Group number: \_\_\_\_\_\_\_

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Last Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ First Name:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Intro to Stats Homework for Week 3

Module 5

Due Friday before the exam

This assignment will cover Sessions 11-14 and be worth 100 points.

You may submit your answers electronically with this sheet via MyCourses, or on paper in class to me on Friday. If you submit answers on paper, please include all associated student names, write legibly, and preserve the order of the questions as listed below.

Please make your submission reasonably convenient for me to grade. That means providing your answers in the correct order and making it clear what part of what you write is your final answer.

**Background:** You have been assigned to your company’s branch office in Atlanta, Georgia. Your boss wants to add a car to the company’s fleet of vehicles in Atlanta but wants a very specific type of car for important executives to drive to meetings. Your assignment is to find a value-priced BMW model 7632 in the secondary market (used car market) to recommend for purchase for the fleet. A value-priced BMW7632 is one that is priced below its expected market value, given various attributes of the vehicle. You have gathered the data for 44 BMW7632s that you could find available for sale online, all from the same model year, and it is in the Excel file BMW7632\_2021.

First, examine the data in Excel file (BMW7632\_2021), noting the definitions of the variables in the table starting in cell R4 on the “cars” tab of the spreadsheet.

Now, let’s explore the connection between mileage and asking price. Construct and run the appropriate regression. Given that mileage is measured in miles and price is measured in dollars, it’s possible that the coefficient might be fairly small. To aid in interpretation, might it be wise to scale the variables? (Hint: yes, it is wise).

**Session 11: Simple Regression (25 points)**

**1) Briefly (in no more than 100 words), report and discuss the coefficient and its interpretation in the appropriate simple regression described above. (You will discuss statistical significance in the next question so no need for that discussion here.)**

**2)** Regressions will always generate coefficients regardless of how impactful or correlated the independent and dependent variables are. But does X really have associated impact on Y?

**Briefly (in no more than 100 words), discuss the statistical significance of your independent variable.** Feel free to discuss it in “stats language”, but also explain it as you might to someone who has never taken a stats class.

**3)** **How much of the variation in the dependent variable is explained by the independent variable?** **Is this a lot or a little? Are you surprised by the result? Explain. Again, please keep to the 100 word limit for the answer.**

**Session 12 (Multiple Regression) (25 points)**

**1) The regression above was pretty simple. Using what we now know about multiple regression, construct and report a more inclusive model below. Note that you do not *necessarily* have to include every available variable. I want you to build the best model you can with the data available.**

**After you report the model below, briefly explain why you built it that way (i.e. justify what you did). Limit your discussion to under 300 words.**

**2) Report the estimated results of the model below and then in under 400 words, summarize the results. In your explanation, parts can be technical (i.e. discussing statistical terms like alpha, beta, statistical significance, r squared, etc.) but be sure that at least the main take-aways would be comprehensible by a general audience.**

**Session 14 (50 points)**

For this final part of the final homework, you’ll generate and answer a research question.

First, go to <https://think.cs.vt.edu/corgis/csv/>

This website is a great collection of interesting data sets.

Once you pick one that looks interesting, generate a research question (it doesn’t have to be extraordinarily profound, in fact, I love interesting/quirky ideas), construct a regression, estimate it, and report your results.

1) What data set are you using?

2) What is the research question?

3) What is the unit of observation?

4) What is the regression model you will estimate?

5) What are the results? (usually a “Copy and paste” looks lazy and cheap, but that would be fine here)

6) In no more than 400 words, briefly summarize the results.

7) Are there any potential statistical issues with your results? In fewer than 100 words, briefly explain.