Fit a generalized linear model

Let's start by ignoring the fact that the data are from multiple years. If time permits we can add year to the model to see how much difference it makes.

1. The response variable is binary. What probability distribution is appropriate to describe the error distribution around a model fit? What is an appropriate link function?

2. Fit a generalized linear model to the data on survival and tarsus length.

3. Use visreg to visualize the model fit. Remember to load the visreg package.

4. Obtain the estimated regression coefficients for the fitted model. What is the interpretation of these coefficients? On a piece of paper, write down the complete formula for the model shown in the visreg plot.

5. Use the coefficients to calculate the predicted survival probability of a song sparrow having tarsus length 20.5 mm. Does the result agree with your plot of the fitted regression curve?

6. The ratio (-intercept/slope) estimates the point at which probability of survival is changing most rapidly. In toxicology this point is known as the LD_{50}. Calculate this value and compare it visually with the fitted curve. Does it agree? Finally, the slope of the curve at a given value for the explanatory variable x is

   $$b \, p(x) \, (1 - p(x))$$

   where b is the slope coefficient of the fitted logistic regression model and $p(x)$ is the predicted probability of survival at that x.

7. Calculate the likelihood-based 95% confidence interval for the logistic regression coefficients.

8. The summary(out) output for the regression coefficients also includes z values and P-values. What caution would you take when interpreting these P-values? Use drop1 to test the null hypothesis of zero slope.

9. If time permits, add year to your logistic regression model as a categorical variable (ensure that year is a factor in your data set). Ignore the interaction between year and tarsus length. Plot the resulting curves using visreg with option xvar="year". Is there any evidence of a difference among years in the relationship between survival and tarsus length in these data? Use emmeans to calculate the model estimates of mean survival in each year.

Crab satellites

The horseshoe crab, *Limulus polyphemus*, has two alternative male reproductive morphs. Some males attach to females with a special appendage. The females bring these males with them when they crawl onto beaches to dig a nest and lay eggs, which the male then fertilizes. Other males are satellites, which are unattached to females but crowd around nesting pairs and obtain fertilizations. What attributes of a female horseshoe crab determine the number of satellite males she attracts on the beaches?

The data here provide measurements of 173 female horseshoe crabs and record the number of satellites she attracted. The data were gathered by Brockman (1996. Satellite male groups in horseshoe crabs, Limulus polyphemus. *Ethology* 102:1-21) and were published by Agresti (2002, *Categorical data analysis*, 2nd ed. Wiley). The variables are female color, spine condition, carapace width (cm), mass (kg), and number of satellite males.

Read and examine the data

1. Read the data from the file. View the first few lines of data to make sure it was read correctly. Use the str command to see the variables and groups.

2. Plot the number of satellites against the width of the carapace, a measure of female body size. Fit a smooth curve to examine the trend.

Fit a generalized linear model

1. What type of variable is the number of satellites? What probability distribution might be appropriate to describe the error distribution around a model fit? What is the appropriate link function?

2. Fit a generalized linear model to the relationship between number of satellite males and female carapace width.

3. Use visreg to examine the relationship on the transformed scale, including confidence bands. This plot reminds u-27(etoitt?)-78 p -14.e