regression in the population

preliminaries

We start with a set of ordered pairs $\{\langle X_1, Y_1\rangle, \langle X_2, Y_2\rangle, \langle X_3, Y_3\rangle, \ldots\}$. X_i are vectors of real numbers, Y_i are real numbers. Neither are random variables.

The CEF minimises $E[u_i^2]$

Some algebraic facts

We write the equality:

$$Y_i = f(X_i) + u_i$$

Where Y_i and X_i are known, but u_i is "unknown" in the sense that it is a function of f.

Minimisation problem

Suppose we want to solve

$$\min_{f(X_i)} E[u_i^2] \leftrightarrow \min_{f(X_i)} E[(Y_i - f(X_i))^2]$$

The solution is $f(X_i) = E[Y_i \mid X_i]$. Suppose we specify $f(X_i)$ as such, we then get:

$$Y_i = E[Y_i \mid X_i] + u_i$$

Now f is known and u_i is known (by the subtraction $u_i = Y_i - E[Y_i \mid X_i]$).

The LRM minimises $E[(e_i + u_i)^2]$

Some algebraic facts

Now we write the following equality:

$$E[Y_i \mid X_i] = \beta X_i + e_i$$

This says that $E[Y_i \mid X_i]$ is equal to a linear function of X_i plus some number e_i.

We then have

$$\begin{aligned} Y_i &= E[Y_i \mid X_i] + u_i \\ &= \beta X_i + e_i + u_i \end{aligned}$$

As before u_i is known, whereas e_i is a function of β.

Minimisation problem

Suppose we want to solve

$$\min_\beta E[(e_i + u_i)^2] \leftrightarrow \min_\beta E[(Y_i - \beta X_i)^2]$$

The solution is such that β is equal to the vector of linear least squares regression coefficients. I won't do the math for the fully general case (multiple regression), but only for univariate regression. It is:

> $$Y_i = \beta_0 + \beta_1 X_i + e_i + u_i$$
>
> $$\min_{\beta} E[(e_i + u_i)^2] \leftrightarrow$$
> $$\min_{\beta} E[(Y_i - \beta_0 + \beta_1 X_i)^2] \leftrightarrow$$
> $$\beta_0 = E[Y_i] - \beta_1 E[X_i] \text{ and } \beta_1 = \frac{cov(Y_i, X_i)}{var(X_i)}$$

Suppose we specify that β is equal to these solution values. Now that β is known (for the univariate case, now that β_0 and β_1 are known), e_i is known too (by the subtraction $e_i = E[Y_i \mid X_i] - \beta X_i$). As before u_i is known.

Thus, in our regression equation,

$$Y_i = \beta_0 + \beta_1 X_i + e_i + u_i$$

all of Y_i, X_i, β_0, β_1, e_i and u_i are known.