3) (A)What fraction of the variation in log wages does the model explain?

Adjusted R2 = 15.2%

(b) Test the hypothesis that

H0 : β1 = β2 = ... = β11 = 0

HA : βj != 0

for some j with α = 0.10.

We can reject the null hypothesis at the alpha = 0.1 level. The F-statistic is 267.9, corresponding to a p value < 0.01. This means that there is at least one predictor with a nonzero effect on log income.

(c) What is the return to an additional year of education? Is this statistically significant? Is it practically significant? Briefly explain.

An additional year of education results in an 8.9% decrease in income. This effect is statistically significant at the p<0.05 level. However, we should be cautious interpreting this because our model doesn’t include factors that are likely to be very linked with income, such as years of experience. Since workers often make a tradeoff between years of education and years of work experience, It seem likely that if we included years of experience in the model, education might have a positive effect on income.

(d) At what age does the model predict an individual will achieve the highest wage?

The coefficient on age is -0.03, so it is maximized at the lowest ages (18 in this dataset)

(e) Does the model predict that men or women will have higher wages, all else equal? Briefly explain why we might observe this pattern in the data.

The coefficient on the dummy variable “female” is -0.6836). This means our model predicts that women have lower wages than men, all else equal. This reflects the “gender pay gap,” in which women make less on average than men. There could be a number of causes, among them discrimination and the career/earnings interruptions of pregnancy and childcare.

(f) Interpret the coefficients on the white, black, and hispanic variables.

The coefficient on white indicates that being white has the effect of raising income by 41.28%. Being black is associated with a 6.54% decrease in income, and being Hispanic is associated with a 66.36% decrease in income.

(g) Test the hypothesis that race has no effect on wages. Be sure to explicitly state the null and alternative hypotheses and show your calculations.

To test this hypothesis, I applied the formula

5. Since the President is considering new education legislation, she asks you to determine whether a college degree is a strong predictor of wages. Write down a model that will allow the returns to education to vary by degree acquired (use the three categories in the previous question). Be sure to include the controls from question 3. Explain/justify why you think your model is the best possible representation of the way the world works.

This model allows us to better understand the effect of receiving a highschool and college diploma. This is a better representation of how the world works than our previous model because it seems likely that income will be affected not only by the number of years of schooling, but also by the degrees achieved. In the previous model, we had educdc as our only education variable, but by including the hs diploma and college diploma binary variables, we can control for the effect of these degrees.

6. Estimate the model you proposed in the previous question and report your results.

Having a highschool diploma is associated with a 50% decrease in income. Having a college diploma is associated with a 205% increase in income. Both of these are significant at the p<0.05 level.

(a) Predict the wages of an 22 year old, female individual (who is neither white, black, nor Hispanic, is not married, has no children, and is not a veteran) with a high school diploma and an all else equal individual with a college diploma. Assume that it takes someone 12 years to graduate high school and 16 years to graduate college.

(b) The President wants to know, given your results, do individuals with college degrees have higher predicted wages than those without? By how much? Briefly explain.

Yes, individuals with college degrees have higher predicted wages than those without. Having a college diploma is associated with a 205% increase in income, all else equal.

(c) The President asked you to look into this question because she is considering legislation that will expand access to college education (for instance, by increasing student loan subsidies). She will only support the legislation if there are cost offsets (if college education increases wages and therefore, future income tax revenues that help reduce the net cost of the subsidy). Given that criteria, how would you advise the President?

If I were being honest with the president, I’d advise her that we need more funding to run additional regressions. The reality is that even though we have disentangled the effects of years of education and receipt of diploma, we don’t know many of the factors that determine whether someone is likely to receive a diploma. For instance, it could be that individuals who attain college diplomas are also more likely to have high income parents. If that’s the case, the observed effect may be due to the positive effects of high parental income rather than from attaining a diploma.

7. There are many ways that this model could be improved. How would you do things differently if you were asked to predict the returns to education given the data available on IPUMS?

I would include additional variables in the regression, being careful to articulate clear theoretical justifications for each and employing a variable selection algorithm to help keep my list of variables to a reasonable size. Some IPUMS variables that seem promising include:

URBAN Urban/rural status

CITIZEN Citizenship status

YRSUSA1 Years in the United States

DEGFIELD Field of degree