The coefficients of an OLS regression of Y_i on a constant and X_i when X_i is binary.

Assume that you have a sample with n units. Let Y_i denote the value of the variable y for unit i, and let X_i denote the value of the variable x for unit i. When you write "ls y c x", E-views computes $\widehat{\beta}_0$ and $\widehat{\beta}_1$, the coefficients of the constant and of X_i in the OLS regression of Y_i on a constant and X_i. It follows from a result you saw during the lectures that

$$\begin{aligned}\widehat{\beta}_1 &= \frac{\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X})}{\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2}\\ \widehat{\beta}_0 &= \overline{Y}-\widehat{\beta}_1\overline{X},\end{aligned}$$

where $\overline{Y}$ denotes the average of the Y_is, while $\overline{X}$ denotes the average of the X_is.

Assume that X_i is a binary variable that is either equal to 0 or to 1. Let n_1 be the number of units with $X_i = 1$, and let $n_0 = n - n_1$ be the number of units with $X_i = 0$. The difference between the average of the Y_is among units with $X_i = 1$ and with $X_i = 0$ is $\frac{1}{n_1}\sum_{i:X_i=1} Y_i - \frac{1}{n_0}\sum_{i:X_i=0} Y_i$. The goal of the exercise is to show that

$$\begin{aligned}\frac{\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X})}{\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2} &= \frac{1}{n_1}\sum_{i:X_i=1} Y_i - \frac{1}{n_0}\sum_{i:X_i=0} Y_i\\ \overline{Y}-\widehat{\beta}_1\overline{X} &= \frac{1}{n_0}\sum_{i:X_i=0} Y_i.\end{aligned}$$

Watch out, these results are only true when X_i is binary.

1) First, let's consider the denominator of $\frac{\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X})}{\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2}$. Show that $\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2 = \overline{X}(1-\overline{X})$. Hint: remember that X_i is a binary variable.

2) Now, let's consider the numerator of $\frac{\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X})}{\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2}$.

a) Show that $\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X}) = \frac{1}{n}\sum_{i=1}^n Y_iX_i - \overline{Y}\,\overline{X}$.

b) Show that $\frac{1}{n}\sum_{i=1}^n Y_iX_i = \overline{X}\frac{1}{n_1}\sum_{i:X_i=1} Y_i$.

c) Use questions a) and b) to show that $\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X}) = \overline{X}\left(\frac{1}{n_1}\sum_{i:X_i=1} Y_i - \overline{Y}\right)$.

d) Show that $\overline{Y} = \overline{X}\frac{1}{n_1}\sum_{i:X_i=1} Y_i + (1-\overline{X})\frac{1}{n_0}\sum_{i:X_i=0} Y_i$.

e) Use questions c) and d) to show that

$$\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X}) = \overline{X}(1-\overline{X})\left(\frac{1}{n_1}\sum_{i:X_i=1} Y_i - \frac{1}{n_0}\sum_{i:X_i=0} Y_i\right).$$

3) Combine the results of questions 1) and 2) e) to show that

$$\widehat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n(Y_i-\overline{Y})(X_i-\overline{X})}{\frac{1}{n}\sum_{i=1}^n(X_i-\overline{X})^2} = \frac{1}{n_1}\sum_{i:X_i=1} Y_i - \frac{1}{n_0}\sum_{i:X_i=0} Y_i.$$

4) Finally, use the result of question 3) to show that

$$\widehat{\beta}_0 = \overline{Y}-\widehat{\beta}_1\overline{X} = \frac{1}{n_0}\sum_{i:X_i=0} Y_i.$$

Conclusion of the exercise. This exercise shows that $\widehat{\beta}_1$, the coefficient of X_i in OLS regression of Y_i on a constant and X_i measures the difference between the average of the Y_is among units with $X_i = 1$ and with $X_i = 0$. Putting it in other words, $\widehat{\beta}_1$ measures the difference between the average Y_i of units whose X_i differs by one unit. $\widehat{\beta}_0$ measures the average of Y_i among units with $X_i = 0$.