CATEGORICAL DATA ANALYSIS
ASSIGNMENT IV
Solution Set

4.1

(a) $\hat{\pi}_i$ = proportion voting for Buchanan in 2000
x_i = proportion voting for Perot in 1996

We have $\hat{\pi}_i = -.0003 + .0304 x_i$. Since the intercept of this fitted model is small, the model implies $\hat{\pi}_i \approx .0304 x_i$. Thus, P% = 3.04%.

(b) Yes. The fitted model estimates π_i to be $\hat{\pi}_i = -.0003 + .0304(.0774) = .0021$. The actual value of π_i is nearly 4 times as large as $\hat{\pi}_i$!

(c) Yes. The fitted model predicts $\hat{\pi}_i$ to be

$$\hat{\pi}_i = \frac{\exp(-7.164 + 12.219(.0774))}{1 + \exp(-7.164 + 12.219(.0774))}$$

$$= .0020.$$

The estimated value of π_i from the logistic regression model is similar to that from the linear probability model.

The results of (b) and (c) imply that the 2000 voting patterns in Palm Beach County were indeed unusual.

4.2

(a) Interpretation of $\hat{\beta}$: As we move from one decade to the next, the probability of a starting pitcher pitching a complete game decreases by roughly $\hat{\beta} = -.0694$.

Interpretation of $\hat{\alpha}$: We might interpret $\hat{\alpha} = .7578$ as approximating the probability of a starting pitcher pitching a complete game in the decade 1890 - 1899. (Note: This interpretation is based on extrapolation outside the scope of the model, and is therefore questionable.)

(b) $\hat{\pi}(10) = .0638$
$\hat{\pi}(11) = -.0056$
$\hat{\pi}(12) = -.0750$

These predictions are not sensible, since $\hat{\pi}(11)$ and $\hat{\pi}(12)$ are negative!

(c) $\hat{\pi}(10) = .1190$
$\hat{\pi}(11) = .0897$
$\hat{\pi}(12) = .0671$

These predictions are much more sensible. However, the predictions are based on extrapolation outside the scope of the model, and only valid under the premise that the model is appropriate from 1990-2019.

4.7

(a) $\hat{\mu} = \exp(-.4284 + .5893(2.44))$
$= 2.7442$

(b) Interpretation of $\hat{\beta}$: For every 1 kg increase in the weight of a female crab, we expect the average number of satellites to increase by a multiplicative factor of $e^{.5893} = 1.8027$.

The Wald CI for β is constructed by using the reported values of $\hat{\beta}$ and the estimated ASE of $\hat{\beta}$:

$\hat{\beta} - 1.96SE = (.5893) - 1.96(.0650) = .4619,$
$\hat{\beta} + 1.96SE = (.5893) + 1.96(.0650) = .7167.$

(c) For the test of $H_0: \beta = 0$ versus $H_a: \beta \neq 0$, the p-value for the Wald test statistic is $< .0001$. Thus, there is strong evidence to support H_a; i.e., the hypothesis of dependence between X and Y.

(d) No. To test $H_0: \beta = 0$ versus $H_a: \beta \neq 0$ using the likelihood-ratio approach, we need two log likelihoods: one for the model $\log \mu = \alpha + \beta x$ and one for the model $\log \mu = \alpha$. The output only provides the output for the model $\log \mu = \alpha + \beta x$.

4.11

The SE for $\hat{\beta}$ in the present model (.048) is much larger than the SE for $\hat{\beta}$ in the §4.3.2 model (.020) because the former model accounts for overdispersion whereas the latter does not.

The SE for $\hat{\beta}$ in the present model is more appropriate. The estimate of the dispersion parameter $\hat{k}^{-1} = 1.106$ is significantly greater than zero (95% Wald CI extends from $1.106 - (1.96)(0.197) = 0.720$ to $1.106 + (1.96)(0.197) = 1.492$).
Thus, overdispersion appears to be present in the phenomenon being modeled.

4.18

$$f(y;k,\mu) = \frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)} \left(\frac{k}{\mu+k}\right)^k \left(1-\frac{k}{\mu+k}\right)^y$$

$$= \underbrace{\left(\frac{k}{\mu+k}\right)^k}_{a(\mu)} \underbrace{\frac{\Gamma(y+k)}{\Gamma(k)\Gamma(y+1)}}_{b(y)}$$

$$\exp\left[y \underbrace{\left\{\log\left(1-\frac{k}{\mu+k}\right)\right\}}_{Q(\mu)} \right]$$

4.29

(a) $\Phi^{-1}(\pi(x)) = \alpha + \beta x$

$\Phi^{-1}(0.5) = \alpha + \beta x$

$0 = \alpha + \beta x$

$-\alpha/\beta = x$

(b) $\pi(x) = \Phi(\alpha + \beta x)$

$$\frac{\partial \pi(x)}{\partial x} = \phi(\alpha + \beta x)\,(\beta)$$

$$\left.\frac{\partial \pi(x)}{\partial x}\right|_{x = -\alpha/\beta} = \phi(\alpha + \beta(-\alpha/\beta))\,(\beta)$$

$$= \beta\,\phi(0)$$

Logit Link:

For the logit link, we have

$$\phi(x) = \frac{\partial}{\partial x}\Phi(x)$$

$$= \frac{\partial}{\partial x}\left(\frac{e^x}{1+e^x}\right)$$

$$= \frac{(1+e^x)(e^x) - (e^x)(e^x)}{(1+e^x)^2}$$

$$= \frac{e^x}{(1+e^x)^2}$$

Thus,

$$\beta \phi(0) = \beta \frac{e^0}{(1+e^0)^2} = \beta \left(\frac{1}{2^2}\right) = \frac{\beta}{4}$$

Probit Link:

For the probit link, we have

$$\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$$

Thus,

$$\beta \phi(0) = \beta \left\{ \frac{1}{\sqrt{2\pi}} e^0 \right\} = \frac{\beta}{\sqrt{2\pi}}$$

(c) Assume $\beta > 0$. Let $X \sim N(-\alpha/\beta, (1/\beta)^2)$. The cdf of X is given by

$$F(x) = P(X \leq x) = P\left(\frac{X-(-\alpha/\beta)}{(1/\beta)} \leq \frac{x-(-\alpha/\beta)}{(1/\beta)} \right)$$

$$= P(Z \leq \alpha + \beta x) = \Phi(\alpha + \beta x).$$

Assume $\beta < 0$. Let $X \sim N(-\alpha/\beta, (-(1/\beta))^2)$. The cdf of X is given by

$$F(x) = P(X \leq x) = P\left(\frac{X-(-\alpha/\beta)}{-(1/\beta)} \leq \frac{x-(-\alpha/\beta)}{-(1/\beta)} \right)$$

$$= P(Z \leq -(\alpha+\beta x)) = \Phi(-(\alpha+\beta x)) = 1 - \Phi(\alpha + \beta x).$$

4.30

$$f(y;\mu) = \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{-\frac{(y-\mu)^2}{2\sigma^2}\right\}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} \exp\left\{\frac{-y^2+2\mu y-\mu^2}{2\sigma^2}\right\}$$

$$= \underbrace{\left\{\frac{1}{\sqrt{2\pi}\,\sigma} \exp(-\mu^2/2\sigma^2)\right\}}_{a(\mu)} \underbrace{\left\{\exp(-y^2/2\sigma^2)\right\}}_{b(y)} \left\{\exp\left[y\,\underbrace{(\mu/\sigma^2)}_{Q(\mu)}\right]\right\}$$

4.34

(a) If $\beta^{(0)}$ is close to $\hat{\beta}$, we can use the first-order Taylor series approximation for $L'(\hat{\beta})$, since the remainder (comprised of higher-order terms) will be negligible. We have

$$L'(\hat{\beta}) \approx L'(\beta^{(0)}) + (\hat{\beta}-\beta^{(0)})\,L''(\beta^{(0)}).$$

Since $L(\beta)$ is maximized at the point $\beta=\hat{\beta}$, we have $L'(\hat{\beta})=0$.

Thus,

$$0 \approx L'(\beta^{(0)}) + (\hat{\beta} - \beta^{(0)}) L''(\beta^{(0)}) \Rightarrow$$

$$-(L'(\beta^{(0)}) / L''(\beta^{(0)})) \approx (\hat{\beta} - \beta^{(0)}) \Rightarrow$$

$$\beta^{(0)} - (L'(\beta^{(0)}) / L''(\beta^{(0)})) \approx \hat{\beta}$$

The preceding suggests an updated estimate $\beta^{(1)}$ of β can be obtained from $\beta^{(0)}$ via the relation

$$\beta^{(1)} = \beta^{(0)} - (L'(\beta^{(0)}) / L''(\beta^{(0)})).$$

(b) In the relation

$$\beta^{(1)} = \beta^{(0)} - (L'(\beta^{(0)}) / L''(\beta^{(0)})),$$

replace $\beta^{(1)}$ with $\beta^{(t+1)}$ and $\beta^{(0)}$ with $\beta^{(t)}$. We then have

$$\beta^{(t+1)} = \beta^{(t)} - (L'(\beta^{(t)}) / L''(\beta^{(t)})).$$

The preceding provides a general updating equation for finding $\hat{\beta}$.

Supplementary Problems

S.1

$$E\left\{\frac{\partial^2 L}{\partial \gamma^2}\right\}$$

$$= \int \left\{\frac{\partial^2 L}{\partial \gamma^2}\right\} f(y;\gamma)\, dy$$

$$= \int \left\{\frac{\partial^2}{\partial \gamma^2} \log f(y;\gamma)\right\} f(y;\gamma)\, dy$$

$$= \int \left\{\frac{\partial}{\partial \gamma}\left(\frac{1}{f(y;\gamma)} \frac{\partial f(y;\gamma)}{\partial \gamma}\right)\right\} f(y;\gamma)\, dy$$

$$= \int \left\{\frac{1}{f(y;\gamma)} \frac{\partial^2 f(y;\gamma)}{\partial \gamma^2}\right\} f(y;\gamma)\, dy$$

$$+ \int \left\{-\frac{1}{(f(y;\gamma))^2} \left(\frac{\partial f(y;\gamma)}{\partial \gamma}\right)^2\right\} f(y;\gamma)\, dy$$

$$= \int \left\{\frac{\partial^2 f(y;\gamma)}{\partial \gamma^2}\right\} dy$$

$$- \int \left\{\frac{1}{f(y;\gamma)} \left(\frac{\partial f(y;\gamma)}{\partial \gamma}\right)\right\}^2 f(y;\gamma)\, dy$$

$$\overset{*}{=} \frac{\partial^2}{\partial \gamma^2} \int f(y;\gamma)\, dy$$

$$- \int \left\{\frac{\partial}{\partial \gamma} \log f(y;\gamma)\right\}^2 f(y;\gamma)\, dy$$

$$= \frac{\partial^2}{\partial \gamma^2}(1) - \int \left\{\frac{\partial L}{\partial \gamma}\right\}^2 f(y;\gamma)\, dy$$

$$= E\left[\left\{\frac{\partial L}{\partial \gamma}\right\}^2\right]$$

* Regularity conditions needed to ensure that integration and differentiation can be interchanged.

S.2

(a) $f(y;\alpha,\lambda)$

$$= \frac{1}{\Gamma(\alpha)\lambda^{\alpha}} y^{\alpha-1} \exp\left(-\frac{y}{\lambda}\right)$$

$$= \exp\left(-\frac{y}{\lambda} + \log y^{\alpha-1} - \log(\Gamma(\alpha)\lambda^{\alpha})\right)$$

$$= \exp\left(-\frac{y}{\lambda} + (\alpha-1)\log y - \log\Gamma(\alpha) - \alpha\log\lambda\right)$$

$$= \exp\left(\alpha\left(y\left(-\frac{1}{\alpha\lambda}\right) - \log(\alpha\lambda)\right) + \alpha\log(\alpha\lambda) - \alpha\log\lambda + (\alpha-1)\log y - \log\Gamma(\alpha)\right)$$

Let $\theta = -(1/(\alpha\lambda))$. The preceding can then be written as follows:

$$\exp\left(\alpha\left(y\theta - \left(-\log(-\theta)\right)\right) + \alpha\log\alpha + (\alpha-1)\log y - \log\Gamma(\alpha)\right)$$

$$= \exp\left\{ \frac{y\theta - b(\theta)}{a(\phi)} + c(y;\phi) \right\},$$

where

$$\theta = -\frac{1}{\alpha\lambda},$$

$$b(\theta) = -\log(-\theta),$$

$$\phi = \alpha, \quad a(\phi) = (1/\phi),$$

$$c(y;\phi) = \phi \log \phi + (\phi - 1)\log y - \log \Gamma(\phi).$$

From Lemma 4.2, we then have

$$\begin{aligned} E(Y) &= b'(\theta) \\ &= \frac{d}{d\theta}\left(-\log(-\theta)\right) \\ &= -\frac{1}{(-\theta)}(-1) \\ &= -\frac{1}{\theta} \\ &= \alpha\lambda \end{aligned}$$

$$\begin{aligned} \mathrm{var}(Y) &= a(\phi)\, b''(\theta) \\ &= a(\phi)\left(\frac{d}{d\theta} b'(\theta)\right) \end{aligned}$$

$$= \left(\frac{1}{\phi}\right)\left(\frac{d}{d\theta}\left(-\frac{1}{\theta}\right)\right)$$

$$= \left(\frac{1}{\phi}\right)\left(\frac{1}{\theta^2}\right)$$

$$= \left(\frac{1}{\alpha}\right)\left((\alpha\lambda)^2\right)$$

$$= \alpha\lambda^2$$

(b) When $\alpha = 1$, the density becomes

$$f(y;\lambda)$$

$$= \frac{1}{\Gamma(1)\,\lambda^{(1)}}\, y^{(1)-1} \exp\left(-\frac{y}{\lambda}\right)$$

$$= \left(\frac{1}{\lambda}\right) \exp\left(-\frac{y}{\lambda}\right)$$

In the formulation of the exponential dispersion family, we have

$$\theta = -\frac{1}{(1)\lambda} = -\frac{1}{\lambda},$$

$$b(\theta) = -\log(-\theta) \quad (\text{or } \log\lambda),$$

$$\phi = 1, \quad a(\phi) = 1,$$

$$c(y;\phi) = (1)\log(1) + ((1)-1)\log y$$
$$- \Gamma(1)$$
$$= 0$$

Additionally,

$$E(Y) = \alpha\lambda = (1)\lambda = \lambda,$$
$$\text{var}(Y) = \alpha\lambda^2 = (1)\lambda^2 = \lambda^2.$$

(c) The canonical link for the gamma distribution would be given by

$$g(\mu) = -(1/\mu),$$

since from part (a), $\mu = E(Y) = \alpha\lambda$, and $\theta = -(1/(\alpha\lambda))$.

(d) The canonical link for the exponential distribution would also be given by

$$g(\mu) = -(1/\mu),$$

since from part (b), $\mu = E(Y) = \lambda$, and $\theta = -(1/\lambda)$.

(e) For the gamma and exponential distribution, $\mu > 0$. Thus, $g(\mu) = -(1/\mu) < 0$.
In the GLM formulation, the systematic component $\eta = \underline{x}'\underline{\beta}$ can assume any value from $-\infty$ to $+\infty$, yet $g(\mu)$ is inherently negative. As a consequence, a fitted GLM based on the canonical link may lead to estimated values of μ which are outside the range of possible values of μ.

5.3

The average variance of the deviance residuals based on a GLM fit using N cases is given by

$$\frac{1}{N}\sum_{i=1}^{N} \mathrm{var}\left(\sqrt{d_i}\ \mathrm{sign}(y_i - \hat{\mu}_i)\right)$$

The large-sample mean of the deviance residuals is approximately zero. Thus, we have

$$\mathrm{var}\left(\sqrt{d_i}\ \mathrm{sign}(y_i - \hat{\mu}_i)\right)$$
$$\approx E\left[\left(\sqrt{d_i}\ \mathrm{sign}(y_i - \hat{\mu}_i)\right)^2\right]$$
$$= E(d_i)$$

The approximate large-sample variance therefore corresponds to

$$\frac{1}{N}\sum_{i=1}^{N} E(d_i) = \frac{1}{N} E\left\{\sum_{i=1}^{N} d_i\right\}$$

$$= \frac{1}{N} E\left\{ D(\underset{\sim}{y}; \hat{\underset{\sim}{\mu}})\right\},$$

where $D(\underset{\sim}{y}; \hat{\underset{\sim}{\mu}})$ is the deviance.

If the fitted GLM is properly specified, the large-sample distribution of $D(\underset{\sim}{y}; \hat{\underset{\sim}{\mu}})$ is approximately $\chi^2_{N-(k+1)}$.

Thus, $E\{D(\underset{\sim}{y}; \hat{\underset{\sim}{\mu}})\} \approx (N-(k+1))$, and the preceding average reduces to

$$\frac{N-(k+1)}{N}.$$

Note: For a GLM formulated based on a density in the one-parameter natural exponential family, $D_S(\underset{\sim}{y}; \hat{\underset{\sim}{\mu}}) \equiv D(\underset{\sim}{y}; \hat{\underset{\sim}{\mu}})$.

S.4

(a) $Q_i(\mu_i; y_i)$

$$= \int_{y_i}^{\mu_i} \left(\frac{y_i - t}{v(t)} \right) dt$$

$$= \int_{y_i}^{\mu_i} \left(\frac{y_i - t}{t} \right) dt \longleftarrow v(\mu_i) = \mu_i$$

$$= \int_{y_i}^{\mu_i} \left(\frac{y_i}{t} - 1 \right) dt$$

$$= (y_i \log t - t) \Big|_{y_i}^{\mu_i}$$

$$= (y_i \log \mu_i - \mu_i) - (y_i \log y_i - y_i)$$

$$= y_i \log (\mu_i / y_i) + (y_i - \mu_i)$$

Thus, for the quasi-likelihood, we have

$$Q(\underline{\mu}; \underline{y}) = \sum_{i=1}^{N} Q_i(\mu_i; y_i)$$

$$= \sum_{i=1}^{N} \left\{ y_i \log (\mu_i / y_i) + (y_i - \mu_i) \right\}$$

(b) $\underline{u}(\underline{\beta}) = \frac{\partial}{\partial \underline{\beta}} Q(\underline{\mu}; \underline{y})$

$$= \sum_{i=1}^{N} \frac{\partial}{\partial \underline{\beta}} Q_i(\mu_i; y_i)$$

Note that we have

$$\frac{\partial}{\partial \underline{\beta}} Q_i(\mu_i; y_i) = \frac{\partial Q_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \underline{\beta}}$$

$$= \frac{\partial Q_i}{\partial \mu_i} \left(\frac{1}{g'(\mu_i)} \right) \underline{x}_i$$

Here,

$$\frac{\partial Q_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i} \int_{y_i}^{\mu_i} \left(\frac{y_i - t}{t} \right) dt = \frac{y_i - \mu_i}{\mu_i},$$

$$g'(\mu_i) = \frac{\partial}{\partial \mu_i} (\log \mu_i) = \frac{1}{\mu_i}$$

Thus, for $\underline{u}(\underline{\beta})$, we obtain

$$\underline{u}(\underline{\beta}) = \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{\mu_i} \right) (\mu_i)(\underline{x}_i)$$

$$= \sum_{i=1}^{N} (y_i - \mu_i)\, \underline{x}_i \leftarrow \mu_i = \exp(\underline{x}_i' \underline{\beta})$$

$$= \sum_{i=1}^{N} (y_i - \exp(\underline{x}_i' \underline{\beta}))\, \underline{x}_i$$

(c) $Q_i(\mu_i; y_i)$

$$= \int_{y_i}^{\mu_i} \left(\frac{y_i - t}{v(t)} \right) dt$$

$$= \int_{y_i}^{\mu_i} \left(\frac{y_i - t}{t^2} \right) dt \longleftarrow v(\mu_i) = \mu_i^2$$

$$= \int_{y_i}^{\mu_i} \left(\frac{y_i}{t^2} - \frac{1}{t} \right) dt$$

$$= \left(-\frac{y_i}{t} - \log t \right) \Big|_{y_i}^{\mu_i}$$

$$= \left(-\frac{y_i}{\mu_i} - \log \mu_i \right) - \left(-1 - \log y_i \right)$$

$$= \log(y_i/\mu_i) - (y_i/\mu_i) + 1$$

Thus, for the quasi-likelihood, we have

$$Q(\underline{\mu}; \underline{y}) = \sum_{i=1}^{N} Q_i(\mu_i; y_i)$$

$$= \sum_{i=1}^{N} \left\{ \log(y_i/\mu_i) - (y_i/\mu_i) + 1 \right\}$$

(d) $\underline{u}(\underline{\beta}) = \frac{\partial}{\partial \underline{\beta}} Q(\underline{\mu}; \underline{y})$

$$= \sum_{i=1}^{N} \frac{\partial}{\partial \underline{\beta}} Q_i(\mu_i; y_i)$$

Again, we have

$$\frac{\partial}{\partial \underline{\beta}} Q_i(\mu_i; y_i) = \frac{\partial Q_i}{\partial \mu_i} \frac{\partial \mu_i}{\partial \eta_i} \frac{\partial \eta_i}{\partial \underline{\beta}}$$

$$= \frac{\partial Q_i}{\partial \mu_i} \left(\frac{1}{g'(\mu_i)} \right) \underline{x}_i$$

Here,

$$\frac{\partial Q_i}{\partial \mu_i} = \frac{\partial}{\partial \mu_i} \int_{y_i}^{\mu_i} \left(\frac{y_i - t}{t^2} \right) dt = \frac{y_i - \mu_i}{\mu_i^2},$$

$$g'(\mu_i) = \frac{\partial}{\partial \mu_i} (\log \mu_i) = \frac{1}{\mu_i}$$

Thus, for $\underline{u}(\underline{\beta})$, we obtain

$$\underline{u}(\underline{\beta}) = \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{\mu_i^2} \right) (\mu_i)(\underline{x}_i)$$

$$= \sum_{i=1}^{N} \left(\frac{y_i - \mu_i}{\mu_i} \right) \underline{x}_i \leftarrow \mu_i = \exp(\underline{x}_i' \underline{\beta})$$

$$= \sum_{i=1}^{N} \left(\frac{y_i - \exp(\underline{x}_i' \underline{\beta})}{\exp(\underline{x}_i' \underline{\beta})} \right) \underline{x}_i$$