Homework 10

STAT 462 (Fall 2020)

UNAUTHORIZED DISTRIBUTION AND/OR UPLOADING OF THIS DOCUMENT IS STRICTLY PROHIBITED.

Clearly label your answers to each question and each sub-question. Your answers MUST be uploaded to Canvas as a <HWx_Yourfirtname.nb.html> file by the deadline.

(1). With the objective of understanding more about spiders, a researcher has collected data on sand grain size on 28 beaches in Japan and observed the presence or absence of the burrowing wolf spider on each beach. Resulting data is in spider.csv file.

(a). What is the response and the predictor variable here?

(b). Is the response ordinal or nominal categorical variable?

(c). What is the estimated regression equation?

(d). Is the sand grain size a significant predictor in determining the presence of wolf spiders in the beaches of Japan? Show all your work answering this research question.

(e). Interpret the regression coefficients in the context of the problem.

(f). Calculate the odds of a certain beach that has sand grain size 0.35 being a wolf spider habitant?

(g). Obtain a 95% confidence interval for the odds ratio. Is your observation consistent with your conclusion from the test you did in part (d). Clearly explain the relationship between two approaches in your answer rather than simply giving a "yes" or "no" answer.

(2). In 1846, the Donner party (Donner and Reed families) left Springfield, Illinois for California in covered wagons. After reaching Fort Bridger, Wyoming, the leaders decided to find a new route to Sacramento. They became stranded in the eastern Sierra Nevada mountains at a place now called Donner Pass (right) when the region was hit by heavy snows in late October. By the time the survivors were rescued on April 21, 1847, 40 out of 87 had died. Data is provided in *Donner.csv* file. Columns of the dataset indicate the age of the person, gender (male=1, female=0) and the survival status (survived=1, dead=0)

(a). Obtain a regression model that would estimate the probability of survival as a function of age and gender. Write the equation of the fitted regression model.

(b). Check the overall goodness of fit of the model.

(c). Interpret the regression coefficients in the context of the problem

(d). After taking age into account, are women more likely to survive harsh conditions than men? Do an appropriate test to answer this research problem. Show all your work including the null and alternative hypothesis to answer this question.

(e) Calculate the odds ratio of survival for men compared to the women for a given age.