Homework Exercise Set 1: STA 7249

1. (4.30 and 4.31) Show that the normal distribution $N(\mu, \sigma^2)$ is in the exponential dispersion family, and identify the components. Formulate the ordinary regression model as a GLM.

2. (4.25) A binomial GLM $\pi_i = \Phi(\sum_j \beta_j x_{ij})$ with arbitrary inverse link function Φ assumes $n_i Y_i$ has a $bin(n_i, \pi_i)$ distribution. Find w_i in $w_i = (\partial\mu_i/\partial\eta_i)^2/\text{Var}(Y_i)$ and hence $\hat{\text{Cov}}(\hat{\boldsymbol{\beta}})$. For logistic regression, show $w_i = n_i\pi_i(1-\pi_i)$.

3. (4.37) In a GLM, suppose $\text{Var}(Y) = v(\mu)$ for $\mu = E(Y)$. Show that the link g satisfying $g'(\mu) = [v(\mu)]^{-1/2}$ has the same weight matrix $\mathbf{W}^{(t)}$ at each cycle. Show that this link for a Poisson random component is $g(\mu) = 2\sqrt{\mu}$.

4. (4.34) Consider the value $\hat{\beta}$ that maximizes a function $L(\beta)$. Let $\beta^{(0)}$ denote an initial guess.

   (a) Using $L'(\hat{\beta}) = L'(\beta^{(0)}) + (\hat{\beta} - \beta^{(0)})L''(\beta^{(0)}) + \cdots$, argue that for $\beta^{(0)}$ close to $\hat{\beta}$, approximately $0 = L'(\beta^{(0)}) + (\hat{\beta} - \beta^{(0)})L''(\beta^{(0)})$. Solve this equation to obtain an approximation $\beta^{(1)}$ for $\hat{\beta}$.

   (b) Let $\beta^{(t)}$ denote approximation t for $\hat{\beta}$, $t = 0, 1, 2, \ldots$. Justify that the next approximation is

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

5. (4.20) For the logistic regression model with a single predictor x and with $\beta > 0$, show (**a**) as $x \to \infty$, $\pi(x)$ is monotone increasing, (**b**) the curve for $\pi(x)$ is the *cdf* of a logistic distribution having mean $-\alpha/\beta$ and standard deviation $\pi/(|\beta|\sqrt{3})$.

6. (4.29) Consider the class of binary models $\pi(x) = \Phi(\alpha+\beta x)$ and $\Phi^{-1}[\pi(x)] = \alpha + \beta x$. Suppose the standard *cdf* Φ corresponds to a probability density function ϕ that is symmetric around 0.

   (a) Show that x at which $\pi(x) = 0.5$ is $x = -\alpha/\beta$.

   (b) Show that the rate of change in $\pi(x)$ when $\pi(x) = 0.5$ is $\beta\phi(0)$. Show that this is $.25\beta$ for the logit link and $\beta/\sqrt{2\pi}$ (where $\pi = 3.14...$) for the probit link.

   (c) Show that the probit regression curve has the shape of a normal *cdf* with mean $-\alpha/\beta$ and standard deviation $1/\mid\beta\mid$.

7. (4.38) For noncanonical links in a GLM, show the observed information matrix may depend on the data and hence differs from the expected information. Illustrate using the probit model.

8. (5.36) Construct the log likelihood function for the model logit$[\pi(x)] = \alpha + \beta x$ with independent binomial outcomes of y_0 successes in n_0 trials at $x = 0$ and y_1 successes in n_1 trials at $x = 1$. Derive the likelihood equations, and show $\hat{\beta}$ is the sample log odds ratio.

9. (4.22) Let Y_i be a $bin(n_i, \pi_i)$ variate for group i, $i = 1, \ldots, N$, with $\{Y_i\}$ independent. Consider the model that $\pi_1 = \cdots = \pi_N$. Denote that common value by π. For observations $\{y_i\}$, show $\hat{\pi} = (\sum y_i)/(\sum n_i)$. When all $n_i = 1$, for testing this model's fit in the $N \times 2$ table, show that $X^2 = n$. Thus, goodness-of-fit statistics can be completely uninformative for ungrouped data.

10. (5.37) A study has n_i independent binary observations $\{y_{i1}, \ldots, y_{in_i}\}$ when $X = x_i$, $i = 1, \ldots, N$, with $n = \sum_i n_i$. Consider the model, logit$(\pi_i) = \alpha + \beta x_i$, where $\pi_i = P(Y_{ij} = 1)$.

    (a) Show that the kernel of the likelihood function is the same treating the data as n Bernoulli observations or N binomial observations.

    (b) For the saturated model (that is, as many parameters as observations), explain why the likelihood function is different for these two data forms. (Hint: The number of parameters differs.) Hence, the deviance reported by software depends on the form of data entry.

    (c) Explain why the difference between deviances for two unsaturated models does not depend on the form of data entry.

    (d) Suppose each $n_i = 1$. Show the deviance depends on $\hat{\pi}_i$ but not y_i. Hence, it is not useful for checking model fit.

11. (6.28) A *threshold model* can motivate the probit model. For it, there is an unobserved continuous response Y^* such that the observed $y_i = 0$ if $y_i^* \leq \tau$ and $y_i = 1$ if $y_i^* > \tau$. Suppose $y_i^* = \mu_i + \epsilon_i$, where $\mu_i = \alpha + \beta x_i$ and where $\{\epsilon_i\}$ are independent from a $N(0, \sigma^2)$ distribution. For identifiability one can set $\sigma = 1$ and the threshold $\tau = 0$. Show the probit model holds and explain why β represents the expected number of standard deviation change in Y^* for a 1-unit increase in x.

12. (6.29) Consider the choice between two options, such as two product brands. Let U_0 denote the *utility* of outcome $y = 0$ and U_1 the utility of $y = 1$. For $y = 0$ and 1, suppose $U_y = \alpha_y + \beta_y x + \epsilon_y$, using a scale such that ϵ_y has some standardized distribution. A subject selects $y = 1$ if $U_1 > U_0$ for that subject.

    (a) If ϵ_0 and ϵ_1 are independent $N(0, 1)$ random variables, show that $P(Y = 1)$ satisfies the probit model.

    (b) If ϵ_y are independent extreme value random variables, with *cdf* $F(\epsilon) = \exp[-\exp(-\epsilon)]$, show $P(Y = 1)$ satisfies the logistic regression model.

13. (6.32) Let $y_i, i = 1, \ldots, n$, denote n independent binary random variables.

(a) Derive the log likelihood for the probit model $\Phi^{-1}[\pi(\mathbf{x}_i)] = \sum_j \beta_j x_{ij}$.

(b) Show the likelihood equations for the logistic and probit regression models are

$$\sum_i (y_i - \hat{\pi}_i) z_i x_{ij} = 0, \qquad j = 0, \ldots, p,$$

where $z_i = 1$ for the logistic case and $z_i = \phi(\sum_j \hat{\beta}_j x_{ij})/\hat{\pi}_i(1-\hat{\pi}_i)$ for the probit case. (When the link is not canonical, there is no reduction of the data in sufficient statistics.)

14. (4.23) Suppose Y_i is Poisson with $g(\mu_i) = \alpha + \beta x_i$, where $x_i = 1$ for $i = 1, ..., n_A$ from group A and $x_i = 0$ for $i = n_A + 1, ..., n_A + n_B$ from group B. Show that for any link function g, the likelihood equations $\sum_{i=1}^N \frac{(y_i - \mu_i) x_{ij}}{\text{Var}(Y_i)} \frac{\partial \mu_i}{\partial \eta_i} = 0, \quad j = 1, \ldots, p$ imply that fitted means $\hat{\mu}_A$ and $\hat{\mu}_B$ equal the sample means.

15. (4.27) Let y_{ij} be observation j of a count variable for group i, $i = 1, \ldots, I$, $j = 1, \ldots, n_i$. Suppose $\{Y_{ij}\}$ are independent Poisson with $E(Y_{ij}) = \mu_i$.

    (a) Show that the ML estimate of μ_i is $\hat{\mu}_i = \bar{y}_i = \sum_j y_{ij}/n_i$.

    (b) Simplify the expression for the deviance for this model. (For testing this model, it follows from Fisher (1970 p. 58, originally published in 1925) that the deviance and the Pearson statistic $\sum_i \sum_j (y_{ij} - \bar{y}_i)^2/\bar{y}_i$ have approximate chi-squared distributions with $df = \sum_i (n_i - 1)$. For a single group, Cochran (1954) referred to $\sum_j (y_{1j} - \bar{y}_1)^2/\bar{y}_1$ as the *variance test* for the fit of a Poisson distribution, since it compares the sample variance to the estimated Poisson variance $\bar{y}_1$.)

16. (4.26) A GLM has parameter β with sufficient statistic S. A goodness-of-fit test statistic T has observed value t_o. If β were known, a P-value is $P = P(T \geq t_o; \beta)$. Explain why $P(T \geq t_o \mid S)$ is the uniform minimum variance unbiased estimator of P.

17. (4.24) For binary data with sample proportion y_i based on n_i trials, we use quasi-likelihood to fit a model using variance function $v(\pi_i) = \phi\pi_i(1-\pi_i)/n_i$. Show that parameter estimates are the same as for the binomial GLM, but the covariance matrix multiplies by ϕ.

18. (4.6) An experiment analyzes imperfection rates for two processes used to fabricate silicon wafers for computer chips. For treatment A applied to 10 wafers, the numbers of imperfections are 8, 7, 6, 6, 3, 4, 7, 2, 3, 4. Treatment B applied to 10 other wafers has 9, 9, 8, 14, 8, 13, 11, 5, 7, 6 imperfections. Treat the counts as independent Poisson variates having means μ_A and μ_B.

    (a) Fit the model $\log \mu = \alpha + \beta x$, where $x = 1$ for treatment B and $x = 0$ for treatment A. Show that $\exp(\beta) = \mu_B/\mu_A$, and interpret its estimate.

Table 1:

Game	No. Made	No. Attempts	Game	No. Made	No. Attempts	Game	No. Made	No. Attempts
1	4	5	9	4	12	17	8	12
2	5	11	10	1	4	18	1	6
3	5	14	11	13	27	19	18	39
4	5	12	12	5	17	20	3	13
5	2	7	13	6	12	21	10	17
6	7	10	14	9	9	22	1	6
7	6	14	15	7	12	23	3	12
8	9	15	16	3	10			

Source: www.nba.com

(b) Test H_0: $\mu_A = \mu_B$ with the Wald or likelihood-ratio test of H_0: $\beta = 0$. Interpret.

(c) Construct a 95% confidence interval for μ_B/μ_A. (*Hint*: First construct one for β.)

(d) Test H_0: $\mu_A = \mu_B$ based on this result: If Y_1 and Y_2 are independent Poisson with means μ_1 and μ_2, then $(Y_1 \mid Y_1 + Y_2)$ is binomial with $n = Y_1 + Y_2$ and $\pi = \mu_1/(\mu_1 + \mu_2)$.

19. (4.13) Table 1 shows the free-throw shooting, by game, of Shaq O'Neal of the Los Angeles Lakers during the 2000 NBA (basketball) playoffs. Commentators remarked that his shooting varied dramatically from game to game. In game i, suppose Y_i = number of free throws made out of n_i attempts is a $bin(n_i, \pi_i)$ variate and $\{Y_i\}$ are independent.

(a) Fit the model, $\pi_i = \alpha$, and find and interpret $\hat{\alpha}$ and its standard error. Does the model appear to fit adequately? (Note: You could check this with a small-sample test of independence of the 23×2 table of game and the binary outcome.)

(b) Adjust the standard error for overdispersion. Using the original SE and its correction, find and compare 95% confidence intervals for α. Interpret.