Final thoughts

- Unlikely that $\hat{\beta}_1$ from simple regression captures <u>true</u> effect of x on y.
  (unbiased)

- Need u & x uncorrelated.

- Randomized trials help in <u>some</u> cases.

Say y: health

x: w/ or w/o insurance

u: unobs. factors such as diet, lifestyle, age, etc.

Random manipulation: coin toss determining x

$\Rightarrow$ on <u>avg.</u> the 2 groups are comparable (provided large enough sample)

Ch: 3 → Multiple Regression

Motivation :

y	indep vars. $x_1, x_2, \ldots$	u
wage	$\boxed{\text{educ.}}$ IQ exper.	more likely to be exogenous under multiple than simple regression
crime	$\boxed{\text{police}}$ unemp. rt. graduation rt.	

k indep. vars. $\quad k = 2$

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

β_1 : Δy for $\Delta x_1 = 1$; $\quad \Delta x_2 = 0$, $\Delta u = 0$

Estimation

Ass^{n}s :

$$E(u) = 0$$

$$E(x_1 u) = 0$$

$$E(x_2 u) = 0$$

OLS estimates of β_0, β_1, and β_2 : $\hat{\beta}_0$, $\hat{\beta}_1$, and $\hat{\beta}_2$

$$\frac{1}{n} \sum_{i=1}^{n} \boxed{\left(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_{i1} - \hat{\beta}_2 x_{i2} \right)} = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} x_{i1} \square = 0$$

$$\frac{1}{n} \sum_{i=1}^{n} x_{i2} \square = 0$$

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_{i1} + \hat{\beta}_2 x_{i2}$$

(fitted value)

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

(residual)

Goodness of fit : $R^2 = \dfrac{SSE}{SST}$

$= 1 - \dfrac{SSR}{SST}$ non $\downarrow$ in k (# x's)

Adjusted R^2 : $\bar{R}^2 = 1 - \dfrac{\frac{SSR}{(n-k-1)}}{\frac{SST}{(n-1)}}$

As $k \uparrow$ $SSR \downarrow$ $(n-k-1) \downarrow$

n : sample size

Expected value

unbiased OLS estimators

$$E(\hat{\beta}_j) = \beta_j \qquad j = 0, 1, \ldots, k$$

$k \rightarrow$ # x's

under certain assumptions

- linear in param.
- random sampling
- variation in each x

  No linear rel^{n} among regressors
- u unrelated to each x

  $E(u \mid x_1, x_2, \ldots, x_k) = E(u) = 0$

$wage = \beta_0 + \beta_1 age + \beta_2 exper + \beta_3 educ + u$

$age = 6 + educ. + exper.$

Omitted Variable Bias

True model satisfying cond^{n}s for unbiasedness :

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + u$$

edu. IQ

wage

smoking alcohol

bwght

x_2 : omitted ; estimate :

$$y = \beta_0 + \beta_1 x_1 + v$$