STA 112
Spring 2022

Exam 2

Dr. Evans
Exam 2

Last Name: _____ **First Name:** _____

I hereby state that I have not communicated with or gained information in any way from other students or any outside resource during this exam. I agree to abide by the rules stated below, and to abide by the Wake Forest Honor Code. All work is my own. I understand that any violation of this agreement will be reported to the Honor Council and will result, at minimum, in a 0 on this exam.

Signature : _____

All work on this exam must be your own.

1. You have 50 minutes to complete the exam.
2. Show all your work on the open ended questions in order to get partial credit. No credit will be given for open ended questions where no work is shown, even if the answer is correct.
3. You are allowed a calculator, however you may not share a calculator with another student during the exam. The calculator must be only a calculator, and may not be connected to the internet.
4. You are allowed to ask clarification questions to me, but you may not ask anyone else.
5. You are **not** allowed a cell phone, even if you intend to use it as a calculator or for checking the time. You are **not** allowed a music device or headphones, notes, books, or other resources.
6. You may **not** communicate with anyone other than myself during the exam.
7. Write clearly and be clear. Make it easy to find your answers.

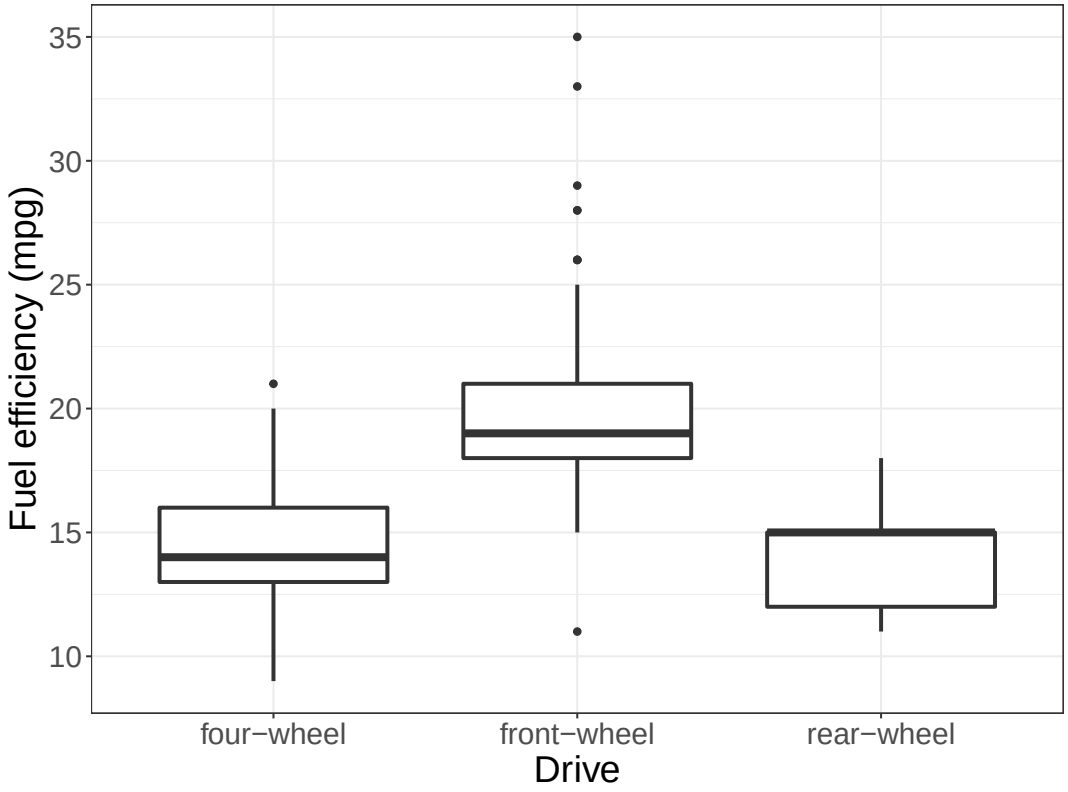
Good luck!

The Data We have a client who is interested in examining which factors influence the fuel efficiency of different cars. The client is curious about the relationship between a car's *engine displacement* (a measure of how big the engine is, measured in liters) and its fuel efficiency (measured in miles per gallon). The client also suspects fuel efficiency may be related to whether the car is front-wheel drive, rear-wheel drive, or four-wheel drive.

We have information on 234 different cars, and the following three variables:

- `mpg`: the car's fuel efficiency, in miles per gallon
- `displacement`: the car's engine displacement (size), in liters
- `drive`: either front-wheel, rear-wheel, or four-wheel

Below are side-by-side boxplots showing the distribution of fuel efficiency for each drive type.



We will begin by looking at the relationship between drive type and fuel efficiency.

1. Write down a population model with drive type as the predictor, and fuel efficiency as the response. Use appropriate notation. (Ignore engine displacement for now).

2. The client fits your model, resulting in the following model output. Using the fitted model, report the estimated average fuel efficiency for cars of each drive type (four-wheel, front wheel, and rear-wheel).

	Estimate	Std. Error
(Intercept)	14.33	0.314
driveFrontWheel	5.64	0.442
driveRearWheel	−0.25	0.71

3. The client wants to test whether there is a relationship between drive type and fuel efficiency. Using your model from Question 1, write down null and alternative hypotheses, in terms of one or more model parameters, for this research question.

4. To test the hypotheses from Question 3, your client starts to create an ANOVA table. Unfortunately, after filling in the sums of squares, the client gets stuck. Help them out by filling in the rest of the ANOVA table. (You can write the table separately if you need more space).

Source	**degrees of freedom (df)**	**Sum of Squares (SS)**	**Mean Squares (MS)**	**F**
Model		1878.8		
Residual		2341.5		
Total				

5. Using your ANOVA table, what is the value of the test statistic you will use to test your hypotheses from Question 3? What distribution will you compare your test statistic with to calculate a p-value (give the name and degrees of freedom for the distribution).

6. The test statistic you calculated corresponds to the data shown in the boxplots on page 1. What could those boxplots look like if the test statistic in Question 5 was *smaller*? Draw a sketch, and explain your reasoning.

The client is now interested in adding engine displacement to the model. In particular, the client wants to know whether a change of 1 liter in engine displacement is associated with the same change in fuel efficiency for four-wheel drive cars, front-wheel drive cars, and rear-wheel drive cars (i.e., is the relationship between displacement and fuel efficiency the same for each drive type).

7. Write down a population model that allows the client to answer this research question. Use appropriate notation, and explain your reasoning for choosing the model.

8. The client fits your model, resulting in the following model output. Interpret the fitted coefficient -2.07 in context.

	Estimate	Std. Error
(Intercept)	22.59	0.814
driveFrontWheel	6.13	1.163
driveRearWheel	−3.01	3.104
displacement	−2.07	0.196
driveFrontWheel:displacement	−1.35	0.370
driveRearWheel:displacement	1.00	0.605

9. The client wants to know whether a change of 1 liter in engine displacement is associated with the same change in fuel efficiency for each drive type. Using your model from Question 7, write down null and alternative hypotheses, in terms of one or more model parameters, for this research question.

10. Which test should the client use to test these hypotheses? (nested F test, t test, or either?) Explain your reasoning, and give the degrees of freedom for the test.

11. Now you want to check assumptions to make sure your hypothesis tests are valid. Explain to the client which diagnostic plot you will use to assess the shape and constant variance assumptions. Then, sketch an example diagnostic plot in which both the shape *and* constant variance assumptions are violated. Make sure to label the axes.

To compare nested regression models we use a nested F test, where the test statistic is given by

$$F = \frac{\frac{1}{\#\text{parameters tested}}(SSE_{reduced} - SSE_{full})}{\frac{1}{n-p}SSE_{full}}$$

where n is the number of observations in the data, and p is the number of parameters in the full model. $SSE_{reduced}$ and SSE_{full} are the residual sum of squares for the reduced and full models; recall that

$$SSE = \sum_{i=1}^{n}(y_i - \widehat{y}_i)^2 \qquad SSModel = \sum_{i=1}^{n}(\widehat{y}_i - \overline{y})^2 \qquad SSTotal = \sum_{i=1}^{n}(y_i - \overline{y})^2$$

and $SSTotal = SSModel + SSE$.

12. The F test you calculated with the ANOVA table in Question 4 is a special case of the nested F test, when we are testing whether there is a relationship between a single categorical predictor and a quantitative response. Using the formula for the nested F statistic given here, explain why.

You are done!!! Whooo!!!