Homework 23: ECO220Y

Required Problems:

(1) Here are some excerpts from *The World Happiness Report (2012)*:

> **Excerpt #1:** If we want to influence the levels of happiness and misery, we need to know what causes them. p. 59
>
> **Excerpt #2:** To isolate the causal effect of each factor is not easy. It clearly requires us to hold as much else as possible constant while we look at the co-movement of well-being and the factor in question. In most cases this cannot be done experimentally. So the next best is to study the same individuals (or countries) over time and see how their well-being moves when different factors change. p. 60
>
> **Excerpt #3:** Much of the evidence we shall quote is of this longitudinal, time-series form. But some insights can also be got from cross-sectional evidence. In this case we are comparing different individuals (or countries) at the same point in time. The problem here is that, when we compare individuals or countries, there are many ways in which they may differ (for example in personality or values) that cannot easily be measured and controlled for when we are examining the effect of those factors that can be measured. p. 60

(a) What is Excerpt #3 talking about?

(A) autocorrelation **(B)** lurking variables **(C)** heteroscedasticity **(D)** simple regression analysis
(E) statistical vs. economic significance

(b) Consider the table of results below. Which kind of data are these? What is the unit of observation? Looking at Panel B, what are the effects "that can be measured"? What are examples of effects that are not measured?

Table 3.1. Regressions to explain average well-being across countries

Independent Variables:	**Dependent Variable: *Life Evaluation***
Panel A	
Log GDP per head	0.81 ***
R^2	0.65
Number of countries	153
Panel B	
Log GDP per head	0.28 **
Health	0.25 **
Education	-0.01
Social Support	0.29 ***
Freedom	0.15 ***
Corruption	-0.18 ***
Divorce etc.	-0.43
R^2	0.80
Number of countries	139

Significance Levels: (1 tailed tests) * 0.05, ** 0.01, *** 0.001

(c) Is the regression in Panel B statistically significant?

(d) Which of the individual coefficients are statistically significant?

(e) Consider another excerpt and the table below summarizing these data. Which kind of data are these? What is the unit of observation? How many of the variables are dummy variables?

Does relative income raise a person's happiness and does absolute income do likewise? To examine the effect of income on happiness, we must eliminate any effect of a person's underlying happiness upon their income. The best way to attempt this is with panel data in which we trace the same individual over many years and examine how changes in the person's income affects their subsequent happiness. Fortunately we have such data from Europe's leading country. In West Germany the German Socio-Economic Panel (GSOEP) has been tracking the same individuals each year since 1984. We can use these data to help us understand the movement of average life satisfaction in that country. p. 61

GSOEP (1984-2009) (West Germany) Summary Statistics

	Mean	S.d.	Min.	Max.
Life satisfaction	7.05	1.72	0	10
Log of Income (monthly)	3.33	0.20	2.70	3.74
Female	0.51	0.49	0	1
Age	42.05	7.27	30	55
Age^2*/1000*	1.82	0.61	0.90	3.02
Age^3*/1000*	81.10	40.43	27	166.37
Single	0.13	0.34	0	1
Widowed	0.01	0.10	0	1
Divorced	0.08	0.28	0	1
Separated	0.02	0.16	0	1
Unemployed	0.06	0.24	0	1
Self Employed	0.05	0.22	0	1
Out of the labor force	0.12	0.33	0	1
Student	0.00	0.07	0	1
Education: high	0.19	0.39	0	1
Education: medium	0.31	0.46	0	1
One child	0.23	0.42	0	1
Two children	0.22	0.41	0	1
Three + children	0.08	0.27	0	1
Health: Excellent	0.09	0.29	0	1
Health: Good	0.46	0.49	0	1
Health: Satisfactory	0.30	0.46	0	1
Health: Poor	0.10	0.30	0	1

(f) Some additional passages that discuss the benefits of panel (i.e. longitudinal) data. *Note:* The questions that you need to answer about **Table 1** and **Table 2** are after these supplemental materials.

But when we have longitudinal data on the same person or the same country we can assume that these unmeasured factors are more similar at each observation, and may have a better chance of tying down what is causing what. p. 60

German Socio-Economic Panel (GSOEP, 1984-2009): Table 1 shows OLS cross-section regressions, while Table 2 shows OLS equations including a fixed-effect for each individual and year, so that the equation estimates the effect of each variable in explaining the different levels of happiness which an individual experiences in each different year. pp. 84 – 86

Table 1. Cross-sectional regressions to explain life satisfaction

	GSOEP: ***Range of life satisfaction 0 – 10***
Log of Income (monthly)	0.60 (0.01)
Female	0.12 (0.00)
Age	0.11 (0.07)
Age^2/1000	-3.55 (1.80)
Age^3/1000	0.03 (0.01)
Single	-0.15 (0.01)
Widowed	-0.18 (0.04)
Divorced	-0.20 (0.01)
Separated	-0.48 (0.02)
Unemployed	-0.63 (0.01)
Self Employed	-0.15 (0.01)
Out of the labor force	-0.03 (0.01)
Student	-0.12 (0.06)
Education: high	0.00 (0.01)
Education: medium	-0.01 (0.01)
One child	-0.02 (0.01)
Two children	-0.03 (0.01)
Three + children	-0.09 (0.01)
Health: Excellent	3.45 (0.03)
Health: Good	2.82 (0.03)
Health: Satisfactory	2.04 (0.03)
Health: Poor	1.26 (0.03)
Fixed effects	No
Time Dummies	Yes
Region Dummies	Yes
Observations	100,945
R^2	0.25

Table 2. Fixed-effects regressions to explain life satisfaction

	GSOEP: ***Range of life satisfaction 0 – 10***
Log of Income (monthly)	0.39 (0.01)
Female	--
Age	-0.16 (0.07)
$Age^2/1000$	2.97 (1.76)
$Age^3/1000$	-0.02 (0.01)
Single	-0.07 (0.03)
Widowed	-0.44 (0.07)
Divorced	0.03 (0.02)
Separated	-0.25 (0.03)
Unemployed	-0.49 (0.01)
Self Employed	-0.01 (0.02)
Out of the labor force	-0.13 (0.01)
Student	-0.14 (0.06)
Education: high	0.07 (0.05)
Education: medium	0.10 (0.03)
One child	0.07 (0.01)
Two children	0.04 (0.02)
Three + children	0.06 (0.02)
Health: Excellent	2.25 (0.03)
Health: Good	1.92 (0.03)
Health: Satisfactory	1.51 (0.03)
Health: Poor	0.93 (0.03)
Fixed effects	Yes
Time Dummies	Yes
Region Dummies	Yes
Observations	100,945
R^2	0.20

Both regressions are highly statistically significant overall. Some questions about these results:

For Table 1: Why is monthly income logged (natural logged)? For each of the dummy variable groups, what is the omitted category? How many *Time Dummies* are there? What is reported in parentheses? Which of the coefficients are statistically significant? How do you interpret the coefficients?

For Table 2: The only difference in the specification between Tables 1 and 2 is that Table 2 includes *Fixed effects*. What are these? Roughly, how many fixed effects are there? Why is there a "—" in the space for the coefficient on *Female*? Which of the coefficients are statistically significant? How do you interpret the coefficients? How do the results compare across Tables 1 and 2?

(g) What can we say about people with high education?

(A) They are no happier than everyone else
(B) The difference in happiness between them and everyone else is not statistically significant
(C) We cannot say they are happier than those with low education once we control for the other factors
(D) All of the above

(h) What can we say about people with no children? Generally, they are slightly ___ than other people.

(A) happier (statistically significant)
(B) happier (not statistically significant)
(C) less happy (statistically significant)
(D) less happy (not statistically significant)
(E) None of the above

(2) What is the meaning of "robust standard errors" (e.g. as in Slide 17 of Lecture 23)? Why are these used? Can you still use these robust standard errors to conduct t tests using the regular formulas given on our Aid Sheets?

(3) Consider a multiple regression analysis of students' grades for a sample of 285 students. The dependent variable – `inter_grade` – is the course grade in a second year intermediate level course (e.g. ECO220Y). The explanatory variables are `female`, `program_a`, `program_b`, `program_c`, and `intro_grade`. The variable `female` is a dummy variable =1 female and 0 otherwise. There are four different programs (e.g. Commerce, Economics Major, Economics Minor, etc.) and the regression includes dummies for three of them (=1 if student in that program and =0 otherwise). Finally, `intro_grade` is the student's course grade in the introductory level course (e.g. ECO100Y). Suppose these variables are observed for 285 students. Review all of the STATA output on the next pages.

```
/* Summary Statistics */
    Variable |       Obs        Mean    Std. Dev.       Min        Max
-------------+--------------------------------------------------------
 inter_grade |       285    75.07018    9.250542         49        100
 intro_grade |       285     79.9193    8.662519         54        100
   program_a |       285    .6491228    .4780841          0          1
   program_b |       285    .1789474    .3839825          0          1
   program_c |       285    .1017544    .3028568          0          1
   program_d |       285    .0701754    .2558918          0          1
      female |       285    .5894737    .4927946          0          1

/* Regression #1 */
      Source |       SS       df       MS              Number of obs =     285
-------------+------------------------------           F(  5,   279) =   67.02
       Model |  13261.3423     5  2652.26845           Prob > F      =  0.0000
    Residual |  11041.2542   279   39.574388           R-squared     =  0.5457
-------------+------------------------------           Adj R-squared =  0.5375
       Total |  24302.5965   284  85.5725229           Root MSE      =  6.2908

------------------------------------------------------------------------------
 inter_grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 intro_grade |   .5208913   .0541569     9.62   0.000     .4142832    .6274995
   program_a |   8.249301   1.652306     4.99   0.000     4.996731    11.50187
   program_b |   4.317033   1.675715     2.58   0.011     1.018382    7.615684
   program_c |  -.4826203   1.830033    -0.26   0.792    -4.085045    3.119804
      female |  -2.269726   .7752339    -2.93   0.004    -3.795777    -.743676
       _cons |   28.70063   4.067882     7.06   0.000     20.69299    36.70826
------------------------------------------------------------------------------
```

/* Regression #2 */

```
      Source |       SS       df       MS              Number of obs =     285
-------------+------------------------------           F(  1,   283) =  238.98
       Model |  11126.4471     1  11126.4471           Prob > F      =  0.0000
    Residual |  13176.1494   283  46.5588319           R-squared     =  0.4578
-------------+------------------------------           Adj R-squared =  0.4559
       Total |  24302.5965   284  85.5725229           Root MSE      =  6.8234

------------------------------------------------------------------------------
 inter_grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
 intro_grade |   .7225616    .046741    15.46   0.000     .6305575    .8145657
       _cons |   17.32356   3.757308     4.61   0.000     9.927743    24.71938
------------------------------------------------------------------------------
```

/* Regression #3 */

```
      Source |       SS       df       MS              Number of obs =     285
-------------+------------------------------           F(  3,   281) =   58.38
       Model |  9331.31343     3  3110.43781           Prob > F      =  0.0000
    Residual |  14971.2831   281  53.2785874           R-squared     =  0.3840
-------------+------------------------------           Adj R-squared =  0.3774
       Total |  24302.5965   284  85.5725229           Root MSE      =  7.2992

------------------------------------------------------------------------------
 inter_grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   program_a |   14.68243   1.718116     8.55   0.000     11.30042    18.06444
   program_b |   6.571569   1.925775     3.41   0.001     2.780793    10.36234
   program_c |   .1327586   2.121585     0.06   0.950    -4.043458    4.308975
       _cons |      64.35   1.632155    39.43   0.000      61.1372     67.5628
------------------------------------------------------------------------------
```

/* Regression #4 */

```
      Source |       SS       df       MS              Number of obs =     285
-------------+------------------------------           F(  1,   283) =    0.06
       Model |  4.80833494     1  4.80833494           Prob > F      =  0.8131
    Residual |  24297.7882   283  85.8579087           R-squared     =  0.0002
-------------+------------------------------           Adj R-squared = -0.0033
       Total |  24302.5965   284  85.5725229           Root MSE      =   9.266

------------------------------------------------------------------------------
 inter_grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      female |   .2640415   1.115745     0.24   0.813    -1.932171    2.460254
       _cons |   74.91453   .8566378    87.45   0.000     73.22834    76.60072
------------------------------------------------------------------------------
```

(a) What fraction of students are female? What fraction of students are in Program B?

(b) What is the predicted intermediate grade of a male student in Program B who earned an 80 introductory grade? What is the predicted intermediate grade of a student who earned an 80 introductory grade?

(c) On average how much to the intermediate grades differ between students in Program A and Program B?

(d) Do female students tend to earn lower grades in the intermediate course? If so, how much? If no, what can you say about gender differences?

(e) How to interpret each of the coefficients in Regression #1?

(f) How do you interpret the constant term in Regression #1? Regression #2? Regression #3? Regression #4?

(4) Last week, you worked on solving some questions from Term Test #4 from March 2017 (http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_4_MAR17.pdf). You are now ready to solve part (d) of Question (4).

(5) Term Test #5 from April 2018 (http://homes.chass.utoronto.ca/~murdockj/eco220/TT220_5_APR18.pdf) has some excellent questions for you to work on that cover this week's material.

> **(a)** To practice making the connection between inference about a difference between two means and a regression analysis using a dummy variable, answer Question (3), all parts.
>
> **(b)** To practice with the Waterloo case study (including quadratics), answer Question (4).