CHAPTER 8

Quasi-Likelihood Methods

For a GLM $\eta_i = g(\mu_i) = \sum_j \beta_j x_{ij}$, the likelihood equations

$$\sum_{i=1}^{n} \frac{(y_i - \mu_i)x_{ij}}{v(\mu_i)} \left(\frac{\partial \mu_i}{\partial \eta_i} \right) = 0, \quad j = 1, \ldots, p, \tag{8.1}$$

depend on the assumed probability distribution for y_i only through μ_i and the variance function, $v(\mu_i) = \text{var}(y_i)$. The choice of distribution for y_i determines the relation $v(\mu_i)$ between the variance and the mean. Higher moments such as the skewness can affect properties of the model, such as how fast $\hat{\boldsymbol{\beta}}$ converges to normality, but they have no impact on the value of $\hat{\boldsymbol{\beta}}$ and its large-sample covariance matrix.

An alternative approach, *quasi-likelihood estimation*, specifies a link function and linear predictor $g(\mu_i) = \sum_j \beta_j x_{ij}$ like a generalized linear model (GLM), but it does not assume a particular probability distribution for y_i. This approach estimates $\{\beta_j\}$ by solving equations that resemble the likelihood equations (8.1) for GLMs, but it assumes only a mean–variance relation for the distribution of y_i. The estimates are the solution of Equation (8.1) with $v(\mu_i)$ replaced by whatever variance function seems appropriate in a particular situation, with a corresponding adjustment for standard errors. To illustrate, a standard modeling approach for counts assumes that $\{y_i\}$ are independent Poisson variates, for which $v(\mu_i) = \mu_i$. However, in the previous chapter we noted that overdispersion often occurs, perhaps because of unmodeled heterogeneity among subjects. To allow for this, we could set $v(\mu_i) = \phi\mu_i$ for some unknown constant ϕ.

In Section 8.1, we present a simple quasi-likelihood (QL) approach for overdispersed Poisson and binomial response variables that merely assumes an inflation of the variance from a standard model. For binary data, Section 8.2 presents alternative approaches that imply overdispersion because of positively correlated Bernoulli trials or because the success probability satisfies a mixture model. In Section 8.3, we show

Foundations of Linear and Generalized Linear Models, First Edition. Alan Agresti.

how to adjust for misspecification of the variance function in finding standard errors of parameter estimates.

8.1 VARIANCE INFLATION FOR OVERDISPERSED POISSON AND BINOMIAL GLMS

This section introduces a simple quasi-likelihood way of adjusting for overdispersion in Poisson and binomial models. This method uses the same estimates as an ordinary GLM but inflates standard errors by taking into account the empirical variability.

8.1.1 Quasi-Likelihood Approach of Variance Inflation

Suppose a standard model specifies a function $v^*(\mu_i)$ for the variance as a function of the mean, but we believe that the actual variance may differ from $v^*(\mu_i)$. To allow for this, we might instead assume that

$$\text{var}(y_i) = \phi v^*(\mu_i)$$

for some constant ϕ. The value $\phi > 1$ represents overdispersion.

When we substitute $v(\mu_i) = \phi v^*(\mu_i)$ in Equation (8.1), ϕ drops out. The equations are identical to the likelihood equations for the GLM with variance function $v^*(\mu_i)$, and estimates of model parameters are also identical. With the generalized variance function,

$$w_i = (\partial\mu_i/\partial\eta_i)^2/\text{var}(y_i) = (\partial\mu_i/\partial\eta_i)^2/\phi v^*(\mu_i),$$

so the asymptotic $\text{var}(\hat{\boldsymbol{\beta}}) = (\boldsymbol{X}^{\text{T}}\boldsymbol{W}\boldsymbol{X})^{-1}$ is ϕ times that for the ordinary GLM.

When a variance function has the form $v(\mu_i) = \phi v^*(\mu_i)$, usually ϕ is also unknown. Let

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\mu}_i)^2}{v^*(\hat{\mu}_i)}$$

be the generalized Pearson statistic (4.17) for the simpler model with $\phi = 1$. When X^2/ϕ is approximately chi-squared, then with p parameters in the linear predictor, $E(X^2/\phi) \approx n - p$. Hence, $E[X^2/(n-p)] \approx \phi$. Using the motivation of estimation by matching moments, $\hat{\phi} = X^2/(n-p)$ is an estimated multiplier to apply to the ordinary estimated covariance matrix.

In summary, this quasi-likelihood approach is simple: fit the ordinary GLM and use its p maximum likelihood (ML) parameter estimates $\hat{\boldsymbol{\beta}}$. Multiply the ordinary standard error estimates by $\sqrt{X^2/(n-p)}$. This method is appropriate; however, only if the model chosen describes well the structural relation between $E(y_i)$ and the explanatory variables. If a large X^2 statistic is due to some other type of lack of fit,

such as failing to include a relevant interaction term, adjusting for overdispersion will not address the inadequacy.

8.1.2 Overdispersed Poisson and Binomial GLMs

We illustrate the quasi-likelihood variance-inflation approach with the alternative to a Poisson GLM in which the mean–variance relation has the form

$$v(\mu_i) = \phi\mu_i.$$

The QL parameter estimates are identical to the ML estimates under the Poisson GLM assumption. With the canonical log link, the adjusted covariance matrix is $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}$ with $w_i = (\partial\mu_i/\partial\eta_i)^2/\mathrm{var}(y_i) = (\mu_i)^2/\phi\mu_i = \mu_i/\phi$. Regardless of the link function, the Pearson statistic is

$$X^2 = \sum_i \frac{(y_i - \hat{\mu}_i)^2}{\hat{\mu}_i},$$

and $\hat{\phi} = X^2/(n-p)$ is the variance-inflation estimate.

An alternative approach uses a parametric model that permits extra variability, such as a negative binomial GLM (Section 7.3). An advantage of that approach is that it is an actual model with a likelihood function.

Overdispersion also can occur for counts from grouped binary data. Suppose y_i is the proportion of successes in n_i Bernoulli trials with parameter π_i for each trial, $i = 1, \ldots, n$. The $\{y_i\}$ may exhibit more variability than the binomial allows. This can happen in two common ways. One way involves heterogeneity, with observations at a particular setting of explanatory variables having success probabilities that vary according to values of unobserved variables. To deal with this, we could use a hierarchical mixture model that lets π_i itself have a distribution, such as a beta distribution. Alternatively, extra variability could occur because the Bernoulli trials at each i are positively correlated. We present models that reflect these possibilities in Section 8.2. Here we consider the simpler variance-inflation approach.

To adjust supposedly binomial sampling (i.e., independent, identical Bernoulli trials), the inflated-variance QL approach uses variance function

$$v(\pi_i) = \phi\pi_i(1-\pi_i)/n_i$$

for the proportion y_i. The QL estimates are the same as ML estimates for the binomial model, and the asymptotic covariance matrix multiplies by ϕ. The $X^2/(n-p)$ estimate of ϕ uses the X^2 fit statistic for the ordinary binomial model with p parameters, which from Equation (5.10) is

$$X^2 = \sum_i \frac{(y_i - \hat{\pi}_i)^2}{[\hat{\pi}_i(1-\hat{\pi}_i)]/n_i}.$$

Although this QL approach with $v(\pi_i) = \phi\pi_i(1-\pi_i)/n_i$ has the advantage of simplicity, it is inappropriate when $n_i = 1$: then $P(y_i = 1) = \pi_i = 1 - P(y_i = 0)$, and necessarily $E(y_i^2) = E(y_i) = \pi_i$ and $\text{var}(y_i) = \pi_i(1-\pi_i)$. For ungrouped binary data, necessarily $\text{var}(y_i) = \pi_i(1-\pi_i)$, and only $\phi = 1$ makes sense. This structural problem does not occur for mixture models or for a QL approach having variance function corresponding to a mixture model (Section 8.2).

8.1.3 Example: Quasi-Likelihood for Horseshoe Crab Counts

The horseshoe crab satellite counts analyzed in Section 7.5 display overdispersion for Poisson GLMs. For example, using the female crab's weight to predict the number of male satellites, the Poisson loglinear fit is $\log \hat{\mu}_i = -0.428 + 0.589x_i$, with $SE = 0.065$ for $\hat{\beta}_1 = 0.589$. Comparing the observed counts and fitted values for the $n = 173$ crabs, Pearson $X^2 = 535.9$ with $df = 173 - 2 = 171$. With the QL inflated-variance approach, $\hat{\phi} = X^2/(n-p) = 535.9/(173-2) = 3.13$. Thus, $SE = \sqrt{3.13}(0.065) = 0.115$ is a more plausible standard error for $\hat{\beta}_1$ in this prediction equation.

```
-------------------------------------------------------------------------
> attach(Crabs)
> fit.pois <- glm(y ~ weight, family=poisson) # ML Poisson loglinear
> summary(fit.pois)
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -0.4284      0.1789    -2.394   0.0167
weight        0.5893      0.0650     9.064   <2e-16

> (X2 <- sum(residuals(fit.pois, type="pearson")^2))
[1] 535.90 # Pearson statistic is sum of squared Pearson residuals
> (phi <- X2/(173 - 2))
[1] 3.13
    # quasi family can use QL inflated Poisson variance directly:
> summary(glm(y ~ weight, family=quasi(link="log",variance="mu")))
            Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -0.4284      0.3168   -1.352     0.178
weight        0.5893      0.1151    5.120  8.17e-07
    (Dispersion parameter for quasi family taken to be 3.134)
-------------------------------------------------------------------------
```

The QL approach with an inflated quadratic variance function yields larger $\hat{\beta}_1$ and *SE* values, similar to what we obtain with a negative binomial (NB2) model.

```
-------------------------------------------------------------------------
> summary(glm(y ~ weight, family=quasi(link="log",variance="mu^2")))
            Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -1.0122      0.3863   -2.621   0.00957
weight        0.8184      0.1542    5.306  3.44e-07
    (Dispersion parameter for quasi family taken to be 1.362496)
> library(MASS)
```

```
> summary(glm.nb(y ~ weight)) # negative binomial (NB2) model
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -0.8647      0.4048   -2.136    0.0327
weight        0.7603      0.1578    4.817  1.45e-06
---
Theta: 0.931  2 x log-likelihood: -748.644 # -916.164 for Poisson fit
-------------------------------------------------------------------------
```

8.2 BETA-BINOMIAL MODELS AND QUASI-LIKELIHOOD ALTERNATIVES

We next describe ways to handle binomial overdispersion that are more satisfying than the variance-inflation approach. We first present a QL method based on correlated Bernoulli trials and then a mixture model that lets success probabilities vary according to values of unobserved variables.

8.2.1 Overdispersion Caused by Correlated Bernoulli Trials

Denote the n_i Bernoulli trials for observation i by $y_{i1}, y_{i2}, \ldots, y_{in_i}$. That is, $P(y_{ij} = 1) = \pi_i = 1 - P(y_{ij} = 0)$, and $y_i = \sum_j y_{ij}/n_i$ is the sample proportion. For independent trials, $n_i y_i \sim \text{bin}(n_i, \pi_i)$, with $\text{var}(y_i) = v(\pi_i) = \pi_i(1 - \pi_i)/n_i$.

Instead of independent trials, suppose that y_{i1} is random but then $y_{ij} = y_{i1}$ for $j = 2, \ldots, n_i$. For instance, in an election, perhaps in each household the head of the household decides how to vote, and then everyone else in the household votes the same way. Then the sample proportion in household i voting for a particular candidate has

$$P(y_i = 1) = \pi_i, \quad P(y_i = 0) = 1 - \pi_i.$$

That is, y_i can take only its extreme possible values. Then $\text{var}(y_i) = \pi_i(1 - \pi_i) > \pi_i(1 - \pi_i)/n_i$, and there is overdispersion relative to the binomial. By contrast, suppose that the observations occur sequentially, and

$$y_{ij} \mid y_{i1}, \ldots, y_{i,j-1} \quad \text{equals} \quad 1 - y_{i,j-1}.$$

That is, trial j for observation i necessarily has the opposite result of trial $j - 1$. Then when n_i is an even number, $P(y_i = 1/2) = 1$, so $\text{var}(y_i) = 0$ and there is underdispersion.

In practice, a more likely scenario than one trial being completely dependent on another one is exchangeability of trials, with a common correlation ρ between each pair of $\{y_{i1}, y_{i2}, \ldots, y_{in_i}\}$, as is often assumed in cluster sampling. When $\text{corr}(y_{is}, y_{it}) = \rho$ for $s \neq t$, then $\text{var}(y_{it}) = \pi_i(1 - \pi_i)$, $\text{cov}(y_{is}, y_{it}) = \rho\pi_i(1 - \pi_i)$, and

$$\text{var}(y_i) = \text{var}\left(\frac{\sum_{t=1}^{n_i} y_{it}}{n_i}\right) = \frac{1}{n_i^2}\left[\sum_{t=1}^{n_i} \text{var}(y_{it}) + 2\sum\sum_{s<t} \text{cov}(y_{is}, y_{it})\right]$$

$$= \frac{1}{n_i^2}\left[n_i\pi_i(1 - \pi_i) + n_i(n_i - 1)\rho\pi_i(1 - \pi_i)\right] = [1 + \rho(n_i - 1)]\frac{\pi_i(1 - \pi_i)}{n_i}.$$

The ordinary variance for a binomial sample proportion results when $\rho = 0$. Overdispersion occurs when $\rho > 0$.

The inflated binomial variance is not a special case of this variance function, unless all n_i are identical. When $n_i = 1$, it is not possible to have overdispersion or underdispersion, and this variance formula is still valid, unlike the inflated binomial variance.

8.2.2 QL with Variance Function for Correlated Bernoulli Trials

For binary count data, a quasi-likelihood approach can use a variance function motivated by the one just found with correlated Bernoulli trials,

$$v(\pi_i) = [1 + \rho(n_i - 1)]\pi_i(1 - \pi_i)/n_i$$

with $|\rho| \leq 1$. The estimates using it differ from ML estimates for an ordinary binomial model, because the multiple of the binomial variance does not drop out of the quasi-likelihood equations (8.1).

For this QL approach, Williams (1982) proposed an iterative routine for estimating $\boldsymbol{\beta}$ and the overdispersion parameter ρ. He let $\hat{\rho}$ be such that the resulting generalized Pearson X^2 statistic (4.17) equals the residual $df = (n - p)$ for the model. This requires an iterative two-step process of (1) solving the quasi-likelihood equations for $\boldsymbol{\beta}$ for a given $\hat{\rho}$, and then (2) using the updated $\hat{\boldsymbol{\beta}}$, solving for $\hat{\rho}$ in the equation that equates

$$X^2 = \sum_{i=1}^{n} \frac{(y_i - \hat{\pi}_i)^2}{[1 + \hat{\rho}(n_i - 1)]\hat{\pi}_i(1 - \hat{\pi}_i)/n_i} = n - p.$$

8.2.3 Models Using the Beta-Binomial Distribution

The *beta-binomial model* is a parametric mixture model that is an alternative to quasi-likelihood generalizations of binomial GLMs. As with other mixture models that assume a binomial distribution at a fixed parameter value, the marginal distribution permits more variation than the binomial. We will see that the variance function for the beta-binomial model has the same form as the one resulting from correlated Bernoulli trials.

The beta-binomial distribution results from a *beta distribution* mixture of binomials. Suppose that (1) given π, s has a binomial distribution, bin(n, π), and (2) π has a beta distribution. The beta pdf is

$$f(\pi; \alpha_1, \alpha_2) = \frac{\Gamma(\alpha_1 + \alpha_2)}{\Gamma(\alpha_1)\Gamma(\alpha_2)} \pi^{\alpha_1 - 1}(1 - \pi)^{\alpha_2 - 1}, \quad 0 \leq \pi \leq 1,$$

with parameters $\alpha_1 > 0$ and $\alpha_2 > 0$, and $\Gamma(\cdot)$ denotes the gamma function. The beta family provides a wide variety of pdf shapes over (0, 1), including uniform

($\alpha_1 = \alpha_2 = 1$), unimodal symmetric ($\alpha_1 = \alpha_2 > 1$), unimodal skewed left ($\alpha_1 > \alpha_2 > 1$) or skewed right ($\alpha_2 > \alpha_1 > 1$), and U-shaped ($\alpha_1 < 1, \alpha_2 < 1$). Let

$$\mu = \frac{\alpha_1}{\alpha_1 + \alpha_2}, \quad \theta = 1/(\alpha_1 + \alpha_2).$$

The beta distribution for π has mean and variance

$$E(\pi) = \mu, \quad \text{var}(\pi) = \mu(1-\mu)\theta/(1+\theta).$$

Marginally, averaging over the beta distribution for π, s has the *beta-binomial distribution*. Its probability mass function is

$$p(s; n, \mu, \theta) = \binom{n}{s} \frac{\left[\prod_{k=0}^{s-1}(\mu + k\theta)\right]\left[\prod_{k=0}^{n-s-1}(1-\mu+k\theta)\right]}{\prod_{k=0}^{n-1}(1+k\theta)}, \quad s = 0, 1, \ldots, n.$$

As $\theta \to 0$, $\text{var}(\pi) \to 0$, and the beta distribution for π converges to a degenerate distribution at μ. Then $\text{var}(s) \to n\mu(1-\mu)$, and the beta-binomial distribution converges to the bin(n, μ). But the beta-binomial can look[1] quite different from the binomial. For example, when $\mu = 1/2$, it is uniform over the integers 0 to n when $\theta = 1/2$ (i.e., when $\alpha_1 = \alpha_2 = 1$ and the beta distribution is uniform), and it is bimodal at 0 and n when $\theta > 1/2$. For the beta-binomial proportion $y = s/n$,

$$E(y) = \mu, \quad \text{var}(y) = [1 + (n-1)\theta/(1+\theta)]\mu(1-\mu)/n.$$

In fact, $\rho = \theta/(1+\theta)$ is the correlation between each pair of the individual Bernoulli random variables that sum to s. The variance function in the beta-binomial and in the QL approach of Section 8.2.2 also results merely from assuming that π has a distribution with $\text{var}(\pi) = \rho\mu(1-\mu)$.

Models using the beta-binomial distribution usually let θ be the same unknown constant for all observations. Models can use any link function for binary data, but the logit is most common. For observation i with n_i trials, assuming that $n_i y_i$ has a beta-binomial distribution with index n_i and parameters (μ_i, θ), the model links μ_i to explanatory variables by

$$\text{logit}(\mu_i) = \boldsymbol{x}_i\boldsymbol{\beta}, \quad i = 1, \ldots, n.$$

Model fitting can employ a variety of methods, including the Newton–Raphson method. See Note 8.4. The beta-binomial distribution is not in the exponential dispersion family, even for known θ. When the linear predictor is correct, the beta-binomial ML estimator $\hat{\boldsymbol{\beta}}$ is not consistent if the actual distribution is not beta-binomial. Quasi-likelihood methods have greater robustness (Liang and Hanfelt 1994).

[1] `distributome.org/V3/calc/BetaBinomialCalculator.html` displays shapes as a function of n, α_1, and α_2.

8.2.4 Example: Modeling Overdispersion in a Teratology Study

Teratology is the study of abnormalities of physiological development. Some teratology experiments investigate effects of dietary regimens or chemical agents on the fetal development of rats in a laboratory setting. Table 8.1 shows results from one such study. Female rats on iron-deficient diets were assigned to four groups. Rats in group 1 were given placebo injections, and rats in other groups were given injections of an iron supplement. This was done on days 7 and 10 in group 2, on days 0 and 7 in group 3, and weekly in group 4. The 58 rats were made pregnant, sacrificed after 3 weeks, and then the total number of dead fetuses was counted in each litter, as was the mother's hemoglobin level. The overall sample proportions of deaths for the four groups were = 0.758 (placebo), 0.102, 0.034, and 0.048. Because of unmeasured covariates and genetic variability, the probability of death may vary among litters within a particular treatment group and hemoglobin level.

Table 8.1 Response Proportions $y = s/n$ Dead in Teratology Study for n Fetuses in Rat Litter in Group i with Mother's Hemoglobin Level h

i	h	y	i	h	y	i	h	y	i	h	y
1	4.1	1/10	2	8.6	1/10	3	11.2	0/8	4	16.6	0/3
1	3.2	4/11	2	11.1	1/3	3	11.5	1/11	4	14.5	0/13
1	4.7	9/12	2	7.2	1/13	3	12.6	0/14	4	15.4	2/9
...											

Source: From Moore and Tsiatis (1991), reproduced with permission of John Wiley & Sons, Inc. Complete data for 58 rats are in the file `Rats.dat` at text website.

Let y_{ij} denote the dead proportion of the n_{ij} fetuses in litter j in treatment group i. Let π_{ij} denote the probability of death for a fetus in that litter. Moore and Tsiatis modeled π_{ij} using only the hemoglobin level or only group indicators as the explanatory variable. Here, we will use hemoglobin level and whether the litter is in the placebo group, to judge whether the death rate differs between the placebo group and the other groups after adjusting for the hemoglobin level.

Let z_i denote an indicator for the placebo group ($z_1 = 1$, $z_2 = z_3 = z_4 = 0$) and let h_{ij} denote the hemoglobin level for litter j in group i. We present four fits for the model

$$\text{logit}(\pi_{ij}) = \beta_0 + \beta_1 z_i + \beta_2 h_{ij}.$$

We first treat $n_{ij}y_{ij}$ as a bin(n_{ij}, π_{ij}) variate and find ML estimates.

```
------------------------------------------------------------------------------
> Rats # data in file Rats.dat at www.stat.ufl.edu/~aa/glm/data.html
  litter  group    h   n  s  # s dead of n fetuses with hemoglobin h
1      1     1   4.1  10  1
2      2     1   3.2  11  4
...
```

```
58    58     4  12.4  17  0
> attach(Rats)
> placebo <- ifelse(group==1, 1, 0)
> fit.ML <- glm(s/n ~ placebo + h, weights=n, data=Rats, family=binomial)
> summary(fit.ML) # ML, assuming independent binomials
Coefficients:
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -0.6239      0.7900   -0.790    0.4296
placebo       2.6509      0.4824    5.495   3.9e-08
h            -0.1871      0.0743   -2.519    0.0118
---
> logLik(fit.ML)
'log Lik.' -121.0219
```

Summing the squared Pearson residuals for the $n = 58$ litters, we obtain $X^2 = 159.815$ with $df = 58 - 3 = 55$, considerable evidence of overdispersion. With the QL inflated-variance approach, $\hat{\phi} = 159.815/55 = 2.906$, so standard errors multiply by $\hat{\phi}^{1/2} = 1.70$. Even with this adjustment for overdispersion and for the hemoglobin level, strong evidence remains that the probability of death is substantially higher for the placebo group.

```
> summary(glm(s/n ~ placebo + h, weights=n, data=Rats,
+             family=quasi(link = "logit", variance="mu(1-mu)")))
Coefficients: # QL inflated-variance approach
            Estimate  Std. Error  t value  Pr(>|t|)
(Intercept)  -0.6239      1.3466   -0.463   0.64495
placebo       2.6509      0.8223    3.224   0.00213
h            -0.1871      0.1266   -1.478   0.14514
---
(Dispersion parameter for quasi family taken to be 2.906)
```

Because of unmeasured covariates, it is natural to permit the probability of death to vary among litters having particular values of z_i and h_{ij}. For the beta-binomial logistic model, $\hat{\rho} = \hat{\theta}/(1 + \hat{\theta}) = 0.237$, so the fit treats

$$\text{var}(y_{ij}) = [1 + 0.237(n_{ij} - 1)]\mu_{ij}(1 - \mu_{ij})/n_{ij}.$$

This corresponds roughly to a doubling of the variance relative to the binomial with a litter size of 5 and a tripling with $n_{ij} = 9$. The log-likelihood shows great improvement over the ordinary binomial GLM.

```
> library(VGAM) # beta-binomial model is available in VGAM package
> fit.bb <- vglm(cbind(s, n-s) ~ placebo + h,
+           betabinomial(zero=2,irho=.2), data=Rats)
 # two parameters, mu and rho; zero=2 specifies 0 covariates for 2nd
```

```
 # parameter (rho); irho is initial guess for rho in beta-bin variance
Coefficients:
               Estimate  Std. Error  z value
(Intercept):1   -0.5009      1.1907   -0.4207
(Intercept):2   -1.1676      0.3251   -3.5918
placebo          2.5601      0.7642    3.3501
h               -0.1546      0.1085   -1.4243
Names of linear predictors: logit(mu), logit(rho)
Log-likelihood: -93.1849
> logit(-1.1676, inverse=T) # Inverse logit is a function in VGAM
[1] 0.2373                  # Estimate of rho in beta-binomial variance
-------------------------------------------------------------------------
```

For the QL approach using beta-binomial-type variance, $\hat{\rho} = 0.1985$. It corresponds to using $v(\pi_{ij}) = [1 + 0.1985(n_{ij} - 1)]\mu_{ij}(1 - \mu_{ij})/n_{ij}$.

```
-------------------------------------------------------------------------
> library(aod)
 # betabin fn. fits beta-bin., quasibin fn. fits QL with beta-bin. var.
>  quasibin(cbind(s, n-s) ~ placebo + h, data=Rats)
            Estimate  Std. Error  z value  Pr(>|z|)
(Intercept)  -0.7237      1.3785  -0.5250    0.5996
placebo       2.7573      0.8522   3.2355    0.0012
h            -0.1758      0.1284  -1.3692    0.1709
Overdispersion parameter: phi 0.1985  # estimate of rho in our notation
-------------------------------------------------------------------------
```

Table 8.2 summarizes results for the four analyses. The QL approaches and the beta-binomial model have similar standard errors, quite different from those for the ordinary binomial ML estimates.

Liang and McCullagh (1993) showed several analyses using the inflated variance and beta-binomial-type variance. A plot of the standardized residuals for the ordinary binomial model against the indices $\{n_i\}$ can provide insight about which is more appropriate. When the residuals show an increasing trend in their spread as n_i increases, the beta-binomial-type variance function may be more appropriate.

Table 8.2 Parameter Estimates (with Standard Errors in Parentheses) for Four Fits of a Model with Logit Link to Table 8.1

	Type of Logistic Model Fit[a]			
Parameter	Binomial ML	QL(1)	QL(2)	Beta-Binomial ML
Intercept	0.62 (0.79)	0.62 (1.35)	0.72 (1.38)	0.50 (1.19)
Placebo	2.65 (0.48)	2.65 (0.82)	2.76 (0.85)	2.56 (0.76)
Hemoglobin	−0.19 (0.07)	−0.19 (0.13)	−0.18 (0.13)	−0.15 (0.11)
Overdispersion	None	$\hat{\phi} = 2.906$	$\hat{\rho} = 0.1985$	$\frac{\hat{\theta}}{1+\hat{\theta}} = 0.237$

[a]Quasi-likelihood (QL) has (1) inflated binomial variance, (2) beta-binomial-type variance.

Figure 8.1 plots these for the teratology data. The apparent increase in their variability as litter size increases suggests that the beta-binomial variance function is plausible.

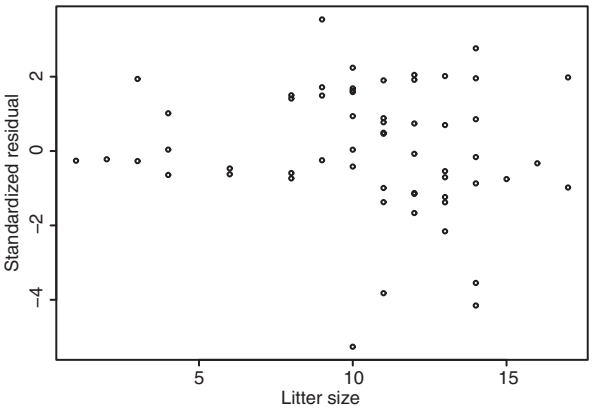

Figure 8.1 Standardized residuals and litter size for binomial logistic model fitted to Table 8.1.

8.3 QUASI-LIKELIHOOD AND MODEL MISSPECIFICATION

For the model $\eta_i = g(\mu_i) = \boldsymbol{x}_i\boldsymbol{\beta}$, the quasi-likelihood parameter estimates $\hat{\boldsymbol{\beta}}$ are the solutions of quasi-score equations

$$\boldsymbol{u}(\boldsymbol{\beta}) = \sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}} \frac{(y_i - \mu_i)}{v(\mu_i)} = \mathbf{0}. \tag{8.2}$$

These equations are the same as the likelihood equations (8.1) when we substitute

$$\frac{\partial \mu_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i}\frac{\partial \eta_i}{\partial \beta_j} = \frac{\partial \mu_i}{\partial \eta_i} x_{ij},$$

but it is more convenient to use Equation (8.2) for the covariance matrix expressions introduced in this section. With the assumption that $\{y_i\}$ has distribution in the exponential dispersion family, these equations are likelihood equations, in which case $v(\mu_i)$ characterizes the distribution. Our interest here, however, is still in variance functions that take us outside that family.

8.3.1 Estimating Equations and Properties of Quasi-Likelihood

The quasi-score equations (8.2) that determine $\hat{\boldsymbol{\beta}}$ in the QL method are called *estimating equations*. The quasi-score function $u_j(\boldsymbol{\beta})$ in Equation (8.2) is called an *unbiased estimating function*; this term refers to any function $h(\boldsymbol{y}; \boldsymbol{\beta})$ of $\boldsymbol{y}$ and $\boldsymbol{\beta}$ such that

$E[h(\boldsymbol{y};\boldsymbol{\beta})] = 0$ for all $\boldsymbol{\beta}$. For an unbiased estimating function, the estimating equation $h(\boldsymbol{y};\hat{\boldsymbol{\beta}}) = 0$ determines an estimator $\hat{\boldsymbol{\beta}}$ of $\boldsymbol{\beta}$. The maximum likelihood estimator is but one example of an estimator that results from estimating equations (namely, likelihood equations) with an unbiased estimating function.

The QL method treats the quasi-score function $\boldsymbol{u}(\boldsymbol{\beta})$ as the derivative of a *quasi-log-likelihood* function. Although this need not be a proper log-likelihood function, the QL estimators that maximize it have properties similar to those of ML estimators: under the correct specification of μ_i and $v(\mu_i)$, they are asymptotically efficient among estimators that are locally linear in $\{y_i\}$. This result generalizes the Gauss–Markov theorem, although in an asymptotic rather than exact manner. The QL estimators $\hat{\boldsymbol{\beta}}$ are asymptotically normal with a model-based covariance matrix approximated by

$$\boldsymbol{V} = \left[\sum_{i=1}^{n} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\mathrm{T}} [v(\mu_i)]^{-1} \left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)\right]^{-1}. \tag{8.3}$$

This is equivalent to the formula for the large-sample covariance matrix of the ML estimator in a GLM, namely $(\boldsymbol{X}^{\mathrm{T}}\boldsymbol{W}\boldsymbol{X})^{-1}$ with $w_i = (\partial\mu_i/\partial\eta_i)^2/\mathrm{var}(y_i)$.

A key result is that the QL estimator $\hat{\boldsymbol{\beta}}$, like ML estimators for GLMs, is consistent for $\boldsymbol{\beta}$ even if $v(\mu_i)$ is misspecified, as long as the specification is correct for the link function and linear predictor. That is, assuming that the model form $g(\mu_i) = \sum_j \beta_j x_{ij}$ is correct, the consistency of $\hat{\boldsymbol{\beta}}$ holds even if the true variance function is not $v(\mu_i)$. Here is a heuristic explanation: when truly $\mu_i = g^{-1}(\sum_j \beta_j x_{ij})$, then from Equation (8.2), $E[u_j(\boldsymbol{\beta})] = 0$ for all j. Also from Equation (8.2), $\boldsymbol{u}(\boldsymbol{\beta})/n$ is a vector of sample means. By a law of large numbers, it converges in probability to its expected value of $\mathbf{0}$. But the solution $\hat{\boldsymbol{\beta}}$ of the quasi-likelihood equations is the value of $\boldsymbol{\beta}$ for which the sample mean is exactly equal to $\mathbf{0}$. Since $\hat{\boldsymbol{\beta}}$ is a continuous function of these sample means, it converges to $\boldsymbol{\beta}$ by the continuous mapping theorem.

8.3.2 Sandwich Covariance Adjustment for Variance Misspecification

In practice, when we assume a particular variance function $v(\mu_i)$, it is likely that the true $\mathrm{var}(y_i) \neq v(\mu_i)$. Then the asymptotic covariance matrix of the QL estimator $\hat{\boldsymbol{\beta}}$ is not $\boldsymbol{V}$ as given in Equation (8.3). To find the actual $\mathrm{var}(\hat{\boldsymbol{\beta}})$, we use a Taylor-series expansion for the quasi-score function in Equation (8.2),

$$\boldsymbol{u}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{u}(\boldsymbol{\beta}) + \frac{\partial \boldsymbol{u}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$

Since $\boldsymbol{u}(\hat{\boldsymbol{\beta}}) = \mathbf{0}$,

$$(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \approx -\left(\frac{\partial \boldsymbol{u}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^{-1} \boldsymbol{u}(\boldsymbol{\beta}),$$

$$\text{so that} \quad \mathrm{var}(\hat{\boldsymbol{\beta}}) \approx \left(\frac{\partial \boldsymbol{u}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^{-1} \mathrm{var}[\boldsymbol{u}(\boldsymbol{\beta})] \left(\frac{\partial \boldsymbol{u}(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}}\right)^{-1}.$$

But $[\partial \boldsymbol{u}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}]$ is the Hessian matrix for the quasi-log-likelihood. So $-[\partial \boldsymbol{u}(\boldsymbol{\beta})/\partial \boldsymbol{\beta}]^{-1}$ is the analog of an inverse observed information matrix for the specified model and approximates the model-based covariance matrix $\boldsymbol{V}$. Also,

$$\text{var}[\boldsymbol{u}(\boldsymbol{\beta})] = \text{var}\left[\sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\text{T}}\frac{(y_i-\mu_i)}{v(\mu_i)}\right] = \sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\text{T}}\frac{\text{var}(y_i)}{[v(\mu_i)]^2}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right).$$

In summary, the actual asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ is

$$\text{var}(\hat{\boldsymbol{\beta}}) \approx \boldsymbol{V}\left[\sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)^{\text{T}}\frac{\text{var}(y_i)}{[v(\mu_i)]^2}\left(\frac{\partial \mu_i}{\partial \boldsymbol{\beta}}\right)\right]\boldsymbol{V}. \tag{8.4}$$

This matrix simplifies to $\boldsymbol{V}$ if $\text{var}(y_i) = v(\mu_i)$.

In practice, the true variance function, $\text{var}(y_i)$, is unknown. With large n we can estimate the asymptotic covariance matrix (8.4) by a sample analog, replacing μ_i by $\hat{\mu}_i$ and $\text{var}(y_i)$ by $(y_i - \hat{\mu}_i)^2$. This estimator of the covariance matrix is valid regardless of whether the model-based variance specification $v(\mu_i)$ is correct, in the sense that n times this estimator converges in probability to the asymptotic covariance matrix of $\sqrt{n}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})$. It is called a *sandwich estimator*, because the empirical evidence is sandwiched between the model-based covariance matrices.

The purpose of the sandwich estimator is to use the data's empirical evidence about variation to adjust the standard errors, in case the true variance function differs substantially from the variance function assumed in the modeling. Inference then uses the asymptotic normality of the estimator $\hat{\boldsymbol{\beta}}$ together with the sandwich-estimated covariance matrix.

8.3.3 Example: Robust Adjustment of Naïve Standard Errors

To illustrate, suppose $\{y_i\}$ are counts and we assume that $v(\mu_i) = \mu_i$, as in Poisson GLMs, but actually $\text{var}(y_i) = \mu_i^2$. Consider the null model, $\mu_i = \beta$, $i = 1, \ldots, n$. Since $\partial \mu_i/\partial \beta = 1$, from Equation (8.2),

$$u(\beta) = \sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \beta}\right)v(\mu_i)^{-1}(y_i - \mu_i) = \sum_{i=1}^{n}\frac{(y_i-\mu_i)}{\mu_i} = \sum_{i=1}^{n}\frac{(y_i-\beta)}{\beta}.$$

Setting this equal to 0 and solving, $\hat{\beta} = (\sum_i y_i)/n = \bar{y}$. The model-based variance (8.3) simplifies to

$$V = \left[\sum_{i=1}^{n}\left(\frac{\partial \mu_i}{\partial \beta}\right)[v(\mu_i)]^{-1}\left(\frac{\partial \mu_i}{\partial \beta}\right)\right]^{-1} = \left[\sum_{i=1}^{n}\mu_i^{-1}\right]^{-1} = \frac{\beta}{n}.$$

If we truly believe that $v(\mu_i) = \mu_i$, a sensible estimate of the variance of $\hat{\beta} = \bar{y}$ is $\widehat{V} = \hat{\beta}/n = \bar{y}/n$.

The actual asymptotic variance (8.4) of $\hat{\beta}$, which incorporates the true variance function, $\text{var}(y_i) = \mu_i^2$, is

$$V\left[\sum_{i=1}^{n}\left(\frac{\partial\mu_i}{\partial\beta}\right)\frac{\text{var}(y_i)}{[v(\mu_i)]^2}\left(\frac{\partial\mu_i}{\partial\beta}\right)\right]V = \frac{\beta}{n}\left[\sum_{i=1}^{n}\frac{\mu_i^2}{(\mu_i)^2}\right]\frac{\beta}{n} = \frac{\beta^2}{n}.$$

This is considerably different from the naive model-based variance when β is not close to 1. In practice, not knowing the true variance function, we obtain a robust estimator of this actual asymptotic variance by replacing $\text{var}(y_i)$ in Equation (8.4) by $(y_i - \bar{y})^2$. Then the sandwich estimator simplifies (using $\mu_i = \beta$) to $\sum_i (y_i - \bar{y})^2/n^2$, which is a sensible estimator of $\text{var}(\bar{y})$ regardless of the model. Using this estimator instead of $\widehat{V} = \bar{y}/n$ protects against an incorrect choice of variance function.

In summary, even with an incorrect specification of the variance function, we can consistently estimate $\boldsymbol{\beta}$. We can also consistently estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\beta}}$ by the sandwich estimator of Equation (8.4). However, we lose some efficiency in estimating $\boldsymbol{\beta}$ when the chosen variance function $v(\mu_i)$ is wildly inaccurate. Also, n needs to be large for the sample sandwich estimator of Equation (8.4) to work well; otherwise, the empirically based standard errors tend to underestimate the true ones (Kauermann and Carroll 2001). If the assumed variance function is only slightly wrong, the model-based standard errors are more reliable. Finally, in practice, keep in mind that, just as the chosen variance function only approximates the true one, the specification for the mean is also only approximate.

8.3.4 Example: Horseshoe Crabs Revisited

For the horseshoe crabs data, in Section 8.1.3 we used the variance-inflation approach to adjust standard errors for overdispersion from using a Poisson loglinear model to predict male satellite counts using the female crab weights. We obtain similar results from the empirical sandwich adjustment for the standard error. In the next chapter (Section 9.6), we will use a generalization of the sandwich covariance matrix as a way of dealing with correlated observations for a multivariate response. The method solves *generalized estimating equations* (GEE). Software for GEE can also perform the analysis described in this section, using empirical variability to find robust standard errors that adjust for variance misspecification. In the following R printout, the naive standard error comes from the ordinary ML fit of the Poisson model.

```
------------------------------------------------------------------------
> library(gee) # sandwich adjustment for generalized estimating equa's
> obs <- c(1:173) # labeling of observations needed for GEE method
> summary(gee(y ~ weight, id=obs, family=poisson, scale.fix=TRUE))
             Estimate  Naive S.E.   Naive z  Robust S.E.  Robust z
(Intercept)   -0.4284      0.1789   -2.3942       0.3083    -1.3896
weight         0.5893      0.0650    9.0638       0.1103     5.3418
------------------------------------------------------------------------
```

CHAPTER NOTES

Section 8.1: Variance Inflation for Overdispersed Poisson and Binomial GLMs

8.1 **QL approach**: Wedderburn (1974) proposed the quasi-likelihood approach with the Pearson moment adjustment for estimating ϕ in the variance inflation approach. Finney (1947) had proposed this for binomial overdispersion. More generally, the QL approach can model ϕ in terms of explanatory variables, thus simultaneously modeling the mean and the variability (McCullagh and Nelder 1989, Chapter 10; Lee et al. 2006, Chapter 3). McCullagh (1983) and Godambe and Heyde (1987) analyzed properties of QL estimators.

8.2 **Poisson overdispersion**: For other QL approaches for Poisson overdispersion, see Cameron and Trivedi (2013, Section 3.2) and Hinde and Demétrio (1998).

Section 8.2: Beta-Binomial Models and Quasi-Likelihood Alternatives

8.3 **Correlated trials**: For extensions of the binomial that permit correlated Bernoulli trials, see Altham (1978) and Ochi and Prentice (1984).

8.4 **Beta-binomial**: Skellam (1948) introduced the beta-binomial distribution. More general beta-binomial models let θ depend on covariates, such as by allowing a different θ for each group of interest (Prentice 1986). For other modeling using this distribution or related QL approaches, see Capanu and Presnell (2008), Crowder (1978), Hinde and Demétrio (1998), Lee et al. (2006), Liang and Hanfelt (1994), Liang and McCullagh (1993), Lindsey and Altham (1998), and Williams (1982).

8.5 **Dirichlet-multinomial**: The beta-binomial generalizes to a *Dirichlet-multinomial* (Mosimann 1962): conditional on the probabilities, the distribution is multinomial, and the probabilities themselves have a Dirichlet distribution. For modeling, see Guimarães and Lindrooth (2007).

Section 8.3: Quasi-Likelihood and Model Misspecification

8.6 **Estimating equations**: Extending Fisher's work, in 1960 Godambe showed that of solutions for unbiased estimating functions, ML estimators are optimal. See Godambe and Heyde (1987), who reviewed the theory of QL estimating equations, and McCullagh (1983).

8.7 **Sandwich**: The sandwich covariance matrix and related results about adjustments for model misspecification using moment-based models evolved from literature in statistics (Huber 1967; Fahrmeir 1990), econometrics (Gourieroux et al. 1984; Hansen 1982; White 1980, 1982), and biostatistics (Liang and Zeger 1986). Cameron and Trivedi (2013, Chapter 2) and Royall (1986) presented motivation and examples.

EXERCISES

8.1 Does the inflated-variance QL approach make sense as a way to generalize the ordinary normal linear model with $v(\mu_i) = \sigma^2$? Why or why not?

8.2 Using $E(y) = E[E(y|x)]$ and $\text{var}(y) = E[\text{var}(y|x)] + \text{var}[E(y|x)]$, derive the mean and variance of the beta-binomial distribution.

8.3 Let y_1 and y_2 be independent negative binomial variates with common dispersion parameter γ.

a. Show that $y_1 + y_2$ is negative binomial with dispersion parameter $\gamma/2$.

b. Conditional on $y_1 + y_2$, show that y_1 has a beta-binomial distribution.

c. State the multicategory extension of (**b**) that yields a Dirichlet-multinomial distribution. Explain the analogy with the Poisson-multinomial result in Section 7.2.1.

8.4 Altham (1978) introduced the discrete distribution

$$f(x;\pi,\theta) = c(\pi,\theta)\binom{n}{x}\pi^x(1-\pi)^{n-x}\theta^{x(n-x)}, \quad x = 0, 1, \ldots, n,$$

where $c(\pi,\theta)$ is a normalizing constant. Show that this is in the two-parameter exponential family and that the binomial occurs when $\theta = 1$. (Altham noted that overdispersion occurs when $\theta < 1$. Lindsey and Altham (1998) used this as the basis of an alternative model to the beta-binomial.)

8.5 Sometimes sample proportions are continuous rather than of the binomial form (number of successes)/(number of trials). Each observation is any real number between 0 and 1, such as the proportion of a tooth surface that is covered with plaque. For independent responses $\{y_i\}$, Bartlett (1937) modeled $\text{logit}(y_i) \sim N(\boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)$. Then y_i itself has a *logit-normal distribution*.

a. Expressing a $N(\boldsymbol{x}_i\boldsymbol{\beta}, \sigma^2)$ variate as $\boldsymbol{x}_i\boldsymbol{\beta} + \sigma z$, where z is a standard normal variate, show that $y_i = \exp(\boldsymbol{x}_i\boldsymbol{\beta} + \sigma z)/[1 + \exp(\boldsymbol{x}_i\boldsymbol{\beta} + \sigma z)]$ and for small σ,

$$y_i = \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}} + \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}\frac{1}{1+e^{\boldsymbol{x}_i\boldsymbol{\beta}}}\sigma z + \frac{e^{\boldsymbol{x}_i\boldsymbol{\beta}}(1-e^{\boldsymbol{x}_i\boldsymbol{\beta}})}{2(1+e^{\boldsymbol{x}_i\boldsymbol{\beta}})^3}\sigma^2 z^2 + \cdots.$$

b. Letting $\mu_i = e^{\boldsymbol{x}_i\boldsymbol{\beta}}/(1+e^{\boldsymbol{x}_i\boldsymbol{\beta}})$, when σ is close to 0 show that

$$E(y_i) \approx \mu_i, \quad \text{var}(y_i) \approx [\mu_i(1-\mu_i)]^2\sigma^2.$$

c. The approximate moments for the logit-normal motivate a QL approach with $v(\mu_i) = \phi[\mu_i(1-\mu_i)]^2$ for unknown ϕ. Explain why this approach provides similar results as fitting an ordinary linear model to the sample logits, assuming constant variance. (The QL approach has the advantage of not requiring adjustment of 0 or 1 observations, for which sample logits do not exist. Papke and Wooldridge (1996) proposed an alternative QL approach using a sandwich covariance adjustment.)

d. Wedderburn (1974) used QL to model the proportion of a leaf showing a type of blotch. Envision an approximation of binomial form based on cutting each leaf into a very large number of tiny regions of the same size

and observing for each region whether it is covered with blotch. Explain why this suggests using $v(\mu_i) = \phi\mu_i(1 - \mu_i)$. What violation of the binomial assumptions might make this questionable? (Recall that the parametric family of beta distributions has variance function of this form.)

8.6 Motivation for the quasi-score equations (8.2): suppose we replace $v(\mu_i)$ by known variance v_i. Show that the equations result from the weighted least squares approach of minimizing $\sum_i[(y_i - \mu_i)^2/v_i]$.

8.7 Before R. A. Fisher introduced the method of maximum likelihood in 1922, Karl Pearson had proposed the *method of moments* as a general-purpose method for statistical estimation[2]. Explain how this method can be formulated as having estimating equations with an unbiased estimating function.

8.8 Ordinary linear models assume that $v(\mu_i) = \sigma^2$ is constant. Suppose instead that actually $\text{var}(y_i) = \mu_i$. Using the QL approach for the null model $\mu_i = \beta$, $i = 1, \ldots, n$, show that $u(\beta) = (1/\sigma^2)\sum_i(y_i - \beta)$, so $\hat{\beta} = \bar{y}$ and $V = \sigma^2/n$. Find the model-based estimate of $\text{var}(\hat{\beta})$, the actual variance, and the robust estimate of that variance that adjusts for misspecification of the variance.

8.9 Suppose we assume $v(\mu_i) = \mu_i$ but actually $\text{var}(y_i) = \sigma^2$. For the null model $\mu_i = \beta$, find the model-based $\text{var}(\hat{\beta})$, the actual $\text{var}(\hat{\beta})$, and the robust estimate of that variance.

8.10 Suppose we assume $v(\mu_i) = \mu_i$ but actually $\text{var}(y_i) = v(\mu_i)$ for some unspecified function v. For the null model $\mu_i = \beta$, find the model-based $\text{var}(\hat{\beta})$, the actual $\text{var}(\hat{\beta})$, and the robust estimate of that variance.

8.11 Consider the null model $\mu_i = \beta$ when the observations are independent counts. Of the Poisson-model-based and robust estimators of the variance of $\hat{\beta} = \bar{y}$ presented in Section 8.3.3, which would you expect to be better (**a**) if the Poisson model truly holds, (**b**) if there is severe overdispersion? Explain your reasoning.

8.12 Let y_{ij} denote the response to a question about belief in life after death (1 = yes, 0 = no) for person j in household i, $j = 1, \ldots, n_i$, $i = 1, \ldots, n$. In modeling $P(y_{ij} = 1)$ with explanatory variables, describe a scenario in which you would expect binomial overdispersion. Specify your preferred method for dealing with it, presenting your reasoning for that choice.

8.13 Use QL methods to construct a model for the horseshoe crab satellite counts, using weight, color, and spine condition as explanatory variables. Compare results with those obtained with zero-inflated GLMs in Section 7.5.

[2]J. Aldrich in *Statistical Science* (**12**: 162–176, 1997) gave a historical overview.

8.14 Use QL methods to analyze Table 7.5 on counts of homicide victims. Interpret, and compare results with Poisson and negative binomial GLMs.

8.15 Refer to Exercise 7.35 on the frequency of sexual intercourse. Use QL methods to obtain a confidence interval for the (**a**) difference, (**b**) ratio of means for males and females.

8.16 For the teratology study analyzed in Section 8.2.4, analyze the data using only the group indicators as explanatory variables (i.e., ignoring hemoglobin). Interpret results. Is it sufficient to use the simpler model having only the placebo indicator for the explanatory variable?

8.17 Table 8.3 shows the three-point shooting, by game, of Ray Allen of the Boston Celtics during the 2010 NBA (basketball) playoffs (e.g., he made 0 of 4 shots in game 1). Commentators remarked that his shooting varied dramatically from game to game. In game i, suppose that $n_i y_i$ = number of three-point shots made out of n_i attempts is a bin(n_i, π_i) variate and the $\{y_i\}$ are independent.

Table 8.3 Data for Exercise 8.17 on Three-Point Shooting in Basketball

Game	y_i	Game	y_i	Game	y_i	Game	y_i	Game	y_i
1	0/4	6	2/7	11	0/5	16	1/3	21	0/4
2	7/9	7	3/7	12	2/5	17	3/7	22	0/4
3	4/11	8	0/1	13	0/5	18	0/2	23	2/5
4	3/6	9	1/8	14	2/4	19	8/11	24	2/7
5	5/6	10	6/9	15	5/7	20	0/8		

Source: http://boston.stats.com/nba. Data at file `Basketball.dat` at text website.

a. Fit the model, $\pi_i = \beta_0$. Find and interpret $\hat{\beta}_0$ and its standard error.

b. Describe a factor that could cause overdispersion. Adjust the standard error for overdispersion. Using the original *SE* and its correction, find and compare 95% confidence intervals for β_0. Interpret.