

Lecture 4

Chapters 1.1-1.4

1.1 The Simple Linear Model

In the stats review, we learned about certain statistics that tell us something about the relationship between two variables:

1. Covariance
2. Correlation

In this chapter will focus on how one variable affects the other through regression analysis.

The simple linear regression model – we hypothesize that X determines Y through some model:

$$Y_i = \beta_1 + \beta_2 X_i + u_i$$

Y-

X-

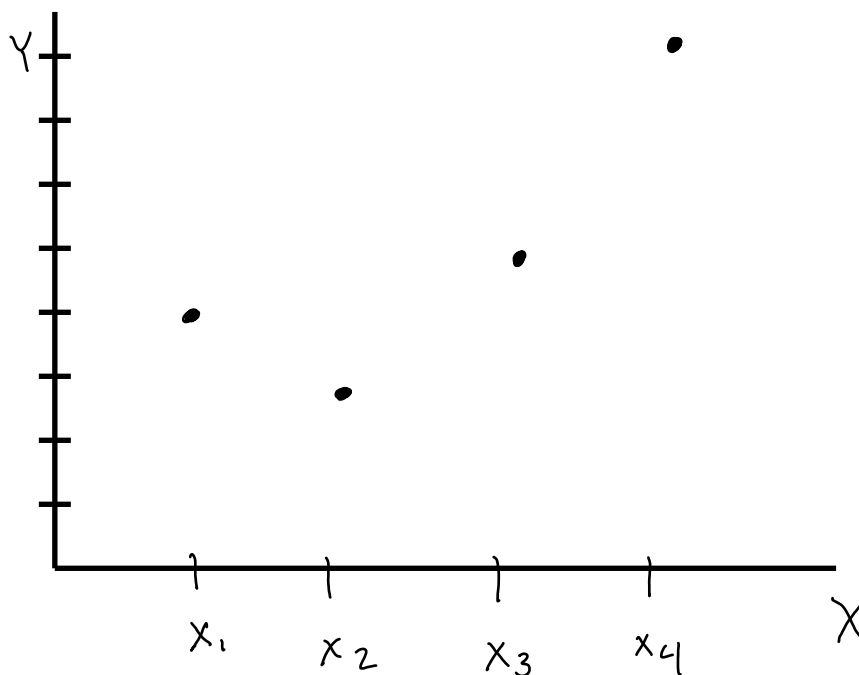
β_1 -

β_2 -

u -

i -

Suppose we have the four points graphed below. We want to find the line that best fits the points. Because of the disturbance term, the points will not fall exactly on the line.



1.2 Least Squares Regression

The line that goes through the data is called the fitted regression line:

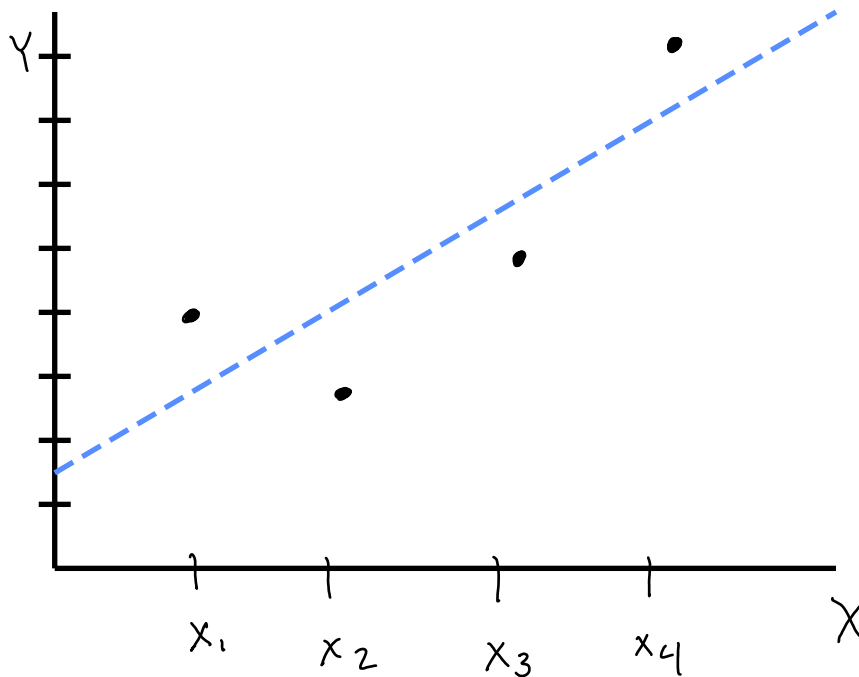
$$\hat{Y}_i = \hat{\beta}_1 + \hat{\beta}_2 X_i$$

\hat{Y}_i –

$\hat{\beta}_1$ –

$\hat{\beta}_2$ –

Residual – The difference between the actual value of Y and the fitted value of Y.



Exercise 1: From the definition of residuals, write an equation where residual is denoted as e_i . Simplify the equation so that it does not contain \hat{Y}_i .

Formula – residual sum of squares, RSS

$$RSS = e_1^2 + e_2^2 + \dots + e_n^2 = \sum_{i=1}^n e_i^2$$

Least Squares Criterion – estimating β_1 and β_2 by minimizing the RSS.

Ordinary least squares estimator – OLS, the estimator that is based on the least squares criterion.

1.3 Derivation of the regression coefficients

Exercise 2: Using the points below and the definitions of residuals and least squares criterion, find the estimates for $\widehat{\beta}_1$ and $\widehat{\beta}_2$.

X	Y
1	4
2	3
3	5
4	8

Exercise 3: Using the general form of Y , find the normal equations for solving for β_1 and β_2 .

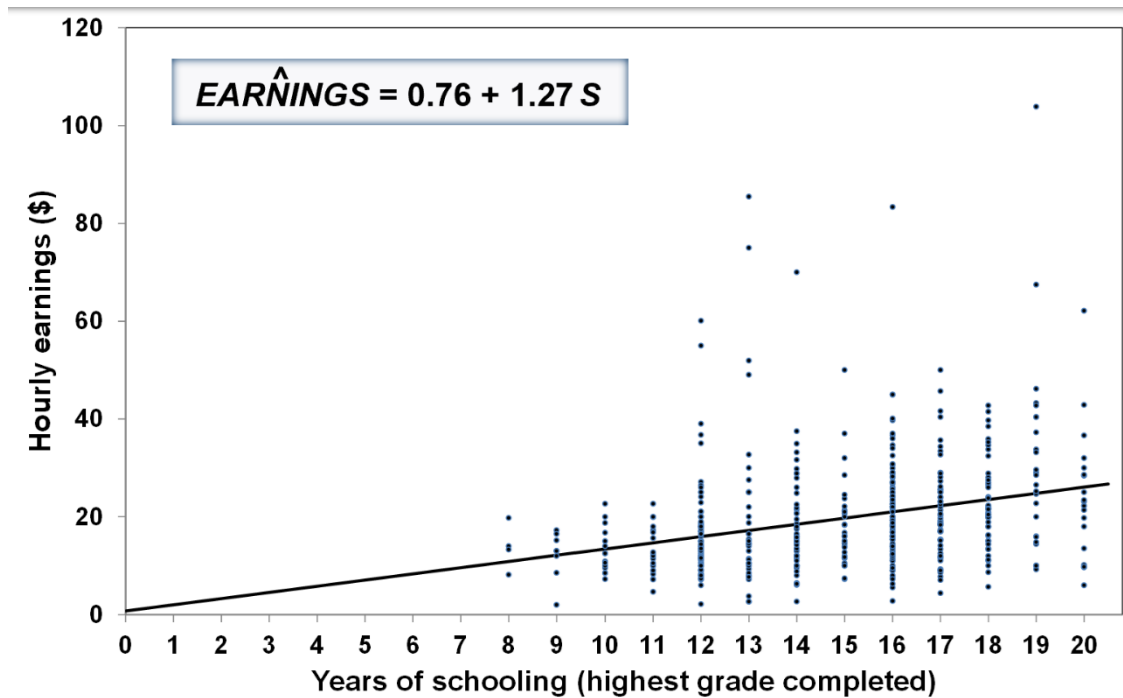
Exercise 4: Use the normal equations we just found and re-solve for the points from exercise 2.

Exercise 5: Solve for the regression coefficients for the following points.

X	Y
1	7
2	8
3	5
4	4

1.4 Interpretation of Regression Coefficients

Suppose we have data on schooling and earnings. We run the regression $Earnings_i = \beta_1 + \beta_2 School_i + u_i$. The output we got is shown below.



Exercise 6: Interpret the β_2 coefficient.

Exercise 7: Interpret the β_1 coefficient. Does this make sense?

Exercise 8: How much would you expect someone to earn if they went to school for 10 years?

Exercise 9: How much would you expect someone to earn if they went to school for 15 years?

Exercise 10: How much would you expect someone's earnings to change by if they went back to school for 2 years?

Exercise 11: Suppose you ran a regression of SAT scores on college GPA and found the fitted model:

$$GPA_i = 2.0 + 0.01SAT_i$$

Interpret the β_2 coefficient.