Econ 103: Introduction to Simple Linear Regression

Manu Navjeevan

UCLA

July 23, 2021

Content Outline

Table of Contents

Linear Regression as Line of Best Fit

Suppose we have two variables, Y and X. We are interested in using data to learning about the relationship between Y and X.

Linear Regression as Line of Best Fit

Suppose we have two variables, Y and X. We are interested in using data to learning about the relationship between Y and X.

Examples:

- How are education and wages related?

Linear Regression as Line of Best Fit

Suppose we have two variables, Y and X. We are interested in using data to learning about the relationship between Y and X.

Examples:

- How are education and wages related?
- How are unemployment and inflation related?

Linear Regression as Line of Best Fit

Suppose we have two variables, Y and X. We are interested in using data to learning about the relationship between Y and X.

Examples:

- How are education and wages related?
- How are unemployment and inflation related?
- What is the relationship between receiving a treatment and a health outcome?

Linear Regression as Line of Best Fit

One way to model the relationship between Y and X would be to try to find the line of best fit between the two variables.

By the line of best fit we mean finding the line, characterized by a slope and an intercept, that minimizes the distance between Y and $\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X$.

Linear Regression as Line of Best Fit

One way to model the relationship between Y and X would be to try to find the line of best fit between the two variables.

By the line of best fit we mean finding the line, characterized by a slope and an intercept, that minimizes the distance between Y and $\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X$.

Formally, we are interested in the parameters β_0 and β_1 that solve

$$\begin{aligned} \beta_0, \beta_1 &= \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - (\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X)\right)^2\right] \\ &= \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right] \end{aligned}$$

Linear Regression as Line of Best Fit

One way to model the relationship between Y and X would be to try to find the line of best fit between the two variables.

By the line of best fit we mean finding the line, characterized by a slope and an intercept, that minimizes the distance between Y and $\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X$.

Formally, we are interested in the parameters β_0 and β_1 that solve

$$\begin{aligned}
\beta_0, \beta_1 &= \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - (\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X)\right)^2\right] \\
&= \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]
\end{aligned}$$

- By $\arg\min$ we just mean we are interested in the **arguments** β_0 and β_1 that minimize

$$\mathbb{E}[(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X)^2]$$

rather than the value $\mathbb{E}[(Y - \beta_0 - \beta_1 \cdot X)^2]$ itself.

Linear Regression as Line of Best Fit

One way to model the relationship between Y and X would be to try to find the line of best fit between the two variables.

By the line of best fit we mean finding the line, characterized by a slope and an intercept, that minimizes the distance between Y and $\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X$.

Formally, we are interested in the parameters β_0 and β_1 that solve

$$\begin{aligned} \beta_0, \beta_1 &= \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - (\tilde{\beta}_0 + \tilde{\beta}_1 \cdot X)\right)^2\right] \\ &= \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right] \end{aligned}$$

- Another way of saying this is that

$$\mathbb{E}[(Y - \beta_0 - \beta_1 \cdot X)^2] < \mathbb{E}[(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X)^2]$$

for any $(\tilde{\beta}_0, \tilde{\beta}_1) \neq (\beta_0, \beta_1)$.

Linear Regression as Line of Best Fit

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Why do we care about these parameters?

- Knowing the line of best fit will help us predict Y using X
  - Will provide the best linear prediction of Y using X.
  - Even though a linear model may seem to simple, ends up being tremendously useful in practice.

Linear Regression as Line of Best Fit

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Why do we care about these parameters?

- We can also interpret the parameters β_0 and β_1 to learn (to a first order degree) about the relationship between Y and X

Linear Regression as Line of Best Fit

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Why do we care about these parameters?

- We can also interpret the parameters β_0 and β_1 to learn (to a first order degree) about the relationship between Y and X
  - Is there a positive or negative relationship between Y and X? $\iff$ Is β_1 positive or negative?

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Why do we care about these parameters?

- We can also interpret the parameters β_0 and β_1 to learn (to a first order degree) about the relationship between Y and X
  - Is there a positive or negative relationship between Y and X? $\iff$ Is β_1 positive or negative?
  - How much can we expect Y to change if we see an increase in X of one unit? $\iff$ What is β_1?

Linear Regression as Line of Best Fit

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg \min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Why do we care about these parameters?

- We can also interpret the parameters β_0 and β_1 to learn (to a first order degree) about the relationship between Y and X
  - Is there a positive or negative relationship between Y and X? $\iff$ Is β_1 positive or negative?
  - How much can we expect Y to change if we see an increase in X of one unit? $\iff$ What is β_1?
  - What is the average value of Y when X is zero? $\iff$ What is β_0?

Linear Regression as Line of Best Fit

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Why do we care about these parameters?

- We can also interpret the parameters β_0 and β_1 to learn (to a first order degree) about the relationship between Y and X
  - Is there a positive or negative relationship between Y and X? $\iff$ Is β_1 positive or negative?
  - How much can we expect Y to change if we see an increase in X of one unit? $\iff$ What is β_1?
  - What is the average value of Y when X is zero? $\iff$ What is β_0?
  - To a first order degree because β_0 and β_1 describe the line of best fit rather than the "true" relationship.
    - No need to worry about this difference for now though.

Linear Regression: The Parameters

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Linear Regression: The Parameters

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Let's solve for β_0 and β_1 by taking first order conditions:

$$\frac{\partial}{\partial \tilde{\beta}_0} : \mathbb{E}\left[Y - \beta_0 - \beta_1 \cdot X\right] = 0$$

$$\frac{\partial}{\partial \tilde{\beta}_1} : \mathbb{E}\left[(Y - \beta_0 - \beta_1 \cdot X) \cdot X\right] = 0$$

Linear Regression: The Parameters

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Let's solve for β_0 and β_1 by taking first order conditions:

$$\begin{aligned} \frac{\partial}{\partial \tilde{\beta}_0} &: \mathbb{E}\left[Y - \beta_0 - \beta_1 \cdot X\right] = 0 \\ \frac{\partial}{\partial \tilde{\beta}_1} &: \mathbb{E}\left[(Y - \beta_0 - \beta_1 \cdot X) \cdot X\right] = 0 \end{aligned}$$

We will return to these first order conditions shortly. For now, after rearranging we get that

$$\begin{aligned} \beta_1 &= \frac{\mathbb{E}[YX] - \mathbb{E}[Y]\mathbb{E}[X]}{\mathbb{E}[X^2] - \mathbb{E}[X]\mathbb{E}[X]} = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)} \\ \beta_0 &= \mathbb{E}[Y] - \beta_1 \mathbb{E}[X] \end{aligned}$$

Linear Regression: The Parameters

We are interested in the parameters β_0 and β_1 that solve

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Let's solve for β_0 and β_1 by taking first order conditions:

$$\frac{\partial}{\partial \tilde{\beta}_0} : \mathbb{E}\left[Y - \beta_0 - \beta_1 \cdot X\right] = 0$$
$$\frac{\partial}{\partial \tilde{\beta}_1} : \mathbb{E}\left[(Y - \beta_0 - \beta_1 \cdot X) \cdot X\right] = 0$$

We will return to these first order conditions shortly. For now, after rearranging we get that

$$\beta_1 = \frac{\mathbb{E}[YX] - \mathbb{E}[Y]\mathbb{E}[X]}{\mathbb{E}[X^2] - \mathbb{E}[X]\mathbb{E}[X]} = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)}$$
$$\beta_0 = \mathbb{E}[Y] - \beta_1 \mathbb{E}[X]$$

Exercise: Show this rearrangement.

Linear Regression: The Error Term

Let's define the random variable

$$\begin{aligned} \epsilon &= Y - (\beta_0 + \beta_1 \cdot X) \\ &= Y - \beta_0 - \beta_1 \cdot X \end{aligned}$$

We can then write

$$Y = \beta_0 + \beta_1 \cdot X + \epsilon.$$

which is the linear regression equation you may have seen before. The random variable ϵ will be important later on as we try to do inference.

Linear Regression: The Error Term

Let's define the random variable

$$\begin{aligned}\epsilon &= Y - (\beta_0 + \beta_1 \cdot X) \\ &= Y - \beta_0 - \beta_1 \cdot X\end{aligned}$$

We call ϵ the linear regression error variable.

Linear Regression: The Error Term

Let's define the random variable

$$\begin{aligned}\epsilon &= Y - (\beta_0 + \beta_1 \cdot X) \\ &= Y - \beta_0 - \beta_1 \cdot X\end{aligned}$$

We call ϵ the linear regression error variable.

Recall that from the first order conditions for β_0 and β_1 we have that

$$\mathbb{E}\big[\underbrace{Y - \beta_0 - \beta_1 \cdot X}_{=\epsilon}\big] = 0$$

$$\mathbb{E}\big[\underbrace{(Y - \beta_0 - \beta_1 \cdot X)}_{=\epsilon} \cdot X\big] = 0$$

These give us the properties that

$$\mathbb{E}[\epsilon] = 0 \text{ and } \mathbb{E}[\epsilon X] = 0.$$

Linear Regression: Model Summary

In total our line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

generate a model betwen Y and X that can be written as

$$Y = \beta_0 + \beta_1 \cdot X + \epsilon \tag{1}$$

where

$$\mathbb{E}[\epsilon] = 0 \text{ and } \mathbb{E}[\epsilon X] = 0.$$

Linear Regression: Model Summary

In total our line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

generate a model betwen Y and X that can be written as

$$Y = \beta_0 + \beta_1 \cdot X + \epsilon \tag{1}$$

where

$$\mathbb{E}[\epsilon] = 0 \text{ and } \mathbb{E}[\epsilon X] = 0.$$

- It is often convenient to work directly with this representation or make assumptions about ϵ.
- You may have seen this representation before, the prior slides go over where this model comes from

Linear Regression: Model Summary

Our line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

are useful for

- Making predictions about Y using X.
    - Predict Y when $X = x$ with $\beta_0 + \beta_1 \cdot x$

Linear Regression: Model Summary

Our line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

are useful for

- Making predictions about Y using X.
  - Predict Y when $X = x$ with $\beta_0 + \beta_1 \cdot x$
- Learning about the relationship between Y and X.
  - Interpret the signs and magnitudes of β_0 and β_1

Questions?

Table of Contents

Linear Regression: Estimation Introduction

As we went over in the last section we are interested in the line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Problem: We do not know know the joint distribution of (Y, X), so we cannot to solve for β_0 and β_1 by evaluating the expectation above.

As we went over in the last section we are interested in the line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Problem: We do not know know the joint distribution of (Y, X), so we cannot to solve for β_0 and β_1 by evaluating the expectation above.

Solution: Use data to estimate the parameters β_0 and β_1.

Solution: Use data to try and estimate the parameters β_0 and β_1.

How do we do this?

Linear Regression: The Estimator

Solution: Use data to try and estimate the parameters β_0 and β_1.

How do we do this?

Intuition:

- Suppose we have access to n randomly collected samples $\{Y_i, X_i\}_{i=1}^n$

Linear Regression: The Estimator

Solution: Use data to try and estimate the parameters β_0 and β_1.

How do we do this?

Intuition:

- Suppose we have access to n randomly collected samples $\{Y_i, X_i\}_{i=1}^{n}$
- We are interested in the line of best fit between Y and X in the population

$$\beta_0, \beta_1 = \arg \min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Linear Regression: The Estimator

Solution: Use data to try and estimate the parameters β_0 and β_1.

How do we do this?

Intuition:

- Suppose we have access to n randomly collected samples $\{Y_i, X_i\}_{i=1}^n$
- We are interested in the line of best fit between Y and X in the population

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

- We estimate the line of best fit between Y and X in the population using the line of best fit between Y_i and X_i in our sample:

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_0 - b_1 \cdot X_i)^2$$

  - Same idea as using $\bar{X}$ to estimate $\mathbb{E}[X]$, etc.

Linear Regression: The Estimator

Solution: Use data to try and estimate the parameters β_0 and β_1.

How do we do this?

Intuition:

- Suppose we have access to n randomly collected samples $\{Y_i, X_i\}_{i=1}^n$
- We are interested in the line of best fit between Y and X in the population

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

- We estimate the line of best fit between Y and X in the population using the line of best fit between Y_i and X_i in our sample:

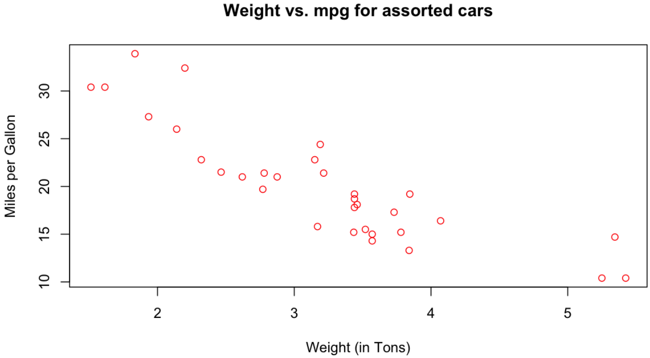
$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{b_0, b_1} \cancel{\frac{1}{n}} \sum_{i=1}^n (Y_i - b_0 - b_1 \cdot X_i)^2$$

  - Same idea as using $\bar{X}$ to estimate $\mathbb{E}[X]$, etc.

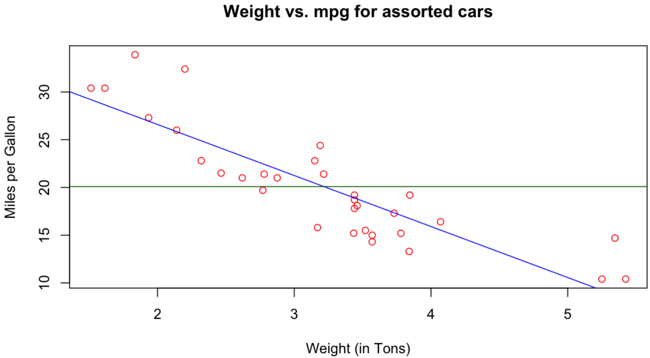
Linear Regression: The Estimator

Let's see how this looks like in practice. Suppose we are interested in the relationship between X, a car's weight, and Y a car's miles per gallon (mpg).

We collect some data $\{Y_i, X_i\}_{i=1}^n$ where each (Y_i, X_i) pair represents the miles per gallon and weight of a particular vehicle in our dataset. We can represent our data using a scatterplot

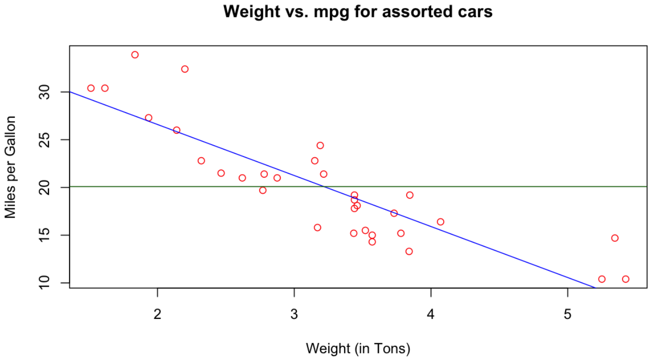


Now to estimate $\hat{\beta}_0, \hat{\beta}_1$ we simply find the line of best fit between the Y_i and X_i 's in our data.



Linear Regression: The Estimator

Now to estimate $\hat{\beta}_0, \hat{\beta}_1$ we simply find the line of best fit between the Y_i and X_i 's in our data.



The blue line represents the line of best fit whereas the green line represents a straight line through $\bar{Y}$. We can see that the blue line is much closer to the data than the green line.

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How do we interpret these estimates?

Linear Regression: The Estimator

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How do we interpret these estimates?

- $\hat{\beta}_0 = 37.2851$: We estimate that the average value of Y when $X = 0$ is 37.2851
  - In context: we estimate that the average mpg for a car that weights 0 tons is 37.2851 miles per gallon

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How do we interpret these estimates?

- $\hat{\beta}_0 = 37.2851$: We estimate that the average value of Y when $X = 0$ is 37.2851
  - In context: we estimate that the average mpg for a car that weights 0 tons is 37.2851 miles per gallon
- $\hat{\beta}_1 = -5.3445$: We estimate that, on average, a one unit increase in X is associated with a 5.3445 unit decrease in Y.
  - In context: we estimate that, on average, a one ton increase in car weight is associated with a 5.3445 unit decrease in miles per gallon.

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How can we use these estimates for prediction?

- Suppose we have a car that weighs 3.5 tons. Based on our estimates, what would we predict its miles per gallon to be?

Linear Regression: The Estimator

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How can we use these estimates for prediction?

- Suppose we have a car that weighs 3.5 tons. Based on our estimates, what would we predict its miles per gallon to be?
    - Our estimated regression line is

$$\text{Predicted MPG} = 37.2851 - 5.3445 \cdot \text{Weight in Tons.}$$

Linear Regression: The Estimator

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How can we use these estimates for prediction?

- Suppose we have a car that weighs 3.5 tons. Based on our estimates, what would we predict its miles per gallon to be?
  - Our estimated regression line is
  $$\text{Predicted MPG} = 37.2851 - 5.3445 \cdot \text{Weight in Tons.}$$
  - Using this line and plugging in we get that
  $$\text{Predicted MPG} = 37.2851 - 5.3445 \cdot 3.5 = 18.5793.$$

Linear Regression: The Estimator

In this case we have that $\hat{\beta}_0 = 37.2851$ and $\hat{\beta}_1 = -5.3445$.

How can we use these estimates for prediction?

- Suppose we have a car that weighs 3.5 tons. Based on our estimates, what would we predict its miles per gallon to be?
  - Our estimated regression line is
  $$\text{Predicted MPG} = 37.2851 - 5.3445 \cdot \text{Weight in Tons.}$$
  - Using this line and plugging in we get that
  $$\text{Predicted MPG} = 37.2851 - 5.3445 \cdot 3.5 = 18.5793.$$
  - We denote this predicted MPG as $\widehat{\text{MPG}}$ and in general will denote our predictions as $\hat{Y}$ so that our estimated regression line can be written
  $$\hat{Y} = \hat{\beta}_0 + \hat{\beta}_1 \cdot X.$$

Linear Regression: The Estimator

Notice a couple things in the above interpretations

- The intercept is often uninterpretable (What car would weigh 0 tons?). For this reason we often focus our analysis on the slope coefficient.
- The interpretation is deliberately not causal. We use "associated with a decrease..." as opposed to "leads to a decrease..."

Linear Regression: Formulas

Now that we've gotten some intuition for what linear regression is doing and how to use our sample to estimate the parameters of interest, let's derive explicit formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Linear Regression: Formulas

Now that we've gotten some intuition for what linear regression is doing and how to use our sample to estimate the parameters of interest, let's derive explicit formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Recall that

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_0 - b_1 \cdot X_i)^2 .$$

Linear Regression: Formulas

Now that we've gotten some intuition for what linear regression is doing and how to use our sample to estimate the parameters of interest, let's derive explicit formulas for $\hat{\beta}_0$ and $\hat{\beta}_1$.

Recall that

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} \left(Y_i - b_0 - b_1 \cdot X_i\right)^2 .$$

Taking first order conditions gives us that

$$\frac{\partial}{\partial b_0} : \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_i) = 0$$

$$\frac{\partial}{\partial b_1} : \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_i) \cdot X_i = 0$$

Rearranging the first equality gives us

$$\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_0 - \frac{1}{n}\sum_{i=1}^{n} \hat{\beta}_1 \cdot X_i = 0$$

Linear Regression: Formulas

Rearranging the first equality gives us

$$\begin{aligned}\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0 - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1 \cdot X_i &= 0 \\ \bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i &= 0\end{aligned}$$

Rearranging the first equality gives us

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n} Y_i - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_0 - \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1 \cdot X_i &= 0 \\
\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i &= 0 \\
\bar{Y} - \hat{\beta}_0 - \hat{\beta}_1 \bar{X} &= 0 \\
\hat{\beta}_0 &= \bar{Y} - \hat{\beta}_1 \bar{X}
\end{aligned}$$

So that what remains is to solve for $\hat{\beta}_1$.

Rearranging the second equality gives us

$$\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \hat{\beta}_0 \frac{1}{n}\sum_{i=1}^{n} X_i - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0$$

Linear Regression: Formulas

Rearranging the second equality gives us

$$\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \hat{\beta}_0 \frac{1}{n}\sum_{i=1}^{n} X_i - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0$$

Using the prior result that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ gives:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - (\bar{Y} - \hat{\beta}_1 \bar{X})\bar{X} - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0$$

$$\left(\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \bar{Y}\bar{X}\right) + \hat{\beta}_1 \left((\bar{X})^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = 0$$

Linear Regression: Formulas

Rearranging the second equality gives us

$$\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \hat{\beta}_0 \frac{1}{n}\sum_{i=1}^{n} X_i - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0$$

Using the prior result that $\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$ gives:

$$\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - (\bar{Y} - \hat{\beta}_1 \bar{X})\bar{X} - \hat{\beta}_1 \frac{1}{n}\sum_{i=1}^{n} X_i^2 = 0$$

$$\left(\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \bar{Y}\bar{X}\right) + \hat{\beta}_1 \left((\bar{X})^2 - \frac{1}{n}\sum_{i=1}^{n} X_i^2\right) = 0$$

So, finally

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \bar{Y}\bar{X}}{\frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2}.$$

Linear Regression: Formulas

Let's make use of the following equalities to represent $\hat{\beta}_1$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}) = \frac{1}{n}\sum_{i=1}^{n}Y_i X_i - \bar{Y}\bar{X}$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n}X_i^2 - (\bar{X})^2$$

Linear Regression: Formulas

Let's make use of the following equalities to represent $\hat{\beta}_1$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \bar{Y}\bar{X}$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$$

Then:

$$\hat{\beta}_1 = \frac{\overbrace{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}^{\text{Sample Covariance between } Y \text{ and } X}}{\underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}_{\text{Sample Variance of } X}}$$

Linear Regression: Formulas

Let's make use of the following equalities to represent $\hat{\beta}_1$

$$\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X}) = \frac{1}{n}\sum_{i=1}^{n} Y_i X_i - \bar{Y}\bar{X}$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = \frac{1}{n}\sum_{i=1}^{n} X_i^2 - (\bar{X})^2$$

Then:

$$\hat{\beta}_1 = \frac{\overbrace{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}^{\text{Sample Covariance between } Y \text{ and } X}}{\underbrace{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}_{\text{Sample Variance of } X}}$$

This ties in nicely as, if we recall from earlier, we found that

$$\beta_1 = \frac{\mathrm{Cov}(Y, X)}{\mathrm{Var}(X)} = \frac{\mathbb{E}[(Y - \mu_Y)(X - \mu_X)]}{\mathbb{E}[(X - \mu_X)^2]}.$$

We have now gone over how use data to obtain estimates $\hat{\beta}_0, \hat{\beta}_1$ of our parameters of interest β_0, β_1.

$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} (Y_i - b_0 - b_1 \cdot X_i)^2$$

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[\left(Y - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot X\right)^2\right]$$

Notice that, while the parameters of interest β_0 and β_1 are fixed quantities, the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of the data; they depend on the specific sample of data collected.

Linear Regression: Randomness

We have now gone over how use data to obtain estimates $\hat{\beta}_0, \hat{\beta}_1$ of our parameters of interest β_0, β_1. Notice that, while the parameters of interest β_0 and β_1 are fixed quantities, the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of the data; they depend on the specific sample of data collected.

Some Questions to Consider:

1. What would happen to our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ if we were to collect a different sample of data?

Linear Regression: Randomness

We have now gone over how use data to obtain estimates $\hat{\beta}_0, \hat{\beta}_1$ of our parameters of interest β_0, β_1. Notice that, while the parameters of interest β_0 and β_1 are fixed quantities, the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of the data; they depend on the specific sample of data collected.

Some Questions to Consider:

1. What would happen to our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ if we were to collect a different sample of data?
2. How can we model the distribution of our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$?

Linear Regression: Randomness

We have now gone over how use data to obtain estimates $\hat{\beta}_0, \hat{\beta}_1$ of our parameters of interest β_0, β_1. Notice that, while the parameters of interest β_0 and β_1 are fixed quantities, the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of the data; they depend on the specific sample of data collected.

Some Questions to Consider:

1. What would happen to our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ if we were to collect a different sample of data?
2. How can we model the distribution of our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$?
3. What happens to this distribution as $n \to \infty$?

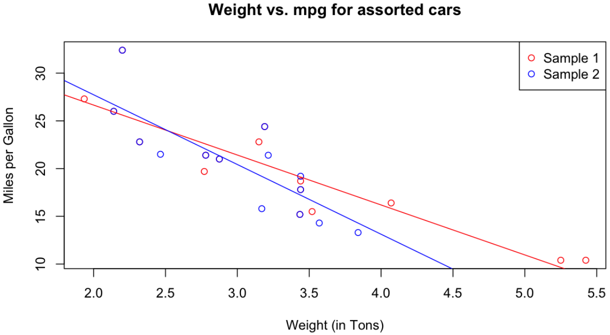
Linear Regression: Randomness

Question: What would happen to our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ if we were to collect a different sample of data?

Linear Regression: Randomness

Question: What would happen to our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ if we were to collect a different sample of data?

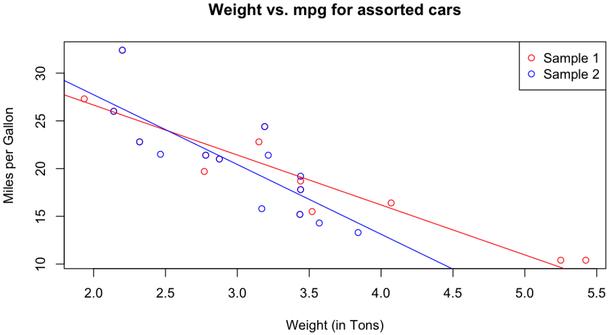
Let's return to the cars data and see how our regression lines look when we consider two different (random) samples.



Linear Regression: Randomness

Question: What would happen to our estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ if we were to collect a different sample of data?

Let's return to the cars data and see how our regression lines look when we consider two different (random) samples.



- Sample 1: $\hat{\beta}_0 = 37.1285$ and $\hat{\beta}_1 = -5.2341.$
- Sample 2: $\hat{\beta}_0 = 42.352$ and $\hat{\beta}_1 = -7.307.$

Linear Regression: Randomness

Key Concept: Because the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of the random sample $\{Y_i, X_i\}_{i=1}^n$ they are themselves random variables.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}$$

Problem: How do we connect $\hat{\beta}_0$ and $\hat{\beta}_1$ to the population parameters β_0 and β_1?

Key Concept: Because the estimators $\hat{\beta}_0$ and $\hat{\beta}_1$ are functions of the random sample $\{Y_i, X_i\}_{i=1}^n$ they are themselves random variables.

$$\hat{\beta}_0 = \bar{Y} - \hat{\beta}_1 \bar{X}$$
$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^n (Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2}$$

Problem: How do we connect $\hat{\beta}_0$ and $\hat{\beta}_1$ to the population parameters β_0 and β_1?

Fundamental Question: Given estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ what can we say about the underlying parameters of interest β_0 and β_1?

Table of Contents

Suppose we are interested in the association between years of education and income. We collect a random sample of size $n = 100$, $\{Y_i, X_i\}_{i=1}^{100}$ and run a simple linear regression of $Y = INC$ against $X = EDU$.

Linear Regression: Motivating Idea

Suppose we are interested in the association between years of education and income. We collect a random sample of size $n = 100$, $\{Y_i, X_i\}_{i=1}^{100}$ and run a simple linear regression of $Y = INC$ against $X = EDU$.

That is, we are interested in the parameters β_0 and β_1 that dictate the line of best fit between income and education in the population

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[(INC - \tilde{\beta}_0 - \tilde{\beta}_1 \cdot EDU)^2\right].$$

or equivalently the parameters from the linear model

$$INC = \beta_0 + \beta_1 \cdot EDU + \epsilon.$$

where $\mathbb{E}[\epsilon \cdot EDU] = 0$.

Linear Regression: Motivating Idea

Using our data $\{Y_i, X_i\}_{i=1}^n$ we find that $\hat{\beta}_1 = 0.5$.

$$\hat{\beta}_0 \hat{\beta}_1 = \arg\min_{b_0, b_1} \frac{1}{n} \sum_{i=1}^{n} \left\{Y_i - b_0 - b_1 \cdot X_i\right\}^2 .$$

Our friend, Prince Harry Estranged of England, however claims that there is no association between education and income, that is that $\beta_1 = 0$.

Question: How can we tell if he is right?

Question: How can we tell if he is right?

Answer: One way would be to find the probability that we would obtain $\hat{\beta}_1 = 0.5$ (or something more extreme) if the true value of β_1 was 0.

$$\Pr(|\hat{\beta}_1| \geq 0.5 | \beta_1 = 0).$$

If this probability is sufficently low, we can reject Former Prince Harry's claim. Otherwise he may be right.

Question: How can we tell if he is right?

Answer: One way would be to find the probability that we would obtain $\hat{\beta}_1 = 0.5$ (or something more extreme) if the true value of β_1 was 0.

$$\Pr(|\hat{\beta}_1| \geq 0.5 | \beta_1 = 0).$$

If this probability is sufficently low, we can reject Former Prince Harry's claim. Otherwise he may be right.

To calculate this probability we will need to know something about the (approximate) distribution of $\hat{\beta}_1$ and how that is related to the true parameter β_1.

Linear Regression: Assumptions

In order to connect the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ to the population parameters, we will need to make some (light) assumptions about the underlying distribution of (Y, X) from which our sample $\{Y_i, X_i\}_{i=1}^n$ is drawn.

Linear Regression: Assumptions

In order to connect the estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ to the population parameters, we will need to make some (light) assumptions about the underlying distribution of (Y, X) from which our sample $\{Y_i, X_i\}_{i=1}^n$ is drawn.

It will be helpful to recall the following definitions here

$$\beta_0, \beta_1 = \arg\min_{\tilde{\beta}_0, \tilde{\beta}_1} \mathbb{E}\left[(Y - \tilde{\beta}_0 - \tilde{\beta}_1)^2\right]$$
$$\epsilon = Y - \beta_0 - \beta_1 \cdot X$$

And see that ϵ is itself a random variable.

Make the following assumptions

1. Random Sampling: Assume that $\{Y_i, X_i\}$ are independently and identically distributed; $(Y_i, X_i) \overset{\text{i.i.d}}{\sim} (Y, X)$
   - Essentially this means that our random sample is "representative of the population"
   - Would be violated if say, we only sampled cars made in Los Angeles and we were trying to make inferences about all cars produced in the US

Linear Regression: Assumptions

Make the following assumptions

1. Random Sampling: Assume that $\{Y_i, X_i\}$ are independently and identically distributed; $(Y_i, X_i) \overset{\text{i.i.d}}{\sim} (Y, X)$
2. Homoskedasticity: Assume that $\mathrm{Var}(\epsilon \mid X = x) = \sigma_\epsilon^2$ for all possible values of x.
   - This means that Y is equally spread around the regression line for all values of X.
   - This is a fairly strong assumption to make and we will relax it later on, but it is helpful for now to provide insight.
   - Conditional variance is similar to the conditional expectation that we went over in our Econ 41 review
$$\mathrm{Var}(\epsilon \mid X = x) = \mathbb{E}[\epsilon^2 | X = x] - (\mathbb{E}[\epsilon | X = x])^2.$$
   - An important implication of this is that
$$\mathrm{Var}(\epsilon(X - \mu_X)) = \mathrm{Var}(\epsilon)\,\mathrm{Var}(X) = \sigma_\epsilon^2 \sigma_X^2.$$

   Questions?

Linear Regression: Assumptions

Make the following assumptions

1. Random Sampling: Assume that $\{Y_i, X_i\}$ are independently and identically distributed; $(Y_i, X_i) \overset{\text{i.i.d}}{\sim} (Y, X)$
2. Homoskedasticity: Assume that $\mathrm{Var}(\epsilon \mid X = x) = \sigma_\epsilon^2$ for all possible values of x.
   - An important implication of this is that
   $$\mathrm{Var}(\epsilon(X - \mu_X)) = \mathrm{Var}(\epsilon)\,\mathrm{Var}(X) = \sigma_\epsilon^2 \sigma_X^2.$$
3. Rank Condition: There must be at least two distinct values of X that appear in the population.
   - Need at least two distinct points to make a line.
   - If there is only one distinct point then our minimization problem is undefined.

Linear Regression: Assumptions

Make the following assumptions

1. Random Sampling: Assume that $\{Y_i, X_i\}$ are independently and identically distributed; $(Y_i, X_i) \overset{\text{i.i.d}}{\sim} (Y, X)$
2. Homoskedasticity: Assume that $\mathrm{Var}(\epsilon \mid X = x) = \sigma_\epsilon^2$ for all possible values of x.
   - An important implication of this is that
   $$\mathrm{Var}(\epsilon(X - \mu_X)) = \mathrm{Var}(\epsilon)\,\mathrm{Var}(X) = \sigma_\epsilon^2 \sigma_X^2.$$
3. Rank Condition: There must be at least two distinct values of X that appear in the population.

And that's it!

Linear Regression: Asymptotic Distribution

Given these assumptions (Random Sampling, Homoskedasticity, Rank Condition) let's try and figure out what the approximate distribution is of $\hat{\beta}_1$.

Recall that

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

Linear Regression: Asymptotic Distribution

Given these assumptions (Random Sampling, Homoskedasticity, Rank Condition) let's try and figure out what the approximate distribution is of $\hat{\beta}_1$.

Recall that

$$\hat{\beta}_1 = \frac{\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

By definition of $\epsilon = Y - \beta_0 - \beta_1 \cdot X$:

$$Y = \beta_0 + \beta_1 \cdot X;$$

and that by the first order conditions of β_0 and β_1:

$$\mathbb{E}[\epsilon] = 0$$
$$\mathbb{E}[\epsilon \cdot X] = 0$$

Linear Regression: Asymptotic Distribution

We will also make use of the following results from our probability review. If Z is a random variables and we have i.i.d observations $Z_1, Z_2, .., Z_n$:

The Law of Large Numbers states that as $n \to \infty$:

$$\bar{Z} \to \mathbb{E}[Z]$$

or, equivalently, $\bar{Z} \approx \mathbb{E}[Z]$ for n large.

The Central Limit Theorem states that as $n \to \infty$, approximately,

$$\sqrt{n}\left(\bar{Z} - \mathbb{E}[Z]\right) \sim N\left(0, \mathrm{Var}(Z)\right)$$

or, equivalently, $\bar{Z} \sim N\left(\mathbb{E}[Z], \mathrm{Var}(Z)/n\right)$.

Linear Regression: Asymptotic Distribution

Starting with:

$$\sqrt{n}\hat{\beta}_1 = \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Linear Regression: Asymptotic Distribution

Starting with:

$$\sqrt{n}\hat{\beta}_1 = \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Expand $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{\epsilon}$, where $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i$:

$$\sqrt{n}\hat{\beta}_1 = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\beta_1(X_i - \bar{X}) + (\epsilon_i - \bar{\epsilon})\right)(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Linear Regression: Asymptotic Distribution

Starting with:

$$\sqrt{n}\hat{\beta}_1 = \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Expand $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{\epsilon}$, where $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i$:

$$\sqrt{n}\hat{\beta}_1 = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\beta_1(X_i - \bar{X}) + (\epsilon_i - \bar{\epsilon})\right)(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Distribute to get:

$$\sqrt{n}\hat{\beta}_1 = \sqrt{n}\beta_1\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} + \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Linear Regression: Asymptotic Distribution

Starting with:

$$\sqrt{n}\hat{\beta}_1 = \frac{\sqrt{n}\frac{1}{n}\sum_{i=1}^{n}(Y_i - \bar{Y})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Expand $Y_i = \beta_0 + \beta_1 X_i + \epsilon_i$ and $\bar{Y} = \beta_0 + \beta_1 \bar{X} + \bar{\epsilon}$, where $\bar{\epsilon} = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i$:

$$\sqrt{n}\hat{\beta}_1 = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(\beta_1(X_i - \bar{X}) + (\epsilon_i - \bar{\epsilon})\right)(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Distribute to get:

$$\sqrt{n}\hat{\beta}_1 = \sqrt{n}\beta_1 \cancel{\frac{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}} + \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

So we have that:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Linear Regression: Asymptotic Distribution

So we have that:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) = \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\epsilon_i - \bar{\epsilon})(X_i - \bar{X})}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2}.$$

Using Law of Large Numbers replace $\bar{\epsilon} \approx \mathbb{E}[\epsilon] = 0$, $\bar{X} \approx \mu_X$, and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \approx \sigma_X^2$:

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \approx \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\epsilon_i(X_i - \mu_X)}{\sigma_X^2}.$$

Linear Regression: Asymptotic Distribution

Finally, note that by Central Limit Theorem, since

$$\mathbb{E}[\epsilon(X_i - \mu_X)] = \mathbb{E}[\epsilon X_i] - \mathbb{E}[\epsilon]\mu_X = 0.$$

we have that (approximately for large n):

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i (X_i - \mu_X) \sim N\left(0, \mathrm{Var}\left(\epsilon(X - \mu_X)\right)\right).$$

Linear Regression: Asymptotic Distribution

Finally, note that by Central Limit Theorem, since

$$\mathbb{E}[\epsilon(X_i - \mu_X)] = \mathbb{E}[\epsilon X_i] - \mathbb{E}[\epsilon]\mu_X = 0.$$

we have that (approximately for large n):

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i (X_i - \mu_X) \sim N\left(0, \operatorname{Var}\left(\epsilon(X - \mu_X)\right)\right).$$

Now note that by Homoskedaticity:

$$\operatorname{Var}(\epsilon(X - \mu_X)) = \sigma_\epsilon^2 \sigma_X^2$$

so that (approximately for large n):

$$\frac{1}{\sqrt{n}} \sum_{i=1}^{n} \epsilon_i (X_i - \mu_X) \sim N\left(0, \sigma_\epsilon^2 \sigma_X^2\right).$$

Putting this all together, we have that, approximately for n large;

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \sim \frac{N(0, \sigma_\epsilon^2 \sigma_X^2)}{\sigma_X^2} = N\big(0, \underbrace{\sigma_\epsilon^2/\sigma_X^2}_{:=\sigma_{\beta_1}^2}\big).$$

where in the last equality we use the fact that $a\,\mathrm{Var}(Z) = \mathrm{Var}(a^2 Z)$.

Putting this all together, we have that, approximately for n large;

$$\sqrt{n}\left(\hat{\beta}_1 - \beta_1\right) \sim \frac{N(0, \sigma_\epsilon^2 \sigma_X^2)}{\sigma_X^2} = N\big(0, \underbrace{\sigma_\epsilon^2/\sigma_X^2}_{:=\sigma_{\beta_1}^2}\big).$$

where in the last equality we use the fact that $a\,\mathrm{Var}(Z) = \mathrm{Var}(a^2 Z)$. Other ways of putting this are, approximately for n large:

$$\hat{\beta}_1 \sim N\left(\beta_1, \sigma_{\beta_1}^2/n\right)$$

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}/\sqrt{n}} \sim N(0,1)$$

where as a reminder $\sigma_{\beta_1} = \sigma_\epsilon/\sigma_X$. This last form is what we will use the most.

Linear Regression: Asymptotic Distribution

Following similar steps we can derive the approximate distribution of $\hat{\beta}_0$ as well as the covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\sqrt{n}\left(\hat{\beta}_1 - \hat{\beta}_1\right) \sim N\left(0, \frac{\sigma_\epsilon^2}{\sigma_X^2}\right)$$

$$\sqrt{n}\left(\hat{\beta}_0 - \beta_0\right) \sim N\left(0, \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{\sigma_X^2}\right)$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}$$

Following similar steps we can derive the approximate distribution of $\hat{\beta}_0$ as well as the covariance between $\hat{\beta}_0$ and $\hat{\beta}_1$:

$$\sqrt{n}\left(\hat{\beta}_1 - \hat{\beta}_1\right) \sim N\left(0, \frac{\sigma_\epsilon^2}{\sigma_X^2}\right)$$

$$\sqrt{n}\left(\hat{\beta}_0 - \beta_0\right) \sim N\left(0, \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{\sigma_X^2}\right)$$

$$\text{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}$$

Important to remember these! The above is just providing intuition on how we get these results.

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

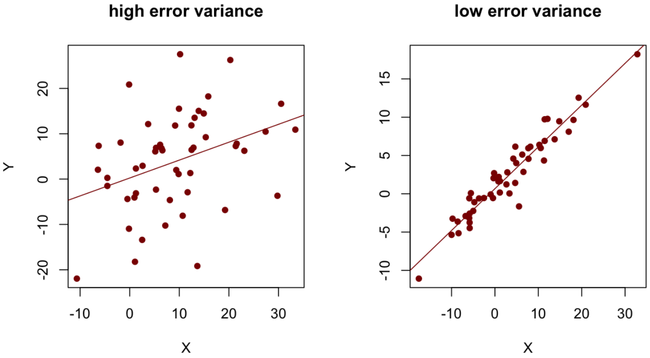
First notice that these variances are increasing with σ_ϵ^2.

Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

First notice that these variances are increasing with σ_ϵ^2.

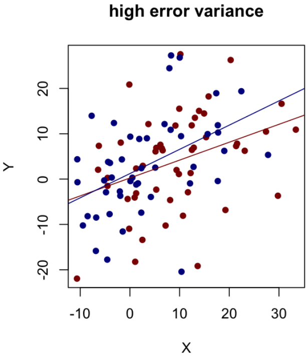


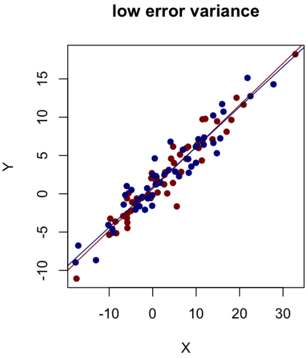
Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

First notice that these variances are increasing with σ_ϵ^2.





Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

First notice that these variances are increasing with σ_ϵ^2.

Intuition: If points are more tightly distributed around the regression line it is easier to tell what the regression line is.

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \;\; \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \;\; \text{and} \;\; \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

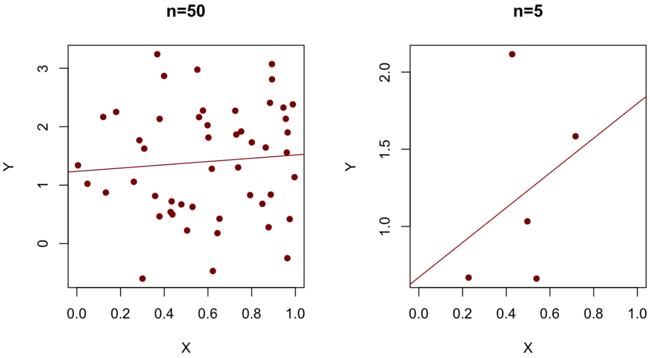
These variances tend to zero as $n \to \infty$; as we collect more data we are closer to the true values β_0 and β_1.

Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

These variances tend to zero as $n \to \infty$; as we collect more data we are closer to the true values β_0 and β_1.

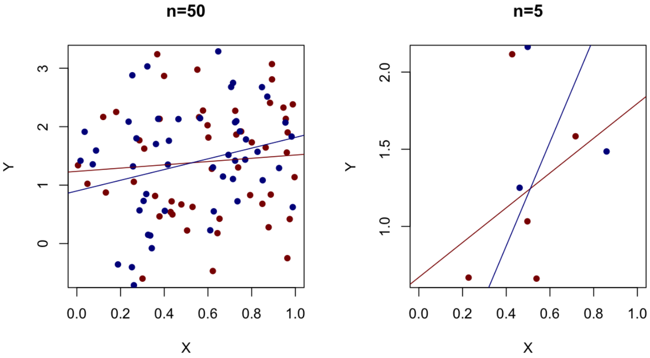


Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

These variances tend to zero as $n \to \infty$; as we collect more data we are closer to the true values β_0 and β_1.



For large n we have that

$$\operatorname{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \operatorname{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \operatorname{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

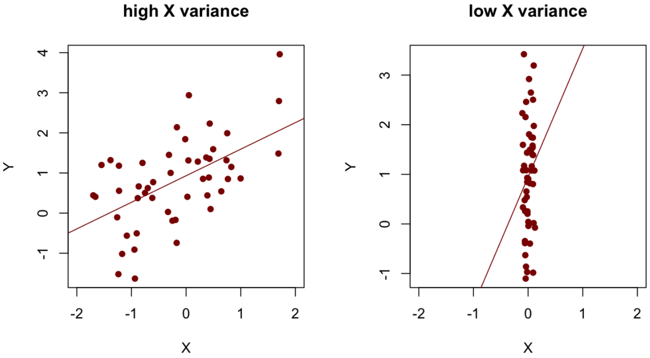
These variances decrase as σ_X^2 increases; as the spread of X increases we can make out the line more clearly.

Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

These variances decrase as σ_X^2 increases; as the spread of X increases we can make out the line more clearly.

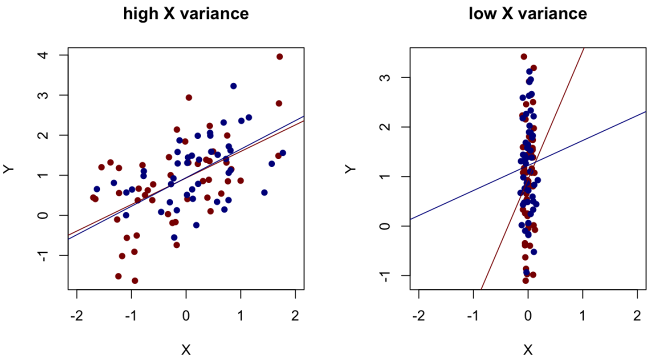


Linear Regression: Asymptotic Variances

For large n we have that

$$\mathrm{Var}(\hat{\beta}_1) = \frac{\sigma_\epsilon^2}{n \cdot \sigma_X^2}, \quad \mathrm{Var}(\hat{\beta}_0) = \sigma_\epsilon^2 \frac{\mathbb{E}[X^2]}{n \cdot \sigma_X^2}, \quad \text{and} \quad \mathrm{Cov}(\hat{\beta}_1, \hat{\beta}_0) = -\sigma_\epsilon^2 \frac{\mathbb{E}[X]}{n \cdot \sigma_X^2}.$$

These variances decrase as σ_X^2 increases; as the spread of X increases we can make out the line more clearly.



Linear Regression: Questions

Questions?

Linear Regression: Variance Estimation

Positive Result: Under homoskedasticity, for n large, we have (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

where

$$\sigma_{\beta_1} = \frac{\sigma_\epsilon^2}{\sigma_X^2}.$$

Positive Result: Under homoskedasticity, for n large, we have (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

where

$$\sigma_{\beta_1} = \frac{\sigma_\epsilon^2}{\sigma_X^2}.$$

Problem: What is σ_{β_1}? How can we estimate it?

Linear Regression: Variance Estimation

Positive Result: Under homoskedasticity, for n large, we have (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

where

$$\sigma_{\beta_1} = \frac{\sigma_\epsilon^2}{\sigma_X^2}.$$

Problem: What is σ_{β_1}? How can we estimate it?

- By LLN we know how to esimate $\mathrm{Var}(X)$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \approx \mathrm{Var}(X).$$

Linear Regression: Variance Estimation

Positive Result: Under homoskedasticity, for n large, we have (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

where

$$\sigma_{\beta_1} = \frac{\sigma_\epsilon^2}{\sigma_X^2}.$$

Problem: What is σ_{β_1}? How can we estimate it?

- By LLN we know how to esimate $\mathrm{Var}(X)$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 \approx \mathrm{Var}(X).$$

- But what about $\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$?

Linear Regression: Variance Estimation

To estimate $\mathrm{Var}(\epsilon)$ we first construct estimated residuals $\hat{\epsilon}_i$ via

$$\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_i.$$

Because $\hat{\beta}_1 \to \beta_1$ and $\hat{\beta}_0 \to \beta_0$ we can say that $\hat{\epsilon}_i \approx \epsilon_i = Y_i - \beta_0 - \beta_1 X_i$ (for n large).

Linear Regression: Variance Estimation

To estimate $\mathrm{Var}(\epsilon)$ we first construct estimated residuals $\hat{\epsilon}_i$ via

$$\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_i.$$

Because $\hat{\beta}_1 \to \beta_1$ and $\hat{\beta}_0 \to \beta_0$ we can say that $\hat{\epsilon}_i \approx \epsilon_i = Y_i - \beta_0 - \beta_1 X_i$ (for n large).

Also by the first order conditions for $\hat{\beta}_0$ we have that

$$-\frac{1}{n}\sum_{i=1}^{n}(\underbrace{Y_i - \hat{\beta}_0 - \hat{\beta}_1 \cdot X_i}_{=\hat{\epsilon}_i}) = 0.$$

so that

$$\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i = \bar{\hat{\epsilon}}_i = 0.$$

Linear Regression: Variance Estimation

Putting this together we can estimate $\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$ by calculating the sample variance of $\hat{\epsilon}_i$:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 - \cancel{(\bar{\hat{\epsilon}}_i)^2}$$

Linear Regression: Variance Estimation

Putting this together we can estimate $\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$ by calculating the sample variance of $\hat{\epsilon}_i$:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 - \cancel{(\bar{\hat{\epsilon}}_i)^2}$$

By $\hat{\beta}_1 \to \beta_1$ and $\hat{\beta}_0 \to \beta_0$ as $n \to \infty$;

$$\approx \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2$$

Linear Regression: Variance Estimation

Putting this together we can estimate $\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$ by calculating the sample variance of $\hat{\epsilon}_i$:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i^2 - \cancel{(\bar{\hat{\epsilon}}_i)^2}$$

By $\hat{\beta}_1 \to \beta_1$ and $\hat{\beta}_0 \to \beta_0$ as $n \to \infty$;

$$\approx \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2$$

By Law of Large Numbers;

$$\approx \mathbb{E}[\epsilon^2]$$

Linear Regression: Variance Estimation

Putting this together we can estimate $\mathrm{Var}(\epsilon) = \sigma_\epsilon^2$ by calculating the sample variance of $\hat{\epsilon}_i$:

$$\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i^2 - \cancel{(\bar{\hat{\epsilon}}_i)^2}$$

By $\hat{\beta}_1 \to \beta_1$ and $\hat{\beta}_0 \to \beta_0$ as $n \to \infty$;

$$\approx \frac{1}{n}\sum_{i=1}^{n} \epsilon_i^2$$

By Law of Large Numbers;

$$\approx \mathbb{E}[\epsilon^2]$$

By $\mathbb{E}[\epsilon] = 0$;

$$= \mathrm{Var}(\epsilon) = \sigma_\epsilon^2$$

Putting all of this together, we can estimate $\sigma^2_{\beta_1} = \frac{\sigma^2_X}{\sigma^2_X}$ via;

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_\epsilon}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2} \approx \sigma^2_{\beta_1}.$$

since for large n

$$\hat{\sigma}^2_\epsilon = \frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i^2 \approx \sigma^2_\epsilon$$

$$\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 \approx \sigma^2_X.$$

Now, since we have that (approximately, for large n):

$$\frac{\hat{\beta}_1 - \beta_1}{\sigma_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

And since, as we have established above, $\hat{\sigma}_{\beta_1} \approx \sigma_{\beta_1}$, for large n we can say that (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

The quantity $\hat{\sigma}_{\beta_1}/\sqrt{n}$ is often referred to as the standard error of $\hat{\beta}_1$.

The quantity $\hat{\sigma}_{\beta_1}/\sqrt{n}$ is often referred to as the standard error of $\hat{\beta}_1$.

The quantity $\hat{\sigma}_{\beta_1}/\sqrt{n}$ is often referred to as the standard error of $\hat{\beta}_1$.

Linear Regression: Variance Estimation

The quantity $\hat{\sigma}_{\beta_1}/\sqrt{n}$ is often referred to as the standard error of $\hat{\beta}_1$.

In general, if we have a parameter θ that we estimate with $\hat{\theta}$, the quantity $\hat{\sigma}_{\theta}/\sqrt{n}$ will be referred to as the standard error of $\hat{\theta}$ where

$$\hat{\sigma}_{\theta} = \sqrt{\mathrm{Var}(\hat{\theta})}.$$

Variance Estimation: Questions

Questions?

Linear Regression: Why Asymptotic Distribution?

Let's return to our example and see why this characterization is useful. Recall that in our example we are interested in the regression parameters from regression $Y = INC$ (income in thousands of dollars) against $X = EDU$ (years of education).

Linear Regression: Why Asymptotic Distribution?

Let's return to our example and see why this characterization is useful. Recall that in our example we are interested in the regression parameters from regression $Y = INC$ (income in thousands of dollars) against $X = EDU$ (years of education).

After collecting a sample size of 100, $\{Y_i, X_i\}_{i=1}^{100}$ we find that:

$$\hat{\beta}_1 = 0.5$$

$$\frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 = 25$$

$$\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = 16$$

Linear Regression: Why Asymptotic Distribution?

Our friend His Majesty Prince Harry claims there is no relationship between education and income, $\beta_1 = 0$. We claim that observing the magnitute of $|\hat{\beta}_1| = 0.5$ is evidence against this claim. Who is right?

Linear Regression: Why Asymptotic Distribution?

Our friend His Majesty Prince Harry claims there is no relationship between education and income, $\beta_1 = 0$. We claim that observing the magnitute of $|\hat{\beta}_1| = 0.5$ is evidence against this claim. Who is right?

- If $\beta_1 = 0$ we would expect $\hat{\beta}_1$ to be close to zero.

Linear Regression: Why Asymptotic Distribution?

Our friend His Majesty Prince Harry claims there is no relationship between education and income, $\beta_1 = 0$. We claim that observing the magnitute of $|\hat{\beta}_1| = 0.5$ is evidence against this claim. Who is right?

- If $\beta_1 = 0$ we would expect $\hat{\beta}_1$ to be close to zero.
- But there is still some randomness in $\hat{\beta}_1$, maybe we got $\hat{\beta}_1 = 0.5$ by chance.

Linear Regression: Why Asymptotic Distribution?

Our friend His Majesty Prince Harry claims there is no relationship between education and income, $\beta_1 = 0$. We claim that observing the magnitute of $|\hat{\beta}_1| = 0.5$ is evidence against this claim. Who is right?

- If $\beta_1 = 0$ we would expect $\hat{\beta}_1$ to be close to zero.
- But there is still some randomness in $\hat{\beta}_1$, maybe we got $\hat{\beta}_1 = 0.5$ by chance.

Want to use the (asymptotic) distribution of $\hat{\beta}_1$ to answer this question.

Linear Regression: Why Asymptotic Distribution?

Our friend His Majesty Prince Harry claims there is no relationship between education and income, $\beta_1 = 0$. We claim that observing the magnitute of $|\hat{\beta}_1| = 0.5$ is evidence against this claim. Who is right?

- If $\beta_1 = 0$ we would expect $\hat{\beta}_1$ to be close to zero.
- But there is still some randomness in $\hat{\beta}_1$, maybe we got $\hat{\beta}_1 = 0.5$ by chance.

Want to use the (asymptotic) distribution of $\hat{\beta}_1$ to answer this question.

- First need to estimate σ_{β_1}.

Using $\hat{\sigma}_{\epsilon}^2 = \frac{1}{n}\sum_{i=1}^{n}\epsilon_i^2 = 25$, and $\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2 = 16)$ we calculate

$$\begin{aligned}\hat{\sigma}_{\beta_1}^2 &= \frac{\hat{\sigma}_{\epsilon}^2}{\frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2} \\ &= \frac{25}{16}\end{aligned}$$

Using $\hat{\sigma}_\epsilon^2 = \frac{1}{n}\sum_{i=1}^n \epsilon_i^2 = 25$, and $\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2 = 16)$ we calculate

$$\begin{aligned}\hat{\sigma}_{\beta_1}^2 &= \frac{\hat{\sigma}_\epsilon^2}{\frac{1}{n}\sum_{i=1}^n (X_i - \bar{X})^2} \\ &= \frac{25}{16}\end{aligned}$$

Using this, we find that $\hat{\sigma}_{\beta_1} = \sqrt{\hat{\sigma}_{\beta_1}^2} = \frac{5}{4}$.

Linear Regression: Why Asymptotic Distribution?

Now recall that for n large we have that (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

Now recall that for n large we have that (approximately)

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}/\sqrt{n}} \sim N(0,1).$$

If the true value of $\beta_1 = 0$ this means that

$$\frac{\hat{\beta}_1}{5/40} = \frac{\hat{\beta}_1}{0.125} \sim N(0,1).$$

Linear Regression: Why Asymptotic Distribution?

Given that if $\beta_0 = 0$, $\hat{\beta}_1/0.125 \sim N(0,1)$, what is the probability of us observing $|\hat{\beta}_1| \geq 0.5$?

Linear Regression: Why Asymptotic Distribution?

Given that if $\beta_0 = 0$, $\hat{\beta}_1/0.125 \sim N(0,1)$, what is the probability of us observing $|\hat{\beta}_1| \geq 0.5$?

$$\Pr\left(|\hat{\beta}_1| \geq 0.5\right) = \Pr\left(|\hat{\beta}_1/0.125| \geq 0.5/0.125\right)$$

Linear Regression: Why Asymptotic Distribution?

Given that if $\beta_0 = 0$, $\hat{\beta}_1/0.125 \sim N(0,1)$, what is the probability of us observing $|\hat{\beta}_1| \geq 0.5$?

$$\begin{aligned}
\Pr\left(|\hat{\beta}_1| \geq 0.5\right) &= \Pr\left(|\hat{\beta}_1/0.125| \geq 0.5/0.125\right) \\
&= \Pr(|Z| \geq 4)
\end{aligned}$$

where $Z \sim N(0,1)$

Given that if $\beta_0 = 0$, $\hat{\beta}_1/0.125 \sim N(0,1)$, what is the probability of us observing $|\hat{\beta}_1| \geq 0.5$?

$$\begin{aligned}\Pr\left(|\hat{\beta}_1| \geq 0.5\right) &= \Pr\left(|\hat{\beta}_1/0.125| \geq 0.5/0.125\right) \\ &= \Pr(|Z| \geq 4)\end{aligned}$$

where $Z \sim N(0,1)$

$$\begin{aligned}&= \Pr(Z \geq 4) + \Pr(Z \leq 4) \\ &= 2\Pr(Z \geq 4)\end{aligned}$$

By symmetry of the normal distribution

Linear Regression: Why Asymptotic Distribution?

Given that if $\beta_0 = 0$, $\hat{\beta}_1/0.125 \sim N(0,1)$, what is the probability of us observing $|\hat{\beta}_1| \geq 0.5$?

$$\begin{aligned}\Pr\left(|\hat{\beta}_1| \geq 0.5\right) &= \Pr\left(|\hat{\beta}_1/0.125| \geq 0.5/0.125\right)\\ &= \Pr(|Z| \geq 4)\end{aligned}$$

where $Z \sim N(0,1)$

$$\begin{aligned}&= \Pr(Z \geq 4) + \Pr(Z \leq 4)\\ &= 2\Pr(Z \geq 4)\end{aligned}$$

By symmetry of the normal distribution

$$\approx 0.00006$$

Linear Regression: Why Asymptotic Distribution?

Using the asymptotic distribution result

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}/\sqrt{n}} \sim N(0,1),$$

we have found that if $\beta_1 = 0$, then $\Pr(|\hat{\beta}_1| \geq 0.5) \approx 0.0006$.

Linear Regression: Why Asymptotic Distribution?

Using the asymptotic distribution result

$$\frac{\hat{\beta}_1 - \beta_1}{\hat{\sigma}_{\beta_1}/\sqrt{n}} \sim N(0,1),$$

we have found that if $\beta_1 = 0$, then $\Pr(|\hat{\beta}_1| \geq 0.5) \approx 0.0006$.

So, given that we observed $\hat{\beta}_1 = 0.5$, it seems very unlikely that $\beta_1 = 0$. We can conclude against Prince Harry's claim.

Asymptotic Distribution: Questions

Questions?

Table of Contents

Linear Regression: Hypothesis Testing

The last exercise where we tested whether Prince Harry's claim made sense was an example of a hypothesis test.

In this section we will formally discuss hypothesis testing.

Linear Regression: Hypothesis Testing

Often in linear regression analysis, we are interested in