Name: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

ECON/ACCT/BUSA 222—Bethany College

**Exam 03**

* There are 110 possible points on this exam. The test is out of 100.
* You have two hours to complete this exam, but you should be able to complete it in less than that
* Please turn off all cell phones and other electronic equipment.
* You are allowed a calculator for the exam. This calculator cannot be capable of storing equations. This calculator cannot double as a cell phone.
* Be sure to read all instructions and questions carefully.
* Remember to show all your work.
* Recall basic logic. “Water is wet” is a true statement. “Water is wet and leopards have stripes” is a false statement.
* You are allowed one 3” by 5” note card with the exam. You are allowed any information you deem important on it. You can also interpret this as having 30 square inches of paper (two 15 square inches, one for each side of the note card).
* *Please print clearly and neatly.*

**Part I: Multiple Choice.** *Choose the best answer to the following.*

4 points each.

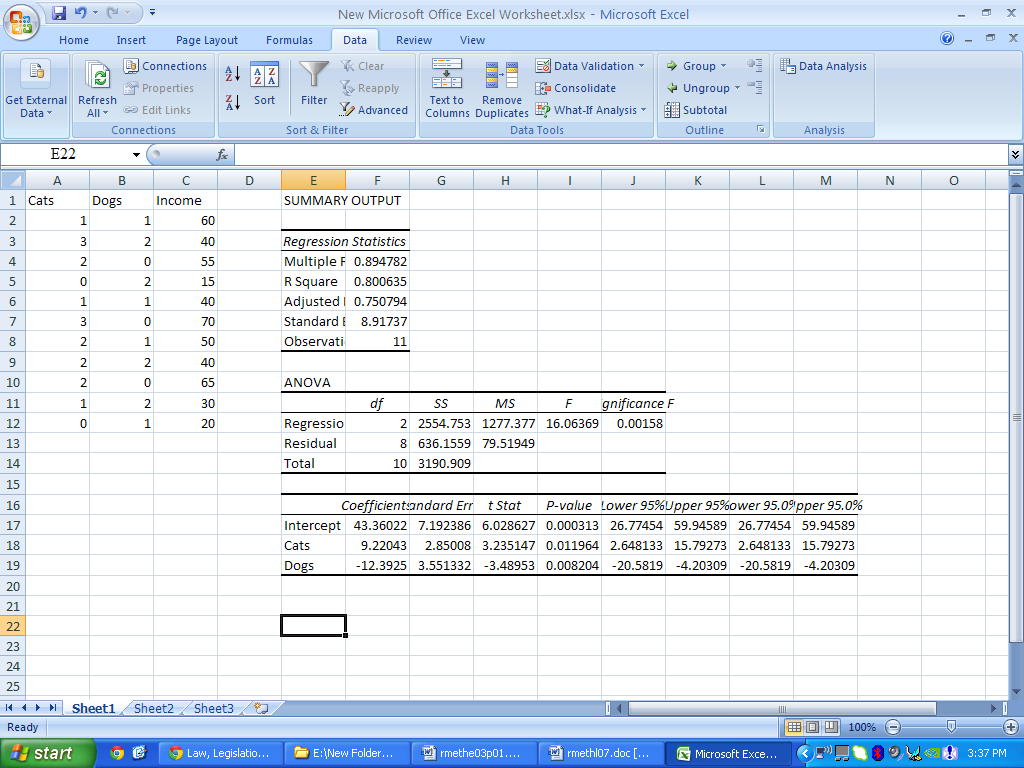
1. Suppose you run a regression with an earner’s income predicting his/her children’s level of education. If you add the earner’s level of education as an explanatory variable, what problem, if any, will you get?
   1. Multicollinearity
   2. Heteroscedasticity
   3. Reverse causation
   4. B & C
   5. None of the above / There is no problem
2. What does R2 represent?
   1. The fraction of the variation the regression explains
   2. Explained Sum of Squares divided by Residual Sum of Squares
   3. Explained Sum of Squares divided by Total Sum of Squares
   4. A & C
   5. None of the above
3. Which of the following would best be represented with a dummy variable (or series of dummy variables)?
   1. Age
   2. Gender
   3. Color of shirt
   4. B & C
   5. None of the above
4. Suppose you ran a regression of AGE predicting INCOME (thousands of dollars). If your estimated line was INCOME = -1.5 + 1.2\*AGE + ε, how much more income do you expect to make for every *two* years you age?
5. -0.3
6. 0.9
7. 1.2
8. 2.4
9. It is impossible to tell given the information provided
10. Suppose you ran a regression and found heteroscedasticity. What should you do first?
11. Start over with new variables
12. Drop as few variables as possible to remove the heteroscedasticity
13. Check how significant your explanatory variables are
14. A & B
15. None of the above
16. As you add explanatory variables to a regression, what always happens?
    1. Your F-stat falls.
    2. The p-values of the variable(s) you started with decrease.
    3. R2 increases.
    4. A & B
    5. None of the above
17. Tyron is interested if the strategy a player prefers in a Rock-Paper-Scissors tournament can be used to predict the player’s age. Assume all players have one and only one favored strategy. Tyron gathers and records his data (the first variable asks if the player prefers Rock, etc.) and a section of the output is indicated below.

|  |  |  |  |
| --- | --- | --- | --- |
| **Age** | **Strategy** | | |
| *Rock?* | *Paper?* | *Scissors?* |
| 14 | 1 | 0 | 0 |
| 18 | 0 | 0 | 1 |
| 12 | 1 | 0 | 0 |
| 31 | 0 | 1 | 0 |
| … | … | … | … |

What, if anything, is wrong with how Tyrone recorded his data?

* 1. He has too many variables.
  2. He will get heteroscedasticity.
  3. He has the same player having more than one favored strategy.
  4. A & C
  5. None of the above / Nothing is wrong with it.

1. Suppose you ran a regression with the natural log (ln) of income (LN(INCOME)) predicting the square footage of that person’s apartment (SQUARE FT) and the result is: SQUARE FT = 600 + 130(LN(INCOME)) + ε. According to your model if a person’s income increased by 1%, how much larger do you predict their house to be?
2. 131.3
3. 731.3
4. (1.01)130 – 1
5. 600 + (1.01)130 – 1
6. It is impossible to tell given the information provided
7. When testing if several different groups have the same mean, why don’t we just take the t-statistic for each pair?
   1. You always use the z-test, not the t-test.
   2. It can take too long.
   3. We might have Type I error.
   4. B & C
   5. None of the above
8. Suppose the coefficient for an independent variable—number of cars per person—is 1.5. Suppose you changed the independent variable so it considers the number of cars per 100 people. What is the coefficient now?
   1. 0.015
   2. 0.15
   3. 15
   4. 150
   5. It is impossible to tell with the information provided
9. Which of the following is an example of a scalar?
   1. A country’s population
   2. The number, pi.
   3. 10 million
   4. B & C
   5. None of the above
10. Consider the following output from Excel, with cats and dogs predicting income in thousands of dollars (the data here are made up):



According this output, what is the estimated line (all values are rounded to three decimal places)?

* 1. INCOME = 0.801 + 0.751 \*CATS + 16.064\*DOGS + ε
  2. INCOME = 0.002 + 0.012\*CATS + 0.008\*DOGS + ε
  3. INCOME = 16.064 + 9.220\*CATS – 12.393\*DOGS + ε
  4. INCOME = 43.360 + 9.220\*CATS – 12.393\*DOGS + ε
  5. None of the above

**Part II: True/False.** *Answer true or false, and justify your answer.*

10 points each.

1. A linear regression with an “error term” (a.k.a. residual) indicates someone made a mistake.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_*\_\_\_\_\_\_\_\_\_\_*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

1. If your estimated line is Y = 0.5 – 5X1 + 2.2X2 + ε, and X is statistically significant, then that means for every additional unit of X1 you get 5 additional units of Y.

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

1. If Harry Kim is interested if ten different airlines have the same portions of flights take off late, he’d definitely want to use a two-factor ANOVA test.

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

*\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_*

**Part III: Short Answer.** *Answer the following.*

16 points each.

1. By now you should have outlined your regression model, even if it is a first draft. Using all your variables, write your regression model in the general form we discussed in class. (HINT: The general form used betas.)

*Don’t forget the last question on the back!*

1. Francis runs a regression with a sample of 33 and with 17 explanatory variables (including the intercept, which is included in *k*). His R2 is 0.70. What is his adjusted R2?