BIOSTAT 653 Homework #5

Due December 11th, 3:10pm, in class.

Problem 1

The incidence of skin cancers is to be monitored in a sample of N patients undergoing PUVA treatment for psoriasis. Consider this simplified version of the problem. Patients initially free of disease are followed annually on 2 successive occasions and the number of new cancers is noted. Let Y_i denote the number of new cancers for t=1,2, for the i'th subject. Our objective is simply to estimate $\mu_t = E(Y_{it})$, for t=1,2, i.e. the annual incidence rates.

1) Consider a GEE model with identity link. Under working independence assumption (i.e. W=I), estimate $\mu = \binom{\mu_1}{\mu_2}$ and $\delta = \mu_2 - \mu_1$. Note that although we have explained a two-stage estimation procedure for GEE in class, GEE works with any working weight/covariance matrices just as in WLS.

2) Use the sandwich variance estimate to obtain $V(\hat{\mu})$ and hence $V(\hat{\delta})$. You should get closed form expressions.

3) With count data, we usually use the log-link, i.e. $\beta_1 = \log(\mu_1)$, $\beta_2 = \log(\mu_2)$. What is the interpretation of $\gamma = \beta_2 - \beta_1$?

4) With the log-link, what is the usual variance function in generalized linear models? Suggest a covariance matrix form based on the usual variance function.

5) Suppose we use GEE with log-link to estimate β and δ again using any working weight/covariance matrix W that does not depend on i. Give out the GEE equations and explain how you would estimate $\mu = \binom{\mu_1}{\mu_2}$.

6) Again using the sandwich variance estimation, and assuming $X_i = I$ and a working variance matrix W, derive an expression for an estimate of $V\left(\hat{\beta}\right)$ using the log-link.

Problem 2

Problem 14.1 on the textbook (page 434-436)