

Models for Three-Dimensional Tables

TABLE 5.2: Socio-economic Status, Parental Encouragement and Educational Aspirations of High School Seniors

Social Stratum	Parental Encouragement	College Plans		Total
		No	Yes	
Lower	Low	749	35	784
	High	233	133	366
Lower Middle	Low	627	38	665
	High	330	303	633
Upper Middle	Low	420	37	457
	High	374	467	841
Higher	Low	153	26	179
	High	266	800	1066
Total		3152	1938	4991

TABLE 5.3: Deviances for Log-linear Models Fitted to Educational Aspirations Data

Model	Deviance	d.f.
$S + E + P$	2714.0	10
$SE + P$	1877.4	7
$SP + E$	1920.4	7
$S + EP$	1092.0	9
$SE + SP$	1083.8	4
$SE + EP$	255.5	6
$SP + EP$	298.5	6
$SE + SP + EP$	1.575	3

Table 9.7 analysed by logistic regression

```
> data97Logistic <- read.table(file="Table9.7Logistic.txt",header=T)
> data97Logistic <- data.frame(data97Logistic)
> data97Logistic
  case ulcer nonasp asp
1    0     0     62  6
2    1     0     39 25
3    0     1     53  8
4    1     1     49  8
```

104

```
> M1 <- glm(cbind(nonasp,asp) ~ case + ulcer + case*ulcer, family=binomial,
data=data97Logistic)
> summary(M1)
```

Call:

```
glm(formula = cbind(nonasp, asp) ~ case + ulcer + case * ulcer,
     family = binomial, data = data97Logistic)
```

Deviance Residuals:

```
[1] 0 0 0 0
```

Coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
(Intercept)	2.3354	0.4275	5.462	4.7e-08	***
case	-1.8907	0.4984	-3.793	0.000149	***
ulcer	-0.4445	0.5715	-0.778	0.436711	
case:ulcer	1.8122	0.7333	2.471	0.013460	*

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 2.1789e+01 on 3 degrees of freedom
Residual deviance: 1.3101e-14 on 0 degrees of freedom
AIC: 23.720

Number of Fisher Scoring iterations: 3

Corresponds to saturated log-linear model!

105

```
> M2 <- glm(cbind(nonasp,asp) ~ case + ulcer, family=binomial, data=data97Logistic)
> summary(M2)
```

Call:

```
glm(formula = cbind(nonasp, asp) ~ case + ulcer, family = binomial,
     data = data97Logistic)
```

Deviance Residuals:

```
      1      2      3      4
1.289 -0.906 -1.540  1.196
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   1.8219     0.3080   5.916 3.3e-09 ***
case          -1.1429     0.3521  -3.246 0.00117 **
ulcer          0.7000     0.3460   2.023 0.04306 *
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 21.789  on 3  degrees of freedom
Residual deviance:  6.283  on 1  degrees of freedom
AIC: 28.003
```

Number of Fisher Scoring iterations: 4

Corresponds to model m4 with all two-factor interactions!

106

```
> M3 <- glm(cbind(nonasp,asp) ~ case, family=binomial, data=data97Logistic)
> summary(M3)
```

Call:

```
glm(formula = cbind(nonasp, asp) ~ case, family = binomial, data = data97Logistic)
```

Deviance Residuals:

```
      1      2      3      4
0.5540 -2.0416 -0.5521  2.3997
```

Coefficients:

```
              Estimate Std. Error z value Pr(>|z|)
(Intercept)   2.1059     0.2831   7.440 1.01e-13 ***
case          -1.1250     0.3490  -3.224 0.00127 **
```

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

```
Null deviance: 21.789  on 3  degrees of freedom
Residual deviance: 10.538  on 2  degrees of freedom
AIC: 30.259
```

Number of Fisher Scoring iterations: 4

Corresponds to model m3 with interaction case * aspirin!

107

```

> M4 <- glm(cbind(nonasp,asp) ~ 1, family=binomial, data=data97Logistic)
> summary(M4)

Call:
glm(formula = cbind(nonasp, asp) ~ 1, family = binomial, data = data97Logistic)

Deviance Residuals:
    1      2      3      4
 2.3005 -3.7649  1.1880  0.9548

Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  1.4631      0.1619   9.038  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 21.789  on 3  degrees of freedom
Residual deviance: 21.789  on 3  degrees of freedom
AIC: 39.51

Number of Fisher Scoring iterations: 4

```

Corresponds to model m2 with main effect aspirin!

108

In general, log-linear and logit models are equivalent as long as the log-linear model

- is saturated on all factors treated as predictors in the logit model, including all possible main effects and interactions among predictors (in our example *SE*),
- includes a main effect for the factor treated as response (in our example *P*), and
- includes a two-factor (or higher order) interaction between a predictor and the response for each main effect (or interaction) included in the logit model (in our example it includes *SP* for the main effect of *S*, and *son on*).

109

Dispersion parameters (Piegorisch 12.1-12.4)

Thus the class becomes

$$f(y) = \exp \left\{ \frac{y\theta - \gamma(\theta)}{\psi} + d(y, \psi) \right\}. \quad (12.2)$$

(In some cases, a set of prior, known weights, w , may be included in the analysis. To do so, simply change all instances where ψ appears to the inverse-weighted dispersion ψ/w . SAS PROC GENMOD handles such a setting via the SCWGT statement. If no weights are to be included, let $w = 1$ for all y .) An additional, important restriction continues to be that the support space for Y cannot depend upon any unknown parameters.

Recall that for any $f(y)$ satisfying (12.1), $E[Y] = -c'(\theta)/b'(\theta)$. It is common to denote this simply as $\mu = E[Y]$. Under (12.2), this becomes

$$\mu = E[Y] = \gamma'(\theta). \quad (12.3)$$

Similarly, under (12.1) $\text{var}[Y] = [b''(\theta)c'(\theta) - b'(\theta)c''(\theta)]/[b'(\theta)]^3$. Under (12.2), this simplifies to

$$\text{var}[Y] = \psi \frac{\partial^2 \gamma(\theta)}{\partial \theta^2}, \quad (12.4)$$

i.e., $\text{var}[Y] = \psi \gamma''(\theta)$. We often find that the variance simplifies to $\text{var}[y] = \psi V(\mu)$, for some function $V(\mu)$ depending only on $\mu = \gamma'(\theta)$. In such instances, we refer to $V(\mu)$ as the *variance function*.

110

$$\begin{aligned} \ell(\hat{\boldsymbol{\vartheta}}_{\max}; \mathbf{y}) - \ell(\hat{\boldsymbol{\vartheta}}; \mathbf{y}) &= \sum \left\{ \frac{y_i \hat{\theta}_{i, \max} - \gamma(\hat{\theta}_{i, \max})}{\psi} + d(y_i, \psi) - \frac{y_i \hat{\theta}_i - \gamma(\hat{\theta}_i)}{\psi} - d(y_i, \psi) \right\} \\ &= \frac{\sum \left\{ y_i (\hat{\theta}_{i, \max} - \hat{\theta}_i) - \gamma(\hat{\theta}_{i, \max}) + \gamma(\hat{\theta}_i) \right\}}{\psi}, \end{aligned} \quad (12.5)$$

Technically, the deviance function, $D(\mathbf{y}, \hat{\boldsymbol{\vartheta}})$, for the current model is twice the *numerator* in (12.5):

$$D(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) = 2 \sum \left\{ y_i (\hat{\theta}_{i, \max} - \hat{\theta}_i) - \gamma(\hat{\theta}_{i, \max}) + \gamma(\hat{\theta}_i) \right\}. \quad (12.6)$$

If $\psi = 1$, then (12.5) and (12.6) coincide (to a factor of 2) and constructions using either form will be equivalent. If the dispersion parameter is not equal to one, however, then its value must be incorporated into the log-likelihood ratio. Traditional GLiMs do so by defining the *scaled deviance* $D^*(\mathbf{y}, \hat{\boldsymbol{\vartheta}})$:

$$D^*(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) = \frac{D(\mathbf{y}, \hat{\boldsymbol{\vartheta}})}{\psi} \quad (12.7)$$

which is simply (12.5) multiplied by two. [A better phrase for (12.7) would be to call it the dispersion-adjusted deviance, since as noted above the term “scale” is used in many different ways for models satisfying (12.2).] Thus $D^*(\mathbf{y}, \hat{\boldsymbol{\vartheta}}) = 2 \{ \ell(\hat{\boldsymbol{\vartheta}}_{\max}; \mathbf{y}) - \ell(\hat{\boldsymbol{\vartheta}}; \mathbf{y}) \}$ is the traditional log-likelihood ratio. This is known to converge in distribution to $\chi^2(\nu)$, where ν is the difference in numbers of parameters fit in $\hat{\boldsymbol{\vartheta}}_{\max}$ and $\hat{\boldsymbol{\vartheta}}$. Note, however, that as was the case with no dispersion parameter, this asymptotic approximation can be poor, and it is not recommended for more than informal use.

111

12.4 Quasi-likelihood

An interesting realization was made by Wedderburn (1974) regarding estimation of the unknown regression parameters, $\boldsymbol{\beta}$, in a GLiM: the ML estimating equation for each β_j has the same form. From Dobson (2002, Eqn. 4.18) we have that the log-likelihood derivative (“score”) is

$$\frac{\partial \ell}{\partial \beta_j} = \sum \frac{[y_i - \mu_i(\boldsymbol{\beta})] x_{ij}}{\text{var}[Y_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right), \quad (12.11)$$

where $\mu_i(\boldsymbol{\beta})$ is written to represent the mean as a function of $\boldsymbol{\beta}$. Recall that we denote the score as $U_j = \partial \ell / \partial \beta_j$. Extended to the larger exponential class in (12.2), U_j simplifies to

$$U_j = \frac{\partial \ell}{\partial \beta_j} \propto \sum \frac{[y_i - \mu_i(\boldsymbol{\beta})] x_{ij}}{V[\mu_i]} \left(\frac{\partial \mu_i}{\partial \eta_i} \right)$$

(since ψ is treated as a constant). $V(\mu_i)$ is the variance function from $\text{var}[Y_i] = \psi V(\mu_i)$. Now, recognize that when $\eta_i = \mathbf{x}_i^T \boldsymbol{\beta}$, $\partial \eta_i / \partial \beta_j$ is simply x_{ij} , so from the chain rule

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij}. \quad (12.12)$$

From this, one can write U_j as

$$U_j = \frac{\partial \ell}{\partial \beta_j} = \sum \frac{[y_i - \mu_i(\boldsymbol{\beta})]}{V[\mu_i]} \left(\frac{\partial \mu_i}{\partial \beta_j} \right) \quad (12.13)$$

(Zeger and Liang, 1992). To find the MLE of β_j , $j=1, \dots, P$, set $U_j = 0$ and solve the equation numerically.

112

Examples: quasi-likelihood estimation

```
quasi> counts <- c(18, 17, 15, 20, 10, 20, 25, 13, 12)
quasi> outcome <- gl(3, 1, 9)
quasi> treatment <- gl(3, 3)
quasi> d.AD <- data.frame(treatment, outcome, counts)
quasi> glm.qD93 <- glm(counts ~ outcome + treatment, family = quasipoisson())
quasi> summary(glm.qD93)
Call:
glm(formula = counts ~ outcome + treatment, family = quasipoisson())
Deviance Residuals:
    1     2     3     4     5     6     7     8     9
-0.67125  0.96272 -0.16965 -0.21999 -0.95552  1.04939  0.84715 -0.09167
-0.96656
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  3.045e+00  1.944e-01  15.665  9.7e-05 ***
outcome2    -4.543e-01  2.299e-01  -1.976   0.119
outcome3    -2.930e-01  2.192e-01  -1.337   0.252
treatment2   8.717e-16  2.274e-01  3.83e-15  1.000
treatment3   4.557e-16  2.274e-01  2.00e-15  1.000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasipoisson family taken to be 1.293300)
Null deviance: 10.5814 on 8 degrees of freedom
Residual deviance: 5.1291 on 4 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 4
```

113

```

quasi> summary(glm.qD93, dispersion = 1)
Call:
glm(formula = counts ~ outcome + treatment, family = quasipoisson())
Deviance Residuals:
    1     2     3     4     5     6     7     8     9
-0.67125  0.96272 -0.16965 -0.21999 -0.95552  1.04939  0.84715 -0.09167
-0.96656
Coefficients:
            Estimate Std. Error  z value Pr(>|z|)
(Intercept)  3.045e+00  1.709e-01  17.815  <2e-16 ***
outcome2    -4.543e-01  2.022e-01  -2.247  0.0246 *
outcome3    -2.930e-01  1.927e-01  -1.520  0.1285
treatment2   8.717e-16  2.000e-01  4.36e-15  1.0000
treatment3   4.557e-16  2.000e-01  2.28e-15  1.0000
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasipoisson family taken to be 1)
Null deviance: 10.5814 on 8 degrees of freedom
Residual deviance: 5.1291 on 4 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 4

```

114

```

> x <- rnorm(100)
> y <- rpois(100, exp(1 + x))
> mql <- glm(y ~ x, family = quasi(var = "mu", link = "log"))
> mq2 <- glm(y ~ x, family = quasipoisson)
> summary(mql)
Call:
glm(formula = y ~ x, family = quasi(var = "mu", link = "log"))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.76467 -0.91554 -0.04998  0.59881  2.55591
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.88677     0.07357  12.05  <2e-16 ***
x            1.08539     0.06099  17.80  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasi family taken to be 1.080803)
Null deviance: 452.76 on 99 degrees of freedom
Residual deviance: 110.91 on 98 degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 5

```

115

```

> mq2 <- glm(y ~ x, family = quasipoisson)
> summary(mq2)
Call:
glm(formula = y ~ x, family = quasipoisson)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.76467 -0.91554 -0.04998  0.59881  2.55591
Coefficients:
            Estimate Std. Error t value Pr(>|t|)
(Intercept)  0.88677    0.07357   12.05  <2e-16 ***
x            1.08539    0.06099   17.80  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasipoisson family taken to be 1.080803)
Null deviance: 452.76  on 99  degrees of freedom
Residual deviance: 110.91  on 98  degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 5

```

116

```

> m <- glm(y ~ x, family = poisson)
> summary(m)
Call:
glm(formula = y ~ x, family = poisson)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-2.76467 -0.91554 -0.04998  0.59881  2.55591
Coefficients:
            Estimate Std. Error z value Pr(>|z|)
(Intercept)  0.88677    0.07077   12.53  <2e-16 ***
x            1.08539    0.05867   18.50  <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for poisson family taken to be 1)
Null deviance: 452.76  on 99  degrees of freedom
Residual deviance: 110.91  on 98  degrees of freedom
AIC: 355.53
Number of Fisher Scoring iterations: 5

```

117


```

> y <- rbinom(100, 1, plogis(x))
> mqb <- glm(y ~ x, family = quasi(var = "mu(1-mu)", link = "logit"), start = c(0, 1))
> summary(mqb)
Call:
glm(formula = y ~ x, family = quasi(var = "mu(1-mu)", link = "logit"),
     start = c(0, 1))
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9713 -0.9877 -0.6172  1.0371  1.9627
Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -0.02093    0.22517  -0.093 0.926123
x             1.02064    0.28555   3.574 0.000547 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for quasi family taken to be 1.038343)
Null deviance: 137.99  on 99  degrees of freedom
Residual deviance: 121.37  on 98  degrees of freedom
AIC: NA
Number of Fisher Scoring iterations: 3

```

118

```

> mb <- glm(y ~ x, family = binomial)
> summary(mb)
Call:
glm(formula = y ~ x, family = binomial)
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.9713 -0.9877 -0.6172  1.0371  1.9627
Coefficients:
              Estimate Std. Error z value Pr(>|z|)
(Intercept) -0.02093    0.22097  -0.095 0.924531
x             1.02064    0.28023   3.642 0.000270 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
(Dispersion parameter for binomial family taken to be 1)
Null deviance: 137.99  on 99  degrees of freedom
Residual deviance: 121.37  on 98  degrees of freedom
AIC: 125.37
Number of Fisher Scoring iterations: 4

```

119