Exercises 4: More Regression with R

Deadline March 3; msq@du.se

1. Consider the following model:

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + X_3 \beta_3 + U,$$

   Make the R matrix (or vector) and the r vector (or value) for the following joint null hypotheses:

   (a) $H_0 : \beta_0 = 0$ & $\beta_1 = 0$ & $\beta_2 = 0$ & $\beta_3 = 0$
   (b) $H_0 : \beta_0 = 0$ & $\beta_1 = 0$
   (c) $H_0 : \beta_0 = 1$ & $\beta_1 = 1$
   (d) $H_0 : \beta_1 = 0$
   (e) $H_0 : \beta_1 + \beta_2 = 1$

2. Consider the data *CPS1985* in the package *AER* and the following two models

$$Y = exp(X\beta + U) \quad (1)$$
$$Y = X\beta + U \quad (2)$$

   where the dependent variable is hourly *wage* and the design matrix X includes the unit vector for the intercept, *education*, *married*, *gender* and a polynomial of degree three of *experience*.

   (a) Make any necessary variable transformations and use lm() and estimate both models.
   (b) For all regression models where the goal is to estimate the effect of a regressor on average Y, we make the 'mean independence' assumption: $E(U|X) = E(U)$. Use some diagnostic to determine which one of the two model lives best up to this assumption. Discuss which model lives best up to $E(U|X) = E(U)$, but also look for other patterns and maybe outliers. Something that looks random is considered preferable.

(c) Select the model you think is the most appropriate according to the diagnostic in the previous question. Compute heteroskedasticity-robust standar errors and P-values.

(d) Continue to work with the selected model. Check for potential outliers. Check if coefficient estimates and/or significant results changes a lot when you remove potential outliers. If not, keep the first results with the "outliers" in the sample.

3. Your Professor (supervisor etc.) tells you that experience should be modelled like a polynomial of degree five. Run the regression of the model you selected in the previous questions and add a polynomial of degree five instead of the degree three polynomial (keep all other variables). Make robust T-testing.

   (a) When the Professor sees the results s/he thinks experience should be modelled linearly. Why does the Professor think so?

   (b) Test the linear model against the model where you have a polynomial of degree five (Joint testing).

   (c) Now test the previous model with a polynomial of degree three agains the model with a polynomial of degree five.

   (d) Make your own conclusions. On statistical ground what degree of polynomial would you select? Degree 1, 3 or 5?

4. Your Professor thinks that marriage has no effect whatsoever on wage since it was insignificant for the regression of the following model:

$$Y = exp(X\beta + U)$$

   where Y is hourly *wage* and the design matrix X includes the unit vector for the intercept, *education*, *married*, *gender* and a polynomial of degree three of *experience*.

   Your Professor is probably a bit naive here. It is common knowledge in empirical economics that males tend to have a wage-premium from marriage, while women may experience the opposite. One way to test this is to include an interaction term between *gender* and *married*. You do this by inserting *gender:married* as a variable in the formula for *lm()*. This variable will be equal to the following dummy:

$$D_i = \begin{cases} 1, & \text{if } i \text{ is a married woman} \\ 0, & \text{otherwise.} \end{cases}$$

(a) Estimate the model with the interaction term included.

(b) Do men have a wage premium from marriage?