Longitudinal Analysis Homework 5

Mairead Dillon

2023-03-13

# a) Conduct an analysis of the longitudinal dataset that focuses on the relationship between the prevalence of respiratory disease and age within each of the two smoking (mother’s) groups. Examine both the mean and correlation structure. Comment on the difficulties of using the correlation coefficient as a measure of dependence between binary outcome variables. Are there restrictions on the value of the correlation coefficient related to the mean of binary outcomes?

The two tables below show the prevalence and standard error of respiratory disease by age in children whose mothers are non-smokers and children whose mothers are smokers. The prevalence is higher in children whose mothers smoked.

Mother = Non-Smoker

Age

Mean

SE

6

0.1600000

0.0196240

7

0.1485714

0.0190383

8

0.1428571

0.0187312

9

0.1057143

0.0164586

Mother = Smoker

Age

Mean

SE

6

0.1657754

0.0272675

7

0.2085561

0.0297896

8

0.1871658

0.0285995

9

0.1390374

0.0253689

The matrix below is the correlation matrix for the model.

1.0000000

0.3504383

0.3083147

0.3029799

0.3504383

1.0000000

0.4695526

0.3185426

0.3083147

0.4695526

1.0000000

0.3763815

0.3029799

0.3185426

0.3763815

1.0000000

The correlation coefficient between respiratory disease and mother’s smoking status is 0.0476. The correlation coefficient between binary variables does not live in the whole range for Pearson’s correlation because it is constrained by the means of the variables.

# b) Analyze the effect of mother’s smoking on prevalence of respiratory disease as a function of age using a marginal regression model. Is there any evidence for an interaction between smoking and age? How do you interpret this interaction? Experiment with a number of different “working” correlation matrices including, at a minimum, independence, exchangeable and autoregressive (1st order). Which if any of these seems most suitable for use with these data?

Under the independence and exchangeable correlation structures, the children whose mother smoked have a 1.4 times greater odds of having respiratory infection compared to children whose mother did not smoke when age and the interaction between age and mother’s smoking status are held constant. Under the autoregressive (1st order) correlation structure, children whose mother smoked have a 1.3 times greater odds of having respiratory infection compared to children whose mother did not smoke when age and the interaction between age and mother’s smoking status are held constant. There is no evidence for an interaction between smoking and age in any of these models. If there were a significant interaction effect, it would be the mean respiratory rate over time for children whose mother smoked. The exchangeable correlation structure seems most suitable for use with this data.

# c) Estimate a conditional model for the probability of infection. Explicitly state your model and modeling assumptions. Interpret the estimated coefficient for mothers’ smoking status.

The model assumptions are:  
1. Linearity  
2. No multicollinearity  
3. No outliers  
The odds of a child having respiratory disease increase by a factor of 1.6 when the child’s mother smoked compared to when the child’s mother did not smoke when age is held constant.

# d) Compare the estimated coefficients for mothers’ smoking status from the fitted models in (b) and (c). Comment upon the magnitude of the estimates and potential problems in the their interpretations.

The estimate in model (c) has a greater magnitude than the estimates in model (b). The marginal models in (b) estimate population odds, while the model in (c) estimates individuals’ odds. The problem with the marginal model is that if some mother’s smoking status changed over the course of the study, those children will be in both populations. The problem with the conditional model is that some children’s mothers did not change smoking status over the course of the study so the mother’s smoking status variable did not change over time for these children. Additionally, it’s unethical to force people to smoke so it’s not possible to force people to smoke at one point in the study and force them not to smoke at another point in order to collect data when mother’s smoking status has each value for each patient.