Introductory Econometrics
Ch2 The Simple Regression Model: Interpretation and Estimation

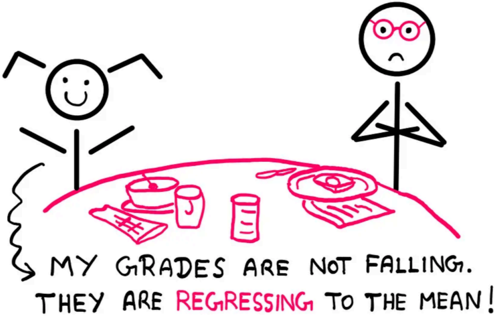
LIU Chenyuan

Spring 2024

Outline

A Little History about "Regression"

- Francis Galton: tall parents tend to have children shorter than them, and short parents often have children who were taller than them.
- He called this "regression to the mean".



Regression Model

- We will derive the linear regression model in three ways.
  1. Descriptive
  2. Causal
  3. Forecasting
- Mathematically they are identical, but conceptually they are very different.

Outline

Descriptive Analysis

Causal Estimation

Forecasting

Conditional Expectations

- Let X and Y be two random variables, e.g. gender and salary.
- In economics, we often are interested in describing the value of Y if X is of some value. Say if I could gather everyone in the world for whom X is some particular value, what would be the expected value of Y?
- We define conditional expectation

$$E(Y|X)$$

  to mean: if I "condition" X to be some value, what is the expected value of Y?

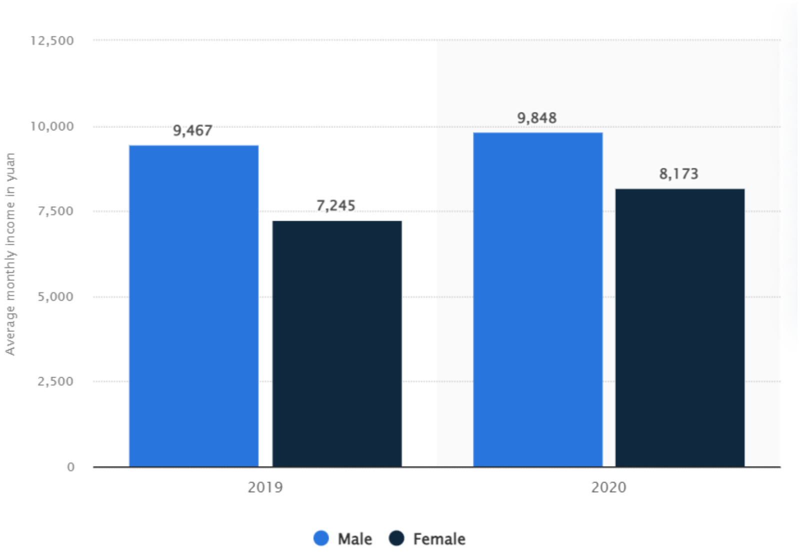
Example: Salary and Gender

- We are interested in how salary varies with gender. Note: this is purely descriptive. We don't say anything about causation.
- Let X denote gender, then we are interested in

$$E(salary|male)$$
$$E(salary|female)$$

- How do we estimate this?
- Recall: to estimate an expectation, we use the mean.
- So we take the mean value of salary for men and the mean value of salary for women.

The 2019 and 2020 Data



Conditional Expectations

- We define conditional expectation

$$E(Y|X)$$

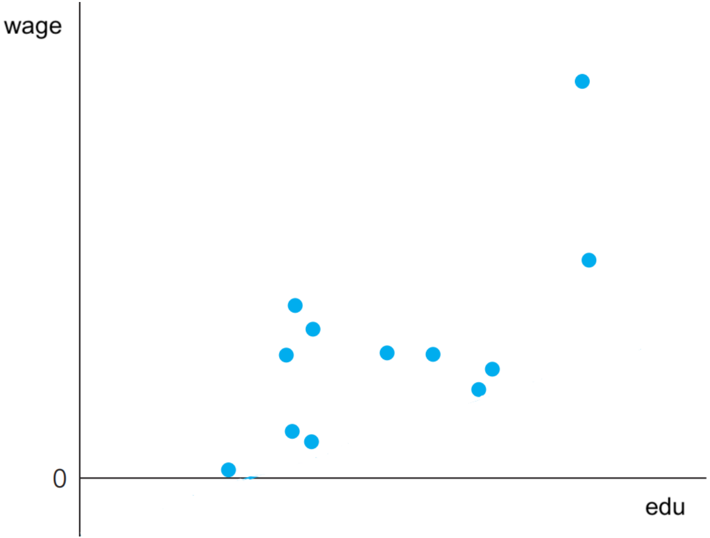
  to mean: if I "condition" X to be some value, what is the expected value of Y?
- Y is a random variable so after choosing X we don't know exactly what Y will be.
- $E(Y|X)$ depends on X, so changing X will change $E(Y|X)$.

What if X is continuous?

- A continuous random variable takes on a continuum of values.
- Often, a variable that can take on many values is treated as continuous because it is convenient.
- Example: education, wage
- How does wage vary with education?

Wage and Education

How do we estimate $E(wage|edu)$?



A Simple Linear Model

- We need a model for $E(y|x)$ when x is continuous.
- Consider a very simple linear model:

$$E(y|x) = \beta_0 + \beta_1 x.$$

A Simple Linear Model

$$E(y|x) = \beta_0 + \beta_1 x.$$

To interpret:

$$\beta_0 = E(y|x = 0),$$
$$\beta_1 = \frac{\partial E(y|x)}{\partial x}.$$

- β_0 is the expected value of y when $x = 0$.
- β_1 is the change in the expected value of y when x increases by 1 unit.

Estimation

How do we estimate it?

- It will be useful to **define**

$$\begin{aligned} u &\equiv y - E(y|x) \\ &= y - \beta_0 - \beta_1 x. \end{aligned}$$

- In other words,

$$y = \beta_0 + \beta_1 x + u.$$

- The fact that $E(y|x) = \beta_0 + \beta_1 x$ implies that

$$\begin{aligned} E(u|x) &= E(y - \beta_0 - \beta_1 x|x) \\ &= E(y|x) - \beta_0 - \beta_1 x \\ &= \beta_0 + \beta_1 x - \beta_0 - \beta_1 x \\ &= 0. \end{aligned}$$

Review: Law of Iterated Expectation

$$E(Y) = E[E(Y|X)].$$

How to understand? Example:

- There are two firms producing the product, $X = 1$ and $X = 2$.
- Let Y denote the product quality. $E(Y|X = 1) = 1$ and $E(Y|X = 2) = 3$.
- Both firms produce an equal amount of products.
- What is $E(Y)$ in the market?

$$\begin{aligned} E(Y) &= E(Y|X = 1)P(X = 1) + E(Y|X = 2)P(X = 2) \\ &= 1 \times 0.5 + 3 \times 0.5 = 2. \end{aligned}$$

- The formal proof is on the Web-Learning page.

Two Conditions of u

From $E(u|x) = 0$, we could derive two conditions about u

1. $E(u) = 0$,
2. $E(ux) = 0$.

Why? Using the law of iterated expectation,
$E(Y) = E[E(Y|X)]$.

$$\begin{aligned} E(u) &= E(E(u|x)) = 0 \\ E(ux) &= E(E(ux|x)) = E(xE(u|x)) \\ &= E(x0) = 0. \end{aligned}$$

Note that in $E(ux|x) = xE(u|x)$ we could treat x as a constant and take it outside of the expectation because the expectation is conditional on x.

The reason why we derive the two conditions,

$$E(u) = 0$$
$$E(ux) = 0.$$

is to use them to estimate β_0 and β_1.

- To estimate an expectation we use a sample mean.
- By the same logic, we could replace the expectations above with sample mean expressions.

- Assume that we observe a random sample with size N
- Let $i = 1, ..., N$ index the unit of observation.
- We write the population regression model as

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

where y_i is the value of y for individual i

- Define the sample regression model

$$y_i = \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{u}_i,$$

where

  - $\hat{\beta}_0$ and $\hat{\beta}_1$ are estimators of β_0 and β_1
  - $\hat{u}_i$ is defined as

$$\hat{u}_i \equiv y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i.$$

The idea of estimation:

- If the sample looks like the population, then we force things to be true in the sample that we know would be true in the population.
- This is the idea of a **sample analogue**.
- In this case, we think $\hat{u}_i$ in the sample should be like u_i in the population.

Method of Moments

- Method of moments: use the sample average to estimate the population expectation
- Use $\frac{1}{N}\sum$ to replace $E[\cdot]$

Population expectations	Sample analogue
$E(u) = 0$	$\frac{1}{N}\sum_{i=1}^{N} \hat{u}_i = 0$
$E(xu) = 0$	$\frac{1}{N}\sum_{i=1}^{N} x_i \hat{u}_i = 0$

We have two unknowns, $\hat{\beta}_0$ and $\hat{\beta}_1$. We have two equations:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0,$$

$$\frac{1}{N}\sum_{i=1}^{N}x_i\hat{u}_i = \frac{1}{N}\sum_{i=1}^{N}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$$

We can solve for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Review: the Summation Operator[1]

$$\sum_{i=1}^{N} x_i = x_1 + x_2 + x_3 + ... + x_N.$$

$$\sum_{i=1}^{N} cx_i = c\sum_{i=1}^{N} x_i.$$

$$\sum_{i=1}^{N}(ax_i + by_i) = a\sum_{i=1}^{N} x_i + b\sum_{i=1}^{N} y_i.$$

$$\sum_{i=1}^{N}(x_i - \bar{x}) = \sum_{i=1}^{N} x_i - N\bar{x} = \sum_{i=1}^{N} x_i - N(\frac{1}{N}\sum_{i=1}^{N} x_i) = 0.$$

$$\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y}) = \sum_{i=1}^{N}(x_i - \bar{x})y_i = \sum_{i=1}^{N}(x_i y_i - \bar{x}\bar{y}).$$

[1] We use $\bar{x}$ to denote the average of x.

Estimation of $\hat{\beta}_0$ and $\hat{\beta}_1$

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} \hat{u}_i \\
&= \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= \frac{1}{N}\sum_{i=1}^{N} y_i - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_0 - \frac{1}{N}\sum_{i=1}^{N}\hat{\beta}_1 x_i \\
&= \bar{y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{x}.
\end{aligned}
$$

Equivalently,

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

$$
\begin{aligned}
0 &= \frac{1}{N}\sum_{i=1}^{N} x_i \hat{u}_i \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - \hat{\beta}_0 \frac{1}{N}\sum_{i=1}^{N} x_i - \hat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} x_i y_i - (\overline{y} - \hat{\beta}_1 \overline{x})\overline{x} - \hat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} x_i^2 \\
&= \frac{1}{N}\sum_{i=1}^{N} (x_i y_i - \overline{x}\overline{y}) - \hat{\beta}_1 \frac{1}{N}\sum_{i=1}^{N} (x_i^2 - \overline{x}^2)
\end{aligned}
$$

After rearrangement:

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2},$$
$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

- The denominator of $\hat{\beta}_1$ is the sample variance of x, while the numerator is the sample covariance of x and y (we can divide $\frac{1}{N-1}$ on both the denominator and the numerator.)
- The regression coefficient, $\hat{\beta}_1$, and the covariance of x and y have the same sign.

Example: Wage and Education

- Suppose we have education and wage data for 400 random individuals. We estimate that

$$E(wage|edu) = -0.9 + 3 \times edu.$$

- People with more years of education have higher expected wages. It could mean any of the following:

1. $edu \rightarrow wage$: I earn a higher wage because I have more years of education.
2. $z \rightarrow edu$, $z \rightarrow wage$: I live in areas where it is easier to attend college and the wage rate is higher.
3. $wage \rightarrow edu$: Because my earning is high, I can "consume" more education.

Example: CEO Earning

- Consider the relation between CEO salary and the return on equity (ROE).
- Suppose we have data on CEO salary and firms' ROE for 50 firms.
- We estimate that

$$E(salary|ROE) = 963 + 18 \times ROE.$$

- It could mean any of the following:

1. $ROE \rightarrow salary$: When the stock price does well, CEOs get a raise.
2. $z \rightarrow salary$, $z \rightarrow ROE$: The economic condition is good, causing both a rise in stock price and CEO's salary.
3. $salary \rightarrow ROE$: By paying the CEO more, they work harder and push up the stock price.

Summary of “Descriptive” Regression

- In the descriptive analysis, the regression model depicts the expectation of y conditional on x.
- The model shows how the two variables are correlated.
- We can always run a regression on two variables. It does not mean they have a causal relationship.
- To give it a causal meaning, we need more assumptions.

Outline

Descriptive Analysis

Causal Estimation

Forecasting

Causal Estimation

To say something about causality, we want a model about how the data are generated. Conceptually,

- For the descriptive analyses, we start with the data and ask which model could help summarize it.
- For the causal analyses, we start with the model and then use it to say what the data will look like.

Simple Regression Model: A Second View

Consider the same simple regression model:

$$y = \beta_0 + \beta_1 x + u.$$

1. β_0 and β_1 are unknown numbers in the nature we want to uncover
2. You choose x
3. Nature chooses u in a way that is unrelated to your choice of x

This is now a causal model.

Example: Education and Wage

We want to know if I have one more year of education, how will that increase my wage?

- y is wage.
- x is education.
- Suppose that in reality, the wage is determined by the following model:

$$y = \beta_0 + \beta_1 x + u.$$

- What is u? Things affect wages other than education.
  - Experience
  - Efforts
  - Family connections
  - Gender
  - Age
  - ...

Estimation

- u represent things affect y other than x.
- Before, u was defined as $y - E(y|x)$, it didn't actually mean anything.
- Now we think of u as some real thing. It's just we can't observe it.
- To estimate the model, we need to know how u is determined.
- The simplest case is that u is assigned at random.
- We can write this as

$$E(u|x) = 0.$$

Zero Conditional Mean

$E(u|x) = 0$ means two things:

1. $E(u|x)$ does not vary with x
2. $E(u|x)$ has a value of 0.

The first is essential, but the second isn't.

Normalization

Why? Suppose we instead assume that

$$E(u|x) = 1.$$

with the model

$$y = \beta_0 + \beta_1 x + u.$$

Then the model is equivalent to another model

$$y = \gamma_0 + \gamma_1 x + \epsilon,$$

where

$$\begin{aligned}
\gamma_0 &= \beta_0 + 1 \\
\gamma_1 &= \beta_1 \\
\epsilon &= u - 1.
\end{aligned}$$

Normalization

$$
\begin{aligned}
E(\epsilon|x) &= E(u-1|x) \\
&= E(u|x) - 1 \\
&= 0.
\end{aligned}
$$

- β_1 and γ_1 are the same in both models, and that's what we care about.
- Thus the fact that we pick 0 is a **normalization**. It makes the model well-defined without posing extra constraints to the data.

Zero Conditional Mean

Now let's look at the first implication of $E(u|x) = 0$. That is, $E(u|x)$ does not vary with x. Take the model:

$$y = \beta_0 + \beta_1 x + u.$$

Fix x and take expectation:

$$E[y|x] = \beta_0 + \beta_1 x + E[u|x].$$

Take derivatives with regard to x

$$\frac{\partial E[y|x]}{\partial x} = \beta_1 + \frac{\partial E[u|x]}{\partial x}.$$

Zero Conditional Mean

$$\frac{\partial E[y|x]}{\partial x} = \beta_1 + \frac{\partial E[u|x]}{\partial x}.$$

- If and only if $E[u|x]$ is a constant, β_1 means "when x changes by one unit, the change in y on average".
- $\frac{\partial E[u|x]}{\partial x} = 0$: when x changes, the other factors affecting y stay the same (all else equal).
- This condition gives the model a causal interpretation: if $E(u|x)$ does not vary when x changes, then any change in y can be attributed to x
- So β_1 reflects the causal effects of x on y.

Estimation

- The estimation is the same as before.
- The population regression model is:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

From $E(u|x) = 0$, we know that

$$E(u_i) = 0$$
$$E(u_i x_i) = 0.$$

Estimation

Take the sample analogues:

$$\frac{1}{N}\sum_{i=1}^{N}\hat{u}_i = \frac{1}{N}\sum_{i=1}^{N}(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

$$\frac{1}{N}\sum_{i=1}^{N}x_i\hat{u}_i = \frac{1}{N}\sum_{i=1}^{N}x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0$$

And the solutions are:

$$\hat{\beta}_1 = \frac{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\frac{1}{N}\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1\bar{x}.$$

- In either the descriptive or the causal case, we get the same estimates.
- The only difference is the interpretation.
- Whether $E(u|x) = 0$ holds depends on contexts.

Example: Education and Wage

Suppose we estimate that

$$wage_i = -0.9 + 3edu_i + \hat{u}_i.$$

- Assume that $E(u|x) = 0$ holds.
- This means for every extra year of education I get, my wage increases by about 3 yuan an hour because of that.
- Thus four years of education is worth 12 yuan per hour.
- If I work 996, then this means 50 weeks $\times$ 12 hours per day $\times$ 6 days per week $= 23,328$ yuan a year.

Outline

Descriptive Analysis

Causal Estimation

Forecasting

Forecasting

- We have data on x_i and y_i
- We know the value of x, and want to predict what the value of y will be tomorrow.

Examples

- Inflation this year, unemployment next year
- Corporate profits today, stock price tomorrow
- My studying hours today, my GPA tomorrow

Forecasting

Use the linear regression model:

$$y_i = \beta_0 + \beta_1 x_i + u_i,$$

and define the predicted value as

$$\hat{y}_i = \hat{\beta}_0 + \hat{\beta}_1 x_i.$$

If we know $x_i = x^*$, we want to predict $\hat{y}^* = \hat{\beta}_0 + \hat{\beta}_1 x^*$

Difference between Forecasting and Causal Estimation

This is different from the causal model because we do not choose x^*. Compare the following:

- If I change the inflation rate to x, what will happen to the unemployment rate next year?
- I observe the inflation rate is x. Given that, what is my best guess about the unemployment rate next year?

Estimation

- Intuitively, I want the predicted value $\hat{y}_i$ to be close to y_i as much as possible.
- We need a measure on the difference between y_i and $\hat{y}_i$
- We pick the following function:

$$(y_i - \hat{y}_i)^2.$$

Estimation: Ordinary Least Squares

Calculate the distance for all data points:

$$H \equiv \sum_{i=1}^{N} (y_i - \hat{y}_i)^2 = \sum_{i=1}^{N} (y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i)^2.$$

We choose $\hat{\beta}_0$ and $\hat{\beta}_1$ to minimize the function H.

The first order conditions:

$$\frac{\partial H}{\partial \hat{\beta}_0} = -\sum_{i=1}^{N} 2(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0,$$
$$\frac{\partial H}{\partial \hat{\beta}_1} = -\sum_{i=1}^{N} 2x_i(y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i) = 0.$$

- Note that the first-order conditions are the same as the moment conditions.
- So the two methods yield the same estimates.

Ordinary Least Squares (OLS) Estimates

Two motivations for the OLS estimates

- Sample analogue
- Minimize the distance between the data and model (least squares)
- Both yield the same estimates

$$\hat{\beta}_1 = \frac{\sum_{i=1}^{N}(x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^{N}(x_i - \bar{x})^2}$$

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}.$$

Quiz: Health Insurance and Life Expectancy

We collect data on 100 individuals on their health insurance status and their age at death. We estimate that

$$L_i = 65 + 7H_i + \hat{u}_i.$$

- H_i is % of medical expenditure covered by insurance, varying between 0 (no insurance) and 1 (free care)
- L_i is the age at death

Explain the meaning of the slope coefficient, 7, in the three interpretations we discussed.

Summary

$$y = \beta_0 + \beta_1 x + u.$$

- Three interpretations of the model:
    1. Descriptive
    2. Causal
    3. Forecasting
- Two views of the estimates
    - Method of moments
    - Ordinary least squares