Econ 103: Introduction to Simple Linear Regression

Manu Navjeevan

UCLA

July 12, 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?

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?

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:

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

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

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

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

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

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 magnitutdes 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.

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.

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

Linear Regression: The Estimator

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 $\{Y_i, X_i\}_{i=1}^n$

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:

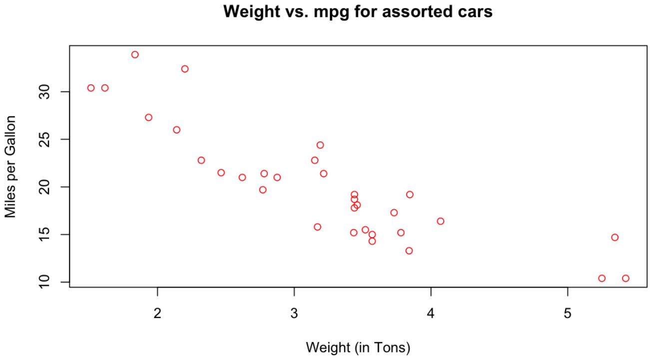
$$\hat{\beta}_0, \hat{\beta}_1 = \arg\min_{b_0, b_1} \cancel{\frac{1}{n}} \sum_{i=1}^{n} (Y_i - b_0 - b_1 \cdot X_i)^2$$

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

Linear Regression: The Estimator

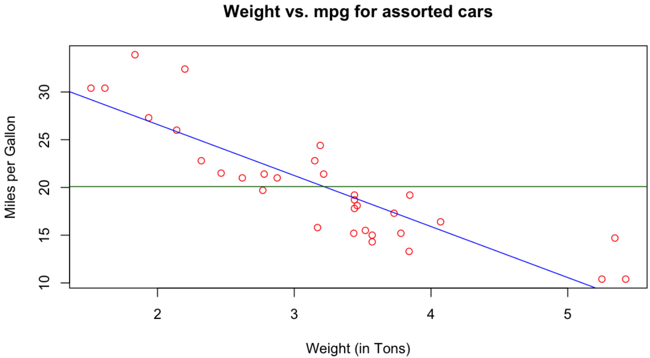
Let's see how this looks like in practice. Suppose we are interested in the relationship between X, a car's weight, and Y a car's miles per gallon (mpg).

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



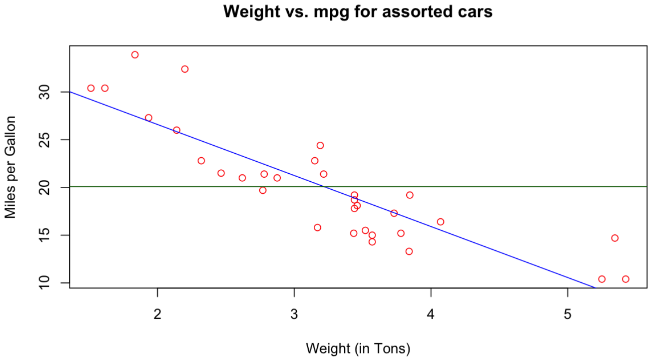
Linear Regression: The Estimator

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.



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.

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?

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

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
- $\hat{\beta}_1 = -5.3445$: 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.

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

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

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

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

Linear Regression: Formulas

Rearranging the first equality gives us

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

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 $\hat{\beta}_1$

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

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

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 $n \to \infty$?

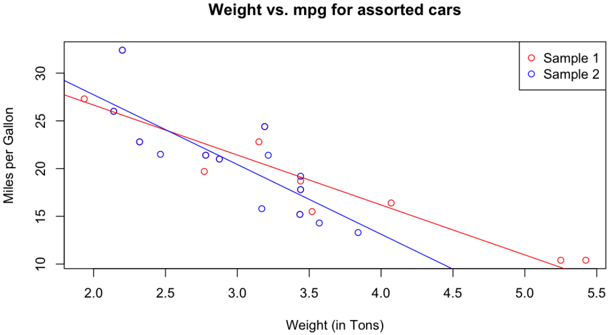
Linear Regression: Randomness

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?

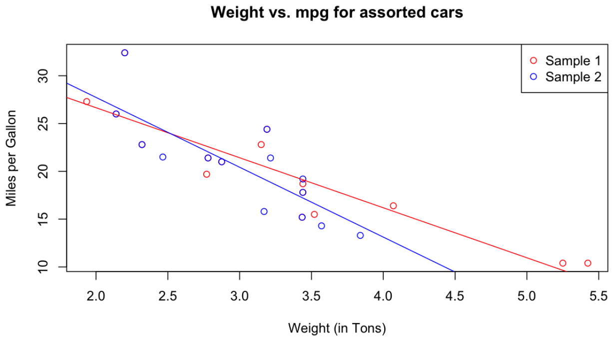
Let's return to the cars data and see how our regression lines look when we consider two different (random) samples.



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 $\hat{\beta}_1 = -7.307$.

Linea

Table of Contents