HW 3

1. Suppose you run an A/B test on a marketing campaign. On day one, 123 out of 500 customers convert (make a purchase) in the new campaign, whereas 63 out of 300 convert on the old campaign. Use a two sample proportions test in R to compute uncertainty in the results. What is the p-value?
    a. .1
    b. .16
    c. .22
    d. .28
2. Based on your answer to question 1, what would you recommend the marketing team do?
3. Use the nycflights13 data and calculate the mean and standard error for the dep_delay (first remove NA from the column). Use the central limit theorem to construct a 95% confidence interval for dep_delay mean. What interval do you get?
    a. [12.40, 12.87]
    b. [12.50, 12.77]
    c. [12.30, 12.97]
    d. [12.20, 13.07]
4. Repeat question 3 but use the bootstrap instead. E.g. resample the values of dep_delay with replacement and calculate the mean for each resampled dataset (do this 10000 times). The 10000 resulting means are called the bootstrap distribution. Next calculate the 95% quantiles for the bootstrap distribution and report them, along with the R code you used. How do they compare to 3?
5. Using the Beerwing.csv data provided on canvas, estimate a linear model with normal errors of the amount of beer consumed for each individual vs. hotwings consumed and gender. What is the coefficient of hotwings?
    a. 2.02
    b. 2.04
    c. 2.06
    d. 2.08
6. Next use the bootstrap to estimate 100 models on resampled data and for each model predict the amount of beer consumed by a male individual who ate 12 wings. What is the max and min predicted number of beers consumed over the 100 models? Report these values along with your code.