**Stat 301 – hands on exercise**

The data in NLSY2.csv contains information on 2005 salaries for individuals being followed by the National Longitudinal Survey of Youth. The study started in 1981 with a random sample of US youth between the ages of 16 and 24. They have been followed regularly since. We looked at these data earlier this semester to assess the value of a college degree. The full data set includes much more information about each individual. We now use the data to evaluate (at least partially) whether men get paid more than women.

Goal: estimate the average difference in 2005 salary between men and women, after adjusting for differences in education.

Variables relevant to the question:

 gender: male or female

 female is a 0/1 indicator variable with 1 for female

 male is a 0/1 indicator variable with 0 for male

 Educ: Years of education (many values, 12 is HS graduate, 16 is college graduate)

 log income: log transformed 2005 income

Questions:

1. What model will allow us to estimate the difference in log salary, after adjusting for differences in education?

Use the female indicator variable to answer questions 2, 3, 5 and 6.

1. What is the difference in log salary, after adjusting for differences in education?
2. Is there evidence of a difference between men and women, after adjusting for differences in education?
3. What model allows us to evaluate whether the Educ slope is the same for men and women?
4. What do the data tell us about equal slopes or not?
5. If we (temporarily) assume unequal slopes, what is the Educ slope for men? for women?
6. What happens if we use the male indicator variable instead of the female indicator?

Answers:

1. -0.643

Note: this means women earn 0.643 smaller log salary, i.e. 47% less at the same education.

1. Yes, p < 0.0001 for testing = 0
2.
3. No evidence of different slopes: p = 0.76
4. men: 0.118, women: 0.114

Note: The equation can be rewritten as

 , so:

the slope for men is = 0.118

the slope for women is = 0.118 + (-0.004) = 0.114

1. Tests of interaction and sex have the same p-value, although signs of the estimates are flipped. Estimates of sex-specific regression coefficients are identical, but computed differently.

|  |  |  |
| --- | --- | --- |
| Quantity | Estimate using female | Estimate using male |
|  | 9.116 | 8.531 |
|  | 0.118 | 0.114 |
|  | -0.585 | 0.585 |
|  | -0.004 | 0.004 |
| Male intercept | 9.116 | 8.531 +0.585 = 9.116 |
| Female intercept | 9.116+(-0.585) = 8.531 | 8.531 |
| Male slope | 0.118 | 0.114+0.004 = 0.118 |
| Female slope | 0.118+(-0.004) = 0.114 | 0.114 |