Final Exam

Michael Huang

EN.625.603.84

Description: The datafile contains data for 2015 for full-time workers with a high school diploma or B.A./B.S. as their highest degree. See the pdf attachment for an overview of the data and variable descriptions. In this exercise, you will investigate the relationship between a worker's age and earnings. (Generally, older workers have more job experience, leading to higher productivity and higher earnings.)

In [ ]:

```
from statsmodels.formula.api import ols

import numpy as np
import pandas as pd

# As excel files do not play nice with Python, converted the data to csv

cps_df = pd.read_csv("CPS2015-1.csv")
cps_df
```

Out[ ]:

	year	ahe	bachelor	female	age
0	2015	11.778846	0	0	26
1	2015	9.615385	0	1	33
2	2015	12.019231	0	0	31
3	2015	18.376068	0	0	32
4	2015	41.836735	0	0	28
...	...	...	...	...	...
7093	2015	96.153847	1	0	25
7094	2015	30.769230	1	0	34
7095	2015	9.230769	0	0	27
7096	2015	13.653846	1	1	27
7097	2015	32.692307	1	1	34

7098 rows × 5 columns

a. Run a regression of average hourly earnings (AHE) on age (Age), gender (Female), and education (Bachelor). If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?

In [32]:

```
model_a = ols("ahe ~ age + female + bachelor", data=cps_df).fit()
print(f"{model_a.summary()=}")

age_coeff_a = model_a.params["age"]
print(f"{age_coeff_a=}")
```

```
model_a.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
==
Dep. Variable:                    ahe   R-squared:                       0.1
90
Model:                            OLS   Adj. R-squared:                  0.1
89
Method:                 Least Squares   F-statistic:                     55
3.4
Date:                Sat, 16 Aug 2025   Prob (F-statistic):          3.46e-3
23
Time:                        06:05:18   Log-Likelihood:                -2703
6.
No. Observations:                7098   AIC:                         5.408e+
04
Df Residuals:                    7094   BIC:                         5.411e+
04
Df Model:                           3
Covariance Type:            nonrobust
==============================================================================
==
                 coef    std err          t      P>|t|      [0.025      0.97
5]
------------------------------------------------------------------------------
--
Intercept      2.0448      1.355      1.509      0.131      -0.611       4.7
00
age            0.5313      0.045     11.788      0.000       0.443       0.6
20
female        -4.1435      0.266    -15.583      0.000      -4.665      -3.6
22
bachelor       9.8456      0.262     37.519      0.000       9.331      10.3
60
==============================================================================
==
Omnibus:                     2458.198   Durbin-Watson:                   1.9
36
Prob(Omnibus):                  0.000   Jarque-Bera (JB):            11294.1
66
Skew:                           1.629   Prob(JB):                         0.
00
Kurtosis:                       8.252   Cond. No.                         31
2.
==============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
ctly specified.
"""
age_coeff_a=0.531275239654071
```

In this model, we simply look at the coefficient for age, since this simply represents the marginal change for if age increases by 1, so this is the same for both if the age increases from 25 to 26 and if the age increases from 33 to 34. The corresponding coefficient as displayed above is 0.531275239654071, so there would be a **0.531275239654071** increase in average hourly earnings.

b. Run a regression of the logarithm of average hourly earnings, ln(AHE), on Age, Female, and Bachelor. If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?

In [26]:
```
cps_df["ln_ahe"] = np.log(cps_df["ahe"])

model_b = ols("ln_ahe ~ age + female + bachelor", data=cps_df).fit()
print(f"{model_b.summary()=}")

age_coeff_b = model_b.params["age"]
print(f"{age_coeff_b=}")
```

```
model_b.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
=============================================================================
==
Dep. Variable:                 ln_ahe   R-squared:                       0.2
08
Model:                            OLS   Adj. R-squared:                  0.2
08
Method:                 Least Squares   F-statistic:                     62
2.4
Date:                Sat, 16 Aug 2025   Prob (F-statistic):               0.
00
Time:                        05:55:27   Log-Likelihood:                -482
1.9
No. Observations:                7098   AIC:                             965
2.
Df Residuals:                    7094   BIC:                             967
9.
Df Model:                           3
Covariance Type:            nonrobust
=============================================================================
==
                 coef    std err          t      P>|t|      [0.025      0.97
5]
-----------------------------------------------------------------------------
--
Intercept      2.0274      0.059     34.220      0.000       1.911       2.1
43
age            0.0242      0.002     12.273      0.000       0.020       0.0
28
female        -0.1776      0.012    -15.274      0.000      -0.200      -0.1
55
bachelor       0.4615      0.011     40.212      0.000       0.439       0.4
84
=============================================================================
==
Omnibus:                      185.302   Durbin-Watson:                   1.9
43
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              309.1
07
Skew:                          -0.236   Prob(JB):                     7.55e-
68
Kurtosis:                       3.906   Cond. No.                         31
2.
=============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
ctly specified.
"""
age_coeff_b=0.02419115791114511
```

In this model, we simply look at the coefficient for age, and in this log-linear model it represents the percentage change for if age increases by 1, so this percentage change would be the same for both if the age increases from 25 to 26 and if the age increases from 33 to 34. The corresponding coefficient as displayed above is 0.02419115791114511, so there would be a **2.419115791114511%** increase in average hourly earnings.

c. Run a regression of the logarithm of average hourly earnings, ln(AHE), on ln(Age), Female, and Bachelor. If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?

In [27]:
```
cps_df["ln_age"] = np.log(cps_df["age"])

model_c = ols("ln_ahe ~ ln_age + female + bachelor", data=cps_df).fit()
print(f"{model_c.summary()=}")

ln_age_coeff_c = model_c.params["ln_age"]
print(f"{ln_age_coeff_c=}")

pct_change_age_25_to_26 = ((26-25)/25) * 100
pct_change_age_33_to_34 = ((34-33)/33) * 100

earnings_change_25_to_26 = ln_age_coeff_c * pct_change_age_25_to_26
earnings_change_33_to_34 = ln_age_coeff_c * pct_change_age_33_to_34

print(f"{earnings_change_25_to_26=}")
print(f"{earnings_change_33_to_34=}")
```

```
model_c.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results                            
==============================================================================
==
Dep. Variable:                 ln_ahe   R-squared:                       0.2
09
Model:                            OLS   Adj. R-squared:                  0.2
08
Method:                 Least Squares   F-statistic:                     62
3.4
Date:                Sat, 16 Aug 2025   Prob (F-statistic):               0.
00
Time:                        05:55:27   Log-Likelihood:                -482
0.8
No. Observations:                7098   AIC:                             965
0.
Df Residuals:                    7094   BIC:                             967
7.
Df Model:                           3                                         
Covariance Type:            nonrobust                                         
==============================================================================
==
                 coef    std err          t      P>|t|      [0.025      0.97
5]
------------------------------------------------------------------------------
--
Intercept      0.3233      0.196      1.649      0.099      -0.061       0.7
08
ln_age         0.7154      0.058     12.368      0.000       0.602       0.8
29
female        -0.1775      0.012    -15.268      0.000      -0.200      -0.1
55
bachelor       0.4615      0.011     40.220      0.000       0.439       0.4
84
==============================================================================
==
Omnibus:                      184.684   Durbin-Watson:                   1.9
43
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              307.7
70
Skew:                          -0.236   Prob(JB):                      1.47e-
67
Kurtosis:                       3.904   Cond. No.                         13
0.
==============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
ctly specified.
"""
ln_age_coeff_c=0.7153749610992853
earnings_change_25_to_26=2.8614998443971413
earnings_change_33_to_34=2.1678029124220766
```

In this model, we now have a log-log model for which the coefficient of the logarithmic term represents the elasticity of the variable of earnings changing with age, that is, the percentage amount that the earnings will change as the age changes. Looking at the coefficient, we see that a 1% increase in age leads to a 0.7153749610992853% change in the earnings, so we can do simple arithmetic by multiplying to convert the percentage change in age to the relevant percentage change in earnings. Doing this, we find that the change for 25 to 26 is **2.8614998443971413%** and the change for 33 to 34 is **2.1678029124220766%**.

d. Run a regression of the logarithm of average hourly earnings, ln(AHE), on Age, Age^2, Female, and Bachelor. If age increases from 25 to 26, how are earnings expected to change? If age increases from 33 to 34, how are earnings expected to change?

In [ ]:

```
cps_df["age_squared"] = cps_df["age"] ** 2

model_d = ols("ln_ahe ~ age + age_squared + female + bachelor", data=cps_df)
print(f"{model_d.summary()=}")

def marginal_age_effect(age):
    age_coeff = model_d.params["age"]
    age_squared_coeff = model_d.params["age_squared"]
    return age_coeff + 2 * age_squared_coeff * age

print(f"{marginal_age_effect(25)=}")
print(f"{marginal_age_effect(33)=}")
```

```
model_d.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results                            
==============================================================================
==
Dep. Variable:                 ln_ahe   R-squared:                       0.2
09
Model:                            OLS   Adj. R-squared:                  0.2
09
Method:                 Least Squares   F-statistic:                     46
8.6
Date:                Sat, 16 Aug 2025   Prob (F-statistic):               0.
00
Time:                        06:17:30   Log-Likelihood:                -481
9.1
No. Observations:                7098   AIC:                             964
8.
Df Residuals:                    7093   BIC:                             968
2.
Df Model:                           4                                         
Covariance Type:            nonrobust                                         
==============================================================================
===
                 coef    std err          t      P>|t|      [0.025      0.9
75]
------------------------------------------------------------------------------
---
Intercept      0.4187      0.672      0.623      0.533      -0.899       1.
736
age            0.1341      0.046      2.929      0.003       0.044       0.
224
age_squared   -0.0019      0.001     -2.403      0.016      -0.003      -0.
000
female        -0.1774      0.012    -15.256      0.000      -0.200      -0.
155
bachelor       0.4616      0.011     40.236      0.000       0.439       0.
484
==============================================================================
==
Omnibus:                      182.315   Durbin-Watson:                   1.9
44
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              302.7
31
Skew:                          -0.234   Prob(JB):                      1.83e-
66
Kurtosis:                       3.897   Cond. No.                      1.07e+
05
==============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
ctly specified.
[2] The condition number is large, 1.07e+05. This might indicate that there
are
strong multicollinearity or other numerical problems.
```

```
"""
marginal_age_effect(25)=0.041100786801177264
marginal_effect_25=0.041100786801177264
marginal_effect_33=0.011336188193903682
```

We have to find the marginal effect by taking the derivative relative to age in the model, i.e.:

> ln(ahe) = age_coeff * age + age_squared_coeff * age^2 + other terms

> d/age ln(ahe) = age_coeff + 2 * age_squared_coeff * age

Then we can find the marginal effect for the two ages by plugging in the values of 25 and 33, respectively, and then multiplying by the coefficient for age to find the percentage change in earnings. Calculating this, we find that the change for 25 to 26 is **4.1100786801177264%** and the change for 33 to 34 is **1.1336188193903682%**.

e. Do you prefer the regression in (c) to the regression in (b)? Explain.

I would prefer the regression in (c) because the regression in (b) only measures absolute changes as a result of the variables being changed, which is not really useful or comparable across different ranges of earners, while the regression in (c) is better in being able to measure both varying returns and relative at different ages which is much more useful. The ability to intuitively understand the elasticity of earnings as age increases is more useful than absolute changes.

f. Do you prefer the regression in (d) to the regression in (b)? Explain.

I would prefer the regression in (d) because as previously stated, the regression in (b) only measures absolute changes as a result of the variables being changed, which is not really useful or comparable across different ranges of earners, while the regression in (d) is better in being able to measure both varying returns and relative at different ages which is much more useful. Being able to measure the marginal effect of age on earnings is more useful than the absolute changes themselves.

g. Do you prefer the regression in (d) to the regression in (c)? Explain.

I would prefer the regression in (d). While both are adaptable in showing effects on earnings that change as age increases, the regression in (d) is more flexible in that it allows for a non-linear relationship between age and earnings, which is likely more realistic and accurate as this relationship is highly complex in the real world. In addition, the elasticity as specified in the regression in (c) is constant over time, which seems to be a bit simplistic.

h. Run a regression of ln(AHE), on Age, Age^2, Female, Bachelor, and the interaction term Female*Bachelor. What does the coefficient on the interaction term measure? Alexis is a 30-year-old female with a bachelor’s degree. What does the regression predict for her value of ln(AHE)? Jane is a 30-year-old female with a high school

degree. What does the regression predict for her value of ln(AHE)? What is the predicted difference between Alexis's and Jane's earnings? Bob is a 30-year-old male with a bachelor's degree. What does the regression predict for his value of ln(AHE)? Jim is a 30-year-old male with a high school degree. What does the regression predict for his value of ln(AHE)? What is the predicted difference between Bob's and Jim's earnings?

In [45]:

```
cps_df["female_bachelor"] = cps_df["female"] * cps_df["bachelor"]

model_h = ols("ln_ahe ~ age + age_squared + female + bachelor + female_bache
print(f"{model_h.summary()=}")

alexis_ln_ahe = (model_h.params["Intercept"] +
                 model_h.params["age"] * 30 +
                 model_h.params["age_squared"] * 30 ** 2 +
                 model_h.params["female"] * 1 +
                 model_h.params["bachelor"] * 1 +
                 model_h.params["female_bachelor"] * 1)
print(f"{alexis_ln_ahe=}")

jane_ln_ahe = (model_h.params["Intercept"] +
               model_h.params["age"] * 30 +
               model_h.params["age_squared"] * 30 ** 2 +
               model_h.params["female"] * 1 +
               model_h.params["bachelor"] * 0 +
               model_h.params["female_bachelor"] * 0)
print(f"{jane_ln_ahe=}")

alexis_jane_diff = np.exp(alexis_ln_ahe) - np.exp(jane_ln_ahe)
print(f"{alexis_jane_diff=}")

bob_ln_ahe = (model_h.params["Intercept"] +
              model_h.params["age"] * 30 +
              model_h.params["age_squared"] * 30 ** 2 +
              model_h.params["female"] * 0 +
              model_h.params["bachelor"] * 1 +
              model_h.params["female_bachelor"] * 0)
print(f"{bob_ln_ahe=}")

jim_ln_ahe = (model_h.params["Intercept"] +
              model_h.params["age"] * 30 +
              model_h.params["age_squared"] * 30 ** 2 +
              model_h.params["female"] * 0 +
              model_h.params["bachelor"] * 0 +
              model_h.params["female_bachelor"] * 0)
print(f"{jim_ln_ahe=}")

bob_jim_diff = np.exp(bob_ln_ahe) - np.exp(jim_ln_ahe)
print(f"{bob_jim_diff=}")
```

```
model_h.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
==
Dep. Variable:                 ln_ahe   R-squared:                       0.2
09
Model:                            OLS   Adj. R-squared:                  0.2
09
Method:                 Least Squares   F-statistic:                     37
5.1
Date:                Sat, 16 Aug 2025   Prob (F-statistic):               0.
00
Time:                        06:39:56   Log-Likelihood:                -481
8.5
No. Observations:                7098   AIC:                             964
9.
Df Residuals:                    7092   BIC:                             969
0.
Df Model:                           5
Covariance Type:            nonrobust
==============================================================================
=======
                      coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
-------
Intercept           0.4119      0.672      0.613      0.540      -0.906
1.729
age                 0.1348      0.046      2.944      0.003       0.045
0.225
age_squared        -0.0019      0.001     -2.416      0.016      -0.003
-0.000
female             -0.1903      0.017    -10.955      0.000      -0.224
-0.156
bachelor            0.4521      0.015     30.379      0.000       0.423
0.481
female_bachelor     0.0235      0.023      1.004      0.315      -0.022
0.069
==============================================================================
==
Omnibus:                      181.391   Durbin-Watson:                   1.9
44
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              300.5
74
Skew:                          -0.234   Prob(JB):                      5.39e-
66
Kurtosis:                       3.893   Cond. No.                      1.07e+
05
==============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
ctly specified.
[2] The condition number is large, 1.07e+05. This might indicate that there
```

```
are
strong multicollinearity or other numerical problems.
"""
alexis_ln_ahe=3.057717242724363
jane_ln_ahe=2.5821294247778046
alexis_jane_diff=8.053656637447899
bob_ln_ahe=3.2245672291181027
jim_ln_ahe=2.7724535719982715
bob_jim_diff=9.144853034268806
```

The coefficient on the interaction term measures the additional effect of having a bachelor's degree for women beyond just the already included effects of separately being a woman and having a bachelor's on earnings, that is, any difference in earnings of having a bachelor's degree between men and women.

We see from the calculations above the following:

- Alexis' predicted ln(AHE) is **3.057717242724363**
- Jane's predicted ln(AHE) is **2.5821294247778046**
- The difference between Alexis's and Jane's earnings is **$8.053656637447899 per hour**
- Bob's predicted ln(AHE) is **3.2245672291181027**
- Jim's predicted ln(AHE) is **2.7724535719982715**
- The difference between Bob's and Jim's earnings is **$9.144853034268806 per hour**

i. Is the effect of age on earnings different for men than for women? Specify and estimate a regression that you can use to answer this question.

In [46]:
```
# Add interaction terms both linearly and squared
cps_df["female_age"] = cps_df["female"] * cps_df["age"]
cps_df["female_age_squared"] = cps_df["female"] * cps_df["age_squared"]

model_i = ols("ln_ahe ~ age + age_squared + female + bachelor + female_age +
print(f"{model_i.summary()=}")

print(f"{model_i.params['female_age']=}")
print(f"{model_i.pvalues['female_age']=}")

print(f"{model_i.params['female_age_squared']=}")
print(f"{model_i.pvalues['female_age_squared']=}")
```

```
model_i.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
==============================================================================
==
Dep. Variable:                 ln_ahe   R-squared:                       0.2
10
Model:                            OLS   Adj. R-squared:                  0.2
09
Method:                 Least Squares   F-statistic:                     31
3.4
Date:                Sat, 16 Aug 2025   Prob (F-statistic):               0.
00
Time:                        06:53:55   Log-Likelihood:                 -481
6.4
No. Observations:                7098   AIC:                             964
7.
Df Residuals:                    7091   BIC:                             969
5.
Df Model:                           6
Covariance Type:            nonrobust
==============================================================================
==========
                         coef    std err          t      P>|t|      [0.025
0.975]
------------------------------------------------------------------------------
----------
Intercept              0.2920      0.882      0.331      0.741      -1.437
2.021
age                    0.1388      0.060      2.312      0.021       0.021
0.256
age_squared           -0.0019      0.001     -1.849      0.064      -0.004
0.000
female                -0.0098      1.363     -0.007      0.994      -2.681
2.661
bachelor               0.4607      0.011     40.142      0.000       0.438
0.483
female_age            -0.0020      0.093     -0.022      0.983      -0.184
0.180
female_age_squared    -0.0001      0.002     -0.077      0.939      -0.003
0.003
==============================================================================
==
Omnibus:                      183.160   Durbin-Watson:                   1.9
45
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              305.9
93
Skew:                          -0.233   Prob(JB):                     3.58e-
67
Kurtosis:                       3.904   Cond. No.                     2.65e+
05
==============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
```

```
ctly specified.
[2] The condition number is large, 2.65e+05. This might indicate that there
are
strong multicollinearity or other numerical problems.
"""
model_i.params['female_age']=-0.0020375336184010606
model_i.pvalues['female_age']=0.9825028123815729
model_i.params['female_age_squared']=-0.00012110805152292913
model_i.pvalues['female_age_squared']=0.9385754060827849
```

After adding in the interaction terms for for female * age, we find both coefficients to be rather low and with very high p-values that are not statistically significant (>0.05), so we can conclude that the effect of age on earnings is likely not different between men and women.

j. Is the effect of age on earnings different for high school graduates than for college graduates? Specify and estimate a regression that you can use to answer this question.

In [48]:

```python
# Add interaction terms both linearly and squared
cps_df["bachelor_age"] = cps_df["bachelor"] * cps_df["age"]
cps_df["bachelor_age_squared"] = cps_df["bachelor"] * cps_df["age_squared"]

model_j = ols("ln_ahe ~ age + age_squared + female + bachelor + bachelor_age
print(f"{model_j.summary()=}")

print(f"{model_j.params['bachelor_age']=}")
print(f"{model_j.pvalues['bachelor_age']=}")

print(f"{model_j.params['bachelor_age_squared']=}")
print(f"{model_j.pvalues['bachelor_age_squared']=}")
```

```
model_j.summary()=<class 'statsmodels.iolib.summary.Summary'>
"""
                            OLS Regression Results
=============================================================================
==
Dep. Variable:                 ln_ahe   R-squared:                       0.2
09
Model:                            OLS   Adj. R-squared:                  0.2
09
Method:                 Least Squares   F-statistic:                     31
2.7
Date:                Sat, 16 Aug 2025   Prob (F-statistic):               0.
00
Time:                        07:02:04   Log-Likelihood:                -481
8.0
No. Observations:                7098   AIC:                             965
0.
Df Residuals:                    7091   BIC:                             969
8.
Df Model:                           6
Covariance Type:            nonrobust
=============================================================================
============
                          coef    std err          t      P>|t|      [0.025
0.975]
-----------------------------------------------------------------------------
------------
Intercept               0.0785      0.966      0.081      0.935      -1.816
1.973
age                     0.1601      0.066      2.432      0.015       0.031
0.289
age_squared            -0.0023      0.001     -2.107      0.035      -0.005
-0.000
female                 -0.1769      0.012    -15.208      0.000      -0.200
-0.154
bachelor                1.1166      1.345      0.830      0.406      -1.520
3.753
bachelor_age           -0.0500      0.092     -0.546      0.585      -0.230
0.130
bachelor_age_squared    0.0009      0.002      0.602      0.547      -0.002
0.004
=============================================================================
==
Omnibus:                      182.523   Durbin-Watson:                   1.9
43
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              303.4
89
Skew:                          -0.234   Prob(JB):                     1.25e-
66
Kurtosis:                       3.898   Cond. No.                     2.87e+
05
=============================================================================
==

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is corre
```

```
ctly specified.
[2] The condition number is large, 2.87e+05. This might indicate that there
are
strong multicollinearity or other numerical problems.
"""
model_j.params['bachelor_age']=-0.04999488380442396
model_j.pvalues['bachelor_age']=0.5853787577766819
model_j.params['bachelor_age_squared']=0.0009324474823376502
model_j.pvalues['bachelor_age_squared']=0.5473168952144802
```

After adding in the interaction terms for for bachelor * age, we find both coefficients to be low and with very high p-values that are not statistically significant (>0.05), although with both of these values holding a bit more value and impact than when compared to the female * age terms. Nonetheless, we can conclude that the effect of age on earnings is likely not different between high school and college graduates.

k. After running all these regressions, summarize the effect of age on earnings for young workers.

For younger workers, the effect of increased age on earnings is generally positive, but the marginal effect of increased age on earnings decreases as age continues to increase. This means that while older workers do indeed tend to earn more than younger workers, the rate at which their earnings increase slows down with age. In addition, there seem to be little differences in this effect of age on earnings when comparing across gender or education differences. Overall, younger workers get more value from increased age in terms of earnings compared to older workers.

Extra Credit (5 points):

In a few sentences, describe something you learned/discovered through this assignment.

I wasn't aware of the marginal effect of age on earnings decreasing over time. I think that I had some thoughts around ageism in certain industries and how this would affect overall employment, which in turn could possibly mean that there may be a tendency for older people to be better paid and therefore for earnings to accelerate more steeply on average for higher ages. The regressions seem to generally reflect that this is not the case. In addition, the effect of gender seemed to generally be rather small on earnings, which I also found to be interesting and could warrant further investigation into furhter breakdowns such as categorizations of jobs and career lengths.