1. We derived the OLS estimator and found that:

$$\hat{\beta}_1 = \frac{\sum_i (x_i y_i) - n\bar{x}\bar{y}}{\sum_i (x_i^2) - \bar{x}^2 n}$$

(a) Show that $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2}$. Trick is to start from here & show that it is equavilent to the above

$$= \frac{\sum_i (x_i y_i - x_i \bar{y} - \bar{x} y_i + \bar{x}\bar{y})}{\sum_i (x_i^2 - \bar{x} x_i - \bar{x} x_i + \bar{x}^2)}$$

Numerator:

$$\sum_i x_i y_i - \bar{y} \underbrace{\sum x_i}_{n\bar{x}} - \bar{x} \underbrace{\sum_i y_i}_{n\bar{y}} + n\bar{x}\bar{y}$$

$$= \sum_i x_i y_i - n\bar{x}\bar{y} - n\bar{x}\bar{y} + n\bar{x}\bar{y}$$

$$= \sum_i x_i y_i - n\bar{x}\bar{y}$$

Denominator:

$$\sum_i (x_i^2 - \bar{x} x_i - \bar{x} x_i + \bar{x}^2)$$

$$\sum_i x_i^2 - \bar{x} \underbrace{\sum_i x_i} - \bar{x} \underbrace{\sum_i x_i} + n\bar{x}^2$$

$$\sum_i x_i^2 - \bar{x} n \bar{x} - \bar{x} n \bar{x} + n\bar{x}^2$$

$$\sum_i x_i^2 - n\bar{x}^2 - \cancel{n\bar{x}^2} + \cancel{n\bar{x}^2}$$

(b) Use the formula for $\hat{\beta}_1$ derived in 1) to show that $\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2}$.

Want to show: $\frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2}$

Because denoms. are equal, we just need to show that the numerators are too

↳ $\underbrace{\sum_i (x_i - \bar{x})(y_i - \bar{y})}_{\text{"LHS"}} = \underbrace{\sum_i (x_i - \bar{x}) y_i}_{\text{"RHS"}}$

$$\text{LHS} = \sum_i x_i y_i - \bar{y} \sum_i x_i - \bar{x} \sum_i y_i + \bar{x}\bar{y}n$$

$$= \sum_i x_i y_i - \bar{x}\underbrace{\bar{y}n}_{\sum_i y_i} - \cancel{\bar{x}\bar{y}n} + \cancel{\bar{x}\bar{y}n}$$

$$= \sum_i x_i y_i - \bar{x} \sum_i y_i$$

$$= \sum_i (x_i y_i - \bar{x} y_i)$$

$$= \sum_i (x_i - \bar{x}) y_i = \text{"RHS"}$$

(c) Use the formula for $\hat{\beta}_1$ in 2) to show that $\hat{\beta}_1 = \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$.

Hint: Note that the left hand-side is the estimate for β_1 and the right hand-side includes the true value of β_1. These will not be exactly equivalent except by chance. You should start this problem by making a substitution for y_i, since $y_i = \beta_0 + \beta_1 x_i + u_i$. This will get the true β_1 and u_i into the equation.

Want to show: $\frac{\sum_i (x_i - \bar{x}) y_i}{\sum_i (x_i - \bar{x})^2} = \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2}$

Use $y_i = \beta_0 + \beta_1 x_i + u_i$

$$= \frac{\sum_i (x_i - \bar{x})(\beta_0 + \beta_1 x_i + u_i)}{\sum_i (x_i - \bar{x})^2}$$

$$= \frac{\sum_i (x_i \beta_0 + x_i^2 \beta_1 + x_i u_i - \bar{x}\beta_0 - \bar{x}\beta_1 x_i - \bar{x} u_i)}{\sum_i (x_i - \bar{x})^2}$$

$$= \frac{\cancel{\beta_0 \underbrace{\sum_i x_i}_{n\bar{x}}} + \beta_1 \sum_i x_i^2 + \sum_i x_i u_i - \cancel{\underbrace{\sum_i \bar{x}}_{\bar{x}n} \beta_0} - \bar{x}\beta_1 \sum_i x_i - \bar{x} \sum u_i}{\sum_i (x_i - \bar{x})^2}$$

Separate into factors

$$\rightarrow \frac{\beta_1 (\sum_i x_i^2 - \bar{x} \sum_i x_i)}{\sum_i (x_i - \bar{x})^2} + \frac{\sum_i x_i u_i - \bar{x} \sum_i u_i}{\sum_i (x_i - \bar{x})^2}$$

↳ Note: $\sum_i (x_i - \bar{x})^2$ in denom.

$= \sum_i (x_i^2 + \bar{x}^2 - 2x_i\bar{x})$

$= \sum_i x_i^2 + n\bar{x}^2 - 2\bar{x}\sum_i x_i$

$= \sum_i x_i^2 + n\bar{x}^2 - 2n\bar{x}^2$

$= \sum_i x_i^2 - n\bar{x}^2 = \sum_i x_i^2 - \bar{x}\sum_i x_i$ Denominator

$$= \beta_1 + \frac{\sum_i x_i u_i - \bar{x} \sum_i u_i}{\sum_i (x_i - \bar{x})^2}$$

$$= \beta_1 + \frac{\sum_i (x_i u_i - \bar{x} u_i)}{\sum_i (x_i - \bar{x})^2}$$

$$= \beta_1 + \frac{\sum_i (x_i - \bar{x}) u_i}{\sum_i (x_i - \bar{x})^2} \quad //$$

2. **Recall:** $\hat{\beta}_1 = \dfrac{\sum_i (x_i y_i) - n\bar{x}\bar{y}}{\sum_i (x_i^2) - \bar{x}^2 n}$ and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$

   Use those formulas to calculate $\hat{\beta}_1$ and $\hat{\beta}_0$ by hand. It may be helpful to draw a plot of the data and try to eyeball the line of best fit in order to double check that your answer makes sense.

x	y
0	1
1	2
1	3
0	2

$\bar{y} = \frac{8}{4} = 2 \quad \bar{x} = \frac{1}{2}$

$\hat{\beta}_0 = 2 - \hat{\beta}_1 \left(\frac{1}{2}\right)$

Need to find $\hat{\beta}_1$

$$\hat{\beta}_1 = \frac{(x_1-\bar{x})(y_1-\bar{y}) + (x_2-\bar{x})(y_2-\bar{y}) + (x_3-\bar{x})(y_3-\bar{y}) + (x_4-\bar{x})(y_4-\bar{y})}{(x_1-\bar{x})^2 + (x_2-\bar{x})^2 + (x_3-\bar{x})^2 + (x_4-\bar{x})^2}$$

$$= \frac{(0-0.5)(1-2) + (1-0.5)(2-2) + (1-0.5)(3-2) + (0-0.5)(2-2)}{(0-0.5)^2 + (1-0.5)^2 + (1-0.5)^2 + (0-0.5)^2}$$

$$= \frac{0.5 + 0 + 0.5 + 0}{0.25 + 0.25 + 0.25 + 0.25} = \frac{1}{1} = \boxed{1 = \hat{\beta}_1}$$

$$\boxed{\hat{\beta}_0 = 2 - 1\left(\tfrac{1}{2}\right) = 1.5}$$

3. **Interpreting regression coefficients**

Consider a dataset obtained from a labor economics study that investigates the impact of years on education on individual's wages. The dataset includes a random sample of workers in a specific region. The following regression equation estimates the relationship between wages (measured in thousands of dollars) and years of education:

$$\text{Wage}_i = \beta_0 + \beta_1 \times \text{Education}_i + u_i$$

From the regression output, you have the following estimates:

$$\text{Wage} = 12 + 3.5 \times \text{Education}$$

(a) **Interpret the Estimates:** Interpret the intercept and slope coefficients in the context of the model

$\beta_0 = 12$ represents the predicted wage (in thousands of \$) for a person w/ zero years of education

$\beta_1 = 3.5$ suggests that each additional year of education is associated w/ an increase in wages of \$3,500.

(b) **Predicted Outcomes:** If an individual has 12 years of education, what is the predicted wage according to the model?

Plug in $X = 12 \longrightarrow$ The predicted wage is $10 + 3.5 \times 12 = 52$ thousand dollars

(c) **Effect of Changing X:** Suppose an individual worker is deciding whether or not to complete their associates degree (two years of education). What would the model predict her change in wage would be? In other words, what is her expected increase in wage if she completes her associates degree?

The predicted change in wage is $3.5 \times 2 = 7$ thousand dollars, reflecting the increase for 2 additional years of education

(d) What must we assume to be true regarding the error term u_i for the OLS estimator to be unbiased? Specifically, I am interested in the third assumption of the classical linear regression model. Give an example of a violation of this assumption in the context of the wage equation.

A3 states that the error term has a mean zero conditional on x_i.

This means that the expected value of the error term is zero given the independent variables.

An example in this context would be if the error term u_i had mean $\underline{1}$

$\hookrightarrow E[u_i | x_i] = \underline{1}$

4. We showed in lecture 03 that:

$$TSS = ESS + RSS + 2\sum_{i=1}^{n} \hat{y}_i \hat{u}_i - 2\bar{y}\sum_{i=1}^{n} \hat{u}_i$$

And I shut down the last two terms to make it so $TSS = ESS + RSS$.

(a) Show that $2\sum_{i=1}^{n} \hat{y}_i \hat{u}_i - 2\bar{y}\sum_{i=1}^{n} \hat{u}_i = 0$
Hint: The property of the residuals may be helpful: $\sum_{i=1}^{n} \hat{u}_i = 0$

$$2\sum_i \hat{y}_i \hat{u}_i - 2\bar{y}\underbrace{\sum_i \hat{u}_i}_{=0 \text{ by assumption}}$$

$$= 2\sum_i \hat{y}_i \hat{u}_i - 2\bar{y}\cdot 0$$

$$= 2\sum_i \hat{y}_i \hat{u}_i$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 \hat{x}_i$$

$$= 2\sum_i (\hat{\beta}_0 + \hat{\beta}_1 \hat{x}_i)\hat{u}_i$$

$$= 2\sum_i \hat{\beta}_0 \hat{u}_i + \sum_i \hat{\beta}_1 \hat{x}_i \hat{u}_i$$

$$= 2\hat{\beta}_0 \underbrace{\sum_i \hat{u}_i}_{=0} + \hat{\beta}_1 \underbrace{\sum_i \hat{x}_i \hat{u}_i}_{Cov(\hat{x}_i, \hat{u}_i) = 0}$$

$$= 2\hat{\beta}_0 \cdot 0 + \hat{\beta}_1 \cdot 0$$

$$= 0$$