

## PROBLEM SET 4

1. A study was conducted in West Java, Indonesia, to determine the effects of vitamin A deficiency in preschool children. The investigators were particularly interested in whether children with vitamin A deficiency were at increased risk of developing respiratory infection, which is one of the leading causes of death in this part of the world.

250 children were recruited in the study, and their age in years, gender (0 =male, 1 =female), and whether they suffered vitamin A deficiency (0 =no, 1 =yes) was recorded at an initial clinic visit (time 0). Also recorded was the response, whether the child was suffering from a respiratory infection (0 =no, 1 =yes). The children then were examined again at 3 month intervals for a year (at 3,6,12, and 15 months after the first visit) and the presence or absence of respiratory infection was recorded at each of these visits. Luckily, all children we seen at all visits, so there were no missing data.

The data file

<http://biosun01.biostat.jhsph.edu/~fdominic/ICHS.dat>

has the following columns:

| Column | Description                                |
|--------|--|
| 1      | Child id                                   |
| 2      | Response (0 or 1 as above)                 |
| 3      | Time (in months, as above)                 |
| 4      | Gender (male = 0, female=1)                |
| 5      | Vitamin A (not deficient =0, deficient =1) |
| 6      | age (in years)                             |

- (a) Let  $\mathbf{y}_i$  be the vector of responses for the  $i$ th child, consisting of elements  $y_{ij}$ , the observations on whether the child has a respiratory infection at time  $t_{ij}$  (recorded in months). Write down a model for  $E(y_{ij})$  in terms of an appropriate link function that is linear in an intercept and include additive terms for time, age, gender, and vitamin A status. Also, write down  $var(y_{ij})$  given the nature of the response.
- (b) The investigators had not taken a course in longitudinal analysis; thus, they were unaware that measurements on the same child might be correlated. They fit the model in (a) without taking correlation into account, treating all the observations from all children as if they were *unrelated*.

Based on this fit, is there sufficient evidence to suggest that the mean pattern of respiratory response is associated with the presence or absence of vitamin A deficiency? State the null hypothesis corresponding to this issue in terms of your model (a), cite the test statistic and p-value on which you base this conclusion, and state your conclusion as a meaningful sentence.

- (c) Because you have taken a course in longitudinal data analysis, the investigators called you in for help with an improved analysis. Extend the model (a) to take into account correlation among repeated measurements on the same subject.
- (d) Fit your model in (c) to the data taking into account the possible structure of correlation among the elements of a data vector. Assuming that your assumed model for correlation is correct, conduct a test of null hypothesis in part (c). State your conclusion as a meaningful sentence.

Do the results agree with those in part (a)? Give a possible explanation for this, citing results from your output to support your explanation.

- (e) Select a correlation model you feel is most plausible based on your inspection, explaining why you chose this model, and fit this model to the data.

[(i)] Is there sufficient evidence to suggest that the probability of respiratory infection changed over the 15 month study period?

[(ii)] Is there sufficient evidence to suggest that it is worthwhile to take gender into account in understanding the risk of respiratory infection in this population of children?

Given these considerations, conduct an analysis of these data. Write a brief report summarizing:

- The statistical model you assumed, and why you choose it
- The analyses you conducted, the assumptions you made and why you made them
- The results, addressing the interests of the investigators as described above.

*Carry out whatever analyses you feel are appropriate. It is important to write a clear and organized report summarizing what you did and why you did it. DO NOT INCLUDE COMPUTER OUTPUT*