MIDTERM EXAM	Biostat/Stat 571

MIDTERM EXAM
February 11, 2000

Biostat/Stat 571
Winter 2000
P. Heagerty

---

1. Define the following terms:
(a) [5pts] Exponential Family
(b) [5pts] Estimating Function
(c) [5pts] Overdispersion
(d) [5pts] Quasilikelihood

2. [10pts] Describe the proportional odds model. Give an interpretation (in language that might be understood by a physician) for the regression parameter $\boldsymbol{\beta}$.

3. The following questions concern the use of regression for independent observations where the response is (Y_i =successes, N_i =total trials) and the covariates are $\boldsymbol{X}_i$. In each of the following situations comment on the CONSISTENCY of $\widehat{\boldsymbol{\beta}}$, the VALIDITY of standard errors, and CORRECTIONS that might be taken that would lead to valid inference.

(a) [5pts] Standard binomial logistic regression, $\text{logit}(\pi_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, when $E(Y_i/N_i) = \pi_i$ is correctly modelled and $\text{var}(Y_i/N_i) = \pi_i(1-\pi_i)/N_i$ but Y_i isn't binomial.

(b) [5pts] Standard binomial logistic regression, $\text{logit}(\pi_i) = \boldsymbol{X}_i\boldsymbol{\beta}$, when $E(Y_i/N_i) = \pi_i$ is correctly modelled but $\text{var}(Y_i/N_i) = \phi \times \pi_i(1-\pi_i)/N_i$.

(c) [5pts] Standard linear regression using $E(Y_i/N_i) = \pi_i = \boldsymbol{X}_i\boldsymbol{\beta}$, when $E(Y_i/N_i) = \pi_i$ is correctly modelled but $\text{var}(Y_i) = N_i\pi_i(1-\pi_i)[1+\rho(N_i-1)]$.

4. In the Vaccine Preparedness Study (VPS) a sample of 5,000 high risk HIV negative participants were enrolled and followed prospectively for 18 months. At baseline, participants were asked to report on the number of partners they had in the last six months that were of unknown HIV serostatus. Suppose it is reasonable to assume Z_i, the (actual) number of partners, is distributed as Poisson with a mean $\lambda_i(\boldsymbol{X}_i)$, that may depend on covariates $\boldsymbol{X}_i$. Interest is in whether drug and alcohol use correlate with (predict) the number of partners. Let Y_i be the reported number of partners and let $\boldsymbol{X}_i$ be covariates of interest.

(a) [10pts] Suppose that we are concerned that a fraction of the participants might be uncomfortable reporting sensitive information and will choose to simply report $Y_i = 0$ even though $Z_i > 0$. Otherwise, participants report correctly, $Y_i = Z_i$. Let $\pi(\boldsymbol{X}_i)$ denote the probability that $Y_i = 0$ when $Z_i > 0$ for a subject with covariate $\boldsymbol{X}_i$. Describe a mixture model for Y_i and derive the mean and variance of Y_i in terms of $\pi(\boldsymbol{X}_i)$ and $\lambda(\boldsymbol{X}_i)$.

(b) [5pts] Suppose we are willing to assume that $\pi(\boldsymbol{X}_i) = \pi_0$, that is, misreporting is independent of covariates. If we are interested in the model $\log(\lambda_i) = \beta_0 + \boldsymbol{\beta}_1 \boldsymbol{X}_i$ then can we use standard Poisson regression estimates based on Y_i to make inference about the parameters $\boldsymbol{\beta}_1$? Justify.

(c) [15pts] Regression summaries for the VPS baseline data (a subset of 2,000 men) are attached. What conclusions can you make regarding drug and alcohol use? Justify your statements. In particular, explain the relationship between amphetamine use and the number of partners.

5. Suppose we have continuous response vectors $\boldsymbol{Y}_i = \text{vec}(Y_{ij})$ for $j = 1, 2, \ldots, n_i$ with covariates $\boldsymbol{X}_i$ ($n_i \times p$ matrix). Furthermore, suppose that the variance of $\boldsymbol{Y}_i$ is assumed to be $\boldsymbol{\Sigma}_i$ ($n_i \times n_i$ matrix).

(a) [5pts] If we are interested in the regression model $E(\boldsymbol{Y}_i) = \boldsymbol{X}_i \boldsymbol{\beta}$ then what is the optimal estimator of $\boldsymbol{\beta}$ based on $\boldsymbol{Y}_i$, $i = 1, 2, \ldots, m$? What is the asymptotic distribution of this estimator if $\boldsymbol{Y}_i$ are independent and $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{\Sigma}_i$? (you can assume $\boldsymbol{\Sigma}_i$ is known)

(b) [10pts] What is the asymptotic distribution of the estimator, $\widehat{\boldsymbol{\beta}}$, given in (a) if $\text{var}(\boldsymbol{Y}_i) = \boldsymbol{M}_i$? How would you estimate the asymptotic variance of $\widehat{\boldsymbol{\beta}}$?

(c) [10pts] If we have a single covariate of interest, x_{ij}, and x_{ij} varies within each subject, then we can calculate the individual regression estimates:

$$\widehat{\boldsymbol{\beta}}_i = \left(\boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{X}_i \right)^{-1} \boldsymbol{X}_i^T \boldsymbol{\Sigma}_i^{-1} \boldsymbol{Y}_i \ .$$

Suppose that our variance model is correct. Show that the optimal weighted least squares estimator based on $\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_m$ is of the form

$$\widehat{\boldsymbol{\beta}} = \sum_{i=1}^{m} w_i \widehat{\boldsymbol{\beta}}_i$$

where the weights are given by the Gauss-Markov theorem.