## **Python Program Output (Terminal)**

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

Dep. Variable: ed R =quared: 0.279 Model: 0.5 Adj. Requared: 0.277 Method: Least Squares F = statistic: 1.46.3 Date: Tue, 07 Nov 2023 Prob (F-statistic): 6.446-26 Date: 1.09-Likelihood: -7025.9 Date: 1.09-Likelihood:	Dep. Variable: ed R-squared: 0.279 Model: OLS Adj. R-squared: 0.277 Method: Least Squares F-statistic: 146.3	
Dep. Variable: ed R =quared: 0.279 Model: 0.5 Adj. Requared: 0.277 Method: Least Squares F = statistic: 1.46.3 Date: Tue, 07 Nov 2023 Prob (F-statistic): 6.446-26 Date: 1.09-Likelihood: -7025.9 Date: 1.09-Likelihood:	Dep. Variable: ed R-squared: 8.279 Model: OLS Adj. R-squared: 0.277 Method: Least Squares F-statistic: 146.3	
Least Squares   F-statistic    1.6.3	Method: Least Squares F-statistic: 146.3	
Date: Tue, 87 Nov 2023 Prob (F-statistic): 6.946-268 [rime: 211:1972 Log-likelihood: -7025.9   1.4070+04   1.4070+		
Time: 21:19:27 Log-Likelihood: -725.9 No. Observations: 3796 AIC: 1.407e+04 Df Residuals: 3795 BIC: 1.414e+04 Df Model: 10  Covariance Type: nonrobust		
No. Observations:   3796   AIČ:   1.497e+04   1.414e+04   1.414e		
10		
Coverage		
Coce   Std err   t   P t   [0.025   0.975		
Intercept 8.8275 0.250 35.271 0.000 8.337 9.318 list -0.0315 0.012 -2.550 0.011 -0.056 -0.007 lytest 0.0938 0.003 20.669 0.000 0.088 0.100 lytest 0.0938 0.003 20.669 0.000 0.088 0.100 lemale 0.1454 0.051 2.874 0.004 0.046 0.245 lack 0.3680 0.071 5.156 0.000 0.228 0.508 lispanic 0.3985 0.074 5.352 0.000 0.223 0.565 lincomehi 0.3995 0.061 6.529 0.000 0.277 0.514 locemehi 0.3995 0.061 6.529 0.000 0.277 0.544 locemehi 0.3995 0.061 6.529 0.000 0.277 0.544 locemehi 0.3995 0.061 6.529 0.000 0.277 0.514 locemehi 0.3995 0.061 0.000 1.000 0.001 0.001 locemehi 0.000 1.000 1.000 0.000 0.001 0.001 locemehi 0.000 0.000 0.000 0.000 0.000 0.001 locemehi 0.000 0.		
	coef std err t P> t  [0.025 0.975]	
Part		
Semale		
naispanic 0.3985 0.871 5.156 0.000 0.228 0.508 ninspanic 0.3985 0.874 5.352 0.000 0.253 0.545 nincomehi 0.3952 0.061 6.529 0.000 0.277 0.514 nincomehi 0.3952 0.061 6.529 0.000 0.277 0.514 nincomehi 0.3952 0.061 0.692 0.000 0.277 0.514 nincomehi 0.6961 0.000 0.067 2.277 0.023 0.001 0.283 nincomehi 0.6961 0.000 0.0651 0.831 nincomehi 0.0001 0.000 0.000 0.0001 0.0001 0.0001 nincomehi 0.0001 0.0		
Ancomehi		
Description 2 (a)  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. Duestion 2(a)  The estimated effect in the regression in Q1 is -0.0734 while the estimated effect in Q2 is -0.0315 duestion 2(b)  Duestion 2(c)  Du		
Add coll 0.0961 0.096 10.129 0.000 0.561 0.831		
tree80 0.0232 0.010 2.409 0.016 0.004 0.042 stumfg80 -0.0518 0.020 -2.608 0.009 -0.091 -0.013 stumfg80 -0.0518 0.020 -2.608 0.009 -0.091 -0.013 stumfg80 -0.0518 0.020 -2.608 0.009 -0.091 -0.013 stumfg80 -0.0518 0.009 Jarque-Bera (JB): 97.867 skew: 0.000 Jarque-Bera (JB): 97.867 skew: 0.000 Prob(JB): 5.60e-22 kurtosis: 2.543 Cond. No. 539.  Notes: 2.543 Cond. No.		
stwings8 -0.0518 0.020 -2.608 0.009 -0.091 -0.013		
Dimnibus: 118.266 Durbin-Watson: 1.924 Prob(Omnibus): 0.000 Jarque-Bera (JB): 97.867 Rickew: 0.320 Prob(JB): 5.60e-22 Rurtosis: 2.543 Cond. No. 539.  Rotes: Rickew: 1.543 Cond. No. 539.		
Problem (Donibus):  0. 4000 Jarque-Bera (JB):  0. 320 Prob(JB):  5.60e-22 State    (Notes:  2. 543 Cond. No.  539.  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)  Question 2(c)  Question 2(d)  Gob's years of schooling are 14.6182  Question 2(e)  Question 2(f)  Glacks complete 9.1955 years of education  dispanics complete 9.2260 years of education		
cokew: 0.320 Prob(JB): 5.60e-22 (Autrosis: 2.543 Cond. No. 539.  Cotes:  (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.  (Autestion 2(a))  (The estimated effect in the regression in Q1 is -0.0734 while the estimated effect in Q2 is -0.0315  (Autestion 2(c))  (Autestion 2(c))  (Autestion 2(d))  (Autestion 2(d))  (Autestion 2(d))  (Autestion 2(d))  (Autestion 2(e))  (Autestion 2(e		
Actrosis:  2.543 Cond. No.  539.  Cond. No.		
Notes:  (1) Standard Errors assume that the covariance matrix of the errors is correctly specified.  (The effect of Dist on ED is -0.0315.)  (The effect of Dist on ED is -0.0315.)  (The estimated 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))  (Question 2(c))  (Question 2(d))  (Question 2(d))  (Question 2(d))  (Question 2(e))  (Question 2(e))  (Question 2(f))  (Question		
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.  Appearance of Dist on ED is -0.0315.  This means that for every 10 mile increase in dist, educational attainment decreases by approx 0.0315.  Appearance of Dist on ED is -0.0315.  Appearance of		
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)  Que80: 0.0232  Stwmfg80:-0.0518  Question 2(d)  Sob's years of schooling are 14.6182  Question 2(e)  Dim's years of schooling are 14.5551  Question 2(f)  Slacks complete 9.1955 years of education  Hispanics complete 9.2260 years of education	Notes: [1] Standard Errors assume that the covariance matrix of the errors is correctly specified. The effect of Dist on ED is -0.0315.	
The estimated effect in the regression in Q1 is -0.0734 while the estimated effect in Q2 is -0.0315  Question 2(c)  Que80: 0.0232  Stwmfg80:-0.0518  Question 2(d)  Sob's years of schooling are 14.6182  Question 2(e)  Dim's years of schooling are 14.5551  Question 2(f)  Slacks complete 9.1955 years of education  Hispanics complete 9.2260 years of education	This means that for every 10 mile increase in dist, educational attainment decreases by appro	x 0.0315.
Question 2(c) Cue80: 0.0232 Stwmfg80:-0.0518 Question 2(d) Sob's years of schooling are 14.6182 Question 2(e) Dim's years of schooling are 14.5551 Question 2(f) Slacks complete 9.1955 years of education Hispanics complete 9.2260 years of education	Question 2(a)	
Cue80: 0.0232 Stwmfg80:-0.0518  Question 2(d)  Sob's years of schooling are 14.6182  Question 2(e)  Dim's years of schooling are 14.5551  Question 2(f)  Slacks complete 9.1955 years of education  Hispanics complete 9.2260 years of education	The estimated effect in the regression in Q1 is $-0.0734$ while the estimated effect in Q2 is $-0.0734$	-0.0315
Stwmfg80:-0.0518 Question 2(d) Gob's years of schooling are 14.6182 Question 2(e) Dim's years of schooling are 14.5551 Question 2(f) Glacks complete 9.1955 years of education Hispanics complete 9.2260 years of education	Question 2(c)	
Obb's years of schooling are 14.6182 Question 2(e) Dim's years of schooling are 14.5551 Question 2(f) Blacks complete 9.1955 years of education Hispanics complete 9.2260 years of education	Cue80: 0.0232 Stwmfg80:-0.0518	
Question 2(e) Dim's years of schooling are 14.5551 Question 2(f) Blacks complete 9.1955 years of education Hispanics complete 9.2260 years of education	Question 2(d)	
Dim's years of schooling are 14.5551 Question 2(f) Blacks complete 9.1955 years of education Hispanics complete 9.2260 years of education	Bob's years of schooling are 14.6182	
Question 2(f) Blacks complete 9.1955 years of education Hispanics complete 9.2260 years of education	Question 2(e)	
complete 9.1955 years of education  Hispanics complete 9.2260 years of education	lim's years of schooling are 14.5551	
Hispanics complete 9.2260 years of education	Question 2(f)	
	Blacks complete 9.1955 years of education	
hites complete 8.8275 years of education	dispanics 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:

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: 13.9559Estimated 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.

### **Ouestion 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:  $\beta 1 = 0$ ) 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.
  - o 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**

Date:

**OLS Regression Results** 

Dep. Variable: ed R-squared: 0.279 Model: OLS Adj. R-squared: 0.277 Method: Least Squares F-statistic: 146.3 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 3785 BIC: Df Residuals: 1.414e+04

Df Model: 10

Covarian	Covariance Type: nonrobust						
	====				======		
	coef	std e	err	t P> t	[0.025	0.975]	
Intercept	8.827	 75	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	8	0.003	29.669	0.000	0.088	0.100
female	0.145	54	0.051	2.874	0.004	0.046	0.245
black	0.3680	0	0.071	5.156	0.000	0.228	0.508
hispanic	0.398	35	0.074	5.352	0.000	0.253	0.545
incomehi	0.39	952	0.061	6.529	0.000	0.277	0.514
ownhome	e 0.1	521	0.06	7 2.27	7 0.023	0.021	0.283
dadcoll	0.696	51	0.069	10.129	0.000	0.561	0.831
cue80	0.023	2	0.010	2.409	0.016	0.004	0.042
stwmfg8	0.0-	518	0.02	0 -2.60	8 0.009	-0.091	-0.013
======		====					
======	===						
Omnibus			118.26	66 Durbi	n-Watson:		1.924
Prob(Omnibus):				0.000 Jarque-Bera (JB):			
Skew:	kew:			0.320 Prob(JB): 5.60e-2			-22
Kurtosis:			2.543	Cond. N	lo.	53	9.
======							

• 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.

### **Ouestion 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.

#### Ouestion 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.
  - o 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).