

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

Adrian Dobra
adobra@uw.edu

1 Introduction

We are now ready to extend the concepts we have seen in 2×2 tables to $I \times J$ tables, where $I \geq 2$ and $J \geq 2$. These tables cross-classify two categorical variables X_1 and X_2 , where $X_1 \in \{1, 2, \dots, I\}$ and $X_2 \in \{1, 2, \dots, J\}$. For example, consider the data in Table 1. It cross-classifies 591 individuals based on their hair and eye color. We denote “hair color” by X_1 and code the four colors of hair with 1, 2, 3 and 4. We denote “eye color” by X_2 and code the four colors of eyes with 1, 2, 3 and 4. Since there are four possible colors of hair, we have $I = 4$. Similarly, we have $J = 4$ since there are four possible colors of eyes.

		Eye Color				Row Totals
		Brown (1)	Blue (2)	Hazel (3)	Green (4)	
Hair Color	Black (1)	68	20	15	5	108
	Brown (2)	119	84	54	29	286
	Red (3)	26	17	14	14	71
	Blond (4)	7	94	10	15	126
Column Totals		220	215	93	63	591

Table 1: Color of hair and eyes data: an example of a 4×4 table.

We will denote the count in cell (i, j) by n_{ij} . This count represents how many persons have hair of color i and eyes of color j . For example, $n_{11} = 68$ means there are 68 persons with black hair and brown eyes. Similarly, $n_{34} = 14$ which means there are 14 persons with red hair and green eyes. The row and the column totals are the one-way marginals of Table 1 corresponding with X_1 and X_2 , respectively. The row totals indicate how many people have been selected in our sample and have black, brown, red or blond hair. Similarly, the column totals indicate how many people have been selected in our sample and have brown, blue, hazel or green eyes.

This table contains $I \cdot J = 16$ cells. These cells will be considered to be ordered column by column, that is, with the leftmost index varying fastest (the convention used in R):

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

The R commands for creating a 4×4 array representing Table 1 are:

```

haireyecolor = c(68,119,26,7,20,84,17,94,15,54,14,10,5,29,14,15)
haireyecolor.array = array(haireyecolor,c(4,4))
> haireyecolor.array
      [,1] [,2] [,3] [,4]
[1,]   68   20   15    5
[2,]  119   84   54   29
[3,]   26   17   14   14
[4,]    7   94   10   15

```

The theoretical results we developed for 2×2 tables extend naturally to $I \times J$ tables. Pretty much all the formulas we wrote in the handout for 2×2 tables will still be true by allowing the index i to take values $\{1, 2, \dots, I\}$ and the index j to take values $\{1, 2, \dots, J\}$. We will always associate the letter/index i with X_1 and the letter/index j with X_2 . For example, we write:

$$\begin{aligned}
P(X_1 = i, X_2 = j) &= p_{ij}, \\
P(X_1 = i) &= p_{i+}, \\
P(X_2 = j) &= p_{+j}, \\
m_{ij} &= n_{++} \cdot p_{ij}
\end{aligned}$$

These are the same notations we introduced in our handout on 2×2 tables.

2 The Log-linear Model of Independence

Under the independence of X_1 and X_2 , we have $p_{ij} = p_{i+} \cdot p_{+j}$ which implies that the estimates of the expected counts under independence are given by

$$\widehat{m}_{ij} = \frac{n_{i+} \cdot n_{+j}}{n_{++}} \quad (1)$$

The log-linear model of independence is written as

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

and involves $1 + I + J$ parameters. We make this model identifiable by imposing either one of the following two types of constraints:

(A) Assume that category 1 of X_1 and X_2 is baseline and set $u_{1(1)} = u_{2(1)} = 0$. *Note: the software (R in this case) will choose an ordering of the categories; this ordering will give the choice of the baseline category. You should know that any category can be taken as baseline. Your model will be the same no matter what this choice is. However, the estimates of the u -terms could be different. Getting different estimates of the u -terms does not imply you made an error or that you are working with a different model.*

(B) Set

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

Since there are a total of $1 + I + J$ parameters and have imposed two constraints, it follows that there are $I + J - 1$ free parameters. Remember that for 2×2 tables (i.e., $I = J = 2$), we argued that we had $2 + 2 - 1 = 3$ free parameters.

From a mathematical point of view, the constraints (A) and (B) are equivalent. However, we favor the constraints of type (B) because they give a better interpretation of the u -terms. Under (B), u is the mean of the logarithms of the expected counts:

$$u = \frac{1}{IJ} \sum_{i=1}^I \sum_{j=1}^J \log m_{ij}$$

Moreover, $u_{1(i)}$ and $u_{2(j)}$ represent deviations from the grand mean u associated with category i of X_1 and with category j of X_2 , respectively. It follows that $u + u_{1(i)}$ is the mean of the logarithms of the expected counts in the J cells at level i of X_1 :

$$u + u_{1(i)} = \frac{1}{J} \sum_{j=1}^J \log m_{ij}$$

Similarly, $u + u_{2(j)}$ is the mean of the logarithms of the expected counts in the I cells at level j of X_2 :

$$u + u_{2(j)} = \frac{1}{I} \sum_{i=1}^I \log m_{ij}$$

The number of degrees of freedom of the log-linear model of independence is the difference between the number of cells in the table and the number of free parameters of the model:

$$I \cdot J - (I + J - 1) = (I - 1) \cdot (J - 1)$$

Remember that for 2×2 tables we argued that the log-linear model of independence has 1 degree of freedom: $(2 - 1) \cdot (2 - 1) = 1$. Let's fit the model of independence to the data in Table 1:

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

```
2 iterations: deviation 2.842171e-14
```

The meaning of the last line of the output is that the Iterative Proportional Fitting (IPF) algorithm has converged in two iterations. The function “loglin” does not actually use the formula (1) to determine the estimates of the expected cell values \widehat{m}_{ij} . Instead, it uses IPF which works for any contingency table and any log-linear model. At the first iteration, IPF reaches the correct estimates of m_{ij} as given in (1). It takes one more iteration for the IPF to see that these estimates do not change, hence IPF determines it has converged, decides to stop and output the estimates of the expected cell values \widehat{m}_{ij} . *Note: please see Problem 2 from Homework 4 for details on IPF for 2×2 tables. The description of IPF from Problem 2 holds for arbitrary 2×2 tables – just allow i and j to take values $\{1, 2, \dots, I\}$ and $\{1, 2, \dots, J\}$, respectively.*

```
> indep.loglin
$lrt
[1] 146.1795
```

```
$pearson
[1] 138.6292
```

```
$df
[1] 9
```

This means that the value of the likelihood ratio test statistic G^2 is 146.18, the value of the X^2 statistic is 138.63 and that the log-linear model of independence has 4 degrees of freedom. In this example we have $I = J = 4$, hence $(I - 1) \cdot (J - 1) = 3 \cdot 3 = 9$.

```
$margin
$margin[[1]]
[1] 1
```

```
$margin[[2]]
[1] 2
```

These are the MSS of the log-linear model that was fit to the data. The first fixed marginal is the one-dimensional marginal corresponding with X_1 (i.e., the row totals). The second fixed marginal is the one-dimensional marginal corresponding with X_2 (i.e., the column totals).

```
$fit
      [,1]      [,2]      [,3]      [,4]
[1,] 40.20305 39.28934 16.99492 11.512690
[2,] 106.46362 104.04399 45.00508 30.487310
[3,] 26.42978 25.82910 11.17259 7.568528
[4,] 46.90355 45.83756 19.82741 13.431472
```

These are estimated expected cell values m_{ij} under the model of independence. Let's compute the estimate of m_{11} directly using equation (1):

$$\widehat{m}_{11} = \frac{n_{1+} \cdot n_{+1}}{n_{++}} = \frac{108 \cdot 220}{591} = 40.20$$

It would be an useful exercise to compute the estimated expected values for other cells, just to make sure you understand what is going on.

```
$param
$param$'(Intercept)'
[1] 3.337455
```

```
$param$'1'
```

[1] -0.1771400 0.7967206 -0.5965913 -0.0229893

\$param\$'2'

[1] 0.5336276 0.5106381 -0.3274005 -0.7168652

These are the estimates of the u -terms: $\widehat{u} = 3.34$, $\widehat{u}_{1(1)} = -0.177$, $\widehat{u}_{1(2)} = 0.796$, $\widehat{u}_{1(3)} = -0.597$, $\widehat{u}_{1(4)} = -0.023$, $\widehat{u}_{2(1)} = 0.534$, $\widehat{u}_{2(2)} = 0.51$, $\widehat{u}_{2(3)} = -0.33$, $\widehat{u}_{2(4)} = -0.717$. Remark that the identifiability constraints (B) are satisfied:

$$\begin{aligned}\widehat{u}_{1(1)} + \widehat{u}_{1(2)} + \widehat{u}_{1(3)} + \widehat{u}_{1(4)} &= 0, \\ \widehat{u}_{2(1)} + \widehat{u}_{2(2)} + \widehat{u}_{2(3)} + \widehat{u}_{2(4)} &= 0\end{aligned}$$

3 The Saturated Log-linear Model

The saturated log-linear model expresses that X_1 and X_2 are not independent through the interaction parameters $u_{12(ij)}$:

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

There are $I \times J$ such terms interaction parameters – one parameter $u_{12(ij)}$ for each cell (i, j) in the table. We make the saturated log-linear model identifiable by imposing one of the following two types of constraints:

(A) Assume that category 1 of X_1 and X_2 is baseline and set

$$\begin{aligned}u_{12(11)} &= 0, \\ u_{12(1j)} &= 0, \text{ for } j = 2, \dots, J \\ u_{12(i1)} &= 0, \text{ for } i = 2, \dots, I\end{aligned}$$

We are setting to zero $1 + (I-1) + (J-1) = I + J - 1$ terms, hence we have $IJ - (I + J - 1) = (I-1)(J-1)$ free interaction terms $u_{12(ij)}$.

(B) Set

$$\begin{aligned}\sum_{i=1}^I u_{12(ij)} &= 0, \text{ for } j = 1, 2, \dots, J, \\ \sum_{j=1}^J u_{12(ij)} &= 0, \text{ for } i = 1, 2, \dots, I\end{aligned}$$

The first J sum-to-zero constraints imply that the sum of all the interaction terms is zero:

$$\sum_{i=1}^I \sum_{j=1}^J u_{12(ij)} = \sum_{j=1}^J \left(\sum_{i=1}^I u_{12(ij)} \right) = 0$$

The same relation is implied by the last I sum-to-zero constraints:

$$\sum_{i=1}^I \sum_{j=1}^J u_{12(ij)} = \sum_{i=1}^I \left(\sum_{j=1}^J u_{12(ij)} \right) = 0$$

This means that we do not have $I+J$ independent constraints. Instead, we have $I+J-1$ independent constraints, which implies that the number of free interaction terms is $IJ - (I+J-1) = (I-1)(J-1)$. Remark that we get the same number of free interaction terms by imposing the baseline constraints (A). In other words, (A) and (B) are mathematically equivalent, but we *prefer* to impose the sum-to-zero constraints (B) to be able to interpret the u-terms in a more appropriate manner.

The total number of free u -terms in the saturated log-linear model is

$$1 + (I - 1) + (J - 1) + (I - 1)(J - 1) = IJ$$

Therefore, the number of degrees of freedom corresponding with the saturated log-linear model is zero (the number of cells in the table is equal with the number of free parameters):

$$IJ - IJ = 0$$

Again, this is why the estimates of the expected cell values under the saturated log-linear model coincide with the observed cell counts:

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

A consequence of this equality is that the likelihood ratio statistic G^2 and the X^2 statistic are zero (*this is true for any table, not just for the hair and eye color 4×4 table*):

$$G^2 = 2 \cdot \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \frac{n_{ij}}{\widehat{m}_{ij}} = 2 \cdot \sum_{i=1}^I \sum_{j=1}^J n_{ij} \log \frac{n_{ij}}{n_{ij}} = 0,$$

$$X^2 = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - \widehat{m}_{ij})^2}{\widehat{m}_{ij}} = \sum_{i=1}^I \sum_{j=1}^J \frac{(n_{ij} - n_{ij})^2}{n_{ij}} = 0$$

We fit the saturated log-linear model for the hair and eye color data in Table 1 using the R command:

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

2 iterations: deviation 0
```

The IPF algorithm found the estimated expected cell values $\widehat{m}_{ij} = n_{ij}$ after one iteration, then it made another iteration and realized that the estimated expected cell values did not change. Therefore IPF determined it has converged and stops. The output looks as follows.

```
$lrt
[1] 0

$pearson
[1] 0

$df
[1] 0
```

This means that $G^2 = 0$, $X^2 = 0$ and the number of degrees of freedom is zero.

```
$margin
$margin[[1]]
[1] 1 2
```

This means that the log-linear model has one MSS and this MSS is the marginal associated with X_1 and X_2 . This is precisely the observed data in Table 1.

```
$fit
      [,1] [,2] [,3] [,4]
[1,]    68   20   15    5
[2,]   119   84   54   29
[3,]    26   17   14   14
[4,]     7   94   10   15
```

These are the estimated expected cell values \widehat{m}_{ij} under the saturated log-linear model. Again, they coincide with the observed cell counts n_{ij} in Table 1.

```
$param
$param$(Intercept)
[1] 3.185513
```

```
$param$'1'
[1] -0.3023313  0.9560417 -0.3431572 -0.3105533
```

```
$param$'2'
[1] 0.3651462  0.5152510 -0.2758441 -0.6045530
```

```
$param$'1.2'
      [,1]      [,2]      [,3]      [,4]
[1,] 0.97117952 -0.4027007  0.10071232 -0.6691911
[2,] 0.27242229 -0.2259892  0.12327315 -0.1697062
[3,] 0.05059425 -0.5243938  0.07254534  0.4012542
[4,] -1.29419606  1.1530837 -0.29653082  0.4376431
```

These are the estimates of the u -terms: first we have the estimate of u , then the estimates of $u_{1(i)}$, followed by the estimates of $u_{2(j)}$ and the estimates of $u_{12(ij)}$. *Please check that these estimates satisfy the sum-to-zero constraints of type (B)!*

4 Testing Independence vs. Interaction

We want to test whether somebody's hair color gives no information about their eye color (i.e., hair and eye color are independent or $X_1 \perp\!\!\!\perp X_2$) versus the possibility that somebody's hair and

eye colors are somehow related (i.e., hair and eye color are not independent or $X_1 \not\perp X_2$). Using the log-linear parametrization we discussed in the previous section, we would like to test the null hypothesis:

$$H_0 : u_{12(ij)} = 0, \text{ for } i = 1, \dots, I \text{ and } j = 1, \dots, J$$

vs. the alternative

$$H_A : u_{12(ij)} \neq 0, \text{ for some } i \in \{1, \dots, I\} \text{ and } j \in \{1, \dots, J\}$$

Remark: the alternative H_A says that at least one interaction term should be zero, not all interaction terms are zero. We plan to test H_0 vs. H_A based on the asymptotic distribution of the likelihood ratio statistic G^2 or the X^2 statistic. The difference between the value of G^2 under the log-linear model of independence (denoted by G_{indep}^2) and the value of G^2 under the saturated log-linear model (this value is *always* zero, hence we will ignore it) follows (asymptotically) a Chi-squared distribution with a number of degrees of freedom equal to the difference between the number of degrees of freedom of the log-linear model of independence $((I - 1) \cdot (J - 1))$ for $I \times J$ tables) and the number of degrees of freedom of the saturated log-linear model (always zero).

Let's test the independence of hair and eye color in Table 1:

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

We obtain an asymptotic p-value equal to zero, hence we reject the hypothesis of independence. We say that the data provides evidence that hair and eye color are indeed related. We perform an asymptotic test based on the X^2 statistic:

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

Once again, we obtain an asymptotic p-value equal to zero. Table 1 has quite large counts in each cell and does not have any structural zeros. Therefore we would expect the asymptotic p-values to be accurate. *Remark: if your table has many zeros (i.e., the table is sparse) you should really doubt the accuracy/validity of any asymptotic p-values you calculate.*

Let's now calculate the exact p-value for testing H_0 vs. H_A . We cannot enumerate all the 4×4 tables consistent with the row and column totals of Table 1, hence we will need to estimate the value of the exact p-value using the R package `exactLoglinTest` written by Brian Caffo. To this end, we will need to organize Table 1 in a file as follows:

```
y Hair Eye
68 1 1
20 1 2
15 1 3
5 1 4
119 2 1
84 2 2
```



```

54 2 3
29 2 4
26 3 1
17 3 2
14 3 3
14 3 4
7 4 1
94 4 2
10 4 3
15 4 4

```

Each row is associated with a cell in the table. The count that appears in each cell is given in the first column of the corresponding row. This is the same format you need to use if you fit log-linear models using the function 'glm' (please see the handout on 2×2 tables). *Remark: once again, the function 'glm' will fit log-linear models based on baseline constraints of type (A). Thus the estimates of the u-terms you obtain from 'glm' not be the same as the estimates you get from 'loglin'.*

Let's say you saved your table as shown above in a file called 'haireyecolor.txt'. We read this file in R, then use the 'mcexact' function to estimate the exact p-value for testing H_0 vs. H_A .

```

haireye.data = read.table('haireyecolor.txt',header=TRUE)
library(mcexact)
set.seed(1)

```

First we calculate the exact p-value using importance sampling:

```
#importance sampling
```

```

mcx = mcexact(y ~ factor(Hair) + factor(Eye), data = haireye.data)
summary(mcx)

```

```

Number of iterations      = 935
t degrees of freedom      = 3
Number of counts          = 16
df                         = 9
Next update has nosim     = 1000
Next update has maxiter   = 1000
Proportion of valid tables = 0.935

```

	deviance	Pearson
observed.stat	1.461795e+02	138.6292
pvalue	2.526520e-30	0.0000
mcse	1.385848e-23	0.0000

We see that we obtained the same values of G^2 (called deviance in the output above) and X^2 (called Pearson in the output above) as we calculated before using 'loglin'. Moreover, we obtained that

the estimated exact p-value is zero for based on both G^2 and X^2 .

Now we perform the same calculation using a Markov chain Monte Carlo algorithm:

```
mcx = mcexact(y ~ factor(Hair) + factor(Eye), data = haireye.data,
              method = "cab", p = .5, nosim = 10 ^ 4, batchsize = 100)
summary(mcx)
```

```
Number of iterations      = 10000
t degrees of freedom      = 3
Number of counts          = 16
df                        = 9
Number of batches         = 100
Batchsize                 = 100
Next update has nosim     = 10000
Proportion of valid tables = 0.962
```

	deviance	Pearson
observed.stat	146.1795	138.6292
pvalue	0.0000	0.0000
mcse	0.0000	0.0000

Understanding what statistical procedures involved in 'mcexact' is beyond the scope of our course. What you need to remember is that you can obtain more accurate estimates by increasing the number of simulations (the parameter 'nosim' that was set to 10000 in the example above) and by varying the parameter 'p' between 0 and 1 (this parameter was set to 0.5 in the example above).

5 The Connection Between Log-linear and Multinomial Models

The log-linear model of independence translates into the null regression of X_1 on X_2 and into the null regression of X_2 and X_1 . That is, if X_1 is independent of X_2 , the log odds of any two categories of X_1 do not depend on X_2 .

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

You see that the expression above does not depend on the category j of X_2 . Therefore, it is fine to write:

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

which means

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

For example, let's say we want to determine the odds of having black hair ($i_1 = 1$) vs. red hair ($i_2 = 3$) based on the estimated u-terms of the log-linear model of independence. We obtain

$$\begin{aligned} \log \frac{P(\text{Hair Color}=\text{Black})}{P(\text{Hair Color}=\text{Red})} &= \widehat{u}_{1(1)} - \widehat{u}_{1(3)}, \\ &= -0.177 - (-0.597) = 0.42 \end{aligned}$$

That is, the odds of a person having black hair vs. red hair are $\exp(0.42) = 1.52$. Notice that this *happens* to be equal with ratio of the number of people in the sample who had black hair (108) and the number of people in the sample who had red hair (71), i.e. $(108/71 = 1.52)$.

Similar relations hold for the odds of two categories j_1 and j_2 of X_2 under the model of independence:

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

For example, the odds of having blue eyes ($j_1 = 2$) vs. hazel eyes ($j_2 = 3$) is estimated to be:

$$\begin{aligned} \log \frac{P(\text{Eye Color}=\text{Blue})}{P(\text{Eye Color}=\text{Hazel})} &= \widehat{u}_{2(2)} - \widehat{u}_{2(3)}, \\ &= 0.511 - (-0.327) = 0.838 \end{aligned}$$

That is, the odds of a person having blue eyes vs. hazel eyes is $\exp(0.838) = 2.31$. Notice that this *happens* to be equal with the ratio of the number of people in the sample who had blue eyes (215) and the number of people in the sample who had hazel eyes (93), i.e. $215/93 = 2.31$.

However, we determined that the independence model does not fit the data in Table 1 well. In other words, we determined that the hair and the eye color of a person are related to each other. This means that the odds of somebody having a hair color vs. another hair color depends on their eye color. Furthermore, this means that we estimate conditional odds based on the saturated log-linear model. It follows that

$$\begin{aligned} \log \frac{P(X_1 = i_1 | X_2 = j)}{P(X_1 = i_2 | X_2 = j)} &= (u + u_{1(i_1)} + u_{2(j)} + u_{12(i_1 j)}) - (u + u_{1(i_2)} + u_{2(j)} + u_{12(i_2 j)}), \\ &= (u_{1(i_1)} - u_{1(i_2)}) - (u_{12(i_1 j)} - u_{12(i_2 j)}) \end{aligned}$$

For example, the odds of having black hair ($i_1 = 1$) vs. red hair ($i_2 = 3$) given each eye color is:

$$\begin{aligned}
\log \frac{P(\text{Hair Color=Black}|\text{Eye Color = Brown})}{P(\text{Hair Color=Red}|\text{Eye Color = Brown})} &= (\widehat{u}_{1(1)} - \widehat{u}_{1(3)}) - (\widehat{u}_{12(11)} - \widehat{u}_{12(31)}), \\
&= (-0.302 - (-0.343)) - (0.971 - 0.051) = -0.879 \\
\log \frac{P(\text{Hair Color=Black}|\text{Eye Color = Blue})}{P(\text{Hair Color=Red}|\text{Eye Color = Blue})} &= (\widehat{u}_{1(1)} - \widehat{u}_{1(3)}) - (\widehat{u}_{12(12)} - \widehat{u}_{12(32)}), \\
&= (-0.302 - (-0.343)) - (-0.403 - (-0.524)) = -0.08 \\
\log \frac{P(\text{Hair Color=Black}|\text{Eye Color = Hazel})}{P(\text{Hair Color=Red}|\text{Eye Color = Hazel})} &= (\widehat{u}_{1(1)} - \widehat{u}_{1(3)}) - (\widehat{u}_{12(13)} - \widehat{u}_{12(33)}), \\
&= (-0.302 - (-0.343)) - (0.101 - 0.073) = 0.013 \\
\log \frac{P(\text{Hair Color=Black}|\text{Eye Color = Green})}{P(\text{Hair Color=Red}|\text{Eye Color = Green})} &= (\widehat{u}_{1(1)} - \widehat{u}_{1(3)}) - (\widehat{u}_{12(14)} - \widehat{u}_{12(34)}), \\
&= (-0.302 - (-0.343)) - (-0.669 - 0.401) = 1.111
\end{aligned}$$

Similarly, the saturated log-linear model implies that the conditional odds of two categories of X_2 given X_1 is expressed as:

$$\begin{aligned}
\log \frac{P(X_2 = j_1|X_1 = i)}{P(X_2 = j_2|X_1 = i)} &= (u + u_{1(i)} + u_{2(j_1)} + u_{12(ij_1)}) - (u + u_{1(i)} + u_{2(j_2)} + u_{12(ij_2)}), \\
&= (u_{2(j_1)} - u_{2(j_2)}) - (u_{12(ij_1)} - u_{12(ij_2)})
\end{aligned}$$

For the data in Table 1, these relations allow us to estimate the conditional odds of somebody having an eye color vs. another eye color conditional on that person's hair color.