér's *V* and the φ-coefficient. You will also learn how to test whether the association is statistically significant with the help of the χ^2-test and the Fisher exact test. Bar plots, mosaic plots and association plots are used as visualization tools for cross-tabulated data. All these concepts and tools will be illustrated by case studies of metaphoric and non-metaphoric uses of the preposition *over* and the verb *see* in different registers.

9.1 Testing independence

Imagine you want to compare frequencies of two near synonymous constructions in two language varieties. An example might be constructions *going to* + *V* and *will* + *V* that express the future, and the British and American varieties of English. The frequencies are obtained from two comparable corpora. In this case, you will have four frequencies:

– *a*: frequency of construction *X* in variety *A;*
– *b*: frequency of construction *Y* in variety *A;*
– *c*: frequency of construction *X* in variety *B;*
– *d*: frequency of construction *Y* in variety *B.*

It is common to represent these frequencies in a table, as shown in Table 9.1. Such tables, which cross-tabulate the levels of two or more categorical variables, are called **contingency tables**. They play a very important role in analysis of categorical data.

Table 9.1 A contingency table

	Variety *A*	Variety *B*	Total
Construction *X*	*a*	*c*	*a* + *c*
Construction *Y*	*b*	*d*	*b* + *d*
Total	*a* + *b*	*c* + *d*	*a* + *b* + *c* + *d*

The rows represent one binary variable, which can be called *Construction*. The columns represent another binary variable, which describes the language variety. The row and column totals are called **marginal frequencies**. The question is whether the variables *Variety* and *Construction* are **independent** or **dependent (associated)**. In other words, is the use of construction *X* and construction *Y* independent from the language variety? Or does variety *A* 'prefer' construction *X* over construction *Y* more strongly than variety *B*, or vice versa?

In this chapter, we will consider the most popular test of independence, the χ^2 ('chi-square', or 'chi-squared') test, as well as the Fisher exact test. The latter is more appropriate when the frequencies are low (see more details in the additional information box in Section 9.2.3). While these well-known tests can tell us if the association is statistically significant, one should also measure the effect size. The effect size measures for categorical variables, such as the odds ratio, Cramér's *V* and the ϕ-coefficient, indicate the strength of association. These measures are analogous to the correlation coefficient, which measures the strength of correlation between two quantitative variables (see Chapter 6). As in correlation analysis, it is important to distinguish between effect size and statistical significance and report both.

9.2 The story of *over* is not over: Metaphoric and non-metaphoric uses in two registers (analysis of a 2-by-2 contingency table)

9.2.1 The data and hypothesis

To be able to reproduce the code in this case study, make sure that you have the following packages installed:

```
> install.packages(c("ggplot2", "reshape", "vcd"))
```

These packages should also be loaded in the beginning of your R session:

```
> library(ggplot2); library(reshape); library(vcd)
```

The case studies in this chapter focus on metaphoric and non-metaphoric uses of words in different registers. In their seminal work, *Metaphors we live by*, Lakoff and Johnson (1980) made a very significant contribution to our understanding how metaphors organize the conceptual system and shape up the language. In metaphors, we normally map a more embodied, concrete source domain onto a more abstract, less directly accessible target domain. For instance, the basic image schemata UP and DOWN are frequently used to speak about emotions (*That boosted my spirits* vs. *I'm feeling down*), quantity (*The prices soar* vs. *The prices plummet*), social status (*She's made it to the top* vs. *He's an underdog*), morality (*She's got high standards* vs. *This is a low thing to do*), etc. (Lakoff & Johnson 1980).

Another influential work in Cognitive Linguistics was Brugman's (1988 [1981]) study of *over*, where she showed, among other things, how more abstract meanings of the

preposition, such as *agonize over a problem* and *take over the responsibilities*, emerge as metaphorical extensions from the more physical ones, e.g. *fly over a city* or *walk over a field*. This case study will investigate how the metaphorical and non-metaphorical meanings of *over* are distributed in the academic and spoken registers of English. The data for this study come from the VU Amsterdam Metaphor Corpus (Steen et al. 2010).[1] This is a small (190,000 words at the moment of writing) corpus that represents four registers of British English: academic texts, conversations, fiction and news. The corpus has been manually annotated by several researchers for metaphoric use according to a metaphor identification protocol and has an online search interface.

Measuring (dis)agreement: Cohen's κ

When one has to decide whether a word in a corpus is used metaphorically or not, or to perform any other kind of semantic annotation, it is useful to collect opinions of different annotators and check if they tend to agree or disagree. A well-known measure of inter-rater agreement is Cohen's κ ('kappa'). It is based on the observed proportions of inter-rater agreement and disagreement compared to the expected proportions. The expected values are used in order to take into account possible biases towards some answers. To compute Cohen's κ for two vectors `Rater1` and `Rater 2`, which contain imaginary raters' scores, one can use the function `ckappa()` in the `psy` package:

```
> install.packages("psy")
> library(psy)
> ckappa(cbind(Rater1, Rater2)) # do not run; only provided as an example
```

For more than two raters, you can use the `lkappa()` function in the same package:

```
> lkappa(cbind(Rater1, Rater2, Rater3)) #do not run
```

Similar to correlation coefficients, the scores range from to −1 (complete disagreement) to 1 (full agreement).

1. Available online at http://www2.let.vu.nl/oz/metaphorlab/metcor/search/index.html (last access 11.06.2015).

The hypothesis of this case study is that the subcorpora with spoken data and academic prose have different distributions of metaphorical and non-metaphorical uses of the preposition. The four corresponding numbers, which are shown in Table 9.2, were obtained by using the online interface. Rare ambiguous cases were discarded.

Table 9.2 Frequencies of metaphoric and non-metaphoric uses of *over* in academic and conversational registers of the VU Amsterdam Metaphor Corpus

	Academic	*Conversations*
Metaphoric use	22	5
Non-metaphoric use	4	12

To create a similar contingency table in R, you can use the `cbind()` command, which can create a matrix from two or more vectors joined as columns:

```
> over <- cbind(c(22, 4), c(5, 12))
> over
      [,1]  [,2]
[1,]  22    5
[2,]  4     12
```

Alternatively, you can use `rbind()`, which combines vectors as rows:

```
> over <- rbind(c(22, 5), c(4, 12))
```

Next, the names of the rows and columns are added:

```
> rownames(over) <- c("met", "nonmet")
> colnames(over) <- c("acad", "conv")
```

The result looks as follows:

```
> over
        acad  conv
met     22     5
nonmet   4    12
```

How to cross-tabulate data in two (or more) factors

In this example, the four frequencies in the contingency table were already available. However, in many cases you will need to cross-tabulate the levels in two (or more) factors. To demonstrate how this can be done, let us create two factors, `f1` and `f2`:

```
> f1 <- factor(c(rep("A", 10), rep("B", 20)))
> f1
[1] A A A A A A A A A A B B B B B B B B B B B B B B B B B B B B
Levels: A B
> f2 <- factor(c(rep("X", 15), rep("Y", 15)))
> f2
[1] X X X X X X X X X X X X X X X Y Y Y Y Y Y Y Y Y Y Y Y Y Y Y
Levels: X Y
```

The first option is to use the familiar `table()` function:

```
> yourTable <- table(f1, f2)
> yourTable
     f2
f1    X    Y
A     10    0
B      5   15
```

Alternatively, you can use `xtabs()` with the formula interface. Note that the place of the response variable on the left of the tilde sign is empty:

```
> yourTable <- xtabs(~ f1 + f2)
> yourTable
     f2
f1   X     Y
A    10     0
B     5    15
```

9.2.2 Visualizations, proportions and measures of effect size: Odds ratios, Cramér's *V* and the φ-coefficient

It is always instructive to begin with visual inspection of the data. There are many ways of visualizing counts for two or more categories. One of the most popular options is the bar plot, which shows the numbers in columns as bars of different heights. We discussed how one can create a bar plot for a one-dimensional table of counts in Chapter 4. A bar plot that represents a two-dimensional table can be created easily as follows:

```
> barplot(over, col = c("grey20", "grey80"), main = "Bar plot
of (non-)metaphoric uses of over", xlab = "Register", ylab =
"Frequency")
> legend("topright", fill = c("grey20", "grey80"), c("metaphoric",
"non-metaphoric"))
```

The result is displayed in Figure 9.1. The plot shows clearly that the proportion of metaphoric uses is greater in the academic register than in the spoken data.

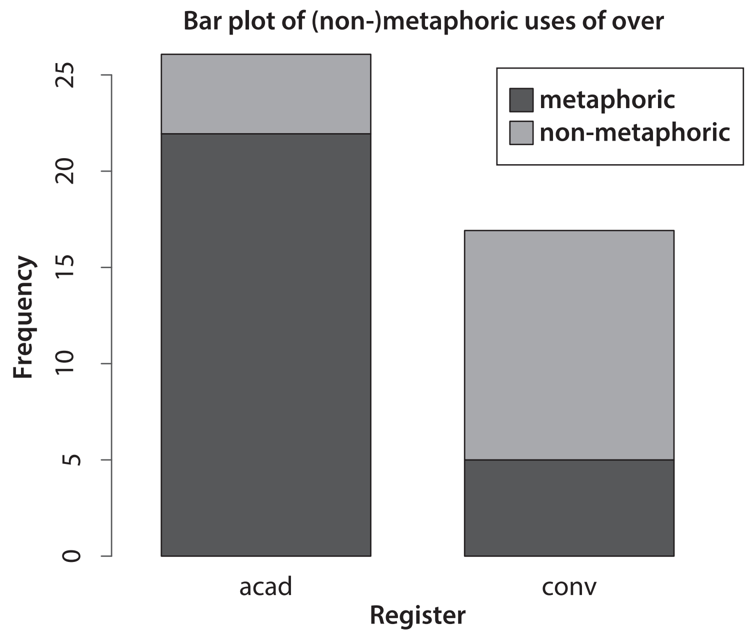
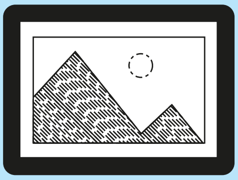


Figure 9.1. Bar plot of metaphoric and non-metaphoric uses of *over* with stacked bars



How to create a bar plot based on a contingency table with the help of `ggplot2`

First, you need to transform your contingency table into a data frame with two factors that specify the register and (non-)metaphoricity, and the third column that contains the counts of all combinations of the first two factors. For this purpose, you can use the function `melt()` from the package `reshape`.

```
> over.df <- melt(over)
> colnames(over.df) <- c("Metaphoricity", "Register", "Frequency")
> over.df
  Metaphoricity  Register  Frequency
1 met            acad      22
2 nonmet         acad      4
3 met            conv      5
4 nonmet         conv      12
```

Now we can create a bar plot of counts with stacked bars shown in Figure 9.1a:

```
> ggplot(over.df, aes(x = Register, y = Frequency, fill =
Metaphoricity)) + geom_bar(stat = "identity", colour = "black")
+ scale_fill_grey()
```

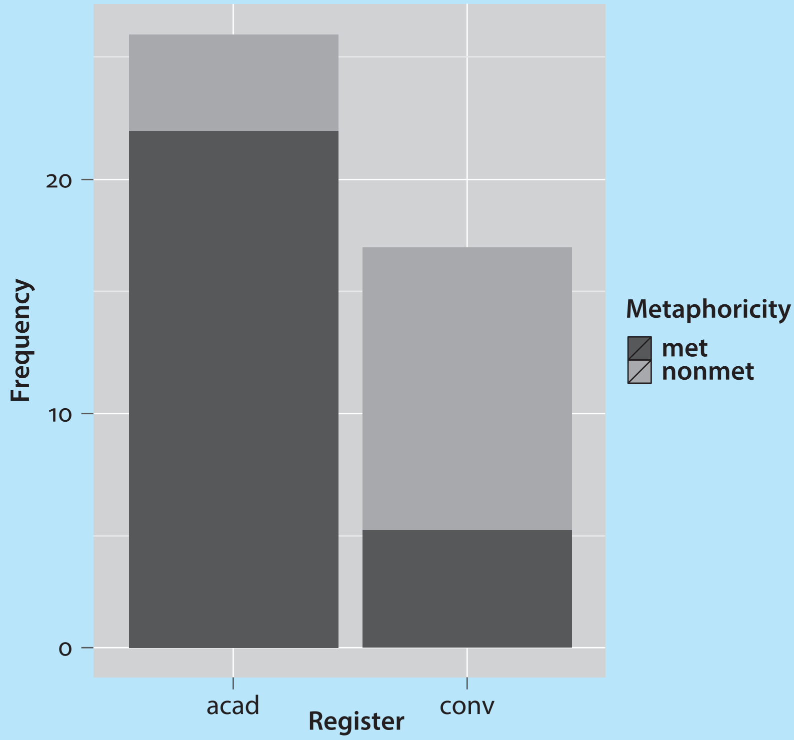


Figure 9.1a. A `ggplot2` version of the stacked bar plot in Figure 9.1

The bars that represent the metaphoric and non-metaphoric uses are stacked. Another option is to create a bar plot where these bars are placed next to one another. To do so, simply add the argument `beside = TRUE`. The result is shown in Figure 9.2.

```
> barplot(over, beside = TRUE, col = c("grey20", "grey80"), main
= "Bar plot of (non-)metaphoric uses of over", xlab = "Register",
ylab = "Frequency")
> legend("topright", fill=c("grey20", "grey80"), c("metaphoric",
"non-metaphoric"))
```

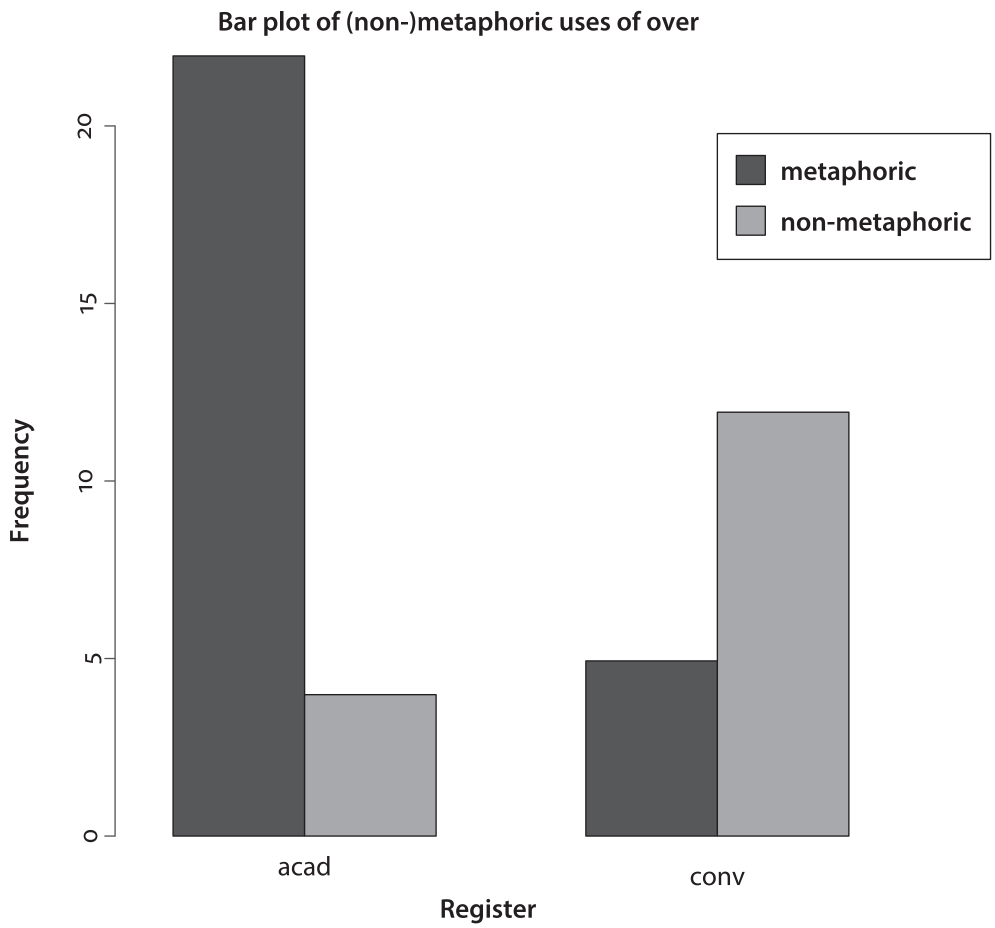
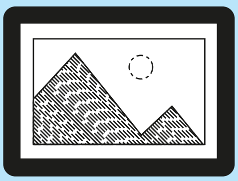


Figure 9.2. Bar plot of metaphoric and non-metaphoric uses of *over* with grouped bars



How to create a bar plot with grouped bars with the help of `ggplot2`

Follow the recommendations as for the plot in Figure 9.1a, but specify `geom_bar(…, position = "dodge")`, as shown below:

```
> ggplot(over.df, aes(x = Register, y = Frequency, fill = Metaphoricity)) + geom_bar(stat = "identity", colour = "black", position = "dodge") + scale_fill_grey()
```

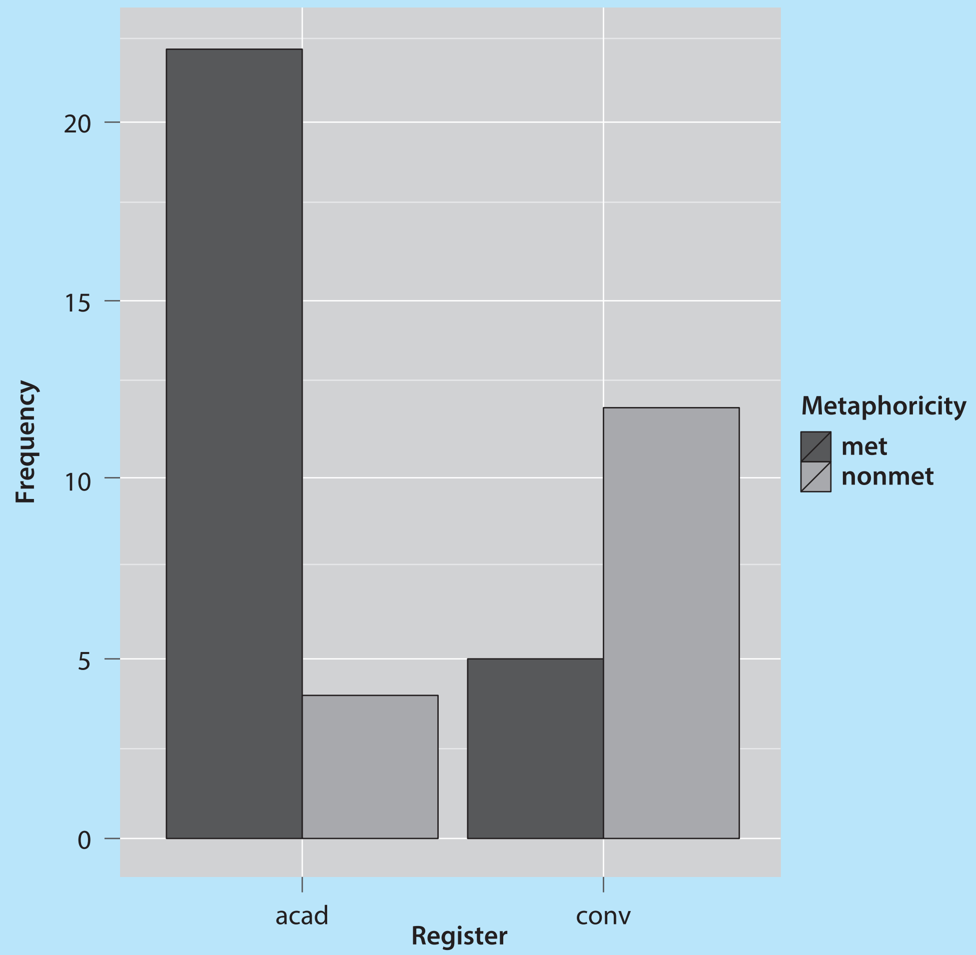


Figure 9.2a. A `ggplot2` version of the bar plot with grouped bars in Figure 9.2

It could be useful to count the proportions of metaphoric and non-metaphoric uses instead of representing the raw frequencies. Chapter 4 showed how one can transform a vector of frequencies into proportions with the help of `prop.table()`. For two-dimensional tables, there exist several opportunities:

– It is possible to compute proportions of the total number of observations in the table (the sum of all cells in the table is regarded as 1). This is the default option;
– One can also compute proportions of each cell in a row (for every row, the sum of cells is 1). To do so, add ‘1’;
– A third option is to compute proportions of every cell in a column (for every column, the sum of cells is 1). To use this option, add ‘2’.

The first (default) option yields the following table:

```
> prop.table(over)
          acad        conv
met       0.51162791  0.1162791
nonmet    0.09302326  0.2790698
```

Below is the outcome of the second option. The sum of the proportions in each row is 1.

```
> prop.table(over, 1)
        acad      conv
met     0.8148148 0.1851852
nonmet  0.2500000 0.7500000
```

Finally, the third option yields the proportions shown below. Now the sum of the proportions in each column is 1.

```
> prop.table(over, 2)
        acad      conv
met     0.8461538 0.2941176
nonmet  0.1538462 0.7058824
```

It is the third option that we are particularly interested in. One can see that the proportion of the metaphoric uses of *over* in the academic register (almost 0.85, or 85%) is much greater than the proportion of metaphoric uses in the conversational register (approximately 0.29, or 29%).

Thus, the proportions are clearly different. But how big is this difference? In statistical terminology, what is the effect size? The simplest measure of effect size for contingency tables is the **odds ratio**. It is a simple ratio of odds. As was mentioned in Chapter 4, odds are the ratio that compares the chances of *X* and the chances of *Y* (or simply non-*X*). If two outcomes are equally frequent, their odds will be 1. If *X* is greater than *Y*, the odds will be greater than 1. If *X* is less than *Y*, the odds will be less than 1. What are the odds of observing a metaphoric use of *over* versus a non-metaphoric use in each given register? Let us begin with the academic subcorpus:

```
> 22/4
[1] 5.5
```

The odds are greater than 1. The odds of metaphoric vs. non-metaphoric uses of *over* in the conversational register are smaller than one:

```
> 5/12
[1] 0.4166667
```

To compute the odds ratio (OR), one should divide the first odds by the second odds:

$$OR = \frac{Odds1}{Odds2}$$

```
> (22/4)/(5/12)
[1] 13.2
```

An odds ratio greater than 1 means that the first odds are greater than the second odds. If an odds ratio is less than 1, then the first odds are less than the second odds. In our case, the

odds ratio is 13.2. That means that the odds of the metaphoric uses of *over* in the academic register are 13.2 times greater than those in the conversational register.

Some other measures of effect size are Cramér's V and the ϕ ('phi') coefficient, which are identical for 2-by-2 tables. To compute them, one can use the `assocstats()` function in the `vcd` package:

```
> assocstats(over)
     X^2 df  P(> X^2)
Likelihood Ratio 13.843 1 0.00019871
Pearson 13.407 1 0.00025064

Phi-Coefficient     : 0.558
Contingency Coeff.:0.488
Cramer's V          : 0.558
```

For 2-by-2 tables, the absolute values of the ϕ-coefficient and Cramér's V range from 0 (no association) to 1 (perfect association). Some statisticians suggest the following guidelines for interpretation of effect size: $0.1 \leq \phi < 0.3$ indicates small effect size; $0.3 \leq \phi < 0.5$ shows a moderate effect; $\phi \geq 0.5$ indicates a strong effect (Sheskin 2011: 535). The value 0.558 shows a strong association. The third score is the Pearson contingency coefficient, but it is less widely used than the other two.

For illustration, let us consider an example of a very strong association. If the proportion of metaphoric uses of *over* were very high in the academic register, and there were very few in the conversational register, the coefficients would be close to 1. That is, if one takes a table `test` with the following values:

```
> test # do not run
       [,1]    [,2]
[1,]   99      1
[2,]   1       99
```

and computes the coefficients, the latter would be very close to the maximum value:

```
> assocstats(test) # do not run
                      X^2 df P(> X^2)
Likelihood Ratio 254.86 1  0
Pearson               192.08 1   0

Phi-Coefficient      : 0.98
Contingency Coeff. :0.7
Cramer's V           : 0.98
```

The main advantage of the ϕ-coefficient or Cramér's V in comparison with the OR is that they give an idea of association strength within the range from 0 to 1 (at least, for 2-by-2 tables). However, the OR shows the direction of association. That is, if an OR is greater than 1, the first odds are greater than the second ones, and if an OR is less than l, the first

odds are smaller than the second ones. Unfortunately, the ϕ-coefficient in `assocstats()` ignores this information. That is, if you swap the columns, the results will be the same, although the direction of association will change:

```
> over1 <- over[, c(2,1)]
> over1
        conv  acad
met       5    22
nonmet   12     4

> assocstats(over1)
                     X^2 df      P(> X^2)
Likelihood Ratio     13.843      1 0.00019871
Pearson              13.407      1 0.00025064

Phi-Coefficient     : 0.558
Contingency Coeff.:0.488
Cramer's V          : 0.558
```

The OR will be, however, different:

```
> (5/12)/(22/4)
[1] 0.07575758
```

In fact, the new OR is the inverse of the old value:

```
> 1/13.2
[1] 0.07575758
```

If you want to use the ϕ-coefficient and retain the information about the direction of association, you can add a minus sign to the coefficient in case the odds ratio is smaller than 1. Negative values of Cramér's *V*, however, are not reported. This measure is used predominantly for larger than 2-by-2 tables, where the direction of association is not specified.

9.2.3 Testing statistical significance: The χ^2-test of independence

The previous analyses show clearly that the proportion of metaphoric uses of *over* is much greater in the academic register, but is the difference statistically significant? To answer this question, we will carry out the χ^2-test. The null hypothesis of the test is that there is no association between the variables, i.e. between the rows and columns in the contingency table. In other words, the event "an observation is in row *i*" is independent of the event that the same observation is in column *j* (Conover 1999: 205). This is why the χ^2-test is regarded as a test of independence.

The χ^2-test is based on the comparison of **observed** and **expected frequencies**. Observed frequencies are the frequencies that you observe in the data, i.e. the actual numbers in the table. Expected frequencies are the frequencies that one can expect if the variables are inde-

pendent, that is, if the null hypothesis were true and there were no differences in the proportions of metaphoric and non-metaphoric uses of the preposition in the two registers. The generic formula for computing the expected frequency in row *i* and column *j* is as follows:

$$E_{ij} = \frac{S_i \, S_j}{n}$$

where S_i is the marginal frequency of row *i*, S_j is the marginal frequency of column *j* and *n* is the total number of observations. For example, the expected frequency of metaphoric uses of *over* in the academic register can be computed as follows:

```
> ((22 + 4)*(22 + 5))/(22 + 4 + 5 + 12)
[1] 16.32558
```

A more convenient way to retrieve all expected frequencies in a contingency table is as follows:

```
> chisq.test(over)$expected
         acad       conv
met      16.325581  10.674419
nonmet   9.674419   6.325581
```

If one transforms the expected frequencies to a table of proportions of metaphoric and non-metaphoric uses in both registers, the proportions in both registers will be equal:

```
> prop.table(chisq.test(over)$expected, 2)
         acad     conv
met      0.627907 0.627907
nonmet   0.372093 0.372093
```

Similarly, it is possible to make another table of proportions, with 1 as the sum for each row (recall that one has to use '1' instead of '2'). The proportions of the registers in the total number of metaphoric uses of the preposition are then identical to those in the total number of non-metaphoric uses, since the proportions in each row are the same:

```
> prop.table(chisq.test(over)$expected, 1)
        acad          conv
met     0.6046512     0.3953488
nonmet  0.6046512     0.3953488
```

To summarize, the expected values are the values that one could observe if there were no difference in the proportions between rows and no difference in the proportions between columns, in accordance with the null hypothesis. The expected values are influenced only by the marginal frequencies of non-metaphoric and metaphoric uses of *over* in the two registers taken together, on the one hand, and by the marginal frequencies of *over* in each register, on the other hand.

There exist only two assumptions of the χ^2-test that should be met:

– *The sample is randomly selected from the population of interest and the observations are independent.*
– *Every observation can be classified into exactly one category according to the criterion represented by each variable* (Conover 1999: 204–205).

There are no reasons to suspect that the assumptions are violated.[2] The function for performing the test is `chisq.test()`.[3] Its use is straightforward:

```
> chisq.test(over) #equivalent to chisq.test(…, correct = TRUE)

        Pearson's Chi-squared test with Yates' continuity correction

data: over
X-squared = 11.1487, df = 1, p-value = 0.0008409
```

As in all previous tests, the p-value below the level of significance, which is conventionally 0.05, allows one to discard the null hypothesis of no association. Note that due to some conceptual reasons one cannot choose between a one-tailed and two-tailed versions of the test.[4] The alternative hypothesis is always bidirectional, saying that the variables are associated, without specifying the direction of this association.

The bottom row also contains two other important statistics: the `X-squared` (the main statistic), and `df` (degrees of freedom). The χ^2-statistic is the sum of all squared deviations of the observed values in the table from the corresponding expected values divided by the expected values:

$$\chi^2 = \sum_{i=1}^{r}\sum_{j=1}^{c}\frac{\left(O_{ij} - E_{ij}\right)^2}{E_{ij}}$$

where O_{ij} is the observed frequency in a table cell and E_{ij} is the expected frequency for that cell. The deviations are squared in order to get rid of the minus sign if the observed frequency

2. It may happen that several occurrences of the preposition come from a text written by the same author. In that case, the corresponding observations would be dependent. However, this assumption is often relaxed in corpus linguistics (e.g. in collostructional analysis).

3. By default, the algorithm uses Yates' continuity correction. There is a lot of controversy about it in the statistical literature. One can run the test without the correction by adding `correct = FALSE`. In practice, this will not result in much difference.

4. Although there are ways of testing a directional hypothesis in case of 2-by-2 tables (e.g. Gries [2013: 171–172] recommends simply dividing the resulting p-value by two), this is seldom done in practice and may create an impression of trying to obtain a significant result at all costs.

is smaller than the expected frequency. This also gives more weight to large deviations. The larger the sum of differences between the observed and expected frequencies, the greater the χ^2-statistic. In addition to the test statistic, the result of the test also depends on the size of the table. The more cells, the higher the chance of finding divergences. This is taken into account in the notion of degrees of freedom, which was introduced in Chapter 1. To calculate the number of degrees of freedom for a contingency table, one should subtract 1 from the number of rows and the number of columns and multiply the results. In the case of a 2-by-2 table, the number of degrees of freedom is determined as follows: $(2 - 1) \times (2 - 1) = 1$. The p-values, as usual, are computed on the basis of both the χ^2-statistic and the degrees of freedom.

When the χ^2-test is not appropriate

There are situations when it is not correct to use the χ^2-test. Consider the following contingency table:

```
> test <- cbind(c(12, 2), c(4, 6))
> test
      [,1]  [,2]
[1,]  12    4
[2,]  2     6
```

If one runs the test, one will receive a warning from R:

```
> chisq.test(test)

     Pearson's Chi-squared test with Yates' continuity correction

data: test
X-squared = 3.6214, df = 1, p-value = 0.05704

Warning message:
In chisq.test(test): Chi-squared approximation may be incorrect
```

If one checks the expected values, one can see that two of them are smaller than 5:

```
> chisq.test(test)$expected
      [,1]       [,2]
[1,]  9.333333   6.666667
[2,]  4.666667   3.333333
```

(*Continued*)

In such situations, one is advised to use the Fisher exact test instead:

```
> fisher.test(test)

     Fisher's Exact Test for Count Data

data: test
p-value = 0.03241
alternative hypothesis: true odds ratio is not equal to 1
95 percent confidence interval:
     0.9560338 114.1109502
sample estimates:
odds ratio
8.055902
```

As one can see, the *p*-value returned by the χ^2-test is greater than 0.05, whereas the one returned by the Fisher exact test is smaller than the significance level. The result retuned by the Fisher test should be preferred because it is exact. Therefore, the association is statistically significant.

Some statisticians suggest more complex rules of thumb. According to these rules, the Fisher exact test should be preferred always when the total number of observations in all cells is smaller than 20. If the total number of observations is from 20 to 40, the Fisher exact test should be preferred if at least one expected frequency is smaller than 5. If the total number of observations is above 40, the Fisher exact test should be preferred if at least one of expected frequencies is smaller than 1 (Sheskin 2011: 646).

To conclude, the analyses have shown that the proportion of metaphoric uses of *over* is greater in the academic register than in the spoken register, and this difference is statistically significant. How can we interpret these results? Let us have a look at the data. In the conversational subcorpus, one sees mostly spatial uses of *over*:

(1) *Forgive me smiling but does everybody do that fall* ***over*** *the step there…*

Compare this with the following concordance lines from the academic subcorpus:

(2) *This session allows for an in depth examination of the drinking diaries* ***over*** *the last five weeks identifying the risky circumstances or situations…*

(3) *In the long debates* ***over*** *Marconi in 1913 and* ***over*** *the Curragh mutiny in 1914 Law was able to nail Asquith 's evasions where a more polished stylist might have…*

These examples demonstrate the temporal use of *over* and the sense 'concerning, referring to', where *over* introduces the object of discussion. Since the academic prose usually relates

to abstract things, such as theories, concepts, ideas, facts, etc., it is natural that the non-physical metaphorical senses of *over* are preferred.

9.3 Metaphorical and non-metaphorical uses of *see* in four registers (analysis of a 4-by-2 table)

9.3.1 The data and hypothesis

For this case study you will need use only one add-on package, `vcd`. If you have not installed it previously, you should do so. Next, load the package:

```
> install.packages("vcd")
> library(vcd)
```

This case study focuses on the verb *see*. Verbs of perception are experientially basic. This is why they play an important role in metaphorical mapping of more abstract events. For instance, the KNOWLEDGE IS SEEING metaphor is pervasive across different languages (Sweetser 1990). Consider the word *idea*, which originates from the Greek *eidon* 'see'. In contemporary English, as in many other languages, verbs of seeing have sense extensions related to knowledge, understanding, opinion and other mental states, e.g. *I see your point; She views the situation differently; This project looks fishy*. Below are examples from the VU Amsterdam metaphor corpus. The example in (4) contains a literal use of the verb see, whereas (5) contains a metaphorical one.

(4) *He **saw** a tall handsome woman dressed with careful and expensive informality in a black cashmere sweater with a silk scarf at the throat and fawn trousers.*

(5) *Maslow* (*1966*) *and Hudson* (*1966 and 1968*) ***saw** science as providing undemanding emotional activity appealing to boys moving from the calm of latency to the turbulence of adolescence.*

This case study explores frequencies of metaphorical and non-metaphorical uses of *see* in four registers of communication available in the corpus. A query about the total number of metaphoric or non-metaphoric uses of the verb *see* in different registers yields the counts shown in Table 9.3.

Table 9.3 The frequencies of metaphoric and non-metaphoric uses of *see* in four registers

	Academic	Conversations	Fiction	News
Metaphoric	44	48	27	17
Non-metaphoric	26	135	98	19

To create a table with the counts in R, one needs two vectors which contain the frequencies shown in the rows of Table 9.3. Vector `see.m` represents the frequencies of metaphoric uses of *see* in different registers, whereas `see.m` contains the frequencies of *see* used non-metaphorically:

```
> see.m <- c(44, 48, 27, 17)
> see.nm <- c(26, 135, 98, 19)
```

Next, the vectors are combined as rows in a matrix with the help of `rbind()`:

```
> see.reg <- rbind(see.m, see.nm)
```

The final step is to add the column names:

```
> colnames(see.reg) <- c("aca", "conv", "fic", "news")
> see.reg
        aca  conv  fic  news
see.m   44    48   27  17
see.nm  26   135   98  19
```

9.3.2 Descriptive statistics and visualizations

The frequencies are visualized in a bar plot with a legend (Figure 9.3):

```
> barplot(see.reg, beside = TRUE, main = "Barplot of (non-)metaphoric
uses of see", xlab = "Registers", ylab = "Frequencies")
> legend("topright", fill = c("grey30", "grey90"), c("metaphoric",
"non-metaphoric"))
```

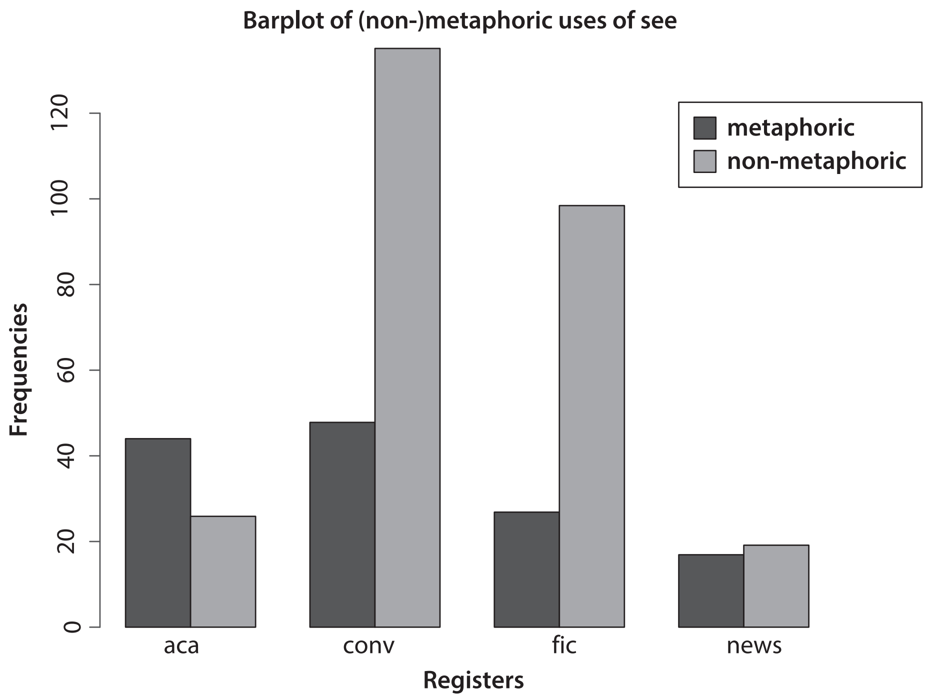


Figure 9.3. Bar plot of metaphoric and non-metaphoric uses of *see* in four registers with grouped bars

The heights of the dark grey and light grey bars represent the frequencies of the metaphoric and metaphoric uses, respectively. The bar plot demonstrates that the proportion of metaphoric uses is the greatest in the academic register, but it is not very clear whether the smallest proportion is found in fiction or conversations. Let us look at the proportions:

```
> prop.table(see.reg, 2)
        aca        conv       fic    news
see.m   0.6285714  0.2622951  0.216  0.4722222
see.nm  0.3714286  0.7377049  0.784  0.5277778
```

The table of proportions tells us that the lowest proportion of metaphoric uses (0.216) is found in fiction. This is somewhat counterintuitive, since fiction is the register where metaphoric language is expected to be used the most frequently. To measure the effect size, we can again use `assocstats()` from the `vcd` package:

```
> assocstats(see.reg)
                 X^2    df  P(> X^2)
Likelihood Ratio 41.000 3   6.5378e-09
Pearson          42.753 3   2.7773e-09
Phi-Coefficient    : 0.321
Contingency Coeff. : 0.306
Cramer's V         : 0.321
```

Although the function again returns three effect size measures, it is only Cramér's *V* that is traditionally reported as the effect size measure for larger than 2-by-2 tables. The main disadvantage of the ϕ-coefficient is that it can have values greater than 1 in case of larger than 2-by-2 tables, which is not desirable. The effect size (0.32) is moderate.[5]

9.3.3 Testing the statistical significance and analysing the residuals: The χ^2-test and mosaic and association plots

The χ^2-test will be used again to test if the register differences in the frequencies of metaphoric and non-metaphoric *see* are statistically significant:

```
> chisq.test(see.reg)

     Pearson's Chi-squared test

data: see.reg
X-squared = 42.7527, df = 3, p-value = 2.777e-09
```

The *p*-value is very small. The null hypothesis, which says that the proportions of metaphoric uses are equal in all registers, can thus be discarded. In other words, the categorical variables 'metaphoric/non-metaphoric use' and 'register' are not independent.

5. In principle, odds ratios can be computed, as well, but only for pairwise comparisons between different registers, not for the entire table.

When analysing medium and large contingency tables, it is often interesting to know which frequencies in the contingency table are greater than what can be expected by chance, and which are smaller. To do so, one can compare the observed and expected frequencies:

```
> chisq.test(see.reg)$observed
        aca  conv  fic  news
see.m   44    48    27  17
see.nm  26   135    98  19
> chisq.test(see.reg)$expected
     aca      conv    fic      news
met      22.99517  60.11594   41.0628  11.82609
nonmet   47.00483  122.88406  83.9372  24.17391
```

For instance, the expected frequency of metaphoric *see* in the academic texts is approximately 23. It is much smaller than the observed frequency 44. Therefore, the metaphoric *see* is overrepresented in the academic register. But which differences are more important, and which are less so? To answer this question, one can check the residuals:

```
> chisq.test(see.reg)$residuals
        aca         conv        fic         news
see.m   4.380270   -1.562652   -2.194561  1.504522
see.nm  -3.063712  1.092973    1.534951   -1.052315
```

The notion of residuals was discussed in Chapters 6 and 7. In a contingency table, the residuals (they are also called Pearson residuals) are the differences between the observed and expected frequencies divided by the squared root of the expected value. For a given cell in a row *i* and a column *j*, the Pearson residual can be computed as follows:

$$e_{ij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}}}$$

where O_{ij} is the observed frequency in the given cell and E_{ij} is the expected frequency. For example, the Pearson residual of the frequency of the metaphoric *see* in the academic register can be computed as follows:

```
> (44 – 22.99517)/sqrt(22.99517)
[1] 4.38027
```

If the observed frequency is greater than expected, the residual is positive. If the observed frequency is smaller than expected, the residual is negative. The greater the absolute value of a residual, the greater the discrepancy between the observed and expected frequencies, and the more it contributes to the test statistic. In fact, the χ^2-statistic is the sum of squared Pearson residuals. Therefore, the metaphoric uses of *see* are strongly overrepresented in the

academic register, and underrepresented in fiction. The non-metaphoric uses are strongly underrepresented in the academic texts.

The observed frequencies and the corresponding Pearson residuals can be visually represented in a mosaic plot with shading. To create a mosaic plot, one can use the `mosaic()` function from the package `vcd`. The result is displayed in Figure 9.4.

```
> mosaic(see.reg, shade = TRUE, varnames = FALSE)
```

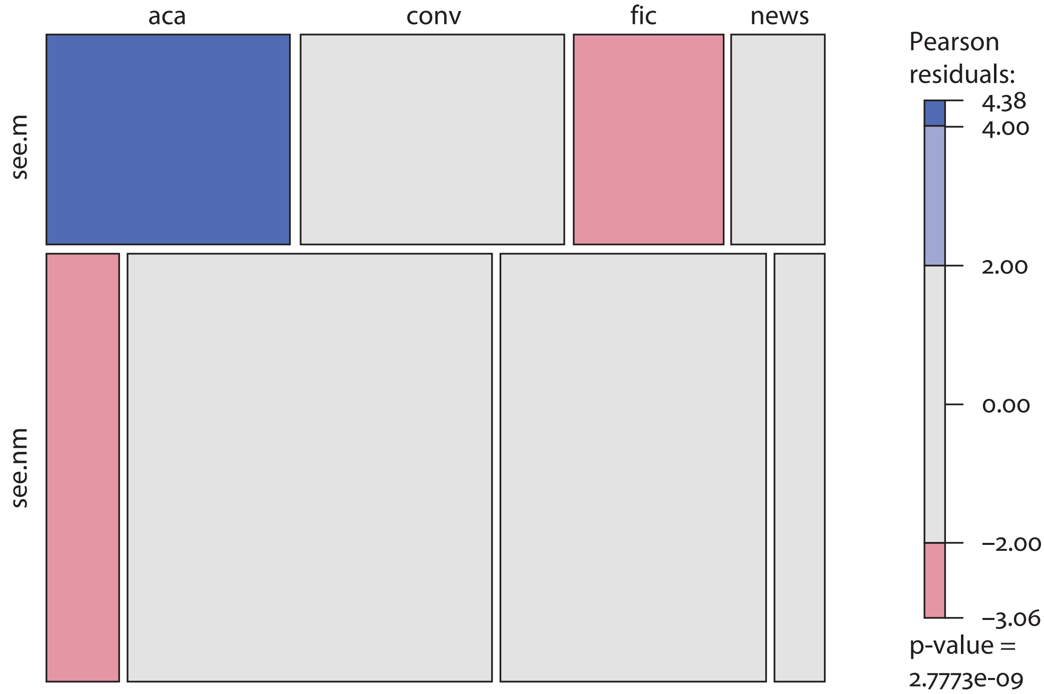


Figure 9.4. Mosaic plot of metaphorical and non-metaphorical uses of *see* in four registers. The colour of shading corresponds to the sign of a residual, and the intensity shows its relative importance

The colour of shading corresponds to the sign of a residual, and the intensity shows its relative importance: the more intensive, the greater the deviation. Unfortunately, the default colour scheme is lost in the black-and-white version, but when you reproduce the plot in R, you should be able to see the colours. Large positive residuals are indicated by blue rectangles if Pearson residuals range between 2 and 4, and by dark blue ones if the residuals are above 4. As for negative residuals, red rectangles indicate values between –2 and –4, whereas dark red ones (not shown in this plot) have residuals below –4.

The plot also reflects the total proportions of the cells, rows and columns in the contingency table. For example, the total area occupied by the metaphoric *see* is smaller than the total area occupied by the non-metaphoric uses because the total number of non-metaphoric occurrences of *see* is larger. On the other hand, the conversations occupy the largest areas both in the metaphoric and non-metaphoric rows, since this register contains the greatest number of instances of *see* in general. In contrast, the smallest total area is occupied by the news.

Another way of visualizing residuals is with the help of a so-called association plot. To create it, you can use another function from the `vcd` package, which is called `assoc()`:

```
> assoc(see.reg, shade = TRUE, varnames = FALSE)
```

The result is shown in Figure 9.5. The plot displays bars that either 'grow' or 'fall'. If a bar 'grows' above the baseline, like a stalagmite, the residual is positive, i.e. the observed frequency is greater than expected. If it 'falls' below the baseline, like a stalactite, the residual is negative, i.e. the observed frequency is smaller than expected. The height of a bar represents the value of the corresponding Pearson residual, whereas the width stands for the squared root of the expected value in the cell. The shading colour and intensity represent the same information as in the mosaic plot. Again, the graph shows that the metaphorical *see* is strongly overrepresented in the academic register and underrepresented in fiction, whereas the non-metaphorical *see* is underrepresented in the academic register.

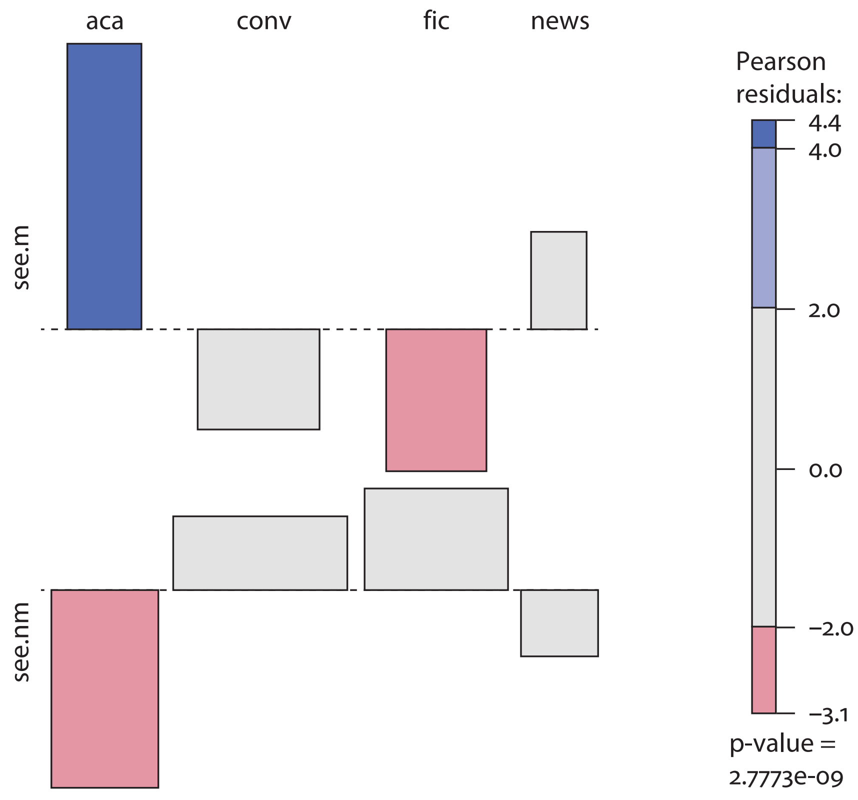


Figure 9.5. Metaphorical and non-metaphorical uses of *see* in four registers: association plot of residuals. The colour of shading and its orientation up- or downward corresponds to the sign of a residual, and the intensity of shading shows its relative importance

One can also determine which residuals represent significant deviations from the expected values at a given level of statistical significance. For this purpose, Agresti (2002: 81) recommends using standardized Pearson residuals, that is, Pearson residuals divided by their standard error, which can be computed as follows:

$$e_{sij} = \frac{O_{ij} - E_{ij}}{\sqrt{E_{ij}\left(1 - p_{i+}\right)\left(1 - p_{+j}\right)}}$$

where O_{ij} is the observed frequency in the given cell, E_{ij} is the expected frequency, p_{i+} is the proportion of the marginal frequency of row *i* with regard to the total number of observations, and p_{+j} is the proportion of the marginal frequency of column *j*. To obtain standardized residuals for a contingency table, you can use the following code:

```
> chisq.test(see.reg)$stdres
        aca       conv      fic       news
see.m   5.864076  -2.552903 -3.205358 1.921457
see.nm  -5.864076 2.552903  3.205358  -1.921457
```

If a standardized residual value is greater than 1.96 or smaller than –1.96, the cell makes a statistically significant contribution to the obtained χ^2-statistic value at the significance level of 0.05. For the significance level of 0.01, the numbers are 2.58 and –2.58, respectively. We find that all cells, except for the ones in the news column, make statistically significant contributions at the significance level of 0.05.

To summarize, the analyses reveal differences in the use of metaphoric and non-metaphoric *see* in different registers. These differences are highly significant ($p < 0.001$), which allows one to discard the null hypothesis of no association between metaphoricity of *see*, on the one hand, and the register, on the other hand. Most importantly, the metaphoric *see* is used more frequently than expected in the academic register, and less frequently than expected in fiction. Since seeing is an important metaphor for thinking, understanding and other mental processes, its highly frequent use in the academic texts is in fact not so surprising. Still, the relative scarcity of metaphors with *see* in fiction may seem counter-intuitive at first. After all, fiction should be considered the most appropriate register for figurative language. A possible explanation is that narratives are often written from the point of view of a character who observes the events. A character's literal, perceptual seeing, like hearing, smelling, etc. provides a window to the world of the narrative.

9.4 Summary

This chapter has demonstrated how one can measure and test associations between two categorical variables. You have learnt how to compute such popular measures of effect size as the odds ratio, Cramér's *V* and the ϕ-coefficient. The concepts of odds and odds ratios will be central in Chapter 12 on logistic regression. Note that only Cramér's *V* is used when the table contains more than two rows and/or columns. The best-known significance test is the χ^2-test, although the Fisher exact test should be used in cases of low expected frequencies. A further discussion of the Fisher exact test and a variety of association measures

follows in Chapter 10 and Chapter 11. You have also learnt how to create and interpret mosaic plots and association plots.

When your table has more than two dimensions, it is recommended to use log-linear analysis. This topic is beyond the scope of this introductory book, but see Field et al. (2012: Ch. 18). If one of the categorical variables can be regarded as the dependent one, then logistic regression may be appropriate (see Chapters 12 and 13).

How to write up the results of an independence test

One can report the results of an independence test using the following template: "We have found a significant association between metaphorical use of the preposition *over* and the register: $\chi^2(1) = 11.15$, $p < 0.001$. The odds of metaphorical use were 13.2 times higher in the academic register than in the conversational register, $\phi = 0.558$". Note that (1) designates the degrees of freedom. It is also important to report the effect size. When reporting effect sizes for larger tables, it is recommended to use Cramér's *V*. For 2-by-2 tables you can choose between ϕ, Cramér's *V* and the odds ratio.