

Log-Linear models: describing count data

Chapter 5 says, R code for contingency tables is given in Chapter 22 In a longitudinal study discussed by Christensen (1997), 2121 people neither exercised regularly nor developed cardiovascular disease during the study. These subjects were cross-classified by three factors: Personality type (A,B), Cholesterol level (normal, high), and Diastolic Blood Pressure (normal, high). The data are given in Table 22.1.

Table 22.1 is a three-way table of counts. The three factors are Personality, Cholesterol level, and Diastolic Blood Pressure. Each factor happens to be at two levels, but that is of no particular consequence for our modelling of the data. We can analyze the data by fitting three-way ANOVA type models to it. However, count data is not normally distributed, so standard ANOVA methods are inappropriate. In particular, random variables for counts tend to have variances that depend on their mean values. Standard sampling schemes for count data are multinomial sampling and Poisson sampling. In this case, we can think of the data as being a sample of 2121 from a multinomial distribution, cf. Section 1.5.

In general we assume that our data are independent Poisson random variables, say,

$$y_h \sim \text{Pois}(\mu_h), h = 1, \dots, n$$

and create models

$$\log(\mu_h) = m(x_h)$$

where x_h is a predictor variable or vector associated with case h . Most often x_h contains factor variables. Our standard models $m(\cdot)$ are all linear in their parameters, so these are called *log-linear models*. Rather than Poisson sampling, most of our data are multinomial or a combination of independent multinomials. For models with an intercept (or ANOVA models), the analysis for Poisson data is the same as the analysis for multinomial data. A similar statement holds for independent multinomial data provided appropriate group effects are included in the model. For a complete discussion of fitting log-linear models with various sampling schemes and the relationship between this approach and ANOVA modeling for normal data, see Christensen (1997).

Log-linear models are more often used for exploring relationships between factors than for prediction. In prediction, one tends to focus on a particular factor (variable) of interest, called the response factor, and use the other factors (variables) to predict or explain the response. If a factor of interest has only two possible outcomes, the logistic models of Chapter 21 can be used for prediction. Log-linear models are more often used to model independence relationships between factors.

Table 22.1: *Personality, Cholesterol, Blood Pressure Data*

y_{ijk}		Diastolic Blood Pressure	
Personality	Cholesterol	Normal	High
A	Normal	716	79
	High	207	25
B	Normal	819	67
	High	186	22

Table 22.2: *Religion and occupations*

Religion	Occupation				Total
	A	B	C	D	
Protestant	210	277	254	394	1135
Roman Catholic	102	140	127	279	648
Jewish	36	60	30	17	143
Total	348	477	411	690	1926

In this chapter we examine log-linear models, we also relate them to logistic regression models and show how to use log-linear models to develop prediction models for factors that have more than two possible outcomes.

22.1 Models for Two-Factor Tables

Consider a 3×4 table such as that given in Table 22.2. These are data extracted from Lazearwitz (1961) and considered in Chapter 5. The data are from three religious groups and give the numbers of people who practice various occupations. The occupations are A, professions; B, owners, managers, and officials; C, clerical and sales; and D, skilled workers.

Consider fitting ANOVA type models to the logs of the expected cell counts. Let y_{ij} denote the observed count in row i and column j of the table and let μ_{ij} denote the expected count in the i, j cell. The two-way model with interaction can be written

$$\log(\mu_{ij}) = \mu + \alpha_i + \eta_j + \gamma_{ij}, \quad i = 1, \dots, 3; j = 1, \dots, 4.$$

An alternative notation is often used for log-linear models that merely changes the names of the parameters,

$$\log(\mu_{ij}) = u + u_{1(i)} + u_{2(j)} + u_{12(ij)}. \quad (22.1.1)$$

This log-linear model imposes no constraints on the table of cell means because there are more parameters than there are cells in the table. Actually, the u , $u_{1(i)}$ and $u_{2(j)}$ terms are all redundant because the $u_{12(ij)}$ terms provide as many parameters for explaining expected cell counts as there are expected cell counts to be explained. This model will fit the data for any two-factor table perfectly! Because it has as many parameters as there are cells, this model is referred to as the *saturated model*.

The log-linear model that includes only main effects is

$$\log(\mu_{ij}) = u + u_{1(i)} + u_{2(j)}. \quad (22.1.2)$$

In terms of Table 22.2, if this model fits the data, it says the data are explained adequately by a model in which religion and occupation are independent, cf. Christensen (1997). If religion and occupation are independent, knowing one's religion gives no new information about a person's occupation. That makes sense relative to the model involving only main effects, because then religion affects only the terms $u_{1(i)}$ and has no effect on the contribution from occupation, which is the additive term $u_{2(j)}$. On the other hand, if model (22.1.1) applies to the data, the interaction terms $u_{12(ij)}$ allow the possibility of different occupation effects for every religious group. (It turns out that the model of independence does not fit these data well.)

22.1.1 Lancaster-Irwin partitioning.

Religion and occupations data are given in a form suitable for computing in Table 22.3. The main effects model (22.1.2) can be rewritten as

$$\log(\mu_{kmj}) = u + u_{1(km)} + u_{2(j)}. \quad (22.1.3)$$

Table 22.3 *Religion and occupations: i is Religion, j is Occupation, k collapses Roman Catholic and Protestant, k and m uniquely define Religions.*

y	i	j	k	m
210	1	1	1	1
277	1	2	1	1
254	1	3	1	1
394	1	4	1	1
102	2	1	1	2
140	2	2	1	2
127	2	3	1	2
279	2	4	1	2
36	3	1	2	2
60	3	2	2	2
30	3	3	2	2
17	3	4	2	2

The reduced table of Table 5.13 can be fitted using the model

$$\log(\mu_{kmj}) = u + u_{1(km)} + u_{2(kj)}. \quad (22.1.4)$$

This model has a separate parameter for each Jewish occupation, so it effectively leaves them alone, and fits an independence model to the Roman Catholics and Protestants, cf. Christensen (1997, Exercise 8.4.3). The pair of subscripts km uniquely define the three religious groups, so in model (22.1.4) the term $u_{1(km)}$ is really a main effect for religions. The terms $u_{2(kj)}$ define main effects for occupations when $k = 1$, i.e., for Roman Catholics and Protestants, but the terms $u_{2(kj)}$ define separate effects for each Jewish occupation when $k = 2$. The key ideas are that k has a unique value for each religious category except it does not distinguish between the categories that are to be collapsed and that k and m uniquely define the religions. Thus m needs to have different values for each religion that is to be collapsed and some common value for the religions that are not to be collapsed.

For fitting models (22.1.2) and (22.1.3) the deviance is $G^2 = 64.342$ on $df = 6$. For fitting model (22.1.4), $G^2 = 12.206$ on $df = 3$ which are the deviance and degrees of freedom for the reduced model. The deviance and degrees of freedom for the collapsed model are obtained by subtraction $G^2 = 64.342 - 12.206 = 52.136$ on $df = 6 - 3 = 3$. These are roughly similar to the corresponding Pearson χ^2 statistics discussed in Chapter 5.

Computing commands

The following R code fits models (22.1.2), (22.1.3), and (22.1.4) and gives fitted values and Pearson residuals. By comparing the fitted values and Pearson residuals for models (22.1.2) and (22.1.3) with each other and results from Section 5.5 you can see that both models fit the independence model. Similarly, the fitted values and Pearson residuals confirm that fitting model (22.1.4) fits the Roman Catholics and Protestants like Table 5.13 while leaving the Jewish entries untouched.

```
occ <- read.table("E:\\Books\\ANREG2\\newdata\\tab22-2.dat",
                 sep=" ", col.names=c("y", "R", "O", "RCP", "Q"))
attach(occ)
occ
#summary(occ)

#Summary tables
r=factor(R)
o=factor(O)
```

```

rcp=factor(RCP)
q=factor(Q)
ts <- glm(y ~ r + o,family = poisson)
tsp=summary(ts)
#tsp
anova(ts)
rpearson=(y-ts$fit)/(ts$fit)^(.5)
rstand=rpearson/(1-hatvalues(ts))^(.5)
invf = c(y,ts$fit,hatvalues(ts),rpearson,
          rstand,cooks.distance(ts))
inf=matrix(invf,I(tsp$df[1]+tsp$df[2]),6,dimnames = list(NULL,
          c("y", "yhat", "lev","Pearson","Stand.","C")))
inf

ts2 <- glm(y ~ o + rcp:q,family = poisson)
tsp2=summary(ts2)
anova(ts2)
rpearson=(y-ts2$fit)/(ts2$fit)^(.5)
rstand=rpearson/(1-hatvalues(ts2))^(.5)
invf = c(y,ts2$fit,hatvalues(ts2),rpearson,
          rstand,cooks.distance(ts2))
inf=matrix(invf,I(tsp2$df[1]+tsp2$df[2]),6,dimnames = list(NULL,
          c("y", "yhat", "lev","Pearson","Stand.","C")))
inf

ts3 <- glm(y ~ o:rcp + rcp:q,family = poisson)
tsp3=summary(ts3)
anova(ts3)
rpearson=(y-ts3$fit)/(ts3$fit)^(.5)
rstand=rpearson/(1-hatvalues(ts3))^(.5)
invf = c(y,ts3$fit,hatvalues(ts3),rpearson,
          rstand,cooks.distance(ts3))
inf=matrix(invf,I(tsp3$df[1]+tsp3$df[2]),6,dimnames = list(NULL,
          c("y", "yhat", "lev","Pearson","Stand.","C")))
inf

```

22.2 Models for Three-Factor Tables

Consider a three-factor table of counts such as Table 22.1. The *saturated* model is the model with three-factor interaction terms,

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

This model has a separate $u_{123(ijk)}$ parameter for each cell in the table, so this model will fit any three-factor table perfectly. Note that the model is grossly overparameterized, an equivalent model is simply

$$\log(\mu_{ijk}) = u_{123(ijk)}.$$

We abbreviate this model as [123]. The only reason for using the overparameterized model is that it is suggestive of some interesting submodels. In the following discussion, statements are made about log-linear models implying independence relationships. See Christensen (1997, Chapter 3) for the validity of these claims.

The model with main effects only is

$$\log(\mu_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)}. \quad (22.2.1)$$

This model implies complete independence of the three factors in the table. Note that a less overparameterized version of model (22.2.1) is

$$\log(\mu_{ijk}) = u_{1(i)} + u_{2(j)} + u_{3(k)}.$$

We can abbreviate the model as [1][2][3].

We can also look at models that include main effects and only one two-factor interaction, for example,

$$\log(\mu_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}. \quad (22.2.2)$$

This model implies that factor 1 is independent of factors 2 and 3. We can also get two other models with similar interpretations by including only the $u_{12(ij)}$ interaction (factor 3 independent of factors 1 and 2) or only the $u_{13(ik)}$ interaction (factor 2 independent of factors 1 and 3). Note that a less overparameterized version of model (22.2.2) is

$$\log(\mu_{ijk}) = u_{1(i)} + u_{23(jk)}.$$

We can abbreviate the model as [1][23].

Now consider models that include two two-factor interactions, for example,

$$\log(\mu_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{13(ik)} + u_{23(jk)}. \quad (22.2.3)$$

This model implies that factor 1 is independent of factor 2 given factor 3. Note that factor 3 is in both interaction terms, which is why the model has 1 and 2 independent given 3. We can also get two other models with similar interpretations by including only the $u_{12(ij)}$ and $u_{23(jk)}$ interactions (factors 1 and 3 independent given factor 2) or only the $u_{12(ij)}$ and $u_{13(ik)}$ interactions (factors 2 and 3 independent given factor 1). Note that a less overparameterized version of model (22.2.3) is

$$\log(\mu_{ijk}) = u_{13(ik)} + u_{23(jk)}.$$

We abbreviate the model as [13][23].

The no three-factor interaction model is

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

A less overparameterized version is

$$\log(\mu_{ijk}) = u_{12(ij)} + u_{13(ik)} + u_{23(jk)},$$

which we can abbreviate as [12][13][23]. This model has no nice interpretation in terms of independence.

22.2.1 Testing Models

Testing models works much the same as it does for linear models except with count data there is no need to estimate a variance. Thus, the tests are similar to looking at only the numerator sums of squares for an analysis of variance or regression test. (This would be totally appropriate in ANOVA and regression if we ever actually knew the value of σ^2 .) The *deviance* of a model is

$$G^2 = 2 \sum_{\text{all cells}} y_{ijk} \log(y_{ijk} / \hat{\mu}_{ijk})$$

where $\hat{\mu}_{ijk}$ is the (maximum likelihood) estimated expected cell count based on the model we are testing. The subscripts in G^2 are written for a three-dimensional table, but the subscripts are irrelevant. What is relevant is summing over all cells.

G^2 gives a test of whether the model gives an adequate explanation of the data relative to the saturated model. (Recall that the saturated model always fits the data perfectly because it has a separate parameter for every cell in the table. Thus the estimated cell counts for a saturated model are always just the observed cell counts.) The degrees of freedom (df) for G^2 are the degrees of freedom for the interaction terms that have been left out. So in an $I \times J \times K$ table, the model that drops out the three-factor interaction has $(I-1)(J-1)(K-1)$ degrees of freedom for G^2 . If we drop out the [12] interaction as well as the three-factor interaction, G^2 has $(I-1)(J-1)(K-1) + (I-1)(J-1)$ degrees of freedom. If the observations y in each cell are reasonably large, the G^2 statistics can be compared to a χ^2 distribution with the appropriate number of degrees of freedom. A large G^2 indicates that the model fits poorly as compared to the saturated model.

Alternatively, one can use the Pearson statistic,

$$X^2 = \sum_{\text{all cells}} \frac{(y_{ijk} - \hat{\mu}_{ijk})^2}{\hat{\mu}_{ijk}}$$

for testing the model. It has the same degrees of freedom as G^2 .

To test a full model (F) against a reduced model (R), we compare deviance G^2 s. In particular, the test statistic is

$$\begin{aligned} G^2(R \text{ versus } F) &\equiv G^2(R) - G^2(F) \\ &= 2 \sum_{\text{all cells}} y_{ijk} \log(y_{ijk}/\hat{\mu}_{Rijk}) - 2 \sum_{\text{all cells}} y_{ijk} \log(y_{ijk}/\hat{\mu}_{Fijk}) \\ &= 2 \sum_{\text{all cells}} y_{ijk} \log(\hat{\mu}_{Fijk}/\hat{\mu}_{Rijk}) \\ &= 2 \sum_{\text{all cells}} \hat{\mu}_{Fijk} \log(\hat{\mu}_{Fijk}/\hat{\mu}_{Rijk}) \end{aligned}$$

where $\hat{\mu}_{Rijk}$ and $\hat{\mu}_{Fijk}$ are the maximum likelihood estimates from the reduced and full models, respectively, and showing the last equality is beyond the scope of this book. The degrees of freedom for the test is the difference in deviance degrees of freedom, $df(R) - df(F)$.

In Chapter 21 we introduced AIC as a tool for model selection. For log-linear models, maximizing Akaike's information criterion (AIC) amounts to choosing the model X that *minimizes*

$$A_X = G^2(X) - [q - 2r],$$

where $G^2(X)$ is the deviance for testing the X model against the saturated model, r is the number of degrees of freedom for the X model and there are $q \equiv n$ degrees of freedom for the saturated model, i.e., $q \equiv n$ cells in the table.

Given a list of models to be compared along with their G^2 statistics and the degrees of freedom for the tests, a slight modification of A_X is easier to compute by hand.

$$\begin{aligned} A_X - q &= G^2(X) - 2[q - r] \\ &= G^2(X) - 2(\text{test degrees of freedom}). \end{aligned}$$

Because q does not depend on the model X , minimizing $A_X - q$ is equivalent to minimizing A_X . Note that for the saturated model, $A - q \equiv 0$.

EXAMPLE 22.2.1. For the personality (1), cholesterol (2), blood pressure (3) data of Table 22.1, testing models against the saturated model gives deviance and AIC values

Model	df	G^2	$A - q$
[12][13][23]	1	0.613	-1.387
[12][13]	2	2.062	-1.938
[12][23]	2	2.980	-1.020
[13][23]	2	4.563	0.563
[1][23]	3	7.101	1.101
[2][13]	3	6.184	0.184
[3][12]	3	4.602	-1.398
[1][2][3]	4	8.723	0.723

Comparing the G^2 values to 95th percentiles of χ^2 distributions with the appropriate number of degrees of freedom, all of the models seem to explain the data adequately.

To test, for example, the reduced model [1][2][3] against a full model [12][13], the test statistic is

$$G^2([1][2][3] \text{ versus } [12][13]) = G^2([1][2][3]) - G^2([12][13]) \\ = 8.723 - 2.062 = 6.661,$$

on $df([1][2][3]) - df([12][13]) = 4 - 2 = 2$. The test statistic is greater than $\chi^2(.95, 2) = 5.991$, so we can reject the model of complete independence [1][2][3]. Complete independence fits the data significantly worse than the model [12][13] in which cholesterol and blood pressure are independent given personality type. So even though [1][2][3] fits adequately relative to the saturated model [123], it fits the data inadequately relative to [12][13].

For the personality data, AIC suggests three attractive models: [3][12], [12][13], and [12][13][23]. Model [12][23] is also not bad. $\square$

With only three factors it is easy to look at all possible models. Model selection procedures become more important when dealing with tables having more than three factors, cf., Christensen (1997) and Section 4.

22.3 Estimation and Odds Ratios

For log-linear models, the primary goal of estimation is to obtain the estimated expected cell counts $\hat{\mu}_h$. Iterative computing methods are needed to find these. The Newton-Raphson method (iteratively reweighted least squares) provides an estimate of the model parameters. The other estimation method is iterative proportional fitting (the Deming-Stephans algorithm) for ANOVA type models. This directly provides estimates $\hat{\mu}_h$ from which parameter estimates can be back calculated. See Christensen (1997) for details. Programs for fitting generalized linear models generally use Newton-Raphson. The BMDP procedure 4F is restricted to ANOVA type models but still uses iterative proportional fitting, and as a result it provides some nice features not available in many other programs.

EXAMPLE 22.3.1. Kihlberg, Narragon, and Campbell (1964), Fienberg (1980), and Christensen (1997) present data on automobile injuries. The model of no three-factor interaction

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

fits the data very well. Below are given the data and the estimated expected cell counts based on the model.

$y_{ijk}(\hat{\mu}_{ijk})$		Accident Type (k)			
		Collision		Rollover	
Injury (j)		Not Severe	Severe	Not Severe	Severe
Driver	No	350 (350.49)	150 (149.51)	60 (59.51)	112 (112.49)
Ejected (i)	Yes	26 (25.51)	23 (23.49)	19 (19.49)	80 (79.51)

One way to examine a fitted model is to look at the estimated odds ratios. We first use the unrestricted estimates of the p_{ijk} 's, i.e., $\hat{p}_{ijk} = y_{ijk}/y_{...}$. The estimated odds of a Not Severe injury when the driver is not ejected in a collision are

$$\frac{y_{111}/y_{...}}{y_{121}/y_{...}} = \frac{y_{111}}{y_{121}} = \frac{y_{111}}{y_{121}} = \frac{350}{150} = 2.333.$$

The estimated odds of a Not Severe injury when the driver is ejected in a collision are

$$\frac{26}{23} = 1.130$$

In a collision, the odds of a not severe injury are

$$2.064 = \frac{2.333}{1.130} = \frac{350(23)}{26(150)}$$

times greater if one is not ejected from the car. Similarly, in a rollover, the odds of a not severe injury are

$$2.256 = \frac{60(80)}{19(112)}$$

times greater if one is not ejected from the car. These odds ratios are quite close to one another. In the no three way interaction model, these odds ratios are forced to be the same. If we make the same computations using the estimated expected cell counts, we get

$$2.158 = \frac{350.49(23.49)}{25.51(149.51)} = \frac{59.51(79.51)}{14.49(112.49)},$$

For both collisions and rollovers, the odds of a severe injury are about twice as large if the driver is ejected from the vehicle than if not. Equivalently, the odds of having a nonsevere injury are about twice as great if the driver is not ejected from the vehicle than if the driver is ejected. It should be noted that the odds of being severely injured in a rollover are consistently much higher than in a collision. What we have concluded in our analysis is that the *relative* effect of the driver being ejected is the same for both types of accident and that being ejected substantially increases one's chances of being severely injured. $\square$

All of the models discussed in Section 2 have interpretations in terms of odds ratios. An odds ratio keeps one of the three indexes fixed, say, k , and looks at quantities like

$$\frac{\hat{p}_{ijk}\hat{p}_{i'j'k}}{\hat{p}_{i'jk}\hat{p}_{ij'k}} = \frac{\hat{\mu}_{ijk}\hat{\mu}_{i'j'k}}{\hat{\mu}_{i'jk}\hat{\mu}_{ij'k}}.$$

In the model of complete independence (22.2.1) these values will always be 1, no matter how they are constructed. In model (22.2.2) where factor 1 is independent of facts 2 and 3, any estimated odds ratio that fixes either the level k or the level j will equal one and all odds ratios that fix i will be the same regardless of the value of i . In model (22.2.3) where factors 1 and 2 are independent given 3, any estimated odds ratio that fixes the level k will equal one and all odds ratios that fix i will be the same regardless of the value of i , similarly all odds ratios that fix j will be the same regardless of the value of j .

EXAMPLE 22.3.2. Consider data from Everitt (1977) and Christensen (1997) on classroom behavior. The three factors are Classroom Behavior (Deviant or Nondeviant), Risk of the home situation: not at risk (N) or at risk (R), and Adversity of the school situation (Low, Medium, or High). The data and estimated expected cell counts for the model $\log(\mu_{ijk}) = u + u_{1(i)} + u_{2(j)} + u_{3(k)} + u_{23(jk)}$, in which behavior is independent of risk and adversity, are given below.

y_{ijk} ($\hat{\mu}_{ijk}$)		Adversity (k)					
		Low		Medium		High	
Risk (j)		N	R	N	R	N	R
Classroom Behavior (i)	Non.	16 (14.02)	7 (6.60)	15 (14.85)	34 (34.64)	5 (4.95)	3 (4.95)
	Dev.	1 (2.98)	1 (1.40)	3 (3.15)	8 (7.36)	1 (1.05)	3 (1.05)

Subject to roundoff error, the estimate of the log odds of nondeviant behavior is

$$\frac{\hat{\mu}_{1jk}}{\hat{\mu}_{2jk}} = \frac{14.02}{2.98} = \frac{6.60}{1.40} = \dots = \frac{4.95}{1.05} = 4.702 = e^{1.548}.$$

Thus, any odds ratio in which either j or k is held fixed always equals 1. The estimate of the log odds of nondeviant behavior is

$$\begin{aligned} \log(\hat{\mu}_{1jk}/\hat{\mu}_{2jk}) &= \log(\hat{\mu}_{1jk}) - \log(\hat{\mu}_{2jk}) \\ &= \hat{u}_{1(1)} - \hat{u}_{1(2)} = 1.548. \end{aligned}$$

The odds of having a home situation that is not at risk depend on the adversity level. Up to roundoff error, the odds satisfy

$$\frac{\hat{\mu}_{i1k}}{\hat{\mu}_{i2k}} = \begin{cases} 14.02/6.60 = 2.98/1.40 = 2.125 & k = 1 \\ 14.85/34.64 = 3.15/7.36 = 0.428 & k = 2 \\ 4.95/4.95 = 1.05/1.05 = 1 & k = 3 \end{cases}.$$

From this you can see that odds ratios that fix i will all equal each other, but will not necessarily equal 1; whereas odds ratios that fix k will again always equal 1.

22.4 Higher Dimensional Tables

Tables with four or more factors can also be modeled. The basic ideas from three dimensional tables continue to apply, but the models become more complicated.

EXAMPLE 22.4.1. Consider again the data of Example 21.7.1 and Table 21.9 on muscle tension change. Here we examine models for expected cell counts but do not identify tension as a response.

The log-linear model of all main effects is

$$\log(\mu_{hijk}) = \gamma + \tau_h + \omega_i + \mu_j + \delta_k. \quad (22.4.1)$$

The model of all two factor interactions is

$$\begin{aligned} \log(\mu_{hijk}) &= \gamma + \tau_h + \omega_i + \mu_j + \delta_k + (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} \\ &\quad + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk}. \end{aligned} \quad (22.4.2)$$

The model of all three factor interactions is

$$\begin{aligned} \log(\mu_{hijk}) &= \gamma + \tau_h + \omega_i + \mu_j + \delta_k + (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} \\ &\quad + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk} \\ &\quad + (\tau\omega\mu)_{hij} + (\tau\omega\delta)_{hik} + (\tau\mu\delta)_{hjk} + (\omega\mu\delta)_{ijk}. \end{aligned} \quad (22.4.3)$$

Removing some redundant parameters gives

$$\log(\mu_{hijk}) = \tau_h + \omega_i + \mu_j + \delta_k, \quad (22.4.1)$$

$$\log(\mu_{hijk}) = (\tau\omega)_{hi} + (\tau\mu)_{hj} + (\tau\delta)_{hk} + (\omega\mu)_{ij} + (\omega\delta)_{ik} + (\mu\delta)_{jk}, \quad (22.4.2)$$

and

$$\log(\mu_{hijk}) = (\tau\omega\mu)_{hij} + (\tau\omega\delta)_{hik} + (\tau\mu\delta)_{hjk} + (\omega\mu\delta)_{ijk} \quad (22.4.3)$$

Corresponding shorthand notations are

$$[T][W][M][D] \quad (1)$$

$$[TW][TM][WM][TD][WD][MD] \quad (2)$$

$$[TWM][TWD][TMD][WMD] \quad (3)$$

Statistics for testing models against the saturated model are given below. The only model considered that fits these data has all three-factor interactions.

Model	df	G^2	P	$A - q$
$[TWM][TWD][TMD][WMD]$	1	0.11	.74	-1.89
$[TW][TM][WM][TD][WD][MD]$	5	47.67	.00	37.67
$[T][W][M][D]$	11	127.4	.00	105.4

To test reduced models, say, $[TWM][TWD][TMD][WMD]$ against the reduced model of all two-factor terms $[TW][TM][WM][TD][WD][MD]$, compute $G^2 = 47.67 - 0.11 = 47.56$ with $df = 5 - 1 = 4$. Clearly, the reduced model does not fit. A reasonable beginning for modelling these data would be to find which, if any, three factor interactions can be removed without harming the model significantly.

EXAMPLE 22.4.2. In later sections we will consider an expanded version of the abortion opinion data of Example 21.8.1 and Table 21.14 that includes another opinion category that was ignored in Chapter 21. The four factors now define a $2 \times 2 \times 3 \times 6$ table. The factors and levels are

Factor	Abbreviation	Levels
Race	R	White, Nonwhite
Sex	S	Male, Female
Opinion	O	Yes = Supports Legalized Abortion No = Opposed to Legalized Abortion Und = Undecided
Age	A	18-25, 26-35, 36-45, 46-55, 56-65, 66+ years

The data are given in Table 22.3.

22.4.1 Computer Commands

The muscle tension data are listed in file 'Tab21-9a.dat' as counts for each cell with indices for tension, weight, muscle type, and drug, respectively. The file is as given below.

```

3 1 1 1 1
21 1 1 1 2
23 1 1 2 1
11 1 1 2 2
22 1 2 1 1
32 1 2 1 2
4 1 2 2 1
12 1 2 2 2
3 2 1 1 1
10 2 1 1 2
41 2 1 2 1
```

Table 22.4: *Abortion Opinion Data*

RACE (<i>h</i>)	SEX (<i>i</i>)	OPINION (<i>j</i>)	AGE (<i>k</i>)					
			18-25	26-35	36-45	46-55	56-65	66+
White	Male	Yes	96	138	117	75	72	83
		No	44	64	56	48	49	60
		Und	1	2	6	5	6	8
	Female	Yes	140	171	152	101	102	111
		No	43	65	58	51	58	67
		Und	1	4	9	9	10	16
Nonwhite	Male	Yes	24	18	16	12	6	4
		No	5	7	7	6	8	10
		Und	2	1	3	4	3	4
	Female	Yes	21	25	20	17	14	13
		No	4	6	5	5	5	5
		Und	1	2	1	1	1	1

```

21 2 1 2 2
45 2 2 1 1
23 2 2 1 2
6 2 2 2 1
22 2 2 2 2

```

This file is different from that used in Chapter 21, Tab21-9.dat. We can fit the log-linear model [WMD][TWM][TWD][TMD] using SAS PROC GENMOD.

```

options ps=60 ls=72 nodate;
data tension;
    infile 'tension.dat';
    input y T W M D;
proc genmod data=tension;
    class T W M D;
    model y = W*M*T T*W*M T*W*D T*M*D / link=log
                                dist=poisson;
run;

```

The “class” command specifies that a variable is not acting like a predictor variable in regression, but rather that it gives indices for specifying the levels of an analysis of variance type factor. To fit other specific models such as [TM][WM][MD] or [T][WM][D], the model statement uses T*M W*M M*D or T W*M D, respectively.

For the abortion data, the data file ‘abort.dat’ has five columns, the first four being indices for race, sex, age, and opinion, and the last being the cell counts. To fit [RSO][RSA][ROA][SOA] with GENMOD use

```

options ps=60 ls=72 nodate;
data abort;
    infile 'abort.dat';
    input R S A O y;
proc genmod data=abort;
    class R S A O;
    model y = R*S*O R*S*A R*O*A S*O*A / link=log
                                dist=poisson;
run;

```

The BMDP-4F commands for fitting the model [RSO][RSA][ROA][SOA] are

```
/ INPUT      FILE = 'C:\LOGLIN\ABORT.DAT' .
              FORMAT = FREE.
              VARIABLES = 5.
/ VARIABLE   NAMES = R, S, A, O, N.
/ TABLE     INDICES = R, S, A, O.
              COUNT = N.
/ STAT       ALL.
/ FIT        MODEL = RSO, RSA, ROA, SOA.
/ PRINT      LINE = 79.
/ END
```

22.5 Ordered Categories

In ANOVA models with one or more factors having quantitative levels, it is useful to model effects with polynomials. Similar results apply to log-linear models.

EXAMPLE 22.5.1. Men from Framingham, Massachusetts were categorized by their serum cholesterol and systolic blood pressure. Consider the subsample that did not develop coronary heart disease during the follow-up period. The data are as follows.

Cholesterol (in mg/100 cc)	Blood Pressure (in mm Hg)				Totals
	<127	127-146	147-166	167+	
<200	117	121	47	22	307
200-219	85	98	43	20	246
220-259	119	209	68	43	439
≥260	67	99	46	33	245
Totals	388	527	204	118	1237

Both factors have ordered levels, although there is no one number associated with each level. We consider quantitative levels 1, 2, 3, 4 for both factors. Obviously, this is somewhat arbitrary. An alternative approach that involves nonlinear modeling is to estimate the quantitative levels for each factor, cf. Christensen (1997) for more information.

Now consider four models that involve the quantitative levels:

Abbreviation	Model
[C][P][C ₁]	$\log(\mu_{ij}) = u + u_{C(i)} + u_{P(j)} + C_{1i} \cdot j$
[C][P][P ₁]	$\log(\mu_{ij}) = u + u_{C(i)} + u_{P(j)} + P_{1j} \cdot i$
[C][P][γ]	$\log(\mu_{ij}) = u + u_{C(i)} + u_{P(j)} + \gamma \cdot i \cdot j$
[C][P]	$\log(\mu_{ij}) = u + u_{C(i)} + u_{P(j)}$

These are called the *row effects*, *column effects*, *uniform association*, and *independence* models, respectively. The fits for the models relative to the saturated model are

Model	df	G ²	A - q
[C][P][C ₁]	6	7.404	-4.596
[C][P][P ₁]	6	5.534	-6.466
[C][P][γ]	8	7.429	-8.571
[C][P]	9	20.38	2.38

The best fitting model is the uniform association model

$$\log(\mu_{ij}) = u + u_{C(i)} + u_{P(j)} + \gamma \cdot i \cdot j.$$

Using the side conditions $u_{C(1)} = u_{P(1)} = 0$, the parameter estimates and large sample standard errors are

Parameter	Estimate	Standard Error
u	4.614	.0699
$u_{C(1)}$	0	—
$u_{C(2)}$	−0.4253	.1015
$u_{C(3)}$	−0.0589	.1363
$u_{C(4)}$	−0.8645	.1985
$u_{P(1)}$	0	—
$u_{P(2)}$	0.0516	.0965
$u_{P(3)}$	−1.164	.1698
$u_{P(4)}$	−1.991	.2522
γ	0.1044	.0293

The estimated cell counts are

Estimated Cell Counts $\hat{\mu}_{ij}$: Uniform Association				
Cholesterol	Blood Pressure			
	<127	127-146	147-166	167+
<200	112.0	131.0	43.1	20.9
200-219	81.3	105.4	38.5	20.8
220-259	130.1	187.4	76.0	45.5
≥ 260	64.5	103.2	46.4	30.8

Because these are obtained from the uniform association model, the odds ratios for consecutive table entries are identical. For example, the odds of blood pressure < 127 relative to blood pressure 127-146 for men with cholesterol < 200 are 1.11 times the similar odds for men with cholesterol of 200-219; up to round off error

$$\frac{112.0/131.0}{81.3/105.4} = \frac{112.0(105.4)}{81.3(131.0)} = e^{0.1044} = 1.11$$

where $0.1044 = \hat{\gamma}$. Similarly, the odds of blood pressure 127-146 relative to blood pressure 147-166 for men with cholesterol < 200 are 1.11 times the odds for men with cholesterol of 200-219:

$$\frac{131.0(38.5)}{105.4(43.1)} = e^{0.1044} = 1.11.$$

Also the odds of blood pressure < 127 relative to blood pressure 127-146 for men with cholesterol 200-219 are 1.11 times the odds for men with cholesterol of 220-259:

$$\frac{81.3(187.4)}{130.1(105.4)} = e^{0.1044} = 1.11.$$

For consecutive categories, the odds of lower blood pressure are 1.11 times greater with lower blood cholesterol than with higher blood cholesterol.

Of course we can also compare nonconsecutive categories. For categories that are one step away from consecutive, the odds of lower blood pressure are $1.23 = e^{2(0.1044)}$ times greater with lower cholesterol than with higher cholesterol. For example, the odds of having blood pressure < 127 compared to having blood pressure of 147 – 166 with cholesterol < 200 are $1.23 = e^{2(0.1044)}$ times those for cholesterol 200 – 219. To check this observe that

$$\frac{112.0(38.5)}{81.3(43.1)} = 1.23.$$

Similarly, the odds of having blood pressure < 127 compared to having blood pressure of 127-146 with cholesterol < 200 are 1.23 times those for cholesterol 220-259. Extending this leads to observing that the odds of having blood pressure < 127 compared to having blood pressure of 167+ with cholesterol < 200 are $2.559 = e^{0.9(0.1044)}$ times those for cholesterol ≥ 260 .

It is of interest to compare the estimated cell counts obtained under uniform association with the estimated cell counts under independence. The estimated cell counts under independence are

Estimated Cell Counts $\hat{\mu}_{ij}$: Independence				
Cholesterol	Blood Pressure			
	<127	127-146	147-166	167+
<200	96.3	130.8	50.6	29.3
200-219	77.2	104.8	40.6	23.7
220-259	137.7	187.0	72.4	41.9
≥ 260	76.85	104.4	40.4	23.4

All of the estimated odds ratios from the independence model are 1. Relative to independence, with $\gamma > 0$, the uniform association model increases the estimated cell counts for cells with 1) high cholesterol and high blood pressure and 2) low cholesterol and low blood pressure. Also, the uniform association model decreases the estimated cell counts for cells with 1) high cholesterol and low blood pressure and 2) low cholesterol and high blood pressure.

EXAMPLE 22.5.1. For the Abortion Opinion data of Table 22.3, the model [RSO][OA] fits well. The ages are quantitative levels. We consider whether using the quantitative nature of this factor leads to a more succinct model. The age categories are 18-25, 26-35, 36-45, 46-55, 56-65, and 66+. For lack of a better idea, the category scores were taken as 1, 2, 3, 4, 5, and 6. Since the first and last age categories are different from the other four, the use of the scores 1 and 6 are particularly open to question. Two new models were considered:

Abbreviation	Model
[RSO][OA]	$\log(\mu_{hijk}) = u_{RSO(hij)} + u_{OA(jk)}$
[RSO][A][O ₁]	$\log(\mu_{hijk}) = u_{RSO(hij)} + u_{A(k)} + O_1jk$
[RSO][A][O ₁][O ₂]	$\log(\mu_{hijk}) = u_{RSO(hij)} + u_{A(k)} + O_1jk + O_2jk^2$

Both of these are reduced models relative to [RSO][OA]. ([RSO][OA] is equivalent to $\log(\mu_{hijk}) = u_{RSO(hij)} + u_{A(k)} + O_1jk + O_2jk^2 + O_3jk^3 + O_4jk^4 + O_5jk^5$ and also to $\log(\mu_{hijk}) = u_{RSO(hij)} + O_1jk + O_2jk^2 + O_3jk^3 + O_4jk^4 + O_5jk^5$. **Is this true?**) To compare models, we need the following statistics

Model	df	G ²
[RSO][OA]	45	24.77
[RSO][A][O ₁][O ₂]	51	26.99
[RSO][A][O ₁]	53	29.33

Comparing [RSO][A][O₁] versus [RSO][OA] gives $G^2 = 29.33 - 24.77 = 4.56$ with degrees of freedom $53 - 45 = 8$. The G^2 value is not significant. Similarly, [RSO][A][O₁][O₂] is an adequate model relative to [RSO][OA]. The test for [O₂] has $G^2 = 29.33 - 27.99 = 1.34$ on 2 df, which is not significant. The model with only [O₁] fits the data well.

22.5.1 Computing Commands

Models with quantitative factors can be fit easily using several computer packages, e.g., R and SAS PROC GENMOD. For example the model $\log(\mu_{hijk}) = u_{RSO(hij)} + u_{A(k)} + O_1jk + O_2jk^2$ can be fitted using SAS PROC GENMOD as given below.

Table 22.5: *Textile faults*

Roll	Length (<i>l</i>)	Faults (<i>y</i>)	Roll	Length (<i>l</i>)	Faults (<i>y</i>)
1	551	6	17	543	8
2	651	4	18	842	9
3	832	17	19	905	23
4	375	9	20	542	9
5	715	14	21	522	6
6	868	8	22	122	1
7	271	5	23	657	9
8	630	7	24	170	4
9	491	7	25	738	9
10	372	7	26	371	14
11	645	6	27	735	17
12	441	8	28	749	10
13	895	28	29	495	7
14	458	4	30	716	3
15	642	10	31	952	9
16	492	4	32	417	2

```

options ps=60 ls=72 nodate;
data abort;
  infile 'abort.dat';
  input R S A O N;
  A1 = A;
  A2 = A * A;
proc genmod data=abort;
  class R S O A;
  model N = R*S*O A O*A1 O*A2 / link=log
          dist=poisson;
run;

```

The terms O_{1jk} and O_{2jk^2} are really interactions between the factor O and the predictor variables age and age squared. In the model statement, they are simply specified as interactions.

22.6 Offsets

Most of our examples have involved multinomial data with ANOVA type models. Now we consider a regression with Poisson data. This example also involves a term in the linear predictor that is known.

EXAMPLE 22.6.1. Consider the data in Table 22.4. This is data from Bissell (1972) on the number of faults in pieces of fabric. It is reasonable to model the number of faults in any piece of fabric as Poisson and to use the length of the fabric as a predictor variable for the number of faults.

In general, we assume the existence of n independent random variables y_h with $y_h \sim \text{Pois}(\mu_h)$. In this example, a reasonable model might be that the expected number of faults μ_h is some number λ times the length of the piece of fabric, say l_h , i.e.,

$$\mu_h = \lambda l_h. \quad (22.6.1)$$

Such a model assumes that the faults are being generated at a constant rate, and therefore the expected number of faults is proportional to the length. We can rewrite model (22.6.1) as a log-linear model

$$\log(\mu_h) = \log(\lambda) + \log(l_h)$$

or

$$\log(\mu_h) = \beta_0 + (1) \log(l_h), \quad (22.6.2)$$

Table 22.6: *Model Fits*

Model	<i>df</i>	Deviance
(22.6.2)	31	64.538
(22.6.3)	30	64.537

Table 22.7: *Parameter Estimates for models (22.6.1) and (22.6.2)*

Model (2)		
Parameter	Estimate	SE
β_0	-4.193	0.05934
Model (3)		
Parameter	Estimate	SE
β_0	-4.173	1.135
β_1	0.9969	0.1759

where $\beta_0 \equiv \log(\lambda)$. If we generalize model (22.6.2) it might look a bit more familiar. Using a simple linear regression structure with $\log(l_h)$ as a predictor variable, we have the more general model

$$\log(\mu_h) = \beta_0 + \beta_1 \log(l_h). \quad (22.6.3)$$

On the original scale, this model has

$$\mu_h = \lambda l_h^{\beta_1}.$$

Model (22.6.2) is the special case of model (22.6.3) with $\beta_1 = 1$, and we would like to test the models. Model (22.6.2) includes a term $\log(l_h)$ that is not multiplied by an unknown parameter. We encountered such terms in normal theory linear models, and dealt with them by constructing a new dependent variable. Dealing with such terms in nonnormal models is more complicated but many computer programs handle them easily. A term like $\log(l_h)$ that is used in the linear predictor of a model like (2) that *is not* multiplied by an unknown parameter is often called an *offset*.

Fitting models (22.6.2) and (22.6.3) gives the results in Table 22.5. The test statistic for the adequacy of model (22.6.2) is

$$G^2 = 64.538 - 64.537 = .001$$

on $31 - 30 = 1$ *df*. Clearly, model (22.6.2) fits these data well. The parameter estimates for both models are given in Table 22.6.

22.6.1 Computer Commands

The data of Table 22.4 are in a file “bis.dat”. SAS GENMOD commands for fitting models (22.6.3) and (22.6.2) are, respectively,

```
options ps=60 ls=72 nodate;
data bis;
  infile 'bis.dat';
  input index 1 f;
  ll = log(1);
proc genmod data=bis;
  model f = ll / link=log
          dist=poisson;
run;
proc genmod data=bis;
  model f = / link=log
```

```

dist=poisson
offset=11;

run;

```

22.7 Relation to Logistic Models

All of the logistic models used in Chapter 7 are equivalent to log-linear models. This is discussed in detail in Christensen (1997). Here, we merely show that log-linear models imply standard logistic models. A logistic regression involves I independent binomial random variables. We can think of this as determining an $I \times 2$ table, where the 2 columns indicate successes and failures.

Note that in this chapter, n denotes the entire number of cells being considered, whereas in Chapter 21 n denoted the number of independent binomials being considered. Thus, in Chapter 21, $n \equiv I$ but in this chapter $n = 2I$, and later when we consider more than two possible responses, n will be a larger multiple of I .

Consider a log-linear model for a two-dimensional table that involves the use of a predictor variable to model interaction. The row effects model of Section 5 for the i th individual (or group) having the j th response is

$$\log(\mu_{ij}) = u_{1(i)} + u_{2(j)} + \eta_j x_i$$

where the usual interaction term $u_{12(ij)}$ is replaced in the model by a more specific interaction term, $\eta_j x_i$. Of course, x_i is the known predictor variable and η_j is an unknown parameter. This is an interaction term because it involves both the i and j subscripts, just like $u_{12(ij)}$. The relationship between the logistic model and the log-linear model is that, with $j = 1$ denoting “success” and p_i denoting the probability of success,

$$\begin{aligned}
 \log\left(\frac{p_i}{1-p_i}\right) &= \log\left(\frac{\mu_{i1}}{\mu_{i2}}\right) \\
 &= \log(\mu_{i1}) - \log(\mu_{i2}) \\
 &= [u_{1(i)} + u_{2(1)} + \eta_1 x_i] - [u_{1(i)} + u_{2(2)} + \eta_2 x_i] \\
 &= [u_{2(1)} - u_{2(2)}] + [\eta_1 x_i - \eta_2 x_i] \\
 &\equiv \beta_0 + \beta_1 x_i
 \end{aligned}$$

where $\beta_0 \equiv [u_{2(1)} - u_{2(2)}]$ and $\beta_1 \equiv [\eta_1 - \eta_2]$.

Similarly, the log-linear model with two predictor variables

$$\log(\mu_{ij}) = u_{1(i)} + u_{2(j)} + \eta_{1j} x_{i1} + \eta_{2j} x_{i2}$$

implies

$$\begin{aligned}
 \log\left(\frac{p_i}{1-p_i}\right) &= \log\left(\frac{\mu_{i1}}{\mu_{i2}}\right) \\
 &= \log(\mu_{i1}) - \log(\mu_{i2}) \\
 &= [u_{1(i)} + u_{2(1)} + \eta_{11} x_{i1} + \eta_{21} x_{i2}] \\
 &\quad - [u_{1(i)} + u_{2(2)} + \eta_{12} x_{i1} + \eta_{22} x_{i2}] \\
 &= [u_{2(1)} - u_{2(2)}] + [\eta_{11} x_{i1} - \eta_{12} x_{i1}] + [\eta_{21} x_{i2} - \eta_{22} x_{i2}] \\
 &\equiv \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2}
 \end{aligned}$$

where $\beta_0 \equiv [u_{2(1)} - u_{2(2)}]$, $\beta_1 \equiv [\eta_{11} - \eta_{12}]$, and $\beta_2 \equiv [\eta_{21} - \eta_{22}]$.

Similar results apply to ANOVA type models. The logit model from Section 21.8

$$\log[p_{hik}/(1-p_{hik})] = (RS)_{hi} + A_k.$$

is equivalent to the log-linear model

$$\log(\mu_{hijk}) = [RSA]_{hik} + [RSO]_{hij} + [OA]_{jk}$$

in which the terms of the logit model become terms in the log-linear model except that the response factor O with subscript j is incorporated into the terms, e.g., $(RS)_{hi}$ becomes $[RSO]_{hij}$ and A_k becomes $[OA]_{jk}$. Also, the log-linear model has an effect for every combination of the explanatory factors, e.g., it includes $[RSA]_{hik}$.

The logit model

$$\log(\mu_{hi1k}/\mu_{hi2k}) = (RS)_{hi} + \gamma k$$

corresponds to the log-linear model

$$\log(\mu_{hijk}) = [RSA]_{hik} + [RSO]_{hij} + \gamma_j k$$

where we have added the highest order interaction term not involving O , $[RSA]$, and made the (RS) and γ terms depend on the opinion level j .

22.8 Multinomial Responses

Logistic regression is used to predict a dependent variable with binary responses. Binary responses involve only two outcomes, either “success” or “failure.” On occasion, we need to predict a dependent variable that has more than two possible outcomes. This can be done by performing a sequence of logistic regressions, although there are several different ways to define such sequences. Another approach is to realize, as in the previous section, that logistic models are equivalent to log-linear models and to use log-linear models to deal with multiple response categories. Christensen (1997) presents a more systematic approach to these problems. Here, we restrict ourselves to presenting some examples.

EXAMPLE 22.8.1. We now examine fitting models to the data from Table 22.3 on race, sex, opinions on abortion, and age. We treat opinions as a three category response variable. In a log-linear model, the variables are treated symmetrically. The analysis looks for relationships among any of the variables. Here we consider opinions as a response variable. This changes the analysis in that we think of having separate independent samples from every combination of the predictor variables. Under this sampling scheme, the interaction among all of the predictors, $[RSA]$, must be included in all models. Table 22.7 presents fits for all the models that include $[RSA]$ and correspond to ANOVA-type logit models.

Using AIC, the best fitting model is clearly $[RSA][RSO][OA]$. The fitted values for $[RSA][RSO][OA]$ are given in Table 22.8.

This log-linear model can be used directly to fit multiple logit models that address specific issues related to the multinomial responses. The method of identifying these logit models is precisely as illustrated in the previous section. For these data we might consider two logit models, one that examines the odds of supporting legalized abortion to opposing legalized abortion and another that examines the odds of opposing legalized abortion to being undecided,

$$\begin{aligned}\log(\mu_{hi1k}/\mu_{hi2k}) &= w_{RS(hi)}^1 + w_{A(R)}^1 \\ \log(\mu_{hi2k}/\mu_{hi3k}) &= w_{RS(hi)}^2 + w_{A(k)}^2.\end{aligned}$$

Alternatively, we might examine the odds of supporting legalized abortion to undecided and the odds of opposing to undecided

$$\begin{aligned}\log(\mu_{hi1k}/\mu_{hi3k}) &= v_{RS(hi)}^1 + v_{A(k)}^1 \\ \log(\mu_{hi2k}/\mu_{hi3k}) &= v_{RS(hi)}^2 + v_{A(k)}^2.\end{aligned}$$

Table 22.8: *Log-Linear Models for the Abortion Opinion Data*

Model	<i>df</i>	G^2	$A - q$
[RSA][RSO][ROA][SOA]	10	6.12	-13.88
[RSA][RSO][ROA]	20	7.55	-32.45
[RSA][RSO][SOA]	20	13.29	-26.71
[RSA][ROA][SOA]	12	16.62	-7.38
[RSA][RSO][OA]	30	14.43	-45.57
[RSA][ROA][SO]	22	17.79	-26.21
[RSA][SOA][RO]	22	23.09	-20.91
[RSA][RO][SO][OA]	32	24.39	-39.61
[RSA][RO][SO]	42	87.54	3.54
[RSA][RO][OA]	34	34.41	-33.59
[RSA][SO][OA]	34	39.63	-28.37
[RSA][RO]	44	97.06	9.06
[RSA][SO]	44	101.9	13.9
[RSA][OA]	36	49.37	-22.63
[RSA][O]	46	111.1	19.1

Table 22.9: *Fitted Values for [RSA][RSO][OA]*

Race	Sex	Opinion	Age					
			18-25	26-35	36-45	46-55	56-65	65+
White	Male	Support	100.1	137.2	117.5	75.62	70.58	80.10
		Oppose	39.73	64.23	56.17	47.33	50.99	62.55
		Undec.	1.21	2.59	5.36	5.05	5.43	8.35
	Female	Support	138.4	172.0	152.4	101.8	101.7	110.7
		Oppose	43.49	63.77	57.68	50.44	58.19	68.43
		Undec.	2.16	4.18	8.96	8.76	10.08	14.86
Nonwhite	Male	Support	21.19	16.57	15.20	11.20	8.04	7.80
		Oppose	8.54	7.88	7.38	7.11	5.90	6.18
		Undec.	1.27	1.54	3.42	3.69	3.06	4.02
	Female	Support	21.40	26.20	19.98	16.38	13.64	12.40
		Oppose	4.24	6.12	4.77	5.12	4.92	4.83
		Undec.	0.36	0.68	1.25	1.50	1.44	1.77

Other possibilities exist. Of these listed, the only odds that seem particularly interesting to the author are the odds of supporting to opposing. In the second pair of models, the category “undecided” is being used as a standard level to which other levels are compared. This seems a particularly bad choice in the context of these data. The fact that undecided happens to be the last opinion category listed in the table is no reason for it to be chosen as the standard of comparison. Either of the other categories would make a better standard.

Neither of these pairs of models are particularly appealing for these data, so we only illustrate a few salient points before moving on. Consider the odds of support relative to opposed. The odds can be obtained from the fitted values in Table 22.8. For example, the odds for young white males are $100.1/39.73 = 2.52$. The full table of odds is given in Table 22.9. Except for nonwhite females, the odds of support are essentially identical to those obtained in Section 21.8 in which undecideds were excluded. The four values vary from age to age by a constant multiple depending on the ages involved. The odds of support decrease steadily with age. The model has no inherent structure among the four race-sex categories; however, the odds for white males and nonwhite males are surprisingly similar. Nonwhite females are most likely to support legalized abortion, white females are next, and males are least likely to support legalized abortion. Confidence intervals for log odds or log odds ratios can be found using methods in Christensen (1997).

Table 22.10: *Estimated Odds of Support versus Oppose*

		Legalized Abortion (Based on the log-linear model [RSA][RSO][OA])					
Race	Sex	Age					
		18-25	26-35	36-45	46-55	56-65	65+
White	Male	2.52	2.14	2.09	1.60	1.38	1.28
	Female	3.18	2.70	2.64	2.02	1.75	1.62
Nonwhite	Male	2.48	2.10	2.06	1.57	1.36	1.26
	Female	5.05	4.28	4.19	3.20	2.77	2.57

Another approach to modelling is to examine the set of three models that consists of the odds of supporting, the odds of opposing, and the odds of undecided (in each case, the odds are defined relative to the union of the other categories).

Finally, we could examine two models, one for the odds of supporting to opposing and one for the odds of undecided to having an opinion. Fitting these two models involves fitting log-linear models to two sets of data. Eliminating all undecideds from the data, we fit [RSA][RSO][OA] to the $2 \times 2 \times 2 \times 6$ table containing only the “support” and “oppose” categories. We essentially did this already in Section 21.8.

We now pool the support and oppose categories to get a $2 \times 2 \times 2 \times 6$ table in which the opinions are “support or oppose” and “undecided.” Again, the model [RSA][RSO][OA] is fitted to the data. For this model, we report only the estimated odds.

		Odds of Being Decided on Abortion					
Race	Sex	Age					
		18-25	26-35	36-45	46-55	56-65	65+
White	Male	116.79	78.52	32.67	24.34	22.26	16.95
	Female	83.43	56.08	23.34	17.38	15.90	12.11
Nonwhite	Male	23.76	15.97	6.65	4.95	4.53	3.45
	Female	68.82	46.26	19.25	14.34	13.12	9.99

The estimated odds vary from age to age by a constant multiple. The odds decrease with age, so older people are less likely to take a position. White males are most likely to state a position. Nonwhite males are least likely to state a position. White and nonwhite females have odds of being decided that are somewhat similar.

Using Table 22.7, the G^2 for [RSA][RSO][OA] is 5.176 on 15 *df*. The G^2 for the smaller model [RSA][RO][SO][OA] is 12.71 on 16 *df*. The difference is very large. Although as seen in Section 21.8, a main effects only logit model fits the support-opposition data quite well, to deal with the undecided category requires a race-sex interaction.

Additional modeling similar to that in Section 21.8 can be applied to the odds of having made a decision on legalized abortion.

22.9 Logistic Discrimination and Allocation

Consider four populations of people determined by age: adult, adolescent, child, infant. These are common distinctions, but the populations are not clearly defined. It is not clear when infants become children, when children become adolescents, nor when adolescents become adults. Nonetheless, most people can clearly be identified as members of one of these four groups. It might be of interest to see whether one can *discriminate* between these populations on the basis of various measurements like height and weight. Another interesting problem is predicting the population of a new individual given only the measurements like height and weight. The problem of predicting the population of a new case is referred to as *allocation*. In a standard discrimination-allocation problem, independent samples are taken from each population. The factor of interest in these problems

Table 22.11: *Cushing's Syndrome Data*

Case	Type	TETRA	PREG	Case	Type	TETRA	PREG
1	A	3.1	11.70	12	B	15.4	3.60
2	A	3.0	1.30	13	B	7.7	1.60
3	A	1.9	0.10	14	B	6.5	0.40
4	A	3.8	0.04	15	B	5.7	0.40
5	A	4.1	1.10	16	B	13.6	1.60
6	A	1.9	0.40	17	C	10.2	6.40
7	B	8.3	1.00	18	C	9.2	7.90
8	B	3.8	0.20	19	C	9.6	3.10
9	B	3.9	0.60	20	C	53.8	2.50
10	B	7.8	1.20	21	C	15.8	7.60
11	B	9.1	0.60				

is the population, but it is not a response factor in the sense used elsewhere in this chapter. The logistic regression approach (or as presented here, the log-linear model approach) to discrimination treats the distribution for each population as a multinomial. While the procedures illustrated are quite straightforward, their philosophical justification is more complex, cf. Christensen (1997).

EXAMPLE 22.9.1. Aitchison and Dunsmore (1975, p. 212) consider 21 individuals with one of 3 types of Cushing's syndrome. Cushing's syndrome involves overproduction of cortisol. The three types considered are

A—adenoma
B—bilateral hyperplasia
C—carcinoma

The case variables considered are the rates at which two steroid metabolites are excreted in the urine. (These are measured in milligrams per day.) The two steroids are

TETRA – Tetrahydrocortisone

and

PREG – Pregnanetriol.

The data are listed in Table 22.10.

The data determine the 3×21 table

Type(<i>i</i>)	1	2	3	4	5	6	7	8	...	16	17	18	19	20	21
A	1	1	1	1	1	1	0	0	...	0	0	0	0	0	0
B	0	0	0	0	0	0	1	1	...	1	0	0	0	0	0
C	0	0	0	0	0	0	0	0	...	0	1	1	1	1	1

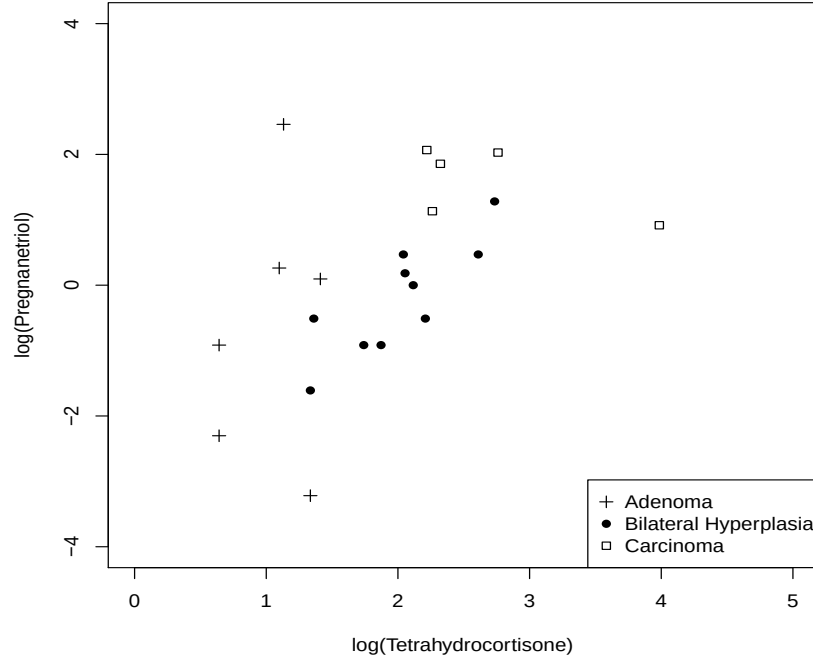
to which we fit a log-linear model.

The case variables TETRA and PREG are used to model the interaction in this table. The case variables are highly skewed so, following Aitchison and Dunsmore, we analyze the transformed variables $TL \equiv \log(\text{TETRA})$ and $PL \equiv \log(\text{PREG})$. The transformed data are plotted in Figure 22.1.

The evaluation of the relationship is based on the relative likelihoods of the three syndrome types. Thus with i denoting the population for any case j , our interest is in the relative sizes of p_{1j} , p_{2j} , and p_{3j} . Estimates of these quantities are easily obtained from the $\hat{\mu}_{ij}$ s. Simply take

$$\hat{p}_{ij} = \hat{\mu}_{ij} / y_{i\cdot} \quad (22.9.1)$$

For a new patient of unknown syndrome type but whose values of TL and PL place him in category j , the most likely type of Cushing's syndrome is that with the largest value among p_{1j} , p_{2j} and p_{3j} . In practice, new patients are unlikely to fall into one of the 21 previously observed categories

Figure 22.1: *Cushing's Syndrome Data.*

but the modeling procedure is flexible enough to allow allocation of individuals having any values of TL and PL .

Discrimination

The main effects model is

$$\log(\mu_{ij}) = \alpha_i + \beta_j \quad i = 1, 2, 3 \quad j = 1, \dots, 21.$$

We want to use TL and PL to help model the interaction, so fit

$$\log(\mu_{ij}) = \alpha_i + \beta_j + \gamma_i(TL)_j + \gamma_i(PL)_j, \quad (22.9.2)$$

$i = 1, 2, 3, j = 1, \dots, 21$.

This model is very similar to a log-linear version of the logit and logistic models discussed earlier. Taking differences gives, for example

$$\log(\mu_{1j}/\mu_{2j}) = (\alpha_1 - \alpha_2) + (\gamma_{11} - \gamma_{12})(TL)_j + (\gamma_{21} - \gamma_{22})(PL)_j$$

which can be written as

$$\log(\mu_{1j}/\mu_{2j}) = \alpha + \delta_1(TL)_j + \delta_2(PL)_j.$$

Although this looks like a logistic regression model, it has a fundamentally different interpretation. A value p_{ij} is the likelihood within population i of observing category j . Having fitted model (22.9.2), the estimate of the log of the likelihood ratio is

$$\log\left(\frac{\hat{p}_{1j}}{\hat{p}_{2j}}\right) = \log\left(\frac{\hat{\mu}_{1j}/y_{1\cdot}}{\hat{\mu}_{2j}/y_{2\cdot}}\right) = \log\left(\frac{\hat{\mu}_{1j}}{\hat{\mu}_{2j}}\right) - \log\left(\frac{y_{1\cdot}}{y_{2\cdot}}\right).$$

Table 22.12: *Estimated Probabilities*

Case	Group			Case	Group		
	A	B	C		A	B	C
1	.1485	.0012	.0195	12	.0000	.0295	.1411
2	.1644	.0014	.0000	13	.0000	.0966	.0068
3	.1667	.0000	.0000	14	.0001	.0999	.0000
4	.0842	.0495	.0000	15	.0009	.0995	.0000
5	.0722	.0565	.0003	16	.0000	.0907	.0185
6	.1667	.0000	.0000	17	.0000	.0102	.1797
7	.0000	.0993	.0015	18	.0000	.0060	.1879
8	.1003	.0398	.0000	19	.0000	.0634	.0733
9	.0960	.0424	.0000	20	.0000	.0131	.1738
10	.0000	.0987	.0025	21	.0000	.0026	.1948
11	.0000	.0999	.0003				

The ratio p_{1j}/p_{2j} is not an odds of Type A relative to Type B. Both numbers are probabilities but they are probabilities from different populations. The correct interpretation of p_{1j}/p_{2j} is as a likelihood ratio, specifically the likelihood of Type A relative to Type B.

The G^2 for model (2) is 12.30 on 36 degrees of freedom. As in logistic regression, although G^2 is a valid measure of goodness of fit, G^2 cannot legitimately be compared to a χ^2 distribution. However, we can test reduced models. The model

$$\log(\mu_{ij}) = \alpha_i + \beta_j + \gamma_i(TL)_j$$

has $G^2 = 21.34$ on 38 degrees of freedom and

$$\log(\mu_{ij}) = \alpha_i + \beta_j + \gamma_i(PL)_j$$

has $G^2 = 37.23$ on 38 degrees of freedom. Neither of the reduced models provides an adequate fit. (Recall that χ^2 tests of model comparisons like these were also valid for logistic regression.)

Table 22.11 contains estimated probabilities for the three populations. The probabilities are computed using equation (1) and model (2).

Table 22.12 illustrates a Bayesian analysis. For each case j , it gives the estimated posterior probability that the case belongs to each of the three syndrome types. The data consist of the observed TL and PL values in category j . Given that the syndrome type is i , the estimated probability of observing data in category j is $\hat{p}_{ij}$. Let $\pi(i)$ be the prior probability that the case is of syndrome type i . Bayes theorem gives

$$\hat{\pi}(i|Data) = \frac{\hat{p}_{ij}\pi(i)}{\sum_{i=1}^3 \hat{p}_{ij}\pi(i)}.$$

Two choices of prior probabilities are used in Table 22.12: probabilities proportional to sample sizes, i.e., $\pi(i) = y_{i\cdot}/y_{\cdot\cdot}$ and equal probabilities $\pi(i) = \frac{1}{3}$. Prior probabilities proportional to sample sizes are *rarely appropriate*, but they relate in simple ways to standard output, so we give them more prominence than they probably deserve. Both of the sets of posterior probabilities are easily obtained. The table of proportional probabilities is just the table of $\hat{\mu}_{ij}$ values from fitting the log-linear model in the usual way. This follows from two facts: first, $\hat{\mu}_{ij} = y_{i\cdot}\hat{p}_{ij}$ and second, the model fixes the column totals so $\hat{\mu}_{\cdot j} = 1 = y_{\cdot j}$. To obtain the equal probabilities values, simply divide the entries in Table 22.11 by the sum of the three probabilities for each case. Cases that are misclassified by either procedure are indicated with a double asterisk in Table 22.12.

Allocation

Model (2) includes a separate term β_j for each case so it is not clear how model (2) can be used to allocate future cases. We begin with logit models and work back to an allocation model. From (2),

Table 22.13: *Probabilities of Classification*

Case	Group	Proportional Prior Probabilities			Equal Prior Probabilities		
		A	B	C	A	B	C
1	A	.89	.01	.10	.88	.01	.12
2	A	.99	.01	.00	.99	.01	.00
3	A	1.00	.00	.00	1.00	.00	.00
4	A	.50	.50	.00	.63	.37	.00
5	**	A	.43	.57	.00	.56	.44
6	A	1.00	.00	.00	1.00	.00	.00
7	B	.00	.99	.01	.00	.99	.01
8	**	B	.60	.40	.00	.72	.28
9	**	B	.58	.42	.00	.69	.31
10	B	.00	.99	.01	.00	.97	.03
11	B	.00	1.00	.00	.00	1.00	.00
12	**	B	.00	.29	.71	.00	.17
13	B	.00	.97	.03	.00	.93	.07
14	B	.00	1.00	.00	.00	1.00	.00
15	B	.01	.99	.00	.01	.99	.00
16	B	.00	.91	.09	.00	.83	.17
17	C	.00	.10	.90	.00	.05	.95
18	C	.00	.06	.94	.00	.03	.97
19	**	C	.00	.63	.37	.00	.46
20	C	.00	.13	.87	.00	.07	.93
21	C	.00	.03	.97	.00	.01	.99

we can model the probability ratio of type *A* relative to type *B*

$$\begin{aligned}
 \log(p_{1j}/p_{2j}) &= \log(\mu_{1j}/\mu_{2j}) - \log(y_{1\cdot}/y_{2\cdot}) \\
 &= (\alpha_1 - \alpha_2) + (\gamma_{11} - \gamma_{12})(TL)_j + (\gamma_{21} - \gamma_{22})(PL)_j - \log(y_{1\cdot}/y_{2\cdot}).
 \end{aligned} \tag{22.9.3}$$

The log-likelihoods of *A* relative to *C* are

$$\begin{aligned}
 \log(p_{1j}/p_{3j}) &= \log(\mu_{1j}/\mu_{3j}) - \log(y_{1\cdot}/y_{3\cdot}) \\
 &= (\alpha_1 - \alpha_3) + (\gamma_{11} - \gamma_{13})(TL)_j + (\gamma_{21} - \gamma_{23})(PL)_j - \log(y_{1\cdot}/y_{3\cdot}).
 \end{aligned} \tag{22.9.4}$$

Fitting model (2) gives the estimated parameters.

Par.	Est.	Par.	Est.	Par.	Est.
α_1	0.0	γ_{11}	-16.29	γ_{21}	-3.359
α_2	-20.06	γ_{12}	-1.865	γ_{22}	-3.604
α_3	-28.91	γ_{13}	0.0	γ_{23}	0.0

where the estimates with values of 0 are really side conditions imposed on the collection of estimates to make it unique.

For a new case with values *TL* and *PL*, we plug estimates into equations (3) and (4) to get

$$\log(\hat{p}_1/\hat{p}_2) = 20.06 + (-16.29 + 1.865)TL + (-3.359 + 3.604)PL - \log(6/10)$$

and

$$\log(\hat{p}_1/\hat{p}_3) = 28.91 - 16.29(TL) - 3.359(PL) - \log(6/5).$$

For example, if the new case has a tetrahydrocortisone reading of 4.1 and a pregnanetriol reading of 1.10 then $\log(\hat{p}_1/\hat{p}_2) = .24069$ and $\log(\hat{p}_1/\hat{p}_3) = 5.4226$. The likelihood ratios are

$$\hat{p}_1/\hat{p}_2 = 1.2721$$

$$\hat{p}_1/\hat{p}_3 = 226.45$$

and by division

$$\hat{p}_2/\hat{p}_3 = 226.45/1.2721 = 178.01.$$

It follows that Type A is a little more likely than Type B and that both are much more likely than Type C

One can also obtain estimated posterior probabilities for a new case. The posterior odds are

$$\frac{\hat{\pi}(1|Data)}{\hat{\pi}(2|Data)} = \frac{\hat{p}_1 \pi(1)}{\hat{p}_2 \pi(2)} \equiv \hat{O}_2$$

and

$$\frac{\hat{\pi}(1|Data)}{\hat{\pi}(3|Data)} = \frac{\hat{p}_1 \pi(1)}{\hat{p}_3 \pi(3)} \equiv \hat{O}_3.$$

Using the fact that $\hat{\pi}(1|Data) + \hat{\pi}(2|Data) + \hat{\pi}(3|Data) = 1$, we can solve for $\hat{\pi}(i|Data)$, $i = 1, 2, 3$. In particular,

$$\begin{aligned}\hat{\pi}(1|Data) &= \left[1 + \frac{1}{\hat{O}_2} + \frac{1}{\hat{O}_3}\right]^{-1} = \frac{\hat{O}_2 \hat{O}_3}{\hat{O}_2 \hat{O}_3 + \hat{O}_3 + \hat{O}_2} \\ \hat{\pi}(2|Data) &= \frac{1}{\hat{O}_2} \left[1 + \frac{1}{\hat{O}_2} + \frac{1}{\hat{O}_3}\right]^{-1} = \frac{\hat{O}_3}{\hat{O}_2 \hat{O}_3 + \hat{O}_3 + \hat{O}_2} \\ \hat{\pi}(3|Data) &= \frac{1}{\hat{O}_3} \left[1 + \frac{1}{\hat{O}_2} + \frac{1}{\hat{O}_3}\right]^{-1} = \frac{\hat{O}_2}{\hat{O}_2 \hat{O}_3 + \hat{O}_3 + \hat{O}_2}.\end{aligned}$$

Using TETRA = 4.10 and PREG = 1.10, the assumption $\pi(i) = y_i./y_{..}$ and more numerical accuracy in the parameter estimates than was reported earlier,

$$\begin{aligned}\hat{\pi}(1|Data) &= .433 \\ \hat{\pi}(2|Data) &= .565 \\ \hat{\pi}(3|Data) &= .002.\end{aligned}$$

Assuming $\pi(i) = 1/3$ gives

$$\begin{aligned}\hat{\pi}(1|Data) &= .560 \\ \hat{\pi}(2|Data) &= .438 \\ \hat{\pi}(3|Data) &= .002.\end{aligned}$$

Note that the values of tetrahydrocortisone and pregnanetriol used are identical to those for case 5; thus the $\hat{\pi}(i|Data)$'s are identical to those listed in Table 22.12 for case 5.

To use the log-linear model approach illustrated here, one needs to fit a 3×21 table. Typically, a data file of 63 entries is needed. Three rows of the data file are associated with each of the 21 cases. Each data entry has to be identified by case and by type. In addition, the case variables should be included in the file in such a way that all three rows for a case include the corresponding case variables, *TL* and *PL*. Model (2) is easily fitted using GENMOD.

It is easy to just fit log-linear or logistic models to data such as that in Table 22.10 and get $\hat{\mu}_{ij}$'s or $\hat{p}_{ij}$'s. If you treat these values as estimated probabilities for being in the various populations, you are doing a Bayesian analysis with prior probabilities proportional to sample sizes. This is rarely an appropriate methodology.

Table 22.14: *Tool Life Factors*

Run	Chaser	Coolant	Speed	Pipe	Rake Angle	Spindle	Head
Run	Ch	Cl	Spd	P	RA	Spn	H

Table 22.15: *Poisson data*

Run	Ch	Cl	Spd	P	RA	Spn	H	Ends cut		
1	1	1	1	1	1	1	1	137	24	58
2	1	1	1	2	2	2	1	89	41	26
3	1	1	2	1	2	1	1	56	34	199
4	1	1	2	2	1	1	1	545	105	106
4	1	1	2	2	1	1	1	122	132	168
4	1	1	2	2	1	2	1	74	66	49
5	1	2	1	1	2	1	1	428	157	188
6	1	2	1	2	1	2	1	352	68	97
7	1	2	2	1	1	1	1	320	750	988
8	1	2	2	2	2	2	2	632	73	529
9	2	1	1	1	2	2	2	44	41	7
10	2	1	1	2	1	2	2	2	3	4
11	2	1	2	1	1	2	2	112	64	59
12	2	1	2	2	2	2	2	59	45	9
13	2	2	1	1	1	2	2	40	7	32
14	2	2	1	2	2	2	2	21	42	120
15	2	2	2	1	2	2	2	14	19	28
15	2	2	2	1	2	2	3	168	34	58
16	2	2	2	2	1	2	3	70	57	18

22.10 Exercises

EXERCISE 22.1. Reanalyze the data using a model based on TETRA and PREG rather than TL and PL .

EXERCISE 22.2. Watkins, Bergman, and Horton (1994) presented data on a complicated designed experiment that generated counts. The dependent variable is the number of ends cut by a tool. The experiment was a half replication of an experiment with five factors each at two levels, i.e., a half rep. of a 2^5 or a 2^{5-1} . The factors in the design of the experiment were the first five factors listed in Table 22.13. There are two different chasers, two different coolants, the two speeds were coded as intermediate (1) and high (2), two different pipes, and two different rake angles. In addition, two covariates were observed. On each run it was noted whether the spindle was left (1) or right (2). In the course of the experiment, two new heads were installed. A new head was installed prior to run number 8 and also prior to the second observation of run 15. The data are given in Table 22.14. It seems reasonable to treat all of the observations as independent Poisson random variables. Analysis of variance type models on seven factors can be performed. Analyze the data.

EXERCISE 22.3. Bisgaard and Fuller (1995) give the data in Table 22.15 on the numbers of defectives per grille in a process examined by Chrysler Motors Engineering. The data are from a fractional factorial design. The factors are

- A Mold Cycle
- B Viscosity
- C Mold Temp
- D Mold Pressure
- E Weight

Table 22.16: *Grille Defectives*

Case	treatment subscripts									Defectives
	a	b	c	d	e	f	g	h	j	
1	0	0	0	0	1	0	1	0	1	56
2	1	0	0	0	1	0	0	1	0	17
3	0	1	0	0	0	1	1	0	0	2
4	1	1	0	0	0	1	0	1	1	4
5	0	0	1	0	1	1	0	1	1	3
6	1	0	1	0	1	1	1	0	0	4
7	0	1	1	0	0	0	0	1	0	50
8	1	1	1	0	0	0	1	0	1	2
9	0	0	0	1	0	1	1	1	1	1
10	1	0	0	1	0	1	0	0	0	0
11	0	1	0	1	1	0	1	1	0	3
12	1	1	0	1	1	0	0	0	1	12
13	0	0	1	1	0	0	0	0	1	3
14	1	0	1	1	0	0	1	1	0	4
15	0	1	1	1	1	1	0	0	0	0
16	1	1	1	1	1	1	1	1	1	0

F Priming

G Thickening Process

H Glass Type

J Cutting Pattern.

Fit a main effects only model. Try to fit a model with all main effects and all two factor interactions.