Homework #1 - Joe Risi

Packages

Question 1

1. I filter out all missing observations. I go from 12144 records to 10200 records.
2. Recode family income into roughly equal sizes. The new categories are as follows:
    - 0 - 9,999
    - 10,000 - 19,999
    - 20,000 - 24,999
    - 25,000 - 34,999
    - 35,000 - 49,999
    - 50,000 - 74,999
    - 75,000 and above
3. I turn sex, byfaminc, and bys45 into dummy variables.
4. I drop the following categories which will serve as the reference categories for the regression:
    - female (sex)
    - 75,000 and above (byfaminc)
    - higher.sch.aftr.coll (bys45)

```
##
## Call:
## lm(formula = bygrads ~ ., data = dataWide)
##
## Residuals:
##      Min       1Q   Median       3Q      Max
## -2.81319 -0.41509  0.05336  0.48159  2.00395
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            3.51841    0.02306 152.560  < 2e-16 ***
## won.t.finish.h.s      -1.17818    0.06214 -18.959  < 2e-16 ***
## will.finish.h.s       -0.92871    0.02587 -35.903  < 2e-16 ***
## voc.trd.bus.aftr.h.s -0.65136    0.02593 -25.115  < 2e-16 ***
## will.attend.college   -0.60532    0.02253 -26.870  < 2e-16 ***
## will.finish.college   -0.24041    0.01626 -14.788  < 2e-16 ***
## Less.than..10.000     -0.34418    0.02976 -11.567  < 2e-16 ***
## `10.000....19.999`    -0.22947    0.02747  -8.355  < 2e-16 ***
## `20.000..24.999`      -0.16325    0.02952  -5.530 3.28e-08 ***
## `25.000..34.999`      -0.15089    0.02622  -5.754 8.95e-09 ***
## `35.000..49.999`      -0.10332    0.02552  -4.049 5.19e-05 ***
## `50.000..74.999`      -0.06352    0.02715  -2.340   0.0193 *
## male                  -0.10190    0.01285  -7.932 2.38e-15 ***
## ---
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.6448 on 10187 degrees of freedom
## Multiple R-squared:  0.2226, Adjusted R-squared:  0.2217
```

```
## F-statistic: 243.1 on 12 and 10187 DF,  p-value: < 2.2e-16
```

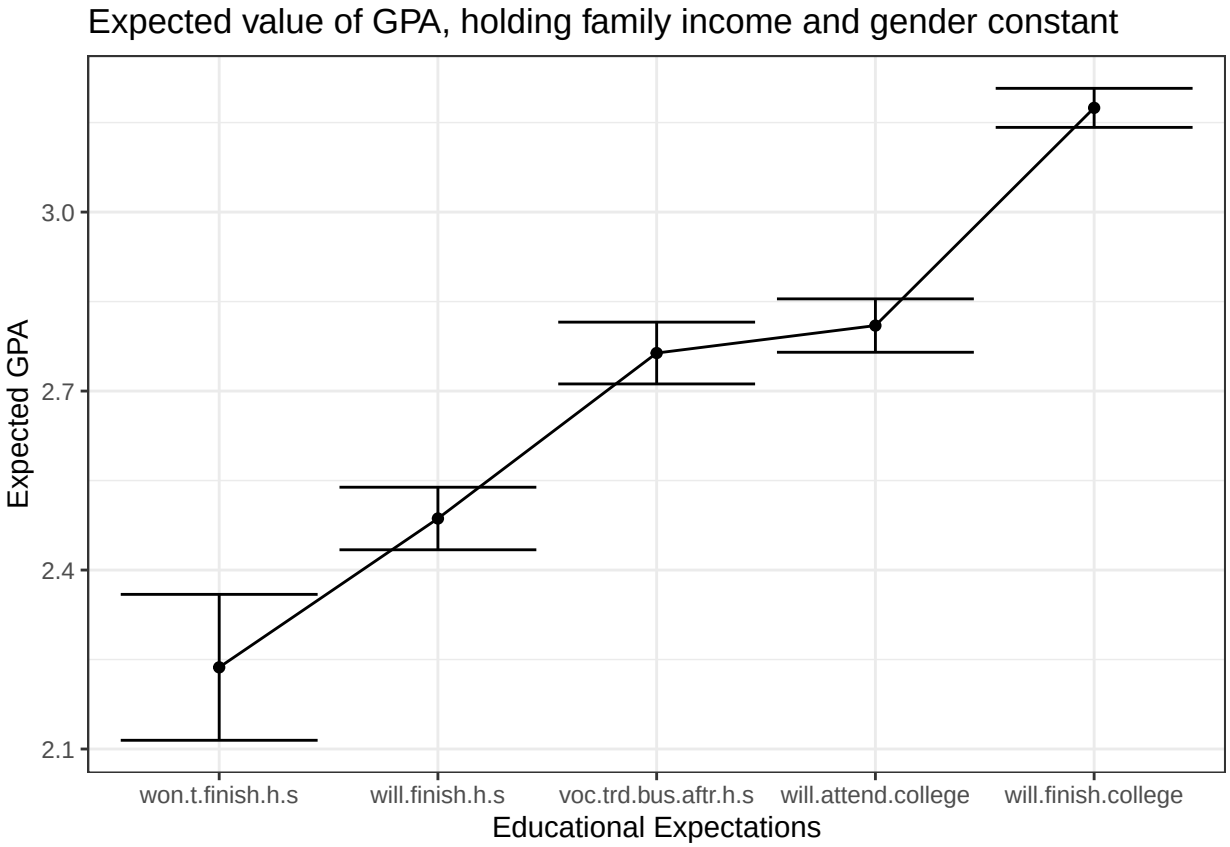
- All results are significant at the 0.05 level. All except one ($50,000 - %74,999) are significant at the 0.001 level.
- Relative to females while holding all other variables in the model constant, being male decreases one's GPA by about 0.1 points, on average.
- Relative to those students who come from families making more $75,000 or more each year while holding all other variables in the model constant:
  - Coming from a family making less than $10,000 decreases your GPA by 0.34418 points on average.
  - Coming from a family making between $10,000 - $19,999 decreases your GPA by 0.22947 points on average.
  - Coming from a family making between $20,000 - $24,999 decreases your GPA by 0.16325 points on average.
  - Coming from a family making between $25,000 - $34,999 decreases your GPA by 0.15089 points on average.
  - Coming from a family making between $35,000 - $49,999 decreases your GPA by 0.10332 points on average.
  - Coming from a family making between $50,000 - $74,999 decreases your GPA by 0.06352 points on average.
- Relative to those students who have expectations of going beyond their college education while holding all other variables in the model constant:
  - Having expectations of not finishing high school decreases your GPA by 1.17818 points on average.
  - Having expectations of just finishing high school decreases your GPA by 0.92871 points on average.
  - Having expectations of going to vocational/trade school decreases your GPA by 0.65136 points on average.
  - Having expectations of attending college decreases your GPA by 0.60532 points on average.
  - Having expectations of finishing college decreases your GPA by 0.24041 points on average.

```
##
## Call:
## lm(formula = c("bygrads.z ~ won.t.finish.h.s + will.finish.h.s + voc.trd.bus.aftr.h.s + ",
## "    will.attend.college + will.finish.college + Less.than..10.000 + ",
## "    `10.000....19.999` + `20.000..24.999` + `25.000..34.999` + ",
## "    `35.000..49.999` + `50.000..74.999` + male"), data = data)
##
## Residuals:
##     Min      1Q  Median      3Q     Max
## -3.8492 -0.5679  0.0730  0.6589  2.7419
##
## Coefficients:
##                       Estimate Std. Error t value Pr(>|t|)
## (Intercept)            0.72593    0.03156  23.005  < 2e-16 ***
## won.t.finish.h.s      -1.61205    0.08503 -18.959  < 2e-16 ***
## will.finish.h.s       -1.27070    0.03539 -35.903  < 2e-16 ***
## voc.trd.bus.aftr.h.s -0.89122    0.03549 -25.115  < 2e-16 ***
## will.attend.college   -0.82823    0.03082 -26.870  < 2e-16 ***
## will.finish.college   -0.32894    0.02224 -14.788  < 2e-16 ***
## Less.than..10.000     -0.47093    0.04071 -11.567  < 2e-16 ***
## `10.000....19.999`    -0.31397    0.03758  -8.355  < 2e-16 ***
## `20.000..24.999`      -0.22336    0.04039  -5.530 3.28e-08 ***
## `25.000..34.999`      -0.20646    0.03588  -5.754 8.95e-09 ***
## `35.000..49.999`      -0.14137    0.03492  -4.049 5.19e-05 ***
## `50.000..74.999`      -0.08691    0.03715  -2.340   0.0193 *
## male                  -0.13942    0.01758  -7.932 2.38e-15 ***
## ---
```

```
## Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
##
## Residual standard error: 0.8822 on 10187 degrees of freedom
## Multiple R-squared:  0.2226, Adjusted R-squared:  0.2217
## F-statistic: 243.1 on 12 and 10187 DF,  p-value: < 2.2e-16
```

The above results represent **y-standardized** coefficients. It does not make sense to standardize my independent variables because they are all dummy variables. It's hard to interpret what a standard deviation in a dummy variable would mean.

- All results are significant at the 0.05 level. All except one ($50,000 - %74,999) are significant at the 0.001 level.
- Relative to females while holding all other variables in the model constant, being male decreases one's GPA by about 0.13942 standard deviations on average.
- Relative to those students who come from families making more $75,000 or more each year while holding all other variables in the model constant:
  - Coming from a family making less than $10,000 decreases your GPA by 0.47093 standard deviations.
  - Coming from a family making between $10,000 - $19,999 decreases your GPA by 0.31397 standard deviations on average.
  - Coming from a family making between $20,000 - $24,999 decreases your GPA by 0.22336 standard deviations on average.
  - Coming from a family making between $25,000 - $34,999 decreases your GPA by 0.20646 standard deviations on average.
  - Coming from a family making between $35,000 - $49,999 decreases your GPA by 0.14137 standard deviations on average.
  - Coming from a family making between $50,000 - $74,999 decreases your GPA by 0.08691 standard deviations on average.
- Relative to those students who have expectations of going beyond their college education while holding all other variables in the model constant:
  - Having expectations of not finishing high school decreases your GPA by 1.61205 standard deviations on average.
  - Having expectations of just finishing high school decreases your GPA by 1.27070 standard deviations on average.
  - Having expectations of going to vocational/trade school decreases your GPA by 0.89122 standard deviations on average.
  - Having expectations of attending college decreases your GPA by 0.82823 standard deviations on average.
  - Having expectations of finishing college decreases your GPA by 0.32894 standard deviations on average.



Because I used categorical variables for everything in my model, I had to set family income and sex equal to their modal values ($35,000 - $49,999 and female) to hold them constant.