VetPath 571 – Spring 2021
Problem Set 2

The International Capybara Racing League is concerned that the scoring system they use is too easy to game, and that it reflects how many races a Capybara is entered into rather than how fast that Capybara actually runs. They have hired you to determine if there is a clear association between the number of races attended and a Capybara's score for the season.

1) Import 'Scoring.csv' into R, and plot races by score. Does there seem to be a relationship? Include your plot in your answer.

2) Now fit a simple regression predicting score by races. What is the coefficient you obtain, and what does it suggest? Do not just cut and paste R output.

3) Plot your model fit. Do you find it satisfactory? Justify your answer.

4) Fit a model with a squared term for races, and plot the model fit. Does this seem like a better fit for the data? How would you interpret your results?

5) Import the 'mers.csv' data set. This is a data set of MERS cases, with an indicator of whether or not they were fatal infections. Note that this is *not* the complete dataset, and it has been edited to fit the purposes of this exercise.

Examine the dataset. Is logistic regression appropriate to study the determinants of whether or not a MERS case was fatal? Why or why not? Justify your answer.

6) It has been reported that being a healthcare worker was a significant risk for *acquiring* MERs due to high levels of contact with infected patients. Does the data suggest their infections are more severe? Justify your answer.

7) Are men more likely to have their MERS infections end in death than women? Justify your answer.

8) There are several potential confounders in the dataset that may bias the estimate in question 7. Evaluate these, select which variables to include in the model, and discuss your reasoning.

9) Plot the crude OR for a fatal MERS infection in men vs. women from Q3 as well as your final adjusted model including all covariates you chose to keep (if any).