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?

Solution

1. Using chain rule and P4Sum, derivative of $\sum_{k=1}^{N}(y_k - a)^2$ wrt to a is $2\sum_{k=1}^{N}(a - y_k)$. Let's

find the value of a at which this derivative is equal to 0:

$$\begin{aligned}
&2\sum_{k=1}^{N}(a-y_k)=0\\
\Leftrightarrow&\sum_{k=1}^{N}(a-y_k)=0\\
\Leftrightarrow&\sum_{k=1}^{N}a-\sum_{k=1}^{N}y_k=0\\
\Leftrightarrow&Na-\sum_{k=1}^{N}y_k=0\\
\Leftrightarrow&Na=\sum_{k=1}^{N}y_k\\
\Leftrightarrow&a=\frac{1}{N}\sum_{k=1}^{N}y_k.
\end{aligned}$$

2nd equivalence: P3Sum, 3rd equivalence: P1Sum.

Derivative is increasing in $a \Rightarrow$ negative to the left of $\frac{1}{N}\sum_{k=1}^{N}y_k$, and positive to the right of $\frac{1}{N}\sum_{k=1}^{N}y_k$. Therefore, $\sum_{k=1}^{N}(y_k-a)^2$ reaches a minimum at $\alpha=\frac{1}{N}\sum_{k=1}^{N}y_k$.

2. No you cannot, as you do not observe the y_ks.

3. Y_i can be equal to:

- y_1 if the ith unit we randomly draw is unit 1. Probability that this happens is $1/N$.
- y_2 if the ith unit we randomly draw is unit 2. Probability that this happens is $1/N$.
- ...
- y_N if the ith unit we randomly draw is unit N. Probability that this happens is $1/N$.

Following the definition of the expectation of a random variable:

$$\begin{aligned}
E(Y_i)=&y_1 1/N+y_2 1/N+...+y_N 1/N\\
=&1/N(y_1+y_2+...+y_N)\\
=&1/N\sum_{k=1}^{N}y_k.
\end{aligned}$$

4. Using chain rule and P4Sum, derivative of $\sum_{i=1}^{n}(Y_i-a)^2$ wrt to a is $2\sum_{i=1}^{n}(a-Y_i)$. Let's

find the value of a at which this derivative is equal to 0:

$$\begin{aligned}
&2\sum_{i=1}^{n}(a-Y_i)=0\\
\Leftrightarrow&\sum_{i=1}^{n}(a-Y_i)=0\\
\Leftrightarrow&\sum_{i=1}^{n}a-\sum_{i=1}^{n}Y_i=0\\
\Leftrightarrow&na-\sum_{i=1}^{n}Y_i=0\\
\Leftrightarrow&na=\sum_{i=1}^{n}Y_i\\
\Leftrightarrow&a=\frac{1}{n}\sum_{i=1}^{n}Y_i.
\end{aligned}$$

2nd equivalence: P3Sum, 3rd equivalence: P1Sum.

Derivative is increasing in $a \Rightarrow$ negative to the left of $\frac{1}{n}\sum_{i=1}^{n}Y_i$, and positive to the right of $\frac{1}{n}\sum_{i=1}^{n}Y_i$. Therefore, $\sum_{i=1}^{n}(Y_i-a)^2$ reaches minimum at $\widehat{\alpha}=\frac{1}{n}\sum_{i=1}^{n}Y_i$.

5.

$$\begin{aligned}
E(\widehat{\alpha}) =&E\left(\frac{1}{n}\sum_{i=1}^{n}Y_i\right)\\
=&\frac{1}{n}\sum_{i=1}^{n}E(Y_i)\\
=&\frac{1}{n}\sum_{i=1}^{n}\frac{1}{N}\sum_{k=1}^{N}y_k\\
=&\frac{1}{n}n\frac{1}{N}\sum_{k=1}^{N}y_k\\
=&\frac{1}{N}\sum_{k=1}^{N}y_k.
\end{aligned}$$

1st equality: P2Expectation and P3Expectation. 2nd equality: question 3. 3rd equality: P1Sum.

6. Our predicted value will be the coefficient of the constant in the regression which is $\frac{1}{n}\sum_{i=1}^{n}Y_i$. **In other words, if we use a regression of y on a constant to predict the value of y for a unit not in the sample, our prediction will just be the average y of units in our sample.**