# Date: Feb. 6, 2019

# Author: Wei Li, Florence Yan, Ziyu Jin

Research question

* In the state of Washington, how do earnings vary by educational attainment?
* Does the premium for higher education vary by race?

Research area

* Washington state

Our first model:

## Call:

## lm(formula = Earnings ~ EducLevel + Gender + Citizenship + Race +

## Marrige\_Status + Employment\_Status + Age + Work\_Type, data = new)

##

## Residuals:

## Min 1Q Median 3Q Max

## -200977 -29621 -8938 14277 788649

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## Intercept 37425.19 1808.26 20.697 < 2e-16 \*\*\*

## CollegeDgr 29011.09 938.03 30.928 < 2e-16 \*\*\*

## ProfessionalDgr 107724.46 2662.64 40.458 < 2e-16 \*\*\*

## MasterDgr 51801.02 1406.12 36.840 < 2e-16 \*\*\*

## DocterDgr 71586.55 3158.02 22.668 < 2e-16 \*\*\*

## Female -22792.18 866.52 -26.303 < 2e-16 \*\*\*

## NonUSCitizen -4574.31 1725.96 -2.650 0.00805 \*\*

## Black -12742.95 2567.96 -4.962 7.02e-07 \*\*\*

## IndianOrAlaska 1098.38 3376.53 0.325 0.74496

## Asian 2972.92 1558.33 1.908 0.05643 .

## OtherRace -16727.40 2259.70 -7.402 1.38e-13 \*\*\*

## MultipleRaces -2270.19 2264.81 -1.002 0.31618

## NotMarried -15686.84 906.41 -17.307 < 2e-16 \*\*\*

## ArmedForcesWork 2971.11 3083.92 0.963 0.33535

## Age 740.19 34.77 21.288 < 2e-16 \*\*\*

## WorkForGovernment -18576.68 1119.41 -16.595 < 2e-16 \*\*\*

## WorkForOwn 3343.65 1565.11 2.136 0.03266 \*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 62330 on 22416 degrees of freedom

## Multiple R-squared: 0.2096, Adjusted R-squared: 0.209

## F-statistic: 371.4 on 16 and 22416 DF, p-value: < 2.2e-16

BPG test results for mod1:

## studentized Breusch-Pagan test

##

## data: mod1

## BP = 1118.1, df = 16, p-value < 2.2e-16

* With a p-value so close to 0, we have strong evidence against the null hypothesis of homoscedasticity.

Change the model and take the log(earnings) to be the dependent variable.

Our second model:

## lm(formula = log(Earnings) ~ EducLevel + Gender + Citizenship +

## Race + Marrige\_Status + Employment\_Status + Age + Work\_Type,

## data = new)

##

## Residuals:

## Min 1Q Median 3Q Max

## -10.3029 -0.3407 0.0413 0.4091 3.2008

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## Intercept 10.3104951 0.0207287 497.402 < 2e-16 \*\*\*

## CollegeDgr 0.4313447 0.0107530 40.114 < 2e-16 \*\*\*

## ProfessionalDgr 1.0103306 0.0305227 33.101 < 2e-16 \*\*\*

## MasterDgr 0.6792253 0.0161188 42.139 < 2e-16 \*\*\*

## DocterDgr 0.8087129 0.0362014 22.339 < 2e-16 \*\*\*

## Female -0.3084009 0.0099333 -31.047 < 2e-16 \*\*\*

## NonUSCitizen -0.1446371 0.0197853 -7.310 2.75e-13 \*\*\*

## Black -0.2183844 0.0294374 -7.419 1.23e-13 \*\*\*

## IndianOrAlaska -0.1010374 0.0387063 -2.610 0.00905 \*\*

## Asian 0.0267050 0.0178636 1.495 0.13495

## OtherRace -0.3069301 0.0259037 -11.849 < 2e-16 \*\*\*

## MultipleRaces -0.0449969 0.0259623 -1.733 0.08308 .

## NotMarried -0.2459349 0.0103905 -23.669 < 2e-16 \*\*\*

## ArmedForcesWork -0.0750892 0.0353520 -2.124 0.03368 \*

## Age 0.0126129 0.0003986 31.644 < 2e-16 \*\*\*

## WorkForGovernment -0.1164410 0.0128321 -9.074 < 2e-16 \*\*\*

## WorkForOwn -0.2304184 0.0179414 -12.843 < 2e-16 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.7145 on 22416 degrees of freedom

## Multiple R-squared: 0.2644, Adjusted R-squared: 0.2639

## F-statistic: 503.6 on 16 and 22416 DF, p-value: < 2.2e-16

BPG test results for mod2:

## studentized Breusch-Pagan test

##

## data: mod2

## BP = 460.41, df = 16, p-value < 2.2e-16

* With a p-value so close to 0, we still have strong evidence against the null hypothesis of homoscedasticity even if we take the log of earnings

So we tried to correct the standard errors by using the sandwich package, and the following is the corrected SEs:

## Intercept CollegeDgr ProfessionalDgr MasterDgr

## 0.0218350851 0.0107127244 0.0366087996 0.0157689189

## DocterDgr Female NonUSCitizen Black

## 0.0370649752 0.0100243392 0.0218232032 0.0308746167

## IndianOrAlaska Asian OtherRace MultipleRaces

## 0.0439593029 0.0184704754 0.0281744096 0.0284546195

## NotMarried ArmedForcesWork Age WorkForGovernment

## 0.0103708991 0.0254209516 0.0004229597 0.0111572006

## WorkForOwn

## 0.0274366661

Our third model (added age^2):

## Call:

## lm(formula = log(Earnings) ~ EducLevel + Gender + Citizenship +

## Race + Marrige\_Status + Employment\_Status + Age + I(Age \*

## Age) + Work\_Type, data = new)

##

## Residuals:

## Min 1Q Median 3Q Max

## -10.4542 -0.3370 0.0405 0.3982 2.9998

##

## Coefficients:

## Estimate Std. Error t value Pr(>|t|)

## (Intercept) 8.687e+00 5.459e-02 159.133 < 2e-16 \*\*\*

## EducLevelB 4.117e-01 1.053e-02 39.083 < 2e-16 \*\*\*

## EducLevelC 9.910e-01 2.985e-02 33.194 < 2e-16 \*\*\*

## EducLevelD 6.479e-01 1.579e-02 41.023 < 2e-16 \*\*\*

## EducLevelE 8.082e-01 3.540e-02 22.829 < 2e-16 \*\*\*

## Gender2 -3.089e-01 9.714e-03 -31.797 < 2e-16 \*\*\*

## Citizenship2 -1.642e-01 1.936e-02 -8.483 < 2e-16 \*\*\*

## Race2 -2.277e-01 2.879e-02 -7.911 2.67e-15 \*\*\*

## Race3 -1.192e-01 3.785e-02 -3.149 0.00164 \*\*

## Race4 2.344e-02 1.747e-02 1.342 0.17968

## Race5 -3.020e-01 2.533e-02 -11.921 < 2e-16 \*\*\*

## Race6 -3.738e-02 2.539e-02 -1.472 0.14094

## Marrige\_Status2 -1.762e-01 1.039e-02 -16.956 < 2e-16 \*\*\*

## Employment\_Status4 5.032e-02 3.479e-02 1.446 0.14807

## Age 9.199e-02 2.509e-03 36.668 < 2e-16 \*\*\*

## I(Age \* Age) -8.961e-04 2.798e-05 -32.029 < 2e-16 \*\*\*

## Work\_TypeB -1.165e-01 1.255e-02 -9.280 < 2e-16 \*\*\*

## Work\_TypeC -2.138e-01 1.755e-02 -12.179 < 2e-16 \*\*\*

## ---

## Signif. codes: 0 '\*\*\*' 0.001 '\*\*' 0.01 '\*' 0.05 '.' 0.1 ' ' 1

##

## Residual standard error: 0.6988 on 22415 degrees of freedom

## Multiple R-squared: 0.2966, Adjusted R-squared: 0.2961

## F-statistic: 556 on 17 and 22415 DF, p-value: < 2.2e-16

* By adding age^2 to the model, we increased the R-squared from 0.26-0.30.