ECO 5350 T. Fomby
Intro. Econometrics Fall 2016

**EXERCISE 4**
**KEY**

**Purpose**: To learn more about bivariate regression. **This exercise is due on Tuesday, September 20.**

Work problems 4, 5, and 6 on page 54 of your textbook and computer exercises C5, C6, and C7 on page 57. I will choose ~~3~~ 2 of these problems/exercises for grading so you need to answer each of the problems and exercises on a separate page or pages so that I can easily grade them. Be sure and put your name on each of the problems/exercises.

I chose problems 4 and C7.

**Problem 4:**

(i) What is the predicted birthweight when cigs = 0? **Answer:** 119.77. What about when cigs = 20? Answer: $\widehat{bwght} = 119.77 - 0.514(20) = 109.49$ . Comment on the difference. **Answer:** Obviously, mothers who smoke cigarettes during pregnancy have, on average, lower birthweight children.

(ii) Does this simple regression necessarily capture a causal relationship between the child's birth weight and the mother's smoking habits? Explain. **Answer:** In this case, smoking precedes the birth of the child and therefore the incidence of smoking can be considered to be causal. However, there are probably many other factors that affect the birthweight of child such as income of the mother, overall health of the mother, and environmental conditions (smog, pollution, etc.). Nevertheless, even with these other factors considered in a multiple regression of birthweight on these factors, we are likely to see that smoking would still maintain its significant role in affecting birthweight.

(iii) To predict a birth weight of 125 ounces, what would *cigs* have to be? **Answer:** Bw = 119.77 – 0.514cigs. Therefore, 125 = 119.77 – 0.514(cigs*) which results in cigs* = -10.175. This is clearly impossible. After all, the above regression equation implies that the "best" average birthrate for non-smokers is 119.77 ounces.

(iv) The proportion of women in the sample who do not smoke while pregnant is about 0.85. Does this help reconcile your finding from part (iii)? **Answer:** We know that it would be nicer, in terms of sample design, to have more smokers in the sample because this would add to the variation in the explanatory variable (*cigs*) and we might get a more accurate estimate of the intercept which gives us the average birthweight of children of non-smoking mothers. It is possible that the intercept estimate might take on a more accurate estimate in this case and indicate the greater likelihood of having larger babies.

**Problem C7:**

(i) The average gift is about 7.44 Dutch guilders. Out of 4,268 respondents, 2,561 did not give a gift, or about 60 percent.

(ii) The average mailings per year is about 2.05. The minimum value is .25 (which presumably means that someone has been on the mailing list for at least four years), and the maximum value is 3.5.

(iii) The estimated equation is

$$\widehat{gift} = 2.01 + 2.65\, mailsyear$$

n = 4,268, $R^2 = 0.0138$

(iv) The slope coefficient from part (iii) means that each mailing per year is associated with – perhaps even "causes" – an estimated 2.65 additional guilders, on average. Therefore, if each mailing costs one guilder, the expected profit from each mailing is estimated to be 1.65 guilders. This is only the average, however. Some mailings generate no contributions, or a contribution less than the mailing cost; other mailings generated much more than the mailing cost.

(v) Because the smallest *mailsyear* in the sample is .25, the smallest predicted value of *gifts* is 2.01 + 2.65(.25) ≈ 2.67. Even if we look at the overall population, where some people have received no mailings, the smallest predicted value is about two. So, with this estimated equation, we never predict zero charitable gifts.

The STATA program that generates the above answers is given by:

```
summarize gift
summarize gift if gift == 0
summarize mailsyear
regress gift mailsyear
predict yhat
summarize yhat
```

Here is the output from that program:

```
  ___  ____  ____  ____  ____ (R)
 /__    /   ____/   /   ____/
___/   /   /___/   /   /___/   14.0   Copyright 1985-2015 StataCorp LP
  Statistics/Data Analysis            StataCorp
                                      4905 Lakeway Drive
                                      College Station, Texas 77845 USA
                                      800-STATA-PC        http://www.stata.com
                                      979-696-4600        stata@stata.com
                                      979-696-4601 (fax)

Single-user Stata perpetual license:
       Serial number:  301406223833
         Licensed to:  Thomas B. Fomby
                       SMU

Notes:
      1.  Unicode is supported; see help unicode_advice.
      2.  New update available; type -update all-

. use "E:\E5350\E5350 f16\Exercises\charity.dta", clear

. do "C:\Users\00008904\AppData\Local\Temp\STD02000000.tmp"

. summarize gift

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        gift |      4,268     7.44447    15.06256          0        250

. summarize gift if gift == 0

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        gift |      2,561           0           0          0          0

. summarize mailsyear

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
   mailsyear |      4,268    2.049555      .66758        .25        3.5

. regress gift mailsyear

      Source |       SS           df       MS      Number of obs   =     4,268
-------------+----------------------------------   F(1, 4266)      =     59.65
       Model |  13349.7251         1  13349.7251   Prob > F        =    0.0000
    Residual |  954750.114     4,266  223.804528   R-squared       =    0.0138
-------------+----------------------------------   Adj R-squared   =    0.0136
       Total |   968099.84     4,267  226.880675   Root MSE        =     14.96

------------------------------------------------------------------------------
        gift |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   mailsyear |    2.649546   .3430598     7.72   0.000     1.976971    3.322122
       _cons |     2.01408   .7394696     2.72   0.006     .5643347    3.463825
------------------------------------------------------------------------------

. predict yhat
(option xb assumed; fitted values)

. summarize yhat

    Variable |        Obs        Mean    Std. Dev.       Min        Max
-------------+---------------------------------------------------------
        yhat |      4,268     7.44447    1.768784   2.676466   11.28749

.
```