CS&SS/STAT/SOC 536: Three-way Contingency Tables

Adrian Dobra
adobra@uw.edu

1 Simpson's Paradox

Why do we need to look at contingency tables that involve more than two variables? The data in Table 1 shows the number of successes and failures of two treatments in a number of males and females. The three observed variables are binary: "Treatment" (1 or 2), "Outcome" (success or failure) and "Gender" (male or female).

		Gender			
		Male		Female	
	Outcome	Success	Failure	Success	Failure
Treatment	1	60	20	40	80
	2	100	50	10	30

Table 1: A $2 \times 2 \times 2$ table illustrating Simpson's paradox.

If we look at the complete $2 \times 2 \times 2$ table, we see that the probability of success of Treatment 1 in males is $\frac{60}{80} = 0.75$ and the probability of success of Treatment 2 in males is $\frac{100}{150} = 0.667$. Similarly, the probability of success of Treatment 1 in females is $\frac{40}{120} = 0.333$, while the probability of success of Treatment 2 in females is $\frac{10}{40} = 0.25$. Therefore, if we consider males and females *separately*, Treatment 1 seems to be *more effective* than Treatment 2.

Now let's look at the effectiveness of the two treatments in all the patients (that is, we ignore the gender of the patients). We look at the 2×2 table that cross-classifies "Treatment" and "Outcome" – see Table 2. This 2×2 table is obtained by collapsing Table 1 across "Gender". The probability of success of Treatment 1 in males and females is $\frac{100}{200} = 0.5$, while the probability of success of Treatment 2 in males and females is $\frac{110}{190} = 0.58$. Therefore Table 2 seems to indicate that the two treatments are *equally effective*.

This is precisely the reason why we need to investigate the relationship between "Treatment" and "Outcome" while taking "Gender" into account. Ignoring "Gender" could lead to unsound conclusions. In other words, we have analyze the three-way dataset in Table 1 instead of relying in a simpler analysis of the two-way dataset in Table 2.

	Outcome	Success	Failure
Treatment	1	100	100
	2	110	80

Table 2: The "Treatment" by "Outcome" marginal of Table 1.

2 Notations and the Color of Hair and Eyes Data

We go back to the example we discussed in the handout on two-way contingency tables. Table 3 cross-classifies the same individuals by gender in addition to their color of hair and eyes. We denote by X_1, X_2 and X_3 the three variables represented in this table, namely Gender, Hair color and Eye color. Gender takes two values: Male (coded with 1) and Female (coded with 2). Similarly, Hair color takes four values: Black, Brown, Red and Blond (coded with 1, 2, 3 and 4). Eye color also takes four values: Brown, Blue, Hazel and Green (coded with 1, 2, 3 and 4).

		Eye Color (X_3)			
Gender (X_1)	Hair Color (X_2)	Brown (1)	Blue (2)	Hazel (3)	Green (4)
Male (1)	Black (1)	32	11	10	3
	Brown (2)	53	50	25	10
	Red (3)	10	10	7	7
	Blond (4)	3	30	5	8
Female (2)	Black (1)	36	9	5	2
	Brown (2)	66	34	29	14
	Red (3)	16	7	7	7
	Blond (4)	4	64	5	8

Table 3: Color of hair and eyes data: an example of a $2 \times 4 \times 4$ table.

Table 3 is an example of a three-way $I \times J \times K$ table. Here $I = 2$, $J = 4$ and $K = 4$. The count in cell (i, j, k) of Table 3 is denoted by n_{ijk}. For example, $n_{111} = 32$, that is, we observed 32 males with back hair and brown eyes.

Table 3 has three two-dimensional (or two-way) marginal totals. The two-way marginals are obtained by collapsing Table 3 across Gender, Hair color or Eye color. For example, by collapsing Table 3 across Gender we obtain Table 1 from the handout on two-way tables. This is a 4×4 table that cross-classifies the samples by Hair color and Eye color. It is called the $(2, 3)$ marginal since it corresponds with variables X_2 and X_3. The counts in the $(2, 3)$ marginal are

$$\{n_{+jk} : 1 \leq j \leq J, 1 \leq k \leq K\},$$

where

$$n_{+jk} = \sum_{i=1}^{I} n_{ijk}$$

The count in the $(1, 1)$ cell of the $(2, 3)$ marginal is $n_{+11} = n_{111} + n_{211} = 32 + 36 = 68$.

By collapsing Table 3 across Hair color we obtain the $(1, 3)$ marginal which is a 2×4 table with counts

$$\{n_{i+k} : 1 \leq i \leq I, 1 \leq k \leq K\},$$

where

$$n_{i+k} = \sum_{j=1}^{J} n_{ijk}$$

Similarly, by collapsing Table 3 across Eye color we obtain the $(1, 2)$ marginal which is a 2×4 table with counts

$$\{n_{ij+} : 1 \leq i \leq I, 1 \leq j \leq J\},$$

where

$$n_{ij+} = \sum_{k=1}^{K} n_{ijk}$$

By collapsing Table 3 across Gender and Hair color we obtain the one-dimensional marginal associated with Eye color. The counts in this one-way table are

$$\{n_{++k} : 1 \leq k \leq K\},$$

where

$$n_{++k} = \sum_{i=1}^{I} \sum_{j=1}^{J} n_{ijk}$$

This one-way marginal can also be obtained by collapsing the $(1, 3)$ marginal across Gender or the $(2, 3)$ marginal across Hair color. The grand total is the sum of all the observed counts:

$$n_{+++} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk}$$

Table 3 has $I \cdot J \cdot K = 32$ cells. In order to represent this table in R we need to make sure we order the cells (i, j, k) such that the leftmost index varies fastest. That is, i changes first, then j, then k. The resulting ordering of the cells of Table 3 is

$$\{(1,1,1),(2,1,1),(1,2,1),(2,2,1),(1,3,1),(2,3,1),(1,4,1),(2,4,1),$$
$$(1,1,2),(2,1,2),(1,2,2),(2,2,2),(1,3,2),(2,3,2),(1,4,2),(2,4,2),$$
$$(1,1,3),(2,1,3),(1,2,3),(2,2,3),(1,3,3),(2,3,3),(1,4,3),(2,4,3),$$
$$(1,1,4),(2,1,4),(1,2,4),(2,2,4),(1,3,4),(2,3,4),(1,4,4),(2,4,4)\}$$

Note: by specifying the ordering above, you actually specify that the first variable is Gender, the second variable is Hair color and the third variable is Eye color. You also specify the numerical label of each category.

In R you create a vector of the observed counts arranged in the ordering explained before. You subsequently create a $2 \times 4 \times 4$ array from this vector:

```
 haireyecolor = c(32,36,53,66,10,16,3,4,11,9,50,34,10,7,30,64,
                          10,5,25,29,7,7,5,5,3,2,10,14,7,7,8,8)
 haireyecolor.array = array(haireyecolor,c(2,4,4))
 , , 1

     [,1] [,2] [,3] [,4]
[1,]   32   53   10    3
[2,]   36   66   16    4

, , 2

     [,1] [,2] [,3] [,4]
[1,]   11   50   10   30
[2,]    9   34    7   64

, , 3

     [,1] [,2] [,3] [,4]
[1,]   10   25    7    5
[2,]    5   29    7    5

, , 4

     [,1] [,2] [,3] [,4]
[1,]    3   10    7    8
[2,]    2   14    7    8
```

R gives the counts in the table sequentially in each 2×4 "slice" of Table 3 corresponding with each category of X_3 (Eye color).

We will follow notations that are consistent with the notations we introduced in the other handouts on contingency tables. We always use index i to denote the categories $\{1, 2, \ldots, I\}$ of X_1. Index j denotes the categories $\{1, 2, \ldots, J\}$ of X_2, while index k denotes the categories $\{1, 2, \ldots, K\}$ of X_3. We use log-linear models to represent associations in the joint distribution of X_1, X_2 and X_3. The cell probabilities of the full $I \times J \times K$ table are denoted by

$$P(X_1 = i, X_2 = j, X_3 = k) = p_{ijk}$$

The cell probabilities in the $(1, 2)$ marginal are

$$P(X_1 = i, X_2 = j) = p_{ij+} = \sum_{k=1}^{K} p_{ijk}$$

The cell probabilities in the $(1, 3)$ marginal are

$$P(X_1 = i, X_3 = k) = p_{i+k} = \sum_{j=1}^{J} p_{ijk}$$

The cell probabilities in the $(2, 3)$ marginal are

$$P(X_2 = j, X_3 = k) = p_{+jk} = \sum_{i=1}^{I} p_{ijk}$$

The cell probabilities in the marginal associated with X_1 are

$$P(X_1 = i) = p_{i++} = \sum_{j=1}^{J} \sum_{k=1}^{K} p_{ijk}$$

The expected cell value for cell (i, j, k) is denoted by m_{ijk}. The expected cell values in the $(1, 2)$ marginal are

$$m_{ij+} = \sum_{k=1}^{K} m_{ijk}$$

The expected cell values in the marginal associated with X_1 are

$$m_{i++} = \sum_{j=1}^{J} \sum_{k=1}^{K} m_{ijk}$$

It will always be true that the sum of the expected cell values is equal with the grand total

$$m_{+++} = \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} m_{ijk} = n_{+++}$$

The expected cell values are obtained by multiplying the cell probabilities by the grand total

$$m_{ijk} = n_{+++} \cdot p_{ijk}$$

For completeness, we write the formulas for the likelihood ratio test statistic G^2

$$\begin{aligned} G^2 &= 2 \sum_{\text{all cells}} (Observed) \log\left(\frac{Observed}{Expected}\right), \\ &= 2 \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} n_{ijk} \log\left(\frac{n_{ijk}}{m_{ijk}}\right) \end{aligned}$$

and for the X^2 test statistic

$$\begin{aligned} X^2 &= \sum_{\text{all cells}} \left(\frac{Observed - Expected}{\sqrt{Expected}}\right)^2, \\ &= \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \left(\frac{n_{ijk} - m_{ijk}}{\sqrt{m_{ijk}}}\right)^2 \end{aligned}$$

G^2 and X^2 are calculated for a log-linear model $\mathcal{M}$ by estimating the expected cell values m_{ijk} under $\mathcal{M}$, then replacing these estimates in the formulas above. Asymptotically (i.e., as the grand total n_{+++} goes to infinity), G^2 and X^2 follow a Chi-squared distribution with a number of degrees of freedom equal to the number of degrees of freedom of $\mathcal{M}$, denoted by $\#df(\mathcal{M})$. We say that $\mathcal{M}$ *fits the data* if the p-value based on the G^2 test statistic

$$P\left(\chi^2_{\#df(\mathcal{M})} \geq G^2\right)$$

or the p-value based on the X^2 test statistic

$$P\left(\chi^2_{\#df(\mathcal{M})} \geq X^2\right)$$

is greater than 0.05 (or 0.01).

3 The Saturated Log-linear Model [123]

We start with the most complex model for $I \times J \times K$ tables, namely the log-linear model that involves all the possible interactions among X_1, X_2 and X_3:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)} + u_{123(ijk)} \tag{1}$$

Any other log-linear model is obtained by setting to zero some of the u-terms in the saturated log-linear model in Equation (1). We constrain our log-linear models to be *hierarchical*, that is, we cannot set to zero higher order interaction terms *before* setting to zero lower order interaction terms. More explicitly, $u_{12(ij)}$, $u_{23(jk)}$ and $u_{13(ik)}$ are called *first order interaction terms* since they express the interaction of two variables. Similarly, $u_{123(ijk)}$ are called *second order interaction terms* since they express the interaction among three variables. One can informally view the main effects $u_{1(i)}$, $u_{2(j)}$ and $u_{3(k)}$ as *zero order* interaction terms since they are associated with one variable. In the context of three-way tables, a model is hierarchical if:

- if $u_{12(ij)} = 0$ then $u_{123(ijk)} = 0$.
- if $u_{23(jk)} = 0$ then $u_{123(ijk)} = 0$.
- if $u_{13(ik)} = 0$ then $u_{123(ijk)} = 0$.

For example, the log-linear model

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{123(ijk)}$$

is not hierarchical since $u_{12(ij)}$ was set to zero, but $u_{123(ijk)}$ was *not* set to zero. *Note: we will* **never** *set to zero u, $u_{1(i)}$, $u_{2(j)}$ and $u_{3(k)}$.* If we would set $u_{1(i)} = 0$, we would also need to set $u_{12(ij)} = 0$ and $u_{123(ijk)} = 0$ to keep the model hierarchical. This means our choice of models would be restricted to

$$\log m_{ijk} = u + u_{2(j)} + u_{3(k)} + u_{23(jk)}$$

or

$$\log m_{ijk} = u + u_{2(j)} + u_{3(k)}$$

These two models to not make sense since variable X_1 is not represented anywhere! *Note: when we write* $u_{1(i)} = 0$ *we mean* $u_{1(i)} = 0$ *for all* $i = 1, 2, \ldots, I$. *Similarly, by writing* $u_{12(ij)} = 0$ *we mean* $u_{12(ij)} = 0$ *for any* $i = 1, 2, \ldots, I$ *and any* $j = 1, 2, \ldots, J$, *and so on.*

We need to explain what is the interpretation of each set of u-terms, what are the constraints associated with it and how many free parameters are in each such set.

$\boxed{u}$ Due to the sum-to-zero constraints we impose on the other sets of u-terms, u is the mean of of logarithms of the expected counts:

$$u = \frac{1}{IJK} \sum_{i=1}^{I} \sum_{j=1}^{J} \sum_{k=1}^{K} \log m_{ijk}$$

This is only one parameter, there are no additional constraints associated with it, which means we count one free parameter when we calculate the degrees of freedom associated with a log-linear model.

$\boxed{\{u_{1(i)} : 1 \leq i \leq I\}}$ The u-term $u_{1(i)}$ represents the deviation from the grand mean u associated with category i of X_1. There are I such parameters since X_1 has I categories. We impose one sum-to-zero constraint:

$$\sum_{i=1}^{I} u_{1(i)} = 0$$

Thus we have $I - 1$ *free* u-terms in this set.

$\boxed{\left\{u_{2(j)} : 1 \leq j \leq J\right\}}$ The u-term $u_{2(j)}$ represents the deviation from the grand mean u associated with category j of X_2. There are J such parameters since X_2 has J categories. We impose one sum-to-zero constraint:

$$\sum_{j=1}^{J} u_{2(j)} = 0$$

Thus we have $J - 1$ *free* u-terms in this set.

$\boxed{\{u_{3(k)} : 1 \leq k \leq K\}}$ The u-term $u_{3(k)}$ represents the deviation from the grand mean u associated with category k of X_3. There are K such parameters since X_3 has K categories. We impose one sum-to-zero constraint:

$$\sum_{k=1}^{K} u_{3(k)} = 0$$

Thus we have $K - 1$ *free* u-terms in this set.

$\boxed{\left\{u_{12(ij)} : 1 \le i \le I, 1 \le j \le J\right\}}$ The u-term $u_{12(ij)}$ represents the deviation from $u + u_{1(i)} + u_{2(j)}$ associated with the interaction between category i of X_1 and category j of X_2. There are $I \cdot J$ such parameters. We impose the following $I + J$ sum-to-zero constraints:

$$\begin{aligned}\sum_{j=1}^{J} u_{12(ij)} &= 0, \text{ for } 1 \le i \le I,\\ \sum_{i=1}^{I} u_{12(ij)} &= 0, \text{ for } 1 \le j \le J\end{aligned}$$

Since both sets of constraints imply that the sum of all $u_{12(ij)}$ is zero, there are only $I + J - 1$ free constraints. Thus we have $IJ - (I + J - 1) = (I - 1) \cdot (J - 1)$ *free* u-terms in this set.

$\boxed{\{u_{13(ik)} : 1 \le i \le I, 1 \le k \le K\}}$ The u-term $u_{13(ik)}$ represents the deviation from $u + u_{1(i)} + u_{3(k)}$ associated with the interaction between category i of X_1 and category k of X_3. There are $I \cdot K$ such parameters. We impose the following (I+K) sum-to-zero constraints:

$$\begin{aligned}\sum_{k=1}^{K} u_{13(ik)} &= 0, \text{ for } 1 \le i \le I,\\ \sum_{i=1}^{I} u_{13(ik)} &= 0, \text{ for } 1 \le k \le K\end{aligned}$$

Since both sets of constraints imply that the sum of all $u_{13(ik)}$ is zero, there are only $I + K - 1$ free constraints. Thus we have $IK - (I + K - 1) = (I - 1) \cdot (K - 1)$ *free* u-terms in this set.

$\boxed{\left\{u_{23(jk)} : 1 \le j \le J, 1 \le k \le K\right\}}$ The u-term $u_{23(jk)}$ represents the deviation from $u + u_{2(j)} + u_{3(k)}$ associated with the interaction between category j of X_2 and category k of X_3. There are $J \cdot K$ such parameters. We impose the following $J + K$ sum-to-zero constraints:

$$\begin{aligned}\sum_{k=1}^{K} u_{23(jk)} &= 0, \text{ for } 1 \le j \le J,\\ \sum_{j=1}^{J} u_{23(jk)} &= 0, \text{ for } 1 \le k \le K\end{aligned}$$

Since both sets of constraints imply that the sum of all $u_{23(jk)}$ is zero, there are only $J + K - 1$ free constraints. Thus we have $JK - (J + K - 1) = (J - 1) \cdot (K - 1)$ *free* u-terms in this set.

$\boxed{\left\{u_{123(ijk)} : 1 \le i \le I, 1 \le j \le J, 1 \le k \le K\right\}}$ The u-term $u_{123(ijk)}$ represents the deviation from $u + u_{1(i)} + u_{2(j)} + u_{3(k)}$ associated with the interaction between category i of X_1, category j of X_2

and category k of X_3. There are $I \cdot J \cdot K$ such parameters. We impose the following $IJ + JK + IK$ sum-to-zero constraints:

$$\begin{aligned}
\sum_{k=1}^{K} u_{123(ijk)} &= 0, \text{ for } 1 \leq i \leq I \text{ and } 1 \leq j \leq J \\
\sum_{j=1}^{J} u_{123(ijk)} &= 0, \text{ for } 1 \leq i \leq I \text{ and } 1 \leq k \leq K \\
\sum_{i=1}^{I} u_{123(ijk)} &= 0, \text{ for } 1 \leq j \leq J \text{ and } 1 \leq k \leq K
\end{aligned}$$

These three sets of constraints imply that the sum of all $\{u_{123(ijk)} : 1 \leq k \leq K\}$ is zero, the sum of all $\{u_{123(ijk)} : 1 \leq j \leq J\}$, the sum of all $\{u_{123(ijk)} : 1 \leq i \leq I\}$ is zero and the sum of all $u_{123(ijk)}$ is zero. After applying the inclusion and exclusion theorem, it follows that there are only $IJ + JK + IK - I - J - K + 1$ free constraints. Therefore this set of u-terms contains

$$IJK - IJ - IK - JK + I + J + K - 1 = IJK - (I-1)(J-1)(K-1)$$

free terms.

The number of free u-terms in the saturated log-linear model from Equation (1) is obtained by adding the number of free u-terms associated with each of the sets we discussed above. This number of free u-terms turns out to be precisely IJK – the number of cells in the table. Therefore the saturated log-linear model has zero degrees of freedom

$$\#df([123]) = 0$$

which implies that the estimated expected cell values under this model are precisely the observed counts:

$$\widehat{m}_{ijk} = n_{ijk}$$

The estimated cell probabilities under the saturated log-linear model are obtained by dividing the observed counts by the grand total

$$\widehat{p}_{ijk} = \frac{n_{ijk}}{n_{+++}}$$

It follows that the values of G^2 and X^2 associated with any log-linear model are equal to zero. Since a Chi-squared random variable is always positive, the p-values based on G^2 and X^2 will be equal to 1. That is, the saturated log-linear model fits the data. In fact, it fits the data perfectly since the expected cell values are equal to the observed counts.

The minimal sufficient statistics (MSS) of the saturated log-linear model are the counts in the entire table

$$\{n_{ijk} : 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K\}$$

which can be viewed as the marginal associated with X_1, X_2 and X_3. For this reason the saturated log-linear model is denoted by

$$[123]$$

For the Color of Hair and Eyes data, the saturated log-linear model says that Gender, Hair color and Eye color are related in any possible way. That is: (i) knowing somebody's color of hair and eyes provides information about their gender; (ii) knowing somebody's color of hair and gender provides information about their color of eyes; and (iii) knowing somebody's color of eyes and gender provides information about their color of hair. This model does not have any interpretation in terms of independence or conditional independence relationships among the three variables. As such, it gives very little insight about what is going on in this dataset.

3.1 R Commands

In R the saturated log-linear model for Color of Hair and Eyes data in Table 3 is fitted with the following command:

```
saturated.loglin = loglin(haireyecolor.array,margin=list(c(1,2,3)),
                                          fit=TRUE,param=TRUE)
```

The call

```
saturated.loglin$param
```

gives the estimates of all the u-terms in Equation (1). More specifically, the estimate of u is

```
$`(Intercept)`
[1] 2.468886
```

The estimates of $\{u_{1(i)} : 1 \leq i \leq I\}$ are

```
$`1`
[1] -0.00924341  0.00924341
```

The estimates of $\{u_{2(j)} : 1 \leq j \leq J\}$ are

```
$`2`
[1] -0.3003665  0.9218819 -0.3304725 -0.2910429
```

The estimates of $\{u_{3(k)} : 1 \leq k \leq K\}$ are

```
$`3`
[1]  0.3772741  0.5113737 -0.2677748 -0.6208730
```

The estimates of $\{u_{12(ij)} : 1 \leq i \leq I, 1 \leq j \leq J\}$ are

```
$`1.2`
           [,1]        [,2]         [,3]       [,4]
[1,]  0.1569309 -0.03058066 -0.004922676 -0.1214276
[2,] -0.1569309  0.03058066  0.004922676  0.1214276
```

The estimates of $\{u_{13(ik)} : 1 \leq i \leq I, 1 \leq k \leq K\}$ are

```
 $`1.3`
           [,1]        [,2]       [,3]        [,4]
[1,] -0.1276105  0.03240871  0.0773343  0.01786752
[2,]  0.1276105 -0.03240871 -0.0773343 -0.01786752
```

The estimates of $\{u_{23(jk)} : 1 \leq j \leq J, 1 \leq k \leq K\}$ are

```
 $`2.3`
            [,1]       [,2]        [,3]       [,4]
[1,]  0.97883369 -0.3823334  0.05526667 -0.6517669
[2,]  0.31193121 -0.1829500  0.17009261 -0.2990738
[3,]  0.02189912 -0.5255398  0.07527125  0.4283694
[4,] -1.31266401  1.0908233 -0.30063053  0.5224713
```

The estimates of $\{u_{123(ijk)} : 1 \leq i \leq I, 1 \leq j \leq J, 1 \leq k \leq K\}$ are

```
$`1.2.3`
, , 1

            [,1]       [,2]       [,3]       [,4]
[1,] -0.07896848  0.0577532 -0.0932252  0.1144405
[2,]  0.07896848 -0.0577532  0.0932252 -0.1144405

, , 2

            [,1]       [,2]       [,3]       [,4]
[1,] -0.07976086  0.2002466  0.1600948 -0.2805806
[2,]  0.07976086 -0.2002466 -0.1600948  0.2805806

, , 3

           [,1]       [,2]        [,3]        [,4]
[1,]  0.1215518 -0.1117202 -0.06316822  0.05333667
[2,] -0.1215518  0.1117202  0.06316822 -0.05333667

, , 4

            [,1]       [,2]         [,3]       [,4]
[1,]  0.03717754 -0.1462796 -0.003701433  0.1128035
[2,] -0.03717754  0.1462796  0.003701433 -0.1128035
```

3.2 The Induced Regressions

The saturated log-linear model implies: (i) X_2 and X_3 are present in the regression associated with X_1; (ii) X_1 and X_3 are present in the regression associated with X_2 and (iii) X_1 and X_2 are

present in the regression associated with X_3. For example, the regression associated with X_2 is:

$$\begin{aligned}
\log \frac{P(X_2 = j_1 | X_1 = i, X_3 = k)}{P(X_2 = j_2 | X_1 = i, X_3 = k)} &= \log \frac{P(X_1 = i, X_2 = j_1, X_3 = k)}{P(X_1 = i, X_2 = j_2, X_3 = k)}, \\
&= \log \frac{p_{ij_1k}}{p_{ij_2k}}, \\
&= \log \frac{m_{ij_1k}}{m_{ij_2k}}, \\
&= \log m_{ij_1k} - \log m_{ij_2k}, \\
&= (u + u_{1(i)} + u_{2(j_1)} + u_{3(k)} + u_{12(ij_1)} + u_{13(ik)} + u_{23(j_1k)} + u_{123(ij_1k)}) - \\
&\quad (u + u_{1(i)} + u_{2(j_2)} + u_{3(k)} + u_{12(ij_2)} + u_{13(ik)} + u_{23(j_2k)} + u_{123(ij_2k)}), \\
&= (u_{2(j_1)} - u_{2(j_2)}) + (u_{12(ij_1)} - u_{12(ij_2)}) + (u_{23(j_1k)} - u_{23(j_2k)}) + (u_{123(ij_1k)} - u_{123(ij_2k)})
\end{aligned}$$

Let's determine the odds of having black hair ($j_1 = 1$) vs. red hair ($j_2 = 3$) for a woman ($i = 2$) who has green eyes ($k = 4$). We write

$$\begin{aligned}
\log \frac{P(X_2 = 1 | X_1 = 2, X_3 = 4)}{P(X_2 = 3 | X_1 = 2, X_3 = 4)} &= (\widehat{u}_{2(1)} - \widehat{u}_{2(3)}) + (\widehat{u}_{12(21)} - \widehat{u}_{12(23)}) + (\widehat{u}_{23(14)} - \widehat{u}_{23(34)}) + (\widehat{u}_{123(214)} - \widehat{u}_{123(234)}), \\
&= (-0.30 - (-0.33)) + (-0.157 - 0.005) + (-0.037 - 0.004) \\
&= -0.173
\end{aligned}$$

That is

$$\frac{P(X_2 = 1 | X_1 = 2, X_3 = 4)}{P(X_2 = 3 | X_1 = 2, X_3 = 4)} = \exp(-0.173) = 0.84 < 1$$

In other words, a woman with green eyes is less likely to have black hair than to have red hair.

3.3 The Iterative Proportional Fitting Algorithm

A statistical package does not know up-front that the estimated expected cell values under the saturated log-linear model are precisely the observed counts. Instead, a statistical package makes use of the Iterative Proportional Fitting (IPF) algorithm to find the estimates of m_{ijk}. IPF starts by setting some initial values for the m_{ijk}. That is, at iteration 0 all the estimated expected cell values are equal to one:

$$m_{ijk}^{[0]} = 1$$

At each iteration, IPF adjusts the estimated expected cell values with respect to each minimal sufficient statistic (MSS) of the saturated log-linear model. We have already explained that the saturated model has only one MSS, namely the actual observed table. As such, at iteration 1, IPF performs the following update:

$$\begin{aligned}
m_{ijk}^{[1]} &= \frac{n_{ijk}}{m_{ijk}^{[0]}} m_{ijk}^{[0]}, \\
&= n_{ijk}
\end{aligned}$$

At iteration 2, IPF performs a similar update:

$$\begin{aligned} m_{ijk}^{[2]} &= \frac{n_{ijk}}{m_{ijk}^{[1]}} m_{ijk}^{[1]}, \\ &= \frac{n_{ijk}}{n_{ijk}} n_{ijk}, \\ &= n_{ijk} \end{aligned}$$

IPF finds the estimates of the expected cell values after the first iteration. After the second iteration these estimates remained the same, hence IPF decides it has converged, stops and outputs these values. *Note: IPF always converges after two iterations for the saturated log-linear model associated with any contingency table.*

4 The Complete Independence Model [1][2][3]

We set to zero the first and second order interaction terms

$$u_{12(ij)} = u_{23(jk)} = u_{13(ik)} = u_{123(ijk)} = 0$$

in the saturated log-linear model in Equation (1) to obtain the model

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} \tag{2}$$

Since the u-terms of this model are associated with at most one variable, the log-linear model in Equation (2) is called the complete independence model. That is, this model says that each variable is independent of the other two variables: $X_1 \perp\!\!\!\perp \{X_2, X_3\}$, $X_2 \perp\!\!\!\perp \{X_1, X_3\}$ and $X_3 \perp\!\!\!\perp \{X_1, X_2\}$. *For the Color of Hair and Eyes data, the model* [1][2][3] *says that Gender, Hair color and Eye color are independent of each other. That is: (i) knowing somebody's gender or color of eyes provides no information about their color of hair; (ii) knowing somebody's gender or color of hair provides no information about their color of eyes; and (iii) knowing somebody's color of hair and color of eyes provides no information about their gender.*

Under complete independence, the joint distribution of X_1 and X_2 and X_3 factorizes as the product of the marginal distributions of X_1 and X_2 and X_3:

$$P(X_1 = i, X_2 = j, X_3 = k) = P(X_1 = i) \cdot P(X_2 = j) \cdot P(X_3 = k)$$

In our notation, we write

$$p_{ijk} = p_{i++} \cdot p_{+j+} \cdot p_{++k}$$

The complete independence model has three MSS, the one-way marginals of the $I \times J \times K$ table:

$$\{n_{i++} : 1 \le i \le I\}, \quad \{n_{+j+} : 1 \le j \le J\}, \quad \{n_{++k} : 1 \le k \le K\}$$

Thus the model is denoted by

[1][2][3]

The number of free parameters in Equation (2) is

$$1 + (I - 1) + (J - 1) + (K - 1)$$

hence the number of degrees of freedom of the complete independence model is

$$\#df([1][2][3]) = IJK - I - J - K + 2$$

For example, for the Hair and Color data, the model of complete independence has

$$2 \cdot 4 \cdot 4 - 2 - 4 - 4 + 2 = 24$$

degrees of freedom.

Since the one-way marginals are the MSS of [1][2][3], it follows that the estimates of marginal cell probabilities are

$$\widehat{p}_{i++} = \frac{n_{i++}}{n_{+++}}, \quad \widehat{p}_{+j+} = \frac{n_{+j+}}{n_{+++}}, \quad \widehat{p}_{++k} = \frac{n_{++k}}{n_{+++}}$$

The expected cell values are therefore estimated as

$$\begin{aligned} \widehat{m}_{ijk} &= n_{+++} \cdot \widehat{p}_{ijk}, \\ &= n_{+++} \cdot \widehat{p}_{i++} \cdot \widehat{p}_{+j+} \cdot \widehat{p}_{++k}, \\ &= n_{+++} \cdot \frac{n_{i++}}{n_{+++}} \cdot \frac{n_{+j+}}{n_{+++}} \cdot \frac{n_{++k}}{n_{+++}}, \\ &= \frac{n_{i++} \cdot n_{+j+} \cdot n_{++k}}{n_{+++}^2} \end{aligned}$$

Note: notice the similarity with the formula for the estimates of the expected cell values under the independence model for two-way tables

$$\widehat{m}_{ij} = \frac{n_{i+} \cdot n_{+j}}{n_{++}}$$

4.1 R Commands

In R the complete independence model for Color of Hair and Eyes Data is fitted with the following command:

```
indep.loglin = loglin(haireyecolor.array,margin=list(1,2,3),
                                  fit=TRUE,param=TRUE)
2 iterations: deviation 5.684342e-14
```

The number of degrees of freedom is obtained with the call

```
> indep.loglin$df
[1] 24
```

The value of the likelihood ratio statistic is

```
 > indep.loglin$lrt
[1] 167.4065
```

while the value of the X^2 statistic is

```
 > indep.loglin$pearson
[1] 164.9951
```

The p-value for testing the null hypothesis

$$H_0 : \quad u_{12(ij)} = u_{23(jk)} = u_{13(ik)} = u_{123(ijk)} = 0$$

based on G^2 is given by

$$P(\chi^2_{24} \geq 167.4065)$$

and is obtained with the R call

```
> 1-pchisq(indep.loglin$lrt,indep.loglin$df)
[1] 0
```

The p-value for testing H_0 based on X^2 is obtained with the call

```
 > 1-pchisq(indep.loglin$pearson,indep.loglin$df)
[1] 0
```

We reject H_0 and conclude that the complete independence model [1][2][3] *does not fit well* the Hair and Color data.

4.2 The Induced Regressions

The complete independence model implies:

- X_2 and X_3 are not present in the regression associated with X_1, i.e.

$$\log \frac{P(X_1 = i_1|X_2 = j, X_3 = k)}{P(X_1 = i_2|X_2 = j, X_3 = k)} = \log \frac{P(X_1 = i_1)}{P(X_1 = i_2)}$$

We also have

$$\begin{aligned} \log \frac{P(X_1 = i_1|X_2 = j, X_3 = k)}{P(X_1 = i_2|X_2 = j, X_3 = k)} &= \log m_{i_1 jk} - \log m_{i_2 jk}, \\ &= (u + u_{1(i_1)} + u_{2(j)} + u_{3(k)}) - (u + u_{1(i_2)} + u_{2(j)} + u_{3(k)}), \\ &= u_{1(i_1)} - u_{1(i_2)} \end{aligned}$$

Remark that the above expression remains the same no matter which values j and k take (i.e., irrespective of the categories of X_2 and X_3), which means that the regression for X_1 induced by the log-linear model [1][2][3] is

$$\log \frac{P(X_1 = i_1)}{P(X_1 = i_2)} = u_{1(i_1)} - u_{1(i_2)}$$

- X_1 and X_3 are not present in the regression associated with X_2, which means that the regression for X_2 induced by the log-linear model [1][2][3] is

$$\log \frac{P(X_2 = j_1)}{P(X_2 = j_2)} = u_{2(j_1)} - u_{2(j_2)}$$

- X_1 and X_2 are not present in the regression associated with X_3, which means that the regression for X_3 induced by the log-linear model [1][2][3] is

$$\log \frac{P(X_3 = k_1)}{P(X_3 = k_2)} = u_{3(k_1)} - u_{3(k_2)}$$

Let's determine the odds of having black hair ($j_1 = 1$) vs. red hair ($j_2 = 3$) for a woman ($i = 2$) who has green eyes ($k = 4$). The complete independence model says that these odds will be the same no matter what hair color that person has and no matter whether the person is male or female. *Note: we already know that the data does not support model* [1][2][3]*, but we will calculate these odds based on complete independence anyways*. We write

$$\log \frac{P(X_2 = 1|X_1 = 2, X_3 = 4)}{P(X_2 = 3|X_1 = 2, X_3 = 4)} \quad = \quad \widehat{u}_{2(1)} - \widehat{u}_{2(3)}$$

We obtain the estimates of $u_{2(1)}$ and $u_{2(3)}$ in R using the call

```
 > indep.loglin$param$'2'
[1] -0.17470699  0.78151645 -0.59415834 -0.01265113
```

It follows that $\widehat{u}_{2(1)} = -0.175$ and $\widehat{u}_{2(3)} = -0.594$. We obtain

$$\frac{P(X_2 = 1|X_1 = 2, X_3 = 4)}{P(X_2 = 3|X_1 = 2, X_3 = 4)} = \exp(-0.175 - (-0.594)) = 1.52 > 1$$

In other words, a woman with green eyes is more likely to have black hair than to have red hair! Remember that, for the saturated log-linear model which fits the data perfectly, we obtained exactly the opposite. In fact, the log-linear model [1][2][3] says that *any* person is more likely to have black hair that to have red hair, i.e.

$$\frac{P(X_2 = 1)}{P(X_2 = 3)} = 1.52$$

Clearly this makes little sense, hence there is no surprise that the complete independence model does not fit the Color of Hair and Eyes Data.

4.3 The Iterative Proportional Fitting Algorithm

The IPF algorithm starts by setting the estimated expected cell values to one:

$$m_{ijk}^{[0]} = 1$$

At each iteration, IPF adjusts the estimated expected cell values with respect to the three minimal sufficient statistic of the complete independence model. The adjustment w.r.t. the one-way marginal associated with X_1 is:

$$m_{ijk}^{[1+\frac{1}{3}]} = \frac{n_{i++}}{m_{i++}^{[0]}} m_{ijk}^{[0]} = \frac{n_{i++}}{JK}$$

The adjustment w.r.t. the one-way marginal associated with X_2 is

$$m_{ijk}^{[1+\frac{2}{3}]} = \frac{n_{+j+}}{m_{+j+}^{[1+\frac{1}{3}]}} m_{ijk}^{[1+\frac{1}{3}]} = \frac{n_{+j+}}{\frac{n_{+++}}{J}} \cdot \frac{n_{i++}}{JK} = \frac{n_{i++} \cdot n_{+j+}}{n_{+++} \cdot K}$$

The adjustment w.r.t. the one-way marginal associated with X_3 is

$$m_{ijk}^{[2]} = \frac{n_{++k}}{m_{++k}^{[1+\frac{2}{3}]}} m_{ijk}^{[1+\frac{2}{3}]} = \frac{n_{++k}}{\frac{n_{+++}}{K}} \cdot \frac{n_{i++} \cdot n_{+j+}}{n_{+++} \cdot K} = \frac{n_{i++} \cdot n_{+j+} \cdot n_{++k}}{n_{+++}^2}$$

Therefore, after one iteration, IPF reaches the correct estimates of the expected cell values (this is precisely the formula we discovered earlier). IPF goes through the same updates one more time to find out the estimates of the expected cell values do not change. Thus IPF decides it has converged after two iterations, stops and outputs the estimates $\widehat{m}_{ijk}$.

In R these estimates are obtained using the call

```
 > indep.loglin$fit
, , 1

         [,1]     [,2]     [,3]     [,4]
[1,] 18.89386 49.15904 12.42097 22.21779
[2,] 21.58314 56.15613 14.18891 25.38017

, , 2

         [,1]     [,2]     [,3]     [,4]
[1,] 18.46446 48.04179 12.13867 21.71284
[2,] 21.09261 54.87985 13.86644 24.80335

, , 3

         [,1]     [,2]     [,3]      [,4]
[1,] 7.986952 20.78087 5.250681  9.392064
[2,] 9.123781 23.73873 5.998041 10.728890

, , 4

         [,1]     [,2]     [,3]     [,4]
[1,] 5.066991 13.18356 3.331077 5.958406
[2,] 5.788205 15.06005 3.805209 6.806500
```

This output says that $\widehat{m}_{111} = 18.89$ or $\widehat{m}_{233} = 5.998$.

5 Models with One Variable Independent of the Other Two

There are three such models:

• [1][23] This is the model of independence of X_1 and $\{X_2, X_3\}$. *For the Color of Hair and Eyes data, this is the model of independence of Gender from Hair color and Eye color. That is, if you know somebody's gender, you do not have any information about their color of hair or color of eyes. The model also says that: (i) if you know somebody's color of hair, you have information about their color of eyes; and (ii) if you know somebody's color of eyes, you have information about their color of hair.*

This model contains an interaction term between X_2 and X_3, but no interaction terms between X_1 and X_2 or between X_1 and X_3:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}$$

The number of free u-terms is

$$1 + (I-1) + (J-1) + (K-1) + (J-1)(K-1)$$

The number of degrees of freedom is obtained by subtracting the number of free u-terms from the number of cells in the table (IJK):

$$\#df([1][23]) = (I-1)(JK-1)$$

If X_1 is independent of X_2 and X_3, the joint distribution of X_1, X_2 and X_3 factorizes as:

$$P(X_1 = i, X_2 = j, X_3 = k) = P(X_1 = i) \cdot P(X_2 = j, X_3 = k)$$

In our notation, we write

$$p_{ijk} = p_{i++} \cdot p_{+jk}$$

The minimal sufficient statistics of the log-linear model [1][23] are the one-way marginal associated with X_1,

$$\{n_{i++} : 1 \le i \le I\}$$

and the $(2,3)$-marginal

$$\{n_{+jk} : 1 \le j \le J, 1 \le k \le K\}$$

The estimates of the expected cell values are given by the formula:

$$\begin{aligned} \widehat{m}_{ijk} &= n_{+++} \cdot \widehat{p}_{i++} \cdot \widehat{p}_{+jk}, \\ &= n_{+++} \cdot \frac{n_{i++}}{n_{+++}} \cdot \frac{n_{+jk}}{n_{+++}}, \\ &= \frac{n_{i++} \cdot n_{+jk}}{n_{+++}} \end{aligned}$$

• [2][13] This is the model of independence of X_2 and $\{X_1, X_3\}$. *For the Color of Hair and Eyes data, this is the model of independence of Hair color from Gender and Eye color. That is, if you know somebody's color of hair, you do not have any information about their gender or their color of eyes. The model also says that: (i) if you know somebody's color of eyes, you have information about their gender; and (ii) if you know somebody's gender, you have information about their color of eyes*.

This model contains an interaction term between X_1 and X_3, but no interaction terms between X_2 and X_1 or between X_2 and X_3:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)}$$

The number of free u-terms is

$$1 + (I-1) + (J-1) + (K-1) + (I-1)(K-1)$$

The number of degrees of freedom is

$$\#df([2][13]) = (J-1)(IK-1)$$

The minimal sufficient statistics of the log-linear model [2][13] are the one-way marginal associated with X_2,

$$\{n_{+j+} : 1 \le i \le J\}$$

and the $(1,3)$-marginal

$$\{n_{i+k} : 1 \le i \le I, 1 \le k \le K\}$$

The estimates of the expected cell values are given by the formula:

$$\begin{aligned} \widehat{m}_{ijk} &= n_{+++} \cdot \widehat{p}_{+j+} \cdot \widehat{p}_{i+k}, \\ &= n_{+++} \cdot \frac{n_{+j+}}{n_{+++}} \cdot \frac{n_{i+k}}{n_{+++}}, \\ &= \frac{n_{+j+} \cdot n_{i+k}}{n_{+++}} \end{aligned}$$

• [3][12] This is the model of independence of X_3 and $\{X_1, X_2\}$. *For the Color of Hair and Eyes data, this is the model of independence of Eye color from Gender and Hair color. That is, if you know somebody's color of eyes, you do not have any information about their color of hair or their gender. The model also says that: (i) if you know somebody's color of hair, you have information about their gender; and (ii) if you know somebody's gender, you have information about their color of hair*. It contains an interaction term between X_1 and X_2, but no interaction terms between X_3 and X_1 or between X_3 and X_2:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)}$$

The number of free u-terms is

$$1 + (I-1) + (J-1) + (K-1) + (I-1)(J-1)$$

The number of degrees of freedom is

$$\#df([2][13]) = (K-1)(IJ-1)$$

The minimal sufficient statistics of the log-linear model [3][12] are the one-way marginal associated with X_3,

$$\{n_{++k} : 1 \leq i \leq K\}$$

and the $(1,2)$-marginal

$$\{n_{ij+} : 1 \leq i \leq I, 1 \leq j \leq J\}$$

The estimates of the expected cell values are given by the formula:

$$\begin{aligned} \widehat{m}_{ijk} &= n_{+++} \cdot \widehat{p}_{++k} \cdot \widehat{p}_{ij+}, \\ &= n_{+++} \cdot \frac{n_{++k}}{n_{+++}} \cdot \frac{n_{ij+}}{n_{+++}}, \\ &= \frac{n_{++k} \cdot n_{ij+}}{n_{+++}} \end{aligned}$$

The Iterative Proportional Fitting (IPF) algorithm finds the estimates $\widehat{m}_{ijk}$ after two iterations. We will not discuss IPF in detail. You should remember that IPF will *always* converge to the estimates $\widehat{m}_{ijk}$ after two iterations if there is a formula that defines these estimates. For three-way tables, only the log-linear model of no second order interaction does not have a formula for its expected cell probabilities.

5.1 R Commands

We fit the log-linear model [1][23] using the command:

```
X1indepX2X3.loglin = loglin(haireyecolor.array,margin=list(1,c(2,3)),
                                                fit=TRUE,param=TRUE)
2 iterations: deviation 5.684342e-14
```

Remark that IPF found the estimates of the expected cell values $\widehat{m}_{ijk}$ after two iterations. The number of degrees of freedom of this model is

```
  X1indepX2X3.loglin$df
[1] 15
```

We see that the formula for calculating the number of degrees of freedom gives the same result:

$$\#df([1][23]) = (I-1)(JK-1) = (2-1)(4 \cdot 4 - 1) = 15$$

The p-value based on G^2 associated with [1][23] is calculated in R like this:

```
1-pchisq(X1indepX2X3.loglin$lrt,X1indepX2X3.loglin$df)
[1] 0.1776609
```

This p-value indicates that [1][23] *fits Color of Hair and Eyes Data well*. We obtain a similar p-value based on the X^2 test statistic:

```
1-pchisq(X1indepX2X3.loglin$pearson,X1indepX2X3.loglin$df)
[1] 0.1885290
```

Next we fit the log-linear model [2][13]:

```
X2indepX1X3.loglin = loglin(haireyecolor.array,margin=list(2,c(1,3)),
                                                fit=TRUE,param=TRUE)
```

The p-value based on G^2 associated with [2][13] is equal to zero,

```
1-pchisq(X2indepX1X3.loglin$lrt,X2indepX1X3.loglin$df)
[1] 0
```

which indicates that [2][13] *does not fit the Color of Hair and Eyes data*. Similarly, we fit the log-linear model [3][12] then calculate the corresponding p-value based on the G^2 test statistic:

```
X3indepX1X2.loglin = loglin(haireyecolor.array,margin=list(3,c(1,2)),
                                           fit=TRUE,param=TRUE)
1-pchisq(X3indepX1X2.loglin$lrt,X3indepX1X2.loglin$df)
[1] 0
```

A p-value equal to zero indicates that [3][12] *does not fit the Color of Hair and Eyes data. It follows that, out of the model of complete independence and the models of independence of one variable of the other two, only the model* [1][23] *fits the Hair and Color data. We learned that Gender is independent of the colors of hair and eyes, and that the colors of hair and eyes are not independent of each other.*

5.2 The Induced Regressions

We will discuss only the regressions induced by the model [1][23] on each of the variables X_1, X_2 and X_3. The regressions induced by the other two models are obtained by permuting the indices of the three variables in the text that follows.

• Since X_1 is independent of X_2 and X_3, the regression of X_1 given X_2 and X_3 does not depend on X_2 and X_3 (i.e., this is the null regression with no predictors):

$$\log \frac{P(X_1 = i_1|X_2 = j, X_3 = k)}{P(X_1 = i_2|X_2 = j, X_3 = k)} = \log \frac{P(X_1 = i_1)}{P(X_1 = i_2)}$$

This regression is expressed using the u-terms of the log-linear model [1][23] as follows:

$$\log \frac{P(X_1 = i_1)}{P(X_1 = i_2)} = u_{1(i_1)} - u_{1(i_2)}$$

For example, for the Color of Hair and Eyes data, this implies that the odds of being male vs. being female does not depend on the color of eyes and hair:

$$\log \frac{P(\text{Gender} = \text{Male})}{P(\text{Gender} = \text{Female})} = \log \frac{P(X_1 = 1)}{P(X_1 = 2)} = \widehat{u}_{1(i_1)} - \widehat{u}_{1(i_2)} = -0.0665 - 0.0665 = -0.133$$

which implies

$$\frac{P(\text{Gender} = \text{Male})}{P(\text{Gender} = \text{Female})} = \exp(-0.133) = 0.875 < 1$$

• Since X_2 interacts with X_3 and it is independent of X_1, the regression of X_2 given X_1 and X_3 depends only on X_3 and does not depend on X_1:

$$\log \frac{P(X_2 = j_1|X_1 = i, X_3 = k)}{P(X_2 = j_2|X_1 = i, X_3 = k)} = \log \frac{P(X_2 = j_1|X_3 = k)}{P(X_2 = j_2|X_3 = k)} = (u_{2(j_1)} - u_{2(j_2)}) + (u_{23(j_1k)} - u_{23(j_2k)})$$

For example, the log-linear model [1][23] says that odds of a woman with green eyes having black hair vs. red hair is equal to the odds of a man with green eyes having black hair vs. red hair, i.e.

$$\log \frac{P(X_2 = 1|X_1 = 1, X_3 = 4)}{P(X_2 = 3|X_1 = 1, X_3 = 4)} = \log \frac{P(X_2 = 1|X_1 = 2, X_3 = 4)}{P(X_2 = 3|X_1 = 2, X_3 = 4)} = \log \frac{P(X_2 = 1|X_3 = 4)}{P(X_2 = 3|X_3 = 4)}$$

These odds are obtained from the estimated u-terms of model [1][23] as follows:

$$\begin{aligned} \frac{P(X_2 = 1|X_3 = 4)}{P(X_2 = 3|X_3 = 4)} &= \exp((\widehat{u}_{2(1)} - \widehat{u}_{2(3)}) + (\widehat{u}_{23(14)} - \widehat{u}_{23(34)})) \\ &= \exp(-0.295 - (-0.335) + (-0.646 - 0.425)) \\ &= 0.357 < 1 \end{aligned}$$

Therefore, according to the log-linear model [1][23], any person with green eyes is about three times more likely to have red hair than to have black hair.

• Since X_3 interacts with X_2 and it is independent of X_1, the regression of X_3 given X_1 and X_2 depends only on X_3 and does not depend on X_2:

$$\log \frac{P(X_3 = k_1|X_1 = i, X_2 = j)}{P(X_3 = k_2|X_1 = i, X_2 = j)} = \log \frac{P(X_3 = k_1|X_2 = j)}{P(X_3 = k_2|X_2 = j)} = (u_{3(k_1)} - u_{3(k_2)}) + (u_{23(jk_1)} - u_{23(jk_2)})$$

For example, the log-linear model [1][23] says that odds of a woman with black hair to have blue eyes vs. green eyes is equal to the odds of a man with black hair to have blue eyes vs. green eyes:

$$\log \frac{P(X_3 = 2|X_1 = 1, X_2 = 1)}{P(X_3 = 4|X_1 = 1, X_2 = 1)} = \log \frac{P(X_3 = 2|X_1 = 2, X_2 = 1)}{P(X_3 = 4|X_1 = 2, X_2 = 1)} = \log \frac{P(X_3 = 2|X_2 = 1)}{P(X_3 = 4|X_2 = 1)}$$

These odds are obtained from the estimated u-terms of model [1][23] as follows:

$$\begin{aligned}\frac{P(X_3 = 2|X_2 = 1)}{P(X_3 = 4|X_2 = 1)} &= \exp((\widehat{u}_{3(2)} - \widehat{u}_{3(4)}) + (\widehat{u}_{23(12)} - \widehat{u}_{23(14)})) \\ &= \exp(0.523 - (-0.628) - 0.41 - (-0.646)) \\ &= 4\end{aligned}$$

Therefore, according to the log-linear model [1][23], any person with black hair is four times more likely to have blue eyes than to have green eyes.

6 Models of Conditional Independence

There are three such models:

• [12][13] This is the model of conditional independence of X_2 and X_3 given X_1. *For the Color of Hair and Eyes data, this model says that somebody's color of hair and color of eyes are independent given their gender. That is, if you already know somebody's gender, knowing their color of eyes does not provide you any additional information about their color of hair. Moreover, if you already know somebody's gender, knowing their color of hair does not provide you any additional information about their color of eyes.*

It contains an interaction term between X_1 and X_2, an interaction term between X_1 and X_3, but no interaction term between X_2 and X_3:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)}$$

The number of free u-terms is

$$1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1) + (I - 1)(K - 1)$$

The number of degrees of freedom is

$$\#df([12][13]) = I(K - 1)(J - 1)$$

Your intuition about this model should be as follows: for each category i_0 of X_1, you have a $J \times K$ contingency table

$$\{n_{i_0jk} : 1 \leq j \leq J, 1 \leq k \leq K\}$$

and the two variable cross-classified in this table are independent. If the log-linear model of independence holds for all such "slices" of the $I \times J \times K$ table, you obtain the conditional independence model [12][13]. The model of independence for each $J \times K$ slice has $(J - 1)(K - 1)$ degrees of freedom. Since you have I such slices, you obtain another justification of the number of degrees of freedom $\#df([12][13])$.

Since X_2 is independent of X_3 given X_1, the joint distribution of X_1, X_2 and X_3 factorizes as:

$$\begin{aligned} P(X_1 = i, X_2 = j, X_3 = k) &= P(X_2 = j, X_3 = k | X_1 = i) \cdot P(X_1 = i), \\ &= P(X_2 = j | X_1 = i) \cdot P(X_3 = k | X_1 = i) \cdot P(X_1 = i), \\ &= \frac{P(X_1 = i, X_2 = j)}{P(X_1 = i)} \cdot \frac{P(X_1 = i, X_3 = k)}{P(X_1 = i)} \cdot P(X_1 = i), \\ &= \frac{P(X_1 = i, X_2 = j) \cdot P(X_1 = i, X_3 = k)}{P(X_1 = i)} \end{aligned}$$

In our usual notation, we write

$$p_{ijk} = \frac{p_{ij+} \cdot p_{i+k}}{p_{i++}}$$

Therefore the expected cell values are estimated as

$$\widehat{m}_{ijk} = \frac{n_{ij+} \cdot n_{i+k}}{n_{i++}}$$

Since there is a formula for the estimates of the expected cell values, the IPF algorithm will find these estimates after two iterations.

• [12][23] This is the model of conditional independence of X_1 and X_3 given X_2. *For the Color of Hair and Eyes data, this model says that somebody's gender and color of eyes are independent given their color of hair. That is, if you already know somebody's color of hair, knowing their color of eyes does not provide you any additional information about their gender. Moreover, if you already know somebody's color of hair, knowing their gender does not provide you any additional information about their color of eyes.*

This model contains an interaction term between X_1 and X_2, an interaction term between X_2 and X_3, but no interaction term between X_1 and X_3:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{23(jk)}$$

The number of free u-terms is

$$1 + (I - 1) + (J - 1) + (K - 1) + (I - 1)(J - 1) + (J - 1)(K - 1)$$

The number of degrees of freedom is

$$\#df([12][23]) = J(I - 1)(K - 1)$$

The expected cell values are estimated using the formula:

$$\widehat{m}_{ijk} = \frac{n_{ij+} \cdot n_{+jk}}{n_{+j+}}$$

• [13][23] This is the model of conditional independence of X_1 and X_2 given X_3. *For the Color*

of Hair and Eyes data, this model says that somebody's gender and color of hair are independent given their color of eyes. That is, if you already know somebody's color of eyes, knowing their color of hair does not provide you any additional information about their gender. Moreover, if you already know somebody's color of eyes, knowing their gender does not provide you any additional information about their color of hair.

This model contains an interaction term between X_1 and X_3, an interaction term between X_2 and X_3, but no interaction term between X_1 and X_2:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}$$

The number of free u-terms is

$$1 + (I-1) + (J-1) + (K-1) + (I-1)(K-1) + (J-1)(K-1)$$

The number of degrees of freedom is

$$\#df([13][23]) = K(I-1)(J-1)$$

The expected cell values are estimated using the formula:

$$\widehat{m}_{ijk} = \frac{n_{i+k} \cdot n_{+jk}}{n_{++k}}$$

6.1 R Commands

We fit the log-linear model [12][13] using the command:

```
X2indepX3givenX1.loglin = loglin(haireyecolor.array,margin=list(c(1,2),c(1,3)),
                                  fit=TRUE,param=TRUE)
2 iterations: deviation 2.842171e-14
```

The IPF algorithm found the estimates of the expected cell values in two iterations since there is a formula for these estimates as we explained earlier. The number of degrees of freedom for this model is:

```
X2indepX3givenX1.loglin$df
[1] 18
```

We can calculate the number of degrees of freedom directly using the formula

$$I(J-1)(K-1) = 2 \cdot (4-1) \cdot (4-1) = 18$$

The p-value based on G^2 associated with [12][13] is calculated in R like this:

```
1-pchisq(X2indepX3givenX1.loglin$lrt,X2indepX3givenX1.loglin$df)
[1] 0
```

and says that this model does not fit the data well. Next we fit the model [12][23]:

```
X1indepX3givenX2.loglin = loglin(haireyecolor.array,margin=list(c(1,2),c(2,3)),
                                 fit=TRUE,param=TRUE)
2 iterations: deviation 2.842171e-14
```

The p-value based on G^2 associated with [12][13] is calculated in R like this:

```
1-pchisq(X1indepX3givenX2.loglin$lrt,X1indepX3givenX2.loglin$df)
[1] 0.4211115
```

and says that this model fits the data well. We fit the third conditional independence model [13][23]:

```
X1indepX2givenX3.loglin = loglin(haireyecolor.array,margin=list(c(1,3),c(2,3)),
                     fit=TRUE,param=TRUE)
2 iterations: deviation 1.421085e-14
```

Its p-value based on G^2 is

```
1-pchisq(X1indepX2givenX3.loglin$lrt,X1indepX2givenX3.loglin$df)
[1] 0.09165593
```

We see that this model also fits the data well, although the fit is not as good as the fit of the model [13][23]. *Note: previously we learned that the model* [1][23] *fits the Color of Hair and Eyes data well. Remark that* [1][23] *is nested in the two conditional independence model that also fit the data:* [12][23] *and* [13][23]. *More explicitly,* [1][23] *is obtained from* [12][23] *by setting* $u_{12(ij)} = 0$. *Similarly,* [1][23] *is obtained from* [13][23] *by setting* $u_{13(ik)} = 0$. *What you should remember: if a simpler/more parsimonious model fits a dataset, a more complex model that embeds it is also likely to fit the data.*

6.2 The Induced Regressions

We will discuss only the regressions induced by the log-linear model [13][23]. The regressions induced by the other two conditional independence models are obtained by permuting the indices of the three variables.

• Since X_1 is independent of X_2 given X_3, the regression of X_1 given X_2 and X_3 depends only on X_3:

$$\log \frac{P(X_1 = i_1|X_2 = j, X_3 = k)}{P(X_1 = i_2|X_2 = j, X_3 = k)} = \log \frac{P(X_1 = i_1|X_3 = k)}{P(X_1 = i_2|X_3 = k)}$$

This regression is expressed using the u-terms of the log-linear model [13][23] as follows:

$$\log \frac{P(X_1 = i_1|X_3 = k)}{P(X_1 = i_2|X_3 = k)} = (u_{1(i_1)} - u_{1(i_2)}) + (u_{13(i_1k)} - u_{13(i_2k)})$$

For example, for the Color of Hair and Eyes data, this implies that the odds of being male ($i_1 = 1$) vs. being female ($i_2 = 2$) depends only on the color of eyes and does not depend on the color of hair. Let's calculate these odds for a person with green eyes ($k = 4$):

$$\begin{aligned}\frac{P(\text{Gender = Male}|\text{Eye color = green})}{P(\text{Gender = Female}|\text{Eye color = green})} &= \exp((\widehat{u}_{1(1)} - \widehat{u}_{1(2)}) + (\widehat{u}_{13(14)} - \widehat{u}_{13(24)})),\\ &= \exp((-0.053 - 0.053) + (0.0017 + 0.0017)),\\ &= 0.902 < 1\end{aligned}$$

Therefore a person with green eyes is more likely to be a woman irrespective of their color of hair.

• Since X_1 is independent of X_2 given X_3, the regression of X_2 given X_1 and X_3 depends only on X_3:

$$\log \frac{P(X_2 = j_1|X_1 = i, X_3 = k)}{P(X_2 = j_2|X_1 = i, X_3 = k)} = \log \frac{P(X_2 = j_1|X_3 = k)}{P(X_2 = j_2|X_3 = k)}$$

This regression is expressed using the u-terms of the log-linear model [13][23] as follows:

$$\log \frac{P(X_2 = j_1|X_3 = k)}{P(X_2 = j_2|X_3 = k)} = (u_{2(j_1)} - u_{2(j_2)}) + (u_{23(j_1k)} - u_{23(j_2k)})$$

For example, for the Color of Hair and Eyes data, this implies that the odds of having black hair ($j_1 = 1$) vs. having red hair ($j_2 = 3$) depends only on the color of eyes and does not depend on gender. Let's calculate these odds for a person with green eyes ($k = 4$):

$$\begin{aligned}\frac{P(\text{Hair = Black}|\text{Eye color = green})}{P(\text{Hair = Red}|\text{Eye color = green})} &= \exp((\widehat{u}_{2(1)} - \widehat{u}_{2(3)}) + (\widehat{u}_{23(14)} - \widehat{u}_{23(34)})),\\ &= \exp(-0.295 + 0.335 - 0.646 - 0.425),\\ &= 0.357 < 1\end{aligned}$$

Therefore a man or a woman with green eyes is about three times more likely to have red hair vs. black hair.

• Since X_1 is independent of X_2 given X_3, the regression of X_3 given X_1 and X_2 depends on both X_1 and X_2 (X_3 has interaction terms with both X_1 and X_2). This regression is expressed using the u-terms of the log-linear model [13][23] as follows:

$$\log \frac{P(X_3 = k_1|X_1 = i, X_2 = j)}{P(X_3 = k_2|X_1 = i, X_2 = j)} = (u_{3(k_1)} - u_{3(k_2)}) + (u_{13(ik_1)} - u_{13(ik_2)}) + (u_{23(jk_1)} - u_{23(jk_2)})$$

For example, for the Color of Hair and Eyes data, this implies that the odds of having blue eyes ($k_1 = 2$) vs. having green eyes ($k_2 = 4$) depends on both gender and the color of hair. Let's calculate these odds for a woman ($i = 2$) with black hair ($j = 1$):

$$\begin{aligned}\frac{P(\text{Eyes = blue}|\text{Gender = Female, Hair = black})}{P(\text{Eyes = green}|\text{Gender = Female, Hair = black})} &= \exp((u_{3(2)} - u_{3(4)}) + (u_{13(22)} - u_{13(24)}) + (u_{23(12)} - u_{23(14)})),\\ &= \exp(0.524 + 0.627 + 0.008 + 0.0017 + 0.963 + 0.646),\\ &= 15.95\end{aligned}$$

The above odds imply that a woman with black hair is extremely likely to have blue eyes rather than having blue eyes.

7 The Model of No Second Order Interaction [12][13][23]

This model is obtained from the saturated log-linear model by setting the second order interaction terms to zero:

$$u_{123(ijk)} = 0$$

It contains an interaction term between X_1 and X_2, an interaction term between X_1 and X_3 and an interaction term between X_2 and X_3:

$$\log m_{ijk} = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{12(ij)} + u_{13(ik)} + u_{23(jk)}$$

The number of free u-terms is

$$1 + (I-1) + (J-1) + (K-1) + (I-1)(J-1) + (I-1)(K-1) + (J-1)(K-1)$$

Therefore the number of degrees of freedom is

$$\#df([12][13][23]) = (I-1)(J-1)(K-1)$$

The minimal sufficient statistics of the no second order interaction model are the two-way marginals of the $I \times J \times K$ table.

The no second order interaction model does not have any interpretation in terms of independence or conditional independence relationships among X_1, X_2 and X_3. In this regard, it is similar to the saturated model, although it has fewer parameters. For this reason this log-linear model is the first model we encountered that does not have an explicit formula for the estimated expected cell values $\widehat{m}_{ijk}$. The Iterative Proportional Fitting (IPF) algorithm is therefore needed to obtain these estimates. As opposed to the other log-linear model that admit formulas for calculating $\widehat{m}_{ijk}$, IPF will converge in more iterations than two iterations by adjusting the current estimates of m_{ijk} with respect to the $(1,2)$-marginal, then with respect to the $(1,3)$-marginal, then with respect to the $(2,3)$-marginal. *Remember: it might be the case that IPF has failed to converge even after a larger number of iterations. If your software gives any indication that this is the case, you must increase the number of iterations IPF is run. If IPF has not reached convergence, the numerical values of the G^2 or X^2 test statistics might be incorrect, therefore your entire analysis might be compromised.*

7.1 R Commands

We fit the log-linear model [12][13][23] using the command:

```
no2nd.loglin = loglin(haireyecolor.array,margin=list(c(1,2),c(1,3),c(2,3)),
                                          fit=TRUE,param=TRUE)
6 iterations: deviation 0.03237332
```

Remark that IPF has converged after six iterations (so far we have always seen IPF converge in exactly two iterations).

The p-value associated with [12][13][23] based on G^2 is calculated as:

```
1-pchisq(no2nd.loglin$lrt,no2nd.loglin$df)
[1] 0.5130111
```

This p-value indicates that the no second order interaction model fits the Color of Hair and Eyes data well. This is no surprise since the two conditional independence models that also fit the data well are nested within it.

7.2 The Induced Regressions

The regressions induced by the log-linear model [12][13][23] are very similar to the regressions induced by the saturated log-linear model [123] in the sense that each variable is present in the regressions associated with the other two variables. Here are the expressions for these regressions as a function of the u-terms of [12][13][23].

$$\begin{aligned}
\log \frac{P(X_1 = i_1|X_2 = j, X_3 = k)}{P(X_1 = i_2|X_2 = j, X_3 = k)} &= (u_{1(i_1)} - u_{2(i_2)}) + (u_{12(i_1 j)} - u_{12(i_2 j)}) + (u_{13(i_1 k)} - u_{13(i_2 k)}), \\
\log \frac{P(X_2 = j_1|X_1 = i, X_3 = k)}{P(X_2 = j_2|X_1 = i, X_3 = k)} &= (u_{2(j_1)} - u_{2(j_2)}) + (u_{12(i j_1)} - u_{12(i j_2)}) + (u_{23(j_1 k)} - u_{23(j_2 k)}), \\
\log \frac{P(X_3 = k_1|X_1 = i, X_2 = j)}{P(X_3 = k_2|X_1 = i, X_2 = j)} &= (u_{3(k_1)} - u_{3(k_2)}) + (u_{13(i k_1)} - u_{13(i k_2)}) + (u_{23(j k_1)} - u_{23(j k_2)})
\end{aligned}$$

8 Model Selection

Figure 1 summarizes the nine log-linear models for three-way table we discussed. The saturated log-linear model [123] will fit any three-way table perfectly and sits at the top of the hierarchy. The model of complete independence [1][2][3] is the simplest model and sits at the bottom of the hierarchy. It is the least likely model to fit your data well. The other models are placed in between. The arrows indicate which u-terms need to be set to zero to obtain a simpler log-linear model from a more complex log-linear model. Two log-linear models are nested within each other if you can start at the most complex one, follow a number of arrows and reach the simpler model. You can perform a test to see which model you would prefer from two nested models.

Not any two models are nested within each other. Any two models that are placed on the same level in Figure 1 are not nested. This means that you cannot set some u-terms to zero to transform one model in the other model (e.g., [12][13] and [13][23] are not nested). Hence there is statistical model that will help you choose between these two models if they both happen to fit your data well. In such cases you will need to formulate a preference based on which model makes more sense for your data (that is, pick the model with the most credible interpretation).

Fortunately hypothesis testing is enough to help us choose a model for the Color of Hair and

Eyes data. We identified four models other than the saturated model that fit these data well: [12][13][23], [13][23], [12][23] and [1][23]. Let's see which model we prefer between [12][13][23] and [13][23]. That is, we test $H_0: \quad u_{12(ij)} = 0$ using a likelihood ratio test:

```
1-pchisq(X1indepX2givenX3.loglin$lrt-no2nd.loglin$lrt,
                 X1indepX2givenX3.loglin$df-no2nd.loglin$df)
0.01370661
```

We took the difference in the values of G^2 associated with the two models. This difference follows an asymptotic Chi-squared distribution with a number of degrees of freedom equal to the difference in the number of degrees of freedom between the two log-linear models. *EXTREMELY IMPORTANT: For mathematical reasons, you should not test H_0 based on the difference of the X^2 test statistic.* We prefer to be cautious and reject H_0 at the 0.05 level. Therefore we eliminate [13][23] from our list of possible models.

Next we want to test [12][13][23] vs. [12][23]. That is, we want to test the null hypothesis $H_0: \quad u_{13(ik)} = 0$. We use the following R command:

```
1-pchisq(X1indepX3givenX2.loglin$lrt-no2nd.loglin$lrt,
                 X1indepX3givenX2.loglin$df-no2nd.loglin$df)
[1] 0.2509857
```

We fail to fail to reject H_0, favor [12][23] and eliminate [12][13][23] from the list of possible models. At this point our list contains two models: [12][23] and [1][23]. We test the null hypothesis $H_0: \quad u_{12} = 0$ using the R call:

```
1-pchisq(X1indepX2X3.loglin$lrt-X1indepX3givenX2.loglin$lrt,
                 X1indepX2X3.loglin$df-X1indepX3givenX2.loglin$df)
 [1] 0.0564755
```

Based on the above p-value, we fail to reject $H_0: \quad u_{12} = 0$, thus we decide that the log-linear model that is most suitable for the Color of Hair and Eyes data is [1][23]. That is, we learned that Gender is independent of the colors of hair and eyes of a person. Moreover, a person's colors of hair and eyes are related to each other.

9 Tables with Structural Zeros

Some contingency tables contain counts equal to zero. If these counts arise by chance, they are called *sampling zeros*. If these counts arise because that particular combination of categories is impossible, they are called *structural zeros*. If you would be able to observe an arbitrarily large number of samples, all sampling zeros will vanish. However, no matter how many samples you will observe, the structural zeros *never* vanish. As such, structural zeros need to receive special attention in your analysis.

The data in Table 4 is a $4 \times 2 \times 2$ cross-classification of health concerns in teenagers (H), their sex (S) and their age group (A). It is analyzed in your textbook on page 148. This table contains

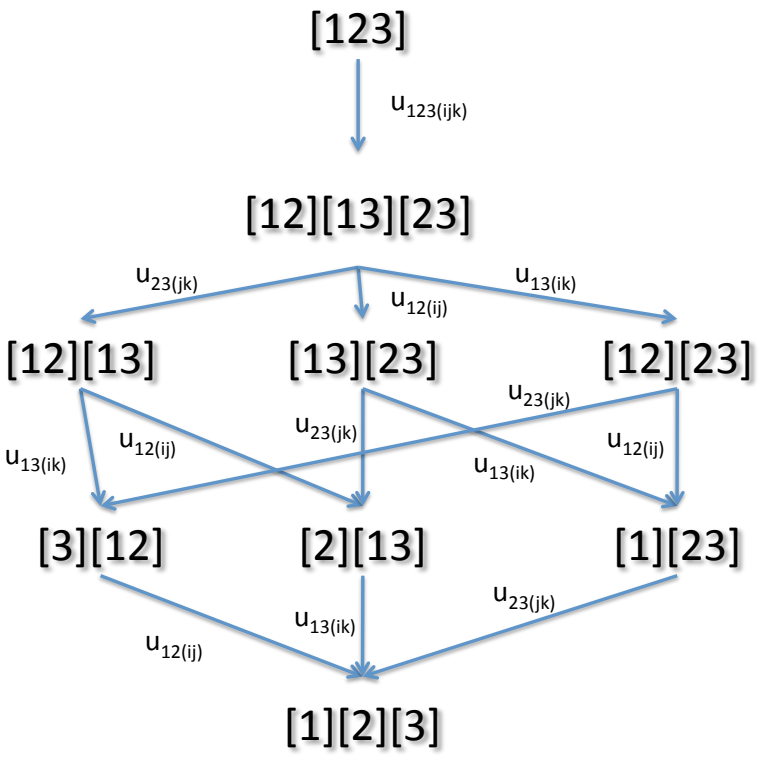

Figure 1: Log-linear models for three-way contingency tables. The arrows and their labels show which u-terms need to be set to zero to obtain one log-linear model from another log-linear model.

two structural zeros induced by the absence of menstrual problems in males. The marginal [HS] (i.e., health concerns vs. sex) also contains a structural zero.Your textbook argues that the number of degrees of freedom associated with a log-linear model needs to be reduced by two unless the [HS] marginal is fitted. In that case the number of degrees of freedom should only be reduced by one. The asymptotic χ^2 p-values for the log-linear models [HS][AH][SA], [AH][SA], [HS][SA] and [HS][AH] corresponding with the Pearson X^2 statistic are given in Table 5 and are based on the calculations reported in Table 8.4, page 149 from your textbook.

Due to the relatively small sample size, it is possible to exhaustively enumerate all the feasible tables associated with the four log-linear models. An examination of the complete set of feasible tables shows that two additional cell counts corresponding with menstrual problems in females are fixed at their observed values of 4 and 8 given the set of marginals [HS][AH][SA], [AH][SA] and [HS][AH]. This implies that the fit of these three models can be assessed in the reduced $3 \times 2 \times 2$ table obtained by eliminating the category "Menstrual problems" altogether. Since the two structural zeros also vanish in this smaller table, there is no need to make any adjustments to the number of degrees of freedom. Usual calculations lead to the same asymptotic p-values obtained in the full $4 \times 2 \times 2$ table.

Table 4: Health concerns of teenagers data from Grizzle and Williams (1972).

	Male		Female	
Health Concerns	12-15	16-17	12-15	16-17
Sex, Reproduction	4	2	9	7
Menstrual Problems	–	–	4	8
How Healthy I am	42	7	19	10
Nothing	57	20	71	31

Table 5: Assessing the fit of four log-linear models for the health concerns of teenagers data.

Log-linear model	Asymptotic χ^2 p-value	Exact χ^2 p-value	Number of tables
[HS][AH][SA]	0.362	0.357 ± 0.001	126
[AH][SA]	0.0107	0.01 ± 0.0002	156240
[HS][SA]	0.087	0.085 ± 0.0004	1252881
[HS][AH]	0.173	0.175 ± 0.0005	6552

The estimated exact p-values for the χ^2 test are given in from Table 5. Remark that the exact and asymptotic p-values are almost equal.

Table 6 is a $4 \times 5 \times 4$ cross-classification of 4345 individuals by occupational groups (O1 – "self-employed, business", O2 – "self-employed, professional", O3 – "teacher", O4 – "salary-employed"), aptitude levels (A) and educational levels (E). It was collected in a 1969 survey of the National Bureau of Economic Research (NBER) – see Table 3-6 page 45 from your textbook.

The horizontal lines denote structural zeros. The ten structural zeros under O3 and E1, E2 are associated with teachers being required to have higher education levels. The other two structural zeros under O2 can be motivated in a similar manner.

We want to assess the fit of the model of all two-way interaction log-linear model. The number of degrees of freedom for this model are calculated by subtracting the number of structural zeros from 36 – the number of degrees of freedom corresponding with a $4 \times 5 \times 4$ table without structural zeros. One needs to add back the number of structural zeros that are present in marginal tables that are among the minimal sufficient statistics of the log-linear model considered. In this case there are two such counts present in the aptitude by educational levels marginal. The resulting number of degrees of freedom is $36 - 12 + 2 = 26$. The observed value of the likelihood-ratio test statistic is $G^2 = 15.91$ which leads to an asymptotic p-value for the all two-way interaction model of 0.938.

The observed value of the χ^2 test statistic is 17.1 which leads to an asymptotic p-value of 0.906.

Table 6: NBER data. The grand total of this table is 4345.

		E1	E2	E3	E4			E1	E2	E3	E4
O1	A1	42	55	22	3	O3	A1	–	–	1	19
	A2	72	82	60	12		A2	–	–	3	60
	A3	90	106	85	25		A3	–	–	5	86
	A4	27	48	47	8		A4	–	–	2	36
	A5	8	18	19	5		A5	–	–	1	14
O2	A1	1	2	8	19	O4	A1	172	151	107	42
	A2	1	2	15	33		A2	208	198	206	92
	A3	2	5	25	83		A3	279	271	331	191
	A4	2	2	10	45		A4	99	126	179	97
	A5	–	–	12	19		A5	36	35	99	79