Biostats 653 – Final

NAME .....

Date: December 15, 2016 *Instructor: Xiang Zhou*

Time: 120 minutes (4:00pm – 6:00pm)

Note that all sub-questions of Question 1 can be answered in a sentence or two. For Questions 2 and 3, you can directly use formulas derived in class. Try not to leave empty space even if you do not know the answer.

USE THE UPDATED FINAL WORD FILE FOR THE EXAM, but use the solution here.

Question 1 (40 pt)

A study was conducted to investigate the effects of an antidepressant drug used to treat individuals who suffer from debilitating panic attacks. Panic attacks are temporary periods (on the order of 10 to 15 minutes) of intense fear and distress that can terrify the sufferer and interfere with his or her day-to-day life. A total of m=300 subjects confirmed to suffer from such attacks were recruited and randomized to three groups:

Group 1 – low-dose antidepressant therapy (100 subjects)

Group 2 – high-dose antidepressant therapy (100 subjects)

Group 3 – placebo (100 subjects)

Before starting his/her assigned treatment, each subject was asked whether or not she/he had suffered at least one panic attack in the previous week (0=no, 1=yes). This was taken as the subject's baseline response (week 0). All subjects then started on their assigned therapies. At 1, 2, 3, and 4 weeks thereafter, each subject visited the clinic and was asked to report whether or not she/he had suffered at least one attach since the last visit (0=no, 1=yes). The gender of each subject (0=female, 1=male) was also recorded.

The following table shows the proportions of subjects reporting suffering at least one attach in the previous week at each time point:

Group	Baseline(week 0)	Week 1	Week 2	Week 3	Week 4
1 (low dose)	0.70	0.61	0.65	0.59	0.54
2 (high dose)	0.60	0.57	0.54	0.51	0.39
3 (placebo)	0.67	0.59	0.64	0.66	0.66

Let Y_{ij} be the indicator of whether or not subject i reported at least one panic attack, $i = 1, \cdots, 300$, in week $t_{ij} = 0,1,2,3,4$, and let

$\delta_{1i} = 1$ if subject i was in Group 1 (low dose), = 0 otherwise

$\delta_{2i} = 1$ if subject i was in Group 1 (high dose), = 0 otherwise

$\delta_{3i} = 1$ if subject i was in Group 1 (placebo), = 0 otherwise

The study team first considered the following marginal model for the probability of at least one panic attack:

$$E(Y_{ij}) = \frac{\exp(\beta_0 + \beta_1 t_{ij}\delta_{1i} + \beta_2 t_{ij}\delta_{1i} + \beta_1 t_{ij}\delta_{1i})}{1 + \exp(\beta_0 + \beta_1 t_{ij}\delta_{1i} + \beta_2 t_{ij}\delta_{1i} + \beta_1 t_{ij}\delta_{1i})}$$

1). (5 pt) In words, state what the above model assumes about the pattern of change of the logarithm of the odds of having at least one panic attack in each group.

The model assumes that the log odds is $\beta_0 + \beta_1 t_{ij}\delta_{1i} + \beta_2 t_{ij}\delta_{1i} + \beta_1 t_{ij}\delta_{1i}$. Thus, the embedded assumption is that the log odds changes smoothly at a constant rate over time in each group, where β_k is the change in log odds per week in group k.

2). (5 pt) The first question the study team wished to address was whether or not the pattern of change of the log odds of having at least one panic attack once treatment is initiated is possibly different for at least one of the three groups. State the null hypothesis and explain how you would test this null.

The null hypothesis is $H_0: \beta_1 = \beta_2 = \beta_3$. We can construct a contrast matrix to test it. (Because we are using GEE, we cannot use likelihood rate test to do so. It is also wrong to have the null as $H_0: \beta_1 = \beta_2 = \beta_3 = 0$.)

3). (5 pt) To complete the model specification, the study team need to make an assumption on the marginal variance $V(Y_i)$ (an n by n matrix). How would you choose $V(Y_i)$?

Because this is a binary data set, I would assume the marginal variance $V(Y_{ij}) = E(Y_{ij})(1 - E(Y_{ij}))$, as diagonal elements inside the diagonal A_i matrix. Then I would also assume an unstructured marginal correlation structure $R(\alpha)$. This way, $V(Y_i) = A_i^{\frac{1}{2}} R(\alpha) A_i^{\frac{1}{2}}$.

4). (5 pt) During the study, some subjects failed to come in week 4. The study team found out that whether a subject came in week 4 only depends on his/her baseline measurement and does not depend on the β's in the model (i.e. a patient tends not to come if he/she has at least one attack in the week before week 0). What is the missing data mechanism?

Missing at random (MAR).

5). (5 pt) Under the missing data mechanism of 1.4, can we ignore the missing data and use only the observed data to fit the marginal model (i.e. is the missingness ignorable)?

MAR is not ignorable under GEE, so we have to use a weighted version of GEE.

6). (5 pt) The study team then considered a conditional model with a random intercept, with the conditional mean defined as:

$$E(Y_{ij}|b_i) = \frac{\exp(\beta_0 + \beta_1 t_{ij}\delta_{1i} + \beta_2 t_{ij}\delta_{1i} + \beta_1 t_{ij}\delta_{1i} + b_i)}{1 + \exp(\beta_0 + \beta_1 t_{ij}\delta_{1i} + \beta_2 t_{ij}\delta_{1i} + \beta_1 t_{ij}\delta_{1i} + b_i)}$$

What is the interpretation of β_1? Is it different from the β_1 in the above marginal model?

β_1 represents the additive increase in log odds per one unit increase of time for a patient in group 1, conditional on the random effects (or for a typical individual with 0 random effect). The interpretation of β_1 is different between GEE and GLMM, as the former has a population level interpretation while the later has a subject specific interpretation.

7). (5 pt) Generally, what are the advantage(s) and disadvantage(s) of a conditional model compared with a marginal model?

GLMM: (1) need distributional assumptions; (2) has a likelihood; (3) computationally difficult; (4) can make inference at a subject level.

GEE: (1) only need assumptions on the two moments; (2) not likelihood based; (3) is computationally easy; (4) have a population level interpretation.

8). (5 pt) Under the same missing data mechanism of 1.4, can we ignore the missing data and use only the observed data to fit the conditional model (i.e. is the missingness ignorable)?

MAR is ignorable under likelihood inference (though the estimates could be less efficient), so we can use observed data directly.

Question 2 (30 pt)

Assume we have observed longitudinal data $y_{ij}, i = 1, \cdots, N, j = 1, \cdots, n_i$. In both question 1 and 2, we will consider a special case when the number of repeated measurement is 1, or $n_i = 1$, for all i. We define $y_i \equiv y_{i1}$. Suppose we have count data $y_1, \cdots, y_N$ which we assume are independently distributed with mean β, or $E(y_i) \equiv \mu_i = \beta$ (i.e. *identity* link with constant mean).

1). (5 pt) Suppose we fit this model under a working variance assumption that $V(y_i) = \mu_i$, as is the case for Poisson data. Obtain the GEE estimator $\hat{\beta}_{GEE}$.

In this simple example,

$$D_i = \frac{\partial \beta}{\partial \beta} = 1, V_i = \mu_i = \beta$$

The GEE is

$$\sum_{i=1}^{N} D_i^T V_i^{-1}(y_i - \beta) = \sum_{i=1}^{N} (y_i - \beta)$$

Set the GEE to zero, we have

$$\hat{\beta}_{GEE} = \bar{y}$$

2). (5 pt) Obtain the model-based variance estimator for $\hat{\beta}_{GEE}$, or $V_1(\hat{\beta}_{GEE})$.

Because

$$F = \sum_{i=1}^{N} D_i^T V_i^{-1} D_i = N\beta^{-1}$$

The model based estimator of variance is

$$\hat{V}\left(\hat{\beta}_{GEE}\right) = \hat{F}^{-1} = \frac{\bar{y}}{N}$$

3). (5 pt) Obtain the robust variance estimator for $\hat{\beta}_{GEE}$, or $V_2(\hat{\beta}_{GEE})$.

Because

$$\widehat{Cov}(y_i) = (y_i - \bar{y})^2$$

We have

$$\hat{G} = \sum_{i=1}^{N} D_i^T V_i^{-1}(y_i - \bar{y})^2 V_i^{-1} D_i = \sum_{i=1}^{N} (y_i - \bar{y})^2 \hat{\beta}^{-2}$$

Therefore, the sandwich/robust estimator of variance is

$$\hat{V}\left(\hat{\beta}_{GEE}\right) = \hat{F}^{-1}\hat{G}\hat{F}^{-1} = \sum_{i=1}^{N} (y_i - \bar{y})^2 / N^2$$

4). (5 pt) Based on your GEE estimator $\hat{\beta}_{GEE}$ in question 2.1, use standard random variable properties to derive a variance estimator for $\hat{\beta}_{GEE}$, or $V_3(\hat{\beta}_{GEE})$, as a function of $V(y_i)$.

$$V(\hat{\beta}_{GEE}) = V(\bar{y}) = \frac{V(y_i)}{N}$$

5). (5 pt) Explain how $V_1(\hat{\beta}_{GEE})$ or $V_2(\hat{\beta}_{GEE})$ can be obtained from $V_3(\hat{\beta}_{GEE})$ based on different assumptions on y_i.

$V(y_i)$ can be estimated by $\sum_{i=1}^{N}(y_i - \bar{y})^2 / N$ (no assumption) or $\bar{y}$ (Poisson assumption).

6). (5 pt) Which of these two variance estimators in 2.2 and 2.3 would you expect to be better (i) if the Poisson model holds, and (ii) if there is severe over-dispersion relative to a Poisson model?

If the Poisson model holds, then I would expect (i) to work better as model-based variance is often more efficient than robust variance estimator. Otherwise, (ii) would work better.

Question 3 (30 pt)

1). (5 pt) As in question 2, now suppose you decide to use a generalized linear mixed model with mean μ and a random intercept b_i to model y_i. Specifically, you assume that y_i, conditional on β, b_i, follows a Poisson distribution, with the conditional mean $\log\big(E(y_i|b_i)\big) = \beta + b_i$. You further assume that b_i follows a normal distribution with mean 0 and variance σ^2. Write down your model and your likelihood. (Note that the probability mass function for a Poisson random variable x is $P(x;\lambda) = \frac{\lambda^x e^{-\lambda}}{x!}$.)

$$y_i \sim Poi\big(e^{\beta+b_i}\big), b_i \sim N(0,\sigma^2)$$

The likelihood is

$$P(y|\mu,\sigma^2) = \prod\nolimits_{i=1}^{N} \int \frac{e^{y_i(\beta+b_i)} e^{-e^{\beta+b_i}}}{y_i!} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{b_i^2}{2\sigma^2}} db_i$$

2). (10 pt) Compute the marginal mean $E(y_i)$ and the marginal variance $V(y_i)$ of the GLMM. Show that $V(y_i) > E(y_i)$; that is, GLMM naturally accounts for over-dispersion.

Based on class notes:

$$E(y_i) = E_{b_i}\big(e^{\beta+b_i}\big) = e^{\beta+\frac{\sigma^2}{2}}$$

$$V(y_i) = V_{b_i}\big(e^{\beta+b_i}\big) + E_{b_i}\big(e^{\beta+b_i}\big) = e^{2\beta}\big(e^{2\sigma^2} - e^{\sigma^2}\big) + e^{\beta+\frac{\sigma^2}{2}} = E(y_i)^2\big(e^{\sigma^2} - 1\big) + E(y_i) > E(y_i)$$

3). (5 pt) What is the main idea behind a Laplace approximation to approximate the above likelihood?

The main idea is to use a Taylor series expansion to approximate the integrand; by doing so, we obtain a normal kernel which can be integrated out directly.

4). (5 pt) In class, we have focused on four different methods to deal with the intractable integration in GLMM. Here, you will develop an alternative approach for inference in GLMM. You notice that the integral of GLMM is often intractable because we assume that the random effects b_i follow a normal distribution. You reason that, instead of normal, perhaps a different assumption on the distribution of the random effects can lead to a tractable integral, at least in some cases. To see this, you decide to assume that the exponential of the random effects, which becomes a positive value, follows a gamma distribution, or $a_i \equiv e^{b_i} \sim Gamma(\alpha, \theta)$, with α, θ known. With this assumption, re-write the likelihood function in terms of the new random effects a_i. (Note that the probability density function for a gamma distribution is $P(x|\alpha,\theta) = \frac{\theta^\alpha x^{\alpha-1} e^{-x\theta}}{\Gamma(\alpha)}$.)

$$P(y|\mu,\sigma^2) = \prod_{i=1}^{N} \int \frac{a_i^{y_i} e^{y_i\beta} e^{-a_i e^\beta}}{y_i!} \frac{\theta^\alpha a_i^{\alpha-1} e^{-a_i\theta}}{\Gamma(\alpha)} da_i$$

5). (5 pt) Inside the integral, you look that the part that involves a_i and you notice that this part looks very familiar. In fact, you notice that this part belongs to some distribution (i.e. is the kernel of that distribution). What distribution is this? Based on this observation, and the fact that any distribution integrates to one, solve the integration in GLMM analytically.

We recognize the kernel as another Gamma distribution, with parameters $(\alpha + y_i, \theta + e^{\beta})$, thus, the integral is

$$P(y|\mu, \sigma^2) = \prod_{i=1}^{N} \frac{e^{y_i\beta}\theta^{\alpha}\Gamma(\alpha + y_i)}{\Gamma(\alpha)(\theta + e^{\beta})^{\alpha+1}y_i!}$$

6). (5 pt) Explain how to estimate β and its standard error afterwards.

After the analytic integration, the likelihood becomes a function of only β, σ^2. We can use EM or NR or other algorithms to perform inference. We can use the information matrix to compute the standard errors for the parameters.