7.1 For table 2.1 on X = gender and Y = belief in an afterlife, Table 7.16 shows the results of fitting the independence loglinear model.


| Table 7.16. Computer Output for Problem 7.1 on Belief in Afterlife |
|--------------------------|-----------|-------------|
| Criteria For Assessing Goodness Of Fit |
| Criterion     | DF | Value       |
| Deviance      | 1  | 0.8224      |
| Pearson Chi-Square | 1 | 0.8246      |

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Std. Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intercept</td>
<td>1</td>
<td>4.5849</td>
<td>0.0752</td>
</tr>
<tr>
<td>gender females</td>
<td>1</td>
<td>0.2192</td>
<td>0.0599</td>
</tr>
<tr>
<td>gender males</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>belief yes</td>
<td>1</td>
<td>1.4165</td>
<td>0.0752</td>
</tr>
<tr>
<td>belief no</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Figure 1:

From Table 7.16, we obtain the results of the goodness-of-fit test. We see that for our deviance and chi-squares terms:

<table>
<thead>
<tr>
<th>Model</th>
<th>Degrees of Freedom</th>
<th>Value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Deviance</td>
<td>1</td>
<td>0.8224</td>
</tr>
<tr>
<td>Pearson Chi-Square</td>
<td>1</td>
<td>0.8246</td>
</tr>
</tbody>
</table>

For our deviance, we desire a value close to our number of degrees of freedom which as we can see occurs in this goodness-of-fit rather well. Similarly, we want the chi-square term to also come close to the degrees of freedom (the mean in this case) in order to show a good fit, which again we see here.

b. Report \(\hat{\lambda}_Y\). Interpret \(\hat{\lambda}_1^Y - \hat{\lambda}_2^Y\)

Further from Table 7.16 we see that the fourth parameter is the term \(\hat{\lambda}_1^Y = 1.4165\) and the fifth parameter is \(\hat{\lambda}_2^Y = 0\). Thus \(\hat{\lambda}_1^Y - \hat{\lambda}_2^Y = 1.4165\) which allows us to calculate the odds of belief in afterlife by \(e^{1.4165} = 4.123\) for either male or females.
7.2 For the saturated model with Table 2.1, software reports the \( \hat{\lambda}^{XY}_{ij} \) estimates:

<table>
<thead>
<tr>
<th>Parameter</th>
<th>DF</th>
<th>Estimate</th>
<th>Std Error</th>
</tr>
</thead>
<tbody>
<tr>
<td>gender*belief females yes</td>
<td>1</td>
<td>0.1368</td>
<td>0.1507</td>
</tr>
<tr>
<td>gender*belief females no</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>gender*belief males yes</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
<tr>
<td>gender*belief males no</td>
<td>0</td>
<td>0.0000</td>
<td>0.0000</td>
</tr>
</tbody>
</table>

Figure 2:

Show how to use these to estimate the odds ratio.

We are able to solve for the odds of a 2 way table with the following equation derived in class:

\[
\log \theta = \lambda^{XY}_{11} + \lambda^{XY}_{22} - \lambda^{XY}_{12} - \lambda^{XY}_{21} \\
= 0.1368 + 0 - 0 - 0 \\
= 0.1368
\]

Thus by simply exponentiating this value we obtain the odds ratio estimate as:

\[e^{0.1368} = 1.1466\]

7.8 Refer to the previous two exercises. PROC GENMOD in SAS reports the maximize log likelihood as 3475.19 for the model of mutual independence (df =11), 3538.05 for the model of homogeneous association (df = 5), and 3539.58 for the model containing all the three-factor interaction terms.

a. Write the loglinear model for each case, and show that the numbers of parameters are 5, 11, and 15

In case of Mutual Independence the model is:

\[
\log \pi_{i,j,k,l} = 3.79255 + 0.26439E + 0.870085S - 0.48551T - 0.12971J
\]

Here, the number of parameters are 5 including the intercept.

The homogeneous association model is:

\[
\log \pi_{i,j,k,l} = 3.4478 - 0.02907E + 1.21082S - 0.64194T + 0.93417J + 0.30212ES \\
+ 0.4092ST - 1022153SJ - 0.55936TJ + 0.19449ET + 0.01766EJ
\]
Here, number of parameters are 11 including the intercept.

The model containing all the three factor interaction terms is:

$$\log \pi_{i,j,k,l} = 3.5637 - 0.2788E + 1.05839S - 0.63483T + 0.76316J + 0.6146ES$$

$$+ 0.41081ST - 0.9629SP - 0.5877TP + 0.2003ET + 0.374EP$$

$$- 0.0236EST - 0.5104ESJ + 0.019STJ + 0.024ETJ$$

Here, number of parameters are 15.

b. According to AIC (see section 5.1.5), which of these models seems best? Why?

In the first model the AIC = 238.7, in the second model the AIC = 125 and in the third model the AIC = 129.93. So, according to AIC the homogeneous association model is the best among these three models. Because we know AIC is an important criteria by which we can compare two or more models. The model having a low AIC compared to all other model is considered to be the best the model.

7.19 Consider loglinear model (WXZ, WYZ).

a. Draw its independence graph, and identify variables that are conditionally independent.

```
  x w z y
```

Hence, X and Y are separated by (W), (Z), and (W,Z). So W and Z are independent given W alone, given Z alone, or given both W and Z. Further, X and Z are independent given W alone or W and Z alone. Finally, W and Y are independent if given Z alone or W and Y.

b. Explain why this is the most general loglinear model for a four-way table for which X and Y are conditionally independent.

We have seen, from the graph of independence, that X and Y are conditionally independent. All terms in the saturated model that are not in the model (WXZ, WYZ) involve X and Y, which permits XY as a conditional association.

7.20 For a four-way table, are X and Y independent, given Z alone, for model (a) (WX, XZ, YZ, WZ), (b) (WX, XZ, YZ, WY)?

In both model (a) and (b) there is no XY term therefore X and Y are conditionally independent.

7.22 Consider model (AC, AM, CM, AG, AR, GM, GR) for the drug use data in Section 7.4.5.
a. Explain why the AM conditional odds ratio is unchanged by collapsing over race, but it is not unchanged by collapsing over gender.

AM conditional odds ratio is unchanged by collapsing over race because Age and Marijuana use are both conditionally independent from Race. But A and M are not independent of Gender. Thus the AM conditional odds ratio is unchanged by collapsing over R but not when collapsing over G.

b. Suppose we remove the GM term. Construct the independence graph, and show that G, R are separated from C,M by A.

Hence, \( \{G, R\} \) are separated from \( \{C, M\} \) by A.

c. For the model in (b), explain why all conditional associations among A,C, and M are identical to those in model (AC, AM, CM), collapsing over G and R.

Consider the sets \( \{C\}, \{A, M\}, \{G, R\} \). For model, (b), every path between C and \( \{G, R\} \) involves a variable in \( \{A, M\} \). Given the outcome on alcohol use and marijuana use, the model states that cigarette use and Marijuana use are independent of both gender (G) and Race (R). Collapsing over the explanatory variables race and gender, the conditional associations between A and C, between C and M, and between A and M are the same as with the model (AC,AM,CM).