Homework Assignment: Week 6

Assignment Due: Sunday, February 13 by Midnight CT

Submit electronic solutions to drop box in BrightSpace

1. Using the mtcars data, build a linear model that will predict weight (wt) from horsepower (hp) + cylinders(cyl) + displacement (disp). Use this model to predict the weight of the cars below and include the confidence intervals for the predictions. Include your R code with your answers

|  |  |  |  |
| --- | --- | --- | --- |
| **Car** | **hp** | **cyl** | **disp** |
| 1 | 110 | 4 | 108 |
| 2 | 170 | 8 | 400 |
| 3 | 175 | 6 | 145 |

2. Explain why even if we know the exact values of the beta’s, that is the population values, that there will still be some variance in the predicted value.

3. A flight from Kansas City to Denver has an on-time performance of 58%. Using R, find the probability that among 15 flights, 12 will arrive on time.

4. Find the maximum likelihood estimate of p in a binomial distribution characterized by 9 successes in 20 trials. Show graphically that this is the maximum. Include your R code with your answers.

5.A young pitcher has three quality pitches, a slider, a curve, and a fast ball. In a particular game, the pitcher has thrown 15 sliders, 31, curves, and 40 fast balls. Given only this information, what are the odds the next pitch will be a curve ball? What is the probability the next pitch will be a slider?

6. Using the heart disease data from the UCI site (the one used in the Logistic Regression (LR) Example), and separate the data set into Train and Test sets (70/30 split). A) Build an LR model and compare its AIC to model using the full data set. B) Test the model and determine it accuracy for predicting healthy/unhealthy patients. Compare this to the model using only the sex variable. Include your R code with your answers.

7. Using the mtcars data set, construct a logistic model (use all of the data) to predict engine configuration, vs, [where 0 = V-shaped and 1 = Straight] as a function of horsepower (hp). A) Provide an interpretation of the betas. B) Calculate the odds of the engine type at zero hp. C) Calculate the odds of the engine type at hp = 240. Include your R code with your answers.

8. Using the output from problem 7, are the beta’s statistically significant? Explain how the Wald Test can be used to determine statistical significance. Include your R code with your answers.

9. Build a decision tree using the Titanic dataset that is in the data module. Predict survival as a function of passenger class (PClass) and Sex. You will need to convert both PClass and Sex to factor variables. Also, the text() command I provided is a bit dated and will use to letters to identify levels of variables with a being the first level, be the second etc. So, Sex = b is displayed, b= Male. For PClass a is 1st class, b second class, etc. Provide an analysis of the data centering on who/what passengers were likely to survive.