Python Program Output (Terminal)

```
Question 1
                            OLS Regression Results
==============================================================================
Dep. Variable:                     ed   R-squared:                       0.007
Model:                            OLS   Adj. R-squared:                  0.007
Method:                 Least Squares   F-statistic:                     28.48
Date:                Tue, 07 Nov 2023   Prob (F-statistic):           1.00e-07
Time:                        21:19:27   Log-Likelihood:                -7632.2
No. Observations:                3796   AIC:                         1.527e+04
Df Residuals:                    3794   BIC:                         1.528e+04
Df Model:                           1
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept     13.9559      0.038    369.945      0.000      13.882      14.030
dist          -0.0734      0.014     -5.336      0.000      -0.100      -0.046
==============================================================================
Omnibus:                     7187.794   Durbin-Watson:                   1.769
Prob(Omnibus):                  0.000   Jarque-Bera (JB):              361.676
Skew:                           0.410   Prob(JB):                     2.90e-79
Kurtosis:                       1.729   Cond. No.                         3.73
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.
Question 1(a)

Intercept:  13.9559

Slope: -0.0734

Question 1(b)

If Bob's HS was 20 miles away from the nearest college, then its predicted that he completed 13.8091 years of education

If Bob lived 10 miles from the nearest college, the prediction changes to 13.8825 years

Question 1(e)

A 95% confidence interval for the slope coefficient is: (-0.1003, -0.0464)

Question 1(f)

If distance to the nearest college is decreased by 20 miles, educational attainment is  14.1026. This is 0.1467 greater than the intercept which is approximately equal to 0.15.
```

```
Question 2
                            OLS Regression Results
==============================================================================
Dep. Variable:                     ed   R-squared:                       0.279
Model:                            OLS   Adj. R-squared:                  0.277
Method:                 Least Squares   F-statistic:                     146.3
Date:                Tue, 07 Nov 2023   Prob (F-statistic):          6.94e-260
Time:                        21:19:27   Log-Likelihood:                -7025.9
No. Observations:                3796   AIC:                         1.407e+04
Df Residuals:                    3785   BIC:                         1.414e+04
Df Model:                          10
Covariance Type:            nonrobust
==============================================================================
                 coef    std err          t      P>|t|      [0.025      0.975]
------------------------------------------------------------------------------
Intercept      8.8275      0.250     35.271      0.000       8.337       9.318
dist          -0.0315      0.012     -2.550      0.011      -0.056      -0.007
bytest         0.0938      0.003     29.669      0.000       0.088       0.100
female         0.1454      0.051      2.874      0.004       0.046       0.245
black          0.3680      0.071      5.156      0.000       0.228       0.508
hispanic       0.3985      0.074      5.352      0.000       0.253       0.545
incomehi       0.3952      0.061      6.529      0.000       0.277       0.514
ownhome        0.1521      0.067      2.277      0.023       0.021       0.283
dadcoll        0.6961      0.069     10.129      0.000       0.561       0.831
cue80          0.0232      0.010      2.409      0.016       0.004       0.042
stwmfg80      -0.0518      0.020     -2.608      0.009      -0.091      -0.013
==============================================================================
Omnibus:                      118.266   Durbin-Watson:                   1.924
Prob(Omnibus):                  0.000   Jarque-Bera (JB):               97.867
Skew:                           0.320   Prob(JB):                     5.60e-22
Kurtosis:                       2.543   Cond. No.                         539.
==============================================================================

Notes:
[1] Standard Errors assume that the covariance matrix of the errors is correctly specified.

The effect of Dist on ED is -0.0315.
This means that for every 10 mile increase in dist, educational attainment decreases by approx 0.0315.

Question 2(a)

The estimated effect in the regression in Q1 is -0.0734 while the estimated effect in Q2 is -0.0315

Question 2(c)

Cue80: 0.0232
Stwmfg80:-0.0518

Question 2(d)

Bob's years of schooling are  14.6182

Question 2(e)

Jim's years of schooling are  14.5551

Question 2(f)

Blacks complete 9.1955 years of education

Hispanics complete 9.2260 years of education

Whites complete  8.8275 years of education
```

Python Program Output (Text)

Question 1

OLS Regression Results

Dep. Variable:	ed	R-squared:	0.007
Model:	OLS	Adj. R-squared:	0.007
Method:	Least Squares	F-statistic:	28.48
Date:	Tue, 07 Nov 2023	Prob (F-statistic):	1.00e-07
Time:	16:43:46	Log-Likelihood:	-7632.2
No. Observations:	3796	AIC:	1.527e+04
Df Residuals:	3794	BIC:	1.528e+04
Df Model:	1		
Covariance Type:	nonrobust		

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	13.9559	0.038	369.945	0.000	13.882	14.030
dist	-0.0734	0.014	-5.336	0.000	-0.100	-0.046

Omnibus:	7187.794	Durbin-Watson:	1.769
Prob(Omnibus):	0.000	Jarque-Bera (JB):	361.676
Skew:	0.410	Prob(JB):	2.90e-79
Kurtosis:	1.729	Cond. No.	3.73

Question 1(a)

- Estimated Intercept: 13.9559
- Estimated Slope: -0.0734
- When colleges are built close to where students go to high school, the average value of years of completed schooling tends to decrease. This indicates **a negative relationship** between the distance to colleges and the average years of completed schooling.

Question 1(b)

- If Bob's high school was 20 miles from the nearest college, using the estimated regression, it's predicted that **Bob completed 13.8091 years of education**.
- If Bob lived 10 miles from the nearest college, **the prediction changes to 13.8825 years.**

Question 1(c)

- No. The R-squared value is about 0.007. This means that **only about 0.7% of the variance in educational attainment is explained by the distance to the nearest college.**

Question 1(d)

- Since the **p-value associated with the coefficient's t-statistic is 0**, it can be concluded that the **estimated regression slope coefficient for "dist" is statistically significant at any significance level**. This is because the p-value is less than the significance levels of 10%, 5%, or 1%.
    - **Can reject the null hypothesis (H0: β1 = 0) in favor of the alternative hypothesis**

Question 1(e)

- The 95% confidence interval for the slope coefficient is: **(-0.1003, -0.0464)**

Question 1(f)

- **The advocacy groups' claim is consistent with the estimated regression.**
    - If distance is decreased by 20 miles (dist = -2), educational attainment is 14.1026.
        - The difference between this estimate and the intercept is approximately 0.15 (exact difference is 0.1467)

Question 2

OLS Regression Results

==========

Dep. Variable:	ed	R-squared:	0.279
Model:	OLS	Adj. R-squared:	0.277
Method:	Least Squares	F-statistic:	146.3
Date:	Tue, 07 Nov 2023	Prob (F-statistic):	6.94e-260
Time:	19:16:29	Log-Likelihood:	-7025.9
No. Observations:	3796	AIC:	1.407e+04
Df Residuals:	3785	BIC:	1.414e+04
Df Model:	10		

Covariance Type: nonrobust

	coef	std err	t	P>|t|	[0.025	0.975]
Intercept	8.8275	0.250	35.271	0.000	8.337	9.318
dist	-0.0315	0.012	-2.550	0.011	-0.056	-0.007
bytest	0.0938	0.003	29.669	0.000	0.088	0.100
female	0.1454	0.051	2.874	0.004	0.046	0.245
black	0.3680	0.071	5.156	0.000	0.228	0.508
hispanic	0.3985	0.074	5.352	0.000	0.253	0.545
incomehi	0.3952	0.061	6.529	0.000	0.277	0.514
ownhome	0.1521	0.067	2.277	0.023	0.021	0.283
dadcoll	0.6961	0.069	10.129	0.000	0.561	0.831
cue80	0.0232	0.010	2.409	0.016	0.004	0.042
stwmfg80	-0.0518	0.020	-2.608	0.009	-0.091	-0.013

Omnibus:	118.266	Durbin-Watson:	1.924
Prob(Omnibus):	0.000	Jarque-Bera (JB):	97.867
Skew:	0.320	Prob(JB):	5.60e-22
Kurtosis:	2.543	Cond. No.	539.

- The **effect of Dist on ED is approx -0.0315**. This means that for every 10 mile increase in distance, educational attainment decreases by approx 0.0315.

Question 2(a)

- The estimated effect in the regression in Q1 is -0.0734 while the estimated effect in Q2 is -0.0315
    - The estimated effect in Q2 is significantly lower than the effect in Q1.
        - Effect is less negative in Q2 than in Q1
        - This indicates that controlling for additional variables in the regression in Q2 significantly changed the estimated impact of Dist on ED.
            - **This difference suggests that the regression in Q1 seems to suffer from omitted variable bias.**

Question 2(b)

- **The coefficient measures how the educational attainment of students is influenced by whether or not their fathers have attended college.**

- The positive coefficient on "DadColl" means that students whose fathers have attended college tend to have higher levels of educational attainment compared to students whose fathers haven't attended.

Question 2(c)

- Cue80
  - The positive coefficient suggests that an increase in the County Unemployment rate in 1980 is associated with an increase in educational attainment.
  - **A one-unit increase in Cue80 is associated with an estimated increase of 0.0232 in educational attainment, all other variables being held constant.**
    - Suggests that the higher level of County Unemployment in 1980 may be associated with a slightly higher level of educational attainment.
- Stwmfg80
  - The negative coefficient suggests that an increase in the State Hourly Wage in Manufacturing in 1980 is associated with a decrease in educational attainment.
  - **A one-unit increase in Stwmfg80 is associated with an estimated decrease of 0.0518 in educational attainment.**
    - Implies that a higher State Hourly Wage in 1980 may be associated with lower levels of educational attainment.
      - Magnitude of Stwmfg80 is greater than Cue80 so the negative effect is larger.
- The signs of these coefficients make sense. **When unemployment is high, more people attend school and when Hourly Wage is high, more people work and thus don't attend school.**

Question 2(d)

- Bob's years of schooling are 14.6182

Question 2(e)

- Jim's years of schooling are 14.5551

Question 2(f)

- This result is consistent with the regressions in Q2.
  - **Controlling for other factors,**
    - **Blacks complete 9.1955 years of education**
    - **Hispanics complete 9.2260 years of education**
    - **Whites complete 8.8275 years of education (intercept).**