Because the null hypothesis is not a 25:25:25:25 distribution across the four categories, we need to calculate the expected frequencies explicitly:

\[
\text{expected} = 556 \times \frac{9}{16}, \frac{3}{16}, \frac{3}{16}, \frac{1}{16} \\
312.75 \quad 104.25 \quad 104.25 \quad 34.75
\]

The expected frequencies are very close to the observed frequencies in Mendel’s experiment, but we need to quantify the difference between them and ask how likely such a difference is to arise by chance alone:

\[
\text{chisq.test(observed,p=c(9,3,3,1),rescale.p=TRUE)}
\]

```
Chi-squared test for given probabilities
data: observed
X-squared = 0.47, df = 3, p-value = 0.9254
```

Note the use of different probabilities for the four phenotypes \( p = c(9,3,3,1) \). Because these values do not sum to 1.0, we require the extra argument \( \text{rescale.p=TRUE} \). A difference as big as or bigger than the one observed will arise by chance alone in more than 92% of cases and is clearly not statistically significant. The chi-squared value is

\[
\sum((\text{observed}-\text{expected})^2/\text{expected})
\]

```
[1] 0.470024
```

and the \( p \)-value comes from the right-hand tail of the cumulative probability function of the chi-squared distribution \( 1 - \text{pchisq} \) with 3 degrees of freedom (4 comparisons –1 for contingency; the total count must be 556)

\[
1 - \text{pchisq}(0.470024,3)
\]

```
[1] 0.9254259
```

exactly as we obtained using the built-in \texttt{chisq.test} function, above.

**Two-by-Two Contingency Tables**

Count data are often classified by more than one categorical explanatory variable. When there are two explanatory variables and both have just two levels, we have the famous two-by-two contingency table (see p. 309). We can return to the example of Mendel’s peas. We need to convert the vector of observed counts into a matrix with two rows:

\[
\text{observed} = \text{matrix(\text{observed}, nrow=2)}
\]

```
observed
 [,1] [,2]
[1,] 315 108
[2,] 101 32
```

Fisher’s exact test (p. 308) can take such a matrix as its sole argument:

\[
\text{fisher.test(\text{observed})}
\]
**Fisher’s Exact Test for Count Data**

- **data:** observed
- **p-value:** 0.819
- **alternative hypothesis:** true odds ratio is not equal to 1
- **95 percent confidence interval:**
  - 0.5667874 to 1.4806148
- **sample estimates:**
  - **odds ratio:** 0.9242126

Alternatively we can use Pearson’s chi-squared test with Yates’ continuity correction:

```r
chisq.test(observed)
```

**Pearson’s Chi-squared test with Yates’ continuity correction**

- **data:** observed
- **X-squared:** 0.0513, df = 1, **p-value:** 0.8208

Again, the \( p \)-values are different with different tests, but the interpretation is the same: these pea plants behave in accordance with Mendel’s predictions of two independent traits, coat colour and seed shape, each segregating 3:1.

**Using Log-linear Models for Simple Contingency Tables**

It is worth repeating these simple examples with a log-linear model so that when we analyse more complex cases you have a feel for what the GLM is doing. Recall that the deviance for a log-linear model of count data (p. 516) is

\[
\text{deviance} = 2 \sum O \ln \left( \frac{O}{E} \right),
\]

where \( O \) is a vector of observed counts and \( E \) is a vector of expected counts. Our first example had 29 males and 18 females and we wanted to know if the sex ratio was significantly male-biased:

```r
observed<-c(29,18)
summary(glm(observed~1,poisson))
```

- **Null deviance:** 2.5985 on 1 degrees of freedom
- **Residual deviance:** 2.5985 on 1 degrees of freedom
- **AIC:** 14.547
- **Number of Fisher Scoring iterations:** 4

Only the bottom part of the summary table is informative in this case. The residual deviance is compared to the critical value of chi-squared in tables with 1 d.f.:

```r
1-pchisq(2.5985,1)
```

[1] 0.1069649

We accept the null hypothesis that the sex ratio is 50:50 (\( p = 0.10696 \)).

In the case of Mendel’s peas we had a four-level categorical variable (i.e. four phenotypes) and the null hypothesis was a 9:3:3:1 distribution of traits: