The coefficient of an OLS regression of Y_i on a constant.

Assume that you have a population with N units. Let y_k denote the value of the variable y for unit k in the population. You do not observe the values of the y_ks, and you would like to predict those values. However, contrary to the examples seen in class, you do not have any other variable x_k that you observe for all units, on which you can base your prediction. Therefore, you cannot regress y_k on x_k, like in the "OLS 1" slides, or on a constant and x_k, like in the "OLS 2" slides. Instead, you are going to regress y_k on a constant.

1. The coefficient of the constant in that regression is $\alpha = \underset{a \in \mathbb{R}}{\text{argmin}} \sum_{k=1}^{N} (y_k - a)^2$. Show that $\alpha = \frac{1}{N} \sum_{k=1}^{N} y_k$.

2. Can you compute α?

Assume that we draw without replacement a random sample of n units from the population, and for those units we measure their value of the variable y. Let Y_1 denote that value for the first unit we draw, let Y_2 denote that value for the second unit we draw, ..., let Y_n denote that value for the nth unit we draw. Y_1, Y_2, ..., Y_n are independent and identically distributed random variables.

3. Show that $E(Y_i) = \frac{1}{N} \sum_{k=1}^{N} y_k$.

4. Based on our random sample, our estimator of α is $\widehat{\alpha} = \underset{a \in \mathbb{R}}{\text{argmin}} \sum_{i=1}^{n} (Y_i - a)^2$. Show that $\widehat{\alpha} = \frac{1}{n} \sum_{i=1}^{n} Y_i$.

5. Show that $\widehat{\alpha}$ is an unbiased estimator of α.

6. If you are asked to predict the value of y for a unit not in your sample, what will be your predicted value for the y of that unit?