

Assignment 3 STAT 315-463: Multivariable Statistical Methods and Applications

Due date: Friday 28 April 2023

- Your assignment needs to show the R code you used, and your well discussed answers to the questions.
- Submit your assignments on Learn.

Background

In `Contraception315.csv`, you are provided with a dataset, modified from a study originally undertaken to ascertain associations concerning contraceptive use among Bangladeshi women. The data available for this assignment consists of data on 453 women in 5 districts. The predictor variables are

- *use*, an indicator for contraceptive use (coded N for no and Y for yes).
- Two geographical location covariates, *district* (5 levels), and *urban* (2 levels), which should be treated as factor variables.
- A continuous covariate for standardised age *age*

The response variable is the number of living children *livch* (0, 1, 2, 3, 4, 5, 6, 7).

Questions

1. Fit a Poisson Regression including all possible predictor variables. (1 mark)
2. Apply either forward or backward selection to determine the most appropriate model for this data. For the chosen model, write down the model equation. (2 marks)
3. For the model chosen in 2, provide an interpretation of the regression coefficients, on both the link scale and the response scale. Include 95 % confidence intervals (4 marks)
4. Describe what the implication of unaccounted overdispersion would be for any inference made. Comment on whether you believe overdispersion is present in this dataset. Describe how you would change your analysis to deal with overdispersion. (4 marks)
5. You are later told that the researchers in comparisons between women without children and women with children. Create a new variable `child` such that a 0 indicates a women with no living children, and 1 indicates a women with at least one living child. (1 mark)
6. Fit a Logistic Regression including all possible predictor variables and `child` as the response. (1 mark)
7. Apply either forward or backward selection to determine the most appropriate model for this data. For the chosen model, write down the model equation. Were the same predictor variables chosen as the Poisson regression (3 marks)
8. If Y is Poisson distributed with parameter λ , then $\Pr(Y > 0) = 1 - e^{-\lambda}$. Use this to construct estimates of the probability of having at least one living child for each women in the study from the Poisson regression model chosen in part 2. Compare this to the estimate of the probability of having at least one living child for each women in the study using the Logistic regression model chosen in part 7. (Note: Due to the number of women in the study, do not provide a table of all the estimated probabilities. Instead focus on graphical visualizations to determine any similarities and differences.) (4 marks)