Univariate Regression: Hypothesis Tests and Confidence Intervals (SW Ch. 5)
Part 2

Dragos Ailoae

dailoae@gradcenter.cuny.edu

Advanced Economics and Business Statistics
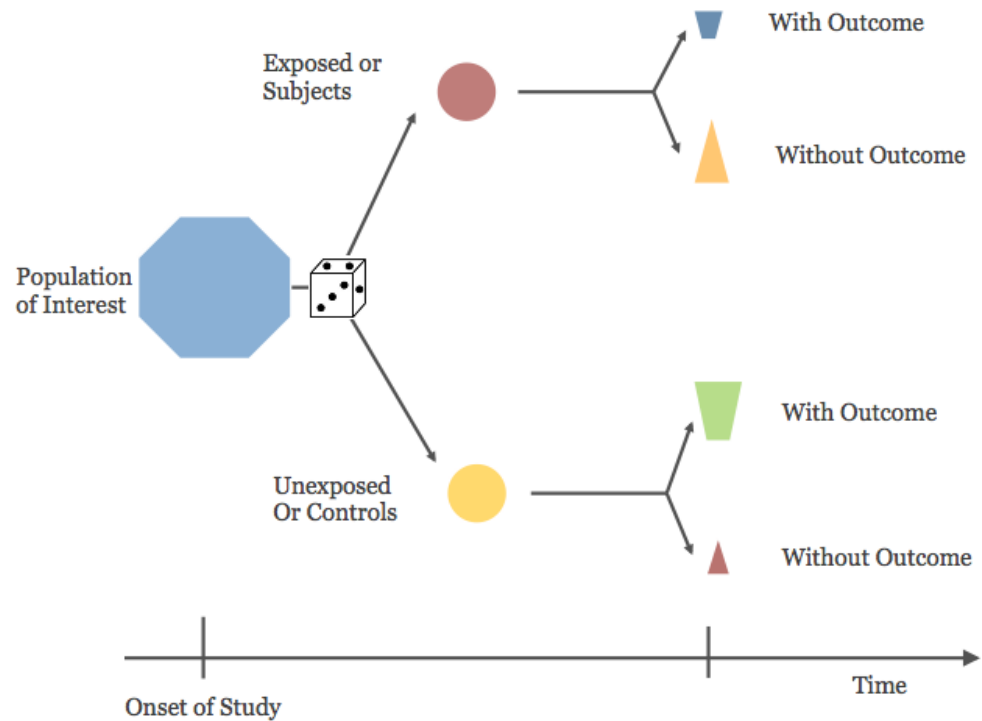ECON-4400w - Fall 2022

Brooklyn College
Oct 26, 2022

Outline

Experimental data (RCTs)

Randomized Control Trials:

- All participants are randomly assigned into **two** groups.

- The control group receives no treatment (or placebo)

- The experimental group receives the treatment.

- After a follow-up period, compare the two groups

Regression when *X* is Binary (Section 5.3)

Sometimes a regressor is binary:

- $X = 1$ if treated (experimental drug), $= 0$ if not

- $X = 1$ if small class size, $= 0$ if not

- $X = 1$ if female, $= 0$ if male

Binary regressors are sometimes called "dummy" variables.

- Oxford dictionary

  dummy /ˈdəmē/ - something designed to resemble and serve as
                            a substitute for the real or usual thing

Interpreting regressions with a binary regressor $Y_i = \beta_0 + \beta_1 X_i + u_i$, where X is binary ($X_i = 0$ or 1):

So far, β_1 has been called a "slope," but that doesn't make sense if X is binary.

How do we interpret regression with a binary regressor?

When $X_i = 0$, $Y_i = \beta_0 + u_i$

- the mean of Y_i is β_0

- that is, $E(Y_i|X_i=0) = \beta_0$

When $X_i = 1$, $Y_i = \beta_0 + \beta_1 + u_i$

- the mean of Y_i is $\beta_0 + \beta_1$

- that is, $E(Y_i|X_i=1) = \beta_0 + \beta_1$

so: $\qquad\qquad \beta_1 = E(Y_i|X_i=1) - E(Y_i|X_i=0)$

$\qquad\qquad\qquad$ = population difference in group means

Excel live session

Interpreting regressions with a binary regressor

$Y_i = \beta_0 + \beta_1 X_i + u_i$, where X is binary $(X_i = 0 \text{ or } 1)$:

Example: Let $D_i = \begin{cases} 1 \text{ if } STR_i < 20 \\ 0 \text{ if } STR_i \geq 20 \end{cases}$

OLS regression: $\widehat{TestScore} = 650.0 + 7.4 \times D$

$$(1.3) \quad (1.8)$$

Tabulation of group means:

Class Size	Average score ($\bar{Y}$)	Std. dev. (s_Y)	N
Small ($STR < 20$)	657.4	19.4	238
Large ($STR \geq 20$)	**650.0**	17.9	182

Difference in means: $\bar{Y}_{small} - \bar{Y}_{large} = 657.4 - 650.0 = \textbf{7.4}$

Standard error $SE = \sqrt{\dfrac{S_s^2}{n_s} + \dfrac{S_l^2}{n_1}} = \sqrt{\dfrac{19.4^2}{238} + \dfrac{17.9^2}{182}} = \textbf{1.8}$

Summary: regression when X_i is binary

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

- β_0 = mean of Y when $X = 0$

- $\beta_0 + \beta_1$ = mean of Y when $X = 1$

- β_1 = difference in group means, $X = 1$ minus $X = 0$

- $SE(\hat{\beta}_1)$ has the usual interpretation

- t-statistics, confidence intervals constructed as usual

- This is another way (an easy way) to do difference-in-means analysis

- The regression formulation is especially useful when we have additional regressors (coming soon)