Worksheet week # 6

1. You have organized a ski trip to the mountains for a group of your friends and, as a true econometrician, you decided to estimate a model of the expenditures of each participant. You suppose that the cost of the trip for each person depends on how many days he or she spent there (some people arrived later and some left earlier) and on what type of skis he or she was going. Some people went on downhill skis, some went on cross-country skis and some people managed to go on both. Since you are friends only with people who like sports, there was nobody who was not skiing (i.e., everybody went on downhill or cross-country skis or both).

   (a) You specify the following model:

   $$cost = \beta_0 + \beta_1 day + \beta_2 DS + \beta_3 CS + \varepsilon \quad ,$$

   where *cost* is the cost of the trip, *days* stands for the number of days the person stays in the mountains, DS is a dummy equal to 1 if the person goes on downhill skis, zero otherwise, and CS is a dummy equal to 1 if the person goes on cross-country skis, zero otherwise.

   i. In terms of the parameters of your model, what is the expected cost of the trip for a person who spends two days in the mountains and goes both on cross-country and downhill skis? What is the expected cost for a person who spends three days in the mountains and goes on downhill skis only?

   ii. You want to test if the two dummy variables are jointly significant in your model. Running the model with the dummies included leads to $R^2 = 0.8$, whereas running it without the dummies gives $R^2 = 0.65$. Knowing that you have 25 observations, test for the joint significance of the two dummies at 95% confidence level.

   (b) A friend of yours is working on the same problem, but he specifies the model in a little bit different way:

   $$cost = \gamma_1 day + \gamma_2 DSO + \gamma_3 CSO + \gamma_4 BS + \varepsilon \quad ,$$

   where *cost* is the cost of the trip, *days* stands for the number of days the person stays in the mountains, DSO is a dummy equal to 1 if the person goes only on downhill skis, zero otherwise, CSO is a dummy equal to 1 if the person goes only on cross-country skis, zero otherwise, and BS is a dummy equal to 1 if the person goes on both skis, zero otherwise. (Hence, your friend's model differs from yours only in the definition of the dummies.)

i. You see that your friend includes in the model the full set of dummies. How can you explain that he is not facing perfect multicollinearity?

ii. In terms of the parameters of your friend's model, what is the expected cost of the trip for a person who spends two days in the mountains and goes both on cross-country and downhill skis? What is the expected cost for a person who spends three days in the mountains and goes on downhill skis only?

2. Consider the following model:

$$\log(price) = \beta_0 + \beta_1 \log(assess) + \beta_2 \log(sqrft) + \beta_3 \log(lotsize) + \beta_4\, d\_bdrms + \varepsilon \quad , \quad (1)$$

where *price* is house price, *assess* is the assessed housing value (before the house was sold), *lotsize* is size of the lot (in feet), *sqrft* is square footage, and *d_bdrms* is a dummy variable indicating if the house has more than 3 bedrooms.

(a) Use the data *housing.gdt* to estimate the model (1). First transform the first four variables in logarithms, then construct the dummy variable as

$$d\_bdrms = \begin{cases} 1 & \text{if } bdrms > 3 \\ 0 & \text{otherwise} \end{cases}$$

and run the regression. Interpret the coefficients.

(b) Now, suppose we would like to test whether the assessed housing price is a rational valuation: if this is the case, then a 1% change in *assess* should be associated with a 1% change in *price*. In addition, *lotsize*, *sqrft*, and *d_bdrms* should not help to explain $\log(price)$, once the assessed value has been controlled for. Define the hypotheses to be tested, the test statistic, and explain how would you conduct the test. Then test for rational valuation in Gretl.

3. Suppose your data produce the regression result $\widehat{y} = 1 + 0.7x$. Consider scaling the data to express them in a different base year dollar, by multiplying observations by 0.8.

(a) If both y and x are scaled, what regression results would you obtain?

(b) If y is scaled but x is not, what regression results would you obtain?

(c) If x is scaled but y is not, what regression results would you obtain?