QM222 SECTION A1: Modeling Business Decisions Midterm

***BOSTON UNIVERSITY***

***Question School of Business***

**Fall 2017**

Sign the following statement. Grades will not be given to students who do not do so.

I have not cheated or helped anyone else cheat on this exam.

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

Signature

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

DO NOT WRITE YOUR NAME ANYWHERE ELSE ON THIS TEST.

NOTE: WE GIVE LOTS OF PARTIAL CREDIT ON TESTS. Always say something.

When we ask for calculations, show all calculations, even those you could do just on the calculator.

Questions asking for explanations and calculations are graded as incorrect if no adequate explanation is given.

Always write something. We award lots of partial credit.

Please report numbers with 3 non-zero decimal places

**Read every question carefully.**

**SECTION I Your Regression**

* Answer the following questions regarding the regressions that you have brought with you. If you have none from your own project, use the regressions at the end of this test (on electricity in Japan).
* Be sure to put your name on the page with your regression. When you complete the test, I will staple your regression sheet to your test.
* If you use your own regression, make sure all variables are defined (including your Y variable) on the sheet with your regressions or in your answers below.

**Answer these questions based on your simple (1 variable) regression:**

(Instead of 1 explanatory variable, you might have two or more categories of the same non-numeric categorical explanatory variable)

1. (4 pts) What does each observation in the data set represent? (in a few words at most)
2. (5 pts) Use the value of the coefficient on your variable in a sentence that explains what it tells us as specifically as possible. In other words, interpret this coefficient. (Do not use statistics terms in your answer. Be specific but concise.) Note: If your “simple” regression includes two (or more) X-variables that are different categories of the same categorical variable, answer this question and the next only about the first of these variables.
3. (4 pts) Does this variable have a statistically significant effect on your dependent variable? Circle one:

YES NO

List three ways that you know based on 3 different numbers in the regression output:

i.

ii.

iii.

1. (4 pts) When a variable has a statistically significant effect, what does that mean, in everyday non- statistics terms?

**Now answer these questions based on both your multiple regression (2 explanatory variables) and your simple regression:**

1. (6 pts) Compare the two coefficients **on the key variable** that enters both regressions.

What is the value of the coefficient that has the (most) omitted variable bias \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Exactly how much is this bias? (Be sure to include a negative sign if it is a negative bias) \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Exactly what do you learn ***from these regressions*** about the **correlation between the two explanatory variables in the multiple regression**? Is the correlation (CIRCLE ONE):

POSITIVE NEGATIVE CAN’T TELL

**Fully explain exactly how you know the amount/sign of the bias and the sign of the correlation (or explain why you can’t tell.)**

1. (4 pts) Name one other possibly confounding factor that you will or should add to this regression to remove some of the omitted variable bias (that remains even in the multiple regression you are using) on the coefficient of the “key” explanatory variable (i.e. the one that is in both regressions above):

Logically, why do you think this is possibly confounding?

1. (4 pts) If you added this new possibly-confounding variable to the multiple regression, how would the coefficient on the key variable (the one in all 3 regressions) change? CIRCLE ONE

THE COEFFICIENT WOULD DECREASE (THIS INCLUDES A NEGATIVE COEFFICIENT BECOMING MORE NEGATIVE)

THE COEFFICIENT WOULD INCREASE (THIS INCLUDES A NEGATIVE COEFFICIENT BECOMING LESS NEGATIVE)

WE CANNOT TELL IN WHICH DIRECTION THE COEFFICIENT WOULD CHANGE.

Explain fully how you know this:

1. (4 pts) Going back to the two original regressions, which one fits the data better? CIRCLE ONE

THE SIMPLE REGRESSION THE MULTIPLE REGRESSION

How do you know?

**SECTION II Marriages made online**

Towards the end of this test is the introduction to SECTION II. Tear it out, read it, then answer these questions:

1. (4 pts) Use the value of the coefficient of online (in the simple regression) in a sentence that explains what it tells us. In other words, interpret this coefficient. (Use every-day non-statistics words in your answer.)
2. (4 pts) I erased the standard error in the row on online. What exactly is the missing standard error?

Standard error: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Show calculations that explain how you got this answer:

USE THE REST OF THIS SHEET AS SCRATCH PAPER

Going back to modeling marital satisfaction, here is a multiple regression that adds yearsmarried to the regression:

Source | SS df MS Number of obs = 18,040

-------------+---------------------------------- F(2, 18037) = 114.09

Model | 311.349854 2 155.674927 Prob > F = 0.0000

Residual | 24610.486 18,037 1.36444453 R-squared = 0.0125

-------------+---------------------------------- Adj R-squared = 0.0124

Total | 24921.8359 18,039 1.38155307 Root MSE = 1.1681

------------------------------------------------------------------------------

marrsatis | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

online | .1555298 .0189214 8.22 0.000 .1184421 .1926176

yearsmarried | -.0490762 .0039301 -12.49 0.000 -.0567797 -.0413728

\_cons | 5.243957 .019531 268.49 0.000 5.205674 5.282239

------------------------------------------------------------------------------

1. (4 pts) James met his wife at work and was married 6 years ago. On average, what do you predict his marital satisfaction to be?

Marital satisfaction: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Show the calculations that lead to this answer:

1. (4 pts) Comparing the coefficient on online above to the coefficient on online in the original simple regression of marrsatis, which of these coefficients include an omitted variable bias? CIRCLE ONE:

SIMPLE REGRESSION MULTIPLE REGRESSION CAN’T TELL

Exactly how much is this bias? Bias is: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Show calculations:

1. (4 pts) **Based on these regressions (only)**, who is more likely to have used online sources to find a spouse, those who got married in 2012 (years married =0) or those who got married in 2005? Explain fully how you know (including any calculations). (Hint: where do you see online *and* yearsmarried in the above regression?)

I suspect that perhaps the relationship between marriage satisfaction and years married is not linear. I therefore add a variable for years married squared into the earlier regression and get this:

Source | SS df MS Number of obs = 18,040

-------------+---------------------------------- F(3, 18036) = 99.26

Model | 404.767917 3 134.922639 Prob > F = 0.0000

Residual | 24517.068 18,036 1.35934065 R-squared = 0.0162

-------------+---------------------------------- Adj R-squared = 0.0161

Total | 24921.8359 18,039 1.38155307 Root MSE = 1.1659

------------------------------------------------------------------------------

marrsatis | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

online | .1596152 .0188924 8.45 0.000 .1225843 .1966462

yearsmarried | -.1840535 .016748 -10.99 0.000 -.2168811 -.151226

yearsmarrsq | .017159 .0020699 8.29 0.000 .0131019 .0212161

\_cons | 5.423128 .029106 186.32 0.000 5.366077 5.480178

------------------------------------------------------------------------------

1. (5 pts) Are we at least 95% certain that the relationship between marriage satisfaction and years married is not linear (holding online constant)? CIRCLE ONE

YES NO CAN’T TELL

Explain how you know (showing any calculations):

1. (4 pts) The average slope of years married in the multiple regression on the previous page was negative (-.0491). Based on the above regression, (on average) is the slope negative *at all of the possible values of years married* **in this data set**? CIRCLE ONE:

YES NO CAN’T TELL

Explain how you know (showing any calculations):

Going back to the original simple regression, you suspect that the effect of online on marital satisfaction depends exactly on the kind of online site used. One of the variables you have is called ***method*** and is a string variable, where each value is one of these four strings:

(1) not on line (2) online on a dating site (3) online on a social network (4) games or other online

1. (4 pts) What stata command(s) do you use to make the dummy variable called socialnetwork for the regression below?

You make socialnetwork and two more additional dummy variables and run this regression:

Source | SS df MS Number of obs = 18,040

-------------+---------------------------------- F(3, 18036) = 31.87

Model | 131.395189 3 43.7983965 Prob > F = 0.0000

Residual | 24790.4407 18,036 1.37449771 R-squared = 0.0053

-------------+---------------------------------- Adj R-squared = 0.0051

Total | 24921.8359 18,039 1.38155307 Root MSE = 1.1724

----------------------------------------------------------------------------------

marrsatis | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-----------------+----------------------------------------------------------------

onlinedatingsite | .2199041 .0259901 8.46 0.000 .168961 .2708471

socialnetwork | .2063277 .0351047 5.88 0.000 .1375192 .2751362

otheronline | .0514609 .02939 1.75 0.080 -.0061463 .1090681

\_cons | 5.037681 .0104586 481.68 0.000 5.017181 5.058181

1. (4 pts) In words, what specific fact do we learn from the coefficient .2063 on socialnetwork in this regression?
2. (4 points) Rank the following groups by how satisfied they are in their marriage by drawing lines between the items in these columns:

Not online Most satisfied

Datingsite Second most satisfied

Social network Second least satisfied

Otheronline Least satisfied

Here is a different regression with a different dependent variable. Specifically, the dependent variable is a dummy for being divorced or separated:

Source | SS df MS Number of obs = 18,040

-------------+---------------------------------- F(2, 18037) = 6.80

Model | .346448384 2 .173224192 Prob > F = 0.0011

Residual | 459.304106 18,037 .025464551 R-squared = 0.0008

-------------+---------------------------------- Adj R-squared = 0.0006

Total | 459.650554 18,039 .025480933 Root MSE = .15958

------------------------------------------------------------------------------

Divorced\_S | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

online | -.003799 .0025849 -1.47 0.142 -.0088656 .0012677

yearsmarried | .0017983 .0005369 3.35 0.001 .0007459 .0028507

\_cons | .0198168 .0026682 7.43 0.000 .0145869 .0250467

------------------------------------------------------------------------------

1. (4 pts) Use the value of the coefficient (-.003799) on online in this regression in a sentence that explains what it tells us as specifically as possible. (Use every-day non-statistics words in your answer.)

However, you suspect that the effect of years married on being divorced or separated might be different for those who found their spouse online versus those who did not find their spouse online, so you make 2 new variables: yearsmarried\_if\_online and yearsmarried\_if\_not\_online .

1. (4 pts) What Stata commands do you use to make these two new variables?

To make yearsmarried\_if\_online:

To make yearsmarried\_if\_not\_online:

1. (5 pts) Here is the result of the regression with these variables:

Source | SS df MS Number of obs = 18,040

-------------+---------------------------------- F(3, 18036) = 5.80

Model | .443085301 3 .1476951 Prob > F = 0.0006

Residual | 459.207469 18,036 .025460605 R-squared = 0.0010

-------------+---------------------------------- Adj R-squared = 0.0008

Total | 459.650554 18,039 .025480933 Root MSE = .15956

-----------------------------------------------------------------------------------

Divorced\_Separa~d | Coef. Std. Err. t P>|t| [95% Conf. Interval]

------------------+----------------------------------------------------------------

online | .0056657 .0055029 1.03 0.303 -.0051205 .016452

yearsmaronline | .0001821 .0009882 0.18 0.854 -.0017548 .002119

yearsmarnotonline | .0024752 .0006395 3.87 0.000 .0012218 .0037286

\_cons | .0169719 .0030415 5.58 0.000 .0110103 .0229334

-----------------------------------------------------------------------------------

In everyday words (no statistics terms) what do you learn from the two coefficients of these new variables?

1. (4 pts) (stretch-question) Am I more than 95% certain that the coefficients of the two new variables in the previous regression are different? CIRCLE ONE:

YES I AM 95% CERTAIN NO I AM NOT 95% CERTAIN

Explain how you know, including any calculations you used:

**SECTION III Final Question**

(6 pts) Social media was designed to connect people. One would think that people who use social media frequently should be less lonely. However, there is growing evidence that among adolescents, greater use of social media (measured in minutes per day) is correlated with more loneliness rather than less loneliness. As a result, many educators are telling parents to limit adolescents’ social media use. You write to these educators expressing concern about their policy suggestion for two reasons.

a. First, you suggest that this correlation could be due to reverse causality, because: (fill in how reverse causality could lead to this correlation, in one or 2 sentences.)

b. Second, you suggest that this correlation could be due to a missing confounding factor, because: (fill in a very possible confounding factor that could lead to this correlation, in 1 or 2 sentences.)

**Introduction for Section II:**

YOU CAN TEAR THIS PAGE OUT OF YOUR TEST TO MAKE IT EASIER FOR YOU TO ANSWER QUESTIONS ABOUT IT.

Are marriages of people who met on line more or less likely to end in divorce? Are marriages of people who met on line happier or less happy? Recent research says that marriages of people who met on line are happier in their marriages, and their marriages last longer.

For instance, in 2012, some researchers asked a nationally representative sample who married between 2005 and 2012 questions about a variety of things including: (1) questions on marital satisfaction that the authors combined into a Marriage Satisfaction Index[[1]](#footnote-1) and whether they were still married to the same person; and (2) how they met – online or not (with more detail as well) Using their data, I have run several regressions on their sample which includes both those who met on line and those who did not and includes both those still married and those no longer married. Variable definitions:

**marrsatis**: Marriage satisfaction, measured 1 through 7, with 1 being extremely unhappy and 7 being perfect.

**online**: A dummy variable =1 if the person met online (either at an online dating site, on a social networking site, or at some other site), 0 otherwise.

**yearsmar**: How many years it has been since the person was married (ranging from 0 for those married in 2012 to 7 for those married in 2005).

**Divorced\_Separated**: A dummy variable =1 if the person was presently divorced or separated from the person they married between 2005-12, 0 otherwise.

They also made multiple dummies to replace the single online dummy by dividing those who found their spouses online (online=1) into these categories:

**datingsite**: =1 if met online on an online-dating site (for example, OKCupid), 0 otherwise.

**socialnetwork**: =1 if met online on a social networking site (for example, Facebook) , 0 otherwise.

**otheronline**: =1 if met online, but neither on an online-dating site nor on an online social networking site, 0 otherwise. For instance, otheronline includes people who met in a multiplayer game, a chatroom, etc.

**yearsmarrsq:** years married squared

First, here is a simple regression:

. regress marrsatis online

Source | SS df MS Number of obs = 18,040

-------------+---------------------------------- F(1, 18038) = 71.64

Model | 98.5943376 1 98.5943376 Prob > F = 0.0000

Residual | 24823.2416 18,038 1.37616374 R-squared = 0.0040

-------------+---------------------------------- Adj R-squared = 0.0039

Total | 24921.8359 18,039 1.38155307 Root MSE = 1.1731

------------------------------------------------------------------------------

marrsatis | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

online | .1608027 8.46 0.000 .1235653 .1980401

\_cons | 5.037681 .0104649 481.39 0.000 5.017169 5.058193

------------------------------------------------------------------------------

**Regression for those without their own regressions for Section I:**

YOU CAN TEAR THIS PAGE OUT OF YOUR TEST TO MAKE IT EASIER FOR YOU TO ANSWER QUESTIONS ABOUT IT.

The dataset for these regressions are about 1000 households in Japan who were offered an opportunity to be part of a new electricity pricing plan designed to give households incentives to avoid using a lot of electricity at times of high demand The dependent variable is electricity consumption, **ac** is a dummy variable for whether or not the house has air conditioning, and **choseplan** is a dummy variable for whether the household chose the plan.

. regress consumption ac

Source | SS df MS Number of obs = 1,000

-------------+---------------------------------- F(1, 998) = 141.00

Model | 15985452.4 1 15985452.4 Prob > F = 0.0000

Residual | 113141747 998 113368.483 R-squared = 0.1238

-------------+---------------------------------- Adj R-squared = 0.1229

Total | 129127199 999 129256.455 Root MSE = 336.7

------------------------------------------------------------------------------

consumption | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

ac | 89.40205 7.528894 11.87 0.000 74.62777 104.1763

\_cons | 653.6924 24.97052 26.18 0.000 604.6916 702.6931

------------------------------------------------------------------------------

. regress consumption ac chose

Source | SS df MS Number of obs = 1,000

-------------+---------------------------------- F(2, 997) = 656.63

Model | 73401869.7 2 36700934.9 Prob > F = 0.0000

Residual | 55725329.1 997 55893.0082 R-squared = 0.5684

-------------+---------------------------------- Adj R-squared = 0.5676

Total | 129127199 999 129256.455 Root MSE = 236.42

------------------------------------------------------------------------------

consumption | Coef. Std. Err. t P>|t| [95% Conf. Interval]

-------------+----------------------------------------------------------------

ac | 55.17814 5.393209 10.23 0.000 44.5948 65.76148

choseplan | -488.913 15.2543 -32.05 0.000 -518.8472 -458.9787

\_cons | 1000.821 20.60855 48.56 0.000 960.3794 1041.262

------------------------------------------------------------------------------

1. Cacioppo,Cacioppa, Gonzaga, Ogburn, and VanderWeele, Marital satisfaction and breakups differ across on-line and off-lin meeting venues. ***PNAS*** 110(25) June 18, 2013. [↑](#footnote-ref-1)