An evolutionary model predicts that the risk of violence in parent-offspring conflict depends on the ages, gender, and other characteristics of the perpetrator and victims. The evolutionary basis for these differences is supposed to be related to parental investment in progeny. Daly and Wilson\(^1\) argue that these predictions are supported by the analysis of data on infanticides, parricides, and filicides. Their data on 525 homicides committed by parents on offspring are categorized in Table 1.

<table>
<thead>
<tr>
<th>Gender of parent/child</th>
<th>Infantile (0-1)</th>
<th>Oedipal (2-5)</th>
<th>Latency (6-10)</th>
<th>Circumpubertal (11-16)</th>
<th>Adult (≥ 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/Male</td>
<td>24</td>
<td>21</td>
<td>21</td>
<td>29</td>
<td>104</td>
</tr>
<tr>
<td>Male/Female</td>
<td>17</td>
<td>27</td>
<td>10</td>
<td>14</td>
<td>47</td>
</tr>
<tr>
<td>Female/Male</td>
<td>53</td>
<td>21</td>
<td>19</td>
<td>9</td>
<td>8</td>
</tr>
<tr>
<td>Female/Female</td>
<td>50</td>
<td>27</td>
<td>5</td>
<td>4</td>
<td>15</td>
</tr>
</tbody>
</table>

Another (graphical) view of the data is provided by a mosaicplot (Figure 1). The R function is called `mosaicplot`. The rectangles in the mosaicplot are proportional to the count in each cell of the table. It’s apparent that there is significant variation among counts. For example, the count in the male/male and adult cell is about twice as large as any other count.

Each row of the contingency table is represented by a column in the mosaicplot. Since the mosaicplot columns are the same height, the mosaicplot displays the relative frequencies of the cell values within a table row (and, equivalently, within a mosaicplot column). The width of the mosaicplot columns indicates the relative frequencies of each row total (in the table) compared to the total number of observations.

Three questions are of primary interest:

1. Ignoring the age of the victims, are there gender differences among parent/child homicides? It appears so, as male/male homicides appear to be substantially more common than female/male or female/female.

2. Ignoring gender composition, are there age differences among the victims of parent/child homicides? It appears so, as there are differences in the width of the bars, implying that there are differences in the gender combinations after summing over ages.

3. Are age and gender composition independent, or does the gender composition depend on the age of the victim? (Equivalently, does the age of the victim depend on the gender composition?). Once again, yes, because when columns are compared, the lengths of the blocks for specific gender combinations are much different.

To address the first question, the marginal distribution of homicides as a function of gender composition is shown in Table 2.

<table>
<thead>
<tr>
<th>Parent/child</th>
<th>Male/male</th>
<th>Male/female</th>
<th>Female/male</th>
<th>Female/female</th>
</tr>
</thead>
<tbody>
<tr>
<td>Count</td>
<td>199</td>
<td>115</td>
<td>110</td>
<td>101</td>
</tr>
</tbody>
</table>

The question of whether there are gender differences amounts to investigating whether the data support the position that there are gender differences versus the position that there are no gender differences.

A test of significance will state the two positions as null and alternative hypotheses.

*The Homogeneity Model*

The homogeneity model is a convenient null hypothesis model which states that there are equal expected numbers of observations across all cells. In other words, the expected cell counts are homogeneous.

According to the homogeneity model, if there are $n$ observations, they will be distributed homogeneously among cells, and if there are $c$ cells, then it is expected that there will be $n/c$ observations in each. Of course, if the data are a sample from a population, then sampling variation will surely produce variation among cells with respect to the observed counts. The extent to which the observed counts diverge from the expected count ($n/c$) amounts to evidence against homogeneity.
The alternative hypothesis \((H_a)\) states that the expected cell counts are not homogenous (and that at least two expected cell counts are different).

Let \(n = \Sigma j n_j\) denote the total number of observations where \(n_j\) denotes the count in cell \(j = 1, \ldots, c\).

The expected count in cell \(j\) is \(e_j = n/c\).

Perhaps the differences between observed and expected counts (e.g., \(n_1 = 199\) versus \(e_1 = 131.25\)) can be attributed to sampling variation; if not, then the differences are sufficiently large to constitute significant evidence against the independence model. A test will be developed more generally for the case of a \(r \times c\) table of counts.

The Independence Model

The independence model is an extension of the homogeneity model to more than one variable. For the homicide study, the independence model specifies that gender composition is independent of age. Thus, the distribution of expected counts across gender composition will be the same in every age group row.

A formal test of the independence model uses the generic hypotheses

\[
H_0 : \text{the row and column variables variables are independent} \\
H_a : \text{the row and column variables variables are not independent.}
\]

Specific (and equivalent) hypotheses are

\[
H_0 : \text{The age distribution of homicides does not depend on gender composition.} \\
H_a : \text{The age distribution of homicides does depend on gender composition.}
\]

and

\[
H_0 : \text{The gender composition distribution of homicides does not depend on age.} \\
H_a : \text{The gender composition distribution of homicides does depend on age.}
\]

The chi-square test statistic

The test statistic is a chi-square test statistic, and it’s distribution is approximately chi-square. Chi-square distributions are (at least informally) very much like \(F\)-distributions except that there is a single parameter called the degrees of freedom instead of the two parameters of the \(F\)-distributions (numerator and denominator degrees of freedom).
The distribution of the chi-square test statistic is approximately chi-square only if the null hypothesis is true. If the alternative hypothesis is true, then the statistic will be larger than expected. Hence, the p-value will be the probability of obtaining a value for the chi-square test statistic as large or larger than was observed. The probability is a right-tail area.

Let
- \( r \) and \( c \) denote the number of rows and columns respectively,
- \( n_{ij} \) denote the observed count in row \( i \) and column \( j \),
- \( e_{ij} \) denote the expected count in row \( i \) and column \( j \).

The statistic contrasts the expected and observed counts:

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}};
\]

The expected cell counts are computed assuming the null hypothesis of independence between rows and columns is correct (discussed in more detail below).

Let \( \chi^2 \) denote the observed value of the statistic and \( \chi^2_{\text{df}} \) denote the chi-square random variable with \( \text{df} \) degrees of freedom. Then,

\[
p\text{-value} = P(\chi^2_{\text{df}} \geq \chi^2),
\]

since the distribution of \( \chi^2 \) is approximately chi-square with \( \text{df} = (r - 1)(c - 1) \) degrees of freedom if \( H_0 \) is true, i.e., \( \chi^2 \sim \chi^2_{(r-1)(c-1)} \).

Returning to the question of whether there are gender composition differences in the homicide study, the following table shows the components of the test statistic.

<table>
<thead>
<tr>
<th>Gender</th>
<th>( n_i )</th>
<th>( e_i )</th>
<th>( \frac{(n_i - e_i)^2}{e_i} )</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/male</td>
<td>199</td>
<td>131.25</td>
<td>34.97</td>
</tr>
<tr>
<td>Male/female</td>
<td>115</td>
<td>131.25</td>
<td>2.01</td>
</tr>
<tr>
<td>Female/male</td>
<td>110</td>
<td>131.25</td>
<td>3.44</td>
</tr>
<tr>
<td>Female/female</td>
<td>101</td>
<td>131.25</td>
<td>6.97</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td></td>
<td></td>
<td>( \chi^2 = 47.39 )</td>
</tr>
</tbody>
</table>

The degrees of freedom are \( \text{df} = (r - 1)(c - 1) = 3 \times 1 \), and the area to the right of 47.4 and under the \( \chi^2_3 \) distribution (i.e., \( P(\chi^2_3 \geq 47.4) \)) is extremely small, and so p-value < .0001. The R function call \( 1-pchisq(47.4, \text{df}=1) \) computes the area to the right and under the
chi-square distribution with 3 degrees of freedom.

There is convincing evidence that the null hypothesis is false and that there are gender differences. The conclusion is that homicide and gender combination are not independent, and that among the four combinations of gender, homicides are more frequent in the male/male combination.

_Needed conditions for the chi-square goodness-of-fit test_ For the computed p-value to be accurate, several hypotheses must be satisfied.

1. The counts are independent. In other words, every event (in this example, a parent/child homicide) occurred independently of all others\(^2\).

2. The counts were obtained from a random (or minimally, a representative) sample from the population of interest. In this study, the homicides that have been tabulated are assumed to be a representative sample from some population (in this case, all parent/child homicides within some socially defined population or strata).

3. The expected cell counts should be at least 5 in every cell of the table\(^3\).

_Comparing observed distributions_

Let’s return to Table 1 examine the gender distributions within each age group. The objective is to answer the question of whether the gender differences are consistent across age classes, or whether gender differences depend on the age class. The first step is to look at the conditional distributions by age class (Table 4).

The conditional distribution of gender composition is obtained by looking only at the data from one age group, and determining the relative frequencies of gender composition across that age group. The term condition arises from having imposed a condition on the data: it must originate from one particular age group.

The adult conditional distribution of gender composition looks much different than the infantile and oedipal conditional distributions (Table 4).

\(^2\)A reasonable assumption in this example.
\(^3\)The expected cell counts are all 131.25.
Table 4: Conditional homicide distributions, by age class.

<table>
<thead>
<tr>
<th>Gender of parent/child</th>
<th>Age Classification of the Child</th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Infantile (0-1)</td>
<td>Oedipal (2-5)</td>
<td>Latency (6-10)</td>
<td>Circumpubertal (11-16)</td>
<td>Adult (≥ 17)</td>
</tr>
<tr>
<td>Male/Male</td>
<td>.17</td>
<td>.22</td>
<td>.38</td>
<td>.52</td>
<td>.60</td>
</tr>
<tr>
<td>Male/Female</td>
<td>.12</td>
<td>.28</td>
<td>.18</td>
<td>.25</td>
<td>.27</td>
</tr>
<tr>
<td>Female/Male</td>
<td>.37</td>
<td>.22</td>
<td>.34</td>
<td>.16</td>
<td>.05</td>
</tr>
<tr>
<td>Female/Female</td>
<td>.35</td>
<td>.28</td>
<td>.09</td>
<td>.07</td>
<td>.09</td>
</tr>
</tbody>
</table>

Figure 1 shows a set of barcharts, organized by age class of the victim. A brief inspection shows that there are large differences in the distribution of counts across gender combination for different age classes. For example, the distribution across gender is relatively uniform in the oedipal age class; in contrast, among the adult age class, most of the homicides are male/male (104) and few are female/male (8).

Figure 1: Homicides classified by gender combination (perpetrator/victim) and age class of victim.
The R function call that produced the figure is `barchart(y~gender|Age, as.table=T, scales=list(rot=45),ylab="Number of homicides`). Figure 1 provides convincing evidence of differences among conditional distributions. A formal test of significance requires only a method of computing the expected cell counts assuming the null hypothesis of independence between column and row variables is correct. The alternative hypothesis is that the conditional distributions are different (not homogeneous).

Before laying out a test comparing the conditional distributions, some notation and comments are needed.

1. Let \( n_i = \sum_j n_{ij} \) denote the row \( i \) total, and \( n_j = \sum_i n_{ij} \) denote the column \( j \) total.

2. The total count is \( n = \sum_i \sum_j n_{ij} \).

3. The null hypothesis states that the distribution of counts is homogeneous across the age classes, and the alternative states that the distribution of counts is not homogeneous across the age classes.

4. The expected number of counts in row \( i \) and column \( j \), \( (e_{ij}) \) is the total number of counts \( n \) multiplied by the probability of belonging to row \( i \) and column \( j \). Suppose that \( y \) represents one of the homicides, and \( R_i \) is the set of row \( i \) homicides, and \( C_j \) is the set of column \( j \) homicides. The expression \( y \in R_i \cap C_j \) states that \( y \) belongs to \( R_i \) and \( C_j \).

According to the independence model, the expected number of observations in \( R_i \cap C_j \) is

\[
e_{ij} = nP(y \in R_i \cap C_j) = nP(y \in R_i)P(y \in C_j).
\]

The justification for setting \( P(y \in R_i \cap C_j) = P(y \in R_i)P(y \in C_j) \) is the independence hypothesis.\(^4\)

5. If \( P(y \in R_i) \) and \( P(y \in C_j) \) are unknown, then the estimates are \( \hat{P}(y \in R_i) = n_i/n \) and \( \hat{P}(y \in C_j) = n_j/n \) where \( n_i \) and \( n_j \) are the totals for row \( i \) and column \( j \). Furthermore, given the row and column totals, and that the counts are truly distributed homogeneously, then the probability that \( y \) belongs to row \( i \) is the proportion of counts

\(^4\)Recall if two events \( A \) and \( B \) are independent, then the probability that both occur is the product of the individual probabilities of occurrence: \( P(A \text{ and } B) = P(A)P(B) \). For example, the probability of getting a head when tossing a coin is \( P(A) = .5 \), and the probability of getting a head on a second toss is \( P(B) = .5 \). The probability that two heads occur is \(.25 = .5 \times .5\).
in row \(i\), and the probability that \(y\) belongs to column \(j\) is the proportion of counts in column \(j\). Thus, the expected number in row \(i\) and column \(j\) (given \(n_i\) and \(n_j\)) is

\[
e_{ij} = n \times \frac{n_i}{n} \times \frac{n_j}{n} = \frac{n_i n_j}{n}.
\]

The test statistic is the chi-square statistic. It contrasts the expected and observed counts:

\[
\chi^2 = \sum_{i=1}^{r} \sum_{j=1}^{c} \frac{(n_{ij} - e_{ij})^2}{e_{ij}}.
\]

The p-value is developed from the following:

- The distribution of \(\chi^2\) is approximately chi-square with \(df = (r - 1)(c - 1)\) degrees of freedom if \(H_0\) is true, where \(r\) and \(c\) are the numbers of rows and columns in the table.
- If \(H_0\) is not true, then the \(\chi^2\) statistic will tend to be unusually large compared to the chi-square distribution.
- The evidence against \(H_0\) and in favor of \(H_a\) is measured by the probability of obtaining a chi-square statistic as large or larger than was observed.
- If \(\chi^2\) is the observed value of the statistic and \(\chi^2_{df}\) represents the chi-square random variable, then

\[
p\text{-value} = P(\chi^2_{df} \geq \chi^2).
\]

**Example** Returning to the personal conflict example, Table 5 shows the expected counts.

<table>
<thead>
<tr>
<th>Gender of parent/child</th>
<th>Infanticile (0-1)</th>
<th>Oedipal (2-5)</th>
<th>Latency (6-10)</th>
<th>Circumpubertal (11-16)</th>
<th>Adult (≥ 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/Male</td>
<td>54.6</td>
<td>36.4</td>
<td>20.8</td>
<td>21.2</td>
<td>66.0</td>
</tr>
<tr>
<td>Male/Female</td>
<td>31.5</td>
<td>21.0</td>
<td>12.0</td>
<td>12.3</td>
<td>38.1</td>
</tr>
<tr>
<td>Female/Male</td>
<td>30.2</td>
<td>20.1</td>
<td>11.5</td>
<td>11.7</td>
<td>36.5</td>
</tr>
<tr>
<td>Female/Female</td>
<td>27.7</td>
<td>18.5</td>
<td>10.6</td>
<td>10.8</td>
<td>33.5</td>
</tr>
</tbody>
</table>

The test statistic is \(\chi^2 = 143.8\), and the p-value is less than .0001. There is overwhelming evidence that distributions are not homogeneous. The next question is where are the greatest deviations from homogeneity? It’s somewhat awkward to compare the expected and observed counts with the intent of finding those cells that deviate the most from observed because the squared deviations are scaled by the expected counts.
A better way is to compute the Pearson cell residuals

\[ v_{ij} = \frac{y_{ij} - e_{ij}}{\sqrt{e_{ij}}}. \]

Note that \( \chi^2 = \sum\sum v_{ij}^2 \). The cell residuals are similar to standard normal variates in distribution. Hence, values between \(-1\) and 1 are commonplace and values less than \(-2\) and greater than 2 are unusually large departures from the independence model.

Table 6 shows the cell residuals. Note that there is a pattern: large residuals are present in the upper right and the lower left of the table, and negative and large in magnitude residuals are present in the upper left and lower right. It can be concluded that male/male homicides are more likely as age of the child increases, and that mothers are more likely to murder the youngest children.

<table>
<thead>
<tr>
<th>Gender of parent/child</th>
<th>Age Classification of the Child</th>
<th>Infantine (0-1)</th>
<th>Oedipal (2-5)</th>
<th>Latency (6-10)</th>
<th>Circumpubertal (11-16)</th>
<th>Adult (≥ 17)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male/Male</td>
<td></td>
<td>−4.1</td>
<td>−2.6</td>
<td>0.0</td>
<td>1.7</td>
<td>4.7</td>
</tr>
<tr>
<td>Male/Female</td>
<td></td>
<td>−2.6</td>
<td>1.3</td>
<td>−0.6</td>
<td>0.5</td>
<td>1.4</td>
</tr>
<tr>
<td>Female/Male</td>
<td></td>
<td>4.2</td>
<td>0.2</td>
<td>2.2</td>
<td>−0.8</td>
<td>−4.7</td>
</tr>
<tr>
<td>Female/Female</td>
<td></td>
<td>4.2</td>
<td>2.0</td>
<td>−1.7</td>
<td>−2.1</td>
<td>−3.2</td>
</tr>
</tbody>
</table>

The R function call `summary(xtabs(Freq~.,data=as.data.frame.table(X)))` computes the chi-square statistic. In this example, `X` is the table of counts, `as.data.frame.table(X)` takes the counts out of tabular form and into a `data.frame` and creates a column with the counts that is called `Freq`. `summary(xtabs)` computes and reports the statistic.

The test statistic above compared groups (age classes) across 4 sub-populations (Male/male, ..., Female/female) to determine if the distribution of homicides were homogeneous. A very similar situation leads to the same test in the investigation of whether two or more categorical variables are independent.

Independence between variables (categorical or quantitative) implies that if both variables are observed on a single sample or experimental unit, then the observed value of one variable does not depend on the value of the other.

Independence between events A and B implies that the knowing that event A has occurred
provides no information about whether event $B$ occurred (e.g., if $A$ is the outcome that a sampled voter is female, the event $B$, Republican primary candidate preference, are independent if knowing that the voter is female provides no information whether the voter’s preference is Romney versus Santorum).

When variables are categorical, independence can be interpreted both ways, since the realization of a categorical variable is an event. For instance, if the categorical variable observed on sampled individuals is candidate preference, then event $A$ might be gender of the sampled individual and event $B$ is preference for Santorum versus Romney. If knowing that the voter is female does not alter the probability of preferring Santorum, then voter preference and gender are independent. Independence probably is not the case in this case. Gary Langer\textsuperscript{5} wrote \textit{Exit poll results found an exact even split between Santorum and Romney among men. By contrast, Romney held an 11-point lead over Santorum among women in Ohio.}

A Danish study conducted in 1983 and 1985 examined opinions of joint sports activities (taking place with the other gender) among high school students. These data provide a second example, and are available in the \texttt{vcd} library. The data set name is \texttt{JointSports}. Study year will be ignored, and the objective is to determine whether genders differ with respect to opinions towards joint sports, and also to determine if opinions shift from grades 1 to 3 (presumably grade 1 corresponds to sophomore in the American school system and grade 3 corresponds to senior).

The function calls

\begin{verbatim}
    tab <- xtabs(Freq ~ gender + opinion + grade, data = JointSports)
    doubledecker(opinion ~ gender + grade, data = tab)
\end{verbatim}

produce the stacked barcharts (\texttt{doubledecker} is available from the \texttt{vcd} library). The \texttt{tab} function constructs a three-way table necessary for producing the barchart. Table 7 below show the three-way table.

\begin{table}
\centering
\begin{tabular}{|c|c|c|c|c|}
\hline
\textit{grade} & \textit{gender} & \textit{opinion} \\
\hline
1st & Boy & very bad & 10 & 20 \\
 & & bad & 30 & 40 \\
 & & indifferent & 50 & 60 \\
 & & good & 70 & 80 \\
 & & very good & 90 & 100 \\
2nd & Girl & very bad & 110 & 120 \\
 & & bad & 130 & 140 \\
 & & indifferent & 150 & 160 \\
 & & good & 170 & 180 \\
 & & very good & 190 & 200 \\
\hline
\end{tabular}
\end{table}

\textsuperscript{5}Mar 6, 2012, \textit{Santorum Struggles With Female Voters In Ohio Primary, Exit Polls Show}, ABC News Blog.
The bar chart reveals that there are substantially fewer boys than girls in the study (510 versus 761) since the bar widths are noticeably different. There is some suggestion that boys are more negative in their opinion toward joint participation than girls, and it appears that there is a shift in opinion between grades among girls, and perhaps also among boys.

Table 7 below can be thought of as a three-way table (obtained from three variables) or as two conditional tables. The conditional table view constructs the left table (boys) by imposing the condition that only boys are used, and constructs the right table (girls) by imposing the condition that only girls are used. The function (part of the vcd library) call was `co_table(tab, margin=1).

Table 7: Distribution of opinions about joint sports participation by gender and grade.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Boy</th>
<th>Girl</th>
</tr>
</thead>
<tbody>
<tr>
<td>1st</td>
<td>72</td>
<td>180</td>
</tr>
<tr>
<td>3rd</td>
<td>54</td>
<td>113</td>
</tr>
<tr>
<td>very good</td>
<td>118</td>
<td>56</td>
</tr>
<tr>
<td>good</td>
<td>73</td>
<td>67</td>
</tr>
<tr>
<td>indifferent</td>
<td>22</td>
<td>25</td>
</tr>
<tr>
<td>bad</td>
<td>22</td>
<td>9</td>
</tr>
<tr>
<td>very bad</td>
<td>6</td>
<td>5</td>
</tr>
</tbody>
</table>

The test for independence of grade and opinion using only the girls data produced $\chi^2 = 11.57$, $df = 4$, p-value = .021. Using the boys data, the test produced $\chi^2 = 5.85$, $df = 4$, and p-value = .210. There’s evidence of a shift in girls attitude between grades 1 and 3, but there’s very little evidence of a shift among the boys.

Differences among gender with respect to opinion should be examined by conditioning on grade and carrying out two tests of significance (rather than using both grades in the same test.) It’s necessary to condition since there is evidence that girls’ opinion shift with grade. (The logic behind conducting separate tests is the same as that which invalidates testing for main effects if interaction is found in the analysis of variance).
By conditioning on grade, differences between grades do not confound the comparisons. Using 1st grade only, a test of independence produced $\chi^2 = 31.79$, df = 4 and p-value < .001. Using 3rd grade only, a test of independence produced $\chi^2 = 17.48$, df = 4 and p-value = .0016. There’s convincing evidence of gender differences. The Pearson residuals shown to the left for grade 1 is display using the commands

```
xt <- xtabs(Freq ~ opinion + gender , data = JointSports, subset=(grade=="1st"))
mosaic(xt,shade=T).
```

The figure indicates that the largest differences between observed and expected counts are in the indifferent category where a greater proportion of boys than females are indifferent and in the very good category where a greater proportion of girls have a very good opinion of joint sports than boys.