CORRECTION - WEEK 6

7 Thursday 02/23/06 - Multivariate Regressions

1. You want to see the impact of gender *gender* on earnings: *wage*.

   (a) Write down the econometric model you want to test.

   $$wage_i = \beta_0 + \beta_1\, gender_i + \varepsilon_i \quad , \quad gender = \begin{cases} 1 & \text{if male} \\ 0 & \text{o/w} \end{cases}$$

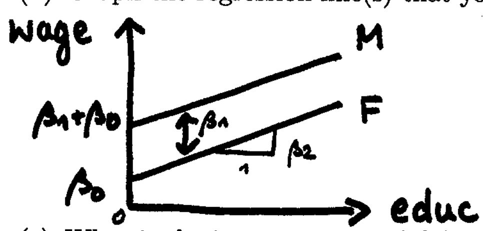
   (b) How do you interpret β_1? And β_0?

   $\beta_1 =$ wage gap , $\beta_0 = E[wage \mid female]$

   (c) Now, let's include another explanatory variable, *educ* into the model. Rewrite your equation, taking into account now that your wage is explained by not only your gender but also your educational level.

   $$wage_i = \beta_0 + \beta_1\, gender_i + \beta_2\, educ_i + \varepsilon_i$$

   (d) Graph the regression line(s) that you get from this new equation.



   - MALE : gender = 1
     $\hat{wage} = \underbrace{\hat\beta_0 + \hat\beta_1}_{\text{intercept}} + \hat\beta_2\, educ$
   - FEMALE : gender = 0
     $\hat{wage} = \underbrace{\hat\beta_0}_{\text{intercept}} + \hat\beta_2\, educ$

   $\boxed{\neq \text{intercepts}}$

   (e) What is the interpretation of β_2?

   if educ ↑ by 1 year, wage ↑ by β_2 units HOLDING EVERYTHG ELSE CST.

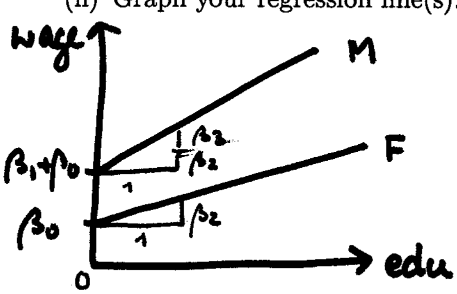
   (f) For now, we have considered that the educational level has the same impact on your wage whether you are a female or a male (the returns to education are the same across genders). Using an interaction variable, write dow a third equation allowing to have different returns to education for men and women.

   $$wage_i = \beta_0 + \beta_1\, gender_i + \beta_2\, educ_i + \beta_3\, educ_i * gender_i + \varepsilon_i$$

   (g) What is the slope for women? And for men?

   $\boxed{\text{WOMEN :}}$ gender = 0 $\Rightarrow$ slope : $\hat\beta_2\, educ + \beta_3 * educ * 0 = \hat\beta_2\, educ$
   slope is $\boxed{\hat\beta_2}$. || $\boxed{\text{MEN}}$, gender = 1 , $\boxed{\text{slope} = \hat\beta_2 + \hat\beta_3}$

   (h) Graph your regression line(s).



   F : $\hat{w} = \hat\beta_0 + \hat\beta_2\, educ$
   M : $\hat{w} = \hat\beta_0 + \hat\beta_1 + (\hat\beta_2 + \hat\beta_3)\, educ$

   $\boxed{\begin{array}{l} \neq \text{intercepts} \\ \neq \text{slopes} \end{array}}$

2. You want to predict the price of houses in the Bay Area. So you collect data on house prices p, the number of rooms n, the construction year *year*, and the geographic area (SF, East Bay, South Bay, Marine County).

(a) You first want to see the impact of the number of rooms and construction year on the price. Write down the econometric model.

$$p_i = \beta_0 + \beta_1 n_i + \beta_2 \, year_i + \varepsilon_i$$

(b) Now you think that the geographic area also has a significant impact on the price of houses. Modify your econometric model to account for that. (*)

$$p_i = \beta_0 + \beta_1 n_i + \beta_2 \, year_i + \beta_3 \, SF_i + \beta_4 \, EB_i + \beta_5 \, SB_i + \varepsilon_i$$

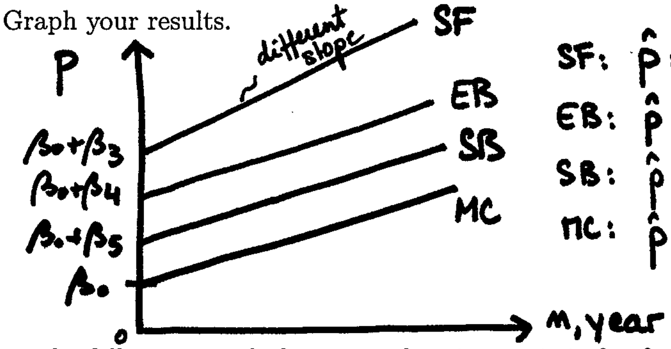
(*) Here we have a qualitative geographic var ⇒ we <u>cannot</u> include it directly. We need to create a dummy for each region and include (n−1) regions in the regression.

Why? because either you belong to one of the (n−1) regions, or if you don't, by default you'll be in the last one ⇒ redundant to include them all.

(c) You suspect that the construction year has a differential impact whether you live in SF or you don't. Modify your econometric model.

$$p_i = \beta_0 + \beta_1 n_i + \beta_2 \, year_i + \beta_3 \, SF_i + \beta_4 \, EB_i + \beta_5 \, SB_i + \beta_6 \, SF_i * year_i + \varepsilon_i$$

(d) Graph your results.



SF: $\hat{p} = \hat{\beta}_0 + \hat{\beta}_3 + (\hat{\beta}_2 + \hat{\beta}_6) year + \hat{\beta}_1 n$

EB: $\hat{p} = \hat{\beta}_0 + \hat{\beta}_4 + \hat{\beta}_2 \, year + \hat{\beta}_1 n$

SB: $\hat{p} = \hat{\beta}_0 + \hat{\beta}_5 + \hat{\beta}_2 \, year + \hat{\beta}_1 n$

MC: $\hat{p} = \hat{\beta}_0 + \hat{\beta}_2 \, year + \hat{\beta}_1 n$

3. You get the following result from your favorite statistical software.

$$crime = 1.2 - 0.76 * policemen + 1.2 * Oakland + 0.65 * Oakland * policemen \qquad (12)$$

(a) Interpret the coefficients.

if no policemen, and don't live in OAK, average crime = 1.2

if not in OAK, 1 more policeman ⇒ ↘ 0.76 crime

if in OAK, on average crime = 1.2+1.2 = 2.4

(b) Draw the regression line(s) on a graph. if in OAK, 1 more policeman ⇒ −0.76+0.65 = −0.11 crime.