Assignment 08

Poisson Regression

This assignment is intended to give you experience working with Poisson regression models to analyze count data. *Do not include any R syntax or output unless it is specifically required in the question.* Please submit your responses to each of the questions below. Please submit your responses to each of the questions below in a printed document. All graphics should be resized so that they do not take up more room than necessary and should have an appropriate caption. All tables should also have an appropriate caption.

This assignment is worth 14 points. Each question is worth 1 point unless otherwise noted.

For this assignment, you will use the file, *risky-behavior.csv*. This file contains data (provided by Gelman & Hill, 2007) collected from a randimized trial targeting opposite sex couples at high risk of HIV infection. The intervention provided counseling sessions related to implementing practies that could reduce their likelihood of contracting HIV. Couples were randomized to (1) a control group; (2) a group in which only the female member of the couple participated in the intervention; or (3) a group in which both members of the couple participated in the intervention. These particular data consist of data from 434 individuals who participated in the study. The variables in this file are:

- `female`: Dummy coded sex variable (0=male, 1=female)
- `intervention`: To which level of intervention was the person randomly assigned? (Control group; Both members of the couple received the intervention; Only the female member of the couple received the intervention)
- `hiv_baseline`: Dummy variable indicating whether the participant was HIV positive at baseline (0=no; 1=yes)
- `unprotected_baseline`: Number of unprotected sex acts at baseline
- `unprotected_follow_up`: Number of unprotected sex acts three months after the completion of the study

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models.* New York: Cambridge University Press.

Exploration

You will begin the analysis by examining the effect of intervention on the number of unprotected sex acts three months after the completion of the study.

1. Compute and report the number of unprotected sex acts three months after the completion of the study for each of the three levels of intervention.
2. Create a histogram of the number of unprotected sex acts three months after the completion of the study for each of the three levels of intervention.
3. Based on the summary measures and plots, what does the sample evidence suggest about whether the intervention is effective? Explain.

Poisson Model 1

Fit a Poisson model (Model 1) that uses the intervention variable as a predictor of the number of unprotected sex acts three months after the completion of the study. In this model use the control group as the reference group.

4. Back-transform and interpret the coefficients associated with the intervention variables from the fitted equation.
5. Using the model-level information about the residual- and null-deviance and *df* reported in the `summary()` output for Model 1, carry out a test of deviance using the `pchisq()` function. Based on this test, is there statistical evidence of an effet of intervention? Explain. (Show your R syntax for this computation.)

Poisson Model 2

Fit a second Poisson model (Model 2) that uses the intervention variable as a predictor of the number of unprotected sex acts three months after the completion of the study. In this model, also control for the pre-study measure of the outcome, whether the participant was HIV positive at baseline, and the participant's sex. Again use the control group as the reference group.

6. Using the model-level information about the residual- and null-deviance compute and report the proportion of deviance accounted for by Model 2. (Show your work.)

Fit a model that only includes the three control variables (pre-study measure of the outcome and the participant's sex). This model can be used as a comparison model to Model 2 in order to allow you to evaluate whether there is an effect of intervention "above and beyond" the effects of the control variables.

7. Compute and report the proportion of deviance accounted for by including the intervention in the model relative to not including intervention in the model. Does intervention seem to explain additional variation in the number of unprotected sex acts three months after the completion of the study? Explain.
8. Using Model 2, evaluate whether or not there is evidence of overdispersion. Explain by referring to the statistical evidence you used to evaluate this.

Negative Binomial Model

Fit a negative binomial model (Model 3) that uses the intervention variable as a predictor of the number of unprotected sex acts three months after the completion of the study. In this model, also control for the pre-study measure of the outcome, whether the participant was HIV positive at baseline, and the participant's sex. Again use the control group as the reference group.

9. Create and report the standing rootogram for Model 2 and Model 3. Comment on which model seems to be more compatible with the observed data.
10. Back-transform and interpret the coefficients associated with the intervention variables from the fitted equation.
11. Based on the results from Model 3, comment on whether the intervention was effective in reducing the number of unprotected sex acts. Also, comment on whether one of the intervention levels was more effective than another.

Model Evaluation

12. Based on the standing rootogram for Model 3, is there evidence that we need to fit a zero-inflated model? Or will the negative binomial suffice? Explain.

13. Examine a plot of the Pearson residuals versus the fitted values for Model 3. Comment on whether you believe the distributional assumptions of the model have been satisfied.

14. These data include responses from both the men and women of the participating couples. Does this give you any concern with regard to the model's assumptions? Explain.