

Homework 5

1. (15 pts)

The file “Solea.txt” is a dataset from a survey of the flatfish *Solea solea*, i.e. sole / Dover sole / black sole. The survey is from the Tagus estuary in Portugal, and was undertaken to better understand what controls sole abundance. Sole has been recorded as presence/absence in the dataset, in the column ‘Solea_solea’. The other columns include many other variables measured at the sites sampled. Later we’ll try a proper exploratory analysis of all this data, but for now just model presence/absence of sole as a function of salinity.

Plot the raw data of sole presence/absence vs. salinity. Break the salinity data into four equally sized bins, and within each bin calculate the proportion of presences. Plot the binned proportions vs. the midpoints of the bins (e.g. if the bin is from 3 to 7, then 5 is the midpoint). Also calculate the standard error of the proportion for each bin, and plot that with error bars.

Fit a binomial glm to test whether presence/absence of sole is driven by salinity. Use `curve()` to plot the fitted logistic relationship on the same plot as the raw data.

What is the effect of salinity on sole presence/absence? Would you consider it a strong effect or a weak effect? Why?

Do a likelihood ratio test to see whether the effect of salinity is significant.

Examine a plot of residuals (deviance residuals or pearson residuals) vs. predicted values. How would you interpret this plot?

2. (15 pts)

The file “heat_shock_subset.csv” includes a subset of the data from the silverleaf whitefly heat shock experiment that you looked at in homework 2 (and the original paper was attached to that assignment).

Each row of the dataset is an experimental replicate. In each replicate, 10 flies were subjected to a heat shock, where they were acclimated to 40°C for 1 hour, then reduced to 25°C for one hour, then shocked at 45°C for 1 hour. The number of survivors at the end is recorded in the column ‘Survival’, and the proportion of survivors is in ‘proportion’. This experiment has a somewhat complex structure, which we’ll analyze fully when we get to mixed models. For now let’s imagine a simpler world where the only important factors are ‘sex’ and ‘region’. The experiment used flies from two regions in Colombia, the Caribbean region (sea level, uniformly hot) and the Southwest region (in the Andes, more variable

temperature). The researchers hypothesized that the flies from these two regions might be locally adapted to thermal conditions, leading to different heat shock tolerances. Experimental replicates were performed on males and females separately, to quantify any effect of sex on heat shock tolerance.

Make an exploratory plot of proportion of survivors vs. sex*region.

Fit a binomial glm where you test for effects of sex and region on the probability of surviving, as well as an interaction between sex and region. Remember there are two different ways to fit a binomial glm where $n > 1$. They are equivalent.

Plot the fitted effects. How do you interpret the results?

Do a likelihood ratio test for the interaction between sex and region.

Oops, we didn't consider overdispersion. Fit the same model using the quasibinomial method. How big is the dispersion parameter?

Test the interaction with the quasibinomial model, using the F-test that accounts for the dispersion parameter. How do you interpret this result, compared to the result from the first model?

Test the main effects of the model (sex and region), using F-tests where the sex*region interaction is no longer in the 'full' model. What's your final take on these results? Are the results consistent with the researchers' hypothesis of local adaptation?