Simple Regression Model

1. The Model

$$y_i = \beta_0 + \beta_1 x_i + u_i \qquad i = 1, \ldots, n$$

where

- y_i = dependent variable
- x_i = independent variable
- u_i = disturbance/error term

Eg: y = wage (measured in 1976 dollars per hr)

x = education (measured in yrs of schooling)

Alternative names for

y	x
explained variable	explanatory variable
response variable	control/treatment
predicted variable	predictor
regressand	regressor

:The joint properties of $(y,x) \Leftrightarrow (u,x)$ are key to determining the statistical properties of any estimator. So having a good idea of what generates the disturbance is important for good empirical work.

: What generates u?

- left out variables
- approximation error for functional form
- measurement error in y
- inherent randomness in relationship between y and x

: (β_0, β_1) are parameters called the "regression coefficients:

- β_0 = intercept
- β_1 = slope coefficient
- each choice of (β_0, β_1) defines a u

  $\Leftrightarrow$ a particular u will define (β_0, β_1)

:ASIDE

Sometimes it is useful to imagine that the slope coefficient varies across observations, i.e.

$$\beta_{1i} = \beta_1(x_i, u_i, v_i)$$

where v_i is independent of (x_i, u_i). Such an underlying model explains why regression coefficients vary from sample to sample. So for example, response to the math-stat review might depend on how much math you know (x_i), persistence and intelligence (u_i), and luck (v_i)

- The (classical) random coefficients model has $\beta_1(v_i)$ –– not very interesting.
- If we have important nonlinearities, then we'll see $\beta_1(x_i)$.
- Tough case is if β_1 varies with the unobservables u_i.

: Linearity is restrictive, but not as bad as it first appears. For example, suppose we have

$$V = c_0 W^{c_1} \varepsilon$$

define $y = \ln(V)$ $x = \ln(W)$ $u = \ln(\varepsilon)$. Then

$$y = \beta_0 + \beta_1 x + u \qquad \beta_0 = \ln(c_0) \;\; \beta_1 = c_1$$

Other transformations (of either the LHS or RHS) may help to induce *linearity in parameters*, which is what we need.

: If we think of (y, x) as random variables, then $f(y|x)$ tells us everything about how probability assessment of y vary with x. Regression models focus on some measure of the "central tendency" of y given information about x.

(A) $E(u) = 0$ and $cov(u, x) = 0$

$$(\Leftrightarrow E(u) = E(ux) = 0)$$

$\therefore (\beta_0, \beta_1)$ are coefficents of BLP

(B) $E(u|x) = 0$

$$(\Leftrightarrow E(ug(x)) = 0 \ \ \forall g(x) \in L^2)$$

$$\therefore E(y|x) = \beta_0 + \beta_1 x$$

conditional mean or "Population Regression Function"

(C) u independent of x

(D) $y = \beta_0 + \beta_1 x$ is a causal relationship (not statistical)

2. Deriving the OLS estimates

Method 1

By definition, OLS estimates satisfy

$$(\widehat{\beta}_0, \widehat{\beta}_1) = \arg \min_{\widetilde{\beta}_0, \widetilde{\beta}_1} \sum_{i=1}^{n} (y_i - \widetilde{\beta}_0 - \widetilde{\beta}_1 x_i)^2$$

Rk: Given any candidate $\left(\widetilde{\beta}_0, \widetilde{\beta}_1\right)$,

- $\widetilde{y}_i = \widetilde{\beta}_0 + \widetilde{\beta}_1 x_i$ is called the fitted value,
- $\widetilde{u}_i \equiv y_i - \widetilde{y}_i$ is called the residual.

I'll reserve $(\widehat{y}_i, \widehat{u}_i)$ for the OLS values.

From the F.O.C., we obtain the "normal equations"

N1 $$\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \quad \Leftrightarrow \sum_{i=1}^{n}\widehat{u}_i = 0$$

N2 $$\sum_{i=1}^{n}x_i(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \quad \Leftrightarrow \sum_{i=1}^{n}x_i\widehat{u}_i = 0$$

Rk: Lots of authors use notation (b_0, b_1) for OLS estimations

Method 2

The BLP generates a disturbance u that satisfies $E(u) = E(xu) = 0$. So (β_0, β_1) of interest satisfies

$$E(y_i - \beta_0 - \beta_1 x_i) = 0$$

$$E(x(y_i - \beta_0 - \beta_1 x_i)) = 0$$

The method of moments approach picks estimators to satisfy the sample counterparts:

N1' $$\frac{1}{n}\sum_{i=1}^{n}(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \quad \Leftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n}\widehat{u}_i = 0$$

N2' $$\frac{1}{n}\sum_{i=1}^{n}x_i(y_i - \widehat{\beta}_0 - \widehat{\beta}_1 x_i) = 0 \quad \Leftrightarrow \quad \frac{1}{n}\sum_{i=1}^{n}x_i\widehat{u}_i = 0$$

Computing $(\widehat{\beta}_0, \widehat{\beta}_1)$

From N1/N1', we get

$$\overline{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{x} \qquad \therefore\ \widehat{\beta}_0 = \overline{y} - \widehat{\beta}_1 \overline{x}$$

Substitute into N2/N2'. If $\sum (x_i - \overline{x})^2 > 0$, then

$$\widehat{\beta}_1 = \frac{\sum (x_i - \overline{x})(y_i - \overline{y})}{\sum (x_i - \overline{x})^2}$$

Rks: Sample Regression function is the estimated PRF (aka fitted regression line)

$$\widehat{y} = \widehat{\beta}_0 + \widehat{\beta}_1 x$$

:Prove

$$\widehat{\beta}_1 = \frac{\sum x_i (y_i - \overline{y})}{\sum (x_i - \overline{x})^2} = \frac{\sum (x_i - \overline{x}) y_i}{\sum (x_i - \overline{x})^2}$$

:We've shown that $(\widehat{\beta}_0, \widehat{\beta}_1)$ is unique provided $\sum(x_i - \overline{x})^2 > 0$. Can you show that the fitted values $\hat{y}_i$ have the same values for ANY solution of the normal equations?

Eg: $y =$ wage (measured in 1976 dollars per hr)

$x =$ education (measured in yrs of schooling)

Suppose the fitted regression line is

$$\widehat{wage} = -0.90 + 0.54\ educ$$

(Notice that I don't report many digits–use lots for accurate calculations but don't present them!)

- $educ = 0$ then $\widehat{wage} = -.90$ (-90 cents per hr)
- $educ = 8$ then $\widehat{wage} = -.90 + 0.54(8) = 3.42$ (\$3.42 per hr)

Q: What's the premium to completing high school vs. grade 8?

Q: What's the premium to completing university vs HS?

Algebra of OLS

Everything follows from the normal equations

1. $\sum \widehat{u}_i = 0$
2. $\sum x_i \widehat{u}_i = 0$

Some implications of 1. & 2.

3. $\overline{y} = \widehat{\beta}_0 + \widehat{\beta}_1 \overline{x}$ (reg line goes through sample mean)
4. $\sum \widehat{u}_i \widehat{y}_i = 0$

Proof:

$$\begin{aligned}\sum \widehat{u}_i \widehat{y}_i &= \sum \widehat{u}_i \left(\widehat{\beta}_0 + \widehat{\beta}_1 x_i\right) \\ &= \widehat{\beta}_0 \sum \widehat{u}_i + \widehat{\beta}_1 \sum x_i \widehat{u}_i \\ &= 0 + 0 \quad \text{by 1. and 2. (resp)}\end{aligned}$$

Rk: The same proof shows $\sum \hat{u}_i(c_0 + c_1 x_i) = 0 \ \forall (c_0, c_1)$

4'. $\sum \hat{u}_i(\hat{y}_i - \overline{y}) = 0$

5. (Analysis of variance)

$$y_i = \hat{y}_i + \hat{u}_i$$

$$\Leftrightarrow y_i - \overline{y} = \hat{y}_i - \overline{y} + \hat{u}_i$$

Therefore

$$\sum (y_i - \overline{y})^2 = \sum (\hat{y}_i - \overline{y})^2 + \sum \hat{u}_i^2 + 2\sum \hat{u}_i(\hat{y}_i - \overline{y})$$

$$= \sum (\hat{y}_i - \overline{y})^2 + \sum \hat{u}_i^2 \quad \text{by 4'}$$

$SST = SSE + SSR$ (Textbook's notation)

Total Sum of Squares $=$ Explained SS$+$Residual SS

6. Coefficient of Determination

$$R^2 = \frac{SSE}{SST} = 1 - \frac{SSR}{SST}$$

:$0 \le R^2 \le 1$

- $R^2 = 1$ says exact lin. rel. between y and x
- $R^2 = 0$ says no lin. rel. between y and x

:R^2 gives the fraction of the variance of y that's "explained" by the model

Exercise:

Show $R^2 = r_{yx}^2 = r_{y\hat{y}}^2$ where

$$r_{yx} = \frac{\sum(y - \overline{y})(x - \overline{x})}{[\sum(y - \overline{y})^2 \sum(x - \overline{x})^2]^{1/2}}$$

7. Computing the explained sum of squares

$$\sum (\hat{y}_i - \overline{y})^2 = \hat{\beta}_1^2 \sum (x - \overline{x})^2$$
$$= \hat{\beta}_1 \sum (y - \overline{y})(x - \overline{x})$$

Rk: This means we can compute $(\hat{\beta}_0, \hat{\beta}_1), \sum \hat{u}_i^2,$ and R^2 from

$$\begin{pmatrix} \sum y_i^2 & \sum y_i & \sum y_i x_i \\ & n & \sum x_i \\ & & \sum x_i^2 \end{pmatrix}$$

(Before proceeding, you should review the Matrix Algebra 1 notes)

Regression Model in Matrix Notation

$$
\begin{aligned}
y_1 &= \beta_1 + \beta_2 x_1 + u_1 \\
y_2 &= \beta_1 + \beta_2 x_2 + u_2 \\
\vdots &= \vdots \\
y_n &= \beta_1 + \beta_2 x_n + u_n
\end{aligned}
$$

Define

$$
y = \begin{bmatrix} y_1 \\ \vdots \\ y_n \end{bmatrix} \quad X_1 = \begin{bmatrix} 1 \\ \vdots \\ 1 \end{bmatrix} \quad X_2 = \begin{bmatrix} x_1 \\ \vdots \\ x_n \end{bmatrix} \quad u = \begin{bmatrix} u_1 \\ \vdots \\ u_n \end{bmatrix}
$$

$$
(nx1) \qquad\qquad (nx1)
$$

$$X = \begin{bmatrix} X_1 & X_2 \end{bmatrix} \quad \beta = \begin{bmatrix} \beta_1 \\ \beta_2 \end{bmatrix}$$

$$(nx2) \qquad\qquad (2x1)$$

:In matrix notation, the model is

$$y = X\beta + u$$

- LHS is a vector in $\mathbb{R}^n$
- RHS is a vector in $\mathbb{R}^n$
- Equality holds component by component

$$y_i = X_i\beta + u_i \quad (X_i \text{ is the } i^{th} \text{ row of } X)$$

$$= 1 \cdot \beta_1 + x_i \cdot \beta_2 + u_i$$

:Derivation of the OLS estimator

$$\widehat{\beta} = \arg\min_{\widetilde{\beta}\in\mathbb{R}^2} \widetilde{u}'\widetilde{u}$$

$$= \arg\min_{\widetilde{\beta}\in\mathbb{R}^2} (y - X\widetilde{\beta})'(y - X\widetilde{\beta})$$

Rk:

$$\widetilde{u}'\widetilde{u} = \begin{bmatrix} \widetilde{u_1} & \cdots & \widetilde{u}_n \end{bmatrix} \begin{bmatrix} \widetilde{u_1} \\ \vdots \\ \widetilde{u}_n \end{bmatrix}$$

$$= \sum_{i=1}^{n} \widetilde{u}_i^2$$

:Normal equations

Define

$$\hat{u} = y - X\hat{\beta} \quad \text{where } \hat{u}, y \in \mathbb{R}^n,\ X \in \mathbb{R}^{nx2},\ \hat{\beta} \in \mathbb{R}^2$$

In matrix notation, the normal equations are

$$X'\hat{u} = 0$$

$$\Leftrightarrow X'(y - X\hat{\beta}) = 0$$

But

$$X' = \begin{bmatrix} 1 & \cdots & 1 \\ x_1 & \cdots & x_n \end{bmatrix} \quad y - X\hat{\beta} = \begin{bmatrix} y_1 - \hat{\beta}_1 - \hat{\beta}_2 x_1 \\ \vdots \\ y_n - \hat{\beta}_1 - \hat{\beta}_2 x_n \end{bmatrix}$$

$$\therefore\ X'(y - X\widehat{\beta}) = \begin{bmatrix} \sum\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right) \\ \sum x_i\left(y_i - \widehat{\beta}_1 - \widehat{\beta}_2 x_i\right) \end{bmatrix} = 0$$

:Matrix notation for OLS estimator

$$\begin{aligned} X'(y - X\widehat{\beta}) &= 0 \\ &\Leftrightarrow X'y - X'X\widehat{\beta} = 0 \\ &\Leftrightarrow X'X\widehat{\beta} = X'y \end{aligned}$$

Rk:

$$X'y = \begin{bmatrix} \sum y_i \\ \sum x_i y_i \end{bmatrix} \quad X'X = \begin{bmatrix} n & \sum x_i \\ \sum x_i & \sum x_i^2 \end{bmatrix}$$

Q: Can we solve $X'X\widehat{\beta} = X'y$ for $\widehat{\beta}$?

A: Yes (always) and the solution is unique iff $\det(X'X) \neq 0$.

Using

$$\det(X'X) = n\sum x_i^2 - \sum x_i \sum x_i$$
$$= n\sum (x_i - \overline{x})^2$$

We see that iff $\sum (x_i - \overline{x})^2 > 0,$ then

$$\exists (X'X)^{-1} \text{ s.t. } (X'X)^{-1}(X'X) = I_2$$

$$\therefore\ (X'X)^{-1}(X'X)\widehat{\beta} = (X'X)^{-1}X'y$$
$$\Leftrightarrow I_2\widehat{\beta} = (X'X)^{-1}X'y$$
$$\Leftrightarrow \widehat{\beta} = (X'X)^{-1}X'y$$

Using

$$(X'X)^{-1} = \frac{1}{n\sum(x_i - \overline{x})^2}\begin{bmatrix} \sum x_i^2 & -\sum x_i \\ -\sum x_i & n \end{bmatrix}$$

We get

$$\begin{bmatrix} \widehat{\beta}_1 \\ \widehat{\beta}_2 \end{bmatrix} = \begin{matrix} \overline{y} - \widehat{\beta}_2 \overline{x} \\ \sum(x_i - \overline{x})(y_i - \overline{y}) / \sum(x_i - \overline{x})^2 \end{matrix}$$

Rk: I'll show how to derive other results using matrix notation when we cover ch3.

:An Important decomposition

Note that we can write

$$\hat{y} = X\hat{\beta} = X(X'X)^{-1}X'y$$
$$\equiv Py$$

where $P = P'$ (symmetric), and $P = P^2$ (idempotent)

Also

$$\hat{u} = y - \hat{y} = (I - P)y$$
$$\equiv My$$

where $M = M'$ (symmetric), and $M = M^2$ (idempotent).

Therefore

$$y \equiv \hat{y} + \hat{u}$$
$$= Py + My$$