Lecture 5. Statistical Inference and the Classical Linear Regression Model

In previous lectures we used an example to review mathematical statistics.

Mathematical statistics is concerned with using data generated by a random experiment to learn about features of that experiment.

Example: Outcome of 100 coin tosses is used to learn about the probability of Head.

In econometrics we want to measure an inexact linear relation between a dependent variable $Y$ and (an) independent variable(s) $X$

$$Y = \alpha + \beta X + u$$

What is the random experiment?

Consider example with

$$Y = \text{House price}$$

$$X = \text{Living area}$$

If we look at a particular city, e.g. Los Angeles, then all houses have a price $Y$ and living area $X$.

Consider the following thought experiment:
For a particular house you are given the living area $X$. You also know the values of $\alpha, \beta$. Using this can you predict the house price $Y$? What is your prediction? Will it be correct? Why (not)?

The determination of $Y$ given the value of $X$ is a random experiment, because you do not know the outcome in advance, because you do not know in advance what value $u$ takes. Hence $u$ is a random variable.

Remember that $u$ captures all omitted variables that determine $Y$ beside $X$. If these variables are $Z_1, \ldots, Z_K$, then

$$u = \gamma_1 Z_1 + \cdots + \gamma_K Z_K$$

Because $u$ is a random variable, it has a probability distribution.
First, note that $u$ can take any value between $-\infty$ and $\infty$. Also if $E(u) \neq 0$ we can always redefine $U$ as $u - E(u)$ and this has mean 0. Note

$$Y = \alpha + E(u) + \beta X + u - E(u)$$

Hence the intercept can always be chosen such that $E(u) = 0$. This is how we choose the intercept and because of this the intercept has no direct interpretation.

Hence we have:

Assumption 1: $u$ is a random variable with $E(u) = 0$

Because of this assumption we know, even before we collect the data $Y_i, X_i, i = 1, \ldots, n$, that the linear relation between $Y$ and $X$ is not exact, because it will be ‘disturbed’ by the random $u$’s

The observed residuals in a scatterdiagram are realizations of the random $u$

The random $u$ is called the error term or random disturbance of the linear relation.

The relation

$$Y = \alpha + \beta X + u$$

in which $u$ is a random error term/disturbance is called the simple linear regression model.

Simple because there is only one explanatory variable/regressor.

The dependent variable is also called regressand.

The independent variables are called explanatory variables, regressors, covariates.

The coefficients $\alpha, \beta$ are called the regression coefficients

Next we discuss further assumptions on the probability distribution of $u$.

The most important assumption concerns the relation between $X$, the variable included in the relation and $u$. 
Remember we can think of $u$ as

$$u = \gamma_1 Z_1 + \cdots + \gamma_K Z_K$$

Focus on the first variable that we call $Z$ and summarize the rest by $v$

$$u = \gamma Z + v$$

The key question is whether $X$ and $Z$ are related or not. If they are we can write

$$Z = \kappa + \lambda X + w$$

This assumes a linear relation (not essential) that is itself not exact.

Substitution gives

$$Y = \alpha + \beta X + \gamma(\kappa + \lambda X + w) + v =$$

$$= \alpha + \gamma \kappa + (\beta + \gamma \lambda) X + \gamma w + v$$

This is again an inexact linear relation. Note that the intercept and residual have changed, but intercept has no interpretation and if both $w$ and $v$ satisfy assumption 1 then so does the combined error term $\gamma w + v$.

More important: The coefficient on $X$ has changed.

Consider this coefficient

$$\beta + \gamma \lambda$$

It is

Direct effect $X$ on $Y$ $(\beta)$ + Indirect effect $X$ on $Y$ $(\lambda \gamma)$

Moreover

Indirect effect $X$ on $Y$ $(\lambda \gamma) = \text{Effect } X \text{ on } Z \times (\lambda) \times \text{Effect } Z \text{ on } Y(\gamma)$

i.e. this is indirect effect through the omitted variable $Z$.

Conclusion: If $u$ contains a (or more) variable(s) that is related to $X$, then the effect that we measure in the linear regression model is NOT the change in $Y$ that results from a change in $X$!
Is this important? Depends on the goal of the analysis.

Consider two questions:

1. I have a house with living area $X$. What is the price that I can expect to get if I sell it?
2. I think of expanding the living area. If I add 1 square foot, how much does that change the price of the house?

In 1, you compare houses with different living areas. The other amenities are different as well, and living area proxies for these other characteristics.

In 2, you look at single house and only the living area changes. In economics this is called a ceteris paribus (all other things held constant) change. It involves an experiment that is not in the data! Such an experiment is called a counterfactual.

For question 1 the interpretation of the coefficient on $X$ is not very important. A regression model in which that is the case is called a reduced form model.

For question 2 the interpretation of this coefficient is critically important: It is the change in $Y$ that results from a ceteris paribus change in $X$. A regression model in which that is the case is called a structural model.

Structural models are needed to evaluate the effect of interventions or policies. For many predictions that do not involve policy changes/interventions the reduced form model suffices.

Because economic theory is about ceteris paribus changes, the measurement of quantities like price elasticities also requires structural models.

How do you know whether you can think of a regression model as a structural model?

Note that this requires that

$$\gamma \lambda = 0$$

Because in general $\gamma \neq 0$ we need that

$$\lambda = 0$$

i.e. no relation between the included and omitted variables.
How can we ensure that?

Two strategies

1. Expand the model, i.e. include $Z$ and all other variables that are potentially related to $X$.
2. Enforce the independence of $X$ and $u$.

For strategy 2 we can perform randomized experiments.

Assume that you want to study the effect of attending USC on income in the 10 years after graduation on all eligible potential students.

Let $X$ be the indicator of this intervention/policy with

$$X = 1 \quad \text{if an individual attends USC}$$
$$X = 0 \quad \text{if not}$$

Assume that you can decide who gets into USC or not. How would you do it if your goal is to measure the effect of attending USC?

To ensure that there is no relation between $X$ and $u$ we select the students that are to attend USC at random.

In that case $\beta$ is the pure effect of attending USC.

Compare this with the effect that you obtain by not selecting the entrants at random, but using the current admissions procedure.

What do you expect? Is the effect smaller, bigger? What does that tell us about the admissions process?

Randomized assignment is common practice in medical research.
We now formalize the assumption

Assumption 2: The distribution of \( u \) is such that there is nor relation between \( u \) and \( X \)

In statistics there are several ways to express lack of relation.
The most important are stochastic independence and 0 covariance/correlation.

Hence assumption 2 in different forms

Assumption 2': \( X \) is not stochastic.

Then by assumption 1

\[
E(Xu) = X E(u) = 0
\]

This is the assumption made in Gujarati. It applies if \( X \) is e.g. time (the regression model is than always a reduced for model) or if \( X \) can be set by you (and you randomize).

Assumption 2'': \( X \) and \( u \) have 0 covariance/correlation, i.e. \( E(Xu) = 0 \)

This applies if both \( X \) and \( u \) are random variables, e.g. if \( X \) is not under our control.

Finally

Assumption 2''': \( X \) and \( u \) are stochastically independent.

This implies Assumption 2''. Assumption 2'' is enough if the relation between \( Y \) and \( X \) is linear. Otherwise we need Assumption 2'''.

It turns out that if we have Assumption 2'' or 2''' then we can treat the \( X \) as set by us as in Assumption 2'.

Lecture 6. The Classical Linear Regression Model

Simple linear regression model

\[ Y = \alpha + \beta X \]

Assumptions in previous lecture

Assumption 1: \( u \) is a random variable with \( E(u) = 0 \)

Assumption 2: The probability distribution of the random error \( u \) is independent of \( X \)

This is implied by

Assumption 2': \( X \) is a deterministic, i.e. non-stochastic, variable

In other words: the observed values \( X_i, i = 1, \ldots, n \) can be treated as \( n \) constants.

In practice assumption 2’ only holds in special cases, e.g. if \( X \) is (calendar) time.

In most cases \( X \) is also the outcome of some random experiment, e.g. in macroeconomics an important relation is that between aggregate consumption \( Y \) and aggregate income \( X \). Both variables are random variables associated with the random experiment that determines the state of the economy in a year.

The random experiment associated with the linear regression model is the determination of \( Y \) given the value of \( X \). It is not important how \( X \) is determined, as long as the assumptions of the regression model (until now assumptions 1 and 2) are satisfied.

Hence we have two random experiments: one that determines \( X \) and one that determines \( Y \) given the outcome of the first experiment.

We are only interested in the second random experiment.

If we have data \( Y_i, X_i, i = 1, \ldots, n \), i.e. \( n \) observations on a dependent variable \( Y \) and an independent variable \( X \), we consider these as outcomes of the \( n \) random experiments

\[ (1) \quad Y_i = \alpha + \beta X_i + u_i \]

with \( i = 1, \ldots, n \).
These random experiments are as follows:

1. Treat \( X_i, i = 1, \ldots, n \) as \( n \) constants (it does not matter how they are determined).
2. For each \( i \), \( u_i \) is a draw from a probability distribution that does not depend on \( X_i \) (assumption 2) and has mean 0 (assumption 1).
3. \( Y_i \) is determined as in (1).

If we choose a particular distribution for \( u_i \), e.g. the normal (mean 0) distribution, then we can write a computer program that generates datasets of size \( n \). This will give repeated samples of size \( n \) that are outcomes of random experiment (1).

The Ordinary Least Squares (OLS) solutions to fitting a straight line in a scatter diagram are

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

\[
\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}
\]

In the Simple Linear Regression random experiment these solutions are estimators of the parameters, here regression coefficients, \( \alpha, \beta \).

These estimators have a sampling distribution that as usual can be used to

- Evaluate the quality of the estimators
- Find confidence intervals
- Perform hypothesis tests

Sampling distribution: Distribution of estimators in repeated samples of size \( n \).
We can use the computer to find the sampling distribution (see above for description of
the random experiment)

Let

\[ \alpha = 1 \]

\[ \beta = 1 \]

\[ X_i, i = 1, \ldots, n \text{ are uniform } [0,1] \text{ random numbers} \]

\[ u_i \text{ has a standard normal distribution, and all } u_i \text{'s are independent and have the same distribution} \]

This completely specifies the Linear Regression random experiment.

The graphs give the sampling distribution of \( \hat{\beta} \) for \( n = 10 \) and 100.

A second set of graphs gives the sampling distribution if we assume that the \( u_i \) 's are independent and have the same distribution, but that distribution is the uniform \( [0, \sqrt{12}] \) distribution.

Which is more relevant: the normal or uniform distribution?
Sampling distr. of $\hat{\beta}$, $n = 100$, $\nu$ uniform
Sampling distr. of $\hat{\beta}$, $n = 10$, $\epsilon$ standard normal
Sampling distr. of $\hat{\beta}$, $n = 100$, $u$ uniform
Sampling distr. of $\hat{\beta}$, $n = 10$, $u$ uniform
If $n=10$, the sampling distribution is close to normal even if the $u_i$'s have a uniform distribution. Variance is larger if $n=10$.

As in the coin tossing experiment, we can derive the sampling distribution of $\hat{\alpha}, \hat{\beta}$ by using the assumptions instead of using the computer to generate samples.

Consider the linear regression model with $\alpha = 0$.

Then

$$\hat{\beta} = \frac{\sum_{i=1}^{n} X_i Y_i}{\sum_{i=1}^{n} X_i^2} = \frac{\sum W_i Y_i}{\sum W_i X_i^2}$$

with

$$W_i = \frac{X_i}{\sum_{j=1}^{n} X_j^2}$$

From the Linear Regression model

$$Y_i = \beta X_i + u_i$$

Substitution gives, using $\sum_{i=1}^{n} W_i X_i = 1$,

$$\hat{\beta} = \sum_{i=1}^{n} W_i Y_i = \sum_{i=1}^{n} W_i (\beta X_i + u_i) = \beta + \sum_{i=1}^{n} W_i u_i$$

Because the $W_i$ can be treated as constants (assumption 2'), we have by assumption 1

$$E(\hat{\beta}) = \beta + E\left(\sum_{i=1}^{n} W_i u_i\right) = \beta + \sum_{i=1}^{n} W_i E(u_i) = \beta$$

Conclusion: The OLS estimator of $\beta$ is unbiased.
The same conclusion holds if the Linear Regression model has an intercept.

Hence, the OLS estimators $\hat{\alpha}, \hat{\beta}$ of $\alpha, \beta$ are unbiased estimators.

Compare with the sampling distribution in the computer experiment.

Next step is to derive the sampling variance of the OLS estimators.

That derivation is simpler if we make two additional assumptions

Assumption 3 (Homoskedasticity)
All $u_i$'s have the same variance

$$Var(u_i) = E(u_i^2) = \sigma^2$$

Assumption 4 (No serial correlation)
The random errors $u_i$ and $u_j$ are not correlated for all $i \neq j$

$$Cov(u_i, u_j) = E(u_iu_j) = 0$$

Discussion assumptions

Heteroskedasticity (Greek for equal dispersion) affects shape of scatterplot

Example: Education and late career income.

No serial correlation is important in time-series data. Most data are

- Cross-section data: variables are for $n$ individuals, households, firms, countries etc. in a particular time period
- Time-series data: variables are for one individual, firm, country etc. in $n$ subsequent time periods (weeks, months, quarters, years)
- Panel data: combination of these two

Remember error term $u$ captures omitted variables. Most time-series are such that the observations in subsequent time periods are correlated. Same is true for omitted variables. Hence in time-series data assumption 4 need not hold.

Serial correlation also affects shape of scatterplot.
Now we use assumptions 3 and 4 to derive the sampling variance of $\hat{\beta}$ (if we assume that $\alpha = 0$)

Remember

$$\hat{\beta} = \beta + \sum_{i=1}^{n} W_i u_i$$

Hence,

$$Var(\hat{\beta}) = E((\hat{\beta} - \beta)^2) = E\left[\left(\sum_{i=1}^{n} W_i u_i\right)^2\right] =$$

$$= E\left[\sum_{i=1}^{n} W_i^2 u_i^2 + 2\sum_{i<j}^{n} W_i W_j u_i u_j\right] =$$

$$= \sum_{i=1}^{n} W_i^2 E(u_i^2) + 2\sum_{i<j}^{n} W_i W_j E(u_i u_j)$$

The second term is 0 by assumption 4. By assumption 3 $E(u_i^2) = \sigma^2$. Also by the definition of $W_i$, we have $\sum_{i=1}^{n} W_i^2 = \frac{1}{\sum_{i=1}^{n} X_i^2}$

Combining this we find

$$Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} X_i^2}$$

Note

- Variance smaller if $\sigma^2$ smaller
- Variance smaller if $\sum_{i=1}^{n} X_i^2$ larger

How does this translate in scatterplot?

What happens to the variance of $n$ is large?
Implication for sampling distribution?
Definition: An estimator with a sampling distribution that becomes a degenerate distribution in the population value of the parameter is called consistent.

The OLS estimator $\hat{\beta}$ is consistent.

Using a similar argument as above we can derive

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

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

$$Cov(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{X}}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sigma^2$$

Moreover, the OLS estimators are consistent.
Lecture 7. Confidence Intervals and Hypothesis Tests in the Simple CLR Model

In lecture 6 we introduced the Classical Linear Regression (CLR) model that is the random experiment of which the data \( Y_i, X_i, i = 1, \ldots, n \) are the outcomes.

The CLR model specifies that the relation between the dependent variable \( Y \) and the independent variable \( X \) is an inexact linear relation, i.e.

\[
Y_i = \alpha + \beta X_i + u_i, \quad i = 1, \ldots, n
\]

and that \( X_i, u_i \) satisfy the assumptions

Assumption 1: \( u_i, i = 1, \ldots, n \) are random variables with \( E(u_i) = 0 \)

Assumption 2: \( X_i, i = 1, \ldots, n \) are deterministic, i.e. non-random, constants.

Assumption 3 (Homoskedasticity)
All \( u_i \)'s have the same variance, i.e. for \( i = 1, \ldots, n \)

\[
Var(u_i) = E(u_i^2) = \sigma^2
\]

Assumption 4 (No serial correlation)
The random errors \( u_i \) and \( u_j \) are not correlated for all \( i \neq j = 1, \ldots, n \)

\[
Cov(u_i, u_j) = E(u_i u_j) = 0
\]

If these assumptions hold, the OLS estimators

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

\[
\hat{\alpha} = \bar{Y} - \hat{\beta} \bar{X}
\]

of the regression coefficients \( \alpha, \beta \) have the following properties
1. The OLS estimators $\hat{\alpha}, \hat{\beta}$ are unbiased, i.e. for the mean of the sampling distribution it holds that

$$E(\hat{\alpha}) = \alpha, \quad E(\hat{\beta}) = \beta$$

For this only assumptions 1 and 2 are needed.

2. The variance of the sampling distribution is

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

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

$$Cov(\hat{\alpha}, \hat{\beta}) = -\frac{\bar{X}}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \sigma^2$$

3. The OLS estimators $\hat{\alpha}, \hat{\beta}$ are consistent, i.e. in large samples the sampling distribution is concentrated in $\alpha, \beta$.

The sampling variance informs us how precise the estimates are. It can be shown that under assumptions 1-4, the OLS estimators are the best estimators, i.e. have the smallest sampling variance, among the unbiased estimators that can be expressed as a linear expression in, i.e. as a weighted average of the $Y_i$.

In jargon: The OLS estimators are Best Linear Unbiased Estimators (BLUE)
For confidence intervals and hypothesis tests, the mean and variance of the sampling distribution are not enough: we need that the sampling distribution belongs to a class of distributions that we can work with, e.g., the normal distribution. This will be the case if the following assumption holds

Assumption 5. The random error terms $u_i, i = 1, \ldots, n$ are random variables with a normal distribution.

Note that

- This normal distribution has mean 0 (assumption 1) and variance $\sigma^2$ (assumption 3) for all $i = 1, \ldots, n$. Hence the $u_i$'s are random variables with identical (normal) distributions.
- The error terms $u_i$ and $u_j$ are uncorrelated (assumption 4) and because they are normal, also stochastically independent.

Why is the normal distribution an obvious choice as a distribution for the error term $u$?

What is the sampling distribution of the OLS estimators if assumption 5 holds?

Consider a CLR model with $\alpha = 0$. We have seen that the OLS estimator $\hat{\beta}$ can be expressed as

$$\hat{\beta} = \beta + \sum_{i=1}^{n} W_i u_i$$

The right-hand side is the sum of a constant ($\beta$) and a linear combination of normal random variables

$$\sum_{i=1}^{n} W_i u_i$$

It can be shown that such a linear combination is also a random variable with a normal distribution.

Conclusion: If assumption 5 holds, the sampling distribution of $\hat{\beta}$ is normal with mean $\beta$ and variance

$$\sigma^2 / \sum_{i=1}^{n} X_i^2.$$
If the CLR model has an intercept we have the same result: If assumption 5 holds, the sampling distribution of the OLS estimators $\hat{\alpha}, \hat{\beta}$ is also normal with mean $\alpha, \beta$ and variance

\[
Var(\hat{\alpha}) = \frac{\sum_{i=1}^{n} X_i^2}{n \sum_{i=1}^{n} (X_i - \bar{X})^2} \sigma^2
\]

\[
Var(\hat{\beta}) = \frac{\sigma^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2}
\]

respectively.

Now we can find a confidence interval for $\beta$. As in the coin tossing example we ‘standardize’ the estimator

\[
Z = \frac{\hat{\beta} - \beta}{\sigma^2} \left( \sum_{i=1}^{n} (X_i - \bar{X}) \right)
\]

Now $Z$ has a normal distribution with mean 0 and variance 1.
We find a 95% confidence interval for $\beta$ by considering the probability

\[ .95 = \Pr(-1.96 < Z < 1.96) = \]

\[
\begin{align*}
&= \Pr \left\{ -1.96 < \frac{\hat{\beta} - \beta}{\frac{\sigma^2}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2}}} < 1.96 \right\} \\
\end{align*}
\]

Hence with probability .95 we have

\[
-1.96 < \frac{\hat{\beta} - \beta}{\frac{\sigma^2}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2}}} < 1.96
\]

or equivalently written as an interval for $\beta$

\[
\hat{\beta} - 1.96 \frac{\sigma^2}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2}} < \beta < \hat{\beta} + 1.96 \frac{\sigma^2}{\sqrt{\sum_{i=1}^{n} (X_i - \overline{X})^2}}
\]

With probability .95, the unknown $\beta$ is in this interval. This refers to repeated samples: in 95% of these samples $\beta$ is in this interval.

The interval is a 95% confidence interval for $\beta$.

What changes if we want a 90% confidence interval? Use Table 1!

What problem do we have with this interval? Can it be computed from the data?
Remember that $\sigma^2$ is the variance of the random error term $u$. We do not have the $u_i$'s, because

$$u_i = Y_i - \alpha - \beta X_i$$

and we do not know $\alpha, \beta$. After we estimate $\alpha, \beta$, we can compute the OLS residuals

$$e_i = Y_i - \hat{\alpha} - \hat{\beta} X_i$$

These are not the same because we use estimators for $\alpha, \beta$.

Remember that the sample mean of the $e_i, i = 1, \ldots, n$ is 0. Hence, the sample variance is

$$\hat{\sigma}^2 = \frac{1}{n} \sum_{i=1}^{n} e_i^2$$

This is an estimator of $\sigma^2$.

The following estimator is preferred

$$s^2 = \frac{1}{n-2} \sum_{i=1}^{n} e_i^2$$

Note that we have estimated 2 parameters $\alpha, \beta$ to compute the $e_i$'s, and this is the reason we use $n-2$. It can be shown that this is an unbiased and consistent estimator of $\sigma^2$.

Now we can derive a computable confidence interval. Start from

$$T = \frac{\hat{\beta} - \beta}{s^2 \sqrt{\sum_{i=1}^{n} (X_i - \bar{X})^2}}$$

i.e. we substitute an (unbiased, consistent) estimator for $\sigma^2$. 
Because the denominator is estimated (and hence varies in samples) $T$ does not have a standard normal distribution, but a Student t distribution. This distribution has a larger variance. See Table 2 for a comparison with the standard normal distribution. As the standard normal, the t distribution is symmetric around 0 and has mean 0.

The t distribution depends on the number of observations. If $n$ is large it is close to the standard normal, but if $n$ is small it much more dispersed. If we have $n$ observations, then we have a t distribution with parameter $n - 2$. In jargon this is number of degrees of freedom of the t distribution.

Note that if $n > 60$ the error that one makes in using the standard normal instead of the t distribution is small.

The starting point for finding a 95% confidence interval is now

$$.95 = \Pr(-c < T < c) =$$

$$= \Pr\left(-c < \frac{\hat{\beta} - \beta}{\frac{s^2}{\sqrt{\sum_{i=1}^{n}(X_i - \bar{X})^2}}} < c\right)$$

If we have 30 observations, then $c = 2.048$.

What is $c$ if we have 62 observations or if we want a 90% confidence interval?

For 30 observations the 95% confidence interval for $\beta$ is

$$\hat{\beta} - 2.048 \sqrt{\frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}} < \beta < \hat{\beta} + 2.048 \sqrt{\frac{s^2}{\sum_{i=1}^{n}(X_i - \bar{X})^2}}$$

This can be computed from the data.
The estimated standard deviation of the OLS estimator $\hat{\beta}$ is

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

This is called the standard error of $\hat{\beta}$ (of course we can also compute the standard error of $\hat{\alpha}$).

It is customary to report both the OLS estimate and its standard error. Reason is clear from formula for confidence interval: boundaries of confidence interval are OLS estimate plus/minus a multiple of the standard error. Multiple close to 2 for 95% (exactly 2 for $n = 62$).

In graphs the sampling distribution of $T$ is given for the linear regression model of lecture 6 with

$$\alpha = 1$$

$$\beta = 3$$

$X_i, i = 1, \ldots, n$ are uniform [0,1] random numbers

$u_i$ has a standard normal distribution or a uniform distribution

The graphs are base on 10000 samples.

- If $n$ is small, the distribution is more dispersed than standard normal
- Even if the error has a uniform distribution we can use the $t$ and standard normal distribution ($n$ large) as an approximation to the sampling distribution of $T$
Sampling distribution t statistic \((n = 30)\), standard normal and t pdf
Sampling distribution t statistic \((n = 30)\), standard normal and t pdf; errors uniformly distributed
Explains why we do not worry much about the validity of assumption 5.

Next we consider testing hypotheses on regression parameters. We use the same setup as in the coin tossing example:

- Null and alternative hypothesis
- Decision rule that chooses between these hypotheses

Null hypothesis

\[ H_0 : \beta = \beta_0 \]

Alternative hypothesis

\[ H_1 : \beta \neq \beta_0 \] (two-sided alternative)

or

\[ H_1 : \beta > \beta_0 \] (one-sided alternative)

or

\[ H_1 : \beta < \beta_0 \] (one-sided alternative)

For instance if \( \beta \) is a price elasticity we may test \( H_0 : \beta = 0 \) against \( H_1 : \beta < 0 \)

\( H_0 : \beta_0 = 0 \) is the null hypothesis that \( X \) has no effect on \( Y \).

Decision rule

Consider

\[
T = \frac{\hat{\beta} - \beta_0}{s^2 \sqrt{\frac{1}{n} \sum_{i=1}^{n} (X_i - \bar{X})^2}}
\]

If \( H_0 \) is true, then \( T \) has a t distribution with \( n-2 \) degrees of freedom. This distribution is symmetric around 0 and most of the time (in repeated samples) we get a value not too far from 0.

If \( H_0 \) is not true, then the distribution of \( T \) will shift to the right (if \( \beta > \beta_0 \)) or to the left (if \( \beta < \beta_0 \)).

The sampling distribution of \( T \) for \( \beta_0 = 3 \) (correct value) and for \( \beta_0 = 0 \) (incorrect value) is plotted (\( n = 30 \) and 10000 samples).
We choose \( H_1 \) (and reject \( H_0 \)) if we obtain an unexpectedly large positive or large negative value for \( T \), otherwise we accept \( H_0 \).

The decision rule or test is

\[
\text{Reject } H_0 \iff |T| = \left| \frac{\hat{\beta} - \beta_0}{\frac{s^2}{\sum_{i=1}^{n} (x_i - \bar{x})^2}} \right| > c
\]

How do we choose \( c \)?
Remember we can make two errors in the decision

- False rejection of $H_0$ or Type I error
- False acceptance of $H_0$ or Type II error

As in coin tossing example we cannot make both small.

Usual approach: Fix probability of Type I error at small value. This gives $c$.

Now if $H_0$ is correct then $T$ has a $t$ distribution with $n - 2$ df.

Then

$$\alpha = \Pr(\text{Type I error}) = \Pr(|T| > c) = 1 - \Pr(|T| < c)$$

From confidence interval we know that for $\alpha = .05$ and $n = 30$, we have $c = 2.048$.

Hence we reject $H_0 : \beta = \beta_0$ if and only if

$$|T| = \frac{|\hat{\beta} - \beta_0|}{s} \sqrt{\frac{n}{\sum_{i=1}^{n} (X_i - \bar{X})^2}} > 2.048$$

Note if $\beta_0 = 0$ we just look at the ratio of the OLS estimate and its standard error. If this is greater than about 2, we reject $H_0 : \beta = 0$

If the alternative is $H_1 : \beta > \beta_0$, we may use a different decision rule. In that case we reject only if we get an unexpectedly large positive value for $T$.

We reject $H_0 : \beta = 0$ if and only if

$$T = \frac{|\hat{\beta} - \beta_0|}{s} \sqrt{\frac{n}{\sum_{i=1}^{n} (X_i - \bar{X})^2}} > c$$

Now $c = 1.701$ for a 5% probability of a Type I error. This is a one-sided test.
Prediction

After we have computed the OLS estimates $\hat{\alpha}, \hat{\beta}$ using the data $Y_i, X_i, i = 1, \ldots, n$ we can use the estimated regression model to predict $Y$ for values of $X$ that have not been observed.

Example: prediction of sales price of house that is new on market.

If that house has living area $X_{n+1}$, then we predict

$$\hat{Y}_{n+1} = \hat{\alpha} + \hat{\beta} X_{n+1}$$

How close is this to $Y_{n+1}$?

Prediction error

$$Y_{n+1} - \hat{Y}_{n+1} = \alpha + \beta X_{n+1} + u_{n+1} - \hat{\alpha} - \hat{\beta} X_{n+1} =$$

$$(\alpha - \hat{\alpha}) + (\beta - \hat{\beta}) X_{n+1} + u_{n+1}$$

Estimation error Unknown future omitted vars

Estimation error is small if we have e.g. many observations ($n$ large)

We cannot reduce $u_{n+1}$

We can obtain a prediction interval for $Y_{n+1}$. This is a confidence interval. A 95% interval is if $n = 30$

$$\hat{Y}_{n+1} - 2.048 s_{n+1} < Y_{n+1} < \hat{Y}_{n+1} + 2.048 s_{n+1}$$

(note we use the t distribution with $n - 2$ df) with

$$s_{n+1}^2 = s^2 \left( 1 + \frac{1}{n} + \frac{(X_{n+1} - \bar{X})^2}{\sum_{i=1}^{n} (X_i - \bar{X})^2} \right)$$

Note that if $n$ is large this is close to $s^2$. Hence a rule of thumb for the prediction interval is

$$\hat{Y}_{n+1} - 2s < Y_{n+1} < \hat{Y}_{n+1} + 2s$$