Lecture 11
Introduction to Regression Analysis

by Hiro Kasahara

Vancouver School of Economics
University of British Columbia

Introduction

- The simple linear regression model is used to study the relationship between two variables.
- It has many limitations, but nevertheless there examples in the literature where the simple linear regression is applied.
- It is also a good starting point to learning the regression technique.

Sample and population

- The econometrician observes random data:

observation	dependent variable	regressor
1	Y_1	X_1
2	Y_2	X_2
$\vdots$	$\vdots$	$\vdots$
n	Y_n	X_n

- A pair X_i, Y_i is called an observation.
- Sample : $\{(X_i, Y_i) : i = 1, \ldots, n\}$.
- The population is the joint distribution of the sample.

The model

- We model the relationship between Y and X using the conditional expectation :

$$E\left(Y_i|X_i\right) = \alpha + \beta X_i.$$

- Intercept : $\alpha = E\left(Y_i|X_i = 0\right)$.
- Slope : β measures the effect of a unit change in X on Y :

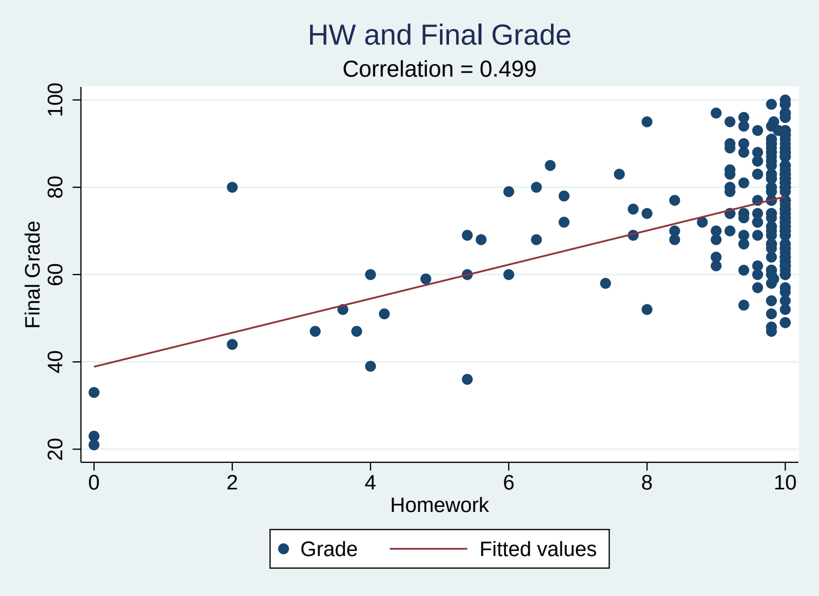
$$\beta = E\left(Y_i|X_i = x+1\right) - E\left(Y_i|X_i = x\right)$$

Ordinary Least Square (OLS) estimator for α and β

- Given the data $\{X_i, Y_i\}_{i=1}^n$, the OLS estimator, $\hat{\alpha}$ and $\hat{\beta}$, solves

$$\min_{\alpha,\beta} \sum_{i=1}^{n} (Y_i - \alpha - \beta X_i)^2.$$

Example: HW Grade and Final Grade



Example: HW Grade and Final Grade

$$\underset{}{\textit{Final}} = \underset{(4.11)}{38.89} + \underset{(0.44)}{3.90} \times \textit{Homework}.$$

```
. regress grade hw

      Source |       SS           df       MS      Number of obs   =       240
-------------+----------------------------------   F(1, 238)       =     79.10
       Model |  11758.7143         1  11758.7143   Prob > F        =    0.0000
    Residual |  35381.0815       238  148.660006   R-squared       =    0.2494
-------------+----------------------------------   Adj R-squared   =    0.2463
       Total |  47139.7958       239  197.237639   Root MSE        =    12.193

------------------------------------------------------------------------------
       grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          hw |   3.897495   .4382306     8.89   0.000     3.034189    4.760802
       _cons |   38.89254   4.113855     9.45   0.000     30.78832    46.99676
------------------------------------------------------------------------------
```

Example: High_HW and Final Grade

- Define

$$\text{High\_HW} = \begin{cases} 0 & \text{if Homework} \leq 0.6 \\ 1 & \text{if Homework} > 0.6 \end{cases}$$

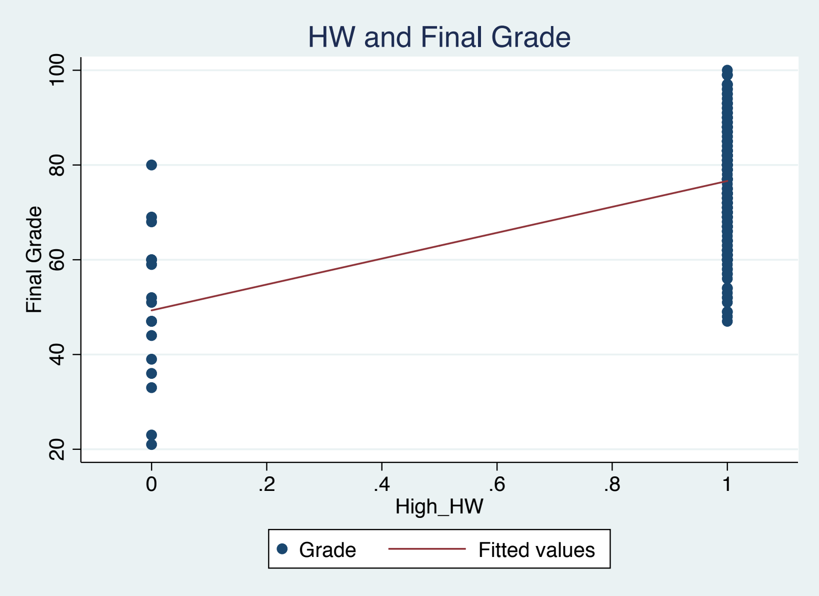
- Regression Model:

$$E(\text{ Final}_i \mid \text{High\_HW}_i\text{ }) = \alpha + \beta \text{ High\_HW}_i.$$

- β represents the difference in population means!

$$\beta = E(\text{Final}_i|\text{High\_HW}_i = 1) - E(\text{Final}_i|\text{High\_HW}_i = 0).$$

Example: High_HW and Final Grade



High_HW = 1 if Homework>0.6

$$\underset{(4.02)}{Final = 49.31} + \underset{(4.09)}{27.31} \times High\_HW.$$

95 percent Confidence Interval: [19.25, 35.38].

```
. regress grade High_HW, robust

Linear regression                              Number of obs     =        240
                                               F(1, 238)         =      44.49
                                               Prob > F          =     0.0000
                                               R-squared         =     0.2363
                                               Root MSE          =     12.299

------------------------------------------------------------------------------
             |               Robust
       grade |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
     High_HW |    27.3125   4.094581     6.67   0.000     19.24625    35.37875
       _cons |    49.3125   4.015515    12.28   0.000     41.40201    57.22299
------------------------------------------------------------------------------
```

95 confidence interval for $\mu_y - \mu_x$

- μ_y = population mean of final grade among students who receive HW grade more than 6 out of 10.
- μ_x = population mean of final grade among students who receive HW grade less than 6 out of 10.
- Sample:

$$\bar{Y} = 76.63, \quad \bar{X} = 49.31, \quad s_y = 11.96, \quad s_x = 16.51.$$

$$n_y = 224, \quad n_x = 16.$$

95 percent Confidence Interval for $\mu_y - \mu_x$

$$\bar{Y} = 76.63, \quad \bar{X} = 49.31, \quad s_y = 11.96, \quad s_x = 16.51.$$
$$n_y = 224, \quad n_x = 16.$$

- $\hat{\mu}_y - \hat{\mu}_x = 76.63 - 49.31 = 27.32$
- 95 percent confidence interval is given by

$$(76.63 - 49.31) \pm 1.96\sqrt{\frac{(11.96)^2}{224} + \frac{(16.51)^2}{16}}$$

or

$$27.32 \pm 8.24 = [19.08, 35.56]$$

Computer usage and academic performance

"The impact of computer usage on academic performance: Evidence from a randomized trial at the United States Military Academy"

by Susan Payne Carter, Kyle Greenberg, Michaels. Walker

Economics of Education Review, 2017.

Computer usage and academic performance

- The United States Military Academy.
- 726 students in the principles of economics course are randomly assigned to control group and treatment group.
- Treatment Group 1 (17 classrooms, 270 students): Laptop and Tablet Computers Permitted.
- Treatment Group 2 (15 classrooms, 208 students): Tablet Permitted with Face-up.
- Control Group (18 classrooms, 248 students): Computer Use Prohibited.
- Outcome: final exam multiple-choice/short-answer scores.

Randomized Experiment

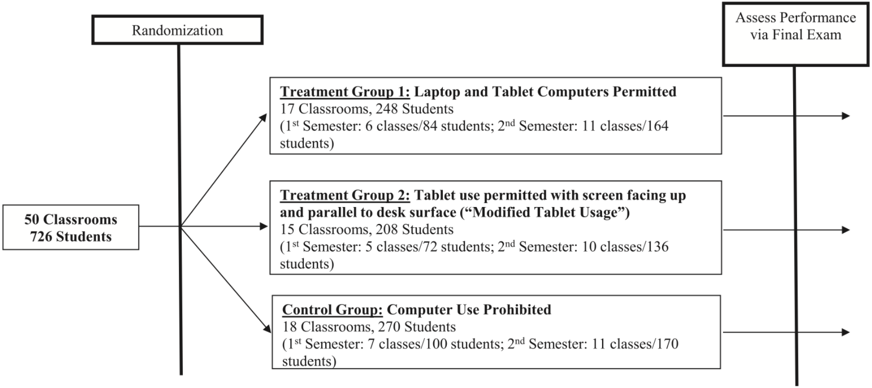


Fig. 1. Experimental design.

Results

Y_i = standardized final exam multiple choice score.

$$X_i = \begin{cases} 0 & \text{if Computers prohibited (Control)} \\ 1 & \text{if Computers permitted (Treatment 1)} \end{cases}$$

$$Y_i = \underset{(0.10)}{-0.25} \times X_i$$

- Moving from computer-permitted class to computer-prohibited class would improve multiple-choice score by 0.25 standard deviations.
- Students who are at 50 percentile would be at 61 percentile.

Results

- The results are robust with respect to adding other controls such as student's demographics (race, gender, age etc.), GPA, ACT scores, instructor fixed effects.
- They also examine the effect on final exam short answer score and found similar results.
- For the effect of "Tablet Permitted with Face-up" (Treatment 2), they found the smaller effect but the effect of using tablet with face up is significantly negative for short answer score.