1 The Econometrics of the Simple Regression Model

- Multiple regression model with k explanatory variables:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + .. + \beta_k X_{ki} + \varepsilon_i,$$

where i subscripts to denote individual observations and we have $i = 1, .., N$ observations.

- In econometrics, lots of uncertainty.

- Uncertain what the regression coefficients, $\alpha, \beta_1, .., \beta_k$ are (and, hence, have to estimate them).

- Uncertain whether a hypothesis (e.g. $\beta_j = 0$) is true (and, hence, have to derive hypothesis testing procedures).

- We are uncertain about what future values of Y might be (and, hence, have to derive procedures for forecasting).

- Probability provides us with a language and a formal structure for dealing with uncertainty.

- In this chapter, we will use probability to do some key statistical derivations.

- To keep formulae simple, will work with simple regression model (i.e. regression model with one explanatory variable) with no intercept:

$$y_i = \beta X_i + \varepsilon_i$$

where $i = 1, .., N$ and X_i is a scalar.

- Derivations for multiple regression model are conceptually similar but formulae get complicated (use of matrix algebra usually involved)

1.1 A Review of Basic Concepts in Probability in the Context of the Regression Model

- See Appendix B for details, here we present basic ideas informally.

- Assume Y is a random variable.

- Regression model provides description about what probable values for the dependent variable are.

- E.g. Y is the price of a house and X is a size of house.

- What if you knew that $X = 5000$ square feet (a typical value in our data set), but did not know Y

- A house with $X = 5000$ might sell for roughly \$70,000 or \$60,000 or \$50,000 (which are typical values in our data set), but it will not sell for \$1,000 (far too cheap) or \$1,000,000 (far too expensive).

- Econometricians use probability density functions (p.d.f.) to summarize which are plausible and which are implausible values for the house.

- Figure 3.1 is example of a p.d.f.: tells you range of plausible values which Y might take when $X = 5,000$.

- Figure 3.1 a Normal distribution

- Bell-shaped curve. The curve is highest for the most plausible values that the house price might take.

- Chapter 1 introduced the ideas of a mean (or expected value) and variance.

- The mean is the "average" or "typical" value of a variable

- Variance as being a measure of how dispersed a variable is.

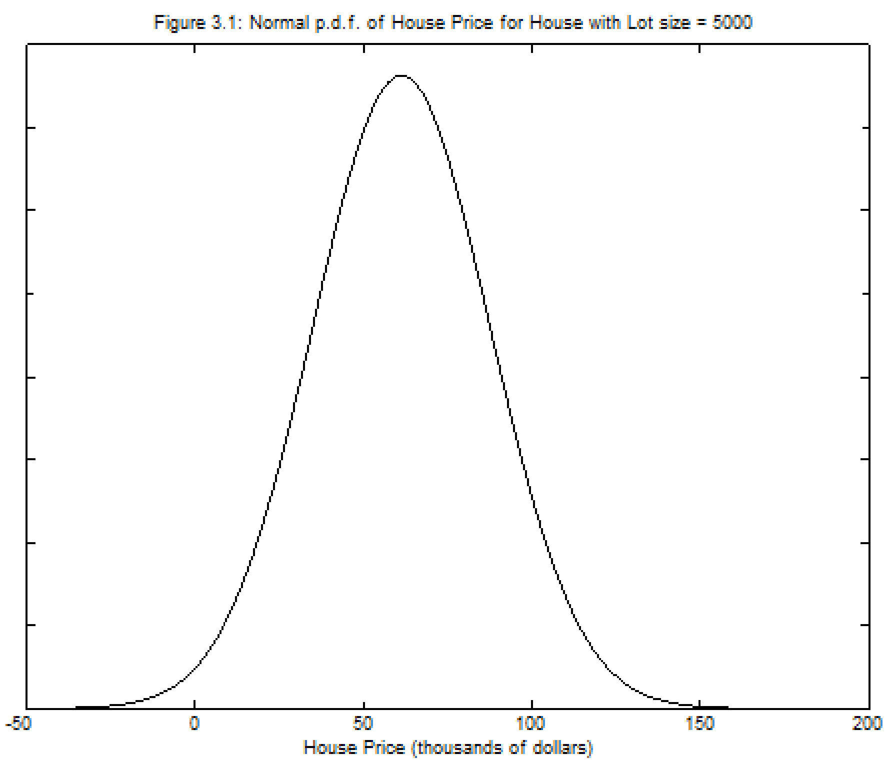
- The exact shape of any Normal distribution depends on its mean and its variance.

- "Y is a random variable which has a Normal p.d.f. with mean μ and variance σ^2" is written:

$$Y \sim N\left(\mu, \sigma^2\right)$$

- Figure 3.1 has $\mu = 61.153 \rightarrow \$61,153$ is the mean, or average, value for a house with a lot size of $5,000$ square feet.

- $\sigma^2 = 683.812$ (not much intuitive interpretation other than it reflects dispersion — range of plausible values)



- P.d.f.s measure uncertainty about a random variable since areas under the curve defined by the p.d.f. are probabilities.

- E.g. Figure 3.2. The area under the curve between the points 60 and 100 is shaded in.

- Shaded area is probability that that the house is worth between $\$60,000$ and $\$100,000$.

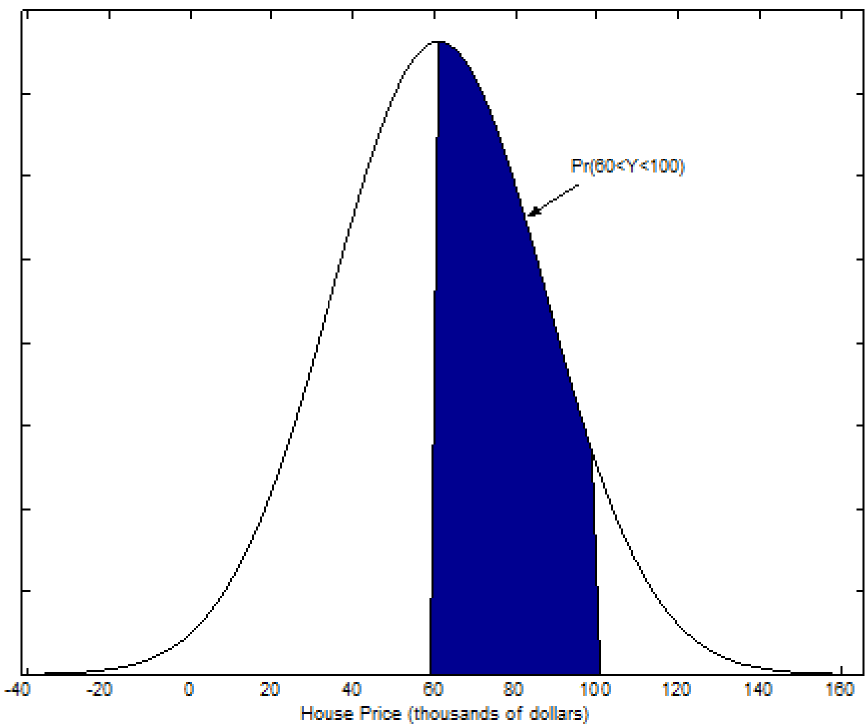
- This probability is 45% and can be written as:

$$\Pr(60 \leq Y \leq 100) = 0.45.$$

- Normal probabilities can be calculated using statistical tables (or econometrics software packages).

- By definition, the entire area under any p.d.f. is 1.

Figure 3.2: Normal p.d.f. of House Price for House with Lot size = 5000



1.1.1 Expected Value and Variance

- In this book, we are repeatedly using expected value and variance operators.

- Best way to learn these is through reading/doing problem sets.

- The *expected value* of X, denoted $E(X)$, can be interpreted as an average or typical value that might occur.

- Expected value is also called the *mean*, often denoted by the symbol μ. Thus, $\mu \equiv E(X)$.

- The *variance*

$$var(X) = E\left[(X-\mu)^2\right] = E\left(X^2\right) - \mu^2$$

- The *standard deviation* is the square root of the variance.

- Variance and standard deviation are commonly-used measures of dispersion of a random variable.

- *Covariance* between two random variables, X and Y, defined as:

$$cov\left(X,Y\right)=E\left(XY\right)-E\left(X\right)E\left(Y\right).$$

- Covariance best motivated through *correlation* between X and Y:

$$corr\left(X,Y\right)=\frac{cov\left(X,Y\right)}{\sqrt{var\left(X\right)var\left(Y\right)}}.$$

- Correlation is degree of association between two random variables. It satisfies $-1 \leq corr(X, Y) \leq 1$ with larger positive/negative values indicating stronger positive/negative relationships between X and Y.

- If X and Y are independent, then $corr(X, Y) = 0$.

1.1.2 Properties of Expected Values and Variance

If X and Y are two random variables and a and b are constants, then:

1. $E(aX + bY) = aE(X) + bE(Y)$
2. $var(Y) = E(Y) - [E(Y)]^2$
3. $var(aY) = a^2 var(Y)$
4. $var(a + Y) = var(Y)$
5. $cov(X, Y) = E(XY) - E(X)E(Y)$
6. $var(aX + bY) = a^2 var(X) + b^2 var(Y) + 2ab\,cov(X, Y)$

7. $E(XY) \neq E(X)E(Y)$ unless $cov(X,Y) = 0$.

8. If X and Y Normally distributed then $a\varepsilon_i + b\varepsilon_2$ is also Normal. "Linear combinations of Normals are Normal".

These properties generalize to the case of many random variables

1.1.3 Using Normal Statistical Tables

- Table for *standard Normal distribution* – i.e. $N(0,1)$ – is in textbook.

- Can use $N(0,1)$ tables to figure out probabilities for the $N\left(\mu,\sigma^2\right)$ for any μ and σ^2.

- If $Y \sim N\left(\mu,\sigma^2\right)$, what are mean and variance of

$$Z = \frac{Y-\mu}{\sigma}.$$

- As an example of a proof using properties of expected value operator:

$$\begin{aligned} E(Z) &= E\left(\frac{Y-\mu}{\sigma}\right) \\ &= \frac{E(Y-\mu)}{\sigma} \\ &= \frac{E(Y)-\mu}{\sigma} \\ &= \frac{\mu-\mu}{\sigma} = 0. \end{aligned}$$

- As an example of a proof using properties of variance:

$$\begin{aligned} var(Z) &= var\left(\frac{Y-\mu}{\sigma}\right) \\ &= \frac{var(Y-\mu)}{\sigma^2} \\ &= \frac{var(Y)}{\sigma^2} = \frac{\sigma^2}{\sigma^2} = 1. \end{aligned}$$

- Thus, Z is $N(0,1)$ and we can use our statistical tables

- Z is often referred to as a *Z-score*.

- *For any random variable, if you subtract off its mean and divide by standard deviation always get a new random variable with mean zero and variance one*

- Example: In Figure 3.2 how did we obtain

$$\Pr(60 \leq Y \leq 100) = 0.45$$

- Remember Figure 3.2 has $Y \sim N(61.153, 683.812)$.

$$\begin{aligned}
&\Pr(60 \leq Y \leq 100) \\
&= \Pr\left(\frac{60-\mu}{\sigma} \leq \frac{Y-\mu}{\sigma} \leq \frac{100-\mu}{\sigma}\right) \\
&= \Pr\left(\frac{60-61.153}{\sqrt{683.812}} \leq \frac{Y-61.153}{\sqrt{683.812}} \leq \frac{100-61.153}{\sqrt{683.812}}\right) \\
&= \Pr(-0.04 \leq Z \leq 1.49)
\end{aligned}.$$

- Now probability involves the standard Normal distribution

- Normal statistical tables say $\Pr(-0.04 \leq Z \leq 1.49) = 0.45$.

- Details: break into two parts as

$$\begin{aligned}
& \Pr(-0.04 \leq Z \leq 1.49) \\
= \; & \Pr(-0.04 \leq Z \leq 0) + \Pr(0 \leq Z \leq 1.49)
\end{aligned}$$

- From table $\Pr(0 \leq Z \leq 1.49) = 0.4319$.

- But since the Normal is symmetric $\Pr(-0.04 \leq Z \leq 0) = \Pr(0 \leq Z \leq 0.04)$=0.0160.

- Adding these two probabilities together gives 0.4479

1.2 The Classical Assumptions for the Regression Model

Now let us return to the regression model.

We need to make some assumptions to do any statistical derivations and start with the *classical assumptions*

1. $E(Y_i) = \beta X_i$.

2. $var(Y_i) = \sigma^2$.

3. $cov(Y_i, Y_j) = 0$ for $i \neq j$.

4. Y_i is Normally distributed

5. X_i is fixed. It is not a random variable.

Compact notation: Y_i are independent $N\left(\beta X_i, \sigma^2\right)$.

An equivalent way of writing the classical assumptions is:

1. $E(\varepsilon_i) = 0$ – mean zero errors.

2. $var(\varepsilon_i) = E\left(\varepsilon_i^2\right) = \sigma^2$ – constant variance errors (homoskedasticity).

3. $cov\left(\varepsilon_i \varepsilon_j\right) = 0$ for $i \neq j$.

4. ε_i is Normally distributed

5. X_i is fixed. It is not a random variable.

1.3 Motivation for Classical Assumptions

- Regression model fits a straight-line through an XY-plot.

- $E(Y_i) = \beta X_i$ is the linearity assumption.

- Second assumption: all observations have the same variance (homoskedasticity).

- Ex. where this might not be a good assumption. House price data. Small houses all the same. Big houses more diverse. If so, house prices might be more diverse for big houses (heteroskedasticity).

- Third assumption: observations uncorrelated with one another.

- This assumption is usually reasonable with cross-sectional data (e.g. in a survey, response of person1 and person 2 are unrelated).

- For time series data not a good assumption (e.g. interest rate now and last month are correlated with one another)

- Fourth assumption (Y is Normal), harder to motivate.

- In many empirical applications, Normality is reasonable.

- Asymptotic theory can be used to relax this assumption. We will not cover this in this course (but see Appendix C and Appendices at end of several chapters)

- Fifth assumption (explanatory variable not a random variable) is good in experimental sciences, but maybe not in social sciences.

- We will talk about relaxing these assumptions in later chapters.

1.4 The Ordinary Least Squares (OLS) Estimator of β

$$y_i = \beta X_i + \varepsilon_i$$

OLS estimator is chosen to minimize:

$$\sum_{i=1}^{N} \varepsilon_i^2$$

This can be done using calculus.

$$\widehat{\beta} = \frac{\sum\limits_{i=1}^{N} X_i y_i}{\sum\limits_{i=1}^{N} X_i^2}$$

1.4.1 Properties of OLS Estimator

$$\widehat{\beta} = \frac{\sum X_i y_i}{\sum X_i^2} = \frac{\sum X_i \left(X_i \beta + \varepsilon_i\right)}{\sum X_i^2} = \beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2} \quad (*)$$

Property 1: OLS is unbiased under the classical assumptions

$$E\left(\widehat{\beta}\right) = \beta$$

Proof:

$$
\begin{aligned}
E\left(\widehat{\beta}\right) &= E\left(\beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= \beta + E\left(\frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= \beta + \frac{1}{\sum X_i^2} E\left(\sum X_i \varepsilon_i\right) \\
&= \beta + \frac{1}{\sum X_i^2} \sum X_i E\left(\varepsilon_i\right) \\
&= \beta
\end{aligned}
$$

Use equation (*) and properties of expected value operator. Remember X_i is not random (hence can be treated as a constant).

Property 2: Variance of OLS estimator under the classical assumptions

$$var\left(\widehat{\beta}\right) = \frac{\sigma^2}{\sum X_i^2}$$

Proof:

$$
\begin{aligned}
var\left(\widehat{\beta}\right) &= var\left(\beta + \frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= var\left(\frac{\sum X_i \varepsilon_i}{\sum X_i^2}\right) \\
&= \left(\frac{1}{\sum X_i^2}\right)^2 var\left(\sum X_i \varepsilon_i\right) \\
&= \left(\frac{1}{\sum X_i^2}\right)^2 \sum X_i^2 var\left(\varepsilon_i\right) \\
&= \left(\frac{1}{\sum X_i^2}\right)^2 \sigma^2 \sum X_i^2 \\
&= \frac{\sigma^2}{\sum X_i^2}
\end{aligned}
$$

Use equation (*) and properties of variance operator. Remember X_i is not random (hence can be treated as a constant).

Property 3: Distribution of OLS estimator under classical assumptions

$$\widehat{\beta} \text{ is } N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right)$$

Proof: Properties 1 and 2 plus "linear combinations of Normals are Normal" theorem.

Property 3 is important since it can be used to derive confidence intervals and hypothesis tests.

The OLS estimator is a random variable and has a p.d.f.
Ex. Figure 3.3 if $\widehat{\beta}$ is $N(2, 1)$

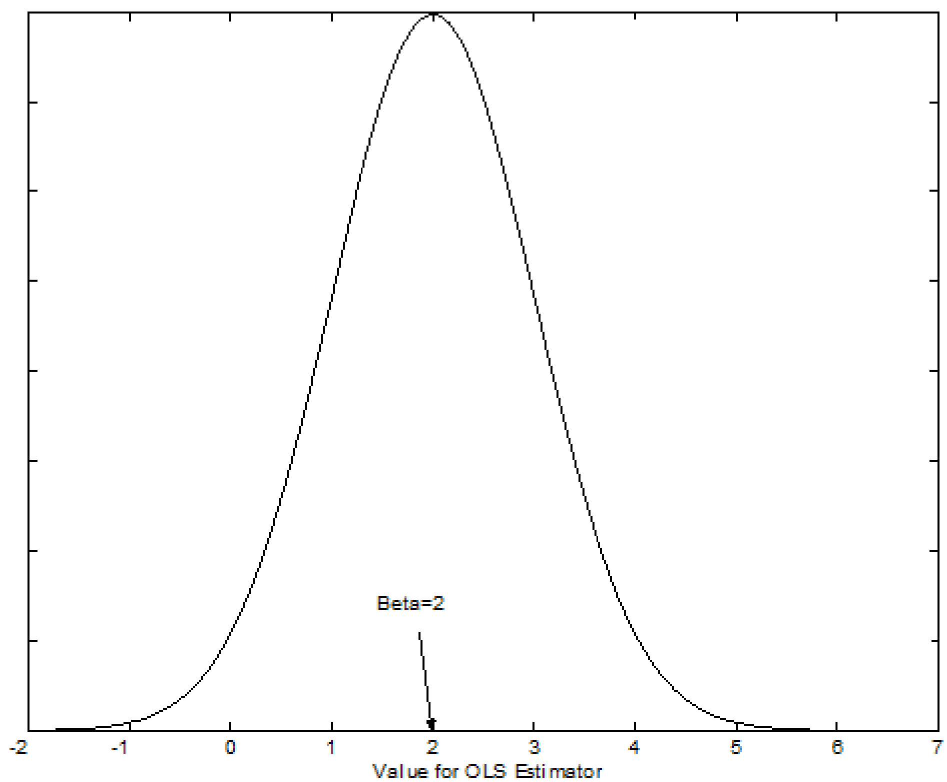
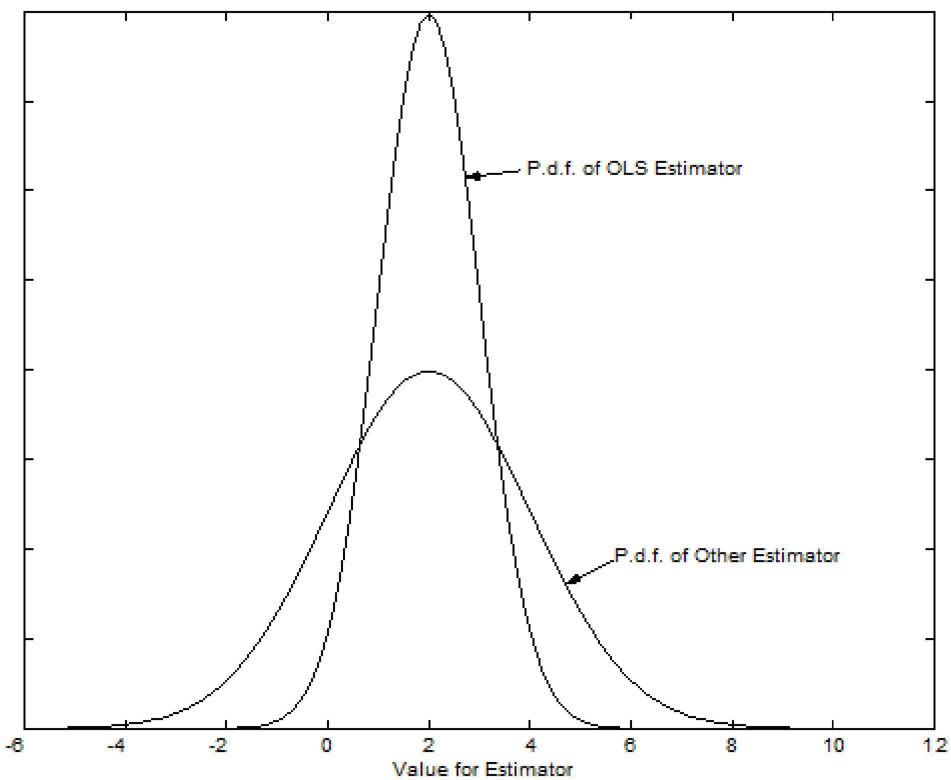


Figure 3.3: The p.d.f. of the OLS Estimator

Want your estimator to be unbiased and want it to have as small a variance as possible.

An unbiased estimator is said to be *efficient* relative to another if it has a smaller variance.

Figure 3.4: The p.d.f.s of the OLS Estimator and A Less Efficient Estimator



Property 4: The Gauss-Markov Theorem

If the classical assumptions hold, then OLS is the best, linear unbiased estimator,

where best = minimum variance

linear = linear in y.

Short form: "OLS is BLUE"

Note: the assumption of Normal errors is NOT required to prove Gauss-Markov theorem. Hence, OLS is BLUE even if errors are not Normal.

Property 5: Under the Classical Assumptions, OLS is the maximum likelihood estimator

Maximum likelihood is another statistical principal for choosing estimators.

Textbook has a discussion of this topic, but I do not have time to cover in lectures.

1.4.2 Deriving a Confidence Interval for β

Assume σ^2 known (discuss relaxing this assumption later).

Use Property 3 to obtain:

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}} \text{ is } N(0,1)$$

Then can use statistical tables for the Normal distribution to make probability statements. For instance,

$$\Pr\left[-1.96 \leq Z \leq 1.96\right] = 0.95$$

To get 95% confidence interval, rearrange the inequalities to put β in the middle:

$$\Pr\left[-1.96 \leq \frac{\widehat{\beta}-\beta}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}} \leq 1.96\right] = 0.95$$

rearranging:

$$\Pr\left[\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\sum X_i^2}} \leq \beta \leq \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\sum X_i^2}}\right] = 0.95$$

Note: $\widehat{\beta}$ is a random variable, β is not a random variable. Hence, we do not say "probability interval" but rather "confidence interval".

95% confidence interval is

$$\left[\widehat{\beta} - 1.96\sqrt{\frac{\sigma^2}{\sum X_i^2}} \leq \beta \leq \widehat{\beta} + 1.96\sqrt{\frac{\sigma^2}{\sum X_i^2}}\right]$$

commonly written as:

$$\widehat{\beta} \pm 1.96\sqrt{\frac{\sigma^2}{\sum X_i^2}}$$

Other confidence levels can be handled by getting different number from Normal tables. For instance, 90% confidence interval would replace "1.96)" by "1.64" in previous equations.

1.4.3 Hypothesis tests about β

Assume σ^2 known (discuss relaxing this assumption later).

Basic idea in testing any hypothesis, H_0:

"The econometrician accepts H_0 if the calculated value of the test statistic is consistent with what could plausibly happen if H_0 is true."

List the general steps in hypothesis testing along with the specific steps for this case.

Step1: Specify a hypothesis, H_0.

H_0: $\beta = \beta_0$ (where β_0 is known, usually $\beta_0 = 0$)

Step 2: Specify a test statistic

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}}$$

Step 3: Figure out distribution of test statistic assuming H_0 is true.

$$Z = \frac{\widehat{\beta} - \beta_0}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}} \text{ is } N(0, 1).$$

Step 4: Choose a level of significance (usually 5%).

0.05

Step 5: Use Steps 3 and 4 to get a critical value.

Critical value = 1.96 (from Normal statistical tables)

Step 6: Calculate your test statistic and compare to critical value. Reject H_0 if absolute value of test statistic is greater than critical value (else accept H_0).

Reject if $|Z| > 1.96$.

1.5 Modifications when σ^2 is unknown

σ^2 appears in previous formula for confidence interval and test statistic. What to do when it is unknown?

1.5.1 Estimation of σ^2

Residuals:

$$\widehat{\varepsilon}_i = y_i - \widehat{\beta} X_i$$

Unbiased estimator of σ^2 is

$$s^2 = \frac{\sum \widehat{\varepsilon}_i^2}{N-1}$$

Property (not proved in this course):

$$E\left(s^2\right) = \sigma^2.$$

Note: The $N-1$ in denominator becomes $N-k-1$ in multiple regression where k is number of explanatory variables.

1.5.2 Confidence interval for β when σ^2 unknown

Replace σ^2 by s^2 in equations from earlier section "Deriving a Confidence Interval for β".

Nothing changes, except:

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\dfrac{\sigma^2}{\sum X_i^2}}} \text{ is } N\left(0, 1\right)$$

is replaced by:

$$Z = \frac{\widehat{\beta} - \beta}{\sqrt{\dfrac{s^2}{\sum X_i^2}}} \text{ is } t\left(N - 1\right),$$

where $t\,(N-1)$ is the Student-t distribution with N-1 degrees of freedom. And must use Student-t statistical tables instead of Normal.

See Appendix B for instruction in using Student-t tables.

Example:

Suppose we have $N = 21$.

Before (with σ^2 known) we derived

$$\widehat{\beta} \pm 1.96\sqrt{\frac{\sigma^2}{\sum X_i^2}}.$$

Now we have to look in $t(20)$ row of Student-t statistical tables and obtain:

$$\widehat{\beta} \pm 2.08\sqrt{\frac{s^2}{\sum X_i^2}}.$$

1.5.3 Hypothesis testing about β when σ^2 unknown

Replace σ^2 by s^2 in equations from earlier section "Hypothesis tests about β".

Nothing changes, except test statistic:

$$Z = \frac{\widehat{\beta} - \beta_0}{\sqrt{\frac{\sigma^2}{\sum X_i^2}}} \text{ is } N(0, 1)$$

is replaced by:

$$Z = \frac{\widehat{\beta} - \beta_0}{\sqrt{\frac{s^2}{\sum X_i^2}}} \text{ is } t(N - 1),$$

where $t(N-1)$ is the Student-t distribution with N-1 degrees of freedom.

Must use Student-t statistical tables instead of Normal to get critical value.

1.5.4 Note on P-values

All relevant computer packages now present P-values for hypothesis tests. This means you do not need to look up critical values in statistical tables (so no emphasis on tables in this course).

Useful (but not quite correct) intuition: "P-value is the probability that H_0 is true"

A correct interpretation: "P-value equals the smallest level of significance at which you can reject H_0"

Example: If P-value is .04 you can reject H_0 at 5% level of significance or 10% or 20% (or any number above 4%). You cannot reject H_0 at 1% level of significance.

Common rule of thumb:

Reject H_0 if P-value less than .05.

1.6 Chapter Summary

The major points and derivations covered in this chapter include:

1. The manner in which the Normal distribution (which is characterized by a mean and variance) is used in the context of the simple regression model.

2. The introduction of the classical assumptions, from which all else in this chapter is derived.

3. The properties of the OLS estimator, including a proof that it is unbiased and a derivation of its distribution (i.e. $\widehat{\beta}$ is $N\left(\beta, \frac{\sigma^2}{\sum X_i^2}\right)$).

4. The Gauss-Markov theorem which says OLS is BLUE under the classical assumptions.

5. The derivation of a confidence interval for β (assuming σ^2 is known).

6. The derivation of a test of the hypothesis that $\beta = 0$ (assuming σ^2 is known).

7. The OLS estimator of σ^2.

8. How the confidence interval and hypothesis test are modified when σ^2 is unknown.