Name: Joseph A. Hediger

HOMEWORK 3: Design Matrices and Maximum Likelihood

WSCI 6390 – 002: Population Parameter Estimation

Due 11:59 PM Tuesday, February 13

Let’s practice (1) building design matrices and (2) getting more comfortable with the differences between probability density/mass functions and maximum likelihood.

INSTRUCTIONS:

There are no R datasets needed for this assignment, though you will need R for a few questions at the end.

ASSIGNMENT:

1. Remember that a linear model is essentially the product of an ***X*** vector and a vector of our Betas (plus a vector of residuals, but let’s ignore that for now). To demonstrate that you understand how matrix math works, can you show me the outcome of multiplying these two matrices together?

=

1. Now let’s put our matrix math in the context of a linear model. Please do the same math, using Betas in the second matrix, and show me the outcome.

=

1. OK, let’s say that we have a dataset with number of species observed as a function of six sites. Can you draw the ***X*** matrix if we are interested in using site 1 as a baseline , and representing the differences between sites 2-6 and site 1, respectively? Also please calculate our matrix, representing our predicted values.

=

1. What if we realize that Site 3 has the greatest number of species and we want to use our betas to determine differences from site 3 rather than site 1? Please adjust the ***X*** matrix. Also please calculate our matrix, representing our predicted values.

=

1. Now we realize, heck, I don’t care about differences among sites, I just want to know average number of detected species per site. This is what “dropping the intercept” looks like. What does the ***X*** matrix look like now? Also please calculate our matrix, representing our predicted values.

=

1. So a forest fire comes through and we are down to **three sites**. Bummer. However, we have data across **two years**! If we have a scenario where we want the intercept to represent site 1 and year 1, and all other betas to represent differences by site and year, respectively, what does the ***X*** matrix look like? Also please calculate our matrix, representing our predicted values.

=

1. Can you describe to me what each Beta represents in regular-speak?
2. Cool. Now let’s say we have those same three sites and annual rainfall for a single year. Can you multiply the ***X*** and matrices together?

=

1. Can you describe to me what each Beta represents in regular-speak?

1. Alright, I’m done with matrix math. Let’s say that you are in a magical land of fluorescent shapeshifting sea snakes with mean length 25 inches and standard deviation 12 inches. Can you use dnorm() in R, as I did in lecture, to tell me what the probability is of finding a magical fluorescent shapeshifting sea snake that has a length of 55 inches?
   1. 0.1%
2. What about 24 inches? Why is this probability higher?
   1. 3.3%
   2. Higher because we are within 1 SD of the mean
3. Can you use pnorm() in R, as I did in lecture, to tell me what the probability is of getting a sea snake that is SMALLER than 18 inches? (a note that you would use lower.tail=TRUE)
   1. 28%
4. In this case, we know the exact parameter values of this sea snake *probability density function* – values that are hard to come across in “real life.” Can you tell me how this scenario differs from one in which we have a sample of 15 snake lengths and want to know the *maximum likelihood estimates (MLEs)* for the mean and standard deviation ?
   1. We want to know the parameters given the data
   2. What is the likelihood of the parameters given the data