# BIOSTAT 653 Homework #5 Solution

**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 denote the number of new cancers for t=1,2, for the i'th subject. Our objective is simply to estimate , 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 and . 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.

With identity link, , and the GEE equations are

Because of working independence assumption, we have . Thus

or equivalently,

1. Use the sandwich variance estimate to obtain and hence . You should get closed form expressions.

Because , the variance estimate formula is simplified to

Because , by Delta method or rules of variance, we have

1. With count data, we usually use the log-link, i.e. . What is the interpretation of ?

is the log of the incidence rate ratio.

1. 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.

We usually assume that , so a natural working variance matrix is

where is the .

1. Suppose we use GEE with log-link to estimate and 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 **.**

With log link, , and assume , we have

and solve to get:

or equivalently,

Remark: For D and (invertible) W independent of i, your GEE estimates will be the sample mean.

1. Again using the sandwich variance estimation, and assuming and a working variance matrix W, derive an expression for an estimate of using the log-link.

**Problem 2**

Problem 14.1 on the textbook (page 434-436)

**Solution**

14.1.1

**DATA** toenail;

INFILE 'toenail-data.txt';

INPUT id $ Y Trt Month Visit;

**RUN**;

**PROC** **GENMOD** DESCENDING DATA=toenail;

CLASS id;

MODEL Y=Month Trt\*Month / DIST=BINOMIAL LINK=LOGIT;

REPEATED SUBJECT=id / TYPE=EXCH;

**RUN**;

| **Analysis Of GEE Parameter Estimates** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Empirical Standard Error Estimates** | | | | | | |
| **Parameter** | **Estimate** | **Standard Error** | **95% Confidence Limits** | | **Z** | **Pr > |Z|** |
| **Intercept** | -0.5782 | 0.1304 | -0.8337 | -0.3226 | -4.43 | <.0001 |
| **Month** | -0.1713 | 0.0296 | -0.2293 | -0.1134 | -5.80 | <.0001 |
| **Month\*Trt** | -0.0777 | 0.0538 | -0.1831 | 0.0277 | -1.44 | 0.1485 |

14.1.2

is the change in log odds of moderate or severe onycholysis per month in the Itraconazole treatment group.

14.1.3

is the difference between changes in log odds of moderate or severe onycholysis per month in the Terbinafine treatment group and the Itraconazole treatment group.

14.1.4

From the result table in 14.1.1, we conclude that the log odds of moderate or severe onycholysis in the Itraconazole treatment group is significantly reduced over the months. In addition, we conclude that there is no statistically significant difference between Itraconazole treatment and Terbinafine treatment on reducing the log odds of moderate or severe onycholysis over the months.

14.1.5

**PROC** **GLIMMIX** METHOD=QUAD(QPOINTS=**50**) DATA=toenail;

CLASS id;

MODEL Y=Month Trt\*Month / DIST=BINOMIAL LINK=LOGIT S;

RANDOM INTERCEPT / SUBJECT=id TYPE=UN;

**RUN**;

| **Solutions for Fixed Effects** | | | | | | |
| --- | --- | --- | --- | --- | --- | --- |
| **Effect** | **Estimate** | **Standard Error** | **DF** | **t Value** | **Pr > |t|** | |
| **Intercept** | -1.6972 | 0.3298 | 293 | -5.15 | <.0001 | |
| **Month** | -0.3885 | 0.04330 | 1612 | -8.97 | <.0001 | |
| **Month\*Trt** | -0.1424 | 0.06493 | 1612 | -2.19 | 0.0284 | |
|  |  |  |  |  |  | |
| **Covariance Parameter Estimates** | | | | | |
| **Cov Parm** | **Subject** | **Estimate** | **Standard Error** | | |
| **UN(1,1)** | **id** | 16.0349 | 3.0395 | | |

14.1.6

The estimated variance for the intercept is , which suggests high individual-level heterogeneity. [R gives you 20.61 apparently. I haven’t taken the time to check why this was though. Exercise for fun.]

14.1.7

is the change in log odds of moderate or severe onycholysis per month in the Itraconazole treatment group, for a typical individual ~~with .~~

14.1.8

is the difference between changes in log odds of moderate or severe onycholysis per month in the Terbinafine treatment group and the Itraconazole treatment group, for a typical individual ~~with .~~

14.1.9

The estimate in GEE is smaller in magnitude [that’s what attenuate means here] compared with the estimate from GLMM. That one slide in your lecture note shows that all the betas should be smaller in magnitude by about the same amount (check this). [If you want to see more about deriving this, you can read this: http://www.jstor.org/pss/2531734]

14.1.10

The parameter estimates become stable once the number of quadrature points is above 20.