

MBA5011 Multivariate Analysis: Model-Based Statistics

Assignment 3

All three questions below use the same data, data (foxes), part of the rethinking package. The urban fox (*Vulpes vulpes*) is a successful exploiter of human habitat. Since urban foxes move in packs and defend territories, data on habitat quality and population density is also included. The data frame has five columns:

- (1) group: Number of the social group the individual fox belong to
- (2) avgfood: The average amount of food available in the territory
- (3) groupsize: The number of foxes in the social group
- (4) area: size of the territory
- (5) weight: Body weight of the individual fox

Question 1 (30%). Fit two bivariate Gaussian regressions, using stan: (1) body weight as a linear function of territory size (area), and (2) body weight as a linear function of groupsize. Plot the results of these regressions, displaying the predicted regression line and the 95% interval of the mean. Is either variable important for predicting fox body weight?

Question 2 (30%). Now fit a multiple linear regression with weight as the outcome and both area and groupsize as predictor variables. Plot the predictions of the model for each predictor, holding the other predictor constant at its mean. What does this model say about the importance of each variable? Why do you get different results than you got in the questions just above?

Question 3 (40%). Finally consider the avgfood variable. Fit two more multiple regressions: (1) body weight as an additive function of avgfood and groupsize, and (2) body weight as an additive function of all three variables, avgfood and groupsize and area. Compare the results of these models to the previous models you've fit, in the first two questions.

- (a) Is avgfood or area a better predictor of body weight? If you had to choose one or the other to include in a model, which would it be? Support your assessment with any tables or plots you choose.
- (b) When both avgfood or area are in the same model, their effects are reduced (closer to zero) and their standard errors are larger than when they are included in separate models. Can you explain this results?