ECN 3972   Name: _____
Exam 2

READ THIS SECTION CAREFULLY.

This exam is for <u>your</u> work ONLY. You may use:
1. Your notes
2. Your book
3. The class videos (made by me)
4. A calculator

You may <u>not</u> use:
1. Your computer (other than downloading, printing, or reading the exam). Note: If you really want, you can type your answers in Word, but I don't really think you will want to with all of the equations.
2. Excel or any other spreadsheet program
3. The internet except D2L notes and videos
4. Other individuals

Failure to follow these guidelines will result in:
1. A grade of zero on the exam
2. A report of academic misconduct to the Office of Student Standards

Please believe me when I say that I take this seriously. I need to be able to trust that this is your work. Just because this is an online class does not change anything – you should treat this as you would an in-class exam where you are allowed to use your book and notes.

Completion of this exam means you have acknowledged the rules listed above and will act accordingly.

You will upload your answers either in Word or a pdf the same way as you submit your homework assignments in ONE document. NO PICTURES! Make sure it is in order!!

Due date: Tuesday, March 30, 2021 at 11:00 pm. No late exams will be accepted!

Please make sure your answers are written clearly and IN ORDER. Make sure every question is labeled with the proper number. Read each question carefully. Make sure your answer is written with the appropriate level of detail. For each hypothesis test, you must write both the null and alternative hypotheses, the decision rule, the calculated test statistic, the critical value of the test statistic and then (finally) the decision. You must have ALL parts! (Don't assume that I will find it – make it clear.)

<u>**Show all work if you want any partial credit.**</u> Point values are in parentheses.

1. You are hired as a consultant to Martin's Dry Cleaners to determine the factors that affect revenues at each of their 35 locations. You estimate a model using OLS with a cross-section data. The estimated model is:

$$Rev_i = \beta_0 + \beta_1\ Pop_i + \beta_2\ Income_i + \beta_3 Comp_i + \varepsilon_i$$

where:
$Sales_i$ = revenue for location i (thousands of $)
Pop_i = population within 2 miles of where location i is located (in hundreds)
$Income_i$ = median income of population in within 2 miles of where i is located
     (thousands of $)
$Comp_i$ = number of competitors within 2 miles of location i

The estimated results are:

Coefficients:		
	Estimate	Std. Error
(Intercept)	34015.78	12770.09
Pop	0.8744	0.3982
Income	0.32778	0.1145
Comp	-1.772	0.911

R-squared: 0.6412	Adjusted R-squared: 0.6129	
F-statistic: 17.22 on 3 and 31 DF		

(6) a. Test the overall significance of the model at the 5% significance level.

(4) b. Provide an interpretation of the coefficient on $Comp_i$.

(6) c. It is believed that dry cleaning is a normal good, meaning that as income rises, people use dry cleaning more. Test this hypothesis at the 5% significance level.

(6) d. At the same time, we believe that the number of competitors within 2 miles of Martin's will reduce revenues. Test this hypothesis (at the 5% level).

(4) e. The R^2 of the model estimated is 0.6412. What does this number tell us? Be specific.

2. Suppose you are interested in determining the factors that affect the average price of a one-minute commercial during the Super Bowl over time. You gather information on the past 35 Super Bowls and estimate the following model:

$$Price_t = \beta_0 + \beta_1 Number_t + \beta_2\, Pats_t + \beta_3\, City_t + \beta_4\, Wins_t + \varepsilon_i$$

where:
$Price_t$ = average price per minute commercial in year t (in thousands of dollars)
$Number_t$ = number of ads sold during the Super Bowl in year t
$Pats_t$ = 1 if the New England Patriots are playing in year t; = 0 otherwise
$City_t$ = 1 if a team from NY or LA is playing in year t; = 0 otherwise
$Wins_t$ = average number of wins of the two teams playing in year t

The results from the estimation are as follows:

Coefficients:

	Estimate	Std. Error
(Intercept)	12.322	2.566
Number	-167.22	67.998
Pats	5.998	3.872
City	22.672	8.771
Wins	21.771	7.832

R-squared: 0.3466, Adjusted R-squared: 0.3387
F-statistic: 15.88 on 4 and 30 DF

(6) a. It is believed that, as the number of advertisement slots available rises, the average price will fall. Test this hypothesis at the 5% level.

(4) b. Interpret the coefficient on $Pats_t$. (Be careful of units!)

(6) c. People want to watch teams who have more wins during the season. This means that advertising prices for the Super Bowl should be higher as the average number of wins for the teams rises. Test this hypothesis at the 5% level.

(6) d. Having a team from LA or NY is expected to raise the average price of an advertisement by $20,000. Test this hypothesis at the 5% level.

3. You are estimating a model of the annual number of hours worked by women in the labor force. Given 100 observations, you estimate the following:

$$Hours_i = \beta_0 + \beta_1 AvgWage_i + \beta_2 Educ_i + \beta_3 Married_i + \varepsilon_t$$

where:
- $Hours_i$ = annual hours worked by the ith observation
- $AvgWage_i$ = average wage earned by observation i during the year (in dollars)
- $Educ_i$ = years of education for the ith observation
- $Married_i$ = 1 if observation i is married; = 0 otherwise

The estimated results are:

Coefficients:

	Estimate	Std. Error
(Intercept)	1102.4	292.88
AvgWage	50.778	22.982
Educ	17.991	6.992
Married	-159.77	54.912

R-squared: 0.7100, Adjusted R-squared: 0.6866

F-statistic: 59.99 on 3 and 96 DF

(4) a. Do each of the coefficients for the three independent variables have the signs you would expect? Explain.

Suppose you believe that having children in the home will reduce a woman's work hours. You decide to test this claim by adding another independent variable to your model that measures the number of children under age 18 living in the home. Your new model is:

$$Hours_i = \beta_0 + \beta_1 AvgWage_i + \beta_2 Educ_i + \beta_3 Married_i + \beta_4 Kids_i + \varepsilon_t$$

where:
- $Kids_i$ = the number of children under age 18 living with observation i
- All other variables are as listed above.

The estimated results are:

Coefficients:

	Estimate	Std. Error
(Intercept)	1287.56	302.88
AvgWage	50.778	22.982
Educ	17.991	6.992
Married	-159.77	54.912
Kids	-98.662	58.72

R-squared: 0.7589, Adjusted R-squared: 0.6812

F-statistic: 64.28 on 4 and 95 DF

(4) b. Interpret the coefficient on $Kids_i$.

(6) c. Test if having children in the home reduces a woman's hours worked. (Test at the 5% level.)

(8) d. Which of these two models is better? Carefully use the four specification criteria to make your decision and also to explain your answer.

4. Let's return to the model we were just using at the end of Question 3:

$$Hours_i = \beta_0 + \beta_1 AvgWage_i + \beta_2 Educ_i + \beta_3 Married_i + \beta_4 Kids_i + \varepsilon_t$$

where:

      $Hours_i$ = annual hours worked by the ith observation
      $AvgWage_i$ = average wage earned by observation i during the year (in dollars)
      $Educ_i$ = years of education for the ith observation
      $Married_i$ = 1 if observation i is married; = 0 otherwise
      $Kids_i$ = the number of children under age 18 living with observation i

The results from the estimation are as follows:

Coefficients:

	Estimate	Std. Error
(Intercept)	1287.56	302.88
AvgWage	50.778	22.982
Educ	17.991	6.992
Married	-159.77	54.912
Kids	-98.662	58.72

R-squared: 0.7589, Adjusted R-squared: 0.6812
F-statistic: 64.28 on 4 and 95 DF

RSS = 1812.71

(4) a. Test the overall significance of the model at the 5% significance level.

Suppose that you believe that the effect of having kids in the home will be larger if the kids are under the age of 5. Therefore, you divide the kids independent variable into two separate independent variables, Under5 and Over5 where:

      $Under5_i$ = the number of children under age 5 living with observation i
      $Over5_i$ = the number of children between ages 5 and 18 living with observation i

You change your model to:

$$Hours_i = \beta_0 + \beta_1 AvgWage_i + \beta_2 Educ_i + \beta_3 Married_i + \beta_5 Under5_i + \beta_6 Over5_i + \varepsilon_t$$

This is an **_unrestricted_** model because it allows the coefficients on Under5 and Over5 to be different from one another. In our previous model, we required these values to be the same (the coefficient on $Kids_i$). Therefore, the first model is the **_restricted_** model. (There is one restriction because the two coefficients are restricted to being equal to one another.) The results from this second model are:

<div style="border:1px solid black; padding:10px;">

Coefficients:

	Estimate	Std. Error
(Intercept)	1287.56	302.88
AvgWage	50.778	22.982
Educ	17.991	6.992
Married	-159.17	54.912
Under5	-162.77	62.72
Over5	-21.88	13.85

R-squared: 0.7922, Adjusted R-squared: 0.7710
F-statistic: 71.22 on 5 and 94 DF

RSS = 1629.22

</div>

(4) b. Create a 95% confidence interval for β_5.

(6) c. Perform a test to see if it is best to separate the effects of having kids in the house by the age of the kids. (In other words, test the restriction that $\beta_5 = \beta_6$.)

5. For the following models, write out the interpretation of the indicated variable:

(4) a. $Y_i = 35 - 15.2X_1 + 11.9X_2$

Interpret the coefficient on X_2.

(4) b. $lnY_i = 1,422 + 3.22lnX_1 - 19.88X_2$

Interpret the coefficient on lnX_1.

(4) c. $lnY_i = 887.2 - 0.836X_1 + 0.451X_2$

Interpret the coefficient on X_2.

(4) d. $Y_i = 377 - 1,922lnX_1 + 21.77X_2$

Interpret the coefficient on lnX_1.