Econ 103: Multiple Linear Regression I

Manu Navjeevan

UCLA

July 30, 2021

Content Outline

The Model:

- Adding more covariates
- Assumptions needed for inference

The Estimator:

- Relation to Single Linear Regression Estimator
- Asymptotic Dsitribution

Inference:

- Hypothesis Tests and Linear Combinations
- Confidence Inervals

Modeling Choices:

- Polynomial Equations, transformations, and interactions
- R^2 and goodness of fit

Table of Contents

The Model: Introduction

So far we have used the model $Y = \beta_0 + \beta_1 X + \epsilon$ defined by the line of best fit parameters

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

to learn about the relationship between a single random variable X and Y and to use X to predict Y.

The Model: Introduction

So far we have used the model $Y = \beta_0 + \beta_1 X + \epsilon$ defined by the line of best fit parameters

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

to learn about the relationship between a single random variable X and Y and to use X to predict Y.

Examples:

- Using education to predict income or interpreting the coeffecient $\hat{\beta}_1$ to learn about the relationship between the two.

The Model: Introduction

So far we have used the model $Y = \beta_0 + \beta_1 X + \epsilon$ defined by the line of best fit parameters

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

to learn about the relationship between a single random variable X and Y and to use X to predict Y.

Examples:

- Using education to predict income or interpreting the coeffecient $\hat{\beta}_1$ to learn about the relationship between the two.
- Learning about the relationship between smoking and heart disease.

The Model: Introduction

However, what happens if we have access to multiple explanatory variables $X_1, \ldots, X_p$?

Examples:

- Suppose we wanted to impact the joint effect of education <u>and</u> experience on age?
- Learn about the relationship between smoking, genetic risk, and heart disease

The Model: Introduction

As before, we may be interested in the parameters of a “line of best fit” between Y and our explantory variables $X_1, \dots, X_p$:

$$\beta_0, \beta_1, \dots, \beta_p = \arg\min_{b_0,\dots,b_p} \mathbb{E}\left[(Y - b_0 - b_1X_1 - b_2X_2 - \cdots - b_pX_p)^2\right].$$

The Model: Introduction

As before, we may be interested in the parameters of a "line of best fit" between Y and our explantory variables $X_1, \ldots, X_p$:

$$\beta_0, \beta_1, \ldots, \beta_p = \arg\min_{b_0,\ldots,b_p} \mathbb{E}\left[(Y - b_0 - b_1 X_1 - b_2 X_2 - \cdots - b_p X_p)^2\right].$$

Again defining $\epsilon = Y - \beta_0 - \beta_1 X_1 - \cdots - \beta_p X_p$ these parameters generate the linear model

$$Y = \beta_0 + \beta_1 X_1 + \cdots + \beta_p X_p + \epsilon$$

where, by the first order conditions for β, $\mathbb{E}[\epsilon] = \mathbb{E}[\epsilon X_j] = 0$ for all $j = 0, 1, \ldots, p$.

The Model: Introduction

Example 1: Let Y be log wages, EDU be years of college education, and EXP be years of experience. Prior to this we have estimated the equation

$$Y = \beta_0 + \beta_1 EDU + \epsilon. \tag{1}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 EDU + \beta_2 EXP + \epsilon. \tag{2}$$

The Model: Introduction

Example 1: Let Y be log wages, EDU be years of college education, and EXP be years of experience. Prior to this we have estimated the equation

$$Y = \beta_0 + \beta_1 EDU + \epsilon. \tag{1}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 EDU + \beta_2 EXP + \epsilon. \tag{2}$$

Note that β_0, β_1 in model (1) will differ from β_0, β_1 in model (2).

The Model: Introduction

Example 1: Let Y be log wages, EDU be years of college education, and EXP be years of experience. Prior to this we have estimated the equation

$$Y = \beta_0 + \beta_1 EDU + \epsilon. \tag{1}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 EDU + \beta_2 EXP + \epsilon. \tag{2}$$

Note that β_0, β_1 in model (1) will differ from β_0, β_1 in model (2).

- In (1) β_0 corresponds to the average log wage for someone with no college education

The Model: Introduction

Example 1: Let Y be log wages, EDU be years of college education, and EXP be years of experience. Prior to this we have estimated the equation

$$Y = \beta_0 + \beta_1 EDU + \epsilon. \tag{1}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 EDU + \beta_2 EXP + \epsilon. \tag{2}$$

Note that β_0, β_1 in model (1) will differ from β_0, β_1 in model (2).

- In (1) β_0 corresponds to the average log wage for someone with no college education
- In (2) β_0 will correspond to the average log wage for someone with no college education and no experience

The Model: Introduction

Example 1: Let Y be log wages, EDU be years of college education, and EXP be years of experience. Prior to this we have estimated the equation

$$Y = \beta_0 + \beta_1 EDU + \epsilon. \tag{1}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 EDU + \beta_2 EXP + \epsilon. \tag{2}$$

Note that β_0, β_1 in model (1) will differ from β_0, β_1 in model (2).

- In (1) β_0 corresponds to the average log wage for someone with no college education
- In (2) β_0 will correspond to the average log wage for someone with no college education and no experience
- In (1) β_1 corresponds to the expected change in log wage for an additional year of college education

The Model: Introduction

Example 1: Let Y be log wages, EDU be years of college education, and EXP be years of experience. Prior to this we have estimated the equation

$$Y = \beta_0 + \beta_1 EDU + \epsilon. \tag{1}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 EDU + \beta_2 EXP + \epsilon. \tag{2}$$

Note that β_0, β_1 in model (1) will differ from β_0, β_1 in model (2).

- In (1) β_0 corresponds to the average log wage for someone with no college education
- In (2) β_0 will correspond to the average log wage for someone with no college education and no experience
- In (1) β_1 corresponds to the expected change in log wage for an additional year of college education
- In (2) β_1 corresponds to the expected change in log wage for an additional year of college education after controlling for years of experience

The Model: Introduction

Example 2: Let Y be the (log) final sales price of a home, $SQFT$ be the square footage of the house, and $DAYS$ be the number of days the house has been on the market. Before we estimated and interpreted the linear model:

$$Y = \beta_0 + \beta_1 SQFT + \epsilon. \tag{3}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 SQFT + \beta_2 DAYS + \epsilon \tag{4}$$

The Model: Introduction

Example 2: Let Y be the (log) final sales price of a home, $SQFT$ be the square footage of the house, and $DAYS$ be the number of days the house has been on the market. Before we estimated and interpreted the linear model:

$$Y = \beta_0 + \beta_1 SQFT + \epsilon. \tag{3}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 SQFT + \beta_2 DAYS + \epsilon \tag{4}$$

- In (3) β_0 is interpreted as the average log sales price for a home with zero square feet (an empty lot) regardless of how long it's been on the market.

The Model: Introduction

Example 2: Let Y be the (log) final sales price of a home, $SQFT$ be the square footage of the house, and $DAYS$ be the number of days the house has been on the market. Before we estimated and interpreted the linear model:

$$Y = \beta_0 + \beta_1 SQFT + \epsilon. \tag{3}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 SQFT + \beta_2 DAYS + \epsilon \tag{4}$$

- In (3) β_0 is interpreted as the average log sales price for a home with zero square feet (an empty lot) regardless of how long it's been on the market.
- In (4) β_0 is interpreted as the average log sales price for a home with zero square feet that has just entered the market

The Model: Introduction

Example 2: Let Y be the (log) final sales price of a home, $SQFT$ be the square footage of the house, and $DAYS$ be the number of days the house has been on the market. Before we estimated and interpreted the linear model:

$$Y = \beta_0 + \beta_1 SQFT + \epsilon. \tag{3}$$

Now, we will consider estimation and inference on the model

$$Y = \beta_0 + \beta_1 SQFT + \beta_2 DAYS + \epsilon \tag{4}$$

- In (3) β_0 is interpreted as the average log sales price for a home with zero square feet (an empty lot) regardless of how long it's been on the market.
- In (4) β_0 is interpreted as the average log sales price for a home with zero square feet that has just entered the market
- In (4) β_1 is interpreted as the average change in sales price for a one unit increase in square footage, holding the number of days on the market constant

The Model: Introduction

Example 3: Finally, let's return to an example from Week 1. Let Y be a measure of anxiety levels, ENG be the number of energy drinks consumed per day, and CLS be the number of courses being taken. Before we may have estimated the model:

$$Y = \beta_0 + \beta_1 ENG + \epsilon \tag{5}$$

Now, we may consider the model

$$Y = \beta_0 + \beta_1 ENG + \beta_2 CLS + \epsilon \tag{6}$$

The Model: Introduction

Example 3: Finally, let's return to an example from Week 1. Let Y be a measure of anxiety levels, ENG be the number of energy drinks consumed per day, and CLS be the number of courses being taken. Before we may have estimated the model:

$$Y = \beta_0 + \beta_1 ENG + \epsilon \tag{5}$$

Now, we may consider the model

$$Y = \beta_0 + \beta_1 ENG + \beta_2 CLS + \epsilon \tag{6}$$

- In (5) we can interpret β_0 as the average anxiety level for someone who drinks no energy drinks

The Model: Introduction

Example 3: Finally, let's return to an example from Week 1. Let Y be a measure of anxiety levels, ENG be the number of energy drinks consumed per day, and CLS be the number of courses being taken. Before we may have estimated the model:

$$Y = \beta_0 + \beta_1 ENG + \epsilon \tag{5}$$

Now, we may consider the model

$$Y = \beta_0 + \beta_1 ENG + \beta_2 CLS + \epsilon \tag{6}$$

- In (5) we can interpret β_0 as the average anxiety level for someone who drinks no energy drinks
- In (6) we can interpret β_0 as the average anxiety level for someone who drinks no energy drinks and takes no classes

The Model: Introduction

Example 3: Finally, let's return to an example from Week 1. Let Y be a measure of anxiety levels, ENG be the number of energy drinks consumed per day, and CLS be the number of courses being taken. Before we may have estimated the model:

$$Y = \beta_0 + \beta_1 ENG + \epsilon \tag{5}$$

Now, we may consider the model

$$Y = \beta_0 + \beta_1 ENG + \beta_2 CLS + \epsilon \tag{6}$$

- In (5) we can interpret β_1 as the expected change in anxiety levels for someone who drinks one more energy drink per day

The Model: Introduction

Example 3: Finally, let's return to an example from Week 1. Let Y be a measure of anxiety levels, ENG be the number of energy drinks consumed per day, and CLS be the number of courses being taken. Before we may have estimated the model:

$$Y = \beta_0 + \beta_1 ENG + \epsilon \tag{5}$$

Now, we may consider the model

$$Y = \beta_0 + \beta_1 ENG + \beta_2 CLS + \epsilon \tag{6}$$

- In (5) we can interpret β_1 as the expected change in anxiety levels for someone who drinks one more energy drink per day
- In (6) we can interpret β_1 as the expected change in anxiety levels for an additional energy drink <u>holding the number of courses being taken constant.</u>

The Model: Introduction

Example 3: Finally, let's return to an example from Week 1. Let Y be a measure of anxiety levels, ENG be the number of energy drinks consumed per day, and CLS be the number of courses being taken. Before we may have estimated the model:

$$Y = \beta_0 + \beta_1 ENG + \epsilon \tag{5}$$

Now, we may consider the model

$$Y = \beta_0 + \beta_1 ENG + \beta_2 CLS + \epsilon \tag{6}$$

- In (5) we can interpret β_1 as the expected change in anxiety levels for someone who drinks one more energy drink per day
- In (6) we can interpret β_1 as the expected change in anxiety levels for an additional energy drink <u>holding the number of courses being taken constant.</u>

Question: How may we expect the signs/magnitutes of the parameters to change when going from model (5) to model (6)?

The Model: Questions

Questions?

Table of Contents

Estimation: Introduction

Before, in single linear regression when we were interested in the population line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{b_0, b_1} \mathbb{E}\left[(Y - b_0 - b_1 X)^2\right],$$

we estimated them by finding the line of best fit through our sample $\{Y_i, X_i\}_{i=1}^n$:

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

Estimation: Introduction

Before, in single linear regression when we were interested in the population line of best fit parameters

$$\beta_0, \beta_1 = \arg\min_{b_0, b_1} \mathbb{E}\left[(Y - b_0 - b_1 X)^2\right],$$

we estimated them by finding the line of best fit through our sample $\{Y_i, X_i\}_{i=1}^n$:

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

$\rightarrow$ Have to estimate these parameters using the sample because we don't know the population distribution of (Y, X)

Now, we are interested in the population line of best fit parameters:

$$\beta_0, \beta_1, \ldots, \beta_p = \arg \min_{b_0, b_1, \ldots, b_p} \mathbb{E}\left[(Y - b_0 - b_1 X_1 - \cdots - b_p X_p)^2\right].$$

Question: How should we estimate these using our sample $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}_{i=1}^n$?

Now, we are interested in the population line of best fit parameters:

$$\beta_0, \beta_1, \ldots, \beta_p = \arg \min_{b_0, b_1, \ldots, b_p} \mathbb{E}\left[(Y - b_0 - b_1 X_1 - \cdots - b_p X_p)^2\right].$$

Question: How should we estimate these using our sample $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}_{i=1}^n$?

Estimate $\beta_0, \ldots, \beta_p$ by finding the line of best fit through our sample:

$$\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p = \arg \min_{b_0, b_1, \ldots, b_p} \frac{1}{n} \sum_{i=1}^n (Y_i - b_0 - b_1 X_{1,i} - \cdots - b_p X_{p,i})^2.$$

Estimation: Introduction

Taking first order conditions for $\hat{\beta}_0, \ldots, \hat{\beta}_p$ above gives us

$$\frac{\partial}{\partial b_0} : \frac{1}{n}\sum_{i=1}^{n} \overbrace{(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i})}^{\hat{\epsilon}_i} = 0$$

$$\frac{\partial}{\partial b_1} : \frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i})X_{1,i} = 0$$

$$\vdots$$

$$\frac{\partial}{\partial b_p} : \frac{1}{n}\sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i})X_{p,i} = 0$$

Estimation: Introduction

Taking first order conditions for $\hat{\beta}_0, \ldots, \hat{\beta}_p$ above gives us

$$\frac{\partial}{\partial b_0} : \frac{1}{n} \sum_{i=1}^{n} \overbrace{(Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i})}^{\hat{\epsilon}_i} = 0$$

$$\frac{\partial}{\partial b_1} : \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i}) X_{1,i} = 0$$

$$\vdots$$

$$\frac{\partial}{\partial b_p} : \frac{1}{n} \sum_{i=1}^{n} (Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i}) X_{p,i} = 0$$

This gives us $p+1$ linear equations to solve for our $p+1$ parameters. Computers can solve these very quickly, but the explicit formulas for $\hat{\beta}_0, \hat{\beta}_1, \ldots, \hat{\beta}_p$ become very cumbersome if we don't use linear algebra notation.

Estimation: Aside

Quickly, it is useful to note the following implication from the first order conditions for $\hat{\beta}_0, \ldots, \hat{\beta}_p$. Define $\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \cdots - \hat{\beta}_p X_{p,i}$. Then

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

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i X_{1,i} = 0$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i X_{2,i} = 0$$

$$\vdots$$

$$\frac{1}{n}\sum_{i=1}^{n} \hat{\epsilon}_i X_{p,i} = 0$$

Just as in single linear regression, however, the solutions for $\hat{\beta}_0, \ldots, \hat{\beta}_p$ depend on the data. That is $\hat{\beta}_0, \ldots, \hat{\beta}_p$ are functions of our sample $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}_{i=1}^n$.

Estimation: Asymptotic Distribution

Just as in single linear regression, however, the solutions for $\hat{\beta}_0, \ldots, \hat{\beta}_p$ depend on the data. That is $\hat{\beta}_0, \ldots, \hat{\beta}_p$ are functions of our sample $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}_{i=1}^n$.

- If we collected a different sample $\{Y_i, X_{1,i}, \ldots, X_{2,i}\}_{i=1}^n$, we would get different values for our estimators $\hat{\beta}_0, \ldots, \hat{\beta}_p$.

Estimation: Asymptotic Distribution

Just as in single linear regression, however, the solutions for $\hat{\beta}_0, \ldots, \hat{\beta}_p$ depend on the data. That is $\hat{\beta}_0, \ldots, \hat{\beta}_p$ are functions of our sample $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}_{i=1}^n$.

- If we collected a different sample $\{Y_i, X_{1,i}, \ldots, X_{2,i}\}_{i=1}^n$, we would get different values for our estimators $\hat{\beta}_0, \ldots, \hat{\beta}_p$.

For hypothesis testing we would still like to know the (approximate) distribution of our estimates $\hat{\beta}_0, \ldots, \hat{\beta}_p$. This will be useful later on as we'd like to calculate objects such as

$$\Pr(|\hat{\beta}_1| > 5 | \beta_1 = -2).$$

In order for the estimates $\hat{\beta}_0, \ldots, \hat{\beta}_p$ to have a stable asymptotic distribution and to converge to the true parameters $\beta_0, \ldots, \beta_p$, we need to make some (light) assumptions about the underlying distribution of $(Y, X_1, \ldots, X_p)$ from which our sample is drawn.

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
  - Needed to make sure that we are making inferences on the correct population

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
  - Question: When would this be violated?

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
- Rank Condition: The right hand side variables $X_1, \ldots, X_p$ are not linearly dependent, i.e we cannot write

$$a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = 0$$

for some constants $a_1, \ldots, a_p$ with at least one $a_k \neq 0$.

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
- Rank Condition: The right hand side variables $X_1, \ldots, X_p$ are not linearly dependent, i.e we cannot write

$$a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = 0$$

  for some constants $a_1, \ldots, a_p$ with at least one $a_k \neq 0$.

  - If this is violated then we can write one random variable as a linear combination of the other ones.

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
- Rank Condition: The right hand side variables $X_1, \ldots, X_p$ are not linearly dependent, i.e we cannot write

$$a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = 0$$

for some constants $a_1, \ldots, a_p$ with at least one $a_k \neq 0$.

  - To see why this is problematic, suppose that we could write $X_1 = 2X_2$. Then these two linear models are equivalent

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon \iff Y = \beta_0 + (2\beta_1 + \beta_2) X_2 + \epsilon.$$

The "line of best fit" solution is then not unique. We can achieve the same fit by setting the coeffecient on X_1 to be $beta_1$ and the coeffecient on X_2 to be β_2 or by setting the coeffecient on X_1 to be zero and the coeffecient on X_2 to be $2\beta_1 + \beta_2$.

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
- Rank Condition: The right hand side variables $X_1, \ldots, X_p$ are not linearly dependent, i.e we cannot write

$$a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = 0$$

for some constants $a_1, \ldots, a_p$ with at least one $a_k \neq 0$.

- Homoskedasticity: $\mathrm{Var}(\epsilon | X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) = \sigma_\epsilon^2$ for all possible $(x_1, \ldots, x_p)$.

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \dots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \dots, X_p)$
- Rank Condition: The right hand side variables $X_1, \dots, X_p$ are not linearly dependent, i.e we cannot write

$$a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = 0$$

for some constants $a_1, \dots, a_p$ with at least one $a_k \neq 0$.

- Homoskedasticity: $\mathrm{Var}(\epsilon | X_1 = x_1, X_2 = x_2, \dots, X_p = x_p) = \sigma_\epsilon^2$ for all possible $(x_1, \dots, x_p)$.
    - Like before we can and should relax this. It is not very important to our results but it makes some closed form equations simpler later one.

Estimation: Asymptotic Distribution

Assumptions needed for valid inference:

- Random Sampling: The data $\{Y_i, X_{1,i}, \ldots, X_{p,i}\}$ is independently and identically sampled from the population distribution $(Y, X_1, \ldots, X_p)$
- Rank Condition: The right hand side variables $X_1, \ldots, X_p$ are not linearly dependent, i.e we cannot write

$$a_1 X_1 + a_2 X_2 + \cdots + a_p X_p = 0$$

for some constants $a_1, \ldots, a_p$ with at least one $a_k \neq 0$.

- Homoskedasticity: $\mathrm{Var}(\epsilon | X_1 = x_1, X_2 = x_2, \ldots, X_p = x_p) = \sigma_\epsilon^2$ for all possible $(x_1, \ldots, x_p)$.

And that's it! Really only need Random Sampling and Rank Condition.

Estimation: Asymptotic Distribution

Under the assumptions Random Sampling and Rank Condition we get the following result for any $\hat{\beta}_k$, $k = 0, 1, \ldots, p$.

Approximately, for large n:

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\beta_k}/\sqrt{n}} \sim N(0,1) \iff \hat{\beta}_k \sim N\Bigg(\beta_k,\ \underbrace{\sigma^2_{\beta_k}/n}_{=\mathrm{Var}(\hat{\beta}_k)}\Bigg).$$

Estimation: Asymptotic Distribution

Under the assumptions Random Sampling and Rank Condition we get the following result for any $\hat{\beta}_k$, $k = 0, 1, \ldots, p$.

Approximately, for large n:

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\beta_k}/\sqrt{n}} \sim N(0,1) \iff \hat{\beta}_k \sim N\Bigg(\beta_k, \underbrace{\sigma^2_{\beta_k}/n}_{=\mathrm{Var}(\hat{\beta}_k)}\Bigg).$$

- The assumption Homoskedasticity simply changes the form of $\sigma^2_{\beta_k}$ and thus it's estimator $\hat{\sigma}^2_{\beta_k}$.

Estimation: Asymptotic Distribution

Under the assumptions Random Sampling and Rank Condition we get the following result for any $\hat{\beta}_k$, $k = 0, 1, \ldots, p$.

Approximately, for large n:

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\beta_k}/\sqrt{n}} \sim N(0,1) \iff \hat{\beta}_k \sim N\Bigg(\beta_k, \underbrace{\sigma^2_{\beta_k}/n}_{=\mathrm{Var}(\hat{\beta}_k)}\Bigg).$$

- The assumption Homoskedasticity simply changes the form of $\sigma^2_{\beta_k}$ and thus it's estimator $\hat{\sigma}^2_{\beta_k}$.
- Unlike in single linear regression we will not go over a general form for $\hat{\sigma}_{\beta_k}$
  - Typically, all that you need to know is that $\hat{\sigma}_{\beta_k}$ (or the standard error, $\hat{\sigma}_{\beta_k}/\sqrt{n}$ or the variance $\hat{\sigma}^2_{\beta_k}/n$) will either be given to us directly or found in R output.

Estimation: Asymptotic Distribution

Under the assumptions Random Sampling and Rank Condition we get the following result for any $\hat{\beta}_k$, $k = 0, 1, \ldots, p$.

Approximately, for large n:

$$\frac{\hat{\beta}_k - \beta_k}{\hat{\sigma}_{\beta_k}/\sqrt{n}} \sim N(0,1) \iff \hat{\beta}_k \sim N\Bigg(\beta_k, \underbrace{\sigma^2_{\beta_k}/n}_{=\text{Var}(\hat{\beta}_k)}\Bigg).$$

- The assumption Homoskedasticity simply changes the form of $\sigma^2_{\beta_k}$ and thus it's estimator $\hat{\sigma}^2_{\beta_k}$.
- Unlike in single linear regression we will not go over a general form for $\hat{\sigma}_{\beta_k}$
  - Typically, all that you need to know is that $\hat{\sigma}_{\beta_k}$ (or the standard error, $\hat{\sigma}_{\beta_k}/\sqrt{n}$ or the variance $\hat{\sigma}^2_{\beta_k}/n$) will either be given to us directly or found in R output.
- In addition, we will be able to estimate the asymptotic covariance between any two estimates $\hat{\beta}_j, \hat{\beta}_k$ for $j, k = 0, 1, \ldots, p$.

Estimation: Asymptotic Variance

To consolidate notation, the variances and covariances are often presented as a Variance-Covariance matrix. For example, when $p = 2$ the variance covariance matrix looks like

$$\text{Cov}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_0) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) \end{pmatrix}.$$

Estimation: Asymptotic Variance

To consolidate notation, the variances and covariances are often presented as a Variance-Covariance matrix. For example, when $p = 2$ the variance covariance matrix looks like

$$\operatorname{Cov}(\hat{\beta}) = \begin{pmatrix} \operatorname{Var}(\hat{\beta}_0) & \operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \operatorname{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \operatorname{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \operatorname{Var}(\hat{\beta}_1) & \operatorname{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \operatorname{Cov}(\hat{\beta}_2, \hat{\beta}_0) & \operatorname{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \operatorname{Var}(\hat{\beta}_2) \end{pmatrix}.$$

- Note that $\operatorname{Cov}(X, Y) = \operatorname{Cov}(Y, X)$ so this is a symmetric matrix

Estimation: Asymptotic Variance

To consolidate notation, the variances and covariances are often presented as a Variance-Covariance matrix. For example, when $p = 2$ the variance covariance matrix looks like

$$\text{Cov}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_0) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) \end{pmatrix}.$$

- Note that $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ so this is a symmetric matrix
- Also note that $\text{Var}(X) = \text{Cov}(X, X)$ which is why we sometimes just call this the Covariance matrix.

Estimation: Asymptotic Variance

To consolidate notation, the variances and covariances are often presented as a Variance-Covariance matrix. For example, when $p = 2$ the variance covariance matrix looks like

$$\text{Cov}(\hat{\beta}) = \begin{pmatrix} \text{Var}(\hat{\beta}_0) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_1) & \text{Cov}(\hat{\beta}_0, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_1, \hat{\beta}_0) & \text{Var}(\hat{\beta}_1) & \text{Cov}(\hat{\beta}_1, \hat{\beta}_2) \\ \text{Cov}(\hat{\beta}_2, \hat{\beta}_0) & \text{Cov}(\hat{\beta}_2, \hat{\beta}_1) & \text{Var}(\hat{\beta}_2) \end{pmatrix}.$$

- Note that $\text{Cov}(X, Y) = \text{Cov}(Y, X)$ so this is a symmetric matrix
- Also note that $\text{Var}(X) = \text{Cov}(X, X)$ which is why we sometimes just call this the Covariance matrix.
- In general the Variance-Covariance matrix will be a $(p + 1) \times (p + 1)$ matrix (one dimension for each of the slope coefficients and the intercept).

Estimation: Asymptotic Variance

Question: What influences the asymptotic variance?

In order to get some intuition for this, we will go over a particular example when $p = 2$. That is when we want to estimate the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

Question: What influences the asymptotic variance?

In order to get some intuition for this, we will go over a particular example when $p = 2$. That is when we want to estimate the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

In this case, we will be able to get some simple closed form expressions for $\sigma^2_{\beta_1}$ and $\sigma^2_{\beta_2}$.

Estimation: Asymptotic Variance

Question: What influences the asymptotic variance?

In order to get some intuition for this, we will go over a particular example when $p = 2$. That is when we want to estimate the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon.$$

In this case, we will be able to get some simple closed form expressions for $\sigma^2_{\beta_1}$ and $\sigma^2_{\beta_2}$.

- Will provide some insight into what drives the asymptotic variance

Estimation: Asymptotic Variance

Before doing so, let's review the correlation coeffecient. Recall that for two random variables X_1 and X_2 the correlation coeffecient ρ_{12} is defined

$$\rho_{12} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}}.$$

Estimation: Asymptotic Variance

Before doing so, let's review the correlation coeffecient. Recall that for two random variables X_1 and X_2 the correlation coeffecient ρ_{12} is defined

$$\rho_{12} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}}.$$

The correlation coeffecient is a measure of the linear dependence between X_1 and X_2

- If $\rho_{12} = 1$ then X_1 and X_2 are perfectly linearly dependent, that is $X_1 = cX_2$ for some constant $c \neq 0$

Estimation: Asymptotic Variance

Before doing so, let's review the correlation coeffecient. Recall that for two random variables X_1 and X_2 the correlation coeffecient ρ_{12} is defined

$$\rho_{12} = \frac{\mathrm{Cov}(X_1, X_2)}{\sqrt{\mathrm{Var}(X_1)}\sqrt{\mathrm{Var}(X_2)}}.$$

The correlation coeffecient is a measure of the linear dependence between X_1 and X_2

- If $\rho_{12} = 1$ then X_1 and X_2 are perfectly linearly dependent, that is $X_1 = cX_2$ for some constant $c \neq 0$
- If $\rho_{12} = 0$ then X_1 and X_2 are have no linear dependence, that is $\mathrm{Cov}(X_1, X_2) = 0$.

Estimation: Asymptotic Variance

With this in mind, the asymptotic variance (under homoskedasticity) $\hat{\sigma}^2_{\beta_1}$ for β_1 in the linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

is given

$$\sigma^2_{\beta_1} = \frac{\sigma^2_\epsilon}{(1-\rho^2_{12})\sigma^2_{X_1}} \iff \sqrt{n}(\hat{\beta}_1 - \beta_1) \sim N(0, \sigma^2_{\beta_1}),$$

where ρ_{12} is the correlation coeffecient between X_1 and X_2.

Notice:

Estimation: Asymptotic Variance

With this in mind, the asymptotic variance (under homoskedasticity) $\hat{\sigma}^2_{\beta_1}$ for β_1 in the linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

is given

$$\sigma^2_{\beta_1} = \frac{\sigma^2_\epsilon}{(1-\rho^2_{12})\sigma^2_{X_1}} \iff \sqrt{n}(\hat{\beta}_1 - \beta_1) \sim N(0, \sigma^2_{\beta_1}),$$

where ρ_{12} is the correlation coeffecient between X_1 and X_2.

Notice:

- As before $\mathrm{Var}(\hat{\beta}_1) = \sigma^2_{\beta_1}/n$ is decreasing with n, $\hat{\beta}_1 \to \beta_1$ as $n \to \infty$.

Estimation: Asymptotic Variance

With this in mind, the asymptotic variance (under homoskedasticity) $\hat{\sigma}^2_{\beta_1}$ for β_1 in the linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

is given

$$\sigma^2_{\beta_1} = \frac{\sigma^2_\epsilon}{(1-\rho^2_{12})\sigma^2_{X_1}} \iff \sqrt{n}(\hat{\beta}_1 - \beta_1) \sim N(0, \sigma^2_{\beta_1}),$$

where ρ_{12} is the correlation coeffecient between X_1 and X_2.

Notice:

- As before $\mathrm{Var}(\hat{\beta}_1) = \sigma^2_{\beta_1}/n$ is decreasing with n, $\hat{\beta}_1 \to \beta_1$ as $n \to \infty$.
- As before $\sigma^2_{\beta_1}$ is decreasing with σ_{ϵ^2} and increasing with $\sigma^2_{X_1}$
  - σ^2_ϵ : If points are closer to the line it is easier to make out the line
  - $\sigma^2_{X_1}$: If points are more spread out, it is easier to make out the line

Estimation: Asymptotic Variance

With this in mind, the asymptotic variance (under homoskedasticity) $\hat{\sigma}^2_{\beta_1}$ for β_1 in the linear model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \epsilon,$$

is given

$$\sigma^2_{\beta_1} = \frac{\sigma^2_\epsilon}{(1-\rho^2_{12})\sigma^2_{X_1}} \iff \sqrt{n}(\hat{\beta}_1 - \beta_1) \sim N(0, \sigma^2_{\beta_1}),$$

where ρ_{12} is the correlation coeffecient between X_1 and X_2.

Notice:

- As before $\mathrm{Var}(\hat{\beta}_1) = \sigma^2_{\beta_1}/n$ is decreasing with n, $\hat{\beta}_1 \to \beta_1$ as $n \to \infty$.
- As before $\sigma^2_{\beta_1}$ is decreasing with σ_{ϵ^2} and increasing with $\sigma^2_{X_1}$
- However, now we see that the variance $\sigma^2_{\beta_1}$ is increasing also as $\rho_{12} \uparrow 1$.
    - Intuition: If X_1 and X_2 are highly correlated, it is difficult to parse out the relationship of X_1 on Y holding X_2 constant.

Estimation: Asymptotic Variance

To estimate $\sigma^2_{\beta_1}$ we can estimate each of it's components.

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_{\epsilon}}{(1 - \hat{\rho}^2_{12})\hat{\sigma}^2_{X_1}}.$$

Estimation: Asymptotic Variance

To estimate $\sigma^2_{\beta_1}$ we can estimate each of it's components.

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_\epsilon}{(1-\hat{\rho}^2_{12})\hat{\sigma}^2_{X_1}}.$$

- For $\hat{\sigma}^2_\epsilon$ generate the estimated residuals $\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \hat{\beta}_2 X_{2,i}$ and calculate the sample variance of the estimated residuals:

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

  - Recall that by the first order conditions for $\hat{\beta}_0$, $\frac{1}{n}\sum_{i=1}^{n}\hat{\epsilon}_i = 0$ so that $\bar{\hat{\epsilon}} = 0$

Estimation: Asymptotic Variance

To estimate $\sigma^2_{\beta_1}$ we can estimate each of it's components.

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_\epsilon}{(1 - \hat{\rho}^2_{12})\hat{\sigma}^2_{X_1}}.$$

- For $\hat{\sigma}^2_\epsilon$ generate the estimated residuals $\hat{\epsilon}_i = Y_i - \hat{\beta}_0 - \hat{\beta}_1 X_{1,i} - \hat{\beta}_2 X_{2,i}$ and calculate the sample variance of the estimated residuals:

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

  - Recall that by the first order conditions for $\hat{\beta}_0$, $\frac{1}{n}\sum_{i=1}^n \hat{\epsilon}_i = 0$ so that $\bar{\hat{\epsilon}} = 0$
- For $\hat{\sigma}^2_{X_1}$ calculate the sample variance of X_1

$$\hat{\sigma}^2_{X_1} = \frac{1}{n}\sum_{i=1}^{n}(X_{1,i} - \bar{X}_1)^2.$$

Estimation: Asymptotic Variance

To estimate $\sigma_{\beta_1}^2$ we can estimate each of it's components.

$$\hat{\sigma}_{\beta_1}^2 = \frac{\hat{\sigma}_\epsilon^2}{(1 - \hat{\rho}_{12}^2)\hat{\sigma}_{X_1}^2}.$$

- To estimate $\hat{\rho}_{12}^2$ recall that

$$\rho_{12} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \implies \hat{\rho}_{12} = \frac{\widehat{\text{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

Estimation: Asymptotic Variance

To estimate $\sigma^2_{\beta_1}$ we can estimate each of it's components.

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_{\epsilon}}{(1 - \hat{\rho}^2_{12})\hat{\sigma}^2_{X_1}}.$$

- To estimate $\hat{\rho}^2_{12}$ recall that

$$\rho_{12} = \frac{\text{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \implies \hat{\rho}_{12} = \frac{\widehat{\text{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

  - We have already covered how to estiamte $\hat{\sigma}^2_{X_1}$. Estimating $\hat{\sigma}^2_{X_2}$ follows the same formula

$$\hat{\sigma}^2_{X_2} = \frac{1}{n}\sum_{i=1}^{n}(X_{2,i} - \bar{X}_2)^2.$$

    Then, take square roots $\hat{\sigma}_{X_1} = \sqrt{\hat{\sigma}^2_{X_1}}$ and $\hat{\sigma}_{X_2} = \sqrt{\hat{\sigma}^2_{X_2}}$.

Estimation: Asymptotic Variance

To estimate $\sigma^2_{\beta_1}$ we can estimate each of it's components.

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_\epsilon}{(1-\hat{\rho}^2_{12})\hat{\sigma}^2_{X_1}}.$$

- To estimate $\hat{\rho}^2_{12}$ recall that

$$\rho_{12} = \frac{\mathrm{Cov}(X_1, X_2)}{\sigma_{X_1}\sigma_{X_2}} \implies \hat{\rho}_{12} = \frac{\widehat{\mathrm{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

  - We have already covered how to estiamte $\hat{\sigma}^2_{X_1}$. Estimating $\hat{\sigma}^2_{X_2}$ follows the same formula

$$\hat{\sigma}^2_{X_2} = \frac{1}{n}\sum_{i=1}^{n}(X_{2,i} - \bar{X}_2)^2.$$

    Then, take square roots $\hat{\sigma}_{X_1} = \sqrt{\hat{\sigma}^2_{X_1}}$ and $\hat{\sigma}_{X_2} = \sqrt{\hat{\sigma}^2_{X_2}}$.

  - To estimate the covariance note $\mathrm{Cov}(X_1, X_2) = \mathbb{E}[(X_1 - \mu_{X_1})(X_2 - \mu_{X_2})]$ so

$$\widehat{\mathrm{Cov}}(X_1, X_2) = \frac{1}{n}\sum_{i=1}^{n}(X_{1,i} - \bar{X}_1)(X_{2,i} - \bar{X}_2).$$

Estimation: Asymptotic Variance

Let's see an example of this. Suppose we are interested in the joint effect of smoking heavily and drinking heavily on liver failure.

That is let $Y \in \{0, 1\}$ denote liver failure, $X_1 \in \{0, 1\}$ denote being a heavy smoker, and $X_2 \in \{0, 1\}$ denote being a heavy drinker and suppose we want to estimate the model

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2.$$

After collecting a sample of size $n = 64$ we estimate $\hat{\sigma}_\epsilon^2 = 0.25$, $\hat{\sigma}_{X_1}^2 = 0.1$, and $\hat{\rho}_{12} = 0.5$, where

$$\hat{\rho}_{12} = \frac{\widehat{\text{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

Question: What is the <u>standard error</u> of $\hat{\beta}_1$?

Estimation: Asymptotic Variance

After collecting a sample of size $n = 64$ we estimate $\hat{\sigma}_\epsilon^2 = 0.25$, $\hat{\sigma}_{X_1}^2 = 0.1$, and $\hat{\rho}_{12} = 0.5$, where

$$\hat{\rho}_{12} = \frac{\widehat{\mathrm{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

Question: What is the standard error of $\hat{\beta}_1$?

Answer: Recall that the standard error is given $\hat{\sigma}_{\beta_1}/\sqrt{n}$. Using the above we get that

$$\hat{\sigma}_{\beta_1}^2 = \frac{\hat{\sigma}_\epsilon^2}{(1 - \hat{\rho}_{12})\hat{\sigma}_{X_1}^2} = \frac{0.25}{(1 - 0.25)0.1} = \frac{10}{3}.$$

The standard error is then $\hat{\sigma}_{\beta_1}/\sqrt{n} = \sqrt{10/3}/\sqrt{64} \approx 0.228$

Estimation: Asymptotic Variance

Now suppose that after collecting a sample of size $n = 100$ we estimate $\hat{\sigma}_\epsilon^2 = 0.25$, $\hat{\sigma}_{X_1}^2 = 0.1$. This time however, we estimate $\hat{\rho}_{12} = 0.75$, where

$$\hat{\rho}_{12} = \frac{\widehat{\text{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

Question: In this case, what is the standard error of $\hat{\beta}_1$?

Estimation: Asymptotic Variance

Now suppose that after collecting a sample of size $n = 100$ we estimate $\hat{\sigma}^2_\epsilon = 0.25$, $\hat{\sigma}^2_{X_1} = 0.1$. This time however, we estimate $\hat{\rho}_{12} = 0.75$, where

$$\hat{\rho}_{12} = \frac{\widehat{\text{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

Question: In this case, what is the standard error of $\hat{\beta}_1$?

Answer: Using the formula above

$$\hat{\sigma}^2_{\beta_1} = \frac{\hat{\sigma}^2_\epsilon}{(1 - \hat{\rho}_{12})^2\hat{\sigma}^2_{X_1}} = \frac{0.25}{(1 - 0.5625)0.1} =\approx 5.714.$$

The standard error is then $\hat{\sigma}_{\beta_1}/\sqrt{n} \approx \sqrt{5.714}/\sqrt{100} = 0.239$.

Estimation: Asymptotic Variance

Now suppose that after collecting a sample of size $n = 100$ we estimate $\hat{\sigma}_\epsilon^2 = 0.25$, $\hat{\sigma}_{X_1}^2 = 0.1$. This time however, we estimate $\hat{\rho}_{12} = 0.75$, where

$$\hat{\rho}_{12} = \frac{\widehat{\text{Cov}}(X_1, X_2)}{\hat{\sigma}_{X_1}\hat{\sigma}_{X_2}}.$$

Question: In this case, what is the standard error of $\hat{\beta}_1$?

Answer: Using the formula above

$$\hat{\sigma}_{\beta_1}^2 = \frac{\hat{\sigma}_\epsilon^2}{(1 - \hat{\rho}_{12})^2 \hat{\sigma}_{X_1}^2} = \frac{0.25}{(1 - 0.5625)0.1} =\approx 5.714.$$

The standard error is then $\hat{\sigma}_{\beta_1}/\sqrt{n} \approx \sqrt{5.714}/\sqrt{100} = 0.239.$

Notice that the standard error is larger now than it was when $n = 64$, despite the fact that our sample size has grown by about 50%!

Table of Contents

Testing single hypothesis about the coeffecients of our regression or linear combinations of coeffecients follows the same procedure.

If we recall, this procedure consists of constructing a test statistic of the form

$$t^* = \frac{\text{Estimator} - \text{Null Hypothesis Value}}{\text{Standard Error of Estimator}},$$

and then either computing a p-value or comparing the test statistic directly to a quantile of the standard normal distribution.

Table of Contents