Stat 363: Binomial Regression

Example: Moth Coloration and Natural Selection. An article in the *Journal of Animal Ecology* by Bishop (1972) investigated whether moths provide evidence of “survival of the fittest” with their camouflage traits. Researchers glued equal numbers of light and dark morph moths in lifelike positions on tree trunks at 7 locations from 0 to 51.2 km from Liverpool. They then recorded the numbers of moths removed after 24 hours, presumably by predators. The hypothesis was that, since tree trunks near Liverpool were blackened by pollution, light morph moths would be more likely to be removed near Liverpool. The data is found in **moth.csv**, and relevant R code can be found under **Moths.Rmd**. Variables include:

* MOPRH = light or dark
* DISTANCE = kilometers from Liverpool
* PLACED = number of moths of a specific morph glued to trees at that location
* REMOVED = number of moths of a specific morph removed after 24 hours

0. What do you think of the study design? Any suggestions for improvement?

1. What are logits, and why would we want to plot logits vs. distance (rather than, say, proportion removed vs. distance)?

2. What can we conclude from the empirical logit plots?

3. Interpret the 3 coefficient estimates from model “breg2”. Note that breg2 and breg2a provide two alternative ways to express the same model…

4. What are the implications to using MORPH rather than “dark” in breg2b?

5. How do the predicted logits from breg2 fit the actual data? Note that the predict() function with type=”link” returns predicted logits, while type=”response” returns predicted probabilities…

6. Interpret the 4 coefficients estimates from model “breg3”.

7. Test the significance of the interaction term in “breg3” in two ways. Do both methods agree?

8. Test the goodness of fit for model “breg3”. What can we conclude about this model?

9. Is there evidence of extra-binomial variation (overdispersion) in “breg3”?

10. Regardless of your answer to (9), repeat (7) after adjusting for overdispersion.

11. Compare confidence intervals for the interaction term in “breg3” with and without adjusting for overdispersion, and with and without using profile likelihoods.

12. What are the implications in “breg5” of including DISTANCE as a factor variable? How does this change model interpretations? Does it lead to an improved model?

13. What happens if we expand the data set to contain one row per moth (968 rows)? Now we can run a logistic regression model. How does the logistic regression model “lreg1” compare to the binomial regression model “breg3”? What are similarities and differences? Would there be any reason to run a binomial regression rather than a logistic regression in a case like this?