# Assignment 3 – Dummies

The purpose of this excise is to have you confirm empirically the relationship that theoretically should occur among alternative simple dummy variable specifications in the linear regression model. This exercise involves examining least squares coefficients in a stylized human capital model where LNWAGE is affected by years schooling (ED) and experience (EX) with slope coefficients that are identical for everyone but where the intercept term differs by race. You should conduct this exercise using the 1987 CPS Outgoing Rotation Groups data. To keep things simple you should use only the male sample.

1. Using least squares, estimate parameter estimates in the simple regression model, estimate parameters in a simple model where LNWAGE is regressed on a constant, years of schooling (ED), years of experience (EX), and years of experience squared (EXSQ). What is the value the estimated coefficient on ED? When can this coefficient be interpreted as the rate of return to a year of schooling? What is the R-squared from this model? Interpret this R-squared.
2. The specification in part 1) assumes that the intercept and slope coefficients are identical for all individuals. Now assume instead that the effect of schooling on wages differs between whites, blacks, and Hispanics; that is years of schooling (ED) and years of experience (EX) by a factor of proportionality; that is, for whites

$$WAGE=α\_{w}.e^{β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ}.e^{ε}$$

for Blacks

$$WAGE=α\_{B}.e^{β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ}.e^{ε}$$

and Hispanics

$$WAGE=α\_{H}.e^{β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ}.e^{ε}$$

Show that this implies that when LNWAGE, rather than WAGE, is the dependent variable, white, Black, and Hispanic workers have different intercept terms but common slope coefficients. Now estimate by least squares parameters in the regression model

$$LNWAGE=γ\_{0}+γ\_{B}BLACK+γ\_{H}HISP+β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ+ε$$

where BLACK is a dummy variable equal to one if the respondent is Black, HISP is a dummy variable equal to one if the respondent is Hispanic, and εis a random disturbance term. Interpret your estimates of$γ\_{0,}γ\_{B,} and γ\_{H}$relating$γ\_{0,}γ\_{B,} and γ\_{H}$to $α\_{W,}α\_{B,} and α\_{H}$above. Formulate and test the null hypothesis that intercept terms for white, Black, and Hispanic workers are identical, using a 5% significance level. What happens to the estimated rate of return to schooling when white, Black, and Hispanic individuals have different intercepts? Can you interpret this?

1. An alternative procedure for formulating the regression relationship in part 2) is to estimate the parameters of the equation

$$LNWAGE=θ\_{W}WHITE+θ\_{B}BLACK+θ\_{H}HISP+β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ+ε$$

where WHITE is a dummy variable equal to 1 if an individual is white, 0 otherwise.[[1]](#footnote-2) Interpret$θ\_{W,}θ\_{B,} and θ\_{H}$and relate them to$γ\_{0,}γ\_{B,} and γ\_{H}$ in part 2). According to econometric theory, what should be the relationship among the estimates of $θ\_{W,}θ\_{B,} and θ\_{H}$, and $γ\_{0,}γ\_{B,} and γ\_{H}$? Why? Are you estimates numerically consistent with this relationship? Why or why not? Formulate and test the null hypothesis that the intercept terms for white, Black, and Hispanic workers are identical. Your test result should be precisely equivalent to that in part 2). Is this the case? Compare the estimates of the rate of return to schooling and R-squared here with those obtained in part 2)? Are they the same? If so, why?

1. A seemingly plausible specification is the equation

$$LNWAGE=τ\_{0}+τ\_{W}WHITE+τ\_{B}BLACK+τ\_{H}HISP+β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ+ε$$

In which LNWAGE is related to a common intercept term $τ\_{0}$, to separate terms $τ\_{W}, τ\_{B}, and τ\_{H}$ for white, Black, and Hispanic workers, and to ED and EX by common coefficients β coefficients. What is wrong with this specification? Why is least squares unable to estimate parameters of this model uniquely? What happens when you try to estimate this model using STATA? Interpret this phenomenon.

1. One other possibility is to estimate parameters in an equation in which the WHITE variable replaces the BLACK variable in the specification estimated in part 2). Such an equation would take the form

$$LNWAGE=δ\_{0}+δ\_{W}WHITE+δ\_{H}HISP+β\_{ED}ED+β\_{EX}EX+β\_{EXSQ}EXSQ+ε$$

Interpret the regression coefficients $δ\_{0,}δ\_{W,} and δ\_{H}$in this equation. What do you expect to be the relationship between the estimates of $δ\_{0,}δ\_{W,} and δ\_{H}$ and $γ\_{0,}γ\_{B,} and γ\_{H}$ in part 2)? Why? Verify your conjectures empirically byestimating the above equation.

1. By putting the race dummy variables into the statistical earnings function what we are doing, in essence, is removing the effect of mean differences in the independent variables on the dependent variable. One way to see this is to compute the means by race of the variables LNWAGE, ED, EX, and EXSQ. Call these means$\overbar{LNWAGE\_{i}}$ , $\overbar{ED\_{i}}$ , $\overbar{EX\_{i}}$ , and $\overbar{EXSQ\_{i}}$ for *i = WHITE, BLACK, HISP*. Once you do this, use these means to form new variables such that

$$LNWAGE^{\*}=LNWAGE-\overbar{LNWAGE\_{i}}$$

$$ED^{\*}=ED-\overbar{ED\_{i}}$$

$$EX^{\*}=EX-\overbar{EX\_{i}}$$

$$EXSQ^{\*}=EXSQ-\overbar{EXSQ\_{i}}$$

Basically, these new variables should be original variables minus the mean variable for the relevant racial group. If I were in the sample and my LNWAGE=4 and the average of LNWAGE for all Black Male workers in 1987 was 2.483037 then the value of $LNWAGE^{\*}$ should be 4 - 2.483037. Once you have created these variables, run the following regression (make sure that you do not include an intercept term)

$$LNWAGE^{\*}=β\_{ED}ED^{\*}+β\_{EX}EX^{\*}+β\_{EXSQ}EXSQ^{\*}+ε$$

Verify that the coefficients on education and experience variables from this regression are equivalent to those from the previous parts of the question.

According to human capital framework there might be differences in the returns to education across different racial groups if there is racial discrimination or if different racial groups differ according to other factors such as the quality of education received, average ability, or rates of labor force participation. In the remainder of this exercise, you will take a preliminary look at this empirical issue by examining whether there are differences in returns to education across racial groups through the use of dummy variables and interaction terms.

1. Take the model you estimated in part 2) and add variables that interact BLACK and HISP with ED. In particular, using least squares procedures, estimate the parameters of the simple human capital model

$$LNWAGE=γ\_{0}+γ\_{B}BLACK+γ\_{H}HISP+β\_{ED}ED+β\_{BLACK\*ED}BLACK\*ED+β\_{HISP\*ED}HISP\*ED+β\_{EX}EX+β\_{EXSQ}EXSQ+ε$$

Interpret the estimates of the regression coefficients$β\_{BLACK\*ED}$and $β\_{HISP\*ED}$ . Test the hypothesis that returns to education differ across racial groups. Do you find strong evidence that returns to education are different for whites, Blacks and Hispanics?

1. An alternative procedure for formulating the regression in part 7) is as follows. Interact the dummy variable WHITE that you created in part 3) with ED. Then estimate the following model

$$LNWAGE=γ\_{0}+γ\_{B}BLACK+γ\_{H}HISP+π\_{ED\*WHITE}ED\*WHITE+π\_{BLACK\*ED}BLACK\*ED+π\_{HISP\*ED}HISP\*ED+π\_{EX}EX+π\_{EXSQ}EXSQ+ε$$

Interpret the $π$ parameters and relate them to the β parameters from part 7). What should be the relationship between the coefficients? Why? Are your estimates numerically consistent with this relationship? Formulate and test the null hypothesis that returns to education are different across racial groups? Your test should be precisely equivalent to that in part 7). Is this the case?

1. A seemingly plausible specification is the equation

$$LNWAGE=γ\_{0}+γ\_{B}BLACK+γ\_{H}HISP+φ\_{ED}ED+φ\_{ED\*WHITE}ED\*WHITE+φ\_{BLACK\*ED}BLACK\*ED+φ\_{HISP\*ED}HISP\*ED+φ\_{EX}EX+φ\_{EXSQ}EXSQ+ε$$

What is wrong with this specification? Why is least squares unable to estimate the parameters in this model uniquely? What happens when you try to estimate this model using STATA? Interpret this phenomenon.

1. In STATA you can estimate this equation using the **reg**command with the **nocon**option. The command line needed to execute the task is exactly that given below

reglnwage white black hisped ex exsq, nocon [↑](#footnote-ref-2)