21.4: Analysis of proportions

4.1 Confidence interval for a single proportion

Methods appropriate for the analysis of proportions are used when the outcome of interest is a binary (‘yes/no’) variable (for example, the proportion of individuals who develop a disease). The standard error of a proportion \( p \), calculated from a sample of \( n \) subjects, is estimated as \( \sqrt{[p(1-p)/n]} \). For example, if the prevalence of splenomegaly in a random sample of 200 children from a population is found to be 0.40 (40%) (i.e. 80 had splenomegaly), the standard error (SE) is given by:

\[
\text{SE}(p) = \sqrt{(0.40 \times 0.60 / 200)} = 0.035 (3.5\%)
\]

The 95% CI for a proportion is given by \( p \pm 1.96 \times \text{SE}(p) \). In the example, the 95% CI is 0.4 ± 1.96 × 0.035 or (0.33, 0.47), i.e. 33–47%. There is a 95% chance that the true prevalence of splenomegaly in the population from which the sample of 200 was taken was between 33% and 47%.

4.2 Difference between two proportions

Suppose now that the objective is to compare the proportions observed in two groups of individuals, as is typically the case in a trial, comparing outcomes in an intervention and control group. The standard error of the difference between two proportions \( p_1 \) and \( p_2 \), based on \( n_1 \) and \( n_2 \) observations, respectively, is estimated approximately as:
\[
\sqrt{\langle \left\{ \overline{\mathrm{p}}(1-\overline{\mathrm{p}})\left[\left(1 / n_{1}\right)+\left(1 / n_{2}\right)\right]\right\} }
\]

where \( \bar{p} = \frac{(n_1 p_1 + n_2 p_2)}{(n_1 + n_2)} \).

For example, if the proportions to be compared are 90/300 (30%) and 135/300 (45%), the observed difference between the two proportions is \(-0.15\), \(\bar{p} = 0.375\), and the standard error of the difference is given by:

\[
\sqrt{\left\{ (0.375 \times 0.625)((1 / 300)+(1 / 300))\right\}}=0.040
\]

The 95% CI for the difference between the proportions is given by \((p_1 - p_2) \pm 1.96 \times SE\). In the example, this gives \((-0.15) \pm 1.96(0.040)\), i.e. \((-0.23, -0.07)\), or \(-23\%\) to \(-7\%\).

To test the null hypothesis that there is no true difference between the two proportions, the data are first arranged in a 2 \times 2 table, as in Table 21.2.

In the table, \(a\) is the number in group 1 who experiences the outcome of interest. The expected value of \(a\), \(E(a)\), and the variance of \(a\), \(V(a)\), are calculated under the hypothesis of no difference between the two groups:

\[
E(a)=\text{mean}_{\{1\}} \times \text{mean}_{\{1\}} / N
\]

<table>
<thead>
<tr>
<th>Group</th>
<th>Outcome</th>
<th>Total</th>
<th>Proportion with outcome</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Yes</td>
<td>No</td>
<td></td>
</tr>
<tr>
<td>1</td>
<td>(a) (90)</td>
<td>(b) (210)</td>
<td>(n_1) (300)</td>
</tr>
<tr>
<td>2</td>
<td>(c) (135)</td>
<td>(d) (165)</td>
<td>(n_2) (300)</td>
</tr>
<tr>
<td>Total</td>
<td>(m_1) (225)</td>
<td>(m_2) (375)</td>
<td>(N) (600)</td>
</tr>
</tbody>
</table>

\[
V(a)=\text{mean}_{\{1\}} \times \text{mean}_{\{2\}} \times \text{mean}_{\{1\}} / \text{mean}_{\{2\}} / (N-1)
\]

The chi-squared \(\chi^2\) statistic is then calculated. This gives a measure of the extent to which the observed data differ from those expected if the two proportions were truly equal:

\[
\chi^2=\left(\left|a-E(a)\right|-0.5\right)^2 / V(a)
\]

where \(|a-E(a)|\) indicates the absolute value of \(a-E(a)\).
The calculated value of $\chi^2$ is compared with tables of the chi-squared distribution with one degree of freedom (df). If it exceeds 3.84, then $p < 0.05$, indicating some evidence of a real difference in the proportions. If it exceeds 6.63, then $p < 0.01$, and there is strong evidence of a difference.

In the example, $a = 90$; $E(a) = (225)(300)/600 = 112.50$; and $V(a) = (300 \times 300 \times 22522 \times 375)/(600 \times 600 \times 599) = 35.215$. Thus $|\chi^2|=|90−112.50|−0.5)/2/35.212=13.74)$. From tables of the chi-squared distribution, a $p$-value of 0.0002 is obtained, indicating a difference as large as that observed would be very unlikely to arise by chance if there really was no difference between the two groups.

If any of the quantities $E(a)$, $E(b)$, $E(c)$, or $E(d)$ (for example, $(E(b)=m_2n_2/N)$) are less than 5.0 and $N$ is less than 40, the $\chi^2$ test is invalid, and a test called 'Fisher’s exact test' should be used instead (Kirkwood and Sterne, 2003).

### 4.3 Ratio of two proportions

The ratio of two proportions is sometimes referred to as the relative risk ($R$). To construct a CI for a relative risk, the natural logarithm of the estimate of the relative risk is computed (Table 21.2):

\[
\log_e(R)=\log_e\left(\frac{p_1}{p_2}\right)=\log_e\left[\frac{a/n_1}{c/n_2}\right]
\]

Its standard error is estimated by:

\[
\text{SE}[\log_e(R)]=\sqrt{\frac{b}{a n_1}+\frac{d}{c n_2}}
\]

The 95% CI for loge($R$) is given by $\log_e(R)\pm 1.96 \times \text{SE}$, and the 95% CI for the relative risk is obtained by taking anti-logarithms.

In the example given in Table 21.2, the relative risk is estimated as $0.30/0.45 = 0.667$, and $\log_e(R)=-0.405$. The SE[loge($R$)] is estimated as $\sqrt{[210/(90 \times 300)] + [165/(135 \times 300)]} = 0.109$, and the 95% CI for loge($R$) is given by $-0.405 \pm 1.96(0.109)$, i.e. $(-0.619, -0.191)$. Taking anti-logarithms, the 95% CI for the relative risk is $(0.538, 0.826)$.

### 4.4 Trend test for proportions

Sometimes, it is of interest to examine whether there is a trend in a series of proportions associated with different levels of some underlying characteristic. For example, consider the proportion of leprosy patients who report regularly to collect their monthly drug supply from a clinic when the accessibility of the clinic is rated as very poor, poor, fair, or good (Table 21.3).

A ‘score’ ($x_i$) is assigned for each kind of clinic, of which the value relates to the level of accessibility. For example, ‘0’ has been assigned to those with ‘very poor’ accessibility and ‘3’ to those with ‘good’ accessibility. A test for the trend in
the proportions $\{a_1/n_1, a_2/n_2, a_3/n_3, a_4/n_4\}$ is provided by testing, as a chi-squared with one df, the expression:

$$
\chi^2 = N \left\{ \left( N \sum a_i x_i \right) - \left( A \sum n_i x_i \right) \right\}^2 /
\left\{ A(N-A) \left\{ \left( N \sum n_i x_i^2 \right) - \left( \sum n_i x_i \right)^2 \right\} \right\}
$$

For example, suppose the data are as shown in Table 21.3 (the respective percentages of regular attenders in the four rows are 20%, 30%, 50%, and 60%). The value of $\chi^2$ is:

$$
\frac{150(150 \times 125) - (63 \times 245)]^2}{63 \times 87 [(150 \times 555) - 245^2]} = 12.95
$$

which is highly significant ($p = 0.0003$), based on a $\chi^2$ test with one df. It may be concluded therefore that there is strong evidence that the regularity of drug collection increases with the accessibility of the clinic.

<table>
<thead>
<tr>
<th>Accessibility of clinic</th>
<th>Collection of drugs</th>
<th>Total</th>
<th>'Score' $x_i$</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Regular</td>
<td>Not regular</td>
<td>$n_1$</td>
</tr>
<tr>
<td>Very poor</td>
<td>$a_1$ (5)</td>
<td>$n_1 - a_1$ (20)</td>
<td>$n_1$ (25)</td>
</tr>
<tr>
<td>Poor</td>
<td>$a_2$ (12)</td>
<td>$n_2 - a_2$ (28)</td>
<td>$n_2$ (40)</td>
</tr>
<tr>
<td>Fair</td>
<td>$a_3$ (25)</td>
<td>$n_3 - a_3$ (25)</td>
<td>$n_3$ (50)</td>
</tr>
<tr>
<td>Good</td>
<td>$a_4$ (21)</td>
<td>$n_4 - a_4$ (14)</td>
<td>$n_4$ (35)</td>
</tr>
<tr>
<td>Total</td>
<td>$A$ (63)</td>
<td>$N - A$ (87)</td>
<td>$N$ (150)</td>
</tr>
</tbody>
</table>