8 Log-linear Models for Contingency Tables

8.1 Introduction

Data can arise in the form of counts of the number of units possessing certain combinations of attributes, or characteristics. These data can be presented in the form of contingency tables.

Here are 3 such examples:

TABLE 1: Yule (1900)

	Wife			
Husband	Tall	Medium	Short	**Totals**
Tall	18	28	14	60
Medium	20	51	28	99
Short	12	25	9	46
Totals	50	104	51	205

TABLE 2: Bishop (1969)

Clinic	*Antenatal care*	Survived	Died	**Totals**
A	Low	176	3	179
	High	293	4	297
B	Low	197	17	214
	High	23	2	25
	Totals	689	26	715

TABLE 3: Ashford & Snowdon (1970)

Age group	Breathlessness		No breathlessness		
in years	Wheeze	No Wheeze	Wheeze	No Wheeze	**Totals**
20-24	9	7	95	1841	1952
25-29	23	9	105	1654	1791
30-34	54	19	177	1863	2113
35-39	121	48	257	2357	2783
40-44	169	54	273	1778	2274
45-49	269	88	324	1712	2393
50-54	404	117	245	1324	2090
55-59	406	152	225	967	1750
60-64	372	106	132	526	1136
Totals	1827	600	1833	14,022	18,282

Table 1 classifies 205 married couples according to the heights of each partner.

Table 2 classifies 715 babies according to their 'survival' (survived or died), clinic attended (clinic A or B), and level of ante-natal care received (low or high).

Table 3 classifies coal miners according to their age (20-24,25-29,. . .,60-64), 'breathing' (breathlessness or no breathlessness), and 'wheeziness' (wheeze or no wheeze).

8.2 Sampling Schemes

We need to carefully consider how the data in these tables were collected as this will determine the appropriate probability models to associate with the data, as well as the types of hypotheses that can be tested for.

Case (a): Nothing Fixed
Here, we simply record the number of individuals falling into particular categories over the time period of interest.
The sample size, as well as the cell counts, are realizations of random variables.

Case (b): Total Sample Size Fixed
Record information on the individuals, according to the classifications, up to a certain (possibly pre-specified) sample size, and then stop.

Case (c): One or more margins Fixed
In table 2, the 'clinic' margin is considered to be fixed alone if information on the level of antenatal care, and survival (in some specified period) on:
179+297=476 babies from clinic A
214+25=239 babies from clinic B

is recorded. Thus 476 and 239 are both fixed by design.
By implication, the sample size is fixed also.

Altenatively, let us suppose that the 'clinic'×'care' two-way margin is fixed. Then the following quantities are fixed by design:
179 for clinic A, low antenatal care
297 for clinic A, high antenatal care
214 for clinic B, low antenatal care
25 for clinic B, high antenatal care.

In general, if we have a contingency table with m classifying factors, then up to $m-1$ margins can be fixed by design.

For the rest of this chapter we will focus our discussion on the analysis of two-way tables. Three-way (and higher-way) tables will be discussed in the next chapter.

8.3 Probability Distributions

Consider a two-way table with two factors, A and B, where the former occurs at J levels, and the latter at K levels. Let Y_{jk} be the frequency for the (j,k)-th cell of the table. Also let

$$Y_{j.} = \sum_{k=1}^{K} Y_{jk}, \quad j=1,\ldots,J,$$

and

$$Y_{.k} = \sum_{j=1}^{J} Y_{jk}, \quad k=1,\ldots,K,$$

be the row and column totals respectively.

Case (a):
Here, it is supposed that the Y_{jk} are mutually independent and follow a Poisson distribution, each with parameter λ_{jk}. Label the realizations of these cell counts by the y_{jk}. Then the joint distribution of the cell counts is:

$$f(\mathbf{y};\boldsymbol{\lambda}) = \prod_{j=1}^{J}\prod_{k=1}^{K} \frac{\lambda_{jk}^{y_{jk}} e^{-\lambda_{jk}}}{y_{jk}!} \tag{1}$$

Also note that if we set $\lambda_{..} = \sum_{j=1}^{J}\sum_{k=1}^{K} \lambda_{jk}$, then the distribution of the sample size is

$$f(n) = \frac{\lambda_{..}^{n} e^{-\lambda_{..}}}{n!} \tag{2}$$

Case (b):
Using (1) and (2), the conditional distribution of the cell counts, given that the total sample size, $\sum_{j=1}^{J}\sum_{k=1}^{K} y_{jk}$, is equal to n, is:

$$
\begin{aligned}
f(\mathbf{y}|n) &= f(\mathbf{y}, n)/f(n) && (3)\\
&= \frac{\prod_{j=1}^{J}\prod_{k=1}^{K}\frac{\lambda_{jk}^{y_{jk}}e^{-\lambda_{jk}}}{y_{jk}!}}{\frac{\lambda_{..}^{n}e^{-\lambda_{..}}}{n!}} = n!\prod_{j=1}^{J}\prod_{k=1}^{K}\frac{\theta_{jk}^{y_{jk}}}{y_{jk}!} && (4)
\end{aligned}
$$

where $\theta_{jk} = \lambda_{jk}/\lambda_{..}$.

Thus, when the sample size is fixed by design, the cell counts follow a *Multinomial distribution* with parameters n and $\{\theta_{jk}\}$.

Case (c):
Suppose that we fix the margin corresponding to factor A. This corresponds to the row totals, $\{y_{j.}\}$, being fixed by design. We say that A is an *explanatory* variable or factor, whereas B is a *response* factor.
By analogy with Case (b), the (joint) distribution of the cell counts at the j-th row of the table, is again Multinomial, with parameters $y_{j.}$, and $\{\theta_{jk}\}$, where $\sum_{k=1}^{K}\theta_{jk} = 1$:

$$
f(y_{j1}, \ldots, y_{jK}|y_{j.}) = y_{j.}!\prod_{k=1}^{K}\frac{\theta_{jk}^{y_{jk}}}{y_{jk}!}
$$

Assuming that the rows are mutually independent, the joint distribution of the cell counts for the whole table is:

$$
f(\mathbf{y}|y_{j.}, j=1, \ldots, J) = \prod_{j=1}^{J} y_{j.}!\prod_{k=1}^{K}\frac{\theta_{jk}^{y_{jk}}}{y_{jk}!} \qquad (5)
$$

where $\sum_{k=1}^{K}\theta_{jk} = 1$ for $j=1, \ldots, J$.

8.4 Log-linear Models

Recall the following result regarding the mean of the Multinomial distribution.

Lemma 8.1
Suppose that the joint distribution of $Y_1, \ldots, Y_N$ is Multinomial with parameters n and $\{\theta_i\}$, thus

$$
f(\mathbf{y}) = n!\prod_{i=1}^{N}\frac{\theta_i^{y_i}}{y_i!}
$$

where the $\{\theta_i\}$ are non-negative, $\sum_{i=1}^{N} \theta_i = 1$, and $\sum_{i=1}^{N} y_i = n$.

Then

$$E[Y_i] = n\theta_i$$

We shall use this general result to compute the expectations of the cell frequencies in cases (b) and (c). In so doing, we also consider the forms of these expressions under certain hypotheses of interest, and then transform these into 'log-linear form'.

Case (a):
Here, we are simply dealing with the expectations of Poisson random variables, and so

$$E[Y_{jk}] = \lambda_{jk} \tag{6}$$

Now under the hypothesis of *independence* between factors A and B, the probability for the (j,k)-th cell of the table, θ_{jk}, can be written as

$$\theta_{jk} = \theta_{j.} \times \theta_{.k}$$

where $\sum_{k=1}^{K} \sum_{j=1}^{J} \theta_{jk} = 1$, and $\theta_{j.}$ and $\theta_{.k}$ are the marginal probabilities of landing in the j-th row and k-th column of the table, respectively. But $\theta_{jk} = \lambda_{jk}/\lambda_{..}$, $\theta_{j.} = \lambda_{j.}/\lambda_{..}$, and $\theta_{.k} = \lambda_{.k}/\lambda_{..}$. Therefore

$$\lambda_{jk} = \lambda_{j.} \times \lambda_{.k}/\lambda_{..}$$

Hence, under independence,

$$E[Y_{jk}] = \lambda_{jk} = \lambda_{j.} \times \lambda_{.k}/\lambda_{..} \tag{7}$$

Taking the logarithm of (6), and with an appropriate re-definition of the resulting terms on the RHS, we have

$$\eta_{jk} = \log E[Y_{jk}] = \mu + \alpha_j + \beta_k + (\alpha\beta)_{jk} \tag{8}$$

Applying the same procedure to the 'independence' expression (7) yields

$$\eta_{jk} = \log E[Y_{jk}] = \mu + \alpha_j + \beta_k \tag{9}$$

By analogy with analysis of variance, the hypothesis of independence between factors A and B is equivalent to $(\alpha\beta)_{jk} = 0$ for $j=1,\ldots,J$, $k=1,\ldots,K$, i.e. the 'interaction' terms are zero.

Case (b):
Under this distribution

$$E[Y_{jk}] = n\theta_{jk} \tag{10}$$

Now under the hypothesis of *independence* between factors A and B, the probability for the (j,k)-th cell of the table, θ_{jk}, can be written as

$$\theta_{jk} = \theta_{j.} \times \theta_{.k}$$

where $\theta_{j.}$ and $\theta_{.k}$ are the marginal probabilities of landing in the j-th row and k-th column of the table, respectively. Hence, under independence,

$$E[Y_{jk}] = n \times \theta_{j.} \times \theta_{.k} \tag{11}$$

Again, taking the logarithms of (10) and (11) yield

$$\eta_{jk} = \log n + \log \theta_{jk} \tag{12}$$

Applying the same procedure to the 'independence' expression (11) yields

$$\eta_{jk} = \log n + \log \theta_{j.} + \log \theta_{.k} \tag{13}$$

For this case, it again turns out that the full model (12), and the independence model (13) can be represented by (8) and (9) respectively.

Case (c):
Under this distribution

$$E[Y_{jk}] = y_{j.}\theta_{jk} \tag{14}$$

where $\sum_k \theta_{jk} = 1$ for $j = 1, \ldots, J$.

Now consider the hypothesis of 'homogeneity', in which the conditional probability for being in the k-th cell given the j-th row, θ_{jk}, is equal to the (unconditional) marginal probability of being in the k-th column of the table, $\theta_{.k}$ say, i.e.

$$\theta_{jk} = \theta_{.k}$$

Hence, under homogeneity,

$$E[Y_{jk}] = y_{j.}\theta_{.k} \tag{15}$$

Taking the log's of (14) and (15) yields

$$\eta_{jk} = \log E[Y_{jk}] = \log y_{j.} + \log \theta_{jk}$$

and

$$\eta_{jk} = \log E[Y_{jk}] = \log y_{j.} + \log \theta_{.k}$$

which can be related to (8) and (9) respectively.

So this time, homogeneity (rather than independence, which has no meaning in this case anyway), requires that all of the 'interaction' terms are equal to zero.

The 'log-linear models' that have arisen from the above discussion actually amount to saying that

$$\eta_i = \log E[Y_i] = \mathbf{x}_i' \, \boldsymbol{\xi}, \quad i = 1, \ldots, N \tag{16}$$

for appropriately chosen $\mathbf{x}_i$ and $\boldsymbol{\xi}$. Thus, we relate the mean of the cell counts, to the linear predictor η_i through the log-link!!! Furthermore, in (a), we have a Poisson error structure for these counts; thus, in this case, we can test hypotheses using the corresponding GLM. The vector of parameters $\boldsymbol{\xi}$ is estimated using maximum likelihood estimation, which, by the invariance property, yields the fitted means, $\hat{\mu}_i = e^{\hat{\eta}_i}$.

But what about cases (b) and (c)?

(Product-) Multinomial data may be analyzed as though they were independent Poisson data, with log-linear predictor, PROVIDED TERMS CORRESPONDING TO THE FIXED MARGINS ARE INCLUDED IN THE MODEL. The deviance will be correct, as well as the estimates.

So, for example, if the row totals are fixed by design, i.e. the $\{y_{j.}\}$ are fixed, then we always fit the terms

$$\mu + \alpha_j$$

The other terms in the linear predictor of the full model are

$$\beta_k + (\alpha\beta)_{jk}$$

corresponding to θ_{jk}: terms in this part of the predictor can be removed to test for the various hypotheses.

As is the case with ANOVA models, our statistical model is over parameterized. Thus, in order to obtain (unique) estimates, we need to impose appropriate constraints. One such choice are the *sum-to-zero* constraints:

$$\sum_{j=1}^{J} \alpha_j = 0, \quad \sum_{k=1}^{K} \beta_k = 0, \quad \sum_{j=1}^{J} (\alpha\beta)_{jk} = 0, \quad \sum_{k=1}^{K} (\alpha\beta)_{jk} = 0.$$

Another choice would be some form of the *corner point* constraints:

$$\alpha_1 = 0, \quad \beta_1 = 0, \quad (\alpha\beta)_{1k} = 0, \ k=1,\ldots,K, \quad (\alpha\beta)_{j1} = 0, \ j=1,\ldots,J.$$

Imposing either set of these constraints on the 'full' model, the number of free parameters are: 1 from μ, $(J-1)$ from the $\{\alpha_j\}$, $(K-1)$ from the $\{\beta_k\}$, and $(J-1)(K-1)$ from the $\{(\alpha\beta)_{jk}\}$, which adds up to JK, which is the total number of cells in the table! Thus, our full model (8) corresponds to the 'saturated' model.

Example 8.2 (Heights)
Suppose that the data in TABLE 1 were collected in such a way that only the total sample size is fixed by design. Determine whether there is any association between the height of the husband and that of the wife.

Solution:

```
> count <- c(18, 20, 12, 28, 51, 25, 14, 28, 9)
> w <- c("T", "T", "T", "M", "M", "M", "S", "S", "S")
> h <- c("T", "M", "S", "T", "M", "S", "T", "M", "S")
> wife <- factor(w)
> husband <- factor(h)
> height <- data.frame(count, wife, husband)
> rm(count, w, h, wife, husband)
> height.glm <- glm(count ~ husband + wife, data = height, family = poisson)
> summary(height.glm)

Call: glm(formula = count ~ husband + wife, family = poisson, data = height)
Deviance Residuals:
        1          2         3          4         5         6          7         8
 0.849001 -0.8698614 0.2303863 -0.4481901 0.1091626 0.3403618 -0.2424412 0.6645306

         9
 -0.7507465

Coefficients:
                  Value Std. Error     t value
(Intercept)  3.01243847 0.07733321  38.9540092
   husband1 -0.38323923 0.08921696  -4.2955873
   husband2 -0.03917869 0.05230846  -0.7489933
      wife1 -0.35628263 0.08547240  -4.1683941
      wife2 -0.12536175 0.05507896  -2.2760370

(Dispersion Parameter for Poisson family taken to be 1 )

    Null Deviance: 50.58898 on 8 degrees of freedom

Residual Deviance: 2.923175 on 4 degrees of freedom

Number of Fisher Scoring Iterations: 3
```

The (scaled) deviance of 2.923175 on the χ^2_4 distribution is not significant. Thus, there is no evidence to reject the hypothesis of independence.

The fitted means can be obtained under this 'independence' model:

```
> fitted(height.glm)
        1        2        3        4        5        6        7        8       9
 14.63415 24.14634 11.21951 30.43902 50.22439 23.33659 14.92683 24.62927 11.4439
```

Note that the estimates of the parameters in our linear predictor are **not** the ones in the ’`Coefficients`’ section of the above output; in fact, they can be obtained by using the `dummy.coef` function.

```
> dummy.coef(height.glm)
$"(Intercept)":
 (Intercept)
    3.012438

$husband:
         M          S           T
 0.4224179 -0.3440605 -0.07835737

$wife:
         M          S          T
 0.4816444 -0.2309209 -0.2507235
```

Thus

$$\widehat{\mu} = 3.012438, \quad \widehat{\alpha}_M = 0.4224179, \quad \widehat{\alpha}_S = -0.3440605, \ldots, \widehat{\beta}_T = -0.2507235$$

By default, these correspond to the ’sum-to-zero’ constraints.

We can also extract parameter estimates using the ’corner-point’ constraints:

```
> options(contrasts = c("contr.treatment", "contr.poly"))
> height.glm <- glm(count ~ husband + wife, data = height, family = poisson)
> dummy.coef(height.glm)
$"(Intercept)":
 (Intercept)
    3.916501

$husband:
 M          S          T
 0 -0.7664785 -0.5007753

$wife:
 M          S          T
 0 -0.7125653 -0.7323679
```

This time

$$\widehat{\mu} = 3.916501, \quad \widehat{\alpha}_M = 0, \quad \widehat{\alpha}_S = -0.7664785, \ldots, \ldots, \widehat{\beta}_T = -0.7323679$$

To return back to the *sum-to-zero* constraints, invoke
`options(contrasts = c("contr.helmert", "contr.poly"))`
and then re-fit the model.