Classwork 3: OLS Numerical Example

Suppose we sampled random variables X and Y to get this data:

X	Y
1	4
2	2
3	1

1. Draw a plot of x_i and y_i and eyeball a line of best fit. What is the slope and intercept of that line (approximately)? (1 point)

2. Use the formulas we derived in Chapter 3 to find the slope and intercept of the OLS line of best fit: $y_i = \beta_0 + \beta_1 x_i + u_i$. Here are the formulas you can use: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$ and $\hat{\beta}_1 = \frac{\sum_{i=1}^n (x_i y_i) - \bar{x}\bar{y}n}{\sum_{i=1}^n (x_i^2) - \bar{x}^2 n}$. Verify that your estimates for β_0 and β_1 are on track by comparing these answers to your answers to question 1. (1 point)

3. Fill out the table below to calculate the OLS fitted values ($\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i$) and residuals ($e_i = y_i - \hat{y}_i$). Verify that the fitted values you calculate are correct by making sure that (x_i, $\hat{y}_i$) lays on the line of best fit you drew in question 1. (2 points)

x	y	$\hat{y}_i$	e_i
1	4	_____	_____
2	2	_____	_____
3	1	_____	_____

4. Two things are always true about OLS residuals e_i: First, $\sum_{i=1}^n e_i = 0$ and second, $\sum_{i=1}^n x_i e_i = 0$. Verify both these things are true in this numerical example. (2 points)

5. Write a proof that shows since $\sum_{i=1}^n e_i = 0$ and $\sum_{i=1}^n x_i e_i = 0$, the covariance of x and e will always be 0. Recall the formula for the estimate of the covariance between two random variables given sample data: $\frac{1}{n-1} \sum_i (x_i - \bar{x})(e_i - \bar{e})$. (2 points)

6. Going back to the numerical example, suppose that the true relationship between X and Y is actually this: $y_i = 5 - x_i + u_i$. So the true values for $\beta_0 = 5$ and $\beta_1 = -1$. Calculate the true disturbance u_i. Is it required that $\sum_{i=1}^n u_i = 0$ or $\sum_{i=1}^n x_i u_i = 0$ like it is with e_i? (2 points)

x	y	u_i
1	4	_____
2	2	_____
3	1	_____