Michael Wells

Problem Set 3

12/17/20

***1b. Give a precise interpretation of the β associated with income. (Hint, if it’s helpful, multiple the regression estimate by 100).***

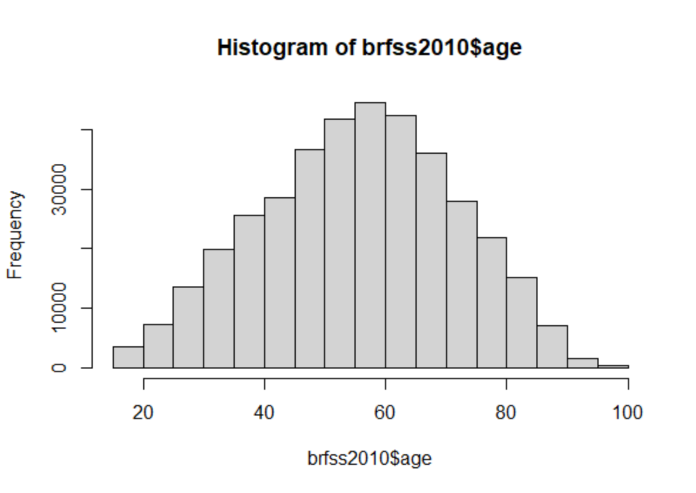
For every $1 increase in income the average probably of that person smoking would decrease by 0.0001635%.

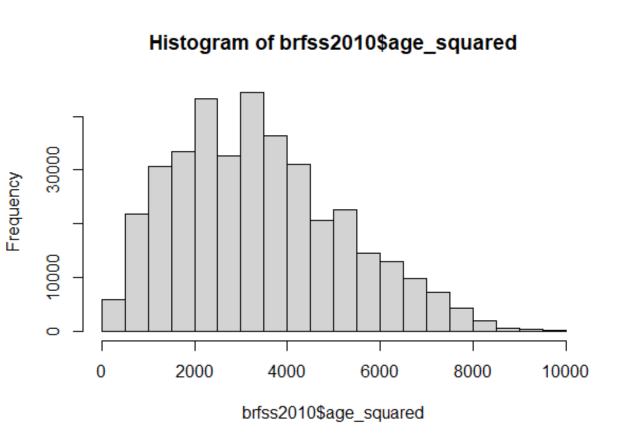
***1c. Can you reject the null hypothesis that the parameter associated with age is 0?***

Yes, p-val associated with age is less than .1.

***1d. Add to the regression model an age squared term. Interpret the resulting βs on age and age squared. What evidence can you put forth that adding age squared to the model improved the model’s ability to explain the variation in smoking?***

Interesting to me that the age estimate is now positive, when in the initial regression it was negative. To me it intuitively makes sense as I would predict smoking is not linear with age. The age estimate would be interpreted saying for every increase in year, the probability of smoking would increase on average by 0.012, but for every increase in age squared, the probability of smoking would decrease on average by 0.0014. Some evidence that adding the square term improved the model is that the R^2 value increased from .08614 to .09905.

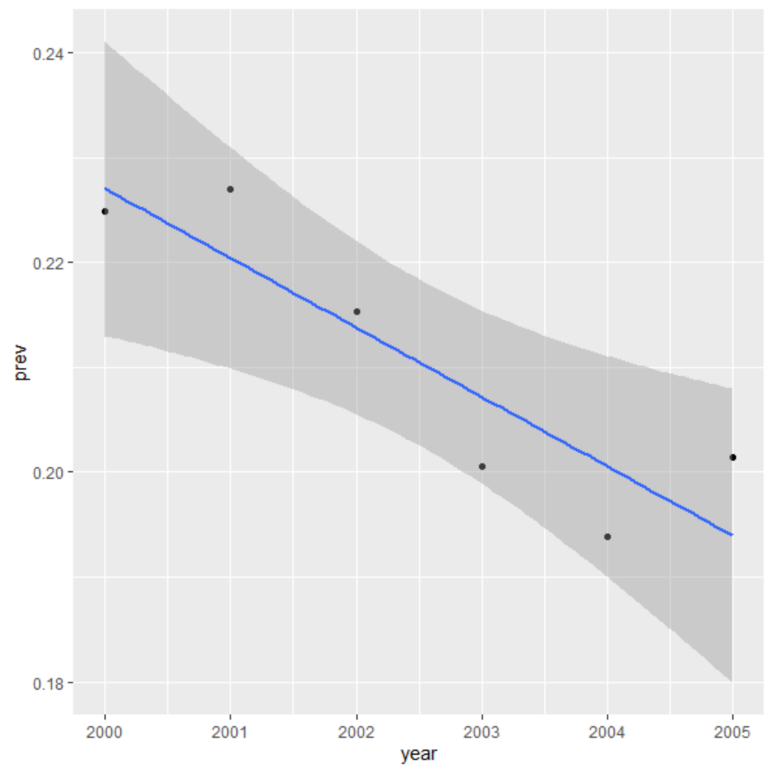




***1f. Interpret each of the education parameter estimates in your regression. How much more or less likely is smoking for someone with education of 4 vs. 1? Can you reject the null hypothesis that there is no difference?***

All of the education parameters show there seems to be significant impact on smoking probability. Each pval < .1 thus we can reject the null hypothesis. Using education of 1 as the reference, the greatest difference is with education level of 4. If you have an education of 4 you are 13% less likely to smoke than if you have an education of 1.

***2b. Generate a graph of smoking prevalence over time.***



***2c. Estimate a regression model of smoking on year in which you allow linear time trends to vary by state. Can you reject the null hypothesis that linear trends in smoking prevalence are the same for each state? Hint - it is helpful to rescale year. Instead of 2000, 2001, etc. you should rescale to 0, 1, etc.***

Yes, you could reject the null as pval = .02985 < .1

***2d. Now create binary variables for each year. Estimate a regression model of smoking on each year in which you allow smoking prevalence to vary by state. Are any of the year\*state interactions statistically different than zero?***

Yes 2002 is one year that the year\*sate interactions are statistically different than zero, pval = .042484 < .1