**DATA 557**

**Winter 2022**

**Homework Assignment 6**

**Instructions**

Submit your solutions **in pdf format** to the dropbox on the canvas page by **1:00PM, Friday March 4**. You may use any program to generate your pdf file. (RStudio is recommended but not required.)

For each question you will be given 1 point for complete credit, ½ point for partial credit, and 0 points for no credit. Assignment of credit will be based on the correctness of your answers as well as your reasoning (when requested as part of the question). You do not need to submit R code for this assignment except where it is requested.

You may work together to help each other solve problems, but you should create your own solutions and hand in your own work without copying others’ work.

**Data: “Sales\_sample.csv” (same one as used in HW 5).**

1. Fit the linear regression model with sale price as response variable and SQFT, LOT\_SIZE, BEDS, and BATHS as predictor variables (Model 1 from HW 5). Calculate robust standard errors for the coefficient estimates. Display a table with estimated coefficients, the usual standard errors that assume constant variance, and robust standard errors.

2. Which set of standard errors should be used? Explain by referring to HW 5.

3. Perform the Wald test for testing that the coefficient of the LOT\_SIZE variable is equal to 0. Use the usual standard errors that assume constant variance. Report the test statistic and p-value.

4. Perform the robust Wald test statistic for testing that the coefficient of the LOT\_SIZE variable is equal to 0. Report the test statistic and p-value.

5. Use the jackknife to estimate the SE for the coefficient of the LOT\_SIZE variable. Report the jackknife estimate of the SE.

6. Use the jackknife estimate of the SE to test the null hypothesis that the coefficient of the LOT\_SIZE variable is equal to 0. Report the test statistic and p-value.

7. Do the tests in Q3, Q4, and Q6 agree? Which of these tests are valid?

8. Remove the LOT\_SIZE variable from Model 1 (call this Model 1A). Fit Model 1A and report the table of coefficients, the usual standard errors that assume constant variance, and robust standard errors.

9. Add the square of the LOT\_SIZE variable to **Model 1** (call this Model 1B). Fit Model 1B and report the table of coefficients, the usual standard errors that assume constant variance, and robust standard errors.

10. Perform the F test to compare Model 1A and Model 1B. Report the p-value.

11. State the null hypothesis being tested in Q10 either in words or by using model formulas.

12. Perform the robust Wald test to compare Model 1A and Model 1B. Report the p-value.

13. Compare the results of the tests in Q10 and Q12. Which test is valid?

The following questions use the LOG\_PRICE variable as in HW 5. Fit models corresponding to Model 1A and Model 1B with LOG\_PRICE as the response variable. Call these models Model 1A\_Log and Model 1B\_Log.

14. Perform the F test to compare Model 1A\_Log and Model 1B\_Log. Report the p-value.

15. State the null hypothesis being tested in Q14 either in words or by using model formulas.

16. Perform the robust Wald test to compare Model 1A\_Log and Model 1B\_Log. Report the p-value.

17. Compare the results of the tests in Q14 and Q16. Do they give the same conclusion?

18. Based on all of the analyses performed, answer the following question. Is there evidence for an association between the size of the lot and sales price? Explain.