Chapter 3

Simple Linear Regression

Flow of study in this chapter

› Population Regression Function

We first try to understand the meaning of demand, supply and equilibrium. How consumers and producers react in a market and how price can be a signal for both parties.

› Sample Regression Function

Commodities and services can be differently elastic. The implication on many studies forward will also be varied by their elasticity.

› Estimation

How to define what people gain from trade and lay out a framework to study who gains or loses when there is a change in a market.

› Assumptions underlying Classical Linear Regression Model (CLRM)

Learn how a political, economic institution can intervene price in a market, what is the implication and results for those actions.

Further reading can be found in Pindyck and Rubinfeld (2018) Part 2, Chapter 3-4.

(1) Creating a population regression function

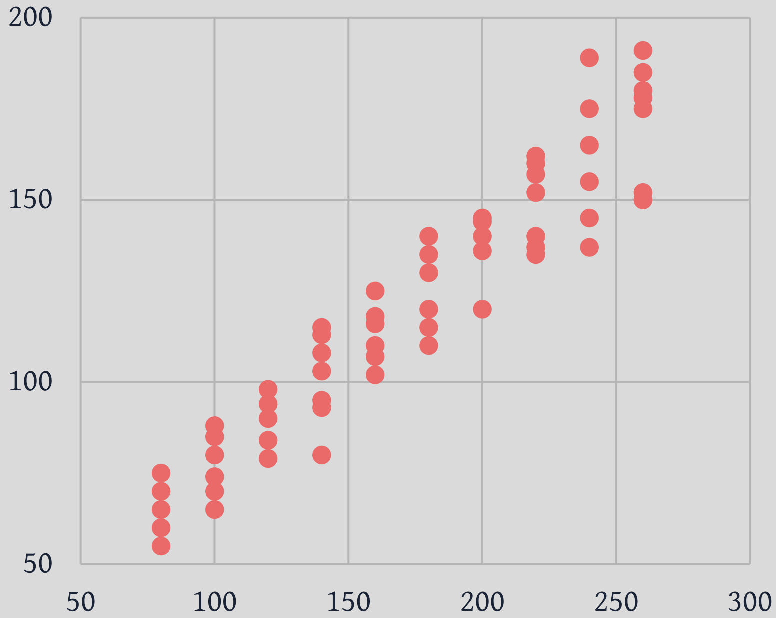
First, let's look at a set of cross-sectional data of household income and consumption. Given that they are both measured in $US.

› X is weekly income

› Y_i is weekly expenditure

X	Y_1	Y_2	Y_3	Y_4	Y_5	Y_6	Y_7
80	55	60	65	70	75	.	.
100	65	70	74	80	85	88	.
120	79	84	90	94	98	.	.
140	80	93	95	103	108	113	115
160	102	107	110	116	118	125	.
180	110	115	120	130	135	140	.
200	120	136	140	144	145	.	.
220	135	137	140	152	157	160	162
240	137	145	155	165	175	189	.
260	150	152	175	178	180	185	191

(1) Creating a regression function

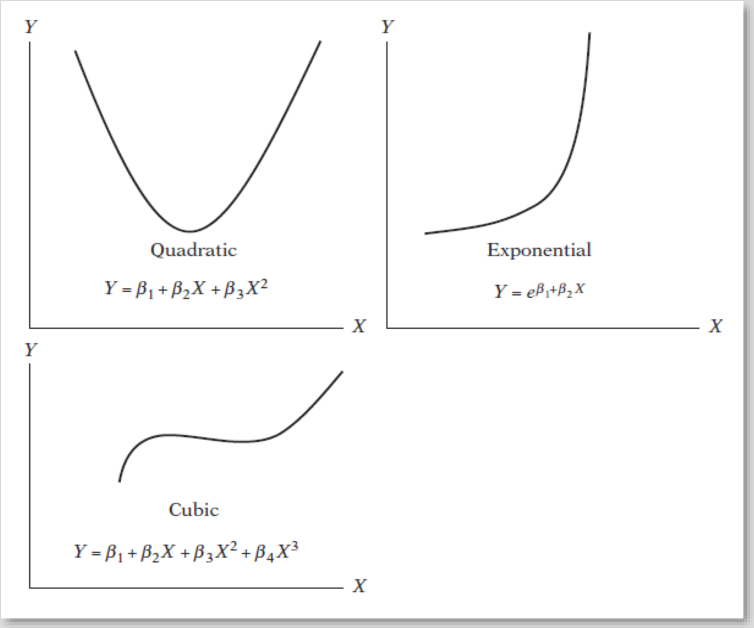


Since we are estimating linear relationship between X and Y, we define a linear **population regression function** (PRF) as

› $E(Y|X_i) = f(X_i) = \beta_1 + \beta_2 X_i$

So, the β_1 and β_2 are the **parameter** or

(2) Notes on linearity

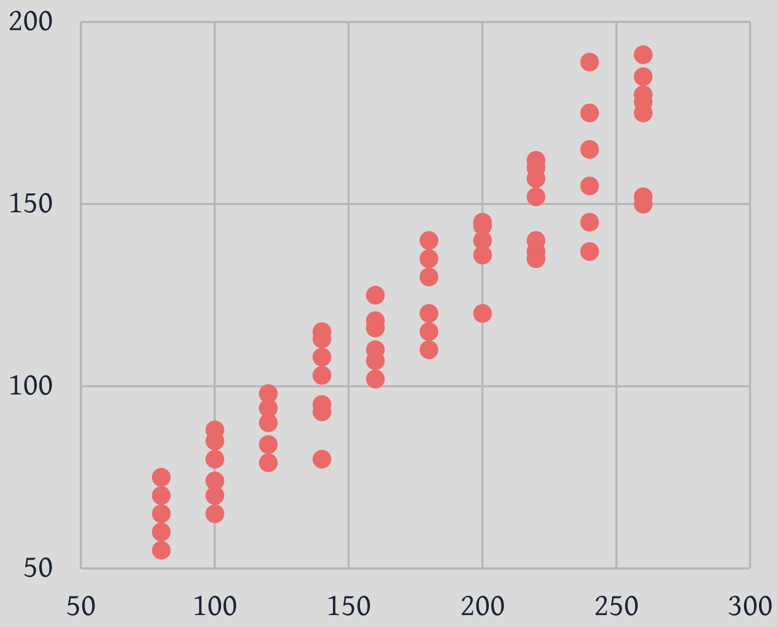


To clear things up, when we mention a linear regression model (**LRM**), we need to specify what are we talking about. There are two types of linearity.

› **Linear in variables**
geometrically, is the function linear or not, considered from the power of X_i.

› **Linear in parameters**
is the function consist of linear parameter or not, considered from the power of β_i.

(3) Stochastic specification



As we know that statistics relation is different from mathematical relation, therefore, we introduce a concept of the **stochastic disturbance** or **stochastic error term** as

› $u_i = Y_i - E(Y|X_i)$ or

› $Y_i = E(Y|X_i) + u_i$

Hence, the stochastic PRF is

› $Y_i = \beta_1 + \beta_2 X_i + u_i$

Now we have two parts of the PRF which are

› **Systematic** or **deterministic** part – which is $\beta_1 + \beta_2 X_i$.

› **Random** or **nonsystematic** part – which is u_i.

(3) Stochastic specification

For instance, let's assume that $\beta_1 = 40$ and $\beta_2 = 0.5$, figure out u_i for the following $E(Y|X_i = 180)$

› $Y_1 = 110 = 40 + 0.5(180) + u_1$ then $u_1 =$

› $Y_3 = 120 = 40 + 0.5(180) + u_3$ then $u_3 =$

› $Y_5 = 135 = 40 + 0.5(180) + u_5$ then $u_5 =$

(3) Stochastic specification

Why do we always have an error term in our equation? Here are some explanations:

› Vagueness of theory

› Unavailability of data
For example, using family wealth to explain consumption behavior but wealth data are usually not available.

› Core variables and peripheral variables
Variable(s) included in our model might be just peripheral ones. Picking a core variable may contribute to our model more.

› Intrinsic randomness in human behavior

(3) Stochastic specification

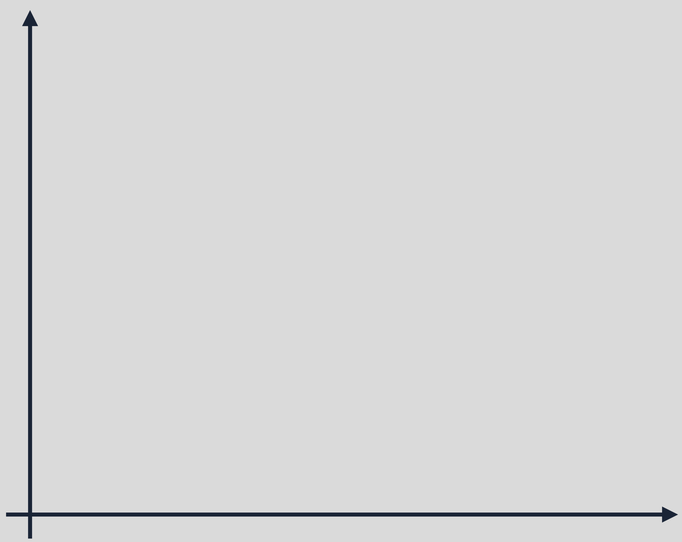
› Poor proxy variables

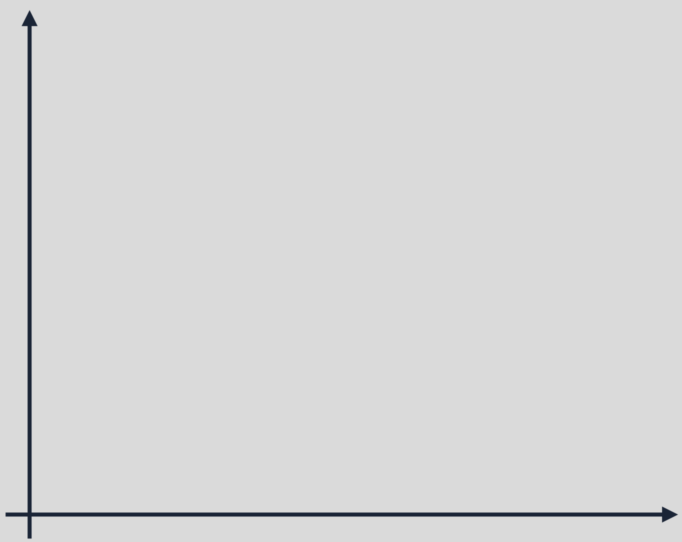
Some intrinsic variable cannot be observed, such as intelligence and skills. Most of the time we rely on another variable as a proxy, such as test score, GPAX, work experience, etc.

› Principle of parsimony

It is what it is if there is no strong theory suggesting adding more variable, keep our model as simple as possible and let the error terms be as they are.

› Wrong functional form





(4) Expected value of the error term

With the error term included in the PRF, taking the expected through this equation

› $Y_i = E(Y|X_i) + u_i$

›

Since the $E(Y|X_i)$ is a constant, therefore

›

So $E(u_i|X_i) = 0$, or we can say that

› $E(u_i|X_i) = \sum_{i=1}^{n} \left(\frac{u_i|X_i}{n}\right) = 0$

(4) Expected value of the error term

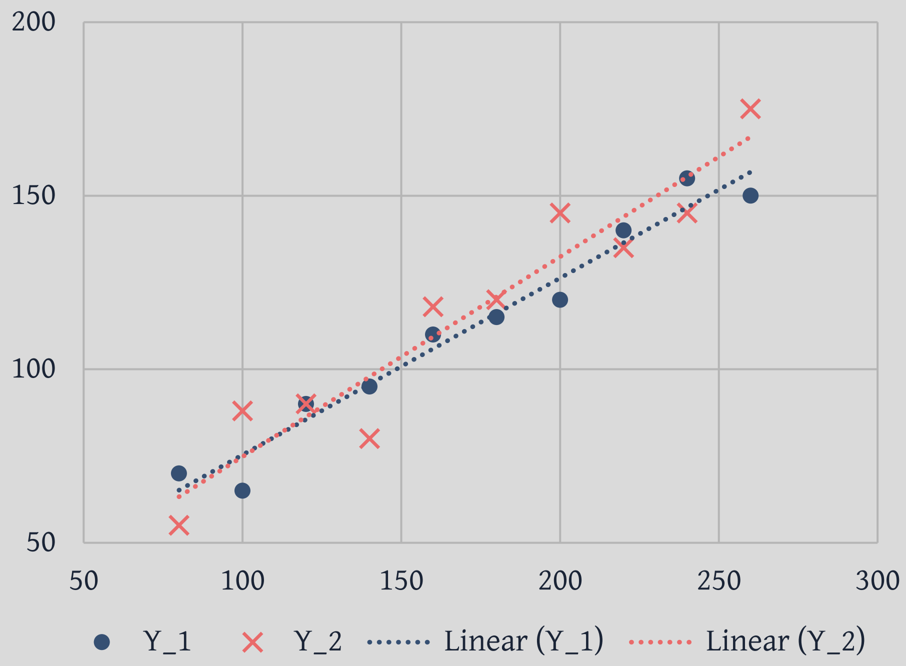
Similar to the property of sum of deviation from the mean, we can see the proof here that it is always zero. For instance,

- $\sum_{i=1}^{n}(x_i - \bar{X}) =$

(1) Creating sample regression functions

In real world scenario, most of the time we cannot collect all the population data. Assumed that we can only collect two sets of data shown in the table.

X	Y_1	Y_2
80	70	55
100	65	88
120	90	90
140	95	80
160	110	118
180	115	120
200	120	145
220	140	135
240	155	145
260	150	175



(1) Creating sample regression functions

Anyway, we still can define our **sample regression function** (SRF) as

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

The stochastic form as

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

where $\hat{Y}_i$ is the estimator of $E(Y|X_i)$

$\hat{\beta}_i$ is the estimator of β_i

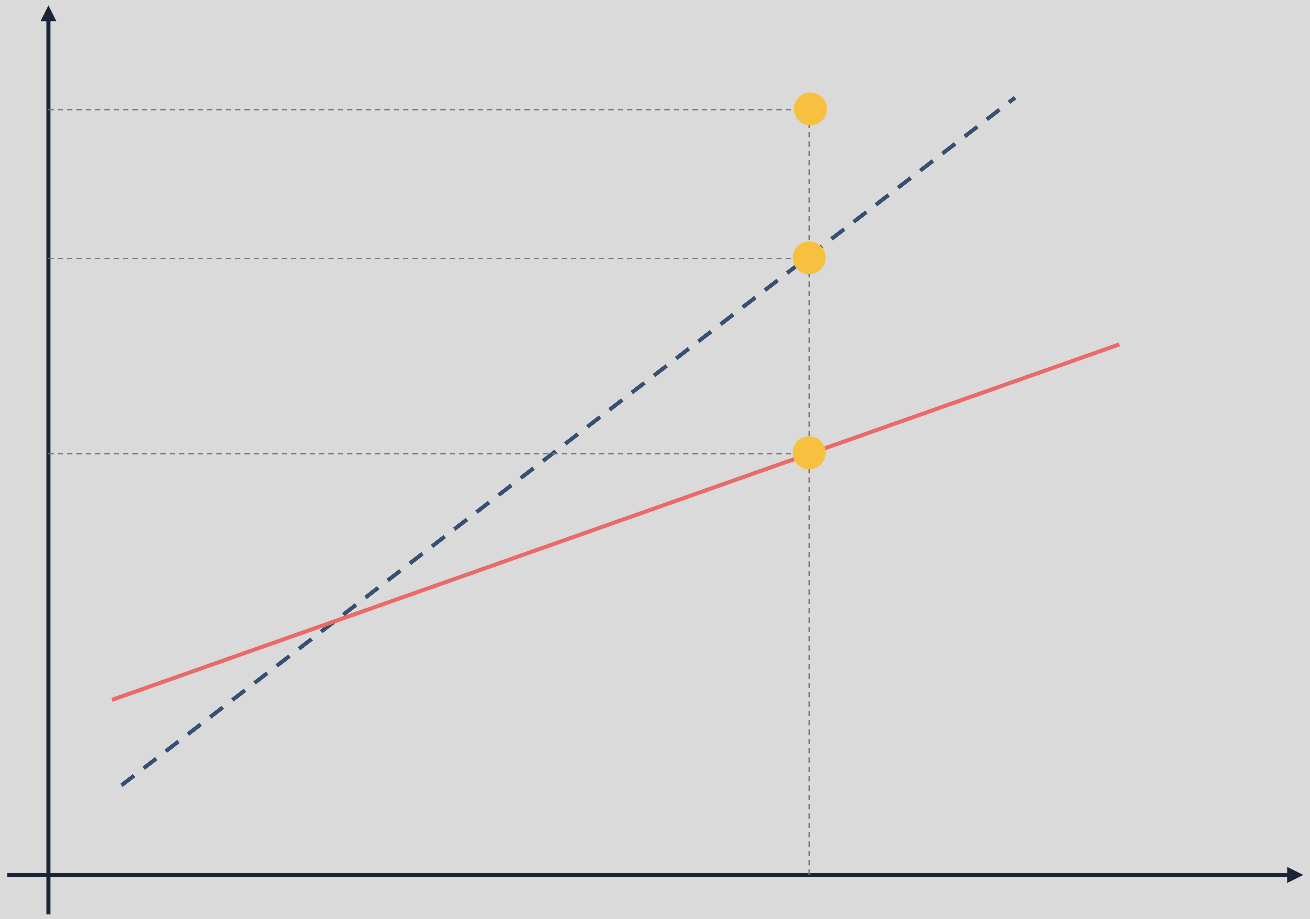
$\hat{u}_i$ is the estimator of u_i

To avoid confusion, since we are referring to β_i quite often

› β_1 and β_2 are called **parameters**.

› $\hat{\beta}_1$ and $\hat{\beta}_2$ are called **estimators**.

(2) Comparing PRF and SRF



Problem statement 1

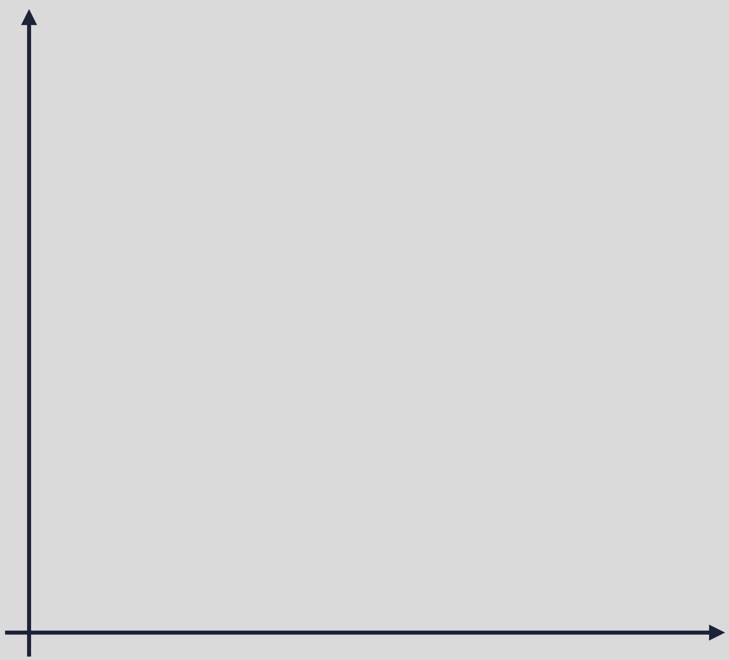
When we obtain a set of data, plot them on a graph, how can we draw a straight line, portraying regression relationship between two variables?

› How to connect each data point with a line that fits best with our data?

› What is/are (a) criteria (ion) that we can rely on?

(1) Ordinary Least Square

This method applies for both PRF and SRF. The intuition is shown here.



Now we try to draw a linear line that minimize sum of the error terms. From

› $Y_i = \hat{\beta}_1 + \hat{\beta}_2 X_i + \hat{u}_i$

Rearranging the equation, we get

›

Setting up the objective function

›

(1) Ordinary Least Square

Solve for $\hat{\beta}_1$.

(1) Ordinary Least Square

Plug in $\hat{\beta}_1$ to solve for $\hat{\beta}_2$.

(1) Ordinary Least Square

$\sum X_i(Y_i - \bar{Y}) = \sum(X_i - \bar{X})(Y_i - \bar{Y})$ because

Eventually, we get

- $\hat{\beta}_1 = \bar{Y} - \hat{\beta}_2\bar{X}$
- $\hat{\beta}_2 = \frac{\sum(X_i - \bar{X})(Y_i - \bar{Y})}{\sum(X_i - \bar{X})^2} = \frac{\sum x_i y_i}{\sum x_i^2}$

where $x_i = (X_i - \bar{X}), y_i = (Y_i - \bar{X}), x_i^2 = (X_i - \bar{X})^2$

(2) Properties of OLS estimators

(1) The OLS estimators are expressed solely in terms of the observables.

(2) They are **point estimators**, instead of interval estimators.

(3) They make the SRF passes through the sample mean.

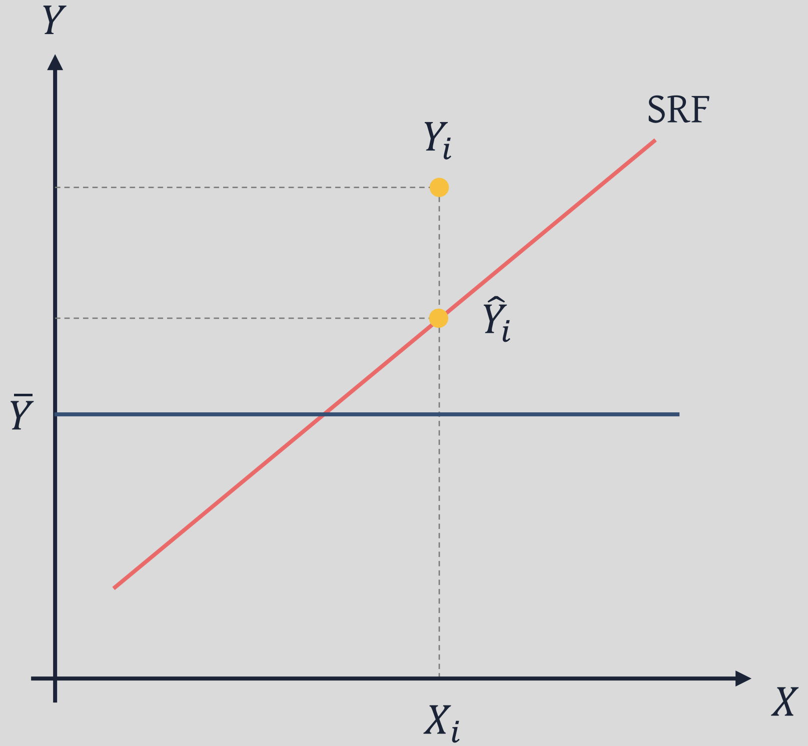
(4) The mean value of $\hat{Y}_i$or $\bar{\hat{Y}} = \bar{Y}$.

(5) The mean value of the residual $\hat{u}_i = 0$.

(6) $\hat{u}_i$ are uncorrelated with both X and $\hat{Y}$.

(3) Coefficient of determination (r^2)

The r^2 is determined by how much the is described by the SRF, or the measurement of **'goodness of fit'** of the fitted regression line comparing to an estimator, $\bar{Y}$.



The intuition is that total sum of squares (TSS) is equal to explained sum of squares (ESS) and residual sum of squares (RSS) or

›

(3) Coefficient of determination (r^2)

Eventually, we get

› $r^2 = \frac{ESS}{TSS} = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2}$ or

› $r^2 = 1 - \frac{RSS}{TSS} = 1 - \frac{\sum \hat{u}_i^2}{\sum(Y_i - \bar{Y})^2}$

There are a few more formulae of r^2 in page 76

Properties of r^2

(1) Non-negativity

(2) $0 \leq r^2 \leq 1$

(4) Sample coefficient of correlation (r)

A formula of r^2 is

› $r^2 = \frac{(\sum x_i y_i)^2}{\sum x_i^2 \sum y_i^2}$

We can define **sample coefficient of correlation** (r) easily by

› $r = \pm\sqrt{r^2} = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2)(\sum y_i^2)}}$

Properties of r

(1) Positive or negative depending on the sign of the term.

(2) $-1 \leq r \leq 1$

(3) Independent of the origin and scale.

(4) If X and Y are statistically independent, $r = 0$, but **not** vice versa.

(5) Does not describe non-linear association or causality.