Minimising Sum of Squared Errors

Guillem Riambau. YSS211. Econometrics, Yale-NUS. Spring 2016.

Our problem: we want to find the relationship between x and y. In other words, what is the average effect of increasing x on y?

One possible solution: Minimising sum of squared residuals.

Note: This is only one solution. There are many more. We will see later in the course why, among the linear ones, this seems optimal.

We know that a linear model means $y_i = \alpha + \beta x_i + \varepsilon_i$

We are looking for estimates of α and β, which we will call $\hat{\alpha}$ and $\hat{\beta}$

α and β are the true population parameters. We cannot observe them, we do not know them. The best we can do is to estimate them using a sample of the population. For instance: let μ_h be the average height of NS men in Singapore. It is very hard to know μ_h, or nearly impossible (Can you measure the height of all NS men at the same day or week?), but we can get a good estimate of it if we get a random sample of NS men and take the average $\bar{h}$. $\bar{h}$ is hence the estimate of the true population parameter μ_h.

What are errors, and what are residuals?

The **error** of an observed value is the deviation of the observed value from the (unobservable) true value of a quantity of interest (for example, a population mean, like μ_h above).

The **residual** of an observed value is the difference between the observed value and the estimated value of the quantity of interest (like $\bar{h}$ above).

In a regression model:

$y_i = \alpha + \beta x_i + \varepsilon_i, \quad \varepsilon_i$ is the error.

But at most we can get estimates of α and β (which we will denote $\hat{\alpha}$ and $\hat{\beta}$), so that $y_i = \hat{\alpha} + \hat{\beta} x_i + \hat{\varepsilon}_i. \quad \hat{\varepsilon}_i$ is the *residual*.

We want to find $\hat{\alpha}$ and $\hat{\beta}$ such that the sum of the ε_i^2 is minimal.

Mathematically,

$$min_{\alpha,\beta} \sum_i^N \varepsilon_i^2 = min_{\alpha,\beta} \sum_i^N (y_i - \alpha - \beta x_i)^2 \tag{1}$$

We need to take derivatives with respect to α and β. Hence, our first order conditions are:

$$(2)\qquad \begin{aligned} [\alpha] \quad & \sum_i^N 2\left(y_i-\alpha-\beta x_i\right)(-1)=0 \\ [\beta] \quad & \sum_i^N 2\left(y_i-\alpha-\beta x_i\right)(-x_i)=0 \end{aligned}$$

Take the first one (FOC with respect to α) and develop it:

$$(3)\qquad \begin{aligned} \sum_i^N 2\left(y_i-\alpha-\beta x_i\right)(-1)=0 & \quad \Longleftrightarrow \\ \sum_i^N \left(y_i-\alpha-\beta x_i\right)=0 & \quad \Longleftrightarrow \\ \sum_i^N \left(y_i-\beta x_i\right)=\sum_i^N \alpha & \quad \Longleftrightarrow \\ \sum_i^N \left(y_i-\beta x_i\right)=N\alpha & \quad \Longleftrightarrow \\ \frac{1}{N}\sum_i^N \left(y_i-\beta x_i\right)=\alpha & \quad \Longleftrightarrow \\ & \bar{y}-\beta\bar{x}=\hat{\alpha} \end{aligned}$$

where $\bar{y}=\frac{1}{N}\sum_i^N y_i$ and $\bar{x}=\frac{1}{N}\sum_i^N x_i$

Now take the second one (FOC with respect to β) and develop it:

$$(4)\qquad \begin{aligned} \sum_i^N 2\left(y_i-\alpha-\beta x_i\right)(-x_i)=0 & \quad \Longleftrightarrow \\ \sum_i^N \left(y_i-\alpha-\beta x_i\right)(x_i)=0 & \quad \Longleftrightarrow \\ \sum_i^N \left(y_i-\alpha\right)(x_i)=\sum_i^N \left(\beta x_i\right)(x_i) & \quad \Longleftrightarrow \\ \sum_i^N \left(y_i x_i-\alpha x_i\right)=\beta\sum_i^N \left(x_i^2\right) & \end{aligned}$$

Plug the last line of (3) (i.e. $\bar{y}-\beta\bar{x}=\hat{\alpha}$) into the FOC with respect to β:

$$
\begin{aligned}
\sum_i^N (y_i x_i - \alpha x_i) = \beta \sum_i^N \left(x_i^2\right) \qquad & \iff \\
\sum_i^N (y_i x_i - (\bar{y} - \beta \bar{x}) x_i) = \beta \sum_i^N \left(x_i^2\right) \qquad & \iff \\
\sum_i^N (y_i x_i - \bar{y} x_i) + \sum_i^N \beta \bar{x} x_i = \beta \sum_i^N \left(x_i^2\right) \qquad & \iff \\
\sum_i^N (y_i x_i - \bar{y} x_i) = \beta \sum_i^N \left(x_i^2\right) - \sum_i^N \beta \bar{x} x_i \qquad & \iff \\
\sum_i^N (y_i x_i - \bar{y} x_i) = \beta \left(\sum_i^N x_i^2 - \sum_i^N \bar{x} x_i \right) \qquad & \iff \\
\frac{\sum_i^N (y_i x_i - \bar{y} x_i)}{\left(\sum_i^N x_i^2 - \sum_i^N \bar{x} x_i \right)} = \hat{\beta} &
\end{aligned} \tag{5}
$$

So now we have β as a function of the y_i and x_i only.

Now, can we develop this expression a bit?

Let's start with the numerator in the last line of (5):

$$
\begin{aligned}
\sum_i^N (y_i x_i - \bar{y} x_i) = \\
\sum_i^N (y_i x_i) - N \bar{y} \bar{x} = \\
\sum_i^N (y_i x_i) - N \bar{y} \bar{x} - N \bar{y} \bar{x} + N \bar{y} \bar{x} = \\
\sum_i^N (y_i x_i) - \sum_i^N (y_i \bar{x}) - \sum_i^N (x_i \bar{y}) + \sum_i^N (\bar{y} \bar{x}) = \\
\sum_i^N ((y_i x_i) - (y_i \bar{x}) - (x_i \bar{y}) + (\bar{y} \bar{x})) = \\
\sum_i^N (y_i - \bar{y}) (x_i - \bar{x})
\end{aligned} \tag{6}
$$

Now let's switch with the denominator in the last line of (5):

$$
\begin{aligned}
&\sum_i^N x_i^2 - \bar{x}\sum_i^N x_i = \\
&\sum_i^N x_i^2 - \bar{x}\sum_i^N x_i - \bar{x}\sum_i^N x_i + \bar{x}\sum_i^N x_i = \\
&\sum_i^N x_i^2 - 2\bar{x}\sum_i^N x_i + \bar{x}\sum_i^N x_i = \\
&\sum_i^N x_i^2 - 2\bar{x}\sum_i^N x_i + \bar{x}N\bar{x} = \\
&\sum_i^N x_i^2 - 2\bar{x}\sum_i^N x_i + N\bar{x}^2 = \\
&\sum_i^N x_i^2 - 2\bar{x}\sum_i^N x_i + \sum_i^N \bar{x}^2 = \\
&\sum_i^N \left(x_i^2 - 2\bar{x}x_i + \bar{x}^2\right) = \\
&\sum_i^N \left(x_i - \bar{x}\right)^2
\end{aligned} \tag{7}
$$

Hence, plugging (7) and (6) into (5), we have

$$
\hat{\beta} = \frac{\sum_i^N \left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum_i^N \left(x_i - \bar{x}\right)^2} \tag{8}
$$

and so

$$
\begin{aligned}
\hat{\alpha} = \bar{y} - \hat{\beta}\bar{x} \qquad &\Longleftrightarrow \\
\hat{\alpha} = \bar{y} - \frac{\sum_i^N \left(y_i - \bar{y}\right)\left(x_i - \bar{x}\right)}{\sum_i^N \left(x_i - \bar{x}\right)^2}\bar{x}
\end{aligned} \tag{9}
$$