Introductory Econometrics
Tutorial 5

PART A: To prepare for this week's quiz read the lecture notes for Topic 3 and Topic 4.

Part B: This part will be covered in the tutorial. It is still a good idea to attempt these questions before the tutorial.

1. Problem C2, Chapter 3 of the textbook: Use the data in HPRICE1.WF1 to estimate the model

$$price = \beta_0 + \beta_1 sqrft + \beta_2 bdrms + u,$$

where *price* is the house price measured in thousands of dollars.

i) Write out the results in equation form.

ii) What is the estimated increase in price for a house with one more bedroom, holding square footage constant?

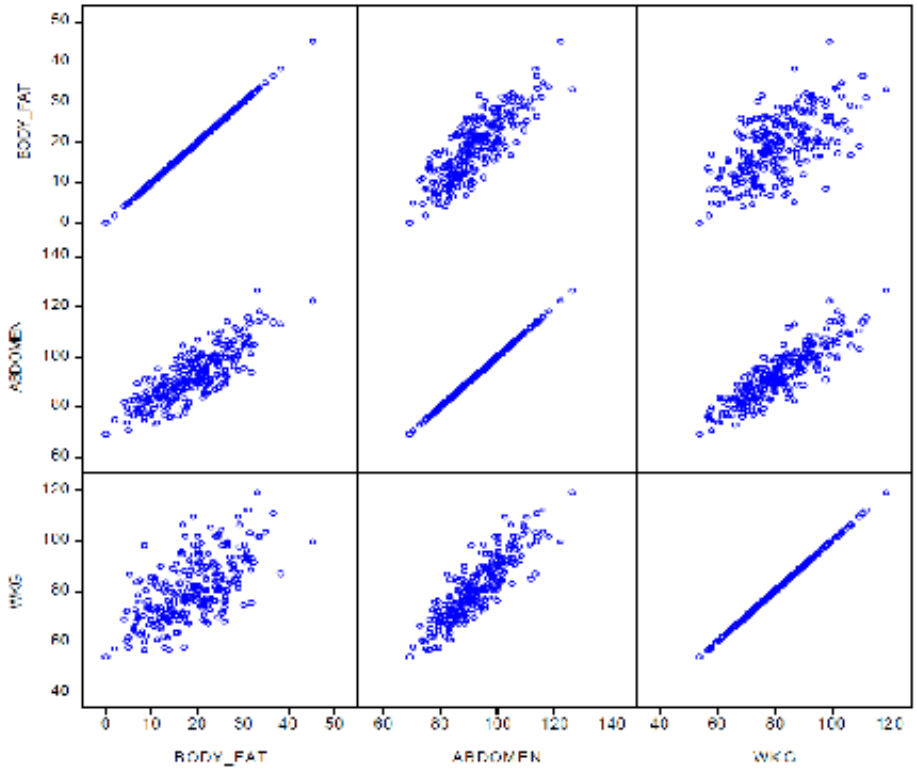
iii) What is the estimated increase in price for a house with an additional bedroom that is 140 square feet in size? Compare this to your answer in part (ii).

iv) What percentage of the variation in price is explained by square footage and number of bedrooms?

v) The first house in the sample has sqrft = 2,438 and bdrms = 4. Find the predicted selling price for this house from the OLS regression line.

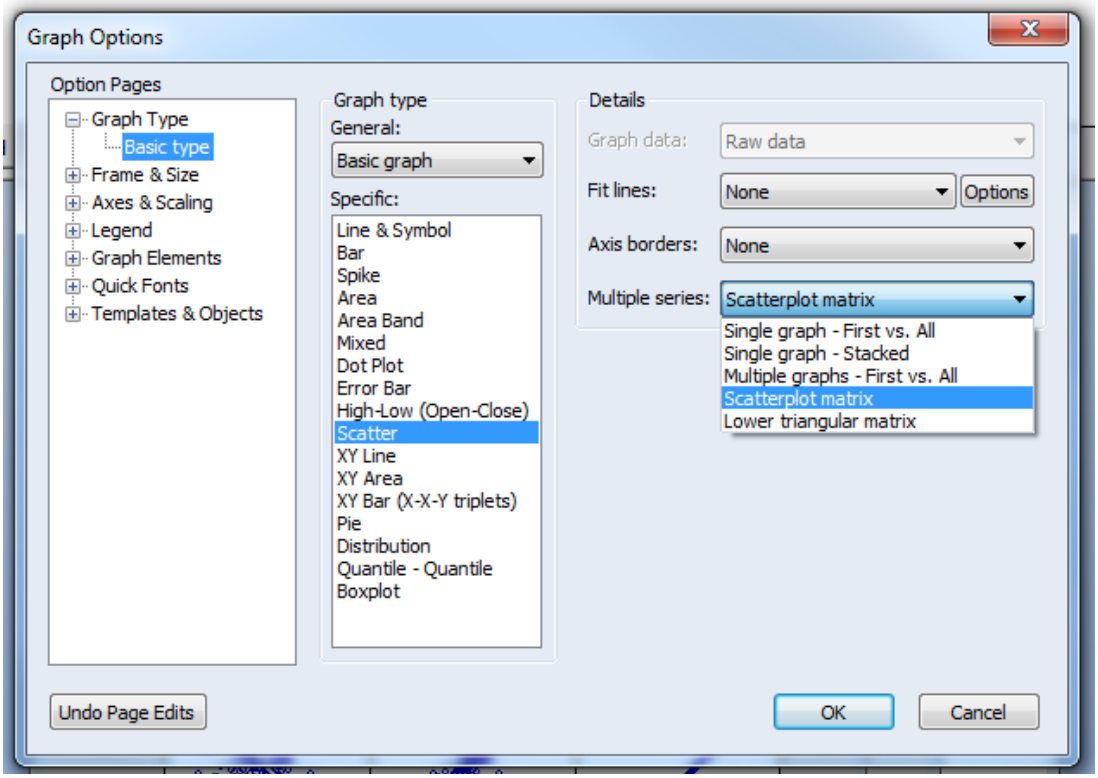
vi) The actual selling price of the first house in the sample was \$300,000 (so price = 300). Find the residual for this house. Does it suggest that the buyer underpaid or overpaid for the house?

2. We would like to make an "app" where users input their easy to measure body characteristics and the app predicts their body fat percentage. We start with making an app for men. We have data on body fat percentage (BODY_FAT), weight in kg (WKG) and abdomen circumference in cm (ABDOMEN) for 251 adult men. The matrix of scatter plots of each pair of these three variables in our sample is given below.



The plots in the first row are: the scatter plot of body fat against body fat (which is the 45 degree line) at the left corner, the scatter plot of body fat against abdomen circumference in the middle, and the scatter plot of body fat against weight in the top right corner. You can create these matrices in Eviews by graphing more than two variables and then choosing scatter plot,

with the scatter plot matrix option, as shown in the screen shot below.



Without estimating any regressions, explain what these plots can tells us about each of the following (the correct answer for one of these is "nothing"):

(a) the sign of the coefficient of ABDOMEN in a regression of BODY_FAT on a constant and ABDOMEN,

(b) the sign of the coefficient of WKG in a regression of BODY_FAT on a constant and WKG,

(c) which of the two regressions explained in parts (a) and (b) is likely to have a better fit,

(d) the sign of the coefficient of WKG in a regression of BODY_FAT on a constant, ABDOMEN and WKG.

3. With the same data as above, we have estimated three regressions:

$$\begin{aligned}
\widehat{BODY_FAT} &= -12.63 + 0.39WKG, \qquad R^2 = 0.385,\ \bar{R}^2 = 0.382 \\
\widehat{BODY_FAT} &= -38.60 + 0.62ABDOMEN, \quad R^2 = 0.681,\ \bar{R}^2 = 0.679 \\
\widehat{BODY_FAT} &= -42.94 + 0.91ABDOMEN - 0.27WKG,\ R^2 = 0.724,\ \bar{R}^2 = 0.722
\end{aligned}$$

(a) The signs and the R^2s of the first two regressions must agree with your answers to parts (a), (b) and (c) of the previous question. If they don't, then discuss these in the tutorial or during consultation hours.

(b) Think about the negative coefficient of WKG in the third equation. Does it make sense? (Hint: yes, it makes very good sense, and it highlights the extra information that multiple regression extracts from the data that simple two variable regressions cannot do). Explain, to a non-specialist audience, what the estimated coefficient of WKG in the third regression tells us.

(c) If weight was measured in pounds rather than kilograms (each kilogram is 2.2 pounds), how would the above regression results change? Check your answers by running the regressions using bodyfatkg.wf1 file.