QM222 SECTION D1: Modeling Business Decisions Midterm

***BOSTON UNIVERSITY***

***School of Management***

**Fall 2015**

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.

**Read every question carefully.**

**SECTION 1 Your Regression**

Answer the following questions regarding the regressions that you have brought with you (or the regressions that Professor Kahn gives to you if you don’t have a regression.)

Be sure to put your name on the page with your regression. When you complete the test, staple your regression sheet to your test. Make sure all variables (including your Y variable) are defined on your regression sheet.

**Answer these questions based on your simpler regression (with the fewest variables):**

1. (4 points) What does each observation in your data set represent? (in a few words at most)
2. (5 points) Use the value of the coefficient on your key variable (the variable that appears in both regressions) in a sentence that explains what it tells us. In other words, interpret this coefficient. (Do not use statistics terms in your answer. Be specific but concise.)
3. (5 points) What specifically do you learn from the t-statistic on this coefficient in *this* regression? (Do not answer what generally the t-test tells us. Answer about what the t-statistic in THIS case.)
4. (5 points) What specifically do you learn from the p-value on this coefficient in this regression output? (The more specifically you can answer this question, the more points you get.)

**Answer these questions based on your multiple regression, the regression that adds a variable that had been omitted in the simpler regression:**

1. (4 points) Use the value of the coefficient on your key variable in the multiple regression (i.e. the key variable that you discussed above) in a sentence that explains what it tells us. In other words, interpret this coefficient. (Do not use statistics terms in your answer. Be specific but concise.)
2. (4 points) What was the sign of the omitted-variable bias in the simpler regression, positive or negative?

CIRCLE ONE: POSITIVE NEGATIVE IMPOSSIBLE TO KNOW

 Explain exactly how you know, including the calculations (if any) that led you to this answer.

1. (4 points) What is the sign of the correlation between your key variable (X1) and the variable omitted in the first regression (X2)?

CIRCLE ONE: POSITIVE NEGATIVE IMPOSSIBLE TO KNOW

Explain exactly how you know, including the calculations (if any) that led you to this answer.

In everyday language (no statistics terms), what does this sign tell us? In your answer, use the real variable names (not X1, X2 ).

1. (3 points) What would be the coefficient a1 if you ran a regression with X2 as the dependent variable and X1 as the explanatory variable? Show your calculations.

VALUE OF a1 \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Calculations:

**SECTION 2 Voting**

On the last page of this test are regressions based on the GSS for the years 1972-2014. The dependent variable ***grass*** is an indicator (dummy) variable that indicates whether or not the survey respondent believes that marijuana should be legalized.

1. **Use regression 1 to answer these questions:**
2. (3 points) How many people answered this question? \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_
3. (4 points) In everyday words (not statistical terms), interpret what the coefficient on year and its t-stat tell us.
4. (4 points) Is the sign on year what you would expect, and why do you say this?

CIRCLE ONE: YES NO

 Explanation:

1. (4 points) In everyday words, interpret what the coefficient on age and its t-stat tells us.
2. (4 points) Is this sign on age what you would expect, ***in light of the fact that we are also controlling for year,***  and why do you say that?

CIRCLE ONE: YES NO MAYBE

Explanation:

1. (4 points) Predict the likelihood that a black woman aged 30 in the year 2014 would believe that marijuana should be legal.

Predicted value: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Show the calculations that led to this answer:

1. (5 points) You strongly believe that people who are parents are less likely to support legalizing marijuana (even controlling for gender, age, race and year.) Also, in this dataset, whites are less likely to have had children.

If I were to add in an indicator variable *child* (=1 if the person ever had a child) to Regression 1, what would happen to the coefficient on white? Explain how you know, including any calculations you did to arrive at this answer. (If you can explain without doing any calculations, then don’t include calculations.)

CIRCLE ONE: GO UP GO DOWN CAN’T TELL DIRECTION

**Explanation:**

1. (4 points) A previous study found than males were 10% more likely to support legalizing marijuana than females (controlling for race, age and year). Are the results in Regression 1 ***significantly*** different from that previous study? Explain, including any calculations needed to answer this question.
2. **Use regression 2 to answer these questions:**
3. (4 points) In regression 2, we have added the variable year-squared. In everyday words, interpret what the coefficient on yearsq and its t-stat tells us.
4. (3 points) How can the coefficient on year be negative in Regression 1 but positive in Regression 2? Explain.
5. **Use regression 3 to answer these questions:**
6. (3 points) In regression 3, we have replaced *age* with a set of indicator variables for the decade in which the person was born.

Actually, the GSS does not include the variable “yearborn.” However, you can calculate it from the variables already there as part of regression 1). What Stata command would you use to create yearborn?

1. (4 points) Here is the “sum” of the variable “yearborn”

 sum yearborn

 Variable | Obs Mean Std. Dev. Min Max

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

 yearborn | 59388 1947.183 20.72048 1883 1996

What Stata command(s) would you use to make the variable born60s (used in Regression 3)?

1. (5 points) Which regression fits best?

CIRCLE ONE: REGRESSION 1 REGRESSION 2 REGRESSION 3

How do you know?

1. (4 points) In everyday words, interpret what the coefficient on born60s and its t-stat (in Regression 3) tells us.
2. (6 points) Based on regression 3, what are the characteristics of the groups least likely to want to legalize marijuana?

Gender: \_\_\_\_\_\_\_\_\_\_\_\_\_\_

Race: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

When born: \_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**SECTION 3 Word Question**

(5 points) The UN’s World Health Organization just announced that eating red meat may cause cancer. They base this on statistical studies that show that ***middle aged people who eat a lot of red meat have higher rates of colon cancer.*** What is an alternative, quite likely, reason that people who eat a lot of red meat have higher colon cancer, one that does NOT suggest that people should stop eating meat? Explain.

**Source: GSS 1972-2014 Definitions:**

grass: an indicator variable =1 if person believes marijuana should be legalized

male: an indicator variable for being male

white : an indicator variable for being Caucasian

age: the age of the respondent

year: the calendar year of the survey

**Regression 1**

 Source | SS df MS Number of obs = 34147

-------------+------------------------------ F( 4, 34142) = 653.44

 Model | 496.737275 4 124.184319 Prob > F = 0.0000

 Residual | 6488.58618 34142 .190047044 R-squared = 0.0711

-------------+------------------------------ Adj R-squared = 0.0710

 Total | 6985.32345 34146 .204572233 Root MSE = .43594

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

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

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

 male | .0799013 .0047614 16.78 0.000 .0705689 .0892337

 white | .0515032 .0062145 8.29 0.000 .0393227 .0636838

 age | -.0043914 .0001358 -32.35 0.000 -.0046575 -.0041253

 year | .0074874 .0001995 37.53 0.000 .0070964 .0078785

 \_cons | -14.50862 .3976841 -36.48 0.000 -15.28809 -13.72914

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

**Regression 2**

 Source | SS df MS Number of obs = 34147

-------------+------------------------------ F( 5, 34141) = 642.52

 Model | 600.775447 5 120.155089 Prob > F = 0.0000

 Residual | 6384.54801 34141 .187005302 R-squared = 0.0860

-------------+------------------------------ Adj R-squared = 0.0859

 Total | 6985.32345 34146 .204572233 Root MSE = .43244

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

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

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

 male | .0787643 .0047234 16.68 0.000 .0695064 .0880222

 white | .0506972 .0061646 8.22 0.000 .0386143 .0627801

 age | -.004464 .0001347 -33.14 0.000 -.004728 -.0042

 year | -1.58008 .0673077 -23.48 0.000 -1.712006 -1.448155

 yearsq | .0003982 .0000169 23.59 0.000 .0003651 .0004313

 \_cons | 1567.731 67.08265 23.37 0.000 1436.246 1699.215

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

**Regression 3**

. regr grass male white year born50s born60s born70s born80s born90

 Source | SS df MS Number of obs = 34256

-------------+------------------------------ F( 8, 34247) = 301.13

 Model | 460.371423 8 57.5464279 Prob > F = 0.0000

 Residual | 6544.5772 34247 .191099285 R-squared = 0.0657

-------------+------------------------------ Adj R-squared = 0.0655

 Total | 7004.94862 34255 .204494194 Root MSE = .43715

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

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

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

 male | .0837273 .0047653 17.57 0.000 .0743871 .0930675

 white | .0465241 .0062205 7.48 0.000 .0343317 .0587164

 year | .00428 .0002297 18.63 0.000 .0038297 .0047302

 born50s | .1464544 .006172 23.73 0.000 .1343571 .1585517

 born60s | .1044433 .0072604 14.39 0.000 .0902128 .1186738

 born70s | .1512566 .0096674 15.65 0.000 .1323081 .1702051

 born80s | .2473589 .0132841 18.62 0.000 .2213217 .2733961

 born90s | .1882654 .0248165 7.59 0.000 .1396242 .2369066

 \_cons | -8.387474 .4569265 -18.36 0.000 -9.283065 -7.491883

Source: GSS 1972-2014

Definitions:

vote is an indicator variable for whether the person voted in the previous presidential election

age is the age in years

educ is the highest level of education achieved in years

. regress vote educ

 Source | SS df MS Number of obs = 55420

-------------+------------------------------ F( 1, 55418) = 3034.41

 Model | 568.521426 1 568.521426 Prob > F = 0.0000

 Residual | 10383.0281 55418 .187358405 R-squared = 0.0519

-------------+------------------------------ Adj R-squared = 0.0519

 Total | 10951.5495 55419 .197613626 Root MSE = .43285

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

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

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

 educ | .0318844 .0005788 55.09 0.000 .0307499 .0330189

 \_cons | .319073 .0076635 41.64 0.000 .3040525 .3340934

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

. regress vote educ age

 Source | SS df MS Number of obs = 55249

-------------+------------------------------ F( 2, 55246) = 4016.38

 Model | 1386.12245 2 693.061225 Prob > F = 0.0000

 Residual | 9533.17372 55246 .172558624 R-squared = 0.1269

-------------+------------------------------ Adj R-squared = 0.1269

 Total | 10919.2962 55248 .197641474 Root MSE = .4154

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

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

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

 educ | .0405764 .0005706 71.12 0.000 .0394581 .0416947

 age | .0072756 .0001057 68.80 0.000 .0070684 .0074829

 \_cons | -.1331514 .0098635 -13.50 0.000 -.152484 -.1138188

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