STAT 308 – Take Home Exam

This take home exam is worth a total of 50 points, and counts for half of your Exam 1 grade, along with the in class portion of the exam. For the problems in which calculations are needed, please include your `R` code with your answers. Turn in this exam to Sakai by Thursday, October 6 at 2:30 pm.

1. For the following question, we will define

   - β_0 and β_1 as the population intercept and slope for the linear model with Y as the response and X as the explanatory variable.

   For a random sample of $n = 40$, we obtain the following sample statistics:

   $$\bar{x} = 2, \bar{y} = -3, s_x = 2.5, s_y = 4.5, r = -0.6.$$

   a. [2 pts] Calculate estimates of the least squares regression line, $\hat{\beta}_0$ and $\hat{\beta}_1$.

   b. [2 pts] Calculate the sum of squared errors, SSE, and the estimate of the regression variance, $s^2_{Y|X}$. (Hint: recall that SSE is the percent of the total sums of squares not explained by the linear model).

   c. [2 pts] Calculate the standard errors of the regression estimates, $s_{\hat{\beta}_0}$ and $s_{\hat{\beta}_1}$.

   d. [2 pts] Calculate a test statistic and p-value for testing $H_0 : \beta_1 = 0$ vs. $H_a : \beta_1 > 0$.

For the following questions, we will use the dataset `nfl` contains information about the regular season games from the 2021 regular season, including

- `Week`: The Week the game took place
- `Day`: The Day of the week the game took place
- `Date`: The Date of the game
- `Time`: The time the game started (in Eastern Time)
- `AwayTeam`: The game's away team
- `HomeTeam`: The game's home team
- `YdsDiff`: The difference in the offensive yards between the home team and away team
- `TODiff`: The difference in turnovers between the home team and away team
- `PtsDiff`: The difference in points scored between the home team and away team

2. Suppose we think that the point difference is not dependent on any other explanatory variables.

   a. [2 pts] What do we expect the point differential between a randomly selected home and away team to be?

b. [4 pts] Calculate a 99% confidence interval for the expected point differential. Interpret this interval in the context of the problem.

c. [4 pts] Bookmakers often maneuver their point spreads to account for the belief that, on average, home teams win by three points. Perform a hypothesis test on the counterargument that home teams do not, on average, win by three points. Be sure to formally state all the necessary information to perform a formal hypothesis test. Assume $\alpha = 0.05$.

3. Now, suppose we think that the observed point differentials can be better explained through a linear relationship with the difference in offensive yards gained.

a. [2 pts] Create a scatterplot of yard differential vs. point differential. Comment on the validity of a linear relationship.

b. [2 pts] Calculate a least squares regression line using R. Formally state the regression line.

c. [2 pts] Determine if there are any influential observations in the given dataset. If you determine that there are, state your new least squares regression line.

For the remainder of this exam, use either the regression line in (b) if you determine there are no influential observations, or the line in (c) if you determine that there are influential observations.

d. [2 pts] What is the estimate of the regression variance?

e. [2 pts] Interpret the estimate of the population slope in the context of the problem.

f. [2 pts] Interpret the estimate of the population intercept in the context of the problem. Does this interpretation make sense? Why or why not?

g. [2 pts] Determine if the assumption of heteroscedasticity is violated.

h. [2 pts] Determine if the assumption of normally distributed residuals is violated.

i. [4 pts] Find the r^2 for this linear regression model. Interpret the r^2 in the context of the problem.

j. [4 pts] Calculate a 99% confidence interval for the population slope. Interpret this interval in the context of the problem.

k. [4 pts] Perform a hypothesis test for a positive linear relationship between yard differential and point differential. Be sure to formally state all the necessary information to perform a formal hypothesis test. Assume $\alpha = 0.01$.

l. [4 pts] Calculate a prediction as well as a 90% prediction interval for the point differential when the home team has 100 more yards than the away team. Interpret the prediction interval in the context of the given problem.