Problem Set 5
ECON 30331

Dan Hungerman

1. In class, we discussed how we can estimate the standard error of a predicted value, $\hat{y}$, evaluated at a set of values $x_1 = c_1, x_2 = c_2, \ldots, x_k = c_k$, by appropriately transforming the x-variables. Departing from my lecture notes, I provided an example where one could estimate $y = \beta_0 + \beta_1 age + \beta_2 age^2 + e$ and then calculate $\hat{y}$ when *age* = 21 by subtracting 21 from the variables *age* and age^2. It was then asked whether I would want to subtract 21, or 21^2 = 441, from the variable age^2. I said I didn't know.

   Your job for this question is to figure it out.

   A. Download the data *house_price.dta* from the website. Run a regression of house price on *age* and age^2. Report the associated coefficients and standard errors.

   B. Immediately after the regression, type "predict phat". This creates a new variable, *phat*, that is the predicted prices from the regression. What is the predicted price for a house that is 88 years old?

   C. Now, by creating transformed variables, run a regression that appropriately estimates the standard error of your estimate in part B. What is the correct way to transform the variable age^2 here?

   D. Now estimate the equation

   $$price = \delta_0 + \delta_1 age + \delta_2 age^2 + \delta_3 sq\_feet + \delta_4 bedrooms + e$$

   What is the expected price of an 88 year-old house with 1,500 square feet and 3 bedrooms? What is the 95% confidence interval for this expected value?

   E. While you were out getting a snack, Mr. Silly took control of your Stata dataset and decided to do something silly: he tried to estimate the expected price for a house that was 3,000 years old, with -100 square feet (that's *negative* one-hundred) and 17 bedrooms. Mr. Silly is delighted to see that Stata can produce an estimate and standard error for this silly expected value. What is the expected value and its standard error?

   F. Is it worrisome that Mr. Silly was able to do that silly stuff in part E? If Stata will give us predicted prices for houses that cannot exist, does this suggest that the entire notion of using a regression to look at predicted prices is a waste of time?

2. Below is a regression I have run on some data (it is a regression on US states, excluding Hawaii and Alaska. The dependent variable is per-capital income, and the X variables are the fraction of a state's population that is over age 65, the fraction black, the fraction foreign-born, and the state's unemployment rate.) The R-squared and adjusted R-squared has been removed

A. What is the R-squared?

B. What is the adjusted R-squared?

C. Suppose I removed the variable for the fraction of the population over 65, `p65`. What would happen to the R-squared? What about the adjusted R-squared?

```
. reg  disposable_income p65 pblack pforeign riu_unemployment

      Source |       SS       df       MS              Number of obs =      48
-------------+------------------------------           F(  4,    43) =    5.43
       Model |  169.901986     4  42.4754965           Prob > F      =  0.0013
    Residual |  336.559814    43  7.82697241           R-squared     =
-------------+------------------------------           Adj R-squared =
       Total |    506.4618    47   10.775783           Root MSE      =  2.7977

------------------------------------------------------------------------------
disposable~e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
         p65 |   .1535071   .2594459     0.59   0.557    -.3697155    .6767297
      pblack |  -.0082877   .0429338    -0.19   0.848     -.094872    .0782966
    pforeign |   .3530978   .0765579     4.61   0.000     .1987041    .5074916
riu_unempl~t |  -.5848798   .7734795    -0.76   0.454     -2.14475    .9749903
       _cons |   20.69081   3.471959     5.96   0.000     13.68894    27.69268
------------------------------------------------------------------------------
```

3. Can an adjusted R-squared ever be larger than the traditional R-squared? (You can assume that we are discussing a setting where $k > 0$ and that the traditional R-squared is greater than zero and less than 1.)

4. Here is a set of three data points:

	y	x
Observation 1	1	3
Observation 2	-4	-10
Observation 3	3	7

A. Consider a regression of $y = b_o + b_1 x + e$ here. What would the OLS estimates of b_0 & b_1 be?

B. Calculate the standard errors for your estimates in part A. (Ignore heteroskedasticity here.)

C. Now calculate the standard errors so that they are robust to heteroskedasticity.